

П.В. НОВИЦКИЙ
И.А. ЗОГРАФ

ОЦЕНКА ПОГРЕШНОСТЕЙ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ



ЭНЕРГОАТОМИЗДАТ

ОГЛАВЛЕНИЕ

Предисловие	
ГЛАВА ПЕРВАЯ. ОБЩИЕ ВОПРОСЫ ТЕОРИИ ПОГРЕШНОСТЕЙ ПРИБОРОВ И ИЗМЕРЕНИЙ	6
ГЛАВА ВТОРАЯ. МЕТОДЫ ВЕРОЯТНОСТНОГО ОПИСАНИЯ ПОГРЕШНОСТЕЙ СРЕДСТВ И РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ	42
ГЛАВА ТРЕТЬЯ. МЕТОДЫ РАСЧЕТНОГО СУММИРОВАНИЯ СОСТАВЛЯЮЩИХ РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ	86
ГЛАВА ЧЕТВЕРТАЯ. МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МНОГОКРАТНЫХ ОТСЧЕТОВ	139
ГЛАВА ПЯТАЯ. МЕТОДЫ ИДЕНТИФИКАЦИИ ФОРМЫ ЗАКОНА РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ	169
ГЛАВА ШЕСТАЯ. МЕТОДЫ ОБРАБОТКИ И ОЦЕНКИ ПОГРЕШНОСТЕЙ ПРИ ОДНОФАКТОРНОМ ЭКСПЕРИМЕНТЕ	211
ГЛАВА СЕДЬМАЯ. МЕТОДЫ ОБРАБОТКИ И ОЦЕНКИ ПОГРЕШНОСТЕЙ ПРИ МНОГОФАКТОРНОМ ЭКСПЕРИМЕНТЕ	243
ГЛАВА ВОСЬМАЯ. ПУТИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ИЗМЕРИТЕЛЬНОГО ЭКСПЕРИМЕНТА	268
Заключение	297
Приложение	299
Список литературы	300

ББК 31.22
Н73
УДК 621.317.08

Редактор *В. Н. Миханкова*

Новицкий П. В., Зограф И. А.

Н73 Оценка погрешностей результатов измерений. — 2-е изд., перераб. и доп. — Л.: Энергоатомиздат. Ленингр. отд-ние, 1991. — 304 с.: ил.
ISBN 5-283-04513-7

Изложены практические методы метрологической оценки результатов измерений (прямых, косвенных, однофакторных и многофакторных) по паспортным данным используемых средств измерений, путем расчетного суммирования составляющих или статистической обработки многократных отсчетов, методы идентификации формы распределений на ЭВМ и пути повышения эффективности эксперимента. Первое издание вышло в 1985 г. Второе издание переработано и дополнено.

Для инженерно-технических работников и исследователей всех специальностей, может быть полезна студентам вузов и слушателям факультетов повышения квалификации.

Н $\frac{2202030000-100}{051(01)-91}$ 45—91

ББК 31.22

ISBN 5-283-04513-7

© Энергоатомиздат, 1985

© П. В. Новицкий, И. А. Зограф, изменения и дополнения, 1991

В настоящей книге систематизированно излагаются практические методы метрологической оценки результатов измерений, начиная с простейшей оценки погрешности результата по паспортным данным используемых средств измерений и кончая обстоятельной статистической обработкой и оценкой погрешностей сложных многофакторных экспериментов.

Пособие предназначено для широкого круга специалистов, занимающихся измерениями в различных областях и не имеющих часто специального образования по метрологии и математической статистике. В связи с этим авторы ставили своей целью дать наиболее доступное изложение материала, не отказываясь, однако, от информации читателя, хотя бы в самой популярной форме, и о более сложных современных проблемах в этой области. Поэтому и подготовленный в вопросах метрологии и математической статистики специалист может найти в этой книге новые для себя мысли и сведения.

Потребность в такой книге для широкого круга исследователей вызвана тем, что измерения — это единственный способ получения количественной информации о величинах, характеризующих те или иные явления и процессы (даже не обязательно физические, а иногда биологические, экономические и т. д.). Поэтому осуществление любых технологических процессов (будь то в машиностроении, нефтедобыче, ирригации, фармакологии или на спортивных соревнованиях), проведение научных исследований в любой отрасли знания всегда сопровождается планированием, постановкой и осуществлением измерений.

В результате этого человечество затрачивает на проведение измерений весьма существенную долю своего труда. Часто говорят о важности экономии цветных металлов или электроэнергии. Однако если суммировать заработную плату всех людей, вынужденных заниматься проведением измерений во всех НИИ, КБ, предприятиях машиностроения, химии, металлургии, космонавтики, опытных сельскохозяйственных станциях, медицинских учреждениях и т. д., то результат, возможно, превысит стоимость всего производства цветных металлов или электроэнергии. Поэтому оценка точности производимых измерений, т. е. качества

этой информационной продукции многомиллионной армии людей, имеет как экономическое, так и непосредственно прикладное значение.

Однако обеспечение высокой точности измерений и ее оценка оказываются весьма не простой задачей. Правда, для ее решения существуют соответствующие метрологические учреждения: Государственный комитет по стандартам с многочисленными научно-исследовательскими центрами и институтами, целая сеть учреждений государственной и ведомственной поверочной службы, метрологические отделы и лаборатории предприятий и т. д. Но эти службы не могут дойти до оценки погрешности каждого проводимого где-то измерения. Они обеспечивают лишь возможность, основу такой оценки. Для этого они создают и хранят эталоны, производят испытания головных образцов средств измерений (СИ), дают разрешение на их выпуск и т. п., следят за тем, чтобы каждое выпущенное СИ имело паспорт, где указывались бы нормируемые для него погрешности, проводят периодическую поверку всех СИ для установления соответствия их погрешностей нормируемым значениям и т. д. Что же касается оценки точности результатов каждого из проводимых измерений, то это приходится делать самим сотрудникам, проводящим измерения, или метрологическим лабораториям предприятий, где такие лаборатории существуют. Поэтому знание практических методов оценки погрешностей результатов измерений необходимо специалистам всех специальностей во всех технологических процессах и научных исследованиях.

Интерес к более точной оценке погрешностей результатов измерений особо возрос в последние годы, так как современное дорогое, быстродействующее и совершенное оборудование, включая ЭВМ, измерительные информационные системы (ИИС), измерительно-вычислительные комплексы (ИВК), требует наиболее рационального его использования, что невозможно осуществить без должной оценки возникающих погрешностей. Без такой оценки невозможно и оптимальное или рациональное планирование измерений.

Дело это, однако, достаточно сложное. Лишь только при самых простейших расчетах погрешностей результатов измерений по паспортным данным СИ оказывается возможным обойтись без использования теории вероятностей и математической статистики. В более же сложных случаях без этого обойтись нельзя, так как погрешность измерений приходится рассматривать как случайную величину, а следовательно, достоверная оценка погрешностей может быть выполнена лишь на основе теории вероятностей.

Исходя из этого предлагаемая читателю книга содержит описание практических методов статистической оценки погрешностей результатов прямых, косвенных, однофакторных и многофакторных измерений, а также рассмотрение вопросов повыше-

ния эффективности эксперимента путем его рационального планирования.

Авторы признательны И. Б. Челпанову, И. А. Назарову и В. Г. Кноррингу за замечания и предложения, которые были учтены при подготовке к печати первого издания 1985 г., а также принявшим участие в написании § 2-2 и 2-5 — И. Б. Челпанову, § 2-4, 3-8, 4-3, 7-3, 8-10 — И. А. Назарову и § 6-6 — И. Л. Мисюченко.

На первое издание книги издательство получило десятки отзывов читателей с просьбой переиздать ее. Авторы благодарны всем читателям, приславшим свои отзывы и замечания на первое издание, и ждут отклика на второе издание по адресу: 191065, Ленинград, Д-65, Марсово поле, 1, Ленинградское отделение Энергоатомиздата.

Авторы

ОБЩИЕ ВОПРОСЫ ТЕОРИИ ПОГРЕШНОСТЕЙ ПРИБОРОВ И ИЗМЕРЕНИЙ

1-1. РАЗНОВИДНОСТИ ПОГРЕШНОСТЕЙ

Качество средств и результатов измерений принято характеризовать указанием их погрешностей. Но так как характер проявления и причины возникновения погрешностей как средств, так и результатов измерений весьма разнообразны, то в практике установилось деление погрешностей на разновидности, за каждой из которых закреплено определенное наименование. Этих наименований около 30, и тот, кто так или иначе связан с измерениями, должен четко усвоить эту терминологию.

Погрешность средств измерений и погрешность результата измерения. Погрешность результата измерения — это число, указывающее возможные границы неопределенности полученного значения измеряемой величины. Погрешность же прибора — это его определенное свойство, для описания которого приходится использовать соответствующие правила. Поэтому полагать, что, воспользовавшись, например, вольтметром класса точности 1,0, т. е. имеющим предел приведенной погрешности, равный 1%, мы получаем и результат измерения с погрешностью, равной 1%, — грубейшая ошибка. Далее все эти вопросы будут пояснены, пока же подчеркнем лишь то, что погрешности средств измерений и погрешности результатов измерений — понятия не идентичные.

Исторически часть наименований разновидностей погрешностей закрепились за погрешностями средств измерений, другая — за погрешностями результатов измерения, а некоторые применяются по отношению и к тем, и к другим. Поэтому, рассматривая далее эти термины, будем обращать внимание на области их применения и отмечать также те случаи, когда один и тот же термин в разных областях имеет несовпадающие значения.

Инструментальные и методические погрешности. Инструментальными (приборными или аппаратными) погрешностями средств измерений называются такие, которые принадлежат данному средству измерений, могут быть определены при его испытаниях и занесены в его паспорт.

Однако, кроме инструментальных погрешностей, при измерениях возникают еще и такие погрешности, которые не могут быть приписаны данному прибору, не могут быть указаны в его пас-

порте и называются *методическими*, т. е. связанными не с самим прибором, а с методом проведения измерений.

Очень часто причиной возникновения методической погрешности является то, что, организуя измерения, нередко измеряют или вынуждены измерять не ту величину, которая в принципе должна быть измерена, а некоторую другую, близкую, но не равную ей. Этот прием замены того, что действительно *п о д л е ж и т* измерению, тем, что несколько отличается от нужного, но проще осуществляется, очень широко используется и при разработке СИ, и в практике организации измерений. Он позволяет создавать наиболее простые, надежные и универсальные приборы и методы измерения. Но когда этот метод уже воплощен в приборе, то погрешности прибора должны быть изучены, определены и внесены в его паспорт. С этого момента вне зависимости от причин возникновения погрешности для пользователя все они являются инструментальными.

Наглядный пример этого — выбор метода построения прибора для измерения запаса горючего в баке автомобиля. Ясно, что суммарная энергия, запасенная в топливе, определяется его массой (а не объемом) и для ее измерения нужны весы. Но совмещение топливного бака с весовым механизмом резко усложняет конструкцию. Поэтому разработчик заменяет весы простейшим поплавковым уровнемером, хотя уровень топлива зависит и от наклона бака, и от температуры и лишь весьма приближенно отражает массу топлива.

Для разработчика погрешности прибора, вызванные наклоном и температурой, представляются «методическими», т. е. обусловленными выбранным методом. Но для пользователя — это инструментальные погрешности данного прибора и они, безусловно, должны указываться в его паспорте.

Если же погрешности вызваны тем, что пользователь сам измеряет не ту величину, которая в действительности его интересует, и вследствие этого возникают погрешности, которые не могли быть изучены разработчиком и внесены в паспорт прибора, то ответственность за установление размера этих уже чисто методических погрешностей целиком лежит на пользователе средств измерений.

Примером такой методической погрешности может служить погрешность, возникающая при измерении напряжения вольтметром. Вследствие шунтирования входным сопротивлением вольтметра того участка цепи, на котором измеряется напряжение, оно оказывается меньшим, чем было до присоединения вольтметра. Поэтому для одного и того же вольтметра, присоединяемого поочередно к разным участкам исследуемой цепи, эта погрешность различна: на низкоомных участках — ничтожна, а на высокоомных — может быть очень большой. Естественно, размер этой переменной погрешности не может быть указан в паспорте прибора и она является методической. Для расчета этой погрешности поль-

зователь должен при каждом конкретном измерении напряжения измерять еще и выходное сопротивление исследуемой цепи между точками, к которым присоединен вольтметр, т. е. производить дополнительное исследование о б ъ е к т а измерения.

Такие исследования совершенно необходимы, когда вообще трудно указать способ измерения, исключающий методическую погрешность. Пусть, например, измерению подлежит температура раскаленных болванок, поступающих из печи на прокатный стан. Спрашивается, где разместить датчик температуры (например, термопару): под болванкой, сбоку или над болванкой? Где бы мы его ни поместили, мы не измерим внутренней температуры тела болванки, т. е. будем иметь существенную методическую погрешность, так как измеряем не то, что нужно, а то, что проще.

Ясно, что подобные погрешности никак не могут быть указаны в паспорте СИ и, следовательно, являются не инструментальными, а методическими.

Относительно термина «методическая погрешность» необходимо отметить, что в последние годы, особенно в теории цифровых приборов, его начали употреблять совсем в другом смысле, а именно для обозначения погрешности квантования, вызванной валожением* в принцип действия прибора методом и не зависящей от разработчика и изготовителя. Погрешность квантования в цифровых приборах действительно обусловлена самим методом цифрового представления информации и является неизбежной. Поэтому разработчик с удовольствием украшает ее наименование дополнением «методическая», стараясь тем самым подчеркнуть, что ее размер не зависит от его усилий. Тем не менее это принципиально неправомерно и в паспорте прибора эта погрешность как согласно ГОСТ 8.009—72, так и ГОСТ 8.009—84, безусловно, должна указываться как инструментальная.

В этой связи имеет смысл привести определения этих понятий, данные во введении к книге Э. И. Цветкова [49]. Он пишет: «Все погрешности, обусловленные отклонением р е а л и з у е м ы х а п п а р а т у р н о преобразований от идеальных, относятся к так называемым аппаратным, или инструментальным, погрешностям». А погрешность квантования можно уменьшить только аппаратно — повышая разрядность АЦП.

«К методическим погрешностям относятся все погрешности, которые могут быть определены и количественно оценены с помощью формального описания (математической модели) измерительной п р о ц е д у р ы. Количественная оценка погрешностей и их характеристик при этом выполняется на о с н о в е р а с ч е т о в или имитационного моделирования».

Таким образом, отличительной особенностью методических погрешностей является то, что они могут быть определены лишь путем создания математической модели или имитационным моделированием и з м е р я е м о г о о б ъ е к т а и не могут быть

найденны сколь угодно тщательным исследованием лишь самого измерительного прибора.

Действительно, определить температуру в центре болванки по измеренной температуре вблизи ее поверхности можно лишь располагая математической моделью теплового поля болванки. Чтобы узнать напряжение на участке цепи, которое было там до подключения вольтметра, нужно располагать не только показаниями вольтметра и значением его входного сопротивления, но и измерить сопротивление между точками присоединения вольтметра, т. е. параметр исследуемого объекта.

Так как методические погрешности не могут быть указаны в паспортных данных СИ, а должны оцениваться самим экспериментатором, то он обязан при выполнении измерений четко различать фактически измеряемую им величину от подлежащей измерению.

Такая оценка достаточно сложна. Она требует постановки обстоятельного экспериментального метрологического исследования принятого метода измерений. Если это установившийся, неизменный в течение длительного времени метод, то в результате такого исследования его погрешности могут быть установлены и занесены в *паспорт* метода. Составление таких аттестационных паспортов погрешности стандартных методов измерений — одна из насущных задач современной отраслевой метрологии.

Основная и дополнительная погрешности СИ. Любой датчик, измерительный прибор или регистратор работают в сложных, изменяющихся во времени условиях. Это прежде всего обусловлено тем, что процесс измерения — это сложное, многогранное явление, характеризующееся множеством воздействующих на прибор (как со стороны объекта, так и внешней среды, источников питания и т. д.) отдельных факторов. Каждый из этих факторов (см. гл. 7, посвященную многофакторным экспериментам) может быть измерен в отдельности, но в реальных условиях прямых измерений действует на измерительный прибор или датчик совместно со всеми остальными факторами. Интересующий нас единственный фактор из всего множества воздействующих мы называем *измеряемой величиной*. Мы требуем от прибора или датчика, чтобы он выделил из всего множества действующих на него величин только ту, которую мы называли измеряемой, и «отстроился» от действия на него всех остальных величин, которые мы именуем *влияющими*, мешающими или помехами.

Естественно, что в этих условиях прибор наряду с чувствительностью к измеряемой величине неминуемо имеет некоторую чувствительность и к неизмеряемым, влияющим величинам. Прежде всего это температура, тряска и вибрации, напряжение источников питания прибора и объекта, коэффициент содержания гармоник питающих напряжений и т. п.

При аттестации или градуировке прибора в лабораторных условиях все значения влияющих величин могут поддерживаться

в узких пределах их измерения (например, температура $(20 \pm \pm 5)^\circ\text{C}$, напряжение питания $\pm 5\%$ номинального, коэффициент гармоник — не более 1% и т. д.). Такие оговоренные в технической документации условия поверки или градуировки принято называть *нормальными*, а погрешность прибора, возникающую в этих условиях, — *основной погрешностью*.

В эксплуатационных условиях, при установке прибора, например, на самолет, ему придется работать при изменении температуры от -60 до $+60^\circ\text{C}$, давления от 1000 до 100 гПа, напряжения питания — на $\pm 20\%$, коэффициента гармоник — от 1 до 10% и т. д., что приведет к появлению погрешностей, естественно, больших, чем в нормальных (лабораторных) условиях или условиях поверки.

Изменения показаний вследствие отклонения условий эксплуатации от нормальных называются *дополнительными погрешностями* и нормируются указанием *коэффициентов влияния* изменения отдельных влияющих величин на изменение показаний в виде Ψ_Θ , %/10 К; Ψ_U , %/(10% $\Delta U/U$) и т. д. Хотя фактически эти функции влияния влияющих факторов, как правило, нелинейны, для простоты вычислений их приближенно считают линейными и возникающие дополнительные погрешности определяют как $\gamma_{\text{доп}} = \Psi \Delta\Theta$, где Ψ — коэффициент влияния, а $\Delta\Theta$ — отклонение от нормальных условий.

Погрешность прибора в реальных условиях его эксплуатации называется *эксплуатационной* и складывается из его основной погрешности и всех дополнительных и может быть, естественно, много больше его основной погрешности (суммирование погрешностей рассмотрено в гл. 3). Таким образом, деление погрешностей на основную и дополнительные является чисто условным и оговаривается в технической документации на каждое средство измерений.

Статические и динамические погрешности, присущие как средствам, так и методам измерений, различают по их зависимости от скорости изменения измеряемой величины во времени. Погрешности, не зависящие от этой скорости, называются *статическими*. Погрешности же, отсутствующие, когда эта скорость близка к нулю, и возрастающие при ее отклонении от нуля, называются *динамическими*. Таким образом, динамические погрешности являются одной из разновидностей дополнительных погрешностей, вызываемой влияющей величиной в виде скорости изменения во времени самой измеряемой величины. Однако вследствие специфических методов нормирования и расчета динамических погрешностей они будут рассмотрены в § 1-5.

Систематические, прогрессирующие и случайные погрешности. *Систематическими* называются погрешности, не изменяющиеся с течением времени или являющиеся не изменяющимися во времени функциями определенных параметров. Основным отличительный признак систематических погрешностей состоит в том, что

они могут быть предсказаны и благодаря этому почти полностью устранены введением соответствующих поправок.

Особая опасность постоянных систематических погрешностей заключается в том, что их присутствие чрезвычайно трудно обнаружить. В отличие от случайных, прогрессирующих или являющихся функциями определенных параметров погрешностей постоянные систематические погрешности внешне себя никак не проявляют и могут долгое время оставаться незамеченными. Единственный способ их обнаружения состоит в *поверке* прибора путем повторной *аттестации* по образцовым мерам или сигналам.

Примером систематических погрешностей второго вида служит большинство дополнительных погрешностей, являющихся не изменяющимися во времени функциями вызывающих их влияющих величин (температур, частот, напряжений и т. п.). Эти погрешности благодаря постоянству во времени функций влияния также могут быть предсказаны и скорректированы введением дополнительных корректирующих преобразователей, воспринимающих влияющую величину и вводящих соответствующую поправку в результат измерения.

Прогрессирующими (или дрейфовыми) называются непредсказуемые погрешности, медленно изменяющиеся во времени. Эти погрешности, как правило, вызываются процессами старения тех или иных деталей аппаратуры (разрядкой источников питания, старением резисторов, конденсаторов, деформацией механических деталей, усадкой бумажной ленты в самопишущих приборах и т. п.). Особенностью прогрессирующих погрешностей является то, что они могут быть скорректированы введением поправки лишь в данный момент времени, а далее вновь непредсказуемо возрастают. Поэтому в отличие от систематических погрешностей, которые могут быть скорректированы поправкой, найденной один раз на весь срок службы прибора, прогрессирующие погрешности требуют непрерывного повторения коррекции и тем более частой, чем меньше должно быть их остаточное значение. Другая особенность прогрессирующих погрешностей состоит в том, что их изменение во времени представляет собой нестационарный случайный процесс и поэтому в рамках хорошо разработанной теории стационарных случайных процессов они могут быть описаны лишь с оговорками.

Случайными погрешностями называют непредсказуемые ни по знаку, ни по размеру (либо недостаточно изученные) погрешности. Они определяются совокупностью причин, трудно поддающихся анализу. Присутствие случайных погрешностей (в отличие от систематических) легко обнаруживается при повторных измерениях в виде некоторого разброса получаемых результатов. Таким образом, главной отличительной чертой случайных погрешностей является их *непредсказуемость* от одного отсчета к другому. Поэтому описание случайных погрешностей

может быть осуществлено только на основе теории вероятностей и математической статистики.

Здесь, однако, необходимы две существенные оговорки. Во-первых, применение методов математической статистики к обработке результатов измерений правомерно лишь в предположении о независимости между собой отдельных получаемых отсчетов. И, во-вторых, большинство приводимых далее формул теории вероятностей правомерно только для непрерывных распределений, в то время как распределения погрешностей вследствие неизбежного квантования отсчетов, строго говоря, всегда дискретны. Таким образом, условия непрерывности и независимости для случайных погрешностей соблюдаются лишь приближенно, а иногда могут и не соблюдаться, т. е. в математике под термином «непрерывная случайная величина» понимается существенно более узкое, ограниченное рядом условий понятие, чем под термином «случайная погрешность» в измерительной технике.

Тем не менее, так как большинство составляющих погрешностей средств и результатов измерений являются случайными погрешностями, то единственно возможным разработанным способом их описания является использование положений теории вероятностей и ее дальнейшего развития применительно к процессам передачи информации в виде теории информации, а для обработки получаемых экспериментальных данных, содержащих случайные погрешности, — методов математической статистики. Поэтому именно эта группа фундаментальных разделов математики является основой для развития современной теории оценок погрешностей средств, процессов и результатов измерений.

С учетом указанных ограничений процесс проявления случайных погрешностей средств и результатов измерений за вычетом систематических и прогрессирующих погрешностей обычно может рассматриваться как центрированный стационарный случайный процесс с использованием для его описания хорошо разработанной в математике теории статистически независимых случайных величин и стационарных случайных процессов.

В заключение описанного деления погрешностей средств и результатов измерений на случайную, прогрессирующую и систематическую составляющие необходимо обратить внимание читателя на то, что такое деление является весьма упрощенным приемом их анализа. Поэтому всегда следует помнить, что в реальной действительности эти составляющие погрешности проявляются совместно и образуют единый нестационарный случайный процесс.

Погрешности адекватности, градуировки и воспроизводимости средств измерений. Присутствие погрешностей приводит к тому, что характеристики СИ (датчиков, приборов, каналов ИИС и ИВК) оказываются неоднозначными. При экспериментальном их определении, т. е. градуировке СИ, получают лишь ряд точек. По этой полосе точек проводят на графике некоторую плавную среднюю кривую, которую и принимают за характеристику СИ,

хотя некоторые экспериментальные точки от нее и отклоняются. Для наименования этих отклонений используется ряд терминов.

Систематически наблюдающиеся отклонения от выбранной в качестве характеристики плавной кривой в общем случае называются погрешностью *адекватности* выбранной функциональной зависимости (прямой линии, параболы, экспоненты и т. д.) фактической характеристике СИ. Если в качестве характеристики выбрана прямая, то погрешность ее адекватности называют погрешностью *линейности* СИ. Если погрешность адекватности меняет свой знак в зависимости от направления предшествующего отсчету изменения входной величины, то такую погрешность прибора или преобразователя называют погрешностью *от гистерезиса* или *вариацией* СИ.

Погрешности СИ могут быть обусловлены недостаточной точностью образцовых СИ, используемых при градуировке. Это может быть вызвано систематической погрешностью образцовых СИ, тогда все отметки шкалы градуируемого СИ будут ошибочными. Это очень опасный случай, так как вся партия приборов, градуированных по неисправному образцовому СИ, есть брак. Поэтому законодательная метрология требует, чтобы погрешность образцовых средств была не более $\frac{1}{3}$ погрешности градуируемых. Если же причиной погрешности градуировки являются случайные погрешности градуируемого и образцового средств, а также самого метода градуировки, то при градуировке будут возникать случайные, неповторяющиеся от отметки к отметке погрешности в их расположении. Казалось бы, что возникающая у градуируемого СИ вследствие этого погрешность является строго систематической, так как ошибочно нанесенная отметка остается на этом месте на весь срок службы СИ, и поэтому может быть устранена внесением в результаты измерений соответствующих поправок. Это действительно так, если прибору придается график поправок для каждой отметки его шкалы. Но если такой график отсутствует или не используется, то возникающая в результатах измерений погрешность является случайной, так как на одной отметке она положительна, на другой — отрицательна, а на третьей — равна нулю. Поэтому для шкалы в целом она является случайной и как таковая должна указываться в паспорте СИ и использоваться при измерениях. Все эти погрешности, как систематические, так и случайные, объединяются общим термином *погрешность градуировки*.

Понятия полосы погрешностей, реальной и номинальной характеристик СИ. Все перечисленные причины появления погрешностей приводят к тому, что многократно снятые характеристики прибора или серии однотипных приборов занимают на графике некоторую полосу. Поэтому в теории измерений используется понятие *полосы неопределенности*, или *полосы погрешностей* данного типа или данного экземпляра прибора, датчика или измерительного канала ИИС и ИВК. Некоторая детерминированная

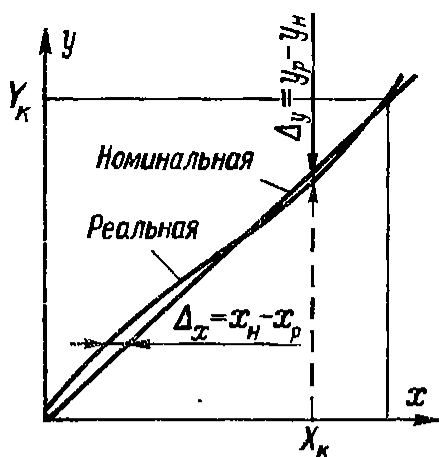


Рис. 1-1

средняя линия этой полосы принимается за *номинальную характеристику* приборов этого типа, указывается в паспорте и используется для определения результатов измерения.

Отсюда погрешность данного измерительного преобразователя, датчика, прибора или канала ИИС есть разность между *реальной* и *номинальной* его характеристиками, т. е. не число, а функция измеряемой величины.

Абсолютная, относительная и приведенная погрешности СИ. Разности между реальной и номинальной характеристиками (рис. 1-1), найденные при заданном значении x в виде $\Delta_y = y_p - y_n$ или при заданном значении y в виде $\Delta_x = x_n - x_p$, суть *абсолютные* погрешности, так как они выражаются в единицах величин x или y . Знак абсолютной погрешности принимается положительным, если реальная характеристика проходит выше номинальной.

Абсолютная погрешность не может, однако, сама по себе служить показателем точности измерений, так как одно и то же ее значение, например $\Delta_x = 0,05$ мм при $x = 100$ мм, соответствует достаточно высокой точности, а при $x = 1$ мм — низкой. Поэтому для характеристики точности результатов измерения вводится понятие *относительной* погрешности $\gamma = \Delta_x/x \approx \Delta_y/y$, выражаемой в относительных единицах или в процентах (x и y — текущие значения входной и выходной величин прибора или преобразователя).

Но эта очень наглядная характеристика точности *результата* измерения не годится для нормирования погрешности СИ, так как при различных значениях x принимает различные значения вплоть до $\gamma = \infty$ при $x = 0$. Поэтому для указания и нормирования погрешности СИ используется еще одна разновидность погрешности, а именно так называемая *приведенная* погрешность. Она определяется как отношение абсолютной погрешности, выраженной в единицах входной Δ_x или выходной Δ_y величин, к *протяженности диапазона* изменения соответственно входной X_K или выходной Y_K величины прибора или преобразователя и выражается в относительных единицах или в процентах, т. е.

$$\gamma_{пр} = \Delta_x/X_K = \Delta_y/Y_K.$$

Ее основное отличие от относительной погрешности состоит в том, что Δ_x или Δ_y относится не к переменной текущей величине x или y , а к постоянной величине протяженности диапазона.

Приведенная погрешность удобна тем, что для многих многопредельных СИ она имеет одно и то же значение как для всех

точек каждого поддиапазона, так и для всех его поддиапазонов, т. е. ее очень удобно использовать для нормирования свойств СИ.

Понятия абсолютной, относительной и приведенной погрешностей существующими стандартами установлены только для СИ, но их удобно использовать и при характеристике погрешностей ре-

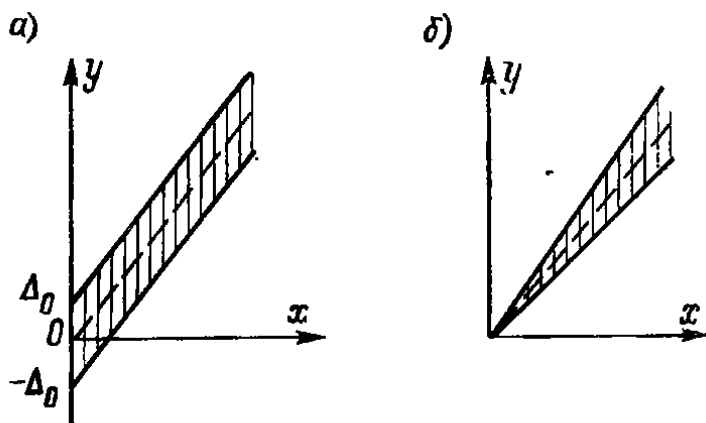


Рис. 1-2

зультатов измерения. В § 6-7 будет показано, что широко используемый в математической статистике показатель тесноты группирования экспериментальных точек вокруг определяемой функциональной зависимости в виде коэффициента ρ множественной корреляции по своему смыслу есть полный аналог понятия приведенной погрешности лишь с той разницей, что он отсчитывается «с другой стороны». Поэтому сумма ρ^2 и γ^2 равна единице. Правда, из-за несколько разного определения этих понятий погрешности это соотношение имеет вид $\rho^2 + 3\gamma^2 = 1$, т. е. $\rho^2 = 1 - 3\gamma^2$ или $\gamma^2 = (1 - \rho^2)/3$ [см. формулы (6-13) и (6-14) и табл. 7-1].

Аддитивные и мультипликативные погрешности. Эти термины служат для описания формы границ полосы погрешностей СИ. При поверке или градуировке СИ (будь то прибор, датчик или канал ИИС) получают ряд значений входной величины x_i и ряд соответствующих им значений выходной величины y_i . Если эти данные нанести на график с координатами x и y , то полученные точки разместятся в границах некоторой полосы (рис. 1-2).

В том случае, когда эти точки лежат в границах линий, параллельных друг другу, как это показано на рис. 1-2, *а*, т. е. **а б с о л ю т н а я** погрешность СИ во всем его диапазоне измерений ограничена **п о с т о я н н ы м** (не зависящим от текущего значения x) **п р е д е л о м** $\pm \Delta_0$, то такая погрешность называется *аддитивной*, т. е. получаемой путем сложения, или *погрешностью нуля*. Это понятие одинаково применимо как к случайным, так и к систематическим погрешностям.

Примерами систематических аддитивных погрешностей являются погрешности от постороннего груза на чашке весов, от неточной установки прибора на нуль перед измерением, от термо-ЭДС в цепях постоянного тока и т. п. Для устранения таких погрешностей во многих СИ предусмотрено механическое или электрическое устройство для установки нуля (корректор нуля).

Примерами случайных аддитивных погрешностей являются погрешность от наводки переменной ЭДС на вход прибора, погрешности от тепловых шумов, от трения в опорах подвижной части измерительного механизма, от ненадежного контакта при

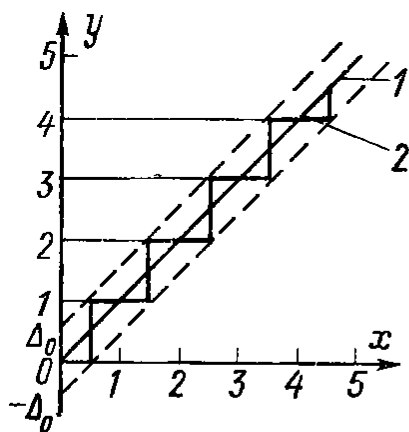


Рис. 1-3

измерении сопротивления, погрешность от воздействия порога трогания приборов с ручным или автоматическим уравниванием и т. п.

Если же положение границ полосы погрешностей имеет вид, показанный на рис. 1-2, б, т. е. ширина полосы возрастает пропорционально росту входной величины x , а при $x = 0$ также равна нулю, то такая погрешность называется *мультипликативной*, т. е. получаемой путем умножения, или *погрешностью чувствительности* вне зависимости от того, является ли погрешность случайной или систематической.

Причинами возникновения мультипликативных погрешностей могут быть: изменение коэффициента усиления усилителя, измерение жесткости мембраны датчика манометра или пружинки прибора, изменение опорного напряжения в цифровом вольтметре и т. д.

Погрешность квантования. Это специфическая разновидность погрешности, возникающая в цифровых приборах и дискретных преобразователях. При плавном изменении входной величины x , например напряжения в пределах от 0 до 5 мВ, цифровой вольтметр с пределом 1000 мВ не может дать других показаний, кроме дискретных значений 0—1—2—3—4 и 5 мВ. Поэтому при возрастании x от 0 до 0,5 мВ прибор, если он хорошо отрегулирован, продолжает показывать $x = 0$. При превышении значения 0,5 мВ прибор дает показание $x = 1$ и сохраняет его до $x = 1,5$ мВ и т. д. Поэтому, хотя его номинальной характеристикой мы считаем прямую 1 (рис. 1-3), его реальная характеристика представляет собой ступенчатую кривую 2. Текущая разность номинальной 1 и реальной 2 характеристик цифрового прибора и составляет погрешность квантования. Границы полосы погрешности квантования показаны на рис. 1-3 штриховыми прямыми, и полоса сохраняет на всем протяжении постоянную ширину, т. е. по форме аналогична полосе погрешностей, представленной на рис. 1-2, а.

Вследствие того, что измеряемая величина x случайным образом может принимать любые промежуточные значения, погрешность квантования также случайным образом принимает значения в интервале от $+\Delta_0$ до $-\Delta_0$. Поэтому погрешность квантования является инструментальной случайной аддитивной статической погрешностью, так как не зависит ни от текущего значения результата измерения величины x , ни от скорости изменения x во времени.

1-2. МЕТОДЫ НОРМИРОВАНИЯ ПОГРЕШНОСТЕЙ СРЕДСТВ ИЗМЕРЕНИЙ

Различные СИ (измерительные приборы и преобразователи, датчики, каналы ИИС и ИВК) обладают погрешностями, харак-

тер проявления которых может быть существенно различным: у одних погрешность практически аддитивная, у других — и аддитивная, и мультипликативная составляющие, у третьих зависимость погрешности от измеряемой величины оказывается еще более сложной. У каждого конкретного СИ имеется случайная и систематическая составляющие погрешности, причем их соотношение также может быть различным. Кроме того, условия работы даже однотипных СИ могут быть различными.

Для того чтобы ориентироваться в метрологических свойствах конкретного СИ, чтобы заранее оценить погрешность, которую внесет данное СИ в конкретный результат, пользуются так называемыми *нормированными* значениями погрешности. Под нормированным значением понимаются погрешности, являющиеся *предельными* для данного типа СИ. При этом как систематическая, так и случайная составляющие погрешности отдельных экземпляров СИ одного и того же типа могут различаться, однако в целом для этого типа СИ погрешности не превосходят гарантированного значения. Таким образом нормируется основная и дополнительная погрешности. Именно эти границы основной погрешности, а также коэффициентов влияния и заносятся в паспорт каждого экземпляра СИ.

Правила, согласно которым назначаются эти границы, значения погрешностей и форма записи, иными словами вся процедура нормирования погрешности средств измерений, основывается на системе стандартов, обеспечивающих единство измерений.

Класс точности средств измерений. Это характеристика, определяющая гарантированные границы значений основных и дополнительных погрешностей, а также другие свойства средств измерений, влияющих на точность. Соответствие погрешности СИ приписанному им классу точности во время эксплуатации проверяется при периодических поверках. Если погрешность оказывается меньше нормированных значений, то СИ продолжает эксплуатироваться, если нет, то подлежит ремонту и регулировке.

Основные способы установления пределов допускаемых погрешностей и обозначения классов точности средств измерений установлены ГОСТ 8.401—80. Основная погрешность СИ нормируется четырьмя различными способами. Чтобы четко уяснить себе эти различия и грамотно использовать нормируемые значения при расчете погрешностей результатов измерения, необходимо рассмотреть характер изменения относительной и абсолютной погрешности СИ в диапазоне значений измеряемой величины и обусловленные этим положения стандартов, регламентирующих нормирование погрешностей средств измерений.

Основное различие в способах нормирования обусловлено разным соотношением аддитивной и мультипликативной составляющих в погрешности тех или иных СИ.

При чисто мультипликативной полосе погрешностей СИ (рис. 1-2, б) абсолютная погрешность $\Delta(x)$ возрастает прямо

пропорционально текущему значению x измеряемой величины. Поэтому *относительная* погрешность, т. е. погрешность *чувствительности* такого преобразователя, $\gamma_s = \Delta(x)/x$ оказывается постоянной величиной при любом значении x и ее удобно использовать для нормирования погрешностей преобразователя и указания его класса точности.

Таким способом нормируются погрешности масштабных преобразователей (делителей напряжения, шунтов, измерительных трансформаторов тока и напряжения и т. п.). Их класс точности указывается в виде значения γ_s , выраженного в процентах. Граница относительной погрешности результата измерения $\gamma(x)$ в этом случае постоянна и при любом x просто равна значению γ_s , а абсолютная погрешность результата измерения рассчитывается по формуле $\Delta(x) = \gamma_s x$.

Если бы эти соотношения оставались справедливыми для всего диапазона возможных значений измеряемой величины x от 0 до X_k (X_k — предел диапазона измерений), то такие измерительные преобразователи были бы наиболее совершенными, так как они имели бы бесконечно широкий рабочий диапазон, т. е. обеспечивали бы с той же погрешностью измерение сколь угодно малых значений x .

Однако реально таких преобразователей не существует, так как невозможно создать преобразователь, полностью лишенный аддитивных погрешностей. Эти погрешности от шума, дрейфа, трения, наводок, вибраций и т. п. неизбежны в любых типах СИ. Поэтому для реальных СИ, погрешность которых нормируется лишь одним числом — погрешностью чувствительности γ_s , — всегда указываются границы рабочего диапазона, в которых такая оценка остается приближенно справедливой.

При чисто аддитивной полосе погрешностей (рис. 1-2, а) остается неизменной для любых значений x граница абсолютной погрешности нуля $\Delta(x) = \Delta_0 = \text{const}$. Но нормировать абсолютное значение Δ_0 неудобно, так как для многопредельных приборов оно будет различным для каждого поддиапазона и в паспорте прибора пришлось бы перечислять эти значения для всех поддиапазонов.

Поэтому нормируют не абсолютное Δ_0 , а приведенное значение этой погрешности: $\gamma_0 = \Delta_0/X_N$, где X_N — так называемое нормирующее значение измеряемой величины. Стандарт 8.401—80 определяет для приборов с равномерной или степенной шкалой, если нулевая отметка находится на краю или вне шкалы, нормирующее значение X_N равным верхнему пределу диапазона измерений. Если же нулевая отметка находится посередине шкалы, то X_N равно протяженности диапазона измерений [например, для амперметра со шкалой от -30 до $+60$ А значение $X_N = 60 - (-30) = 90$ А].

Значение приведенной погрешности γ_0 , выраженное в процентах, используется для обозначения класса точности таких СИ.

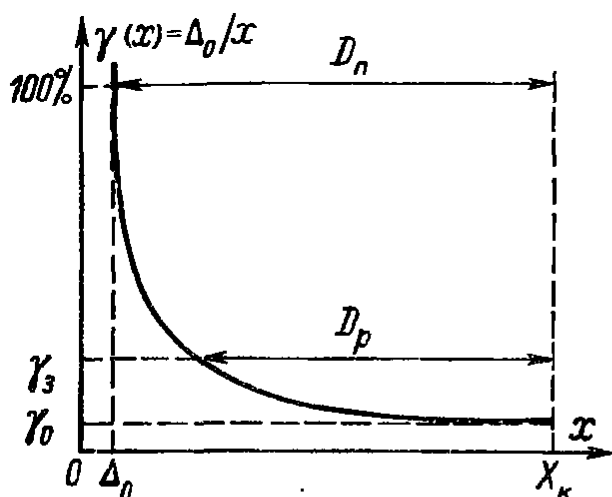


Рис. 1-4

Однако полагать, как уже указывалось, что вольтметр класса точности 1,0 обеспечивает во всем диапазоне измерений получение результатов с погрешностью $\pm 1\%$, — грубейшая ошибка. В действительности текущее значение относительной погрешности $\gamma(x) = \Delta_0/x$, т. е. растет обратно пропорционально x и изменяется по гиперболе (рис. 1-4). Таким образом, относительная погрешность $\gamma(x)$ равна классу точности прибора γ_0 лишь на последней отметке шкалы (при $x = X_k$). При $x = 0,1 X_k$ она в 10 раз больше γ_0 , а при дальнейшем уменьшении x стремится к бесконечности.

При уменьшении измеряемой величины x до значения абсолютной погрешности нуля Δ_0 относительная погрешность результата измерения достигает $\gamma(x) = \Delta_0/x = \Delta_0/\Delta_0 = 1 = 100\%$. Такое значение измеряемой величины, когда $x = \Delta_0$ и $\gamma(x) = 100\%$, называется *порогом чувствительности СИ*.

Отсюда *полный диапазон* D_n измеряемых величин для любого преобразователя ограничивается снизу порогом чувствительности, а сверху — пределом измерений. Так как в области малых значений x погрешность измерений очень велика, то *рабочий диапазон* D_r ограничивают снизу таким значением x , где относительная погрешность измерений $\gamma(x)$ не превосходит еще некоторого заранее заданного значения γ_3 , равного, например, 4, 10 или 20%. Таким образом, рабочий диапазон назначается достаточно произвольно (см. рис. 1-4) и составляет только некоторую часть полного диапазона СИ. В начальной же части шкалы измерения недопустимы, в чем и заключается отрицательное влияние аддитивной погрешности, не позволяющее использовать один и тот же преобразователь для измерения как больших, так и малых измеряемых величин.

При одновременном присутствии аддитивной и мультипликативной составляющих полоса погрешностей имеет трапецеидальную форму (рис. 1-5, а), а текущее значение абсолютной погреш-

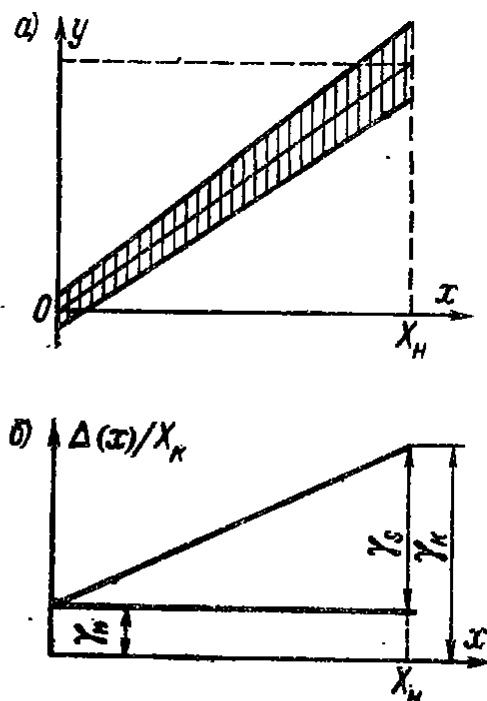


Рис. 1-5

ности $\Delta(x)$ в функции измеряемой величины x описывается соотношением

$$\Delta(x) = \Delta_0 + \gamma_s x, \quad (1-1)$$

где Δ_0 — аддитивная, а $\gamma_s x$ — мультипликативная составляющие абсолютной погрешности.

Если все члены уравнения (1-1) разделить на предел измерений X_K , то для приведенного значения погрешности получим

$$\gamma_{пр}(x) = \frac{\Delta(x)}{X_K} = \frac{\Delta_0}{X_K} + \gamma_s \frac{x}{X_K}. \quad (1-2)$$

Приведенное значение погрешности в начале диапазона (при $x = 0$) обозначим через $\Delta_0/X_K = \gamma_H$. Тогда соотношение (1-2) примет вид

$$\gamma_{пр}(x) = \gamma_H + \gamma_s \frac{x}{X_K}$$

и изобразится графиком рис. 1-5, б.

Таким образом, при наличии у СИ и аддитивной, и мультипликативной составляющих погрешности его приведенная погрешность линейно возрастает от $\gamma_H = \Delta_0/X_K$ в начале диапазона (при $x = 0$) до значения $\gamma_K = \gamma_H + \gamma_s$ в конце диапазона (при $x = X_K$).

Относительная погрешность результата измерения исходя из выражения (1-1) составляет

$$\gamma(x) = \frac{\Delta(x)}{x} = \frac{\Delta_0}{x} + \gamma_s = \gamma_s + \gamma_H \frac{X_K}{x}, \quad (1-3)$$

т. е. при $x = X_K$ она будет $\gamma(x) = \gamma_H + \gamma_s = \gamma_K$, а по мере уменьшения x возрастает до бесконечности. Но отличие $\gamma(x)$ от чисто аддитивной погрешности состоит в том, что заметное возрастание $\gamma(x)$ начинается тем позже, чем меньше γ_H по сравнению с γ_s .

Для иллюстрации этого явления на рис. 1-6 изображены кривые возрастания $\gamma(x)$ при уменьшении x при разных γ_s/γ_H для частного случая $\gamma_H + \gamma_s = \gamma_K = 2\% = \text{const.}$ Из этих кривых видно, как расширяется рабочий диапазон СИ по мере увеличения отношения γ_s/γ_H , т. е. уменьшения Δ_0 и приближения полосы погрешностей, приведенной на рис. 1-5, а, к чисто мультипликативной полосе (рис. 1-2, б).

Так, если заданное значение погрешности γ_s , ограничивающее нижнюю границу рабочего диапазона, принять в нашем примере $\gamma_s = 4\%$, то при $\gamma_s/\gamma_K = 0$, т. е. при $\gamma_s = 0$ и чисто аддитивной полосе погрешностей, представленной на рис. 1-2, а, рабочий диапазон будет двукратным (от 50 до 100%). При $\gamma_s/\gamma_H = 3$ он становится уже пятикратным (от 20 до 100%), а при $\gamma_s/\gamma_H = 20$ — становится двадцатикратным (от 5 до 100%). В последнем случае в интервале от 100 до 10% диапазона прибора погрешность результатов измерения почти не изменяется, т. е. боль-

шие и малые значения x измеряются с одной и той же относительной погрешностью.

Форму полосы погрешностей, которая изображена на рис. 1-5, а следовательно, и вытекающие из этого свойства имеют высокоточные потенциометры постоянного тока, цифровые вольтметры и другие высокоточные приборы. Формальным отличительным признаком для

них является то, что их класс точности согласно ГОСТ 8.401—80 обозначается не одним, а двумя числами, записываемыми через косую черту, т. е. в виде условной дроби γ_k/γ_n , в числителе которой указывается (в процентах) приведенная погрешность γ_k в конце диапазона измерений, а в знаменателе — приведенная погрешность γ_n в нуле диапазона.

Специальные формулы нормирования погрешностей средств измерений. Кроме перечисленных разновидностей нормирования погрешностей средств измерений (путем указания классов точности в виде γ_s , γ_0 , γ_k/γ_n), ГОСТ 8.401—80 разрешает использовать так называемые специальные формулы нормирования погрешностей. Дело заключается в том, что некоторые СИ не могут быть нормированы описанными выше способами, так как имеют более сложный вид полосы погрешностей, чем приведенный на рис. 1-2 и 1-5.

Это, например, цифровые частотомеры, погрешность которых зависит не только от измеряемой величины x , но и от времени T , отводимого для измерения этой частоты. Мосты для измерения сопротивлений отличаются тем, что имеют не только нижний порог чувствительности (т. е. такое малое измеряемое сопротивление, когда погрешность достигает 100%, например, из-за неопределенности контактных сопротивлений), но и верхний порог чувствительности (когда погрешность при измерении очень больших сопротивлений вновь достигает 100%, например, из-за приближения измеряемого сопротивления к сопротивлению изоляции между зажимами самого моста). В этом случае погрешность результатов измерения описывается трехчленной формулой вида

$$\varphi(x) = \Delta_0/x + \varphi_s + x/\Delta_\infty, \quad (1-4)$$

где Δ_∞ и Δ_0 — верхний и нижний пороги измеряемых сопротивлений.

Во всех подобных случаях необходимо внимательно изучать документацию на соответствующий прибор и пользоваться для вычисления погрешности результата измерения приводимыми в ней специальными формулами.

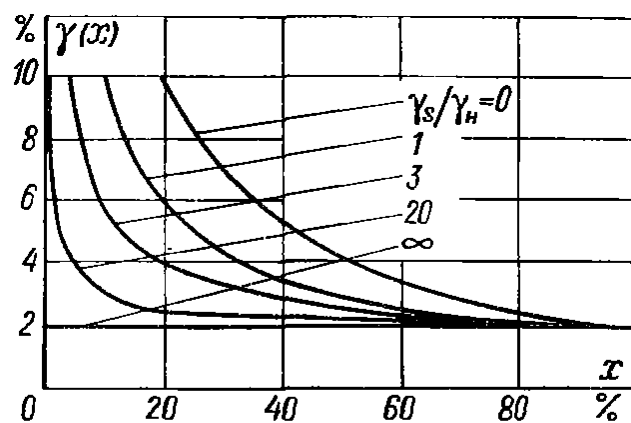


Рис. 1-6

Обозначения классов точности средств измерений. Согласно ГОСТ 8.401—80 для указания нормированных значений погрешности чувствительности γ_s , приведенной аддитивной погрешности γ_0 , приведенных погрешностей в начале γ_n и конце γ_k диапазона измерений не могут использоваться произвольные числа. Выраженные в процентах, они могут иметь значения 6—4—2,5—1,5—1,0—0,5—0,2—0,1—0,05—0,02—0,01—0,005—0,002—0,001 и т. д. Значение класса точности прибора маркируется на его шкале. Для того чтобы различить, какая из погрешностей обозначена в качестве класса точности, используются следующие условные обозначения.

Если класс точности прибора установлен по значению погрешности чувствительности γ_s , т. е. форма полосы погрешности условно принята чисто мультипликативной, обозначаемое на шкале значение класса точности обводится кружком. Например, (1,5)

обозначает, что $\gamma_s = 1,5\%$.

Если же полоса погрешностей принята аддитивной и прибор нормируется приведенной погрешностью нуля γ_0 (таких приборов большинство), то класс точности указывается без каких-либо подчеркиваний (например, просто 1,5).

Наконец, на приборах с резко неравномерной шкалой, например омметрах, класс точности прибора указывается в долях от длины шкалы и обозначается $\sqrt{1,5}$.

Обозначение класса точности в виде, например, 0,02/0,01 указывает, что погрешность прибора нормирована по двучленной формуле с $\gamma_n = 0,01\%$ и $\gamma_k = 0,02\%$.

Таким образом, обозначение класса прибора дает достаточно полную информацию для вычисления приближенной оценки погрешностей результатов измерения. Точность такой простейшей приближенной оценки будет в последующих главах (см. § 4-1) обсуждена более подробно.

Некоторые практические замечания по использованию нормируемых значений погрешности СИ для вычисления погрешности результатов измерения. Хотя ГОСТ 8.401—80 направлен на то, чтобы нормирование погрешностей СИ производилось единообразно, в измерительной практике такого единообразия пока еще нет, так как используется большое число хороших высокоточных приборов, которые были выпущены еще до введения этого стандарта, закупаются и широко используются приборы иностранного производства, нормированные, естественно, не в соответствии с ГОСТ 8.401—80, и т. д.

Например, погрешность высокоточных потенциометров постоянного тока нормируется чаще всего двучленной формулой (1-1), а класс точности прибора указывается в виде одного числа —

его относительной погрешности чувствительности. В этом случае указание класса точности в виде одного числа γ_s не является признаком того, что прибор не имеет аддитивной составляющей погрешности и потребитель обязан быть внимательным при расчете погрешностей результатов измерения, чтобы не допустить ошибки.

При нормировании погрешностей сложных СИ двучленной формулой (1-3) ГОСТ 8.401—80 предусматривает несколько иное ее написание, когда текущее значение относительной погрешности $\gamma(x)$ выражается не через значение аддитивной γ_n и мультипликативной γ_s составляющих предела допускаемых погрешностей, как в формуле (1-3), а через указываемые в обозначении класса точности приведенные погрешности в начале γ_n и в конце γ_k диапазона измерений. В этом случае, учитывая, что $\gamma_k = \gamma_n + \gamma_s$, соотношение (1-3) получает вид

$$\gamma(x) = \gamma_k + \gamma_n (X_k/x - 1). \quad (1-5)$$

Практически этим соотношением более удобно пользоваться для вычисления $\gamma(x)$ по известным x , X_k , γ_n и γ_k , чем соотношением (1-3).

У весьма широкодиапазонных приборов, например мостов для измерения сопротивлений, в их технической документации вместо указания коэффициентов трехчленной формулы (1-4) часто приводятся просто диапазоны, в которых погрешность результата измерения не превосходит указанного значения. Например, указывается, что относительная погрешность не превосходит:

0,5% в диапазоне от 10^3 до 10^4 Ом;

1% — от 5 до 10^5 Ом;

5% — от 0,5 до 10^6 Ом;

10% — от 0,2 до $2 \cdot 10^6$ Ом;

20% — от 0,1 до $4 \cdot 10^6$ Ом.

Как правило, эти данные достаточно точно соответствуют трехчленной формуле (1-4). Поэтому по ним можно определить коэффициенты Δ_0 , Δ_∞ и γ_s формулы (1-4) и использовать ее для аналитического определения $\gamma(x)$ при любом произвольном значении x . Так, например, приведенным выше данным соответствует $\Delta_0 = 0,02$ Ом, $\Delta_\infty = 20 \cdot 10^6$ Ом и $\gamma_s = 0,5\%$. Отсюда для любого x погрешность (в процентах)

$$\gamma(x) = 100 [0,02/x + 0,5/100 + x/(20 \cdot 10^6)].$$

1-3. РАСЧЕТ ОЦЕНКИ

ИНСТРУМЕНТАЛЬНОЙ СТАТИЧЕСКОЙ ПОГРЕШНОСТИ

РЕЗУЛЬТАТА ИЗМЕРЕНИЯ ПО ПАСПОРТНЫМ ДАННЫМ

ИСПОЛЬЗУЕМОГО СРЕДСТВА ИЗМЕРЕНИЙ

Результат измерения имеет ценность лишь тогда, когда можно оценить его интервал неопределенности, т. е. степень достоверности. Поэтому согласно ГОСТ 8.011—72 «Показатели точности

измерений и формы представления результатов измерений» сообщение о любом результате измерений обязательно должно сопровождаться указанием его погрешности.

Погрешность результата прямого однократного измерения зависит от многих факторов, но, в первую очередь, определяется, естественно, погрешностью используемых средств измерений. Поэтому в первом приближении погрешность результата измерения можно принять равной погрешности, которой в данной точке диапазона измерений характеризуется используемое средство измерений.

Так как погрешности средств измерений изменяются в диапазоне, то вычисление должно производиться по соответствующим формулам (см. § 1-2). Вычисляться должна как абсолютная, так и относительная погрешности результата измерения, так как первая из них нужна для округления результата и его правильной записи, а вторая — для однозначной сравнительной характеристики его точности.

Для разных характеристик нормирования погрешностей СИ эти вычисления производятся по-разному, поэтому рассмотрим три характерных случая.

1. Класс точности прибора указан в виде одного числа γ_s , заключенного в кружок. Тогда относительная погрешность результата (в процентах) $\gamma(x) = \gamma_s$, а абсолютная его погрешность $\Delta(x) = \gamma_s x / 100$.

2. Класс точности прибора указан одним числом γ_0 (без кружка). Тогда абсолютная погрешность результата измерения $\Delta(x) = \gamma_0 X_K / 100$, где X_K — предел измерений, на котором оно производилось, а относительная погрешность измерения (в процентах) находится по формуле

$$\gamma(x) = \frac{\Delta(x)}{x} = \gamma_0 \frac{X_K}{x}, \quad (1-6)$$

т. е. в этом случае при измерении, кроме отсчета измеряемой величины x , обязательно должен быть зафиксирован и предел измерений X_K , иначе впоследствии нельзя будет вычислить погрешность результата.

3. Класс точности прибора указан двумя числами в виде γ_K / γ_H . В этом случае удобнее вычислить относительную погрешность результата по формуле (1-5), а уже затем найти абсолютную погрешность как $\Delta(x) = \gamma(x) x / 100$.

При использовании этих формул полезно помнить, что в формулы для определения $\gamma(x)$ значения γ_s , γ_0 , γ_H и γ_K подставляются в процентах, поэтому и относительная погрешность результата измерения получается также в процентах.

Однако для вычисления абсолютной погрешности $\Delta(x)$ в единицах x значение $\gamma(x)$ (в процентах) надо не забыть разделить на 100.

1-4. ПРАВИЛА ОКРУГЛЕНИЯ ЗНАЧЕНИЙ ПОГРЕШНОСТИ И РЕЗУЛЬТАТА ИЗМЕРЕНИЙ

Рассчитывая значения погрешности по формулам (1-5) и (1-6), особенно при пользовании электронным калькулятором, значения погрешностей получают с большим числом знаков. Однако исходными данными для расчета являются нормируемые значения погрешности СИ, которые указываются всего с одной или двумя значащими цифрами. Вследствие этого и в окончательном значении рассчитанной погрешности должны быть оставлены только первые одна-две значащие цифры. При этом приходится учитывать следующее. Если полученное число начинается с цифр 1 или 2, то отбрасывание второго знака приводит к очень большой ошибке (до 30—50%), что недопустимо. Если же полученное число начинается, например, с цифры 9, то сохранение второго знака, т. е. указание погрешности, например, 0,94 вместо 0,9, является дезинформацией, так как исходные данные не обеспечивают такой точности.

Исходя из этого на практике установилось такое правило: если полученное число начинается с цифры, равной или большей $\sqrt{10} \approx 3$, то в нем сохраняется лишь один знак; если же оно начинается с цифр, меньших 3, т. е. с цифр 1 и 2, то в нем сохраняют два знака. В соответствии с этим правилом установлены и нормируемые значения погрешностей средств измерений: в числах 1,5 и 2,5% указываются два знака, но в числах 0,5; 4; 6% указывается лишь один знак.

В итоге можно сформулировать три правила округления рассчитанного значения погрешности и полученного экспериментального результата измерения.

1. Погрешность результата измерения указывается двумя значащими цифрами, если первая из них равна 1 или 2, и одной, — если первая есть 3 и более.

2. Результат измерения округляется до того же десятичного разряда, которым оканчивается округленное значение абсолютной погрешности.

3. Округление производится лишь в окончательном ответе, а все предварительные вычисления проводят с одним-двумя лишними знаками.

Пример. На вольтметре класса точности 2,5 с пределом измерений 300 В был получен отсчет измеряемого напряжения $x = 267,5$ В.

Расчет погрешности удобнее вести в следующем порядке: сперва необходимо найти абсолютную погрешность, а затем — относительную. Абсолютная погрешность $\Delta(x) = \gamma_0 X_K / 100$; при $\gamma_0 = 2,5\%$ и $X_K = 300$ В. это дает $\Delta(x) = \frac{2,5 \cdot 300}{100} = 7,5$ В ≈ 8 В; относительная $\varphi(x) = \frac{\Delta_0}{x} \cdot 100 = \frac{7,5}{267,5} \cdot 100 = 2,81\% \approx 2,8\%$.

Так как первая значащая цифра значения абсолютной погрешности (7,5 В) больше трех, то это значение должно быть округлено по обычным правилам округления до 8 В, но в значении относительной погрешности (2,81%) первая значащая цифра меньше 3, поэтому здесь должны быть сохранены в ответе два десятичных разряда и указано $\varphi(x) = 2,8\%$. Полученное значение $x = 267,5$ В должно быть округлено до того же десятичного разряда, которым оканчивается округленное значение абсолютной погрешности, т. е. до целых единиц вольт.

Таким образом, в окончательном ответе должно быть сообщено: «Измерение произведено с относительной погрешностью $\varphi(x) = 2,8\%$. Измеренное напряжение $x = (268 \pm 8)$ В или $x = 268 \text{ В} \pm 8 \text{ В}$.

При этом более наглядно указать пределы интервала неопределенности измеренной величины в виде $x = 260 \div 276$ В или $260 \text{ В} < x < 276 \text{ В}$.

Наряду с изложенными правилами округления значений погрешностей результатов измерения иногда (например, в [37]) предлагаются более обоснованные, но и более сложные правила. Недостаток изложенных правил состоит в том, что относительная погрешность* от округления изменяется скачком при переходе, например, от числа 0,29, когда она составляет $(0,30 - 0,29)/0,30 = 3\%$, к числу 0,3, когда она будет $(0,4 - 0,3)/0,3 = 30\%$. Для устранения столь резкого скачка относительной погрешности округления предлагается каждую декаду возможных значений округляемой погрешности делить на три части: от 0,1 до 0,2, от 0,2 до 0,5 и от 0,5 до 1,0, и в каждой из этих частей использовать свой шаг округления, соответственно равный 0,02, 0,05 и 0,1. Тогда ряд разрешенных к употреблению округленных значений погрешностей получает вид: 0,10—0,12—0,14—0,16—0,18—0,20—0,25—0,30—0,35—0,40—0,45—0,5—0,6—0,7—0,8—0,9—1,0. Бесспорное преимущество такого ряда состоит в том, что погрешность от округления на границах участков изменяется лишь от 5 до 10%. Однако при использовании такого правила округления погрешности последние цифры результата, оставляемые после округления, также должны соответствовать приведенному ряду.

1-5. ПРОСТЕЙШАЯ ОЦЕНКА ДИНАМИЧЕСКИХ ПОГРЕШНОСТЕЙ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

Все рассмотренное выше относилось к определению статических погрешностей результатов измерений, когда предполагалось, что измеряемая величина остается во времени неизменной. На практике, однако, наоборот, сама потребность в измерении тех или иных величин чаще всего обусловлена тем, что они не остаются постоянными, а изменяются во времени. Принято говорить, что в этом случае речь идет не об измерении той или иной

величины, а об измерении процесса ее изменения во времени.

Так как результат измерения величины отражается числом, то результат измерения процесса представляет собой ряд последовательных чисел, отражающих значения измеряемой величины в последовательные моменты времени. Таким образом, результат измерения процесса в виде конечного набора чисел отражает этот процесс с каким-то приближением, т. е. с погрешностью. Эта погрешность носит название *погрешности восстановления*.

Это название обусловлено тем, что, просто глядя на таблицу результатов, человек не может представить себе характер измеренного процесса. Он наносит полученные точки на график и по ним пытается восстановить характер этого процесса.

Методы восстановления могут быть различными — точки можно соединить ступенчатой линией (кусочно-постоянная функция), прямыми (кусочно-линейная) или дугами парабол и т. д., и погрешность восстановления будет разной. Но совершенно ясно, что погрешность восстановления будет большой, если точки расположены во времени далеко друг от друга, и малой, — если они расположены часто.

Выбор средств измерений по их бытродействию прежде всего определяется именно бытродействием, т. е. обеспечиваемой *частотой дискретизации во времени*. Ее возможное наибольшее значение полностью определяется методом регистрации данных в используемой аппаратуре. Так, стрелочные показывающие приборы имеют время установления показаний 4 с плюс время записи результата наблюдателем (примерно 2 с), поэтому период дискретизации в этом случае $t_0 \geq 6$ с. При использовании бытродействующих цифровых приборов и записи результатов наблюдателем в большинстве случаев можно считать $t_0 \geq 2$ с, электрифицированная пишущая машинка обеспечивает скорость печати 7 знаков в секунду, но если каждый отсчет регистрируется тремя десятичными цифрами плюс пробел, то $t_0 \approx 0,5$ с.

Более бытродействующими регистраторами на носитель, пригодный для ввода данных в ЭВМ, являются ленточные перфораторы ПЛ-80 (80 знаков в секунду) и ПЛ-150 (150 знаков в секунду). При регистрации на каждый отсчет трех десятичных знаков плюс пробел перфоратор ПЛ-80 позволяет регистрировать $80/4 = 20$ отсчетов в секунду, а ПЛ-150, соответственно, $150/4 = 37,5$ отсчета в секунду, т. е. $t_0 = 0,05$ или $t_0 = 0,027$ с $\approx 0,03$ с.

Большую скорость регистрации данных могут обеспечить лишь бытродействующие АЦП, работающие совместно с ЭВМ. В этом случае t_0 может составлять, например, 30 мкс, т. е. обеспечивается регистрация 30 000 отсчетов в секунду.

Наибольшее бытродействие достигается путем осуществления так называемого канала прямого доступа в память ЭВМ, минуя ее процессор. В этом режиме, например, ЭВМ «Электроника-100» или «Электроника-60» обеспечивают ввод (12—16)-разрядных

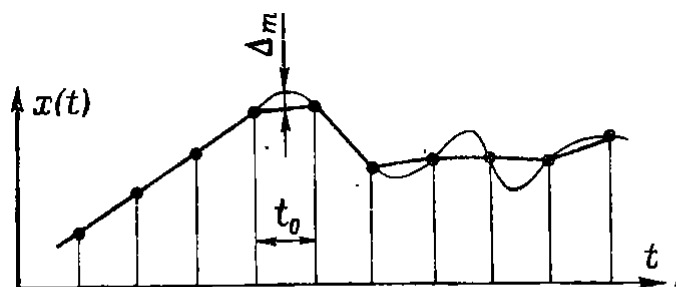


Рис. 1-7

двоичных чисел, т. е. четырехразрядных десятичных, с частотой 500 000 чисел в секунду.

Для того чтобы получить представление о том, какие процессы и с какой погрешностью восстановления могут быть зарегистрированы при использовании перечисленных техниче-

ских средств, рассмотрим пример. Предположим, что восстановление кривой процесса по зарегистрированным отсчетам производится методом линейной интерполяции, т. е. полученные точки просто соединяются между собой отрезками прямых линий. В этом случае плавные участки, близкие к прямым линиям, восстанавливаются с малыми погрешностями, а максимальная погрешность восстановления получается на участках с максимальной кривизной (см. рис. 1-7). Известно, что любую кривую $x(t)$ на некотором участке можно разложить по степеням t , т. е. описать многочленом. В простейшем случае, используя лишь первые члены разложения, участок кривой между отсчетами можно представить в виде параболы, тогда погрешность линейной интерполяции будет представлять собой разность между этой параболой и ее хордой, соединяющей смежные отсчеты. Как известно, парабола имеет наибольшее отклонение от хорды в середине интервала интерполяции t_0 с абсолютным значением (Δ_m на рис. 1-7)

$$\Delta_m = x''(t) t_0^2 / 8,$$

где $x''(t)$ — значение второй производной процесса $x(t)$, т. е. оценка его кривизны. Отсюда максимальное значение погрешности восстановления наблюдается на участках кривой с наибольшей кривизной (в области максимумов и минимумов процесса на рис. 1-7).

Если задать не абсолютную погрешность Δ_m , а ее приведенное значение $\gamma_m = \Delta_m / X_K$, где X_K — предел измерений, то можно определить максимальный допустимый период дискретизации t_0 , при котором погрешность восстановления не будет превышать γ_m :

$$t_0 \leq \sqrt{8 X_K \gamma_m / x''(t)_{\max}}.$$

Так как любую сложную кривую можно разложить на ряд гармонических составляющих, определим необходимый период дискретизации для синусоидального процесса. При $x(t) = X_K \sin \omega t$ оценка текущей кривизны $x''(t) = -\omega^2 X_K \sin \omega t$ и ее максимальное значение $x''(t)_{\max} = \omega^2 X_K$. Отсюда необходимый период дискретизации для синусоидального процесса

$$t_0 \leq \frac{\sqrt{8 \gamma_m}}{\omega} = \frac{T}{2\pi} \sqrt{8 \gamma_m}. \quad (1-7)$$

Таблица 1-1

Метод регистрации	t_0 , с	Период или частота процесса при γ_m , %, равной		
		0,1	1	10
Запись в журнал с показывающего прибора	6	7 мин	2,2 мин	42 с
Цифропечать	0,5	35 с	11 с	3,5 с
Перфоратор ПЛ-150	0,027	1,9 с	0,6 с	0,2 с
АЦП с ЭВМ	$30 \cdot 10^{-6}$	500 Гц	1,5 кГц	5 кГц
«Прямой ввод» в память малых ЭВМ	$2 \cdot 10^{-6}$	7 кГц	23 кГц	70 кГц
Канал ввода больших ЭВМ	10^{-6}	14 кГц	46 кГц	140 кГц

Соотношение (1-7) воспринимается более наглядно, если из него вычислить число точек n , приходящихся на каждый период T синусоидального процесса:

$$n = T/t_0 \geq \pi/\sqrt{2\gamma_m}. \quad (1-8)$$

Это соотношение дает:

γ_m , %	0,1	1	10	20
n	70	22	7	5

Таким образом, для восстановления синусоидального процесса с максимальной погрешностью 1% при равномерной дискретизации необходимо иметь 22 отсчета на период процесса, но для представления с погрешностью 0,1% нужно не менее 70 отсчетов на каждый период, а для $\gamma_m = 20\%$ достаточно пяти отсчетов на период.

Исходя из соотношения (1-8), можно подсчитать минимальный период или максимальную частоту процесса, который может быть зарегистрирован с заданной максимальной погрешностью γ_m разными из перечисленных выше средств. Данные о максимальных погрешностях приведены в табл. 1-1 и свидетельствуют о том, что без использования ЭВМ или специальных самописцев, магнитописцев или осциллографов могут быть зарегистрированы лишь очень медленные процессы (с периодом 0,2—2 с).

Из выражения (1-7) или (1-8) получаем

$$\gamma_m \leq \frac{\sigma^2}{2} \frac{t_0^2}{T^2} = \frac{\sigma^2}{2} t_0^2 f^2 \quad (1-9)$$

т. е. динамическая погрешность восстановления γ_m возрастает в квадрате частоты восстанавливаемого процесса.

Особенности измерения сложных несинусоидальных процессов. На практике чаще всего приходится измерять существенно несинусоидальные процессы, содержащие гармонические составляющие или высокочастотные составляющие шумов, помех или наводок. В этих случаях динамическая погрешность восстановления процесса по дискретным отсчетам резко возрастает, о чем исследователь должен всегда помнить.

Рассмотрим это свойство погрешности восстановления на конкретном примере. Так, в табл. 1-1 указано, что при использовании АЦП с периодом дискретизации $t_0 = 30$ мкс исследуемый процесс с частотой $f_1 = 500$ Гц восстанавливается с $\gamma_{m1} \approx 0,1\%$. Действительно, рассчитывая γ_{m1} по формуле (1-9), получаем

$$\gamma_{m1} = \frac{\pi^2}{2} t_0^2 f_1^2 = \frac{\pi^2}{2} (30 \cdot 10^{-6} \cdot 500)^2 = 10^{-3} = 0,1\%,$$

что часто можно считать достаточно высокой точностью восстановления. Однако если в кривой этого процесса содержится дополнительно еще 10-я гармоника с частотой $f_{10} = 5000$ Гц и амплитудой в 0,1 основной волны, она будет восстанавливаться с относительной погрешностью γ_{m10} , в 100 раз большей, чем γ_{m1} , т. е. равной 10%. Правда, так как амплитуда этой гармоники в 10 раз меньше амплитуды основной волны, то приведенное значение этой погрешности составит лишь $\gamma_{m10} = 1\%$. Тем не менее результирующая погрешность восстановления всего процесса будет в 10 раз (!) больше, чем погрешность восстановления $\gamma_{m1} = 0,1\%$ процесса, не содержащего этой высокочастотной составляющей.

Погрешность восстановления для основной волны и ее гармоник является систематической (она всегда отрицательна, см. рис. 1-7, и приводит к уменьшению восстанавливаемой амплитуды кривой), однако если высокочастотная составляющая вызвана шумом или другими помехами и не синхронна с основной волной, то и погрешность восстановления оказывается случайной и наблюдается в виде случайного разброса отсчетов.

При ручной регистрации наблюдений подобный разброс данных будет сразу замечен экспериментатором и он примет соответствующее решение о ходе эксперимента. Рассмотренное явление особенно опасно при автоматическом вводе данных в ЭВМ и подчеркивает крайнюю важность метрологического анализа динамических погрешностей в этом случае.

Простейшая оценка динамических погрешностей при использовании аналоговых средств регистрации. Как видно из данных табл. 1-1, с применением ручной записи показаний, цифрорепренти или перфораторов могут быть зарегистрированы только очень медленные процессы с периодом 0,6 с — 2,2 мин (при $\gamma_m \approx \approx 1\%$). Поэтому для регистрации процессов, протекающих с частотами от 1 Гц до 5—50 кГц, широко используются аналоговые средства регистрации: самопишущие приборы с чернильной за-

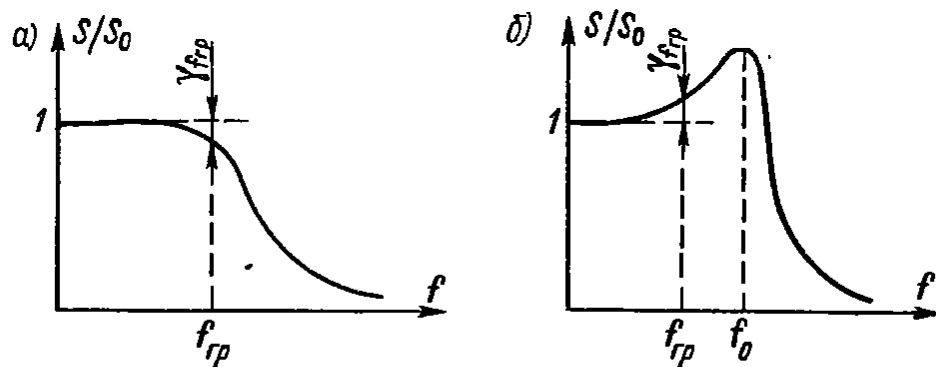


Рис. 1-8

писью, светолучевые и электронные осциллографы с фотоприставками.

Динамические погрешности таких приборов, а также используемых в комплекте с ними датчиков и усилителей принято нормировать указанием их амплитудно-частотной характеристики, представляющей собой график зависимости от частоты f отношения их чувствительности S при частоте f к чувствительности S при $f = 0$ (см. ГОСТ 8.256—77).

В большинстве случаев эти характеристики имеют вид кривых, показанных на рис. 1-8. При апериодической частотной характеристике (рис. 1-8, а) (показывающие приборы, усилители, датчики температуры и т. п.) чувствительность S прибора или датчика монотонно понижается с ростом частоты f воспринимаемого процесса. Частотная погрешность γ_f есть разность между ординатами частотной характеристики и постоянным уровнем $S/S_0 = 1$, показанным на рис. 1-8, а штриховой прямой. Она всегда отрицательна и увеличивается с ростом частоты f . Ее численное значение может быть найдено из этого графика для любой частоты и использовано для оценки точности регистрации или введения поправки в результат измерения.

Частотная характеристика, изображенная на рис. 1-8, б, характерна для колебательных систем с малым успокоением (гальванометров, светолучевых осциллографов, датчиков манометров, акселерометров и т. д.). Она имеет резонансный пик вблизи собственной частоты f_0 колебательной системы и положительную частотную погрешность γ_f .

Для приборов и датчиков с такими частотными характеристиками нормируется рабочий диапазон частот, простирающийся от $f = 0$ до такой частоты $f_{гр}$, где γ_f достигает некоторого граничного значения $\gamma_{f_{гр}}$. Так как граничное значение частотной погрешности достигается только в конце рабочего диапазона частот, то внутри его частотные погрешности оказываются много меньше этого значения.

Располагая частотной характеристикой прибора или датчика, можно найти частотную погрешность для любого значения частоты регистрируемого процесса внутри рабочего диапазона частот. Так, например, при частотной характеристике, приведенной

на рис. 1-8, б, частотная погрешность может быть рассчитана по формуле

$$\gamma_f \approx (1 - 2\beta^2) (f/f_0)^2, \quad (1-10)$$

где β — степень успокоения колебательной системы; f_0 — ее собственная частота.

При отсутствии успокоения ($\beta \approx 0$), что характерно для датчиков, не имеющих специальных средств успокоения, частотная погрешность

$$\gamma_f \approx (f/f_0)^2. \quad (1-11)$$

Так же легко может быть вычислена частотная погрешность и для апериодических (неколебательных) преобразователей невысоких порядков. Так, например, термopapa или термометр сопротивления могут быть представлены апериодическим звеном первого порядка (с одной постоянной времени). Для них частотная погрешность может быть приближенно выражена как

$$\gamma_f \approx -\frac{1}{2} (f/f_c)^2, \quad (1-12)$$

где $f_c = (2\pi\tau)^{-1}$ — так называемая *частота среза* частотной характеристики, а τ — постоянная времени.

Практическое использование формул (1-11) и (1-12) рассмотрим на двух конкретных примерах.

1. Пусть для регистрации пульсирующего давления используется мембранный датчик (тензометрический, пьезоэлектрический, емкостный или индуктивный) с собственной частотой $f_0 = 5$ кГц. Какие процессы и с какой погрешностью могут быть им измерены? Полагая, что степень успокоения датчика $\beta \approx 0$, и используя соотношение (1-11), получаем, что при частоте измеряемого процесса $f = 50$ Гц его частотная погрешность $\gamma_f = 0,01\%$, но при $f = 100$ Гц уже $\gamma_f = 0,04\%$, при $f = 500$ Гц $\gamma_f = 1\%$, а при $f = 1000$ Гц $\gamma_f = 4\%$, т. е. рабочий диапазон частот датчика оказывается уже исчерпанным.

2. Пусть периодические колебания температуры измеряются с помощью термopapa или термометра сопротивления средней инерционности с постоянной времени $\tau = 1$ мин = 60 с. Спрашивается, каков рабочий диапазон частот такого датчика? Для этого преобразуем формулу (1-12), заменив f на $1/T$, где T — период измеряемого процесса; тогда получим

$$\gamma_f \approx -\frac{1}{2} \frac{f^2}{f_c^2} = -\frac{1}{2} \frac{(2\pi\tau)^2}{T^2} = -\frac{2\pi^2\tau^2}{T^2}.$$

Подставляя в это выражение разные значения периода T измеряемых колебаний, получим частотную погрешность $\gamma_f = 0,14\%$ при периоде колебаний $T = 2$ ч, $\gamma_f = 0,5\%$ — при $T = 60$ мин, $\gamma_f = 2\%$ — при $T = 30$ мин, $\gamma_f = 5\%$ — при $T =$

≈ 20 мин, т. е. рабочий диапазон частот можно считать исчерпанным.

Соотношения (1-11) и (1-12) показывают, что частотная погрешность возрастает пропорционально квадрату частоты, что, как и при дискретных отсчетах [см. формулу (1-9)], приводит к очень неблагоприятным соотношениям при регистрации несинусоидальных процессов. Поэтому изложенные выше предостережения остаются актуальными и в этом случае.

1.6. ИЗМЕНЕНИЕ ПОГРЕШНОСТИ СРЕДСТВ ИЗМЕРЕНИЙ ВО ВРЕМЯ ИХ ЭКСПЛУАТАЦИИ

Как бы тщательно ни был изготовлен и отрегулирован прибор к моменту выпуска его на приборостроительном заводе, с течением времени в элементах схемы и механизме неизбежно протекают разнообразные процессы старения и погрешность его неуклонно возрастает. Поэтому нормирование гарантированных в паспорте СИ пределов допускаемой погрешности производится заводом-изготовителем с 1,25—2,5-кратным запасом на старение. Такое превышение пределов допускаемой погрешности над фактическим значением погрешности СИ в момент их выпуска с производства или из *ремонта* является по существу единственным практическим способом обеспечения долговременной метрологической стабильности средств измерений.

Это обстоятельство должно быть четко известно потребителю средств измерений, так как его приходится принимать во внимание при решении многих вопросов организации процессов измерений, поддержания СИ в работоспособном состоянии, оценки допускаемых при измерении погрешностей и т. д.

Характер возрастания погрешности СИ во времени показан на рис. 1-9, где приведены результаты проверок на всех цифровых

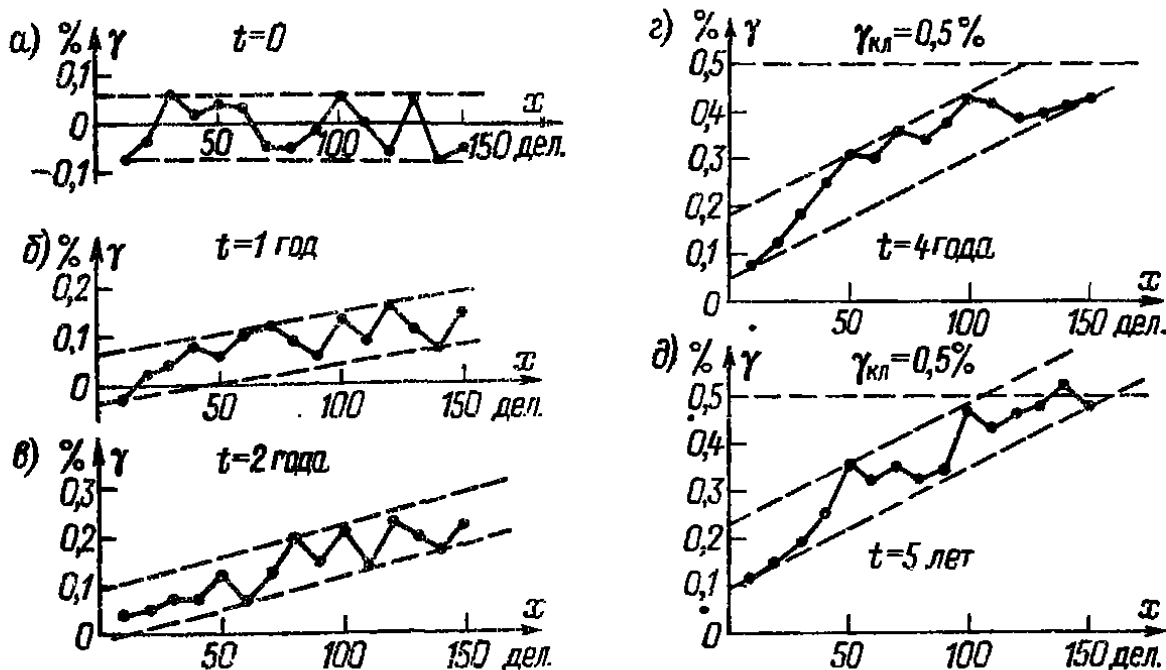


Рис. 1-9

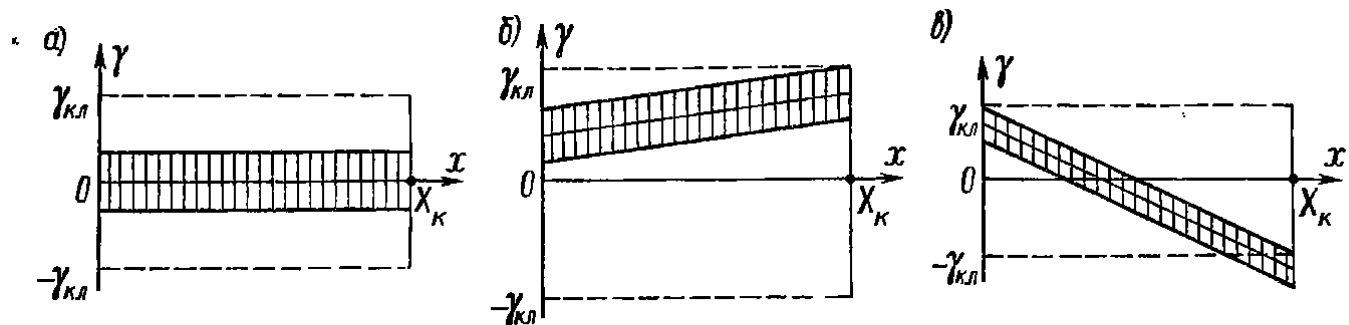


Рис. 1-10

отметках шкалы прибора типа М105 класса точности 0,5 за первые пять лет его эксплуатации. У нового, только что изготовленного прибора (рис. 1-9, а при $t = 0$) полоса его погрешностей располагается симметрично относительно нуля в границах $\pm 0,09\%$. Систематическая погрешность отсутствует, так как она устранена благодаря только что проведенной на заводе регулировке или градуировке шкалы прибора, а случайная погрешность составляет одну пятую часть от нормированного предела $\gamma_{\text{кл}} = 0,5\%$.

Изменение погрешности с возрастом t прибора, наблюдаемое при последующих ежегодных поверках, происходит в виде прогрессирующего смещения и поворота полосы погрешностей, т. е. в виде непрерывного возрастания систематической составляющей погрешности прибора, в то время как размер случайной составляющей определяется шириной полосы погрешностей и остается практически неизменным.

Из кривых рис. 1-9 видно, как постепенно с возрастом t прибора расходуются обеспеченный при изготовлении запас погрешности на старение. Так, если при $t = 0$ он составлял $0,4\%$ (из нормированного значения $\gamma_{\text{кл}} = 0,5\%$), то в возрасте прибора $t = 2$ года максимальная погрешность прибора на 120-м делении шкалы достигла $0,23\%$ и запас стал лишь двукратным. При $t = 4$ года запас на 100-м делении составлял лишь $0,07\%$, т. е. всего $1/7$ от нормированного $\gamma_{\text{кл}} = 0,5\%$, а при $t = 5$ лет запас был уже полностью израсходован и погрешность прибора на 140-м делении превысила допускаемую.

Аналогичный характер имеет и процесс накопления прогрессирующей погрешности с возрастом цифровых приборов и измерительных каналов измерительных информационных систем (ИИС) или измерительно-вычислительных комплексов (ИВК). Как правило, ИИС и ИВК выполняются достаточно высококачественно, т. е. при изготовлении полосе погрешностей канала стремятся придать вид, показанный на рис. 1-10, а. Однако накопление прогрессирующей погрешности приводит, как и у других СИ, к смещению и повороту их полосы погрешностей, т. е. к постепенному расходованию запаса погрешности, созданного при изготовлении. Так, поверка одной из ИИС типа К200 показала, что через несколько лет после выпуска полоса погрешностей имела вид, пред-

ставленный на рис. 1-10, б, т. е. система находилась на пороге выхода из нормированного допуска. А проверка одной из больших ИИС через 5 лет после ее выпуска дала для каналов картину, представленную на рис. 1-10, в.

Таким образом, характер проявления прогрессирующей погрешности с возрастом СИ является единым для всех СИ и пользователь средств измерений не может его игнорировать.

Математическое описание динамики прогрессирующей погрешности средств измерений. Динамика погрешности СИ подробно рассмотрена в монографии [29] на основе анализа фактических данных для отечественных и зарубежных аналоговых и цифровых приборов и ИИС.

Так как заранее не известно, на каком делении шкалы погрешность прибора превысит нормируемое для него значение $\gamma_{\text{нл}}$ (на 100, 120 или 140-м на рис. 1-9), то процесс возрастания прогрессирующей погрешности в работе [29] рассматривался как нестационарный случайный процесс, состоящий из пучка реализаций, соответствующих траекториям возрастания погрешности на каждой из цифровых отметок шкалы прибора.

Далее рассматривалось текущее значение $\gamma(t)$ верхней границы этого пучка траекторий в виде его верхней 95%-ной квантили, а достижение ею границы допуска $\gamma_{\text{нл}}$ принималось за метрологический отказ прибора. Оказалось, что для всех рассмотренных электроизмерительных аналоговых и цифровых приборов текущее значение $\gamma(t)$ приведенной погрешности (в процентах) описывается соотношением

$$\gamma(t) = \gamma_0 + \frac{v_0}{a} (e^{at} - 1) = \gamma_0 + v_0 \tau (1 - e^{-t/\tau}),$$

где t — возраст прибора с момента его изготовления, лет; τ — постоянная времени процесса метрологической стабилизации прибора, лет; $a = -1/\tau$ — отрицательное ускорение процесса старения, 1/год; v_0 — начальная скорость возрастания приведенного значения прогрессирующей погрешности, %/год; γ_0 — значение приведенной погрешности прибора в момент выпуска из производства, %.

Эта зависимость графически представлена кривой 1 на рис. 1-11. При $t = 0$ она выходит из точки с ординатой γ_0 , скорость ее возрастания постепенно замедляется с отрицательным ускорением a , и при $t \rightarrow \infty$ она стремится к установившемуся значению $\gamma_\infty = \gamma_0 + v_0 \tau$. Это широко известный процесс стабилизации характеристик СИ с их возрастом, который протекает тем быстрее, чем меньше постоянная времени $\tau = -1/a$.

Если бы нормируемый класс точности СИ назначался из условия $\gamma_{\text{нл}} = \gamma_\infty$, то погрешность СИ $\gamma(t)$ достигла бы значения $\gamma_{\text{нл}}$ лишь при $t = \infty$, т. е. метрологические отказы за весь срок службы СИ практически бы отсутствовали. Однако в погоне за указываемой в паспорте точностью СИ изготовители назначают $\gamma_{\text{нл}} <$

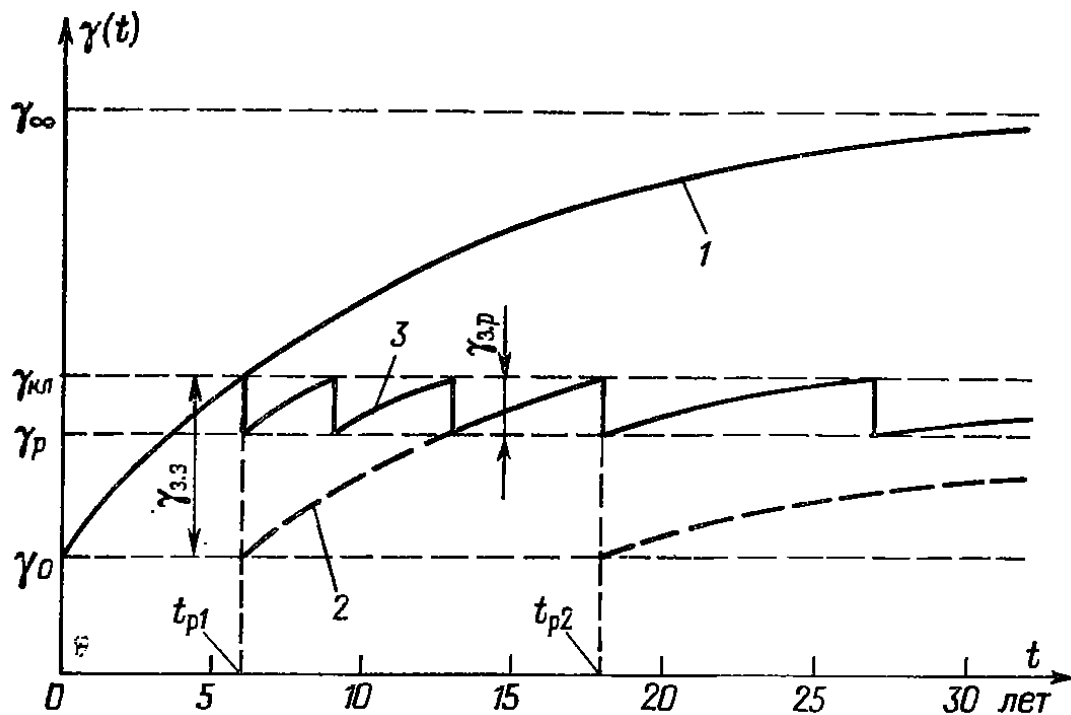


Рис. 1-11

$< \gamma_{\infty}$. В итоге в возрасте прибора t_{p1} , равном, например, 6 лет, погрешность $\gamma(t)$ достигает назначенного значения $\gamma_{кл}$, прибор бракуется и направляется в первый метрологический ремонт и регулировку. Если при ремонте его погрешность вновь доводится до значения γ_0 , то ее возрастание в дальнейшем идет согласно штриховой кривой 2 на рис. 1-11 и при t_{p2} , равном, например, 18 лет, он подвергается второму метрологическому ремонту, и т. д.

Так как процесс метрологического старения идет с замедлением, то текущее значение частоты метрологических отказов $\omega(t)$ так же убывает с возрастом t прибора, как производная от $\gamma(t)$, т. е.

$$\omega(t) = \omega_0 e^{-t/\tau},$$

где $\omega_0 = v_0/(\gamma_{кл} - \gamma_0) = v_0/\gamma_{з.з}$ — частота метрологических отказов на момент времени $t = 0$, 1/год; v_0 — начальная скорость возрастания погрешности, %/год; $\gamma_{з.з}$ — заводской запас погрешности (см. рис. 1-11), обеспечиваемый при выпуске с производства, %.

Значения параметров процесса метрологического старения, полученные для некоторых аналоговых и цифровых приборов в монографии [29], показывают, что размер γ_{∞} у аналоговых приборов составляет от 1,2 до 2,6% и превосходит размер назначаемого $\gamma_{кл}$ в 1,1—3 раза. У цифровых приборов это соотношение еще больше: так, у измерительных каналов ИИС типа К200 $\gamma_{\infty} = 0,68\%$, т. е. в 3,4 раза превышает $\gamma_{кл} = 0,2\%$, а у цифрового вольтметра фирмы «Солартрон» типа 7071 $\gamma_{\infty} = 0,0185\%$, т. е. в 9,25 раза превосходит $\gamma_{кл} = 0,002\%$.

Среднее значение наработки на первый метрологический ремонт t_{p1} , обеспечиваемое заводским запасом на старение, у анало-

говых приборов составляет от 4 до 14 лет, а для цифрового вольтметра «Солартрон 7071» $t_{pi} = 1$ год.

Таким образом, на тех предприятиях, где ремонтные службы обеспечивают ремонтные запасы, примерно равные заводским, погрешность приборов во время эксплуатации изменяется в соответствии с кривой 2 (рис. 1-11), когда за весь срок службы аналогового прибора происходит лишь 1—2 метрологических ремонта, например, в возрасте 6 и 18 лет, как это показано на рис. 1-11. Так, наблюдения, проведенные в 1965—1975 гг. за двумя тысячами электроизмерительных приборов, эксплуатировавшихся в цехах и лабораториях ЛПЭО «Электросила», показали, что начальное значение погрешности γ_0 новых или только что отремонтированных магнитоэлектрических приборов классов точности 0,5 и 0,2 составляло примерно 0,4 от класса точности.

Однако ремонтные службы многих других предприятий, как показала статистика, приведенная в работе [29], допускают при ремонте выпуск приборов с запасом погрешности, в 2—3, а иногда и в 6 раз меньшим заводского. Это обусловлено тем, что размер запаса погрешности при ремонте не оговорен и даже не упоминается ни в одном из нормативных документов Госстандарта и предъявлять рекламации к службам ремонта, в отличие от заводо-изготовителей, нет юридических оснований.

В итоге из ремонта приборы выходят с погрешностью $\gamma_p \gg \gamma_0$, т. е. с ремонтным запасом $\gamma_{p.з} \ll \gamma_{з.з}$. Вследствие этого изменение текущей погрешности с возрастом прибора получает вид кривой 3, показанной на рис. 1-11. До первого ремонта прибор работает с заводским запасом погрешности $\gamma_{з.з}$ и частота метрологических отказов невелика, после же первого ремонта он получает ремонтный запас $\gamma_{p.з}$ в 2—6 раз меньший заводского. В то же число раз после первого же ремонта возрастает частота метрологических отказов, и прибор бракуется и направляется в ремонт почти ежегодно. Кроме многократной перегрузки ремонтных и поверочных служб такой «способ» эксплуатации СИ не гарантирует от того, что погрешность прибора превосходит нормированное значение $\gamma_{кл}$ задолго до ежегодной поверки и большую часть года прибор используется с погрешностью больше нормированной.

При анализе статистических данных возрастание частоты метрологических отказов после первого ремонта из-за того, что $\gamma_{p.з} \ll \gamma_{з.з}$, воспринимается как положительное ускорение процесса старения, что и отмечалось некоторыми авторами. Однако такую трактовку этого явления и положительные значения ускорения процесса старения приборов, приведенные, например, в методике [23], теперь следует считать ошибочными, так как положительное ускорение частоты отказов — это не свойство СИ данного типа, а характеристика плохой работы ремонтных служб, ремонтировавших эти СИ.

На основании изложенного можно сделать выводы, важные для пользователей средств измерений:

1) метрологическая исправность СИ в течение ряда лет обеспечивается наличием запаса нормированного значения погрешности по отношению к ее фактическому значению;

2) заводы-изготовители СИ во избежание рекламаций выполняют начальный запас на старение вплоть до 2,5-кратного, когда $\gamma_0 = 0,4\gamma_{кл}$ и $\gamma_{в.з} = 0,6\gamma_{кл}$. Минимально допустимым является 1,25-кратный запас, когда $\gamma_0 = 0,8\gamma_{кл}$ и $\gamma_{в.з} = 0,2\gamma_{кл}$;

3) хотя эти нормы не регламентированы какими-либо нормативными документами Госстандарта, приборопользователь должен требовать от ремонтных служб их выполнения; так как при получении из ремонта прибора с $\gamma_p = (0,85 \div 0,9) \gamma_{кл}$ нет гарантии того, что такого запаса хватит даже на один год эксплуатации.

■

1-7. СООТНОШЕНИЕ СЛУЧАЙНОЙ И СИСТЕМАТИЧЕСКОЙ СОСТАВЛЯЮЩИХ В РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ

Знание этого соотношения необходимо пользователю СИ во многих случаях. Так, например, при статистической обработке данных путем усреднения многократных отсчетов могут быть уменьшены лишь случайные погрешности, а систематические — не уменьшаются при любом числе усредняемых отсчетов. При использовании же ручной или автоматической коррекции, наоборот, могут быть уменьшены лишь систематические погрешности, а случайные не только не уменьшаются, но даже возрастают, и т. д.

Соотношение между случайной и систематической составляющими видно из кривых, представленных на рис. 1-9—1-11. Действительно, размер случайной составляющей соответствует половине ширины полосы погрешностей на рис. 1-9 и 1-10. При эксплуатации, как видно из рис. 1-9, он остается приблизительно неизменным. Описанное выше (§ 1-6) исследование 2000 приборов, эксплуатировавшихся в цехах и лабораториях ЛПЭО «Электросила», показало, что эта с л у ч а й н а я погрешность приборов составляет примерно $\gamma_0 = 0,4\gamma_{кл}$.

Возрастающая со временем прогрессирующая погрешность СИ для каждого конкретного результата измерения является с и с т е м а т и ч е с к о й. На протяжении ремонтного интервала она возрастает от 0 до $0,6\gamma_{кл}$, поэтому в среднем для всего парка эксплуатируемых приборов может быть принята $\theta = 0,3\gamma_{кл}$.

Эти средние соотношения для случайной и систематической составляющих погрешности справедливы не только для электроизмерительных приборов, но и для измерительных каналов ИИС и ИВК, так как они обусловлены лишь тем, что при выпуске из ремонта или производства погрешность СИ не удастся сделать меньше, чем $\gamma_0 = 0,4\gamma_{кл}$, а при достижении суммарной погреш-

ностью значения $\gamma_{\text{кл}}$ такой канал ИИС или ИВК из эксплуатации исключается.

Конечно, если к эксплуатации после ремонта или некачественного изготовления допускаются СИ или каналы ИИС с погрешностью $\gamma = 0,8\gamma_{\text{кл}}$ и текущее значение погрешности изменяется по кривой 3 (рис. 1-11) от $0,8\gamma_{\text{кл}}$ до $\gamma_{\text{кл}}$, то среднее значение систематической составляющей погрешности оказывается равным не $0,3\gamma_{\text{кл}}$, а $0,5\gamma_{\text{кл}}$, т. е. больше, чем случайная составляющая.

Это становится особенно опасным, если в ИИС или ИВК использована программа статистического усреднения многократных отсчетов и полученному результату присваивается погрешность, в $\sqrt{n}$ меньшая погрешности канала, где n — число усредняемых отсчетов. Систематическая погрешность в конце межремонтного интервала независимо от качества ремонта достигает значения $0,6\gamma_{\text{кл}}$ и остается после усреднения той же самой. В этих условиях уменьшенная в $\sqrt{n}$ раз и сообщаемая получателю информации погрешность усредненного результата становится чистейшим обманом, так как фактическая погрешность результатов при любом большом n остается равной $0,6\gamma_{\text{кл}}$.

Если же ИИС, ИВК или датчики, используемые в их каналах, эксплуатируются длительное время без проведения периодических метрологических проверок, то накопившиеся прогрессирующие погрешности могут быть еще больше. Использование в этих условиях программ статистического усреднения многократных отсчетов может оказаться полным самообманом.

Приведенные примеры показывают, что при оценке погрешностей измерений, проводимых средствами измерений, длительное время не подвергавшимися проверке, особое внимание следует уделять оценке возможных накопившихся прогрессирующих погрешностей.

1-8. ОПРЕДЕЛЕНИЕ И КОРРЕКЦИЯ ПОГРЕШНОСТЕЙ РЕЗУЛЬТАТОВ В ПРОЦЕССЕ ИЗМЕРЕНИЙ

Каждый результат измерения должен сопровождаться сообщением о его погрешности и тем самым четко указывать интервал его неопределенности. Как уже отмечалось в § 1-3, более наглядно приводить результат измерения не в виде $x \pm \Delta$, а в виде явного указания границ интервала неопределенности $a < x < b$, т. е., например, вместо $x = 1,5 \pm 0,7$ — в виде $0,8 < x < 2,2$.

Методы и формулы, изложенные в § 1-3 и 1-5, позволяют это сделать, пользуясь лишь нормированными характеристиками СИ и полученной в результате измерения оценкой измеряемой величины. Эта процедура сводится к элементарным вычислениям по простейшим формулам. Но если в процессе эксперимента общее число зафиксированных результатов достигает нескольких десятков (не говоря уже о тех случаях, когда их десятки тысяч), такие

вычисления для каждого из результатов становятся обременительными и поэтому практически не проводятся.

Индивидуальная оценка погрешностей всех результатов прямых однократных измерений особенно важна при автоматизации измерений, когда эти результаты без участия экспериментатора вводятся в ЭВМ и используются для дальнейших вычислений. При ручных измерениях экспериментатор интуитивно оценивает качество получаемых данных (по наблюдаемому разбросу, по тому, получен ли результат в конце шкалы прибора или на первых ее отметках, и другим признакам). При автоматических измерениях такой субъективный контроль отсутствует.

Однако использование ИВК открывает возможность автоматического вычисления погрешности для каждого отдельного однократного измерения по приведенным выше простейшим формулам. Благодаря простоте этих вычислений они занимают очень малую часть машинного времени, а исходные данные (метрологические характеристики измерительных каналов) требуют ничтожную часть памяти ЭВМ. Итог же получается очень эффективным — каждый выводимый на печать результат измерений в соседнем столбце таблицы снабжается указанием погрешности, с которой он получен, или границами интервала его неопределенности.

Сообщаемый потребителю интервал неопределенности каждого из полученных результатов оперативно информирует его о качестве измерений, хотя, строго говоря, вычисленная по нормируемым метрологическим характеристикам СИ погрешность результата может быть как больше, так и меньше ее действительного значения.

При использовании новых, только что изготовленных СИ в зависимости от размера запаса на старение вычисленная погрешность может быть в 2,5—1,25 раза больше ее фактического значения и приближается к нему только в конце межремонтного интервала. А меньше может быть потому, что погрешность результата измерения, как это указывалось в § 1-1, определяется не только инструментальной погрешностью СИ, но и методическими погрешностями, допускаемыми самим экспериментатором. Анализ размера методических погрешностей, как подчеркивалось в § 1-1, лежит на ответственности экспериментатора.

Исключение прогрессирующих погрешностей. Скорость изменения во времени прогрессирующих погрешностей носит случайный характер. Поэтому методом обнаружения их накопившихся значений является поверка приборов или измерительных каналов ИИС по образцовым сигналам и мерам. При этом исключение прогрессирующих погрешностей может выполняться как вручную, так и автоматически.

Если до накопления прогрессирующих погрешностей характеристика канала имела вид прямой 1 (рис. 1-12), а после накопления этих погрешностей приобрела вид прямой 2, то для восста-

новления номинальной характеристики необходимо провести две операции: 1) коррекцию нуля, т. е. смещение нижнего конца характеристики 2 в начало координат и 2) коррекцию чувствительности, т. е. смещение верхнего конца характеристики 2 в точку X_k, Y_k .

Для выполнения этих операций большинство СИ, у которых существенную роль играет прогрессирующая погрешность, снабжаются корректором нуля.

Приборы с не стабильной во времени чувствительностью (электронные, цифровые приборы, потенциометры и т. п.) имеют, кроме корректора нуля, также приспособления для проверки и коррекции чувствительности. Казалось бы, что в результате этих двух операций как аддитивные, так и мультипликативные погрешности устраняются и остается лишь случайная составляющая погрешности. Но это не совсем так.

Во-первых, у многих приборов есть локальные отклонения их характеристики от номинальной. Поэтому при совмещении их характеристики с номинальной в двух точках (в нуле и в конце или в другой точке диапазона измерений) она может отклоняться от этой прямой в остальных точках диапазона и возникающая погрешность будет повторяться при каждом измерении, т. е. является систематической.

Во-вторых, производя коррекцию нуля или чувствительности, мы пользуемся для этого показаниями того же самого прибора с присущей ему случайной погрешностью, т. е. фиксируем данную реализацию случайной составляющей в качестве систематической составляющей для последующих измерений. Вследствие этого размер остаточной систематической составляющей погрешности прибора всегда имеет тот же порядок, что и случайная составляющая погрешности.

Такая коррекция исключает все накопившиеся прогрессирующие погрешности вне зависимости от причин их возникновения. Но после проведения коррекции идет новое накопление погрешностей. Поэтому для поддержания погрешности в определенных пределах операция коррекции должна периодически повторяться. Путем ежегодных поверок поддерживается точность всех средств измерений (см. рис. 1-11). Если период поверок уменьшить до одного дня, часа, минуты или секунды, то точность можно существенно повысить, но для этого весь процесс коррекции должен быть полностью автоматизирован. Однако необходимо иметь в виду, что исключение прогрессирующих погрешностей сопровождается удвоением дисперсии случайных погрешностей.

Для проверки этого соотношения в поверочной лаборатории ЛПЭО «Электросила» был поставлен следующий эксперимент. Опытному поверителю было предложено провести многократную

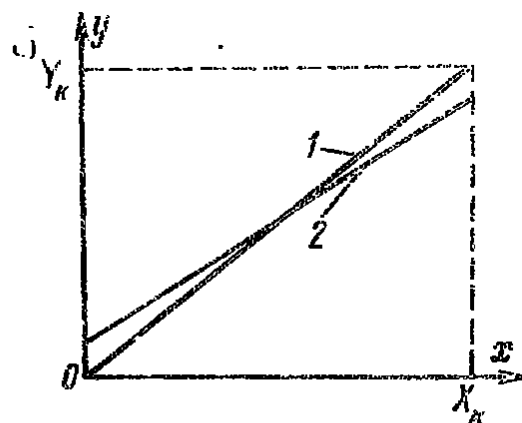


Рис. 1-12

поверку прибора на одной и той же числовой отметке шкалы один раз, устанавливая заново корректором указатель на нулевую отметку перед каждой поверкой, а другой раз — с однажды установленным и заклеенным корректором. Во втором случае дисперсия отсчетов по образцовому потенциометру оказалась ровно в два раза меньшей, чем в первом. Следовательно, устанавливая каждый раз заново указатель на нулевую отметку, поверитель вносит точно такую же случайную погрешность, какая характерна для данного прибора, и общая дисперсия удваивается. Это особенно важно иметь в виду при использовании методов автоматической коррекции. Поэтому вопрос о возможном снижении погрешности результатов измерения требует анализа соотношения случайных и систематических составляющих погрешностей (см. § 4-7 и 8-8—8-10).

✽

ГЛАВА ВТОРАЯ

МЕТОДЫ ВЕРОЯТНОСТНОГО ОПИСАНИЯ ПОГРЕШНОСТЕЙ СРЕДСТВ И РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

✽

2-1. НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Для характеристики частоты появления различных значений случайной величины X (в нашем случае погрешности прибора или результата измерения с учетом и ее систематической составляющей) теория вероятностей предлагает пользоваться указанием закона распределения вероятностей различных значений этой величины. При этом различают два вида описания законов распределения: интегральный и дифференциальный.

Интегральным законом, или *функцией распределения вероятностей* $F(X)$ случайной величины X , называют функцию, значение которой для каждого x является вероятностью события, заключающегося в том, что случайная величина X принимает значения, меньшие x , т. е. функцию $F(x) = P[X < x]$. Это неубывающая функция x , изменяющаяся от $F(-\infty) = 0$ до $F(+\infty) = 1$. Она существует для всех случайных величин, как дискретных, так и непрерывных.

Для случайной величины с непрерывной и дифференцируемой функцией распределения $F(x)$ можно найти *дифференциальный закон распределения* вероятностей, выражаемый как производная от $F(x)$, т. е. как $p(x) = F'(x)$. Эта зависимость называется *кривой плотности распределения вероятностей*. Она всегда неотрицательна и подчинена условию нормирования в виде

$$\int_{-\infty}^{+\infty} p(x) dx = 1,$$

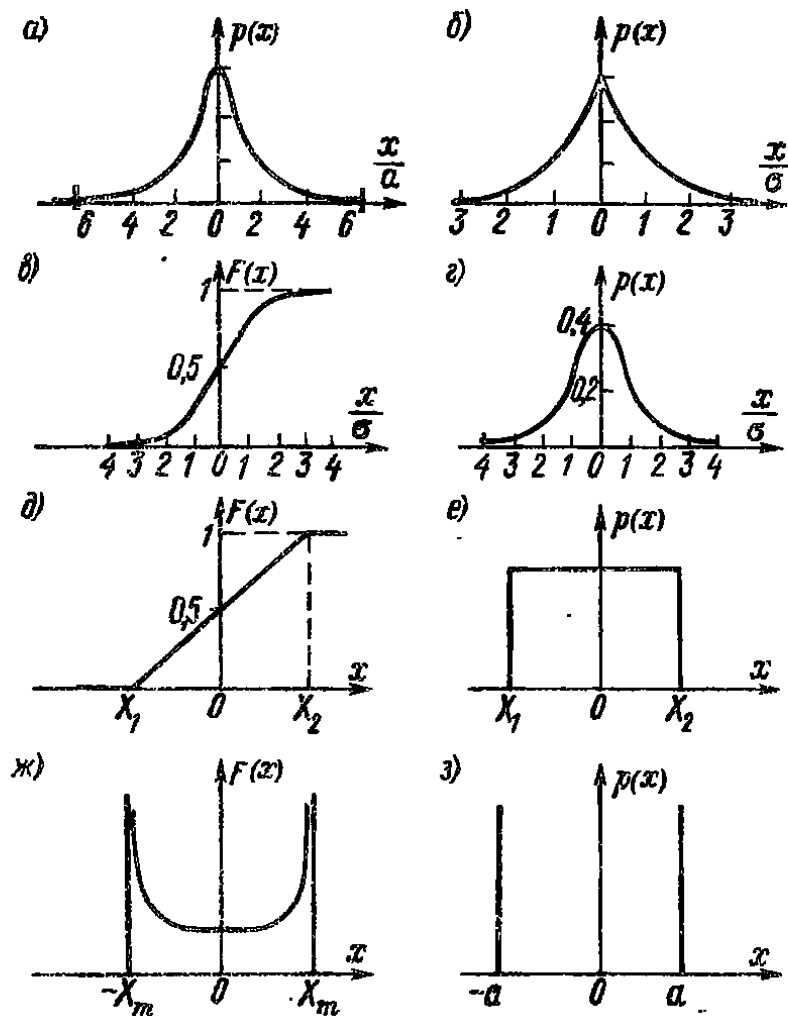


Рис. 2-1

что непосредственно следует из свойств интегральной функции распределения $F(x)$.

Примеры законов распределения. Одним из простейших законов распределения является *распределение Коши*, плотность вероятностей для которого

$$p(x) = \frac{a}{\pi(a^2 + x^2)} = \frac{1}{a\pi[1 + (x/a)^2]}.$$

Это распределение (рис. 2-1, а) близко к предельно пологому, так как при более пологих, чем $1/x^{1+\alpha}$ (где α — сколь угодно малая положительная величина), спадах площадь под кривой бесконечна и не может быть приравнена единице, т. е. не выполняется условие нормирования, и такие кривые не могут описывать плотность распределения вероятностей.

Другим законом распределения, с более быстроспадающей плотностью при отклонении x от центра распределения, является *распределение Лапласа* (рис. 2-1, б) с плотностью

$$p(x) = \frac{1}{2} e^{-|x|},$$

т. е. двустороннее экспоненциальное распределение.

Наиболее часто используемым в теории вероятностей законом

распределения является *нормальный* (распределение Гаусса), плотность вероятности которого описывается выражением

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x}{\sigma} \right)^2 \right],$$

т. е. спадает по мере удаления от $x = 0$ еще быстрее, чем при законе распределения Лапласа. Интегральный закон этого распределения показан на рис. 2-1, в, а кривая плотности — на рис. 2-1, г.

Если непрерывная случайная величина принимает значения лишь в пределах некоторого конечного интервала от X_1 до X_2 с постоянной плотностью вероятностей, то такой закон распределения называют *равномерным*. Его функция распределения (рис. 2-1, д) на участке от $-\infty$ до X_1 равна нулю, на участке от X_1 до X_2 линейно возрастает от 0 до 1, а на участке от X_2 до $+\infty$ равна 1. Плотность вероятностей такого распределения представлена на рис. 2-1, е и записывается как

$$\begin{cases} p(x) = 1/(X_2 - X_1) = \text{const} \text{ при } X_1 < x < X_2; \\ p(x) = 0 \text{ при } x < X_1 \text{ и } x > X_2. \end{cases}$$

Распределение отсчетов синусоидально изменяющейся во времени величины $x = X_m \sin \omega t$, если моменты этих отсчетов равномерно распределены во времени, называется *арксинусоидальным*. Его плотность описывается выражением

$$p(x) = 1/(\pi \sqrt{X_m^2 - x^2})$$

и представлена на рис. 2-1, ж.

Распределение, при котором встречаются с равными вероятностями только два дискретных значения случайной величины $+a$ и $-a$, называется *дискретным двузначным распределением*. Его плотность распределения вероятностей представлена на рис. 2-1, з и описывается аналитически:

$$p(x) = \frac{1}{2} \delta(x - a) + \frac{1}{2} \delta(x + a),$$

где δ — дельта-функция Дирака.

Понятие центра распределения. Координата центра распределения определяет положение случайной величины на числовой оси. Однако дать строгое определение этого понятия далеко не просто. Распределения погрешностей приборов или результатов измерений, как правило, являются симметричными. Поэтому применительно к распределениям вероятностей погрешностей центр распределения может быть определен как центр симметрии распределения.

Координата центра распределения может быть определена несколькими способами. Наиболее общим, а следовательно, и наи-

более фундаментальным является определение центра из принципа симметрии, т. е. как такой точки на оси x , слева и справа от которой вероятности появления различных значений случайной величины равны между собой и составляют $P_1 = P_2 = 0,5$. Такое значение x называется *медианой*. На графике интегрального закона распределения (рис. 2-1, в или д) абсцисса медианы соответствует пересечению кривой уровня $F(x) = 0,5$.

Координата центра может быть определена и по-иному, а именно как *центр тяжести* распределения, т. е. такая абсцисса $\bar{X}$, относительно которой опрокидывающий момент равен нулю, т. е.

$$\bar{X} = \int_{-\infty}^{+\infty} x p(x) dx.$$

Центр распределения, найденный таким путем, носит название *математического ожидания*. При дискретных отсчетах x_i вычисление интеграла, определяющего математическое ожидание, заме-

няют вычислением *среднего арифметического*: $\bar{X} = \sum_{i=1}^n x_i/n$.

При симметричной кривой плотности распределения одной из возможных оценок центра распределения может служить абсцисса *моды* распределения, т. е. максимума плотности. Однако есть распределения, у которых не существует моды. Например, равномерное распределение (рис. 2-1, е). В этих случаях определение центра как моды распределения лишено смысла.

То же самое относится и к понятию математического ожидания. У распределения Коши (рис. 2-1, а), а также у распределений, необходимых при вычислении погрешностей косвенных измерений (см. табл. 3-3), математического ожидания не существует, так как определяющий его интеграл расходится. Понятие же центра распределения правомерно для всех распределений.

При вероятностном описании погрешности координата центра распределения определяет значение систематической составляющей погрешности, т. е. вероятностное описание погрешностей включает в себя и указание ее систематической составляющей.

На рис. 2-1 все распределения были показаны с координатой центра $X_{ц} = 0$. При $X_{ц} \neq 0$ несколько изменяется и аналитическое описание плотности распределения вероятностей. Так, плотность распределения Коши при $X_{ц} \neq 0$ будет

$$p(x) = \left\{ a\pi \left[1 + \left(\frac{x - X_{ц}}{a} \right)^2 \right] \right\}^{-1},$$

а плотность распределения Гаусса

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x - X_{ц})^2}{2\sigma^2} \right]$$

и т. д.

Если из всех наблюдавшихся значений погрешности вычесть систематическую составляющую, т. е. перенести начало координат в центр распределения, то такое распределение называется *центрированным*.

Моменты распределения. Для описания различных свойств распределений используют также параметры законов распределения, называемые моментами. Моменты, найденные без исключения систематической составляющей, называются *начальными*, а найденные для центрированных распределений, — *центральными*.

Первый начальный момент называется математическим ожиданием и был уже рассмотрен выше. Центральный момент k -го порядка для непрерывной случайной величины выражается интегралом

$$\mu_k = \int_{-\infty}^{+\infty} (x - X_n)^k p(x) dx.$$

Второй центральный момент называется дисперсией случайной величины и относится к параметрам, характеризующим рассеяние отдельных ее значений от центра распределения:

$$\mu_2 = D = \int_{-\infty}^{+\infty} (x - X_n)^2 p(x) dx.$$

Дисперсия имеет размерность квадрата случайной величины и выражает как бы мощность рассеяния относительно постоянной составляющей. Поэтому для более наглядной характеристики самого рассеяния пользуются корнем квадратным из дисперсии, т. е. действующим значением рассеяния, которое называется *средним квадратическим отклонением* (сокращенно с. к. о.) и имеет размерность самой случайной величины: $\sigma = \sqrt{D}$.

Третий центральный момент μ_3 характеризует асимметрию, т. е. скошенность распределения: когда один спад — крутой, а другой — пологий. Для симметричных относительно центра распределений он равен нулю. Третий момент имеет размерность куба случайной величины, поэтому для относительной характеристики асимметрии используют безразмерный коэффициент асимметрии, равный третьему моменту, деленному на куб с. к. о.: $s = \mu_3/\sigma^3$.

Четвертый центральный момент μ_4 характеризует протяженность распределения, а отнюдь не остроту его вершины, как это часто ошибочно указывается. Его относительное значение $\varepsilon = \mu_4/\sigma^4$ называется *эксцессом* распределения и для разных законов может иметь значения от 1 (для дискретного двузначного — рис. 2-1, в) до ∞ (для распределения Коши — рис. 2-1, а). Для островершинного треугольного распределения $\varepsilon = 2,4$, а для кругловершинного нормального $\varepsilon = 3,0$. Во многих пособиях по теории вероятностей вводится величина *коэффициента эксцесса* $\gamma_2 = \varepsilon - 3$, которая для менее протяженных распределений (треугольного, равномерного, арксинусоидального) — отрицатель-

на (от -2 до 0), а для более протяженных, чем нормальное, — положительна (от 0 до ∞). Но такое, сдвинутое на 3 , значение γ_2 не входит ни в одну из формул теории вероятностей и поэтому в данной книге использоваться не будет. Для классификации распределений по их форме удобнее использовать другую функцию от эксцесса, а именно $\kappa = 1/\sqrt{\epsilon}$, изменяющуюся для любых распределений лишь в пределах от 0 до 1 . Эту величину будем именовать *контрэксцессом*.

Деформация законов распределения при суммировании случайных величин. Особенность законов распределения таких случайных величин, как погрешности приборов и результатов измерений, состоит в их большом разнообразии. Это вызвано тем, что результирующая погрешность прибора или результата измерения складывается из ряда составляющих. Если эти составляющие рассматривать как случайные величины, то суммирование погрешностей сводится к суммированию случайных величин. Но при суммировании случайных величин законы их распределения резко изменяют свою форму.

Закон распределения суммы независимых случайных величин $p(x) = p(x_1 + x_2)$, имеющих распределения $p_1(x)$ и $p_2(x)$, называется *композицией* и выражается интегралом свертки:

$$p(x) = \int_{-\infty}^{+\infty} p_1(z) p_2(x - z) dz.$$

Изменение формы законов распределения при образовании композиции показано на рис. 2-2.

Так, при суммировании двух равномерно распределенных погрешностей (рис. 2-2, а) с шириной распределений $a > b$ результирующая погрешность имеет распределение в форме трапеции с верхним основанием $a - b$ и нижним $a + b$. Эту деформацию можно представить себе более наглядно как «размыв» резко ограниченных концов более широкого распределения (шириной a) на величину протяженности b менее широкого распределения как это показано штриховыми линиями на рис. 2-2, а.

Композиция двух одинаковых (с шириной a) равномерных распределений является треугольной (так называемое *распределение Симпсона*), так как в этом случае верхнее основание трапеции обращается в нуль, а нижнее — в $2a$.

Подобным же образом образуется композиция равномерного и нормального распределений (рис. 2-2, б), лишь с тем отличием, что подъем и спад по краям результирующего распределения происходит по кривой интегрального закона нормального распределения, аналогично тому, как на рис. 2-2, а он происходил по кривой интегрального закона равномерного распределения (по прямой линии).

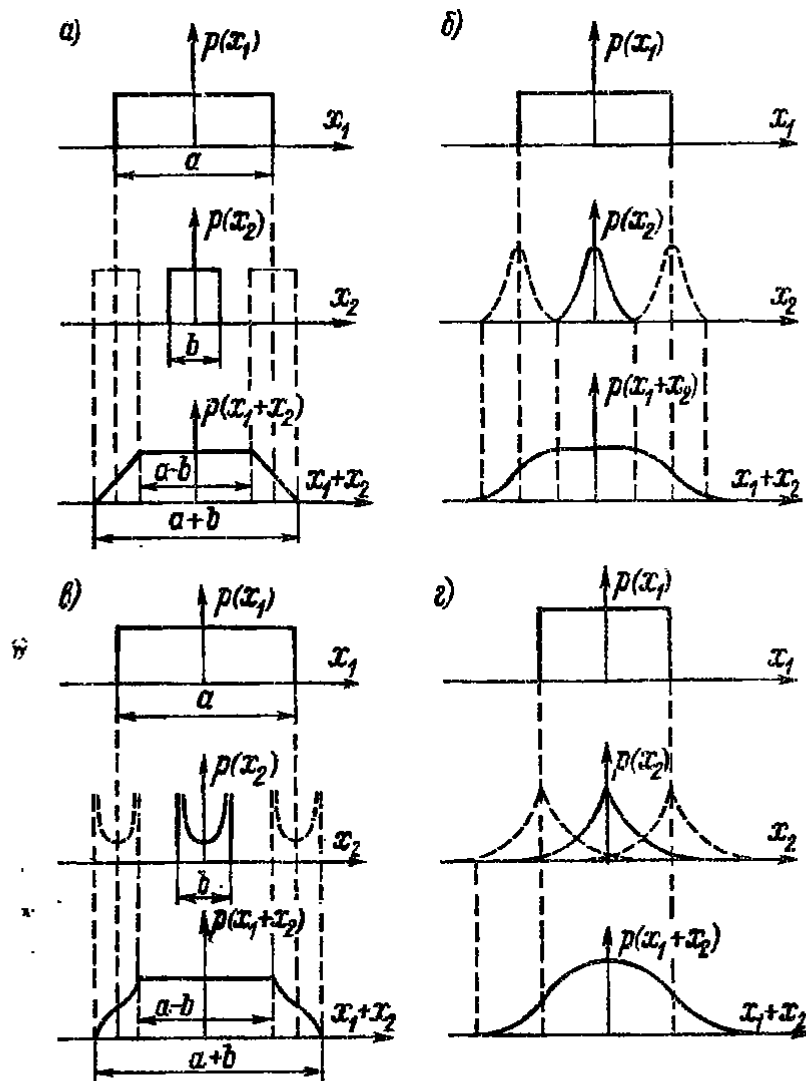


Рис. 2-2

Образование композиции равномерного распределения шириной a и арксинусоидального распределения шириной b показано на рис. 2-2, в. Композиция представляет собой криволинейную трапецию с верхним основанием $a - b$, нижним $a + b$ и спадами по кривым интегрального закона арксинусоидального распределения (функции арксинуса).

Композиция равномерного распределения и распределения Лапласа (двустороннее экспоненциальное распределение на рис. 2-1, б) показана на рис. 2-2, г и имеет длинные, полого спадающие «хвосты» кривой результирующего распределения.

Распределения, показанные на рис. 2-2, построены без соблюдения относительного масштаба кривых по вертикали. Этот масштаб определяется каждый раз тем, что площадь под любой из кривых плотности должна быть равна единице.

2-2. ВЕРОЯТНОСТНЫЕ ОЦЕНКИ ШИРИНЫ РАСПРЕДЕЛЕНИЯ

Для оценки величины разброса случайных погрешностей относительно центра, т. е. ширины распределения, на практике используются различные приемы, приводящие к существенно разным

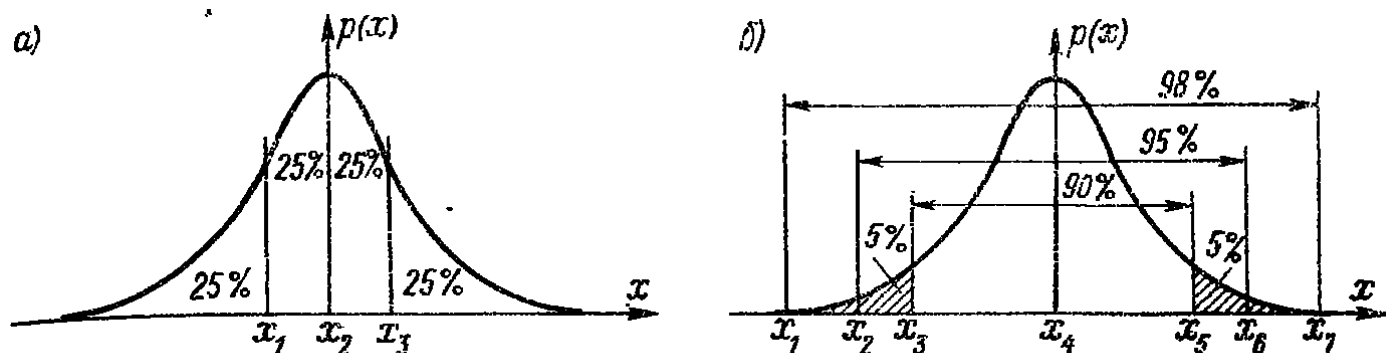


Рис. 2-3

результатам. Поэтому целесообразно сопоставить эти приемы и уяснить себе их особенности.

«Предельная», или «максимальная», оценка случайной погрешности. Она теоретически правомерна только для ограниченных распределений (равномерного, трапецеидального, треугольного, арксинусоидального и т. п.). Для этих распределений (рис. 2-1, е—з и 2-2, а, в) действительно существует такое значение $\pm X_m$, которое ограничивает с обеих сторон возможные значения случайной величины. Однако эти распределения являются лишь теоретической идеализацией и реальные распределения погрешностей, строго говоря, им никогда не соответствуют. Кривые плотности реальных распределений погрешностей (рис. 2-2, б, г и 2-3, б, г), за редкими исключениями, не имеют четко выраженных границ. И поэтому указание для них «предельных», или «максимальных», значений неправомерно. На практике такая оценка есть указание наибольшего по модулю отклонения, встретившегося в данном, произвольно ограниченном ряду наблюдений, так как с увеличением объема выборки экспериментальных данных «предельные» значения монотонно возрастают. Предельная погрешность прибора, найденная экспериментально по 100 отсчетам, всегда будет большей, чем найденная по первым 10 отсчетам.

Главным недостатком такой оценки является бессмысленность арифметического суммирования таких «предельных» значений, так как получаемая сумма может превышать действительные погрешности в несколько раз (см. конец § 3-5).

Квантильные оценки случайной погрешности. Площадь, заключенная под кривой плотности распределения (рис. 2-3), согласно правилу нормирования, равна единице, т. е. отражает вероятность всех возможных событий. Эту площадь можно разделить на некоторые части вертикальными линиями. Абсциссы таких линий называют *квантилями*. Так, $x = x_1$ на рис. 2-3, а есть 25%-ная квантиль, так как площадь под кривой $p(x)$ слева от нее составляет 25% всей площади, а справа — 75%. Между x_1 и x_3 , т. е. 25%- и 75%-ной квантилями, которые принято называть *сгибами* (или *квартилями*) данного распределения, заклю-

чено 50% всех возможных значений погрешности, а остальные 50% лежат вне этого промежутка.

Медиана ($x = x_2$ на рис. 2-3, а) — это 50%-ная квантиль, так как она делит площадь под кривой $p(x)$ на две равные части.

На рис. 2-3, б $x = x_3$ есть 5%-ная квантиль, так как площадь под кривой $p(x)$ слева от нее составляет 5% всей площади. Соответственно значения x_1, x_2, x_6 и x_7 на рис. 2-3, б — это 1%-, 2,5%-, 97,5%- и 99%-ная квантили. Их удобно обозначать соответственно как $x_{0,01}, x_{0,025}, x_{0,975}$ и $x_{0,99}$. Интервал значений x между $x_3 = x_{0,05}$ и $x_4 = x_{0,95}$ охватывает 90% всех возможных значений случайной величины и называется *интерквантильным промежутком с 90%-ной вероятностью*. Его протяженность $d_{0,9} = x_{0,95} - x_{0,05}$. Интерквантильный промежуток $d_{0,95} = x_{0,975} - x_{0,025}$ включает в себя 95% всех возможных значений случайной величины и т. д.

На основании такого подхода вводится понятие квантильных оценок погрешности, т. е. значений погрешности с заданной доверительной вероятностью P_d , как границ интервала неопределенности $\pm \Delta_d = \pm d_d/2$, на протяжении которого встречается P_d процентов всех значений погрешности, а $1 - P_d$ процентов общего числа их значений остаются за границами этого интервала.

Таким образом, доверительное значение случайной погрешности есть ее максимальное значение с указанной доверительной вероятностью P_d , т. е. сообщение, что часть реализаций погрешности с вероятностью $1 - P_d = q$ может быть и больше указанного значения погрешности.

Так как квантили, ограничивающие доверительный интервал погрешности, могут быть выбраны различными, то при сообщении такой оценки должно одновременно обязательно указываться значение принятой доверительной вероятности P_d . Удобнее всего для этого обозначение доверительной погрешности снабжать индексом, численно равным принятой доверительной вероятности, т. е. писать, например $\Delta_{0,9}$ при $P_d = 0,9$, $\Delta_{0,95}$ при $P_d = 0,95$ и т. д.

Исторически сложилось так, что в разных областях знаний используют различные значения доверительной вероятности, равные 0,5; 0,8; 0,9; 0,95 и 0,99. Так, в высокоответственной области расчета артиллерийской стрельбы общепринятой является так называемая *срединная ошибка*, т. е. погрешность с доверительной вероятностью $P_d = 0,5$, когда 50% всех возможных отклонений меньше ее, а другие 50% — больше (см. рис. 2-3, а). Доверительная вероятность $P_d = 0,8$ является общепринятой в теории и практике оценки надежности средств автоматики, электронной и измерительной техники.

Погрешность $\Delta_{0,9}$ обладает тем уникальным свойством, что для широкого класса наиболее употребительных законов распределения вероятностей (см. § 2-6) только она имеет однозначное соотношение со средним квадратическим отклонением в виде $\Delta_{0,9} =$

$= 1,6\sigma$ вне зависимости от вида закона распределения. Поэтому ГОСТ 11.001—73 при отсутствии данных о виде закона распределения для определения двусторонней доверительной вероятности предписывал использовать только $P_d = 0,9$.

При наличии у прибора, кроме чисто случайной составляющей погрешности, еще и систематической погрешности θ выход возможных значений погрешности за границы доверительного интервала $\pm(\theta + \Delta_{0,9})$ становится практически односторонним (см. § 3-4). Для односторонней вероятности выхода за пределы интервала $\pm\Delta_d$ при отсутствии данных о виде закона распределения ГОСТ 11.001—73 предписывал использование доверительной вероятности $P_d = 0,95$. Доверительная вероятность $P_d = 0,99$ используется лишь при указании погрешности первичных и рабочих эталонов.

Достоверность определения доверительного значения погрешности по экспериментальным данным. Достоинство доверительного значения погрешности состоит в том, что оно может быть достаточно просто оценено прямо по экспериментальным данным. Пусть проведена серия из n измерений. Из наблюдавшихся n случайных погрешностей составляют вариационный ряд, располагая их в порядке возрастания: $\Delta_{(1)} \leq \Delta_{(2)} \leq \Delta_{(3)} \leq \dots \leq \Delta_{(n)}$. Далее используется предположение, что каждый из членов вариационного ряда является оценкой соответствующих квантилей, которые делят весь интервал возможных вероятностей (от 0 до 1) на $n + 1$ частей с равными значениями вероятности, иными словами, вероятности попадания значений погрешности в каждый из интервалов $(-\infty, \Delta_{(1)})$, $(\Delta_{(1)}, \Delta_{(2)})$, ..., $(\Delta_{(n-1)}, \Delta_{(n)})$ и $(\Delta_{(n)}, +\infty)$ предполагаются одинаковыми, а следовательно, равными $1/(n + 1)$. Отсюда каждое из наблюдавшихся значений $\Delta_{(i)}$ может быть принято как оценка $[1/(n + 1)] \cdot 100\%$ -ной квантили.

Таким образом, практическое определение Δ_d сводится к тому, что из всех полученных отсчетов отбрасываются наиболее удаленные от центра, а следовательно, самые ненадежные отсчеты. Если при переменном n отбрасывается постоянная относительная доля всех отсчетов, то определяемое по крайним членам оставшегося вариационного ряда значение Δ_d , в отличие от Δ_m , с ростом длины n серии отсчетов не возрастает, а стабилизируется и оказывается тем более устойчивым, чем больше объем выборки n , не уступая по простоте своего определения «максимальному» значению Δ_m .

При этом следует иметь в виду, что по ограниченным экспериментальным данным мы получаем не точные доверительные значения, а лишь их приближенные значения — *оценки*. Достоверность квантильных оценок резко повышается с понижением значений P_d , а при постоянном P_d — с ростом числа отсчетов n . Поэтому квантильные оценки с большими доверительными вероятностями могут быть найдены только при большом числе отсчетов. Действительно, так как вариационный ряд из n членов определяет

границы $n + 1$ интервалов, вероятность попаданий в которые принимается нами одинаковой, то при отбрасывании лишь интервалов $(-\infty, \Delta_{(1)})$ и $(\Delta_{(n)}, +\infty)$ оценка погрешности может быть определена с доверительной вероятностью, не большей, чем

$$P_d \leq (n - 1)/(n + 1).$$

При небольших объемах выборки n фактическая доверительная вероятность может быть существенно меньшей, т. е. достоверность оценки Δ_d , найденной таким путем, очень мала. Для определения оценки Δ_d с большей достоверностью с каждого из концов вариационного ряда должны быть отброшены не только пустые интервалы от $-\infty$ до $\Delta_{(1)}$ и от $\Delta_{(n)}$ до $+\infty$, но и какое-то число фактических отсчетов. Располагая рядом из n отсчетов и отбрасывая с каждого из концов ряда по $n_{отб}$ отсчетов, можно определить Δ_d с доверительной вероятностью, не большей, чем

$$P_d \leq (n - 1 - 2n_{отб})/(n + 1). \quad (2-1)$$

Отсюда число отсчетов n , необходимое для определения по экспериментальным данным Δ_d с заданной вероятностью P_d , будет не меньшим, чем

$$n \geq (1 + P_d + 2n_{отб})/(1 - P_d) \approx [2(1 + n_{отб})]/(1 - P_d), \quad (2-2)$$

и для различных значений P_d и $n_{отб} = 1$ приведено ниже:

P_d	0,8	0,9	0,95	0,98	0,99	0,995	0,997
n	20	40	80	200	400	800	1333

Таким образом, по экспериментальным данным легко определить значение Δ_d лишь с доверительной вероятностью $P_d \leq 0,95$ ($n \approx 80$), а определение $\Delta_{0,99}$ или $\Delta_{0,997}$ практически трудноосуществимо (нужно $n \geq 400 \div 1333$). При этом необходимо обратить внимание читателя на то, что объем выборки n , рассчитанный по формуле (2-2), обеспечивает лишь выполнение неравенства (2-1), т. е., взяв, например, выборку объемов $n = 80$ и отбросив с каждой стороны по одному отсчету, получим, что доверительная вероятность не может быть большей, чем 0,95. При этом нет никаких оснований утверждать, что она равна 0,95 (так же как утверждать, что она равна 0,8 или 0,3).

Тем не менее очень часто доверительные погрешности рассчитывают, вводя ничем не обоснованное предположение о том, что вид закона распределения погрешностей будто бы точно известен. В частности, используют прием, заключающийся в вычислении по небольшой выборке в 20—30 отсчетов оценки среднего квадратического отклонения σ , а затем указывают погрешность с доверительной вероятностью $P_d = 0,997$, равную $\Delta_{0,997} = 3\sigma$ на основании предположения о нормальности закона распределения.

Из приведенного выше анализа ясно, что такой прием является некорректным вне зависимости от того, допускается ли он сознательно или неосознанно. Следует иметь в виду, что реальные законы распределения погрешностей приборов весьма разнообразны и часто очень далеки от нормального. (Это далее будет подробно рассмотрено.) Для установления действительного хода кривой распределения на ее краях необходимо проведение испытаний, число которых должно быть тем больше, чем большим выбирается значение доверительной вероятности [см. формулу (2-2)]. При малом числе отсчетов (20—30) какие-либо сведения о ходе кривой в области квантилей, соответствующих $P_d = 0,95 \div 0,99$ (не говоря уже о $P_d = 0,997$), отсутствуют и утверждения о ходе кривой распределения в этой неисследованной области лишены каких-либо оснований.

Основным недостатком доверительного значения погрешности Δ_d при произвольно выбираемых P_d , как и «максимальной» погрешности Δ_m , является невозможность их суммирования, так как доверительный интервал суммы не равен сумме доверительных интервалов слагаемых. Поэтому приведенные выше применительно к Δ_m рассуждения остаются в силе и для Δ_d .

Среднее квадратическое отклонение σ случайной величины (сокращенно с. к. о.). Это положительное значение квадратного корня из ее дисперсии

$$\sigma = \sqrt{D} = \sqrt{\int_{-\infty}^{+\infty} (x - X_d)^2 p(x) dx},$$

где D — дисперсия, т. е. второй центральный момент случайной величины, а $p(x)$ — плотность распределения. Для определения оценок дисперсии по экспериментальным данным пользуются соотношением

$$D = \sum_{i=1}^n \frac{(x_i - X_d)^2}{n-1},$$

где x_i — значения отдельных отсчетов; X_d — координата центра распределения; n — объем выборки.

Отсюда оценка с. к. о. определяется как

$$\sigma = \sqrt{\sum_{i=1}^n (x_i - X_d)^2 / (n-1)}.$$

Основным достоинством оценки разброса случайных величин средним квадратическим значением σ является возможность определения дисперсии суммы статистически независимых величин

как $D_\Sigma = \sum_{i=1}^n D_i$ или $\sigma_\Sigma^2 = \sum_{i=1}^n \sigma_i^2$ независимо от разнообразия

законов распределения каждой из суммируемых величин и деформации знаков распределения при образовании композиций.

Таким образом, для того чтобы отдельные составляющие погрешности средств измерений можно было суммировать расчетным путем, они должны быть предварительно представлены своими средними квадратическими значениями σ , а не «максимальными» Δ_m или доверительными Δ_d значениями. При этом открывается возможность расчетным путем не только складывать любое число составляющих погрешности, что необходимо при анализе точности косвенных измерений или сложных измерительных устройств, но и достаточно точно вычитать погрешности, что необходимо при синтезе методов измерений или сложных устройств с заданной результирующей погрешностью. Действительно, если $\sigma_z = \sqrt{\sigma_1^2 + \sigma_2^2}$, то $\sigma_2 = \sqrt{\sigma_z^2 - \sigma_1^2}$. Это правомерно, однако, как выше подчеркивалось, для независимых случайных величин. Суммируемые или вычитаемые составляющие погрешности могут быть иногда и взаимно коррелированными. В этом случае приведенные соотношения заметно усложняются, что будет более подробно рассмотрено в § 3-1.

2-3. НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ИНФОРМАЦИИ

Вероятностная теория информации является дальнейшим развитием теории вероятностей применительно к процессам получения и передачи информации и, в частности, к процессам измерения. «Основная идея теории информации, — писал К. Шеннон в работе [50], — состоит в том, что с информацией можно обращаться почти так же, как с такими физическими величинами, как масса или энергия». Но «почти так же» не означает «точно так же». Действительно, характеристикой транспортной способности тока с произвольной формой кривой при переносе вещества (например, в электролитической ванне) является его среднее значение

$$I_{\text{ср}} = \frac{1}{T} \int_0^T i(t) dt. \quad (2-3)$$

Аналогом такого линейного функционала в теории погрешностей является определение интервала неопределенности результата измерения с помощью квантильной оценки с заданной вероятностью, когда за интервал неопределенности принимается интервал, в который попадает просто определенный процент всех наблюдаемых отсчетов.

При использовании того же тока для транспортировки не вещества, а энергии, его оценка в виде $I_{\text{ср}}$ оказывается неправомерной, так как количество переносимой энергии обусловлено

не средним, а действующим значением, определяемым другим функционалом в виде

$$I^2 = \frac{1}{T} \int_0^T i^2(t) dt. \quad (2-4)$$

Аналогом этого квадратического функционала в теории погрешностей является функционал для дисперсии, а аналогом действующего значения переменной составляющей тока — понятие с. к. о. σ погрешности.

Однако, как утверждает Н. Винер в работе [62], «информация есть информация, а не вещество и не энергия». Поэтому если функционал (2-3) неправилен для описания транспортировки энергии, а функционал (2-4) — для транспортировки вещества, то тем меньше оснований считать их строго правомерными для характеристики процессов «транспортировки информации». Видимо, для однозначной характеристики процессов передачи информации существует свой, особый функционал.

Действительно, исходя из исторического опыта установления интервала неопределенности результатов измерений следует, что этот интервал не имеет однозначного соотношения с оценкой с. к. о. σ , т. е. энергетическая оценка, оценка мощности помехи не определяет однозначно ее мешающее, дезинформирующее влияние. Для коррекции этой неоднозначности и приходится при разных законах распределения принимать разные значения доверительной вероятности, пытаясь интуитивно учесть еще какое-то свойство погрешности, кроме ее мощности, характеризующей с. к. о. σ .

Так, при равномерном законе распределения погрешности интервал неопределенности принимают равным всей ширине этого распределения, т. е. полагают доверительную вероятность $P_d = 1$. При нормальном законе распределения обычно переходят к доверительной вероятности $P_d = 0,95$. Следуя этой тенденции, при распределении погрешности по еще более пологому экспоненциальному распределению Лапласа не логично ли перейти к $P_d = 0,9$, а для предельного пологого распределения Коши использовать $P_d = 0,8$?

Анализ дезинформационного действия случайных помех с различными законами распределения вероятностей действительно привел К. Шеннона к выводу, что вносимая помехой дезинформация определяется не только мощностью этой помехи, т. е. ее с. к. о. σ , но еще зависит от вида закона распределения этой помехи.

Формально это положение К. Шеннон сформулировал в виде своей 16-й теоремы [50], которая утверждает, что если помеха в вероятностном смысле не зависит от сигнала, то независимо

от закона распределения и мощности сигнала дезинформационное действие помехи определяется ее энтропией

$$H(X) = - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx, \quad (2-5)$$

т. е. функционалом, действительно отличным по своему виду от двух предыдущих (2-3) и (2-4).

Значение этого функционала для теории измерений стало ясно после того, как в работах [12, 13, 27—30, 32, 34, 36, 40] на его основании была разработана система понятий, аппарата анализа, классификация и методы суммирования составляющих погрешностей, которые и излагаются далее.

Информационное описание измерения. Согласно К. Шеннону, количество информации I определяется как разность энтропий: $I = H(X) - H(X/X_{из})$, где $H(X)$ — энтропия (мера неопределенности) измеряемой величины до ее измерения, а $H(X/X_{из})$ (эта запись читается как «энтропия X при условии $X_{из}$ ») — энтропия действительного значения x измеряемой величины (мера интервала неопределенности) вокруг полученного после измерения показания $X_{из}$, т. е. энтропия погрешности измерения.

Эти оценки неопределенности в виде энтропии до и после измерения могут быть вычислены по соотношению (2-5) на основании вероятностного описания ситуаций до и после измерения. Покажем это на конкретном примере. Пусть для измерения величины x был использован прибор со шкалой от X_1 до X_2 (например, амперметр со шкалой от -50 А до $+50$ А). Тогда вероятностное описание ситуации до измерения состоит в том, что вероятность получить показания прибора в интервалах от $-\infty$ до X_1 и от X_2 до $+\infty$ равна нулю, т. е. плотность распределения вероятностей $p(x)$ в этих интервалах также равна нулю. Следовательно, показание можно ожидать лишь в интервале от X_1 до X_2 и, если предположить, что оно с равной вероятностью может оказаться в любой части этого диапазона, то вероятностное описание ситуации до измерения изобразится равномерным распределением x в пределах от X_1 до X_2 , показанным на рис. 2-4, и может быть записано как

$$\begin{cases} p(x) = 1/(X_2 - X_1) & \text{при } X_1 \leq x \leq X_2; \\ p(x) = 0 & \text{при } x < X_1 \text{ и } x > X_2. \end{cases}$$

Отсюда энтропия $H(X)$ до измерения согласно (2-5)

$$H(X) = - \int_{X_1}^{X_2} \frac{1}{X_2 - X_1} \ln \frac{1}{X_2 - X_1} dx = \ln(X_2 - X_1).$$

Таким образом, до измерения интервал неопределенности предстоящего отсчета простирается от X_1 до X_2 , а шенноновская энтропия есть логарифмическая мера длины этого интервала.

После проведения измерения мы получаем отсчет X_n . Однако вследствие погрешности прибора, равной $\pm\Delta$, можем лишь утверждать, что действительное значение измеряемой величины лежит где-то в пределах интервала неопределенности шириной $d = 2\Delta$. Если прибор обладает погрешностью с равномерным распределением, то ситуация после измерения описывается распределением, показанным на рис. 2-4, с шириной $d = 2\Delta$ и плотностью $p(x) = 1/(2\Delta)$.

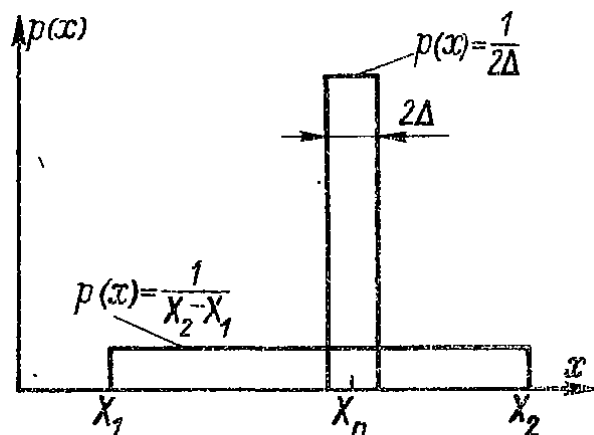


Рис. 2-4

Таким образом, в понятиях теории информации смысл измерения состоит в сужении интервала неопределенности от $X_2 - X_1$ до измерения до $d = 2\Delta$ — после измерения, т. е. в $N = (X_2 - X_1)/(2\Delta)$ раз.

Энтропия результата измерения после получения показания X_n

$$H(X/X_n) = - \int_{X_n - \Delta}^{X_n + \Delta} \frac{1}{2\Delta} \ln \frac{1}{2\Delta} dx = \ln 2\Delta,$$

т. е. также является логарифмической мерой нового интервала неопределенности.

Количество информации, полученное в результате измерения, равно разности исходной и оставшейся энтропий, т. е.

$$I = H(X) - H(X/X_n) = \ln(X_2 - X_1) - \ln(2\Delta) = \ln \frac{X_2 - X_1}{2\Delta} = \ln N.$$

Число $N = (X_2 - X_1)/(2\Delta)$ показывает, сколько интервалов неопределенности длиной $d = 2\Delta$ укладывается во всем диапазоне $X_2 - X_1$, т. е. какое *число различных градаций* измеряемой величины позволяет получить данный прибор или метод измерения.

Энтропийный интервал неопределенности. Соотношения $I = \ln N$ и $N = (X_2 - X_1)/d$ справедливы при любом законе распределения погрешности, если только интервал неопределенности d будет найден через энтропию. Поэтому В. И. Рабинович и М. П. Цапенко [36] предложили называть число N числом эквивалентных делений, различных в диапазоне $X_2 - X_1$ при данном законе $p(x)$ распределения погрешности, а d — эквивалентным в энтропийном смысле делением. Нам же представляется более удобным именовать величину N *числом различных градаций* измеряемой величины, а d — *энтропийным интервалом неопределенности* результата измерения.

Основное достоинство информационного подхода к математическому описанию случайных погрешностей состоит в том, что

размер энтропийного интервала неопределенности может быть вычислен строго математически для любого закона распределения погрешности как величина, стоящая под знаком логарифма в выражении для энтропии $H(X/X_{\Pi})$, устраняя тем самым исторически сложившийся произвол, неизбежный при волевом назначении различных значений доверительной вероятности.

Покажем это на конкретном примере, заимствованном из работы К. Шеннона [50]. Так, например, для нормально распределенной погрешности

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right);$$

$$\ln p(x) = -\ln(\sigma \sqrt{2\pi}) - x^2/(2\sigma^2).$$

Отсюда энтропия погрешности

$$\begin{aligned} H(X/X_{\Pi}) &= - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx = \int_{-\infty}^{+\infty} p(x) \left(\ln \sigma \sqrt{2\pi} + \frac{x^2}{2\sigma^2} \right) dx = \\ &= \ln \sigma \sqrt{2\pi} \int_{-\infty}^{+\infty} p(x) dx + \frac{1}{2\sigma^2} \int_{-\infty}^{+\infty} x^2 p(x) dx. \end{aligned}$$

Учитывая, что $\int_{-\infty}^{+\infty} p(x) dx = 1$ и по определению дисперсии $\int_{-\infty}^{+\infty} x^2 p(x) dx = \sigma^2$, получаем $H(X/X_{\Pi}) = \ln \sigma \sqrt{2\pi} + 1/2 =$
 $= \ln \sigma \sqrt{2\pi} + \ln \sqrt{e} = \ln \sigma \sqrt{2\pi e}$, т. е. интервал неопределенности d результата измерения, найденный через энтропию в соответствии с теорией информации, однозначно (без каких-либо предположений о выборе уровня доверительной вероятности) равен $d = \sqrt{2\pi e} \sigma \approx 4,133\sigma$, а число различных градаций результата измерения при равномерном распределении вероятности различных значений измеряемой величины

$$N = (X_2 - X_1)/d = (X_2 - X_1)/(4,133\sigma).$$

Подобным же образом энтропийный интервал неопределенности результата измерения может быть однозначно найден для любого выраженного аналитически закона распределения погрешности. Например, при распределении погрешности по треугольному закону Симпсона $H(X/X_{\Pi}) = \ln(\sqrt{6e}\sigma)$ и $d = \sqrt{6e}\sigma \approx 4,04\sigma$.

Разделение диапазона $X_2 - X_1$ на отдельные различимые градации на основании формальных положений теории информации в виде функционала (2-5) для энтропии представлено на рис. 2-5. Здесь диапазон $X_2 - X_1$ разбит на интервалы длиной d , вычис-

ленные указанным выше путем, и относительно центра каждого такого интервала, как начала координат, построена кривая соответствующего закона распределения погрешности: равномерного, треугольного и нормального. Отсюда видно, что следствие шенноновского определения энтропии помехи состоит в том, что только при равномерном распределении погрешности границы интервалов неопределенности, логарифм числа которых есть количество

получаемой при измерении информации $I = \ln N$, совпадают с границами распределения погрешности, т. е. отдельные полосы погрешностей лишь соприкасаются между собой. При треугольном, а тем более при неограниченных распределениях интервалы неопределенности определяются лишь той частью распределения, где сосредоточена основная масса этих погрешностей.

Таким образом, то различие в интервалах неопределенности при равномерном распределении погрешности и при неограниченных распределениях погрешности, которое исторически пытались преодолеть назначением соответствующих значений доверительной вероятности, математически строго и наглядно описывается при использовании в теории погрешностей информационного подхода.

Энтропийное значение случайной погрешности. При практическом использовании изложенного информационного подхода для оценки точности результатов измерений привычнее оперировать не со значениями энтропийного интервала неопределенности результата измерения d , а с половиной этого интервала, присвоив ей согласно работе [30] наименование *энтропийного значения погрешности* Δ_9 .

Формальным определением энтропийного значения случайной величины являются соотношения

$$H(X/X_H) = \ln d = \ln(2\Delta_9),$$

отсюда $d = 2\Delta_9 = e^{H(X/X_H)}$ и $\Delta_9 = \frac{1}{2} e^{H(X/X_H)}$.

Соотношение между энтропийным Δ_9 и средним квадратическим σ значениями погрешности различно для разных законов распределения, и его удобно характеризовать значением энтропийного коэффициента $k = \Delta_9/\sigma$ данного закона распределения.

Так, для равномерного распределения $\Delta_9 = \Delta_m = \sqrt{3}\sigma \approx 1,73\sigma$ и, следовательно, $k = 1,73$. Для нормального распределения, как было показано выше, $\Delta_9 = \sigma\sqrt{2\pi e}/2 = \sigma\sqrt{\pi e}/2 =$

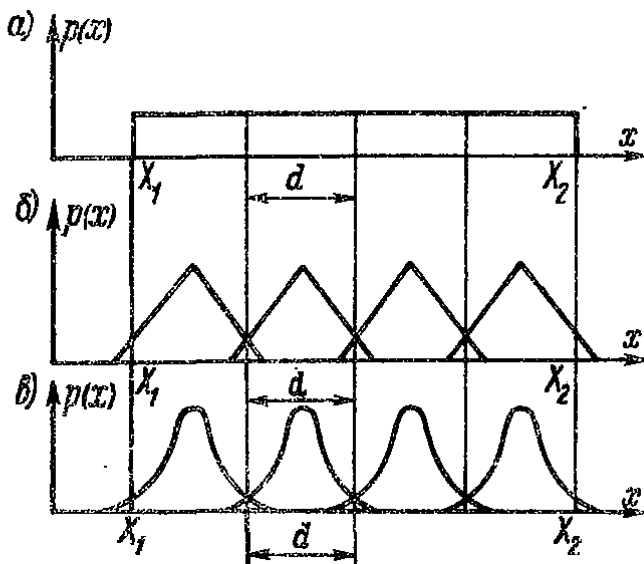


Рис. 2-5

Таблица 2-1

α	v	x	P_9	$\hat{P}_9$	$\nabla, \%$
∞	1,8	0,745	1,00	1,00	0
2	3	0,577	0,961	0,96	-0,1
1	6	0,408	0,934	0,93	-0,4
0,707	10,8	0,304	0,925	0,917	-0,8
0,5	25,2	0,199	0,896	0,906	+1,0
Коши	00	0	0,899	0,899	0

$= 2,066\sigma$ и $k = 2,066$. Для треугольного распределения Симпсона $k = \sqrt{6e}/2 \approx 2,02$, для распределения Лапласа $k = 1,93$, для арксинусоидального распределения $k = \pi/\sqrt{8} = 1,11$ и т. д.

Еще К. Шеннон показал, что максимальное возможное значение энтропийного коэффициента $k = 2,066$ имеет нормальное распределение. Поэтому для наиболее часто встречающихся на практике распределений энтропийное значение погрешности колеблется от $\Delta_9 = 1,11\sigma$ у арксинусоидального распределения до $\Delta_9 = 2,066\sigma$ у нормального распределения и при известном законе распределения может быть найдено как $\Delta_9 = k\sigma$.

Соотношение между энтропийным и доверительным значениями погрешностей. Как было видно из рис. 2-5, энтропийный интервал неопределенности $d = 2\Delta_9$ охватывает лишь ту часть распределения, в которой сосредоточена основная часть возможных значений случайной погрешности, в то время как некоторая их доля остается за границами этого интервала. Поэтому для любого распределения может быть указано такое значение доверительной вероятности P_9 , при котором энтропийное и доверительное значения погрешности совпадают.

Впервые эта задача была поставлена и решена в общем виде В. Н. Ивановым [12]. В частном случае для класса экспоненциальных распределений и распределения Коши это соотношение может быть получено следующим образом. В табл. 2-1 приведены значения P_9 для этих распределений. Если эти значения P_9 нанести на график рис. 2-6 в функции от значений контрэксцесса x этих распределений, то они расположатся близко к некоторой плавной кривой, которую можно аппроксимировать формулой вида

$$\hat{P}_9 = 0,899 + 0,1818/v \approx 0,899 + x^2/5,5 \quad (2-6)$$

и использовать для перевода энтропийных значений погрешности в доверительные. Погрешности такой аппроксимации указаны в табл. 2-1.

О единицах количества информации, энтропии и относительной и абсолютной погрешности. Абсолютные погрешности Δ_m , Δ_d ,

σ , Δ , выражаются в единицах измеряемой величины. Если эти оценки применяются для выражения относительных или приведенных погрешностей, то при определении $\varphi = \Delta/x$ или $\delta = \sigma/x$ значения Δ или σ и значения x должны подставляться в одних и тех же единицах. Это правило остается в силе и при использовании соотношений теории информации.

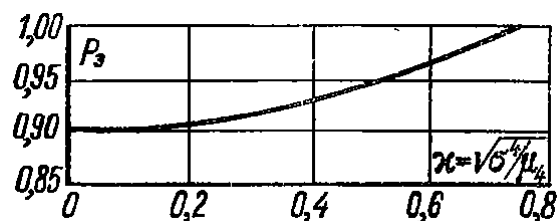


Рис. 2-6

Единицы энтропии и количества информации одни и те же, однако численное значение энтропии или информации зависит от основания используемых логарифмов. При теоретическом анализе, интегрировании и дифференцировании математических выражений наиболее удобно использовать натуральные логарифмы и тогда энтропия и количество информации получаются в так называемых *натуральных единицах* — *нитах*. При анализе цифровых машин и других устройств, работающих в двоичном коде, удобнее использовать двоичные логарифмы и тогда энтропия и количество информации получаются в *двоичных единицах* — *битах*. И, наконец, при анализе измерительных устройств, работающих в подавляющем большинстве случаев в десятичном коде, удобнее использовать десятичные логарифмы и *десятичные единицы* энтропии и информации — *дитах*.

Соотношения между этими единицами приблизительно следующие: 1 дит = 2,3 нит = 3,3 бит, 1 нит = 1,45 бит = 0,43 дит и 1 бит = 0,69 нит = 0,3 дит.

Количество информации в битах, дитах или нитах не зависит от единиц Δ и x , но они должны подставляться в соответствующие выражения для энтропии безразлично в каких, но обязательно в одних и тех же единицах. Действительно, $I = H(X) - H(\Delta) = \lg(X_2 - X_1) - \lg 2\Delta = \lg [(X_2 - X_1)/(2\Delta)]$, где X и Δ должны быть в одних и тех же единицах. В выражение для плотности распределения $p(\Delta)$ или в ее график единицы Δ входят по обоим осям, так как масштаб по оси абсцисс имеет единицу x , а по оси ординат — $1/x$. Поэтому численное значение энтропии

$$H(\Delta) = - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx$$
 или в простейшем случае $H(\Delta) = \ln(2\Delta)$ зависит от единиц Δ , совпадающих с единицами X . Энтропийное значение погрешности $\Delta_0 = 0,5e^{H(\Delta)}$ автоматически получается в тех же единицах, в которых была отложена погрешность по оси Δ графика плотности распределения, согласно соотношению $\Delta_0 = 0,5e^{H(\Delta)} = 0,5e^{\ln 2\Delta} = 0,5 \cdot 2\Delta = \Delta$.

Сделанные замечания относятся не только к результатам измерений, но и к результатам счета. Если, например, сумма в 10 000 руб. 00 коп. была сосчитана с точностью до 1 коп., то при определении количества информации, содержащейся в этом сообщении, обе величины должны быть выражены или в копейках

(тогда $X = 10^6$, $\Delta = 1$, $H(X) = 6$ дит, $H(\Delta) = 0$ дит и $I = 6 - 0 = 6$ дит), или в рублях (тогда $X = 10^4$, $\Delta = 10^{-2}$, $H(X) = 4$ дит, $H(\Delta) = -2$ дит и $I = 4 - (-2) = 6$ дит). Отсюда также следует, что результаты измерения, представленные числами 999, 0,999 или 0,000999, несут равные количества информации, так как первый дан с интервалом неопределенности $d = 1$ (последующий, не указанный знак, мог бы иметь значение от 0 до 9), второй — с $d = 10^{-3}$, а третий с $d = 10^{-6}$. Поэтому количество информации в первом равно $3 - 0 = 3$ дит, во втором $0 - (-3) = 3$ дит и в третьем $(-3) - (-6) = 3$ дит, т. е., попросту говоря, «три десятичных знака». Однако количество информации в результате, отсчитанном по цифровому вольтметру один раз в виде 999 000 мкВ, а другой раз в виде 000 999 мкВ, в первом случае равно 6 дит, а во втором 3 дит, так как указание значащих нулей равносильно тому, что последний знак результата отличается от нуля менее чем на 1 мкВ.

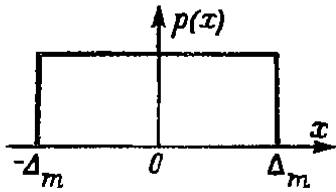
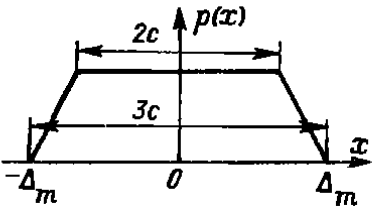
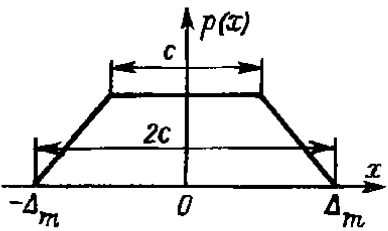
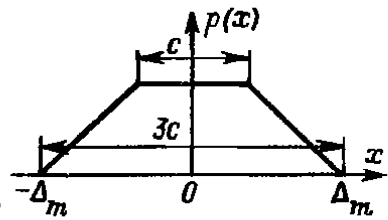
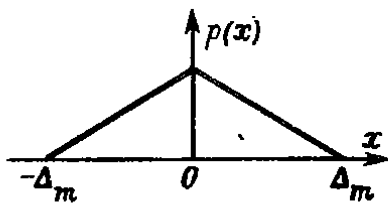
2.4. АНАЛИТИЧЕСКИЕ МОДЕЛИ И ПАРАМЕТРЫ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ

Для использования на практике вероятностного подхода к оценке погрешностей как средств, так и результатов измерений, прежде всего бывает необходимо установить для данной конкретной погрешности вид аналитической модели закона распределения. Распределения же достаточно разнообразны: одни ограничены, другие не ограничены, одни имеют плоскую вершину, другие круглую, третьи острую, а иные и две круглых или острых вершины.

Дж. Кендал и А. Стьюарт [14, с. 54] предлагали классифицировать формы распределений на пять типов: 1) симметричные одномодальные, 2) симметричные двухмодальные, 3) косые, 4) крайне косые и 5) все остальные. Применительно к погрешностям достаточно рассмотреть лишь два первых типа распределений, но целесообразно подразделить их на более мелкие классы, а именно: трапецеидальные (т. е. плосковершинные), уплощенные (т. е. приближенно плосковершинные), семейство распределений Стьюдента (включая распределение Коши), экспоненциальные, двухмодальные кругло- и островершинные распределения, выделив отдельно класс арксинусоидальных распределений.

Класс трапецеидальных распределений. Как было показано на рис. 2-2, а, трапецеидальное распределение образуется как композиция распределений при суммировании двух равномерно распределенных случайных величин. Поэтому равномерное распределение — это предельный случай трапецеидального, когда одна из суммируемых случайных величин исчезающе мала по сравнению с другой. Однако если ширина b меньшего из суммируемых распределений (см. рис. 2-2, а) даже в 5—10 раз меньше ширины a более широкого распределения и ее вес в дисперсии

Таблица 2-2

Но- мер пп.	Вид распределения	b/a	Δ_m/σ	e	κ	k
1		0	$\sqrt{3} \approx 1,73$	1,8	0,745	1,73
2		1/5	$\sqrt{4,15} \approx 2,04$	1,9	0,728	1,83
3		1/3	$\sqrt{4,8} \approx 2,19$	2,016	0,704	1,94
4		1/2	$\sqrt{5,2} \approx 2,32$	2,184	0,677	2,00
5		1	$\sqrt{6} \approx 2,45$	2,4	0,645	2,02

суммарного распределения пренебрежимо мал (составляет всего 1/25—1/100), тем не менее влияние на форму и параметры суммарного распределения этой, казалось бы, ничтожной добавки оказывается весьма существенным.

Для иллюстрации этого свойства композиций равномерных распределений в п. 1 табл. 2-2 приведены параметры равномерного, а в п. 2 — трапецидального распределения, образованного из двух равномерных составляющих с соотношением $b/a = 1/5$. Среднее квадратическое отклонение этого суммарного распределения всего на 2% больше, чем с. к. о. распределения с шириной a , но форма распределения приобретает вид трапеции с отношением оснований 2 : 3, что влечет за собой изменение отношения Δ_m/σ от $\sqrt{3} \approx 1,73$ до $\sqrt{4,15} \approx 2,04$, существенно изменяются эксцесс e

(с 1,8 до 1,9) и энтропийный коэффициент k (с 1,73 до 1,83).

В пп. 3—5 табл. 2-2 также приведены параметры трапецидальных распределений с отношением b/a , равным $1/3$, $1/2$ и 1 , которые потребуются в дальнейшем при решении задач расчетного суммирования погрешностей.

Равномерное распределение имеют: погрешность квантования в цифровых приборах, погрешность от округления при расчетах, при отсчете показаний аналоговых приборов, погрешность от трения в стрелочных приборах с креплением подвижной части на кернах и подпятниках, а также в самоуравновешивающихся мостах и потенциометрах со следящим электромеханическим приводом, погрешность определения момента времени для каждого из концов временного интервала в электронных цифровых хронометрах и частотомерах и т. д.

Суммируясь между собой, эти погрешности образуют трапецидальные распределения с разным отношением оснований трапеции. Так, например, общая погрешность протяженности временного интервала в электронных цифровых частотомерах оказывается распределенной по треугольному закону Симпсона, так как образуется из двух равных равномерно распределенных погрешностей определения его концов.

Равномерное распределение имеют дополнительные погрешности от колебания влияющих величин. Так как функции влияния принимаются, как правило, линейными, а коэффициенты влияния — постоянными, то распределение вероятностей дополнительной погрешности Δ_Θ как неслучайной (систематической) линейной функции случайного аргумента Θ повторяет с масштабом (по оси $\Delta_\Theta = \psi\Theta$) в виде коэффициента влияния ψ закон распределения вероятностей влияющей величины Θ .

Так, например, изменение напряжения питания вследствие постепенного разряда гальванических источников тока можно приближенно считать линейной функцией времени. Поэтому, полагая, что измерения могут быть с равной вероятностью проведены в любой момент времени, а стало быть при любом из значений питающего напряжения, закон распределения возникающей от этого погрешности можно считать равномерным.

Также можно считать равномерным распределение погрешности от изменения температуры окружающей среды для приборов, работающих в цеховых или лабораторных условиях при односменной работе. За ночь помещение остывает, например, до 20°C , а в течение рабочей смены нагревается, например, до 24°C . Поэтому распределение вероятности различных температур окружающей среды оказывается равномерным со средним значением 22°C и максимальным значением отклонения $\Delta(\Theta) = \pm 2^\circ\text{C}$. Умножая эти значения на соответствующий коэффициент влияния, получаем параметры распределения возникающей при этом температурной погрешности.

В подобных же условиях оказывается, например, датчик, установленный на двигателе внутреннего сгорания. До запуска двигатель имеет температуру $+20^{\circ}\text{C}$. После запуска нагревается до $+80^{\circ}\text{C}$. Кривая изменения температуры двигателя во времени вероятнее всего является экспонентой. Но если испытания проводятся с перерывами в работе двигателя, то начальные участки этих экспонент можно считать близкими к прямым линиям и распределение вероятностей различных температур от 20 и до 80°C приближенно считать равномерным, а среднюю температуру $(20 + 80)/2 = 50^{\circ}\text{C}$ принять за центр распределения.

Колебания напряжения питания переменным током от сети крупных энергосистем подчиняются приблизительно треугольному распределению. Поэтому если известно, что питающее напряжение 220 В колеблется в пределах 5% , то его закон распределения следует считать треугольным с максимальным отклонением $\pm 11\text{ В}$.

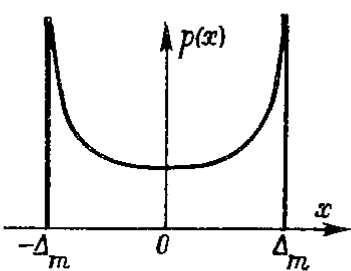
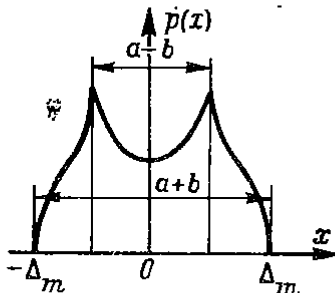
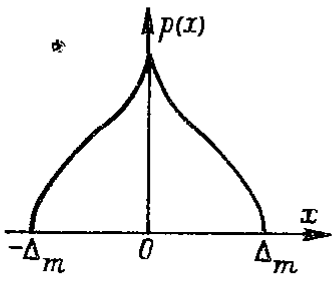
Класс арксинусоидальных распределений. Одной из составляющих погрешности, характерной для электрических средств измерений как электрических, так и неэлектрических измеряемых величин, является погрешность от наводки на входе прибора или линии связи синусоидального напряжения силовых цепей с частотой 50 или 400 Гц . Эта помеха, складываясь с полезным сигналом, создает, как правило, аддитивную погрешность и в ряде случаев ограничивает порог чувствительности измерительного устройства.

Распределение такой погрешности называется *арксинусоидальным* и имеет плотность

$$p(x) = [\pi X_m \sqrt{1 - (x/X_m)^2}]^{-1} = (\pi \sqrt{X_m^2 - x^2})^{-1}. \quad (2-7)$$

Кривая такого распределения была приведена на рис. 2-1, *ж* и повторена в п. 1 табл. 2-3, где указаны его основные параметры. Среднее квадратическое, т. е. действующее, значение синусоиды общеизвестно: $\sigma = X_m/\sqrt{2}$. Его четвертый момент $\mu_4 = \frac{3}{8}X_m^4$, эксцесс $\varepsilon = 1,5$ и контрэксцесс $\kappa = 0,816$. Энтропия этого распределения $H(x) = \ln(\pi X_m/2)$, следовательно, энтропийное значение $\Delta_s = \pi X_m/4$ и энтропийный коэффициент $k = \pi/(2\sqrt{2}) \approx 1,11$. В этой связи интересно отметить, что энтропийное значение синусоиды во столько же раз больше ее действующего значения ($k = 1,11$), во сколько раз ее действующее значение больше средневыпрямленного ($\pi/(2\sqrt{2}) = 1,11$).

На практике, однако, напряжение наводки на входе прибора или линии связи редко имеет чисто синусоидальную форму кривой. Чаще всего оно загрязнено присутствием высших гармоник. Распределение суммы двух синусоидально изменяющихся во времени с разными частотами величин является *композицией двух арксинусоидальных распределений*. Если одна из этих величин имеет размах, равный a , а другая — b , то их композиция обра-

Но- мер пп.	Вид распределения	b/a	Δ_m/σ	v	η	k
1		0	$\sqrt{2} \approx 1,41$	1,5	0,816	$\frac{\pi}{2\sqrt{2}} \approx 1,11$
2		1/3	$\frac{4}{\sqrt{5}} \approx 1,79$	1,72	0,752	1,76
3		1	2	2,25	0,667	1,88

зуется подобно тому, как это было показано для ограниченных распределений на рис. 2-2, а и в, и результирующее распределение получает вид фигуры, показанной в п. 2 табл. 2-3, с нижним основанием, равным $a + b$, и двумя пиками на расстоянии между собой, равном $a - b$. Параметры такого распределения приведены в п. 2 табл. 2-3. При $a = b$ расстояние между пиками становится равным нулю и распределение получает вид древнерусского шлема (п. 3 табл. 2-3).

Класс экспоненциальных распределений. Как показал И. А. Назаров в 1965 г. в работе [27], широкий класс симметричных распределений может быть описан единой аналитической моделью вида

$$p(x) = \frac{\alpha}{2\lambda\sigma\Gamma(1/\alpha)} \exp\left(-\left|\frac{x - X_{\text{ц}}}{\lambda\sigma}\right|^{\alpha}\right), \quad (2-8)$$

где $\lambda = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$; σ — с. к. о.; $X_{\text{ц}}$ — координата центра; $\Gamma(z)$ — гамма-функция; α — некоторая характерная для данного распределения постоянная — его показатель степени.

В иностранной литературе первое упоминание о такой модели содержится в диссертации В. М. Гентлемана (Gentleman W. M.

Robust estimation of multivariate location by minimizing p-th powers deviations. — Dissertation. — Princeton University and Memorandum. MM 65-1215-16, 1965).

Для иллюстрации влияния показателя степени α на форму описываемого распределения положим $X_n = 0$, а произведение $\lambda\sigma = 1$. Тогда

$$p(x) = \frac{\alpha}{2\Gamma(1/\alpha)} \exp(-|x|^\alpha) = A(\alpha) \exp(-|x|^\alpha), \quad (2-9)$$

где $A(\alpha)$ — нормирующий множитель распределения, зависящий от его показателя степени α .

При $\alpha < 1$ аналитические модели (2-8) и (2-9) описывают распределения с очень пологими спадами, близкие по своим свойствам к распределению Коши. При $\alpha = 1$ она соответствует распределению Лапласа, при $\alpha = 2$ — нормальному распределению Гаусса, при $\alpha \gg 2$ она описывает распределения, по своим свойствам близкие к трапецеидальным, и, наконец, при $\alpha \rightarrow \infty$ соответствует равномерному распределению.

Эта обобщенная модель интересна тем, что обычно в учебниках по теории вероятностей распределения Лапласа, нормальное и равномерное, рассматриваются разрозненно, без какой-либо взаимосвязи. Соотношения же (2-8) и (2-9) показывают, что все они являются представителями единого большого класса *экспоненциальных распределений*. При этом единственным параметром, характеризующим их форму, а следовательно, и их свойства, является показатель степени α описывающей их симметричной двусторонней экспоненты.

Действительно, эксцесс всех этих распределений выражается единой формулой также через показатель степени экспоненты α :

$$e = \Gamma(1/\alpha) \Gamma(5/\alpha) / [\Gamma(3/\alpha)]^2, \quad (2-10)$$

а контрэксцесс, соответственно,

$$\kappa = \Gamma(3/\alpha) / \sqrt{\Gamma(1/\alpha) \Gamma(5/\alpha)}. \quad (2-11)$$

Энтропийный коэффициент экспоненциальных распределений также является однозначной функцией α :

$$k = \frac{1}{\alpha} e^{1/\alpha} \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \Gamma(1/\alpha). \quad (2-12)$$

Таким образом, показатель степени α однозначно определяет все параметры формы этих распределений. При этом значения α могут быть не только равными 1, 2 и ∞ , что соответствует распределениям Лапласа, Гаусса и равномерному, но принимать и любые другие целые и дробные значения, т. е. модель (2-8) класса экспоненциальных распределений исключительно удобна для описания распределений погрешностей приборов и измерений.

Значения параметров таких распределений приведены

Таблица 2-4

Номер пп.	Вид распределения	α	ν	η	k
1	$p(x) = \frac{1}{48} e^{-\sqrt[4]{ x }}$	1/4	458	0,0467	0,085
2	$p(x) = \frac{1}{12} e^{-\sqrt[3]{ x }}$	1/3	107,25	0,0966	0,424
3	$p(x) = \frac{1}{4} e^{-\sqrt{ x }}$	1/2	25,2	0,199	1,35
4	$p(x) = \frac{1}{2} e^{- x }$ (распределение Лапласа)	1	6	0,408	1,92
5	$p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ (распределение Гаусса)	2	3	0,577	2,066
6	$p(x) = \frac{7}{2\Gamma(1/7)} e^{- x ^7}$	7	1,962	0,714	1,87
7	Равномерное распределение	∞	1,8	0,745	1,73

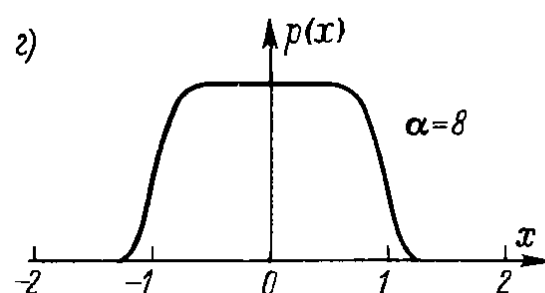
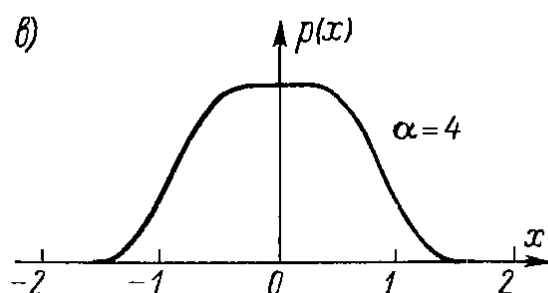
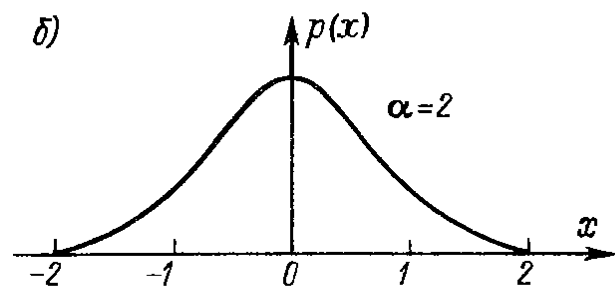
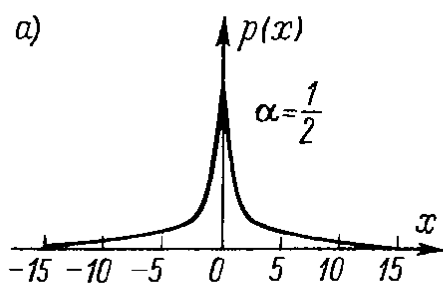


Рис. 2-7

в табл. 2-4, а на рис. 2-7 показано изменение формы этих распределений с изменением показателя степени α .

Доцентом высшего Машинно-электротехнического института (Варна, Болгария) З. Таушановым и его сотрудниками было подробно исследовано [41] распределение *погрешности градуировки* шкал аналоговых электроизмерительных приборов. Погрешности, допускаемые в процессе градуировки, для каждой градуируемой отметки шкалы прибора остаются, естественно, неизменными в течение всей жизни прибора, т. е. являются систематическими. Однако по совокупности всех отметок шкалы они являются случайными, так как для различных отметок шкалы могут быть как положительными, так и отрицательными или равными нулю. Это распределение погрешностей, допущенных при градуировке, изучалось З. Таушановым для приборов, изготовленных в ЧССР, ГДР, СССР и НРБ. Закон распределения этих погрешностей оказался одним и тем же. Он имеет вид кривой рис. 2-7, а и описывается уравнением $p(x) = 0,534 \exp(-|x|^2)$. Параметры этого распределения приведены в п. 6 табл. 2-4.

После опубликования в работе [27] формулы (2-8) и объединения всех этих распределений в единый большой класс экспоненциальных распределений казалось, что наконец найдена единая модель для описания всех распределений погрешностей приборов и результатов измерений. Действительно, она описывает и традиционно используемые нормальное и равномерное распределения. Описывает и распределение Лапласа, которому подчиняется распределение погрешностей потенциометров постоянного тока, высокоточных цифровых вольтметров, и т. п., т. е. всех тех приборов, погрешности которых определяются подгонкой номинальных значений элементов многоуровневой кодирующей сетки. Таким образом, казалось, что это универсальное обобщение описывает все разновидности распределений погрешностей приборов и измерений. Однако в работе [3] были собраны данные о параметрах 219 фактических распределений погрешностей, исследованных разными авторами, при измерении как электрических, так и неэлектрических величин самыми разнообразными (электрическими) приборами. В результате этого исследования оказалось, что 111 распределений, т. е. примерно 50%, действительно принадлежат классу экспоненциальных распределений. Но 63 распределения, т. е. 30%, имеют вид распределений с плоской вершиной и пологими длинными спадами и не могут быть описаны как экспоненциальные. Оставшиеся же 45 распределений, т. е. 20%, оказались двухмодальными, т. е. также не относящимися к классу экспоненциальных распределений.

Класс уплощенных распределений типа шапо (по-французски *le chapeau* — шляпа). Распределения такой формы, как это было показано на рис. 2-2, б и г, образуются как композиция равномерного и какого-либо из экспоненциальных распределений. Распределения класса шапо отличаются от экспоненциальных рас-

пределений с $\alpha > 2$ тем, что одновременно с почти плоской вершиной имеют длинные, медленно спадающие «хвосты», в то время как экспоненциальные распределения при $\alpha \gg 2$ обрываются тем круче, чем более плоской является их вершина.

Такие распределения погрешностей, как было показано в работе [8], характерны для цифровых средств измерений невысокой точности (например, каналов ИИС типа К-200 с цифровым вольтметром Ф203 класса точности 0,2/0,1, три знака отсчета). Они имеют широкую, почти плоскую вершину, соответствующую ступени квантования АЦП, а на краях спадают по интегральным кривым экспоненциальных распределений с α , равной от 2 до 1/4, т. е. являются композициями равномерного и указанных экспоненциальных распределений.

Основными параметрами, определяющими форму таких распределений, являются: показатель относительного содержания в композиции равномерной составляющей $C_p = \sigma_p / \sigma_{\text{экс}}$ (где σ_p — с. к. о. равномерной составляющей, а $\sigma_{\text{экс}}$ — с. к. о. экспоненциальной составляющей) и показатель степени α экспоненциальной составляющей.

При этом относительный вес дисперсии $\sigma_{\text{экс}}^2$ экспоненциальной составляющей в суммарной дисперсии композиции $\sigma_p^2 + \sigma_{\text{экс}}^2$, равный $p = \sigma_{\text{экс}}^2 / (\sigma_p^2 + \sigma_{\text{экс}}^2)$, может составлять всего несколько процентов, однако отличие распределения класса шапо от равномерного будет весьма существенным из-за наличия длинных, медленно спадающих «хвостов» такой композиции. Так, по данным работы [8], где приводятся результаты исследования более 100 распределений погрешностей в разных точках диапазона каналов ИИС К-200, вес дисперсии экспоненциальной составляющей изменяется от 0 до 13% при $\alpha = 1/3$ и до 5% при $\alpha = 1/4$.

Другой особенностью распределений класса шапо является то, что они при том же значении эксцесса имеют значения энтропийного коэффициента существенно меньшие, чем у экспоненциальных распределений. Вследствие этих существенных различий для описания распределений класса шапо не могут быть использованы аналитические модели экспоненциальных распределений.

Класс двухмодальных распределений. В практике измерений кривая плотности распределения погрешностей имеет достаточно выраженный максимум, совпадающий с координатой центра распределения, как это было показано на рис. 2-1, а, б, г или рис. 2-7, а, б. Такие распределения называются *одномодальными*. Трапецеидальные распределения (исключая треугольное) не имеют моды, т. е. являются *безмодальными*. Однако иногда погрешности оказываются распределенными с кривой плотности, имеющей два симметричных относительно центра максимума. Такие распределения называются *симметричными двухмодальными*.

В качестве аналитической модели для описания симметричных двухмодальных распределений может использоваться композиция дискретного двузначного распределения (см. рис. 2-1, в) и экспо-

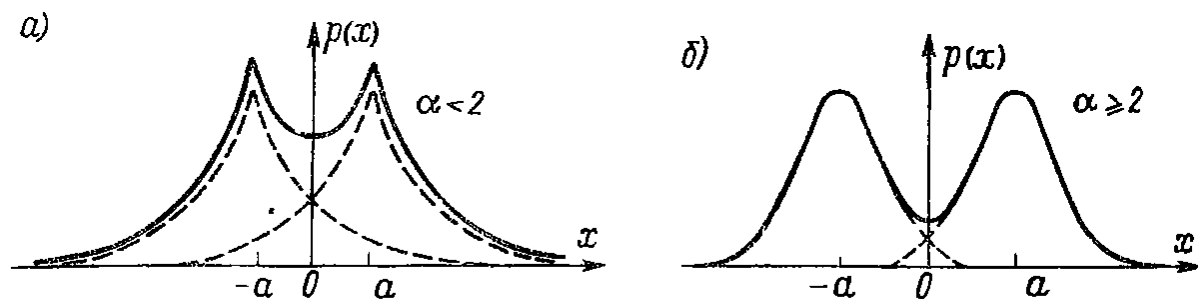


Рис. 2-8

ненциальных распределений с произвольным значением показателя степени α . Образование таких композиций показано на рис. 2-8. Дискретное двузначное распределение в 50% случаев имеет значение $x = -a$ и в 50% случаев — значение $x = +a$. Поэтому композиция его с экспоненциальным распределением складывается из суммы двух экспоненциальных, сдвинутых на $\pm a$ относительно центра (на рис. 2-8 они показаны штриховыми линиями) с площадью под каждой из кривых $P_1 = P_2 = 0,5$. Суммируясь при каждом из значений x , они и дают кривую плотности такой композиции. При показателе степени экспоненциальной составляющей $\alpha < 2$ получаются островершинные двухмодальные распределения (рис. 2-8, а), а при $\alpha \geq 2$ — кругловершинные (рис. 2-8, б).

Основными параметрами, определяющими форму таких распределений, являются показатель относительного содержания в композиции дискретной составляющей $C_d = \sigma_d / \sigma_{\text{экс}} = a / \sigma_{\text{экс}}$ (где σ_d — с. к. о. дискретной составляющей, равное ее полуразмаху a ; $\sigma_{\text{экс}}$ — с. к. о. экспоненциальной составляющей) и показатель степени α экспоненциальной составляющей.

В работе [3] показатель C_d предлагалось именовать *глубиной антимодальности* таких распределений, так как при $C_d = 0$ провал на вершине кривой плотности отсутствует, а при $C_d \neq 0$ этот провал тем глубже, чем больше C_d . У реально встречающихся распределений погрешностей глубина антимодальности C_d колеблется в пределах от 0 до 2, а показатель степени α экспоненциальной составляющей изменяется от $1/2$ до 2. Так, подробное исследование распределений погрешности высокоточных цифровых вольтметров типа Щ-1411 [7] показало, что они имеют вид, представленный на рис. 2-8, а, с параметрами $\alpha \approx 1$ и $C_d \approx 1,7$.

Кривую распределения, подобную приведенной на рис. 2-8, б, имеют погрешности от механического гистерезиса упругих элементов приборов и датчиков. При возрастании измеряемой величины превалируют отрицательные погрешности порядка $-a$, а при убывании — положительные порядка $+a$. В итоге общая кривая распределения приближается к двугорбой.

Подобное же двухмодальное распределение имеют температурные погрешности приборов, работающих в течение всего года на открытом воздухе. Значения температур атмосферного воздуха подробно изучены на протяжении столетий и публикуются в «Ме-

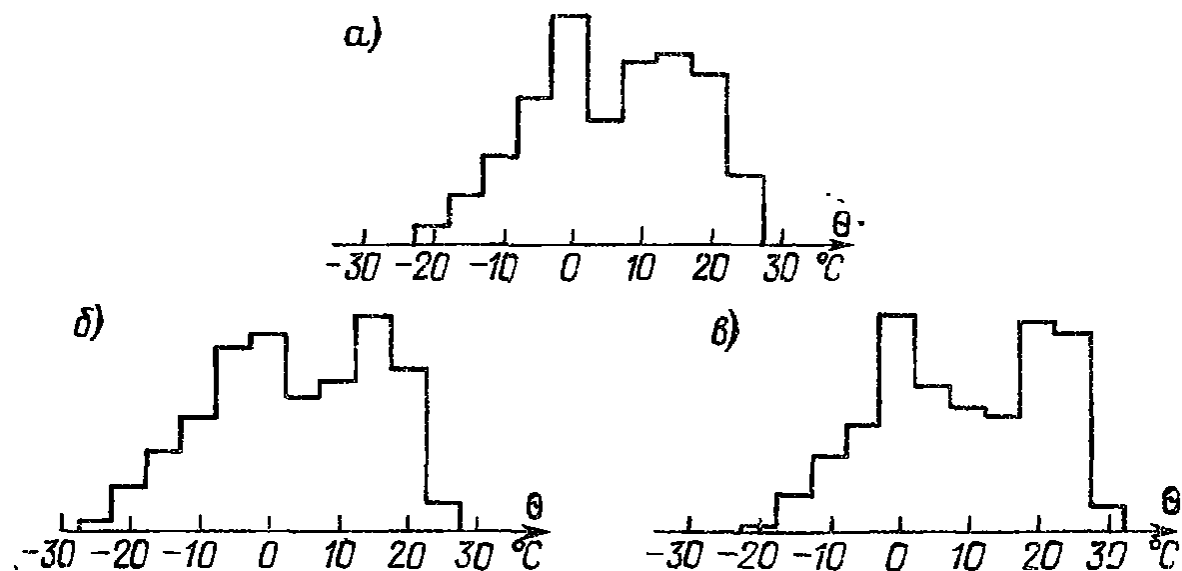


Рис. 2-9

трологическом ежегоднике». Закон распределения этих температур достаточно устойчив как для разных лет, так и для различных географических пунктов. В качестве примера приведены гистограммы распределения температур для Москвы (рис. 2-9, а), Пензы (рис. 2-9, б) и Астрахани (рис. 2-9, в). Распределения являются двухмодальными, но несколько асимметричными. Форма распределения и энтропийное значение отклонений от центра сохраняются практически постоянными для любых лет и географических пунктов: $k = 1,95$, $\Delta_0 = [20 \pm (1 \div 2)]^\circ\text{C}$, среднее же значение зависит от географического пункта (в Москве $+6^\circ\text{C}$, Пензе $+4^\circ\text{C}$, Астрахани $+9^\circ\text{C}$). Эти данные сведены в ГОСТ 16350—80.

Семейство законов распределения Стюдента. Эти законы распределения описывают плотность вероятности значений среднего арифметического, вычисленного по выборке из n случайных отсчетов из нормально распределенной генеральной совокупности. Применение этих распределений будет обсуждаться в гл. 4, посвященной методам статистической обработки многократных отсчетов. Здесь же обратим внимание читателя лишь на то, что это не один какой-то «закон распределения Стюдента», а целое семейство законов, так как вид этого распределения зависит от числа n отсчетов, по которым рассчитывается среднее значение.

В центрированном и нормированном виде семейство распределений Стюдента описывается выражением

$$p(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi} \Gamma\left(\frac{v}{2}\right) \left(1 + \frac{x^2}{v}\right)^{\frac{v+1}{2}}}, \quad (2-13)$$

где $\Gamma(z)$ — гамма-функция, а v — так называемое число степеней свободы, зависящее от числа n усредняемых отсчетов: $v = n - 1$.

Номер. вп.	Число n усредняе- мых отсче- тов	Число ν степеней свободы	σ	κ	k
1	5	4	∞	0	1,9003
2	6	5	9	0,333	1,9717
3	7	6	6	0,408	2,0053
4	8	7	5	0,447	2,013
5	11	10	4	0,500	2,047
6	∞	∞	3	0,577	2,066

Для нормированных распределений Стьюдента с числом степеней свободы $\nu \geq 4$ справедливы соотношения:

$$\sigma = \sqrt{\frac{n-1}{n-3}} = \sqrt{\frac{\nu}{\nu-2}}; \quad (2-14)$$

$$\varepsilon = \frac{\mu_4}{\sigma^4} = 3 \frac{n-3}{n-5} = 3 \frac{\nu-2}{\nu-4}; \quad (2-15)$$

$$\kappa = \sqrt{\frac{\nu-4}{3(\nu-2)}}, \text{ отсюда } \nu = \frac{4-6\kappa^2}{1-3\kappa^2}; \quad (2-16)$$

$$k = \frac{\sqrt{\pi(\nu-2)} \Gamma\left(\frac{\nu}{2}\right)}{2\Gamma\left(\frac{\nu+1}{2}\right)} e^{(\nu+1)\beta(\nu)},$$

где

$$\beta(\nu) = \sum_{m=0}^{+\infty} \frac{(-1)^m}{\nu+m}, \text{ т. е. } \beta(1) = \ln 2, \beta(2) = 1 - \ln 2,$$

$$\beta(\nu) = \left[1 - \frac{1}{2} + \frac{1}{3} \dots - \frac{(-1)^{\nu-1}}{\nu-1} - \ln 2 \right] (-1)^\nu.$$

Параметры распределений Стьюдента с $\nu \geq 4$ приведены в табл. 2-5.

Как видно из соотношения (2-14), особенность распределений Стьюдента состоит в том, что при $n \leq 3$ ($\nu \leq 2$) их с. к. о. σ становится равным бесконечности, т. е. дисперсионная оценка ширины разброса перестает работать (не существует). Она одинаково будет равна бесконечности у распределений как с бóльшим, так и с меньшим (например, в 10 раз) разбросом. Оценка четвертого момента, а следовательно, и эксцесса, перестает существовать еще раньше. Уже при $n = 5$ ($\nu = 4$) согласно формуле (2-15) $\varepsilon = \infty$ и поэтому не имеет смысла. Таким образом, классический аппарат моментов для оценки ширины и формы распределений Стьюдента с малым числом степеней свободы оказывается неработо-

способным и их ширина и форма могут быть оценены лишь с использованием доверительных и энтропийных оценок. Этим распределения Стюдента резко отличаются от всех других рассмотренных ранее законов распределений.

По значениям эксцесса (от $\varepsilon = 3$ до $\varepsilon = \infty$) распределения Стюдента с числом степеней свободы от $\nu = 4$ до $\nu = \infty$ совпадают с распределениями класса экспоненциальных с показателями степени от $\alpha = 0$ до $\alpha = 2$, но резко отличаются от них по значениям энтропийного коэффициента, который у распределений Стюдента значительно больше, чем у экспоненциальных распределений. Так, например, экспоненциальное распределение Лапласа с показателем степени $\alpha = 1$ имеет такой же эксцесс ($\varepsilon = 6$, $\kappa = 0,408$), как и распределение Стюдента с числом степеней свободы $\nu = 6$ (ср. п. 4 табл. 2-4 и п. 3. табл. 2-5), но его энтропийный коэффициент ($k = 1,92$) еще далек от предельного значения $k = 2,066$, в то время как у распределения Стюдента с $\nu = 6$ он равен $k = 2,0053$. По мере увеличения числа степеней свободы эти различия монотонно уменьшаются и при $\nu \rightarrow \infty$ распределение Стюдента стремится к нормальному с $\varepsilon = 3$, $\kappa = 0,577$ и $k = 2,066$.

Закон распределения Коши (кривая плотности вероятностей приведена на рис. 2-1, а) важен для теории оценки погрешностей результатов измерений тем, что ему подчиняется, например, распределение отношения двух нормально распределенных центрированных случайных величин.

В последнее время в измерительной технике получает все большее применение для измерений использование шумовых явлений (шумовые термометры, приборы на основе шумов Баркгаузена и др.). Но если, например, измерение постоянного активного электрического сопротивления производится на основе соотношения $R = U/I$ с использованием источника шумового напряжения, имеющего нормальное распределение, то при одновременном отсчете U и I их отношение U/I будет иметь распределение Коши, что влечет за собой ряд существенных особенностей.

Распределение Коши — это предельное распределение семейства распределений Стюдента с минимально возможным числом степеней свободы $\nu = 1$. Подставляя в выражение (2-13) $\nu = 1$, получим

$$p(x) = \frac{\Gamma(1)}{\sqrt{\pi} \Gamma(1/2) (1+x^2)} = \frac{1}{\pi (1+x^2)} \quad (2-17)$$

или в общем (т. е. ненормированном и нецентрированном) виде

$$p(x) = \frac{1}{a\pi \left[1 + \left(\frac{x - X_{\text{ц}}}{a} \right)^2 \right]}, \quad (2-18)$$

где a — параметр ширины, а $X_{\text{ц}}$ — координата центра этого распределения. Кривая плотности распределения Коши

(рис. 2-1, а) с первого взгляда кажется очень похожей на кривую плотности нормального распределения, однако в действительности это совсем не так, ибо ее свойства резко отличны от свойств экспоненциальных распределений. Так, дисперсия отсчетов при таком законе распределения вероятностей принципиально не может быть указана, так как определяющий ее интеграл расходится. На практике это означает, что оценка дисперсии и с. к. о., определяемые по полученным экспериментальным данным, будут неограниченно возрастать по мере увеличения объема n этих данных. Естественно, что использование такой оценки неправомерно. Четвертого момента μ_4 у распределения Коши также не существует, так как определяющий его интеграл расходится, как и интеграл, определяющий дисперсию. Эксцесс распределения Коши $\varepsilon = \mu_4/\sigma^4$ также равен бесконечности, а следовательно, контрэксцесс $\kappa = 1/\sqrt{\varepsilon} = 0$.

Оценка координаты центра X_n распределения Коши в виде среднего арифметического всех наблюдаемых отсчетов также неправомерна, так как ее рассеяние, определяемое как $\sigma_x = \sigma/\sqrt{n}$ (см. § 4-2), при $\sigma = \infty$ равно бесконечности, т. е. распределение Коши не имеет даже определенного значения математического ожидания.

Однако если по практически полученным экспериментальным данным, например при измерении активного сопротивления по падению напряжения на нем от шумового тока, построить кривую распределения плотности вероятности, то получим фигуру (рис. 2-1, а), ясно показывающую и положение центра распределения на числовой оси, и ширину разброса экспериментальных данных. Таким образом, шумовые измерения отнюдь не являются неправомерными. Но классический метод моментов теории вероятностей не способен дать оценку параметров таких распределений. Если вместо математического ожидания для определения координаты центра использовать медиану, то координата центра X_n таких отсчетов при достаточном объеме выборки будет надежно определена (см. § 4-3).

Оценку ширины разброса экспериментальных данных при подобных распределениях возможно произвести только на основе теории информации. Действительно, интеграл, определяющий энтропию, а следовательно, и энтропийный интервал неопределенности, и энтропийное значение погрешности, сходится и для таких распределений. Энтропия распределения Коши согласно выражениям (2-5) и (2-17)

$$\begin{aligned}
 H(X) &= - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx = - \int_{-\infty}^{+\infty} \frac{\ln \{1/[\pi (1+x^2)]\}}{\pi (1+x^2)} dx = \\
 &= 2 \int_0^{\infty} \frac{\ln \pi (1+x^2)}{\pi (1+x^2)} dx = \frac{2}{\pi} \int_0^{\pi/2} \frac{\ln \pi (1+\operatorname{tg}^2 t)}{1+\operatorname{tg}^2 t} \frac{dt}{\cos^2 t} =
 \end{aligned}$$

$$= \frac{2}{\pi} \ln \pi \int_0^{\pi/2} dt + \frac{2}{\pi} \int_0^{\pi/2} \ln \cos^2 t \, dt = \ln \pi + \frac{4\pi}{2\pi} \ln 2 = \ln(4\pi),$$

а при использовании вместо (2-17) выражения (2-18) получим $H(X) = \ln(4\pi a)$. Отсюда энтропийный интервал неопределенности при распределении Коши $d = 4\pi a$ и энтропийное значение погрешности $\Delta_0 = 2\pi a$. Границы энтропийного интервала неопределенности $\pm \Delta_0$ помечены на рис. 2-1, а жирными штрихами и согласно кривой рис. 2-6 или формуле (2-6) соответствуют доверительной вероятности $P_0 = 0,899 \approx 0,9$.

2-5. ТОПОГРАФИЧЕСКАЯ КЛАССИФИКАЦИЯ АНАЛИТИЧЕСКИХ МОДЕЛЕЙ СИММЕТРИЧНЫХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

В учебниках по теории вероятностей обычно (за исключением книги [14]) не дается какой-либо систематизации или классификации законов распределений. Приводятся примеры распределений (нормальное, треугольное, равномерное, Стюдента и т. д.) без какой-либо связи между собой. Поэтому у изучающего не складывается общей картины возможных разновидностей форм законов распределений. Проведенная в предыдущем параграфе попытка объединения распределений, необходимых для описания погрешностей приборов и результатов измерений, в шесть крупных классов не решает задачи четкой систематизации распределений по их форме, необходимой практику для выбора моделей законов распределения. Остается неясным, какие из них близки между собой по форме и при необходимости могут быть заменены друг другом, а какие, наоборот, далеки друг от друга, и т. д.

Наиболее удобной для практики была бы такая систематизация аналитических моделей законов распределения погрешностей, которая явно показывала бы их взаимную близость или удаленность, а в идеале позволяла бы оценить эту близость или удаленность численно. Эта систематизация, а на ее основе и классификация, могла бы быть построена, исходя из представлений хорошо разработанной к настоящему времени теории распознавания образов. Закон распределения $p(x)$ как функция характеризуется набором признаков $u_1, u_2, \dots, u_k$. Выбор характера признаков, их смысла, числа и формы аналитического представления ничем не ограничен. Признаки могут быть однородными и неоднородными, локальными (относящимися к отдельным точкам кривых распределений) и интегральными (выражающимися через интегралы, в подынтегральные выражения которых входит закон распределения). Чтобы признаки характеризовали только форму закона распределения, они должны быть безразмерными и не зависеть от смещения центра распределения. В качестве локальных признаков могут быть взяты, например, отношения интерквантильных интервалов, соответствующих заданным значениям вероятно-

стей, или вероятности, соответствующие интервалам $\pm t\sigma$ при нескольких заданных значениях коэффициента t . Интегральные признаки проще всего формируются как линейные функционалы от функции $p(x)$. В частности, ими могут служить нормированные моменты распределений

$$\mu_k = \frac{1}{\sigma^k} \int_{-\infty}^{+\infty} (x - X_{\Pi})^k p(x) dx.$$

Распространен формальный подход к выбору признаков, когда признаки выбираются как нормированные коэффициенты разложения закона распределения в ряд по какой-либо системе функций (например, в ряд Грама—Шарлье, Хаара и т. п.). При ортогональном разложении коэффициенты разложения являются линейными функционалами

$$u_i = \frac{1}{N_i} \int_{-\infty}^{+\infty} f_i(x) p(x) dx,$$

где $f_i(x)$ — заданные функции, по которым производится разложение; N_i — нормирующие множители.

В качестве одного из основных преимуществ использования коэффициентов разложения указывают на возможность все более точного представления закона распределения конечным отрезком ряда при увеличении числа членов разложения. Однако задача обратного восстановления закона по признакам не ставится: при заведомо малом числе признаков это неосуществимо. При выборе признаков желательно, чтобы признаки по отдельности возможно более полно отражали в количественной форме представляющие интерес особенности формы закона распределения. При практическом построении признаков часто оказывается необходимым их нелинейное преобразование: $U_1 = g_1(u_1)$, $U_2 = g_2(u_2)$ и т. д., с целью приведения к заданным диапазонам (например, к стандартному интервалу 0—1), повышения чувствительности в определенных диапазонах и т. п.

После выбора совокупности признаков целесообразно ввести в рассмотрение в общем случае *многомерное пространство признаков* $u_1, u_2, \dots$ (по ортогональным осям откладываются численные значения признаков). В пространстве признаков каждый конкретный закон представляется *изображающей точкой* с координатами $u_1, u_2, \dots$. Если взять два близких друг другу по форме закона распределения, то им будут соответствовать близкие изображающие точки. Желательно, чтобы был справедлив и обратный переход — близким изображающим точкам должны соответствовать близкие в определенном смысле формы законов распределения.

В отношении простоты описания преимущество имеют такие способы задания признаков, при которых число признаков мини-

мально, а построенная систематизация достаточно полно отражает желаемые особенности формы распределений. Поэтому ограничим свою задачу рассмотрением лишь симметричных распределений, когда левая половина кривой плотности является точным зеркальным отображением ее правой половины.

По числу максимумов в кривой плотности, называемых *модами*, законы распределения можно разделить на безмодальные (равномерное, трапецеидальные), одномодальные, двухмодальные и полимодальные. Полимодальные законы распределения, имеющие более двух мод, исключим из рассмотрения. Будем полагать, что если из экспериментальных данных получается трехмодальное и более распределение, то это вызвано лишь случайностью малой выборки, а распределение генеральной совокупности случайной величины является плавным и не имеет более двух мод. Таким образом, предлагаемую классификацию ограничим лишь безмодальными, одномодальными и двухмодальными распределениями.

Обсудим теперь выбор признаков, характеризующих форму распределений. При использовании второго и четвертого центральных моментов форма закона распределения численно характеризуется значением эксцесса ε . Но эксцесс различных распределений колеблется в бесконечных пределах (от 1 до ∞), из-за чего этот параметр неудобен. Поэтому произведем его нелинейное преобразование в значение контрэксцесса $\kappa = 1/\sqrt{\varepsilon}$, которое для любых распределений заключено в пределах от 0 (при $\varepsilon = \infty$) и до 1 (при $\varepsilon = 1$). Таким образом, в качестве первого признака примем значение контрэксцесса κ .

Однако классификация распределений по одному контрэксцессу является недостаточной. Рассматривая параметры различных законов распределений, читатель уже мог заметить, что совершенно разные законы распределения могут иметь совпадающие значения эксцесса и контрэксцесса. Так, кругловершинное распределение Стюдента с числом степеней свободы $\nu = 6$ (табл. 2-5) и островершинное распределение Лапласа (табл. 2-4) имеют точно совпадающие значения $\varepsilon = 6$ и $\kappa = 0,408$. Эксцесс трапецеидальных распределений с отношением оснований от 1 : 1 до 1 : 3 изменяется от 1,8 до 2,2 (табл. 2-2), а эксцесс композиций арксинусоидальных распределений изменяется от 1,5 до 2,25, т. е. значения эксцессов этих двух классов распределений в значительной мере перекрываются, и т. д.

В качестве второго независимого признака формы закона распределения вероятности предлагается принять значение энтропийного коэффициента $k = \Delta_z/\sigma$, который для любых законов распределения изменяется в пределах от 0 до $\sqrt{(\pi e)/2} \approx 2,066$.

При использовании этих двух признаков изображающая точка с координатами k и κ будет всегда находиться в пределах прямоугольника, ограниченного значениями k от 0 до 2,066 и значе-

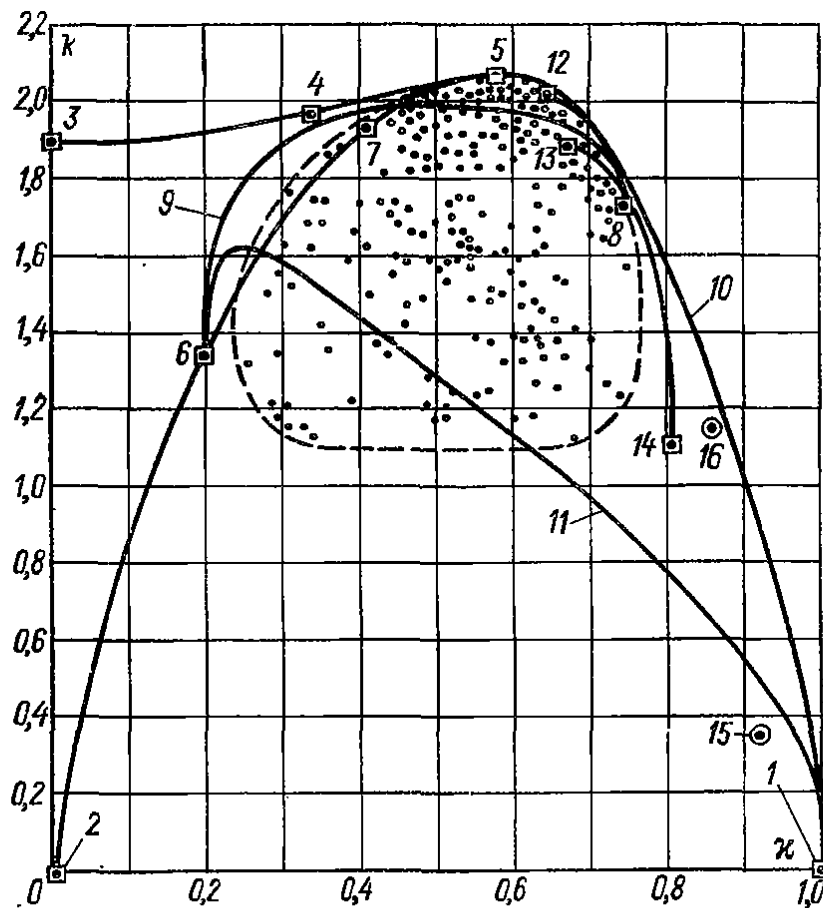


Рис. 2-10

ниями x от 0 до 1. Такая область на плоскости признаков изображена на рис. 2-10. Разместим на ней все распределения рассмотренных выше классов. Начнем с экстремальных распределений, т. е. двузначного дискретного распределения и семейства распределений Стюдента. Они занимают в этой области самое правое и самое левое граничные положения.

Дискретное двузначное распределение было показано на рис. 2-1, з и состоит из двух импульсов δ -функций при $x = -a$ и $x = +a$. Шенноновскую энтропию такого распределения можно определить на основании следующих рассуждений. Если при двухмодальном распределении (рис. 2-8, б), состоящем из двух нормальных распределений с $P_1 = P_2 = 0,5$ и с центрами $-a$ и $+a$, с. к. о. σ_n нормальных распределений будет убывать по сравнению с a , то при $\sigma_n \ll a$ они практически уже не будут пересекаться между собой и шенноновская энтропия будет определяться лишь суммой их интервалов неопределенности, т. е. будет $H(X) = \ln(2\sigma_n \sqrt{2\pi e})$, а суммарный интервал неопределенности — $d = 2\Delta_0 = 2\sigma_n \sqrt{2\pi e}$. Среднее квадратическое отклонение же такого распределения σ_2 при $\sigma_n \ll a$ будет стремиться к значению $\sigma_2 = a$. Отсюда энтропийный коэффициент этого распределения $k = \Delta_0/\sigma = \sigma_n \sqrt{2\pi e}/a$. При $\sigma_n \rightarrow 0$ это распределение будет стремиться к дискретному двузначному распределению, а его $k \rightarrow 0$.

Четвертый момент такого распределения $\mu_4 = a^4$, поэтому эксцесс $\varepsilon = \mu_4/\sigma^4 = a^4/a^4 = 1$, а следовательно, контрэксцесс

$\kappa = 1$. Таким образом, дискретному двузначному распределению в рассматриваемой области соответствует изображающая точка 1 с координатами $k = 0$ и $\kappa = 1$ (рис. 2-10).

Геометрическое место точек, соответствующих на рис. 2-10 семейству распределений Стьюдента, представляет собой линию с резким изломом, проходящую через точки 2—3—4—5. Действительно, распределение Коши, представляющее собой распределение Стьюдента с числом степеней свободы $\nu = 1$, имеет, как было показано в § 2-4, $\kappa = 0$, $\sigma = \infty$ и $\Delta_\sigma = 2\pi a$. Следовательно, для него $k = \Delta_\sigma/\sigma = 0$ и соответствующая ему на рис. 2-10 точка 2 имеет координаты $\kappa = 0$ и $k = 0$. Распределение Стьюдента при числе степеней свободы $\nu = 4$ (см. табл. 2-5) имеет $\kappa = 0$ и $k = 1,900$ (точка 3 на рис. 2-10). При $\nu = 5$ распределение Стьюдента имеет $\kappa = 0,333$ и $k = 1,97$ (точка 4 на рис. 2-10). И наконец, при $\nu \rightarrow \infty$ распределение Стьюдента стремится к нормальному с $\kappa = 0,577$ и $k = 2,006$ (точка 5 на рис. 2-10).

Геометрическое место точек, соответствующих классу экспоненциальных распределений, представляет собой на рис. 2-10 кривую, проходящую через точки 2—6—7—5—8. Экспоненциальное распределение с $\alpha \rightarrow 0$ имеет $\kappa \rightarrow 0$ и $k \rightarrow 0$, т. е. соответствует точке 2. Точка 6 с $\kappa = 0,199$ и $k = 1,35$ (см. табл. 2-4) соответствует экспоненциальному распределению с $\alpha = 1/2$, точка 7 — распределению с $\kappa = 408$ и $k = 1,92$, т. е. распределению Лапласа, точка 5 — нормальному распределению и, наконец, точка 8 с $\kappa = 0,745$ и $k = 1,73$ — равномерному (т. е. экспоненциальному с $\alpha \rightarrow \infty$).

Семейству распределений типа шапос экспоненциальной составляющей с $\alpha = 1/2$ на рис. 2-10 соответствует геометрическое место точек в виде кривой 9. Она начинается (при относительном весе дисперсии равномерной составляющей $p_{\text{равн}} = 0$ и соответственно весе экспоненциальной $p_{\text{экс}} = 1$) в точке 6, соответствующей чисто экспоненциальному распределению, а заканчивается в точке 8, соответствующей чисто равномерному распределению.

Положение геометрических мест точек, соответствующих двумодальным распределениям, существенно зависит от показателя степени экспоненциальной составляющей, входящей в такие композиции. Так, кругловершинным двухмодальным распределениям с экспоненциальной составляющей в виде нормального распределения (с $\alpha = 2$) на рис. 2-10 соответствует кривая 10, начинающаяся в точке 5, соответствующей нормальному распределению, и заканчивающаяся в точке 1, соответствующей дискретному двузначному распределению. Островершинным двухмодальным распределениям соответствует кривая 11, начинающаяся в точке 6 и заканчивающаяся в точке 1.

Классу трапецеидальных распределений на рис. 2-10 соответствует кривая, соединяющая точку 12 (тре-

угольное распределение Симпсона с $\kappa = 0,645$ и $k = 2,02$, см. табл. 2-2) и точку 8, соответствующую равномерному распределению.

И наконец, классу арксинусоидальных распределений на рис. 2-10 соответствует кривая, соединяющая точку 14 ($\kappa = 0,816$ и $k = 1,11$) с точкой 13, соответствующей композиции двух равных арксинусоидальных распределений с $\kappa = 0,667$ и $k = 1,88$ (см. табл. 2-3). (Точки 16 и 15 будут пояснены ниже.)

Предлагаемое представление аналитических моделей законов распределения в виде изображающих точек на плоскости признаков позволяет достаточно наглядно видеть близость или удаленность как различных моделей между собой, так и близость изображающих точек, соответствующих экспериментальным распределениям погрешности, к той или иной аналитической модели.

Для получения ответа на вопрос о том, какие из рассмотренных аналитических моделей пригодны для описания распределений погрешностей средств и результатов измерений, на рис. 2-10 нанесены изображающие точки, соответствующие параметрам 219 распределений погрешностей измерений с помощью различных СИ, собранных в упоминавшейся выше работе [3]. Эти точки на рис. 2-10 обведены штриховым овалом.

При необходимости на плоскости параметров в пределах рассматриваемой области могут быть выделены подобласти, соответствующие различным классам распределений.

Так, при оценке погрешностей средств и результатов измерений закон их распределения чаще всего принимают нормальным или равномерным. Из рис. 2-10 видно, что точки, соответствующие экспериментально полученным распределениям, действительно наиболее плотно группируются вблизи линии 5—8, соответствующей композиции этих распределений. Но для описания таких распределений одновременно пригодны несколько моделей: это трапеции (линия 8—12), распределения класса шапо, т. е. композиции равномерного с нормальным (линия 5—8) и другими экспоненциальными распределениями (правая часть кривой 9), и даже кругловершинные двухмодальные распределения (верхний участок кривой 10). При необходимости одна из этих моделей с успехом может быть заменена другой.

Для описания же экспериментальных распределений, точки которых на рис. 2-10 попадают в левую верхнюю область, ограниченную штриховой линией, могут быть использованы лишь модели островершинных экспоненциальных распределений (линия 5—7—6). Внутреннюю часть овала реальных распределений погрешностей можно описать лишь с помощью моделей распределений класса шапо в ее верхней части (правая часть кривой 9) или моделей островершинных двухмодальных распределений (левая часть кривой 11, пересекающая нижнюю часть штрихового овала).

Одновременно на рис. 2-10 видно, что погрешности обследованных 219 средств измерений никогда не были близки к распределению Стьюдента (линия 2—3—4) или к введенным ГОСТ 8.011—72 моделям для двухмодальных распределений (названных в этом стандарте «антимодальная I» и «антимодальная II»), изображающие точки которых помечены на рис. 2-10 числами 16 и 15.

В заключение следует отметить, что подобный способ изображения для классификации распределений в свое время предлагался Пирсоном, но в качестве координат признаков он предлагал использовать эксцесс и коэффициент асимметрии распределений. Этот способ не получил широкого распространения, так как все симметричные распределения располагались по оси эксцесса при асимметрии, равной нулю. Лишь использование понятия энтропийного коэффициента позволило получить вторую независимую координату и развернуть эту классификацию в плоскости признаков k — κ .

Кроме того, необходимо особо подчеркнуть, что по точке на плоскости признаков k — κ нельзя однозначно восстановить закон распределения. Совпадение или близость двух точек (как это было показано на примере линии 5—8, которой одновременно соответствуют несколько разных моделей) отнюдь не гарантируют полного совпадения законов или полного соответствия экспериментальных данных теоретической модели. Близость имеет место лишь в некотором «среднем» смысле. Тем не менее, как показал опыт исследований [3, 7, 8, 32, 40, 41, 42], описанная классификация симметричных распределений оказалась очень полезной на практике, так как позволяет по рассчитанным оценкам k и κ получить представление о возможном виде аналитической модели распределения, что может быть выполнено ЭВМ (см. § 5-8) даже без участия оператора.

2.6. ФОРМУЛЫ ДЛЯ РАСЧЕТА ПОГРЕШНОСТИ С ЗАДАННОЙ ДОВЕРИТЕЛЬНОЙ ВЕРОЯТНОСТЬЮ

Государственный стандарт 8.011—72 требует для характеристики точности результатов измерений указания или границ интервала, в котором эта погрешность находится с заранее заданной вероятностью, или указания самой функции распределения погрешности.

До тех пор, пока считали, что все случайные погрешности приборов или результатов измерения должны быть распределены только нормально, для определения погрешности Δ_d с заданной доверительной вероятностью P_d рекомендовалось пользоваться таблицей квантилей t стандартного нормального распределения, т. е. определять их как $\Delta_{0,9} = 1,65\sigma$, $\Delta_{0,95} = 1,96\sigma$, $\Delta_{0,99} = 2,58\sigma$, $\Delta_{0,997} = 2,97\sigma$ и $\Delta_{0,999} = 3,29\sigma$, где σ — найденная по достаточно большой выборке оценка с. к. о.

Однако по мере накопления данных о фактических распределениях погрешностей стало очевидным, что они весьма разнообразны и очень часто далеки от нормального. Это большое разнообразие законов распределения погрешностей обуславливает практическую сложность определения доверительных значений погрешностей, так как необходимо иметь таблицы квантилей для всех разновидностей распределений. В связи с этим в последние годы появились предложения различных авторов, пытавшихся найти упрощенное решение этой задачи.

Так, еще в 1966—1968 гг. в работах [40, 32] был приведен расчетный график, показывающий, что для класса экспоненциальных распределений независимо от их вида при значениях контрэксцесса $\kappa = 0,3 \div 0,74$ можно приближенно считать $\Delta_{0,9} \approx 1,65\sigma$, а при $\kappa = 0,4 \div 0,7$ — соответственно $\Delta_{0,95} \approx 2\sigma$. Аналогичные результаты были получены Б. Л. Рывкиным (Труды ВНИИМ, вып. 126 (186), 1971, с. 15—25).

В 1969 г. в работе [11] было проведено сравнение интегральных кривых равномерного, трапецеидального, треугольного и нормального распределений и путем их усреднения получена некоторая средняя зависимость $t = \Delta_d/\sigma$ как функция от P_d . Эта зависимость была включена в виде рекомендуемого приложения 2 в ГОСТ 8.009—72 для определения t в пределах P_d от 0,7 до 0,99 без установления вида распределения. Согласно этой кривой $\Delta_{0,9} = 1,64\sigma$, $\Delta_{0,95} = 1,9\sigma$ и $\Delta_{0,99} = 2,1\sigma$. Существенным было то, что без знания вида распределения указание погрешности с доверительной вероятностью, большей, чем $P_d = 0,99$, запрещалось, так как кривая $t = f(P_d)$ обрывалась при $P_d = 0,99$. Это было очень важным нововведением, так как до этого времени было широко принято приводить совершенно необоснованное значение $\Delta_{0,997}$ как $\Delta = 3\sigma$. После введения ГОСТ 8.009—72 такое указание погрешностей практически прекратилось.

В 1977 г. в работе [33] на основе сравнения аналитических интегральных кривых различных законов распределения (от Лапласа до равномерного и даже некоторых двухмодальных) авторы рекомендовали следующие приближенные (с погрешностью $\pm 10\%$) соотношения: $\Delta_{0,9} = 1,6\sigma$, $\Delta_{0,95} = 1,8\sigma$ и $\Delta_{0,997} = \varepsilon\sigma$ (где $\varepsilon = \mu_4/\sigma^4$ — эксцесс распределения). Однако указание приближенных соотношений для определения некоторых точек интегральных кривых при двух-трех значениях доверительной вероятности не решает полностью поставленной задачи.

В работе [31] было показано, что среди различных произвольно назначаемых значений доверительной вероятности есть одно значение, обладающее уникальным свойством. Оказывается, что интегральные кривые для широкого класса симметричных, высокоэнтропийных ($k > 1,7$) распределений (равномерного, треугольного, трапецеидальных, нормального, экспоненциальных с $\alpha \geq 2/3$ и двухмодальных с небольшой глубиной антимодальности $S_d = a/\sigma < 1,5$) в области 0,05-й и 0,95-й квантилей (рис. 2-11)

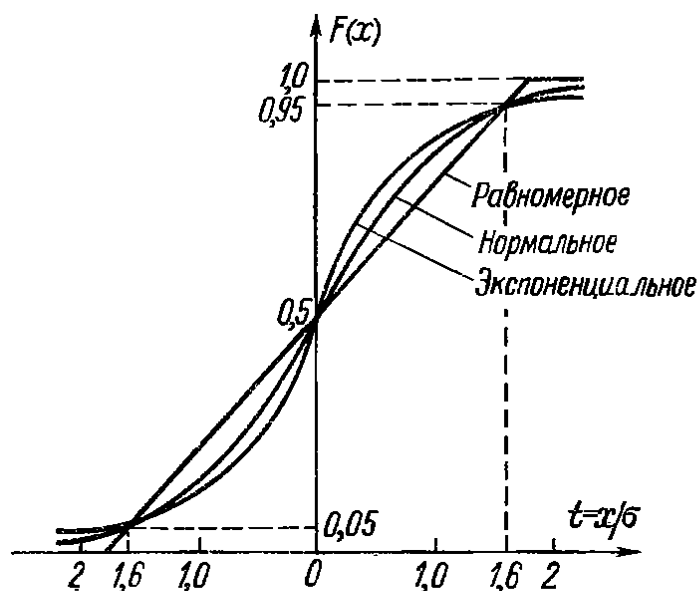


Рис. 2-11

пересекаются между собой в очень узком интервале значений $x/\sigma = 1,6 \pm 0,05$. Поэтому с погрешностью в $0,05\sigma$ можно считать, что 0,05-я и 0,95-я квантили для любых из этих распределений могут быть определены как $X_{0,05} = X_{ц} - 1,6\sigma$ и $X_{0,95} = X_{ц} + 1,6\sigma$, где $X_{ц}$ — координата центра распределения. Отсюда значение погрешности, определенное как $\Delta_{0,9} = 1,6\sigma$ (с погрешностью $\pm 0,05\sigma$), для любых из этих распределений является погрешно-

стью с 90%-ной доверительной вероятностью. Именно эта общая точка интегральных кривых ряда классов распределений и была найдена Б. Л. Рывкиным, а также в работах [11, 32, 33, 40].

При $F(x) > 0,95$, т. е. при $P_d > 0,9$ интегральные кривые для разных законов распределения резко расходятся между собой. Производить их усреднение, как это было сделано в приложении 2 ГОСТ 8.009—72, не имеет смысла. Поэтому это приложение из ГОСТ 8.009—72 было изъято как необоснованное, однако подобная усредненная зависимость вновь введена в ГОСТ 8.009—84.

Задача же, наоборот, состоит в том, чтобы вместо большого числа таблиц квантилей разнообразных распределений, которые безусловно не могут быть сведены к единой «средней» кривой, найти хотя бы приближенное аналитическое описание веера интегральных кривых этих распределений в области $t > 1,6$ и $F(x) > 0,95$ на рис. 2-11 от какого-либо числового параметра формы распределения.

В качестве такого параметра, определяемого формой распределения, может быть взято, например, значение эксцесса ε , как это было сделано авторами [33] в последнем из предложенных ими соотношений. Таким образом, решение этой задачи состоит в нахождении аппроксимирующих формул вида $t = f(\varepsilon, P_d)$ для близких классов распределений. А эта близость или удаленность легко определяется благодаря использованию топографической классификации, полученной на основе энтропийного подхода. Поэтому возможность создания аппроксимирующих формул является также одним из практических результатов такой классификации. Если от этих выражений не требовать слишком высокой точности, то, исходя из классификации рис. 2-10, можно заключить, что такое выражение может быть единым как для класса экспоненциальных распределений, так и для распределений Стюдента с достаточно большим числом степеней свободы ($n > 8$, $\alpha > 0,45$), а также и для всего класса трапецеидальных распреде-

лений, ибо значения энтропийного коэффициента (отражающего форму этих распределений) при одном и том же значении контрэксцесса для них мало различаются между собой.

Что же касается распределений класса шапо, то из характера кривых топографической классификации следует, что для них функция от эксцесса должна быть совсем иной, чем для класса экспоненциальных распределений.

То же самое относится и к двухмодальным распределениям: формулы для кругловершинных (кривая 10 на рис. 2-10) и островершинных (кривая 11) распределений должны быть различными, так как при одном и том же значении эксцесса эти кривые далеко отстоят друг от друга.

Для классов экспоненциальных и трапецеидальных распределений, а также распределений Стьюдента при $n > 8$, в качестве конкретных моделей, соответствующих области реально встречающихся распределений погрешностей, примем: распределение Лапласа ($\epsilon = 6$, $\kappa = 0,4$), нормальное ($\epsilon = 3$, $\kappa = 0,577$), трапецеидальное с отношением верхнего и нижнего оснований $1 : 2$ ($\epsilon = 2$, $\kappa = 0,7$) и равномерное ($\epsilon = 1,8$, $\kappa = 0,745$) распределения. Семейство аналитических зависимостей $P_d = f(t, \epsilon)$ для указанных четырех распределений изображено на рис. 2-12 сплошными линиями и представляет собой изображенный в увеличенном масштабе участок кривых рис. 2-11 при $F(x) > 0,95$ и $t > 1,6$.

Это семейство кривых с погрешностью, не превосходящей 4%, в пределах значений P_d от 0,9 до 0,99 аппроксимируется выражением

$$t = 1,62 [3,8 (\epsilon - 1,6)^{2/3}]^{\lg \lg [1/(1-P_d)]}, \quad (2-19)$$

а с погрешностью 8% оно пригодно для P_d от 0,99 до 0,999.

Семейство аппроксимирующих кривых в соответствии с (2-19) показано на рис. 2-12 штриховыми линиями, которые удовлетворительно совпадают с теоретическими. Поэтому выражение (2-19) может использоваться для расчета Δ_d как $\Delta_d = t\sigma$, где $t = f(\epsilon, P_d)$.

Аналогичное исследование для кругловершинных двухмодальных распределений, представляющих собой композиции нормального и двузначного дискретного распределений, показывает, что на участке ϵ от 3 до 1,3 (κ от 0,6

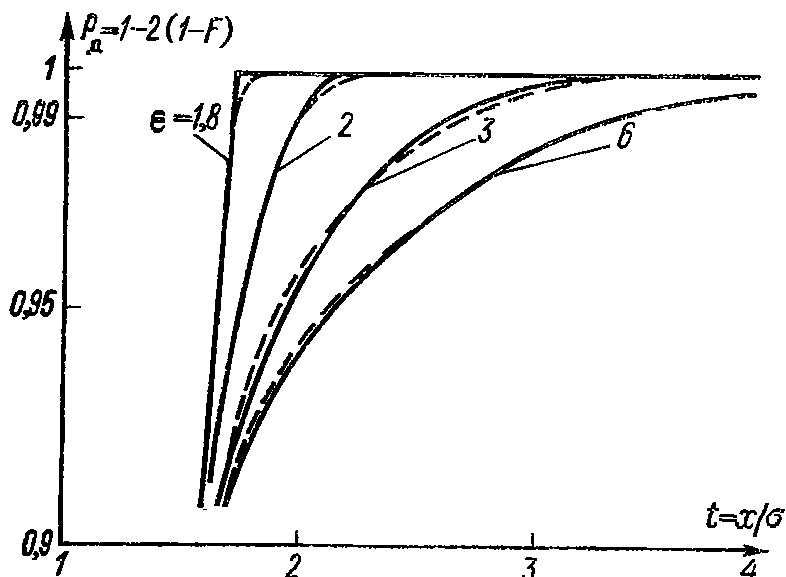


Рис. 2-12

до 0,9) для P_d от 0,9 до 0,999 с погрешностью 10% подобная зависимость аппроксимируется формулой

$$t = 1,6 \{3,6 [1 + \lg(\varepsilon - 1)]\}^{\lg \lg [1/(1-P_d)]}. \quad (2-20)$$

Для распределений класса шапо, образующихся как композиции экспоненциального распределения с $\alpha = 1/2$ и равномерного распределения в интервале значений ε от 6 до 1,8 с погрешностью 8%, искомое соотношение аппроксимируется формулой

$$t = 1,56 [1,12 + (\varepsilon - 1,8)^{0,58} / \sqrt{10}]^{\lg [0,1/(1-P_d)]}. \quad (2-21)$$

И, наконец, для островершинных двухмодальных распределений, образующихся как композиция распределения Лапласа и дискретного двузначного распределения, аналогичная зависимость в интервале значений ε от 6 до 1,8 для P_d от 0,9 до 0,999 с погрешностью 5% аппроксимируется формулой

$$t = 1,23 \left[1 + \sqrt{\frac{\varepsilon - 1}{2,5}} \lg \frac{0,175}{1 - P_d} \right]. \quad (2-22)$$

Таким образом, используя соотношения (2-19)—(2-22), можно с достаточной для практики точностью, не прибегая к использованию таблиц, вычислять доверительные значения погрешностей для всех классов практически встречающихся распределений погрешностей. Однако для выбора нужной формулы необходимо знать вид класса распределения погрешности. Этот вопрос будет подробно рассмотрен далее на практическом примере в § 3-5.

ГЛАВА ТРЕТЬЯ

МЕТОДЫ РАСЧЕТНОГО СУММИРОВАНИЯ СОСТАВЛЯЮЩИХ РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ

3-1. ОСНОВЫ ТЕОРИИ РАСЧЕТНОГО СУММИРОВАНИЯ ПОГРЕШНОСТЕЙ

Задача определения расчетным путем оценки результирующей погрешности по известным оценкам ее составляющих называется задачей суммирования погрешностей и возникает во многих случаях в практике измерений. Так, для определения погрешности даже отдельного прибора или измерительного преобразователя необходимо суммировать все составляющие его погрешности (основной, от колебаний температуры, от колебания напряжения питания и др.). При создании измерительных каналов ИИС и

ИВК встает задача суммирования погрешностей нескольких измерительных преобразователей, образующих данный измерительный канал.

При определении погрешности как прямых, так и косвенных измерений к погрешностям используемых средств измерений должны быть добавлены методические погрешности, погрешности, появляющиеся при отсчете показаний, расшифровке осциллограмм, должен быть учтен иногда очень сложный механизм трансформации погрешностей каждого из результатов прямых измерений в результирующую погрешность результата косвенного измерения и т. д. Таким образом, задача расчетного суммирования погрешностей — это одна из основных задач как при создании средств измерений, так и при оценке погрешностей результатов самих измерений.

Трудность проведения такого суммирования заключается в том, что все составляющие погрешности должны рассматриваться как случайные величины, принимающие в каждой частной реализации самые разнообразные значения. С точки зрения теории вероятностей они могут быть наиболее полно описаны своими законами распределения, а их совместное действие — соответствующим многомерным законом распределения. Однако в такой постановке задача суммирования погрешностей практически неразрешима уже для 3—4 составляющих (не говоря уже о 30—40), так как операции с такими многомерными законами непреодолимо сложны. Поэтому практический путь решения задачи суммирования состоит в том, чтобы вместо определения многомерных законов распределения подобрать для характеристики составляющих такие числовые оценки (например, среднее квадратическое или энтропийное значение, эксцесс, контрэксцесс, энтропийный коэффициент и т. д.), оперируя с которыми, можно было бы определить соответствующие числовые оценки результирующей погрешности без определения многомерных или результирующих одномерных законов распределения рассматриваемых случайных величин.

При этом оказывается необходимым учитывать, что: 1) числовые характеристики законов распределения составляющих (например, σ_i , k_i или t_i) могут не оставаться постоянными в функции от значений измеряемой величины, т. е. изменяться в диапазоне ее изменения; 2) отдельные составляющие погрешности могут быть коррелированы между собой; 3) при суммировании случайных величин законы их распределения существенно деформируются, т. е. форма закона распределения суммы может резко отличаться от формы распределения составляющих.

Правила суммирования погрешностей основываются на том предположении (ограничении), что погрешность по абсолютному значению всегда много меньше самой измеряемой величины. Поэтому, например, изменение погрешности в функции от изменения значений самой измеряемой величины может учитываться самым простейшим способом, а именно путем разделения всех

суммируемых составляющих погрешности на аддитивные и мультипликативные. Далее предполагается, что сумма аддитивных составляющих дает значение аддитивной части результирующей погрешности, а сумма мультипликативных составляющих — мультипликативной.

В пределах диапазона изменения измеряемой величины не более десятикратного изменение результирующей погрешности может быть с достаточной точностью представлено прямой линией. Поэтому достаточно найти значения результирующей погрешности лишь в начале и конце такого диапазона и описать результирующую погрешность простейшей линейной двучленной формулой.

При диапазоне изменения измеряемой величины более десятикратного он может быть разбит на участки и результирующая погрешность определена в начале и конце каждого участка.

Для устранения влияния деформации формы законов распределения при суммировании погрешностей все суммируемые составляющие исходно представляются своими с. к. о. σ и все операции расчетного суммирования производятся только над этими *средними квадратическими значениями* (с. к. о.) погрешностей.

Учет взаимных корреляционных связей между суммируемыми составляющими производится путем использования различных правил суммирования для жестко коррелированных и слабо коррелированных составляющих.

В результате суммирования с. к. о. исходных составляющих получают с. к. о., соответственно, аддитивной и мультипликативной составляющих результирующей погрешности. Среднее квадратическое отклонение аддитивной составляющей характеризует результирующую погрешность в начале диапазона измерений (при $x \approx 0$), а для определения с. к. о. результирующей погрешности в конце диапазона измерений с. к. о. аддитивной и мультипликативной составляющих должны быть просуммированы. Если диапазон измерений простирается на несколько порядков измеряемой величины, то такое суммирование производится в нескольких точках диапазона, а затем принимается решение о методе описания изменения результирующей погрешности во всем диапазоне.

Чаще всего результирующую погрешность желательно выразить не в виде с. к. о., а в виде некоторого (энтропийного или доверительного) интервала неопределенности. Этот последний переход от с. к. о. σ_x к энтропийному γ_e или доверительному γ_d значениям погрешности является с теоретической точки зрения самой трудной операцией при суммировании погрешностей. Дело в том, что $\gamma_e = k_x \sigma_x$ и $\gamma_d = t_x \sigma_x$, где энтропийный коэффициент k_x или квантильный множитель t_x зависят от формы закона распределения результирующей погрешности, а вся излагаемая методика с самого начала была нацелена на то, чтобы обойтись без точного определения результирующего закона распределения суммы всех составляющих (см. § 3-2—3-4).

Дисперсия суммы коррелированных и некоррелированных случайных величин. Из теории вероятностей известно, что дисперсия суммы двух случайных величин в общем случае

$$D(x_1 + x_2) = D(x_1) + D(x_2) + 2K_{x_1 x_2}$$

где $D(x_1)$ — дисперсия x_1 ; $D(x_2)$ — дисперсия x_2 ; $K_{x_1 x_2} = \rho \sigma_1 \sigma_2$ — их взаимный корреляционный момент; ρ — коэффициент корреляции.

Отсюда с. к. о. суммы этих величин $\sigma_z = \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}$.

Если эти величины между собой не коррелированы, то $\rho = 0$ и $\sigma_z = \sqrt{\sigma_1^2 + \sigma_2^2}$, т. е. в подобном случае для определения оценки результирующей погрешности в виде σ_z суммируемые составляющие должны быть оценены также средними квадратическими значениями σ_1 и σ_2 и эти оценки должны суммироваться геометрически (чем мы и пользовались выше, говоря о суммировании статистически независимых погрешностей).

Однако если x_1 и x_2 жестко и положительно коррелированы между собой ($\rho = +1$), т. е. x_2 принимает значения, лишь строго пропорциональные x_1 , то всякое положительное отклонение $+\Delta x_1$ сопровождается также положительным отклонением $+\Delta x_2$ и отклонение $\Delta(x_1 + x_2)$ складывается как $\Delta x_1 + \Delta x_2$. Это формально следует и из формулы для σ_z при $\rho = +1$, ибо $\sigma_z = \sqrt{\sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2} = \sigma_1 + \sigma_2$.

Если же при возрастании x_1 значения x_2 , наоборот, линейно убывают, то $\rho = -1$ и $\sigma_z = \sqrt{\sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2} = |\sigma_1 - \sigma_2|$.

Таким образом, оценки жестко коррелированных погрешностей ($\rho = \pm 1$) должны суммироваться не геометрически, а алгебраически с учетом их знаков.

Понятия корреляционной связи и коэффициента корреляции. Они излагаются во многих курсах теории вероятностей, однако для человека, впервые сталкивающегося с ними, эти понятия, как показывает опыт, долгое время остаются достаточно туманными. А так как в дальнейшем в этой книге нам еще не раз придется обращаться к использованию понятия коэффициента корреляции, то позволим себе несколько отклониться от основного вопроса и обсудить это специфическое понятие.

Если при изменении величины x другая величина y изменяется так, что каждому значению x_i соответствует совершенно определенное значение y_i , то такую связь называют функциональной. Однако на практике, вследствие влияния случайных погрешностей при измерениях, мы наблюдаем такие зависимости в несколько ином виде. Ограничимся рассмотрением только линейной зависимости, когда на графике экспериментальных данных такая связь явно просматривается (т. е. видно, что в среднем на рис. 3-1, б $y \approx kx$), но строгого соответствия между x_i и y_i (как на рис. 3-1, а) нет. Одному и тому же значению x_i в разных реализациях могут соответствовать различные значения y_i в интервале $L_{\Delta y}$.

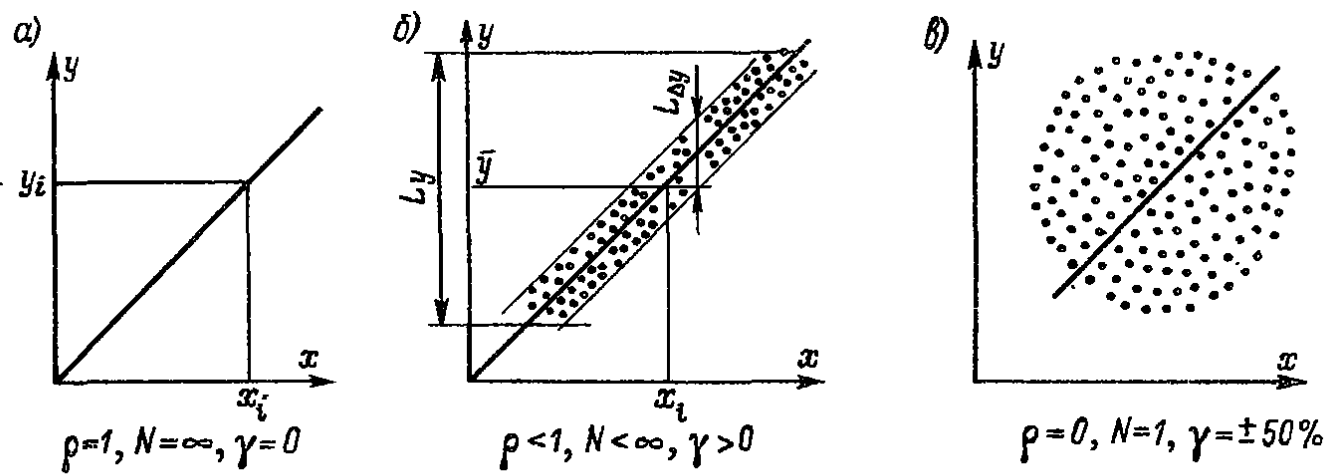


Рис. 3-1

Такие нежесткие линейные связи величин в теории вероятностей в отличие от функциональных называются корреляционными. Теснота корреляционной связи может быть оценена относительной вытянутостью поля экспериментальных точек, т. е. отношением ширины полосы точек $L_{\Delta y}$ вдоль оси y к протяженности всего поля L_y вдоль той же оси (см. рис. 3-1, б). В теории измерений $N = L_y/L_{\Delta y}$ понимается как число различных градаций измеряемой величины (см. рис. 1-3). В теории вероятностей каждый из размеров L_y и $L_{\Delta y}$ принято характеризовать, соответственно, значением с. к. о. отдельных точек от их среднего арифметического значения $\bar{y}$ в виде σ_y и с. к. о. отдельных точек от линии $y = kx$ в виде $\sigma_{\Delta y}$.

Используемая в измерительной технике приведенная погрешность есть отношение половины ширины полосы неопределенности $L_{\Delta y}/2$ к длине диапазона L_y , т. е. $\gamma = L_{\Delta y}/(2L_y) = 1/(2N)$. Она, как и всякая другая оценка погрешности, является негативной характеристикой тесноты корреляционной связи, т. е. относительной вытянутости поля экспериментальных точек. В теории вероятностей для той же самой цели принято использовать позитивную оценку в виде так называемого *коэффициента корреляции*

$$\rho = 1/\sqrt{1 + (\sigma_{\Delta y}/\sigma_y)^2}.$$

Учитывая, что при измерениях погрешность составляет, как правило, единицы или доли процента, т. е. $\sigma_{\Delta y} \ll \sigma_y$, практически можно пользоваться приближенным соотношением

$$\rho = \sqrt{1 - (\sigma_{\Delta y}/\sigma_y)^2} = \sqrt{1 - (2\gamma)^2}.$$

[Несколько более точное соотношение между ρ и γ , учитывающее различие законов распределения y и Δy , обсуждается в § 6-7, формулы (6-13) и (6-14).]

Таким образом, коэффициент корреляции ρ и приведенная погрешность γ — это два понятия, характеризующие одно и то же свойство, и потому они находятся в строгом соотношении между

собой и могут равноправно использоваться для характеристики этого свойства.

Однако значения ρ весьма своеобразно связаны с обычно используемыми в измерительной технике значениями приведенной погрешности γ . При идеальной линейной функциональной связи x и y (рис. 3-1, а) $\gamma = 0$, $N = \infty$ и $\rho = 1$. При мощности шума $\sigma_{\Delta y}^2$, равной мощности σ_x^2 сигнала $\sigma_y^2 = \sigma_x^2 + \sigma_{\Delta y}^2 = 2\sigma_{\Delta y}^2$, $(\sigma_{\Delta y}/\sigma_y)^2 = 0,5$ и $\rho^2 = 0,5$, так как всегда $(\sigma_{\Delta y}/\sigma_y)^2 + \rho^2 = 1$, то следовательно, $\rho = \sqrt{0,5} \approx 0,7$.

При $\gamma = \pm 50\%$ интервал неопределенности составляет $2\gamma = 100\%$, т. е. разброс экспериментальных данных занимает весь диапазон изменения y . В этом случае (рис. 3-1, в) $N = \sigma_y/\sigma_{\Delta y} = 1$ и $\rho = 0$.

Реально исследуемые функциональные связи измеряемых величин или характеристики СИ с приведенной погрешностью 2,5—0,25%, обеспечивающие различие 20—200 градаций измеряемой величины (см. табл. 7-1), характеризуются коэффициентами корреляции от 0,999 до 0,99999 соответственно.

При решении задач расчетного суммирования погрешностей нас будет интересовать лишь разграничение сильной и слабой корреляционной связи суммируемых погрешностей. Условной границей между сильной и слабой корреляционной связью суммируемых погрешностей можно считать случай, когда мощность сигнала σ_x^2 и мощность шума $\sigma_{\Delta y}^2$ равны между собой и $\rho = \sqrt{0,5} \approx 0,7$.

Практические правила расчетного суммирования составляющих результирующей погрешности.

1. Как следует из изложенного, для определения оценки результирующей погрешности должны учитываться взаимные корреляционные связи различных составляющих погрешности (принадлежащих часто различным узлам канала, различным стадиям процесса измерения и т. п.). Поэтому исходными данными для более точного расчета должны служить оценки именно всех отдельных составляющих погрешности, а не оценки некоторых суммарных погрешностей (отдельных преобразователей, суммарная методическая погрешность и т. п.).

2. Так как суммировать с учетом корреляционных связей можно лишь с. к. о. составляющих, то для каждой составляющей должно быть найдено по исходным данным ее с. к. о. В большинстве случаев для этого необходимо знание или предположение о виде закона распределения каждой из составляющих.

3. Эти составляющие подразделяются на аддитивные и мультипликативные и суммируются отдельно.

4. Из суммируемых составляющих выделяются группы сильно коррелированных между собой составляющих погрешности, и внутри этих групп производится алгебраическое суммирование их оценок.

При этом следует заметить, что вычислить точные значения коэффициентов взаимной корреляции составляющих погрешности, как правило, чрезвычайно трудно. Поэтому по степени коррелированности погрешности следует разделять лишь на два вида — сильно коррелированные ($\rho = 1 \div 0,7$) и слабо коррелированные ($\rho = 0 \div 0,7$). К первым относят погрешности, вызванные одной и той же причиной (общим источником питания, примерно одинаковыми изменениями температуры и т. д.), когда тесные корреляционные связи просматриваются логически и для них принимают ρ равным $+1$ или -1 . Погрешности же, между которыми такие взаимосвязи не обнаруживаются, относят к некоррелированным, и для них принимают $\rho = 0$.

5. После того как все группы сильно коррелированных погрешностей выделены и внутри их произведено алгебраическое суммирование, суммарные по группам и оставшиеся вне групп погрешности можно считать уже некоррелированными и складывать по правилу $\sigma_x^2 = \sum \sigma_i^2$.

Для определения с. к. о. при $x = 0$ складываются лишь аддитивные составляющие, а для определения с. к. о. погрешности в конце диапазона измерений — все суммируемые составляющие.

6. В общем случае может быть сформулировано правило пренебрежения малыми составляющими при суммировании погрешностей. Так как дискретность округления окончательного результата (см. § 1-4) всегда больше 3%, то могут быть опущены: одна малая составляющая, если она в 5 раз меньше наибольшей из суммируемых составляющих, две составляющие, если они в 6 раз меньше, три составляющие, если они в 7 раз меньше, и четыре, если они в 8 раз меньше наибольшей. Но делать такое заключение можно только после суммирования коррелированных составляющих и приведенных числовых значений погрешности к одному виду, лучше всего к с. к. о.

7. Для перехода от с. к. о. погрешности к энтропийному $\Delta_{\sigma} = k\sigma_x$ или доверительному $\Delta_d = t\sigma_x$ значениям должно быть тем или иным путем вынесено суждение о форме закона распределения результирующей погрешности и тем самым выбрано значение энтропийного коэффициента k_x или квантильного множителя t_x .

8. При определении энтропийного значения погрешности для него также может быть указано соответствующее значение доверительной вероятности в соответствии с формулой (2-6).

3-2. МЕТОДИКА РАСЧЕТА ЭНТРОПИЙНОГО ЗНАЧЕНИЯ РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ

Для перехода к интервальной оценке в виде доверительного $\Delta_d = t_x\sigma_x$ или энтропийного $\Delta_{\sigma} = k_x\sigma_x$ значений необходимо знание не самого закона распределения результирующей погреш-

ности, а лишь его одного числового параметра в виде квантильного множителя t_z или энтропийного коэффициента k_z .

Наиболее строго, без каких-либо допущений может быть определено значение k_z результирующего закона распределения, поэтому определение энтропийного значения результирующей погрешности является наиболее точным методом расчетного суммирования погрешностей. Однако этот метод суммирования достаточно трудоемкий. Поэтому при недостатке времени или исходных данных предпочтительнее использовать хотя и менее строгие теоретически, но более простые практически методы расчетного суммирования погрешностей, основанные на приближенном определении квантильных множителей t_z для той или другой доверительной вероятности (см. § 3-3).

При наиболее полном анализе механизма образования результирующей погрешности для каждой из составляющих должно быть найдено ее среднее квадратическое значение σ_i и принят (найден или приписан) тот или иной закон распределения вероятностей с известными значениями энтропийного коэффициента k_i , эксцесса ε_i и контрэксцесса κ_i .

Если составляющие погрешности были исходно заданы своими энтропийными значениями Δ_i и энтропийными коэффициентами k_i , то процедура суммирования (при $\rho = 0$) может быть представлена следующими соотношениями: $\sigma_1 = \Delta_1/k_1$; $\sigma_2 = \Delta_2/k_2$; ...; $\sigma_n = \Delta_n/k_n$ и $\sigma_z = \sqrt{\sum \sigma_i^2}$, а энтропийное значение результирующей погрешности определится как $\Delta_{\Sigma} = k_z \sigma_z$.

Таким образом, для определения Δ_{Σ} по σ_z достаточно знания всего лишь одной числовой характеристики результирующего закона распределения в виде его энтропийного коэффициента k_z .

Для суммы жестко коррелированных погрешностей ($|\rho| \approx 1$) нахождение энтропийного коэффициента облегчается тем, что законы распределения каждой из этих погрешностей повторяют форму закона распределения вероятностей вызывающей их общей влияющей величины. Так, например, если распределение вероятностей различных значений температуры в условиях лаборатории равномерно, то и все возникающие температурные погрешности (при условии линейной зависимости погрешности от температуры) распределены равномерно и различаются только шириной своих распределений. Сумма жестко коррелированных погрешностей распределена опять-таки по тому же самому закону, так как мгновенные значения всех ее составляющих совпадают по знаку и пропорциональны друг другу.

Таким образом, при суммировании жестко коррелированных погрешностей деформации законов распределения не происходит и энтропийный коэффициент распределения суммы равен энтропийному коэффициенту составляющих.

Задача определения энтропийного коэффициента композиции некоррелированных погрешностей по энтропийным коэффициентам и относительным весам дисперсий каждого из них

в суммарной дисперсии достаточно сложна. Ее аналитическое решение известно лишь для частных случаев суммирования двух нормальных, двух равномерных и равномерного и нормального распределения.

При суммировании двух случайных величин, распределенных нормально, их композиция представляет собой также нормальное распределение. Поэтому энтропийный коэффициент композиции в этом частном случае просто равен энтропийным коэффициентам суммируемых составляющих. Это один из редких случаев, когда при образовании композиции форма распределений не изменяется. (Из симметричных распределений это имеет место лишь при суммировании случайных величин с нормальным распределением и распределением Коши.)

При суммировании двух равномерно распределенных случайных величин энтропийное значение суммарной погрешности (см. [32, с. 92])

$$\Delta_{\Sigma} = \sqrt{3}\sigma_1 e^{\sigma_2/(2\sigma_1)}, \quad (3-1)$$

отсюда энтропийный коэффициент результирующего распределения

$$k_{\Sigma} = \sqrt{3e^C/(1+C^2)}, \quad (3-2)$$

где $C = \sigma_1/\sigma_2$ при $\sigma_1 \leq \sigma_2$ или $C = \sigma_2/\sigma_1$ при $\sigma_2 \leq \sigma_1$.

Если одна составляющая погрешности распределена равномерно, а другая — нормально, то строгое решение задачи существенно осложняется. Аппроксимирующее выражение для энтропийного коэффициента такой композиции (см. [32, с. 95])

$$k_{\Sigma} = \sqrt{3(\pi e/3)^{V_{p_H}}/(1+p_H)}, \quad (3-3)$$

где $p_H = \sigma_H^2/(\sigma_P^2 + \sigma_H^2)$, а σ_P и σ_H — с. к. о. суммируемых, соответственно, равномерной и нормальной составляющих.

Сводка зависимостей энтропийного коэффициента k_{Σ} от соотношения суммируемых составляющих и их энтропийных коэффициентов может быть представлена в виде семейства графиков. Эти зависимости не выражаются простыми аналитическими соотношениями, но в работах [7, 13, 27, 28, 32, 34, 40] эти задачи численно решены для композиций всех рассмотренных выше видов законов распределений. Результаты решений представлены на рис. 3-2 и 3-3, где по оси абсцисс отложены значения относительного веса дисперсии σ_2^2 второго из суммируемых распределений в полной дисперсии $p = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$, по оси ординат — значение энтропийного коэффициента k_{Σ} образующейся при этом композиции.

Так, если обе суммируемые составляющие распределены нормально, то композиция двух нормальных распределений всегда нормальна и значение $k_{\Sigma} = 2,066$ при любых значениях веса p

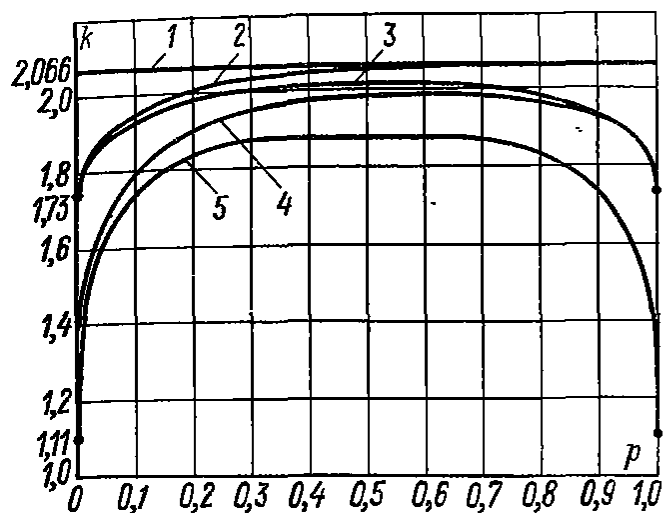


Рис. 3-2

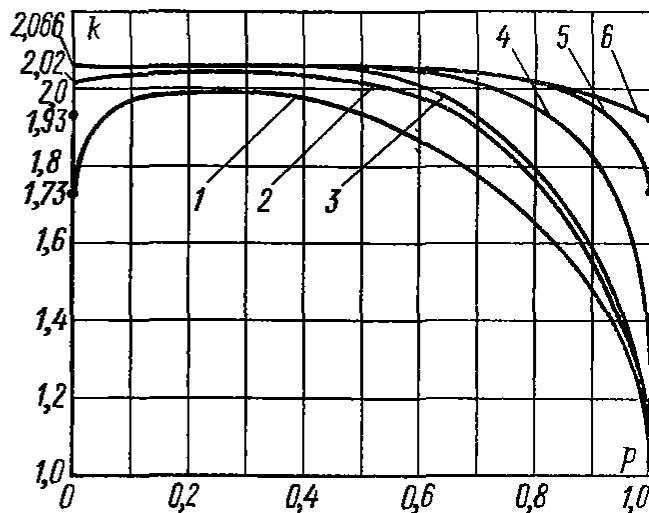


Рис. 3-3

на графике рис. 3-2 изображается горизонтальной прямой 1. При суммировании равномерно распределенной погрешности ($k_1 = 1,73$) с нормально распределенной ($k_2 = 2,066$) изменение k_{Σ} соответствует кривой 2.

Если обе суммируемые погрешности распределены равномерно, то при возрастании с. к. о. второго слагаемого от $\sigma_2 = 0$ до $\sigma_2 = \sigma_1$ их композиция имеет форму трапеции и изменяется от равномерной с $k_{\Sigma} = 1,73$ до треугольной с $k_{\Sigma} = 2,02$. При дальнейшем возрастании σ_2 от $\sigma_2 = \sigma_1$ до $\sigma_2 \rightarrow \infty$ вид распределения проходит все эти формы в обратном порядке. Поэтому изменение k_{Σ} в этом случае соответствует выражению (3-2) и кривой 3 (рис. 3-2).

При суммировании двух арксинусоидально распределенных погрешностей k_{Σ} изменяется аналогичным образом согласно кривой 5 (рис. 3-2), от $k_{\Sigma} = 1,11$ при $\sigma_2 = 0$ ($p = 0$) до $k_{\Sigma} = 1,88$ при $p = 0,5$ и далее вновь убывает до $k_{\Sigma} = 1,11$ при $p = 1$. Суммирование арксинусоидально распределенной составляющей с σ_1 и равномерно распределенной составляющей с σ_2 соответствует кривая 4, начинающаяся при $p = 0$ и $k_{\Sigma} = 1,11$, достигающая максимума $k_{\Sigma} = 1,98$ и заканчивающаяся при $p = 1$ и $k_{\Sigma} = 1,73$.

На рис. 3-3 кривые 1—3 соответствуют суммированию равномерного, треугольного и нормального распределений с дискретным двузначным распределением, а кривые 4—6 — суммированию нормального распределения соответственно с арксинусоидальным, равномерным и экспоненциальным.

Несмотря на то, что кривые рис. 3-2 и 3-3 построены только для нескольких видов законов распределения, их сетка настолько густа, что позволяет на глаз интерполировать значения k_{Σ} для композиций любых законов распределения с известным энтропийным коэффициентом, тем более, что значения энтропийного коэффициента точнее, чем до 0,1 (т. е. примерно 5%), уточнять не имеет смысла.

Однако при расчете суммарной погрешности на ЭВМ пользование кривыми неудобно. Для этого предпочтительнее использовать

вместо рис. 3-2 и 3-3 приближенные формулы, аппроксимирующие эти кривые. Так, семейство симметричных кривых 1, 3, 5 рис. 3-2 с погрешностью 1,5% при $k \geq 1,8$ и с погрешностью 2,5% при $k \geq 1,11$ описывается соотношением

$$k_{\Sigma} = k + 1,15 (k_{\Sigma} - k)^{0,75} [p(1-p)]^{0,21k^{0,18}},$$

где k — энтропийный коэффициент суммируемых составляющих; p и $1-p$ — веса суммируемых составляющих; $k_{\Sigma} = \sqrt{\pi e/2} = 2,066$ — энтропийный коэффициент нормального распределения.

Несимметричные кривые вида кривой 4 на рис. 3-2, начинающиеся при $k = k_1$ и оканчивающиеся при $k = k_2$, где $k_1 < k_2$, с погрешностью 2% описываются соотношением

$$k_{\Sigma} = k_1 + (k_2 - k_1) [p + a_0 p^{a_1} (1-p)^{a_2}],$$

где p — вес составляющей с $k = k_1$; $1-p$ — вес составляющей с $k = k_2$; $a_0 = 0,817 (1 + k_1/k_2)$; $a_1 = 0,75 (k_1/k_2)^{4,5}$; $a_2 = 1,23 \times (k_1/k_2)^2$.

Несимметричные кривые вида кривой 2 на рис. 3-2 и кривых 4—6 на рис. 3-3 для суммирования нормально распределенных погрешностей с погрешностями, распределенными с энтропийным коэффициентом k , с погрешностью 2% при $k_{\Sigma} \geq 1,9$; 4% до $k_{\Sigma} \geq 1,8$; 10% до $k_{\Sigma} \geq 1,5$ и 15% до $k_{\Sigma} \geq 1,4$ описываются соотношением

$$k_{\Sigma} = k_{\Sigma} - p^{1,4(5,7-k)} [0,14 + 0,4 (k_{\Sigma} - k)^2],$$

где p — вес составляющей с энтропийным коэффициентом k .

Если после определения энтропийного значения результирующей погрешности потребуется выразить его в виде доверительного значения погрешности, то для этого можно воспользоваться формулой (2-6). Однако при этом следует иметь в виду, что вследствие неточности используемой оценки с. к. о. σ_{Σ} или энтропийного коэффициента k_{Σ} и оценка P_d доверительной вероятности будет также иметь соответствующий интервал неопределенности. Поэтому полученное значение P_d необходимо округлять и выражать не более чем двумя знаками (т. е. в пределах от 0,90 до 0,99 даже в том случае, если по формуле оно получится, например, $P_d = 1,02$).

3.3. МЕТОДИКА РАСЧЕТА

РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ С ПРОИЗВОЛЬНЫМ ЗНАЧЕНИЕМ ДОВЕРИТЕЛЬНОЙ ВЕРОЯТНОСТИ

Описываемый ниже метод расчета результирующей погрешности с произвольным значением заданной доверительной вероятности в литературе неизвестен. Его использование стало возможным лишь после того, как были получены достаточно точные аппроксимирующие формулы, позволяющие рассчитать значения

квантильного множителя i для различных значений P_d и разных классов распределений (см. § 2-6). Для того чтобы воспользоваться этими формулами для определения доверительного значения результирующей погрешности, необходимо располагать оценкой эксцесса ε распределения результирующей погрешности.

Так как формулы (2-19)—(2-22) имеют разный вид для разных классов распределений (экспоненциальных, уплощенных, двухмодальных островершинных и двухмодальных кругловершинных), то необходимо решить, к какому из этих классов относится данное распределение результирующей погрешности.

Значение эксцесса распределения суммы двух независимых случайных величин можно вычислить аналитически. Действительно, дисперсия композиции двух распределений σ_Σ^2 определяется дисперсиями $\sigma^2(x)$ и $\sigma^2(y)$ как $\sigma_\Sigma^2 = \sigma^2(x) + \sigma^2(y)$.

Четвертый момент композиции

$$\begin{aligned}\mu_{4\Sigma} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y)^4 p(x) p(y) dx dy = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4) p(x) p(y) dx dy = \\ &= \mu_4(x) + 4\mu_3(x)\mu_1(y) + 6\mu_2(x)\mu_2(y) + 4\mu_1(x)\mu_3(y) + \mu_4(y).\end{aligned}$$

Для симметричных (не скошенных) распределений $\mu_3(x) = 0$ и $\mu_3(y) = 0$, при этом $\mu_{4\Sigma} = \mu_4(x) + 6\sigma^2(x)\sigma^2(y) + \mu_4(y)$.

Эксцесс суммарного распределения

$$\varepsilon_\Sigma = \frac{\mu_{4\Sigma}}{\sigma_\Sigma^4} = \frac{\mu_4(x) + 6\sigma^2(x)\sigma^2(y) + \mu_4(y)}{[\sigma^2(x) + \sigma^2(y)]^2}.$$

Обозначая вес дисперсии первого распределения в общей дисперсии как

$$p = \sigma^2(x)/[\sigma^2(x) + \sigma^2(y)], \quad (3-4)$$

получим

$$\varepsilon_\Sigma = \frac{\mu_4(x)}{\sigma^4(x)} p^2 + 6p(1-p) + \frac{\mu_4(y)}{\sigma^4(y)} (1-p)^2$$

или окончательно

$$\varepsilon_\Sigma = \varepsilon(x) p^2 + 6p(1-p) + \varepsilon(y) (1-p)^2. \quad (3-5)$$

Таким образом, для расчета результирующей погрешности этим методом для каждой из суммируемых составляющих надо знать оценки $\sigma(x)$, $\varepsilon(x)$ и $\sigma(y)$, $\varepsilon(y)$. Далее находится $\sigma_\Sigma = \sqrt{\sigma^2(x) + \sigma^2(y)}$, по выражению (3-4) определяется вес дисперсии одной из составляющих, а по (3-5) находится оценка ε_Σ суммарной погрешности.

Если суммируемых составляющих больше двух, то суммирование и определение ε_Σ производится последовательно — первой

со второй, затем полученной с третьей, вновь полученной с четвертой и т. д. Таким образом могут быть определены расчетные значения σ_x и ε_x при суммировании любого числа составляющих.

Нестрогость этого метода состоит в том, что данное (или полученное расчетом) значение эксцесса ε_x не определяет однозначно вида закона распределения.

Формулы для расчета $t = f(\varepsilon_x, P_d)$ были получены в § 2-6 для следующих классов распределений (см. рис. 2-10): формула (2-19)—для экспоненциальных и трапецеидальных распределений, т. е. распределений, отображающие точки которых располагаются на топографической классификации рис. 2-10 по дуге 6—7—5—12—8; формула (2-20) — для двухмодальных кругловершинных распределений, располагающихся на рис. 2-10 по дуге 5—1, а формула (2-12) — для островершинных двухмодальных распределений, располагающихся на рис. 2-10 по дуге 6—1. И, наконец, формула (2-21) предназначена для уплощения распределений типа шапо, располагающихся на рис. 2-10 по дуге 6—8 кривой 9.

Поэтому, получив при расчете, например, $\varepsilon_x = 2,8$, т. е. $\kappa = 1/\sqrt{\varepsilon} = 0,6$, и нанеся на рис. 2-10 вертикаль при $\kappa = 0,6$, мы видим, что она одновременно пересекает все четыре кривые, и по какой из четырех формул должен производиться расчет, на основании этих данных указать невозможно.

Поэтому этот метод удобно использовать лишь совместно с предыдущим. Действительно, если совместно с ε_x будет рассчитано и k_x , то по топографической классификации рис. 2-10 будет легко решить, к какой из кривых ближе данное распределение с точки зрения погрешности и какая из формул может быть применена.

3-4. ВОЗМОЖНЫЕ УПРОЩЕНИЯ МЕТОДИКИ СУММИРОВАНИЯ ПОГРЕШНОСТЕЙ

Наиболее трудными моментами изложенной методики суммирования погрешностей являются нахождение с. к. о. всех составляющих по известным их интервальным оценкам и определение интервальной оценки результирующей погрешности по полученному в результате расчета ее с. к. о., так как и тот и другой переход требуют знания формы закона распределения. Отсюда пути возможного упрощения методики суммирования погрешностей сводятся к использованию упрощенных методов осуществления этих переходов.

Один из возможных методов такого упрощения может основываться на том, что при суммировании большого числа составляющих закон распределения результирующей погрешности все более и более приближается к нормальному. Однако без соответствующего анализа (хотя бы определения k_x или κ_x) степень близости распределения результирующей погрешности к нормаль-

ному установить невозможно. Поэтому предположение о близости распределения результирующей погрешности к нормальному достаточно рискованно даже и при большом числе суммируемых составляющих. Тем не менее при недостатке времени и невысоких требованиях к точности получаемого результата использование такого предположения вполне возможно.

Методика такого суммирования до момента расчета оценок σ_z в начале и в конце диапазона измерений не будет ничем отличаться от описанных выше методик, однако переход от σ_z к γ_z будет выполняться по формуле $\gamma_d = t_d \sigma_z$, где t_d — квантильный множитель, определяемый по таблицам квантилей нормального распределения (см. стр. 142) или по формуле (2-19), полагая $v = 3$. Такой прием существенно уменьшает трудоемкость расчетного суммирования погрешностей, но может вносить существенные ошибки, если фактическое распределение погрешностей далеко от нормального. (Например, при фактически арксинусоидальном распределении ошибка может достигать $2,07/1,11 = 1,8$ раза.) Поэтому использовать этот метод следует весьма осмотрительно.

В качестве другого пути упрощения перехода от σ_z к γ_z следует указать возможность использования для результирующей погрешности значения доверительной вероятности $P_d = 0,9$, при котором для большой группы классов различных распределений (ограничения были указаны в § 2-6) сохраняется постоянным соотношение $\Delta_{0,9}/\sigma = 1,6$, так как при суммировании погрешностей любого сочетания распределений из этой группы классов результирующее распределение также будет принадлежать этой же группе классов и для него будет справедливо соотношение $\gamma_{z0,9} = 1,6\sigma_z$.

Это уникальное свойство погрешности при $P_d = 0,9$ открывает возможность резкого упрощения метода расчетного суммирования погрешностей. Так, если суммируемые составляющие за-

даны своими значениями $\Delta_{0,9i}$, то $\sigma_i = \Delta_{0,9i}/1,6$; $\sigma_z = \sqrt{\sum_{i=1}^n \sigma_i^2}$ и $\Delta_{0,9z} = 1,6\sigma_z$ или просто

$$\Delta_{0,9z} = \sqrt{\sum_{i=1}^n \Delta_{0,9i}^2}. \quad (3-6)$$

Исходя из изложенного, предпочтительным значением доверительной вероятности при нормировании случайных погрешностей является $P_d = 0,9$, тем более, что оценка $\Delta_{0,9}$ определяется по экспериментальным данным (см. § 2-2) с гораздо большей точностью, чем, например, $\Delta_{0,97}$ или $\Delta_{0,99}$.

Используя доверительные границы $\pm \Delta_d$ погрешности, необходимо иметь в виду следующее. Эти границы располагаются симметрично лишь при отсутствии у прибора или измерительного канала систематической составляющей погрешности θ . Если

$\theta \neq 0$, то границы погрешности оказываются несимметричными. Так, например, если $\gamma_d = \pm 0,4\%$, а $\theta = +0,1\%$, то одна граница оказывается равной $-0,4 + 0,1 = -0,3\%$, а другая $+0,4 + 0,1 = +0,5\%$. Знак погрешности θ чаще всего неизвестен, поэтому ввести поправку невозможно. Пользоваться же при дальнейших расчетах такими несимметричными границами погрешностей крайне неудобно. Поэтому на практике вместо использования несимметричных границ переходят к указанию симметричных границ, равных по модулю большей из несимметричных, т. е. вместо указания «погрешность находится в пределах от $-0,3\%$ до $+0,5\%$ » переходят к указанию «погрешность находится в пределах $\pm 0,5\%$ ». Вероятность выхода погрешности за такие симметричные границы, естественно, в два раза меньше, так как такой выход происходит практически только с одной стороны, а не с обеих сторон. В результате, если $\gamma_d = \pm 0,4\%$ была определена с $P_d = 0,9$, то $\gamma_d = \pm 0,5\%$ есть погрешность с доверительной вероятностью $P_d = 0,95$. Таким образом, при $\theta \neq 0$, а точнее при $\theta > 0,66\sigma$

$$\Delta_{0,95} = (|\theta| + \Delta_{0,9}) = (|\theta| + 1,6\sigma),$$

т. е. результирующая погрешность $\Delta_{0,95}$ очень просто определяется через $|\theta|$ и $\Delta_{0,9}$ случайной составляющей.

Однако дальнейшие упрощения методики суммирования, как-то: пренебрежение делением погрешностей на коррелированные и некоррелированные (а следовательно, суммирование составляющих, которые в действительности должны вычитаться) или подразделением на аддитивные и мультипликативные, уже недопустимы, так как они могут привести к существенно ошибочным результатам.

3.5. ПРИМЕР РАСЧЕТА ПОГРЕШНОСТИ ИЗМЕРИТЕЛЬНОГО КАНАЛА

Хотя изложенные выше правила суммирования погрешностей не очень сложны, но практическое выполнение таких расчетов требует как привлечения дополнительных исходных сведений, так и принятия по ходу расчета целого ряда эвристических решений и вызывает у исполнителя подчас серьезные затруднения. Целесообразно рассмотреть ход такого решения на конкретном примере.

Пусть требуется рассчитать результирующую погрешность измерительного канала, созданного самим экспериментатором из серийно выпускаемых узлов. При этом канал пусть состоит всего из трех узлов — реостатного датчика D , усилителя $У$ и регистратора P (рис. 3-4). Погрешность реостатного датчика с сопротивлением $R_d = 200$ Ом нормирована в его паспорте одним числом границ максимального значения приведенной погрешности $\gamma_{md} = 0,15\%$, т. е. предполагается, что его погрешность аддитивна.

Датчик питается напряжением 200 мВ через стабилизатор напряжения с коэффициентом стабилизации $K = 25$ от общего в усилителе стандартного нестабилизированного блока питания. Усилитель U предназначен для обеспечения линейности характеристики всего канала, для чего имеет входное сопротивление, много большее сопротивления датчика. Об усилителе известно, что он выполнен в виде эмиттерного повторителя и благодаря глубокой отрицательной обратной связи коэффициент влияния колебаний напряжения U его питания на ко-

эффициент усиления снижен до значения $\psi_{Uy} = +0,3\%/(10\% \Delta U/U)$, а влияние температуры приводит к смещению его нуля на $\psi_{\theta y} = +0,2\%/10 \text{ K}$.

В качестве регистратора использован электронный самопишущий автоматический потенциометр класса точности 0,5 с записью регистрируемого процесса на широкую бумажную ленту. Время прохода регистратором всей шкалы составляет 0,5 с. Изменение температуры вызывает смещение нуля регистратора на $\psi_{\theta \text{рег}} = -0,1\%/10 \text{ K}$.

Параллельно с аналоговым регистратором (для последующей обработки данных эксперимента на ЭВМ) включен цифровой вольтметр класса точности 0,2/0,1 с отсчетом трех десятичных знаков (например, Ф203), к которому через транскриптор подключен перфоратор ПЛ-150, регистрирующий данные измерений с частотой 5 отсчетов в секунду.

Установка питается от сети 220 В, 50 Гц с нестабильностью напряжения $\pm 15\%$. Объект испытаний, на котором установлен датчик, находится в испытательном цехе или ангаре, где обычно поддерживается температура около 20°C , но летом может достигать 35°C , а зимой при открывании въездных ворот может падать до 5°C , т. е. может изменяться в пределах $(20 \pm 15)^\circ\text{C}$.

Усилитель и регистрирующая аппаратура установлены в лаборатории, где температура колеблется от 18 до 24°C .

В качестве линии связи с датчиком используется стационарно проложенная в траншеях цеха проводка, каждая жила которой имеет сопротивление около 2 Ом. Напряжение наводки частотой

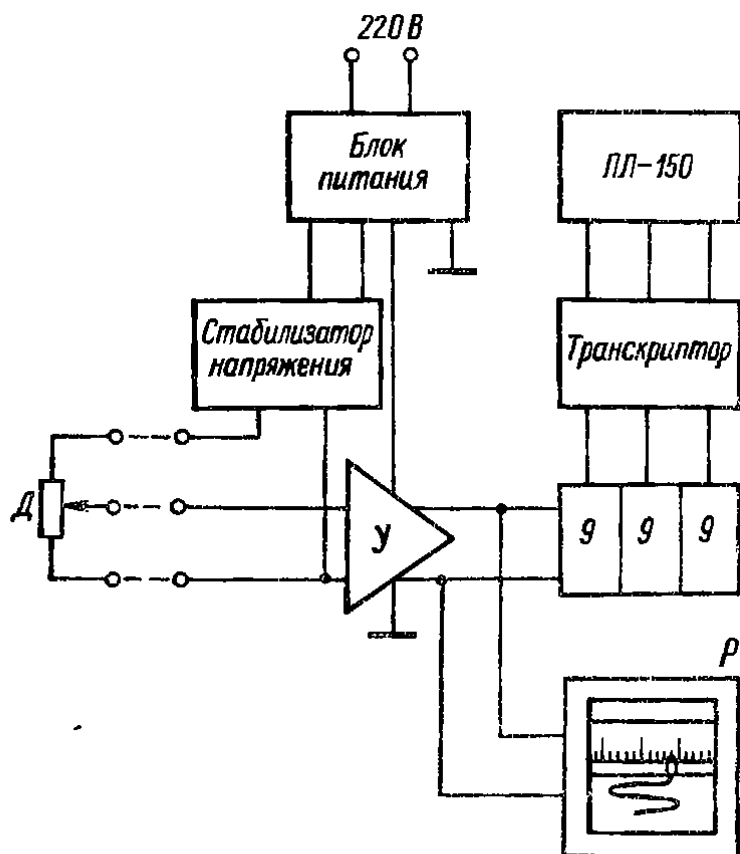


Рис. 3-4

50 Гц, измеренное между заземленным проводом линии и одной из изолированных жил с помощью электронного вольтметра или осциллографа с входным сопротивлением 1 МОм, составляет 1,6 В.

Читателя, возможно, уже утомили перечисленные подробности, но они совершенно необходимы для сколь-нибудь достоверной оценки погрешности предстоящих измерений. Особенно это относится к погрешностям от наводок, которые, естественно, не могут быть указаны в паспортных данных используемых средств измерений.

При расчете результирующей погрешности канала прежде всего, как указывалось выше, каждой из составляющих погрешности нужно приписать соответствующий закон распределения, найти с. к. о. и разделить погрешности на аддитивные и мультипликативные.

Условимся все расчеты как максимальных погрешностей, так и с. к. о. вести в относительных приведенных значениях и сохранять при промежуточных округлениях один лишний недостоверный знак в их значениях, с тем чтобы округление по правилам, изложенным в § 1-4, произвести лишь над окончательным результатом.

Погрешность от наводки на линию связи может быть приближенно рассчитана следующим образом. При присоединении к линии электронного вольтметра с входным сопротивлением 1 МОм на его входе возникало падение напряжения 1,6 В, т. е. через него проходил ток $I = 1,6/10^6 = 16 \cdot 10^{-7}$ А. При замыкании линии на датчик сопротивлением $R_d = 200$ Ом падение напряжения на нем от этого тока составит $U = 16 \cdot 10^{-7} \cdot 200 = 3,2 \cdot 10^{-4} = 0,32$ мВ. Измерить такое напряжение трудно, поэтому мы и находим его расчетным путем. Таким образом, приведенная погрешность от наводки при максимальном сигнале с датчика, равном 200 мВ, будет $\gamma_{\text{нав}} = 0,32/200 = 0,16\%$.

Так как исходное значение наводки в 1,6 В было отсчитано по шкале вольтметра, т. е. являлось действующим (средним квадратическим) значением, то полученная оценка погрешности от наводки и есть ее с. к. о., т. е. $\sigma_{\text{нав}} = 0,16\%$. Ее максимальное значение $\gamma_{m \text{ нав}} = \sigma \sqrt{2} = 0,16 \cdot \sqrt{2} = 0,225\%$, а закон распределения — арксинусоидальный. Следовательно (см. табл. 2-3), энтропийный коэффициент этого распределения $k_{\text{нав}} = 1,11$, контрэксцесс $\kappa = 0,816$, эксцесс $\varepsilon = 1,5$.

Погрешность, возникающая от наводки посторонних напряжений на вход измерительного канала, как правило, является аддитивной, так как не зависит от величины измеряемого сигнала. Но данный конкретный случай является исключением. Здесь, как видно из схемы рис. 3-4, сопротивление, на которое замкнут вход измерительного усилителя, есть сопротивление нижней части реостатного датчика. При входной измеряемой величине, равной нулю, это сопротивление также равно нулю. Поэтому напряжение

наводки возрастает линейно с ростом входного сигнала, т. е. в данном случае, как исключение из общего правила, погрешность от наводки оказывается мультипликативной.

Основная погрешность датчика нормирована по паспорту максимальным значением $\gamma_{\text{мд}} = 0,15\%$. Для того чтобы от этого значения перейти к с. к. о., необходимо знание вида закона распределения погрешности. Одной из составляющих погрешности датчика является погрешность дискретности, обусловленная конечным числом витков его обмотки, по которым скользит подвижный контакт. Эта погрешность имеет, как было указано выше (см. § 2-5), равномерное распределение. Но если обмотка датчика имеет, например, 500 витков, то погрешность дискретности, соответствующая $\pm 0,5$ витка, составляет $0,001 = 0,1\%$. А ему нормируется погрешность $0,15\%$. Следовательно, есть еще какие-то составляющие погрешности, которые нам неизвестны (погрешность линейности, т. е. неравномерность намотки, люфт в опорах оси подвижного контакта и т. п.). Но так как они незначительно увеличили результирующую погрешность, то превалирующей, видимо, является погрешность дискретности и поэтому общее распределение погрешности можно считать близким к равномерному и приближенно принять равномерным. Тогда $\gamma_{\text{мд}} = 0,15\%$ можно считать половиной ширины этого равномерного распределения и найти с. к. о. как $\sigma_{\text{д}} = \gamma_{\text{мд}}/\sqrt{3} = 0,15/\sqrt{3} = 0,087\%$. Для равномерного распределения $k = 1,73$, $\kappa = 0,745$ и $\varepsilon = 1,8$.

Температурная погрешность датчика в его паспорте не указана, так как у самого датчика она отсутствует (коэффициент деления напряжения не зависит от температуры при одинаковых температурных коэффициентах обоих сопротивлений делителя). Но у нас датчик с $R_{\text{д}} = 200$ Ом включен последовательно с двумя жилами медной линии сопротивлением в 2 Ом каждая. При изменении температуры в цехе, где проложена линия связи, в диапазоне $(20 \pm \pm 15^\circ\text{C})$ и температурном коэффициенте меди $\alpha_{\text{в}} = +4\%/10^\circ\text{K}$ изменение сопротивления каждой из жил составит $\Delta R = \Delta\theta\alpha_{\text{в}}R = \frac{4}{100} \cdot \frac{15}{10} \cdot 2 = 120 \cdot 10^{-8} = 0,12$ Ом, что по отношению к $R_{\text{д}} = 200$ Ом составляет $0,06\%$, т. е. величину, соизмеримую с другими погрешностями. Эта погрешность может быть исключена изменением схемы включения датчика (заменой питающего датчик стабилизатора напряжения на стабилизатор тока и переходом с трехпроводной линии на четырехпроводную). Но если этого не сделано, то возникающая погрешность хотя бы приближенно должна быть учтена при расчете результирующей погрешности канала. Приближение может состоять в следующем. Если подвижный контакт датчика находится точно в среднем положении, то погрешность отсутствует. Она максимальна лишь при крайних положениях этого контакта. Но при крайнем верхнем (на схеме) положении контакта эта погрешность, по-видимому, будет невелика по сравнению с другими мультипликативными погрешно-

стями и ею можно в первом приближении пренебречь. Но при крайнем нижнем положении контакта возникающая погрешность смещения нуля должна быть оценена.

Для перехода от вычисленного выше максимального значения этой погрешности $\gamma_{т\theta д} = 0,06\%$, возникающего при предельных отклонениях температуры до 5 или 35 °С, к с. к. о. необходимо знать закон распределения температуры в испытательном цехе. Какие-либо данные об этом у нас отсутствуют. Известно лишь, что в большинстве случаев температура там близка к 20 °С и крайне редко достигает значений 5 и 35 °С. Примем совершенно эвристическое предположение, что она распределена нормально и 4 дня в году бывает ниже 5 °С и 4 дня в году — выше 35 °С, а остальные $365 - 8 = 357$ дней, т. е. $357/365 = 0,98$ случаев, не выходит за эти пределы. По таблице нормального распределения (см. таблицу на стр. 142) находим, что вероятности $P = 0,98$ соответствуют границы в $\pm 2,3\sigma$. Отсюда искомая $\sigma_{\theta д} = 0,06/2,3 = 0,026\%$, а параметры закона распределения (см. табл. 2-4) $k = 2,066$, $\varepsilon = 3$, $\kappa = 0,577$.

Погрешность датчика от колебаний напряжения питания является чисто мультипликативной и распределена по тому же закону, что и отклонения напряжения сети от своего номинального значения 220 В. Как было указано в § 2-5, распределение напряжения сети близко к треугольному с принятыми выше пределами $\pm 15\%$. Стабилизатор снижает размах колебаний напряжения в $K = 25$ раз, т. е. на выходе стабилизатора распределение также треугольное, но с размахом $15\%/25 = 0,6\%$. Поэтому максимальное значение этой погрешности $\gamma_{тU д} = 0,6\%$. Среднее квадратическое отклонение для треугольного распределения (см. табл. 2-2) $\sigma = \gamma_m/\sqrt{6}$, поэтому $\sigma_{U д} = 0,6/\sqrt{6} = 0,245\%$. Параметры этого распределения: $k = 2,02$, $\varepsilon = 2,4$, $\kappa = 0,65$.

Погрешность коэффициента усиления усилителя является мультипликативной и распределена также по треугольному закону, так как вызывается колебаниями напряжения питания. Ее максимальное значение составляет $\gamma_{тU у} = 0,03\% \cdot 15 = 0,45\%$, а с. к. о. $\sigma_{U у} = 0,45/\sqrt{6} = 0,184\%$. Параметры распределения уже были указаны выше.

Погрешность смещения нуля усилителя при колебании температуры является аддитивной, а закон ее распределения повторяет закон распределения температуры в лаборатории, где установлены усилитель и регистраторы. Как указывалось в § 2-5, закон распределения температуры в лаборатории в пределах от 18 до 24 °С можно считать равномерным со средним значением 21 °С и размахом ± 3 К. Максимальное значение этой погрешности при $\psi_{\theta у} = -0,2\%/10$ К составляет $\gamma_{т\theta у} = 0,2\% \cdot \frac{3}{10} = 0,06\%$, а с. к. о. (см. табл. 2-2) $\sigma_{\theta у} = 0,06/\sqrt{3} = 0,034\%$. Параметры равномерного распределения уже были указаны выше.

Основная погрешность аналогового регистратора определяется его классом точности. Однако в отличие от датчиков, погрешность которых, как правило, нормируется без запаса на старение, погрешность всех электроизмерительных приборов согласно стандарту нормируется с 25%-ным запасом на старение, т. е. фактическая погрешность нового, только что выпущенного заводом прибора составляет не больше чем $0,8\gamma_{\text{кл}}$. Если же прибор уже ремонтировался на предприятии-потребителе, то этот запас зависит от уровня метрологической службы на данном предприятии. Так, например, обследование многих тысяч электроизмерительных приборов на ЛПЭО «Электросила» им. С. М. Кирова показало, что службы главного метролога создали на предприятии такой климат, что ремонтные службы выпускают электроизмерительные приборы, имеющие погрешность $0,4\gamma_{\text{кл}}$, т. е. тот же уровень, который имеют и ведущие приборостроительные заводы. Благодаря этому приборам обеспечивается большой межремонтный срок службы, пока их погрешность возрастает от $0,4\gamma_{\text{кл}}$ до $\gamma_{\text{кл}}$, а в среднем по объединению погрешность находящихся в эксплуатации приборов составляет $0,7\gamma_{\text{кл}}$.

Если же ремонт производится так, что погрешность прибора еле-еле укладывается в класс, то прибор большую часть межповерочного срока работает с погрешностью, превышающей класс точности, и бракуется при следующей ежегодной поверке.

Будем считать, что в нашем примере используется только что полученный с завода новый регистратор и при классе точности 0,5 его погрешность составляет не более $\gamma_{\text{т рег}} = 0,8 \cdot 0,5 = 0,4\%$. Погрешность потенциометра определяется прежде всего погрешностью реохорда и поэтому она аддитивна, а закон ее распределения, как и у реостатного датчика, будем считать равномерным с шириной $\pm 0,4\%$. Тогда с. к. о. этой погрешности $\sigma_{\text{рег}} = 0,4/\sqrt{3} = 0,23\%$.

Температурная погрешность регистратора проявляется в виде смещения нуля на $\psi_{\text{о рег}} = -0,1\%/10 \text{ К}$, также аддитивна и при принятом выше равномерном законе распределения температуры шириной $\pm 3 \text{ К}$ ее с. к. о. составляет $\sigma_{\psi \text{ рег}} = 0,03/\sqrt{3} = 0,017\%$.

Погрешность цифрового вольтметра (ЦВ) нормирована двучленной формулой. Ее приведенное значение равно 0,1% при $x = 0$ и линейно возрастает до 0,2% в конце шкалы. Но при использовании в канале измерения ЦВ или АЦП чаще всего возникает следующая ситуация. Максимальный сигнал датчика в нашем случае равен 200 мВ. При номинальном коэффициенте усиления усилителя, равном единице, выходное напряжение также равно 200 мВ. Подобрать самописец на такой предел измерений или подогнать его предел с помощью делителя не составляет трудности. Пределы же измерений цифровых приборов, как правило, кратны 10. Так, например, указанный выше ЦВ типа Ф203 имеет пределы

измерений 100, 10 и 1 В. При использовании в рассматриваемом канале измерений на пределе 1 В = 1000 мВ он обеспечивает удобный отсчет измерений величины непосредственно в милливольтах, но расчет его погрешности в этом случае имеет некоторые особенности. Так как такое неполное использование рабочего диапазона ЦВ или АЦП очень часто встречается в практике измерений, рассмотрим его более подробно.

Итак, в цифровом приборе с диапазоном измерений от 0 до 1000 мВ используется лишь часть этого диапазона от 0 до 200 мВ, так как выходное напряжение в 200 мВ является предельным выходным напряжением данного канала измерений. Особенность расчета погрешности в данном случае состоит в том, что приведенное значение погрешности должно рассчитываться для предела измерений именно 200 мВ, в то время как приведенная погрешность цифрового прибора дана для предела измерений 1000 мВ. Для этого должна быть вычислена абсолютная погрешность ЦВ в точках начала и конца этого нового диапазона и отнесена к его концу.

Абсолютная погрешность при $x = 0$ ЦВ класса c/d

$$\Delta_0 = \frac{d}{100} X_{\text{н. цв}} = \frac{0,1}{100} \cdot 1000 = 1 \text{ мВ},$$

а ее приведенное значение к пределу измерений $X_{\text{н. кан}}$ данного канала $\gamma_{\text{н. цв}} = 1 \text{ мВ} / 200 \text{ мВ} = 0,5\%$. Абсолютная погрешность ЦВ при $x = 200 \text{ мВ}$

$$\Delta_{200} = \Delta_0 + \frac{c-d}{100} x = 1 + \frac{0,2-0,1}{100} \cdot 200 = 1,2 \text{ мВ},$$

а ее приведенное значение к пределу измерений канала

$$\gamma_{\text{н. цв}} = \frac{\Delta_{200}}{X_{\text{н. кан}}} = \frac{1,2 \text{ мВ}}{200 \text{ мВ}} = 0,6\%.$$

Таким образом, неполное использование диапазона ЦВ приводит к существенному возрастанию приведенной погрешности измерения, в данном примере с 0,2/0,1 до 0,6/0,5. Однако вследствие того, что оно обеспечивает удобство цифрового отсчета непосредственно в единицах измеряемой величины, такой прием широко используется на практике. Тем более важен правильный расчет погрешностей для этого случая.

Для перехода от максимальной погрешности ЦВ или АЦП к с. к. о. необходимо знание вида закона распределения этой погрешности. В § 2-5 указывалось, что это распределение является композицией равномерного распределения погрешности квантования и очень полого спадающего распределения суммы погрешностей нуля-органа, аналоговых узлов ЦВ и АЦП и разброса отдельных ступеней кодирующей сетки, которое в среднем можно считать экспоненциальным распределением с показателем степени $\alpha = 0,5$, т. е. вида $p(x) = 0,25e^{-\sqrt{|x|}}$. У высокоточных ЦВ и АЦП

с большим числом разрядов экспоненциальное распределение (как отмечалось в § 2-5) преобладает над равномерным, а у низкочастотных — наоборот. Так, например, у ЦВ типа Ф203 дисперсия экспоненциальной составляющей равна от 2 до 13% общей дисперсии [8]. Поэтому при грубой оценке погрешностей этой составляющей можно пренебречь и считать распределение близким к равномерному. Для более точного расчета погрешностей дисперсию этой составляющей можно в среднем принять равной 8% или $1/13$ общей дисперсии.

Таким образом, если полагать в нашем примере распределение погрешности ЦВ равномерным, то с. к. о. будет $\sigma_{н. ЦВ} = 0,5/\sqrt{3} = 0,289\%$; $\sigma_{к. ЦВ} = 0,6/\sqrt{3} = 0,346\%$.

При учете экспоненциальной составляющей погрешности ЦВ следует считать состоящей из двух частей: при $x = 0$

$$\sigma_{1 н. ЦВ} = \sqrt{0,289^2 \frac{12}{13}} = 0,278\%;$$

$$\sigma_{2 н. ЦВ} = \sqrt{0,289^2 \frac{1}{13}} = 0,080\%,$$

и в конце диапазона измерений канала

$$\sigma_{1 к. ЦВ} = \sqrt{0,346^2 \frac{12}{13}} = 0,332\%;$$

$$\sigma_{2 к. ЦВ} = \sqrt{0,346^2 \frac{1}{13}} = 0,096\%,$$

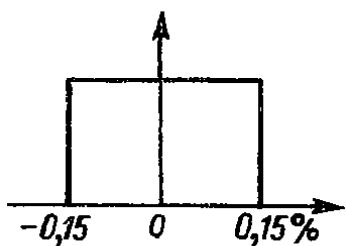
соответственно с равномерным распределением ($\sigma_{1 ЦВ}$) и экспоненциальным с $\alpha = 0,5$ распределением с параметрами (см. табл. 2-4) $k = 1,35$, $x = 0,2$, $v = 25,2$ ($\sigma_{2 ЦВ}$).

Итак, мы разделили все составляющие погрешности на аддитивные и мультипликативные, приписали им законы распределения и вычислили с. к. о. Этот результат для наглядности дальнейших действий представлен на рис. 3-5, где буквами А и М в кружках отмечены соответственно аддитивная и мультипликативная составляющие погрешности.

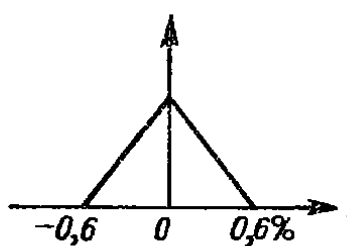
Суммирование погрешностей. Расчет результирующей погрешности канала сводится к вычислению приведенной погрешности при $x = 0$, которая складывается только из аддитивных составляющих, и в конце диапазона, которая складывается из всех составляющих. Эти операции нам придется проделать дважды: один раз — для канала с аналоговым регистратором, а другой раз — для канала с цифровым регистратором.

Выбор метода суммирования (складывать алгебраически или геометрически) зависит от того, являются ли суммируемые погрешности коррелированными или независимыми. Чтобы не допустить ошибок, целесообразно сразу выделить коррелированные погрешности и произвести их алгебраическое сложение. Коррелирован-

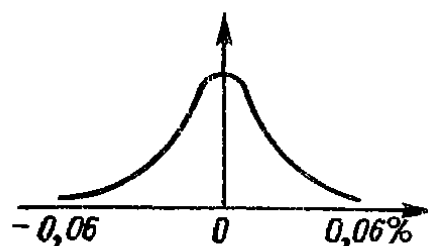
Датчик



$$\textcircled{A} \sigma_{\mu} = 0,087\%, \quad k = 1,73$$

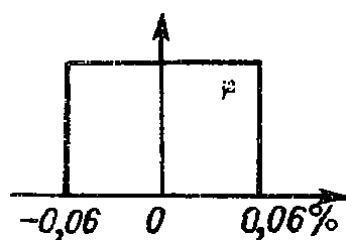


$$\textcircled{M} \sigma_{\mu_d} = 0,245\%, \quad k = 2,02$$

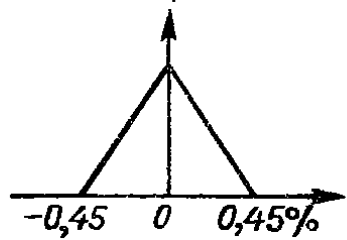


$$\textcircled{A} \sigma_{\theta_d} = 0,026\%, \quad k = 2,066$$

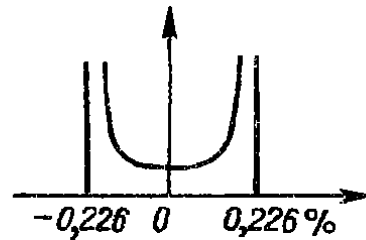
Усилитель



$$\textcircled{A} \sigma_{\theta_y} = 0,034\%, \quad k = 1,73$$

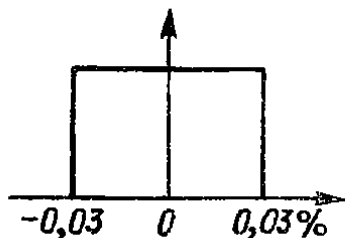


$$\textcircled{M} \sigma_{\mu_y} = 0,184\%, \quad k = 2,02$$

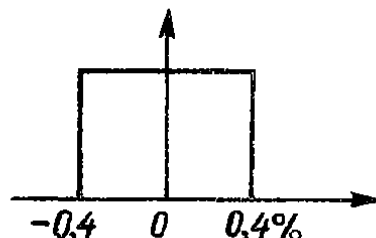


$$\textcircled{M} \sigma_{\text{нов}} = 0,16\%, \quad k = 1,11$$

Аналоговый регистратор

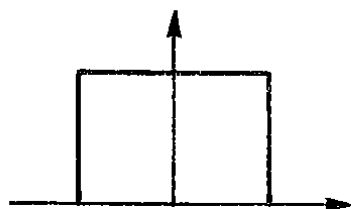


$$\textcircled{A} \sigma_{\theta_{\text{рег}}} = 0,017\%, \quad k = 1,73$$



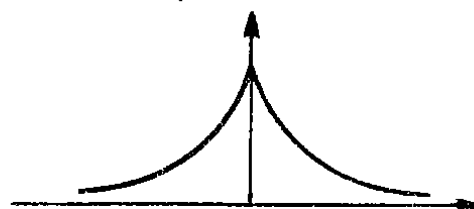
$$\textcircled{A} \sigma_{\text{рег}} = 0,23\%, \quad k = 1,73$$

Цифровой регистратор



$$\textcircled{A} \sigma_{1\text{н.цв}} = 0,287\%,$$

$$\textcircled{A} + \textcircled{M} \sigma_{1\text{к.цв}} = 0,332\%, \quad k = 1,73$$



$$\textcircled{A} \sigma_{2\text{н.цв}} = 0,080\%,$$

$$\textcircled{A} + \textcircled{M} \sigma_{2\text{к.цв}} = 0,096\%, \quad k = 1,35$$

ными являются те погрешности, которые вызываются одной и той же общей причиной, а поэтому имеют одинаковую форму закона распределения, которая остается справедливой и для их алгебраической суммы.

В нашем примере это погрешность датчика и усилителя от колебаний напряжения питания U , имеющая треугольный закон распределения, а также погрешность усилителя и регистратора от колебания температуры Θ в лаборатории, имеющая равномерный закон распределения. Но погрешность от колебания температуры датчика и температуры усилителя — это не коррелированные погрешности, так как их вызывает не одна и та же температура, а разные — температура в цехе и температура в лаборатории.

Для алгебраического суммирования коррелированных погрешностей необходимо установить их знаки. Так, коэффициент влияния на погрешность коэффициента усилителя от колебаний напряжения питания является положительным ($\psi_{Uy} = 0,3\%/(10\%\Delta U/U)$), т. е. коэффициент усиления с увеличением напряжения питания возрастает. Также положительным является коэффициент влияния на погрешность от колебания напряжения питания датчика. Поэтому результирующее значение этих погрешностей равно просто их сумме:

$$\sigma_{U(x+y)} = |\sigma_{Ud} + \sigma_{Uy}| = 0,245 + 0,184 = 0,429 \approx 0,43 \%,$$

а закон распределения этой суммарной погрешности сохраняется треугольным.

Температурная погрешность эмиттерного повторителя положительна ($\psi_{\Theta y} = +0,2\%/10\text{ K}$), так как с ростом температуры падение напряжения промежутка база — эмиттер падает, а следовательно, напряжение на эмиттерной нагрузке возрастает. Температурная же погрешность аналогового регистратора указана как отрицательная: $\psi_{\Theta \text{ рег}} = -0,1\%/10\text{ K}$. Следовательно, $\sigma_{\Theta(y+\text{рег})} = |\sigma_{\Theta y} - \sigma_{\Theta \text{ рег}}| = 0,034 - 0,017 = 0,017\%$.

Теперь, после учета корреляционных связей, все полученные погрешности можно суммировать как независимые.

Погрешность канала с аналоговым регистратором при $x = 0$ складывается из четырех составляющих: $\sigma_{\text{рег}} = 0,23\%$, $\sigma_d = 0,087\%$, $\sigma_{\Theta d} = 0,026\%$ и $\sigma_{\Theta(y+\text{рег})} = 0,017\%$. Однако $\sigma_{\Theta d}$ в 9 раз, а $\sigma_{\Theta(y+\text{рег})}$ — в 13,5 раза меньше, чем $\sigma_{\text{рег}}$. Так как суммирование под корнем будет производиться над квадратами этих величин, то их вклад в результат будет соответственно в $9^2 = 81$ и $13,5^2 = 183$ раза меньше. Отсюда ясно, что этими погрешностями можно пренебречь и опустить их из дальнейшего рассмотрения (см. п. 6 в конце § 3-1).

Итак, с. к. о. погрешности нуля канала с аналоговым регистратором определяется как

$$\begin{aligned} \sigma_n &= \sqrt{\sigma_{\text{рег}}^2 + \sigma_d^2} = \sqrt{0,23^2 + 0,087^2} = \sqrt{5,29 \cdot 10^{-2} + 7,57 \cdot 10^{-3}} = \\ &= 0,2459 \approx 0,25 \%. \end{aligned}$$

Если же не пренебрегать двумя опущенными составляющими, то было бы $\sigma_n = 0,2479 \approx 0,25\%$.

Обе суммируемые составляющие погрешности распределены равномерно, поэтому результирующее распределение является трапецеидальным. Для определения эксцесса и энтропийного коэффициента этого распределения нужно рассчитать вес дисперсии второго из слагаемых в общей дисперсии. Нужные для этого числа уже были получены (под корнем) при расчете σ_n . Поэтому $p = \sigma_d^2 / \sigma_n^2 = 7,57 \cdot 10^{-3} / 0,0605 = 0,12$; $(1 - p) = 0,88$ и эксцесс этого распределения в соответствии с выражением (3-5) будет $\varepsilon_n = \varepsilon_2 p^2 + 6p(1 - p) + \varepsilon_1(1 - p^2) = 1,8 \cdot 0,12^2 + 6 \cdot 0,12 \cdot 0,88 + 1,8 \cdot 0,88^2 = 2,08$, а контрэксцесс $\kappa = 0,69$.

Энтропийный коэффициент композиции двух равномерных распределений определяется по кривой 3 (рис. 3-2) при $p = 0,12$: $k_n = 1,93$. Отсюда энтропийное значение приведенной погрешности нуля канала $\gamma_n = k_n \sigma_n = 1,93 \cdot 0,25 = 0,4825 \approx 0,5\%$.

Для того чтобы получить более точное представление о виде закона распределения этой погрешности, можно обратиться к топографической классификации распределений, представленной на рис. 2-10, и найти там точку с координатами $\kappa_n = 0,69$ и $k_n = 1,93$. При необходимости представить полученную энтропийную оценку погрешности в форме доверительной погрешности по соотношению (2-4) можно рассчитать соответствующее ей значение доверительной вероятности. В результате этого получим $P_d = 0,899 + \kappa^2 / 5,5 = 0,899 + 0,69^2 / 5,5 = 0,98$, т. е. $\gamma_n = 0,5\%$ соответствует $\gamma_{0,98}$.

Для расчета погрешности в конце диапазона канала к полученному значению $\sigma_n = 0,25\%$ нужно добавить мультипликативные составляющие $\sigma_U (д+у) = 0,43\%$ и $\sigma_{нав} = 0,16\%$. Среди этих значений погрешности нет пренебрежимо малых, поэтому все они должны быть по очереди просуммированы.

При суммировании σ_n и σ_U (индекс «д + у» опустим)

$$\sigma_{n+U} = \sqrt{0,25^2 + 0,43^2} = \sqrt{0,0625 + 0,1849} = \sqrt{0,2474} = 0,497 \approx 0,5\%.$$

Погрешность от колебаний напряжения питания (σ_U) распределена по треугольному закону, а суммарная погрешность нуля (σ_n) — по трапецеидальному. Кривой для суммирования таких распределений на рис. 3-2 нет. Однако на рис. 3-3 есть кривая 2 для суммирования треугольного распределения с дискретным двузначным. Воспользуемся этой кривой следующим образом. Будем считать исходным распределением треугольное (σ_U), а добавляемым к нему — трапецеидальное (σ_n). Тогда нужная нам кривая будет проходить всегда выше кривой 2 на рис. 3-3, но она не может быть выше кривой 6, на рис. 3-3, соответствующей нормальному распределению. Узкая полоса между этими кривыми в

их начальной части и ограничивает возможное положение нужной нам кривой.

Вес дисперсии σ^2 в суммарной дисперсии $p = \sigma_{\text{н}}^2 / \sigma_{\text{н+U}}^2 = 0,25^2 / 0,2474 = 0,25$; $(1 - p) = 0,75$, эксцесс этого распределения $\epsilon_{\text{н+U}} = 2,08 \cdot 0,25^2 + 6 \cdot 0,25 \cdot 0,75 + 2,4 \cdot 0,75^2 = 2,6$ и контраэксцесс $\kappa = 0,62$. Согласно рис. 3-3 значению $p = 0,25$ соответствует $k_{\text{н+U}} = 2,05$, т. е. распределение оказывается достаточно близким к нормальному.

Этот числовой пример иллюстрирует процесс, который получил в теории вероятностей обобщение в виде «центральной предельной теоремы». Действительно, по мере суммирования все большего числа погрешностей эксцесс получающихся композиций все более и более приближался к эксцессу нормального распределения $\epsilon_{\text{нор}} = 3$ ($\epsilon_{\text{д}} = 1,8$, $\epsilon_{\text{н}} = 2,08$, $\epsilon_{\text{н+U}} = 2,6$). Но для завершения суммирования погрешностей теперь мы должны прибавить к полученной сумме погрешность наводки ($\sigma_{\text{нав}}$), распределенную по очень низкоэнтропийному арксинусоидальному закону ($k_{\text{нав}} = 1,11$, $\epsilon_{\text{нав}} = 1,5$, $\kappa_{\text{нав}} = 0,816$).

Кривая 4 на рис. 3-3, соответствующая суммированию нормального с арксинусоидальным распределением, на всем своем протяжении является падающей. Правда, это падение до тех пор, пока $\sigma_{\text{нор}} \geq \sigma_{\text{арк}}$, очень медленное, а при $\sigma_{\text{арк}} > \sigma_{\text{нор}}$ становится более крутым. Среднее квадратическое отклонение погрешности в конце диапазона канала

$$\begin{aligned}\sigma_{\text{н}} &= \sqrt{\sigma_{\text{н+U}}^2 + \sigma_{\text{нав}}^2} = \sqrt{0,2474 + 0,16^2} = \sqrt{0,2474 + 0,0256} = \\ &= \sqrt{0,273} = 0,5225 \approx 0,52 \, \%. \end{aligned}$$

Вес дисперсии $\sigma_{\text{нав}}^2$ с арксинусоидальным законом распределения составляет $p = \sigma_{\text{нав}}^2 / \sigma_{\text{к}}^2 = 0,0256 / 0,273 = 0,09$. При этом значении веса снижение энтропийного коэффициента композиции почти не заметно. Поэтому $k_{\text{к}} \approx k_{\text{н+U}} = 2,05$ и энтропийное значение погрешности в конце диапазона канала $\gamma_{\text{к}} = k_{\text{к}} \sigma_{\text{к}} = 2,05 \cdot 0,52 = 1,066 \approx 1,1 \, \%$. Значение эксцесса этого распределения $\epsilon_{\text{к}} = 1,5 \cdot 0,09^2 + 6 \cdot 0,09 \cdot 0,91 + 2,6 \cdot 0,91^2 = 2,66$ и $\kappa = 0,61$. Таким образом, если при добавлении второй и третьей составляющих погрешности эксцесс увеличивался на $2,6 - 2,08 = 0,52$ и $2,08 - 1,8 = 0,28$, то при добавлении четвертой — всего на $2,66 - 2,6 = 0,06$, т. е. изменение формы распределения в сторону нормального ($\kappa \epsilon = 3$) почти прекратилось. Более того, если бы вес четвертой низкоэнтропийной составляющей был бы большим, то при ее добавлении наблюдалось бы не приближение к нормальному распределению, а наоборот, удаление от него. Центральная предельная теорема верна при сложении бесконечно большого числа близких по значению погрешностей, но при суммировании 3 — 4 погрешностей закон распределения может быть еще очень далек от нормального.

Доверительная вероятность, соответствующая полученному энтропийному значению $\gamma_{\text{н}}$, согласно формуле (2-4), $P_{\text{д}} = 0,899 + 0,1818/2,66 = 0,97$. Таким образом, при оценке погрешностей результатов измерений с вероятностью $P_{\text{д}} = 0,97 \div 0,98$ следует ожидать погрешности $\gamma_{\text{в}} = 0,5\%$ и $\gamma_{\text{н}} = 1,1\%$, т. е. при произвольном значении x погрешность результатов измерений будет $\varphi(x) = 0,5 + 0,6x/X_{\text{н}}$. При необходимости указания погрешности с другими значениями $P_{\text{д}}$ можно воспользоваться найденными выше значениями ε и формулами из § 2-6.

Однако если задача состоит в назначении предела допускаемой погрешности при ежегодных поверках канала, т. е. занесении в официальные документы значения класса точности канала, то он не может быть указан как 1,0/0,5, так как в этом случае может оказаться, что при следующей ежегодной поверке погрешность канала превысит указанный предел. Чтобы этого не происходило, класс точности должен указываться с запасом на старение не менее 25%. Так как $1,1/0,8 = 1,4\%$ и $0,5/0,8 = 0,63\%$, то ближайшим разрешенным ГОСТ 8.401—80 будет класс точности 1,5/1,0. Учитывая, что старение средств измерений происходит, как правило, не быстрее $0,1\gamma_{\text{кл}}$ в год (см. § 1-6), такой запас нормируемой погрешности канала обеспечит межремонтный период около 4 лет.

Погрешность канала с цифровым регистратором включает в себя вместо погрешностей аналогового самописца погрешности цифрового прибора. Поэтому погрешность начала диапазона канала в этом случае будет складываться из двух составляющих погрешности ЦВ (с равномерным распределением с $\sigma_{1\text{н.ЦВ}} = 0,278\%$ и с экспоненциальным распределением с $\sigma_{2\text{н.ЦВ}} = 0,080\%$) и погрешности датчика с равномерным распределением и $\sigma_{\text{д}} = 0,087\%$. Составляющими же $\sigma_{\text{од}} = 0,026\%$ и $\sigma_{\text{оу}} = 0,034\%$ можно пренебречь, так как даже большая из них в 8 раз меньше, чем $\sigma_{1\text{н.ЦВ}} = 0,278\%$.

Начнем суммирование с двух равномерно распределенных составляющих. Тогда составляющая $\sigma_{\text{д+ЦВ}} = \sqrt{0,087^2 + 0,278^2} = \sqrt{7,57 \cdot 10^{-3} + 7,72 \cdot 10^{-2}} = \sqrt{8,48 \cdot 10^{-2}} = 0,291\%$. Вес дисперсии второй составляющей $p = \sigma_{\text{ЦВ}}^2 / \sigma_{\text{д+ЦВ}}^2 = 7,73 \cdot 10^{-2} / (8,48 \cdot 10^{-2}) = 0,91$, эксцесс этого распределения $\varepsilon_{\text{д+ЦВ}} = 1,8 \cdot 0,91^2 + 6 \cdot 0,91 \times 0,09 + 1,8 \cdot 0,09^2 = 2,0$, $k_{\text{д+ЦВ}} = 0,71$ и по кривой 3 рис. 3-2 $k_{\text{д+ЦВ}} = 1,93$.

Для суммирования этого распределения с экспоненциальной составляющей погрешности ЦВ с $\alpha = 0,5$ и $k_{2\text{ЦВ}} = 1,35$ соответствующей кривой на рис. 3-2 и 3-3 нет. Поэтому ее придется воссоздать по аналогии с имеющимися кривыми. На рис. 3-3 при $p = 0$ она должна начинаться в точке $k_{\text{д+ЦВ}} = 1,93$, а заканчиваться при $p = 1$ в точке $k_{2\text{ЦВ}} = 1,35$. При $p = 0,1 \div 0,2$ она пойдет вверх, как и все кривые на рис. 3-3, но достигнет максимума в области при $p \approx 0,3$ аналогично кривой 4 (рис. 3-2), имеющей максимум при $p = 0,7$. Этот максимум, по-видимому, не должен

превышать $k = 2,02$, характерного для максимума кривой 3 на рис. 3-2, а спад ее в области $p = 0,9 \div 1,0$ должен быть аналогичен спаду кривых 4 и 5 на рис. 3-3. Этих рассуждений достаточно, чтобы ориентировочно провести эту кривую.

Среднее квадратическое отклонение погрешности в начале диапазона канала

$$\sigma_{\text{н}} = \sqrt{\sigma_{\text{д+1цв}}^2 + \sigma_{\text{2цв}}^2} = \sqrt{0,291^2 + 0,08^2} = \\ = \sqrt{8,48 \cdot 10^{-2} + 6,4 \cdot 10^{-3}} = \sqrt{9,12 \cdot 10^{-2}} = 0,302 \approx 0,3 \, \%.$$

Вес экспоненциальной составляющей $p = \sigma_{\text{2цв}}^2 / \sigma_{\text{н}}^2 = 6,4 \times 10^{-3} / (9,12 \cdot 10^{-2}) = 0,07$, эксцесс этого распределения $\epsilon_{\text{н}} = 2,0 \cdot 0,93^2 + 6 \cdot 0,93 \cdot 0,07 + 25,4 \cdot 0,07^2 = 2,24$ и $\kappa = 0,67$. Отсюда значение энтропийного коэффициента по построенной нами кривой составляет $k_{\text{н}} = 2,02$ и энтропийное значение погрешности в начале диапазона канала с цифровым регистратором $\gamma_{\text{н}} = k_{\text{н}} \sigma_{\text{н}} = 2,02 \cdot 0,302 = 0,61 \approx 0,6 \, \%$.

Для определения погрешности в конце диапазона канала нужно к составляющим погрешности ЦВ в конце диапазона канала с равномерным ($\sigma_{\text{1к. ЦВ}} = 0,332 \, \%$) и экспоненциальным ($\sigma_{\text{2к. ЦВ}} = 0,096 \, \%$) распределением прибавить мультипликативные погрешности от колебаний напряжения питания ($\sigma_{U(\text{д+у})} = 0,429 \, \%$) с треугольным распределением и погрешность от наводки ($\sigma_{\text{нав}} = 0,16 \, \%$) с арксинусоидальным распределением. Погрешностью датчика можно пренебречь, так как отношение $\sigma_{U(\text{д+у})} / \sigma_{\text{д}} = 0,429 / 0,087 \approx 5$.

Просуммируем сначала самые низкоэнтропийные из этих составляющих, а именно: $\sigma_{\text{нав}} = 0,16 \, \%$ с $k_{\text{нав}} = 1,11$ и $\sigma_{\text{2к. ЦВ}} = 0,096 \, \%$ с $k_{\text{2к. ЦВ}} = 1,35$: $\sigma_{\text{нав+2к. ЦВ}} = \sqrt{0,16^2 + 0,096^2} = \sqrt{2,56 \cdot 10^{-2} + 9,22 \cdot 10^{-3}} = \sqrt{3,48 \cdot 10^{-2}} = 0,187 \, \%$.

Вес дисперсии экспоненциальной составляющей $p = \sigma_{\text{2к. ЦВ}}^2 / \sigma_{\text{нав+2к. ЦВ}}^2 = 9,22 \cdot 10^{-3} / (3,48 \cdot 10^{-2}) = 0,26$, эксцесс $\epsilon_{\text{нав+2к. ЦВ}} = 1,5 \cdot 0,74^2 + 6 \cdot 0,74 \cdot 0,26 + 25,4 \cdot 0,26^2 = 3,69 \approx 3,7$ и $\kappa_{\text{нав+2к. ЦВ}} = 0,52$.

Для определения значения энтропийного коэффициента воспользуемся кривой 4 рис. 3-2, соответствующей суммированию арксинусоидальной ($k = 1,11$) и равномерной ($k = 1,73$) составляющих. Нужная нам кривая в своей начальной части совпадает с этой кривой, а при $p \rightarrow 1$ проходит несколько ниже. При малых значениях p ($p = 0,26$) расхождение будет малым. Отсюда $k_{\text{нав+2к. ЦВ}} = 1,94$.

Теперь возьмем треугольное распределение с $\sigma_U = 0,429 \, \%$ и $k_U = 2,02$ и сложим с ним полученную составляющую:

$$\sigma_{U+\text{нав+2к. ЦВ}} = \sqrt{0,429^2 + 0,187^2} = \sqrt{1,84 \cdot 10^{-1} + 3,48 \cdot 10^{-2}} = \\ = \sqrt{0,2188} = 0,468 \, \%.$$

Вес дисперсии второй составляющей $p = 3,48 \cdot 10^{-2} / 0,2188 = 0,16$, эксцесс $\varepsilon_{U_{\text{нав}+2\text{ЦВ}}} = 2,4 \cdot 0,84^2 + 6 \cdot 0,84 \cdot 0,16 + 3,7 \cdot 0,16^2 = 2,6$ и $\kappa = 0,62$.

При столь малом весе второй составляющей для суммирования можно воспользоваться начальным участком кривой 2 (рис. 3-3), откуда $k_{U_{\text{нав}+2\text{ЦВ}}} = 2,04$, т. е. распределение достаточно близко к нормальному.

И наконец, последнее суммирование будем рассматривать как сложение равномерно распределенной составляющей $\sigma_{1\text{к. ЦВ}} = 0,332\%$ с близкой к нормальной составляющей с $k = 2,04$. Среднее квадратическое отклонение погрешности в конце диапазона

$$\sigma_{\text{к}} = \sqrt{0,332^2 + 0,468^2} = \sqrt{0,1102 + 0,2188} = \sqrt{0,3290} = 0,57 \%$$

Вес близкой к нормальному распределению составляющей $p = 0,2188 / 0,3290 = 0,66$, эксцесс $\varepsilon_{\text{к}} = 1,8 \cdot 0,34^2 + 6 \cdot 0,34 \cdot 0,66 + 2,6 \cdot 0,66^2 = 2,7$ и $\kappa_{\text{к}} = 0,61$.

На кривой 2 рис. 3-2 такому весу соответствует k , почти совпадающий с k нормального распределения, следовательно, и в нашем случае k будет почти совпадать с $k_{\text{к}} = 2,04$. Таким образом, энтропийное значение погрешности в конце диапазона канала с цифровым регистратором будет $\Psi_{\text{к}} = k_{\text{к}} \sigma_{\text{к}} = 2,04 \cdot 0,57 = 1,16 \approx 1,2\%$, а общая формула для вычисления приведенной погрешности результатов измерения при любом x может быть записана как $\Psi(x) = 0,6 + 0,6x/X_{\text{к}}$.

Но так же как и в предыдущем случае, если по результатам этого расчета необходимо назначить пределы допускаемых погрешностей канала при его ежегодных поверках, это должно быть сделано не менее чем с 25 %-ным запасом по отношению к вычисленным погрешностям. Так как $0,6/0,8 = 0,75$ и $1,2/0,8 = 1,5$, то для внесения в официальные документы в соответствии с ГОСТ 8.401—80 класс точности канала должен быть указан как 1,5/1,0, т. е. оказывается таким же, как у канала с аналоговым регистратором.

Динамические погрешности канала с аналоговым и цифровым регистратором являются дополнительными и обычно не суммируются с остальными погрешностями, а просто ограничивают частотный диапазон измеряемой величины в области высоких частот ее изменения. Поэтому рассчитаем динамические погрешности и укажем рабочий диапазон частот измеряемой величины при использовании цифрового и аналогового регистраторов. (Остальные узлы канала по сравнению с инерционностью регистрирующих устройств можно считать безынерционными.)

Относительно цифровой регистрации выше было указано, что она обеспечивает получение пяти отсчетов измеряемой величины в секунду. Динамическая погрешность восстановления сигнала по таким дискретным отсчетам согласно формуле (1-9) (см. § 1-4) $\gamma_{\text{д}} \leq \pi^2 t_0^2 f^2 / 2$. При $t_0 = 0,2$ с погрешность восстановления в зави-

симости от периода изменения и частоты сигнала согласно (1-9) составит:

$T, \text{ с.}$	20	10	5	3,6	2
$f, \text{ Гц}$	0,05	0,1	0,2	0,27	0,5
$\gamma_d, \%$	0,05	0,2	0,8	1,5	5,0

Отсюда видно, что при медленных изменениях измеряемой величины ($T \geq 20 \text{ с}$) частотная погрешность ничтожна, но при уменьшении периода до $T = 3,6 \text{ с}$ она уже равна основной погрешности канала. Таким образом, цифровая регистрация согласно (1-9) обеспечивает частотный диапазон от 0 до 0,27 Гц.

Аналоговый регистратор в виде электронного самопишущего автоматического потенциометра со следящим электромеханическим приводом обладает той особенностью, что его подвижная часть имеет постоянную максимальную скорость передвижения, обеспечивая проход всей его шкалы, как было указано выше, за 0,5 с. На пределе измерений $X_m = 200 \text{ мВ}$ это соответствует максимальной скорости изменения напряжения во времени $\dot{X}_m = 20 \text{ мВ}/0,5 \text{ с} = 400 \text{ мВ/с}$.

Если абсолютная скорость изменения поданного на самописец электрического сигнала меньше этой величины $\dot{X}_m$, то регистрация происходит без искажений и динамическая погрешность равна нулю. Если же скорость будет больше, то следящий прибор не будет успевать отслеживать изменения сигнала и погрешности будут очень велики. Это явление и ограничивает частотный диапазон аналогового регистратора.

При входном синусоидальном сигнале $x = X_m \sin \omega t$ скорость его изменения $\dot{x} = \omega X_m \cos \omega t$, а максимальное значение этой скорости $\dot{X}_m = \omega X_m = 2\pi f X_m$. Отсюда граничная частота регистрации $f_{гр} = \dot{X}_m / (2\pi X_m)$. При $\dot{X}_m = 400 \text{ мВ/с}$ и $X_m = 200 \text{ мВ}$ это дает $f_{гр} = 400 / (2\pi \cdot 200) = 0,32 \text{ Гц}$. Таким образом, рабочий диапазон частот аналогового регистратора в нашем примере оказывается несколько шире, чем рабочий диапазон частот цифровой регистрации.

Сравнение приближенных методов суммирования погрешностей. Изложенный метод суммирования с учетом всех вероятностных характеристик и свойств суммируемых погрешностей является наиболее точным, но и достаточно трудоемким. При недостатке времени можно использовать упрощенные методы суммирования, но надо отдавать себе отчет о неизбежных при этом неточностях.

Анализ вида распределений составляющих необходим для нахождения по нормированным значениям их с. к. о., и им пренебрегать не следует. Но анализ образующихся после сложения композиций при недостатке времени можно опустить, определяя результирующую погрешность при доверительной вероятности $P_d = 0,9$ как $\gamma_{0,9} = 1,6 \sigma_{\Sigma}$, или при суммировании не менее четырех составляющих в отсутствие доминирующих погрешностей

считать, что распределение их суммы близко к нормальному и $\gamma_{0,95} = 1,96\sigma_{\Sigma}$.

В нашем примере для канала с цифровым регистратором такой метод суммирования дал бы: $\sigma_{\Sigma} = 0,302\%$ и $\gamma_{\Sigma 0,95} = 0,302 \times 1,96 = 0,59\%$; $\sigma_{\kappa} = 0,57\%$ и $\gamma_{\kappa 0,95} = 0,57 \cdot 1,96 = 1,12\%$, т. е. оценки несколько меньше энтропийных, ибо последние соответствовали $P_{\Sigma} = 0,98$. Однако после введения 25%-ного запаса и округления до рекомендованных ГОСТ 8.401—80 значений класс точности канала получился бы тем же самым (1,5/1,0).

При крайней нехватке времени можно воспользоваться соотношением $\gamma_{\Sigma 0,9} = \sqrt{\sum \gamma_{i 0,9}^2}$ и допустить, что все исходные максимальные погрешности были заданы при $P_{\Sigma} = 0,9$. Это, безусловно, преувеличение, поэтому полученное значение будет соответствовать доверительной вероятности, большей, чем 0,9. Но раздельное суммирование аддитивных и мультипликативных составляющих и в этом случае является обязательным.

В нашем примере для аналогового канала максимальные значения аддитивных составляющих были равны: $\gamma_{\Sigma} = 0,15\%$; $\gamma_{\Theta \Sigma} = 0,06\%$; $\gamma_{\Theta y} = 0,06\%$; $\gamma_{\Theta \text{рег}} = -0,03\%$; $\gamma_{\text{рег}} = 0,4\%$. Учет корреляционных связей остается также обязательным, поэтому $\gamma_{\Theta (y+\text{рег})} = 0,06 - 0,03 = 0,03\%$. Но результирующая погрешность без какого-либо анализа вида законов распределения определяется как

$$\gamma_{\Sigma} = \sqrt{0,15^2 + 0,06^2 + 0,03^2 + 0,4^2} = 0,43\% \approx 0,4\%.$$

Максимальные значения мультипликативных составляющих были равны: $\gamma_{U_{\Sigma}} = 0,6\%$; $\gamma_{U_y} = 0,45\%$; $\gamma_{\text{нав}} = 0,226\%$. Вследствие тесной корреляционной связи $\gamma_{U (\Sigma+y)} = 0,6 + 0,45 = 1,05\%$. Погрешность в конце диапазона измерений канала $\gamma_{\Sigma} = \sqrt{0,45^2 + 1,05^2 + 0,226^2} = 1,16 \approx 1,2\%$, и общая характеристика точности канала получается 1,2/0,4 вместо 1,0/0,5 при точном расчете. При введении 25%-ного запаса (1,2/0,8 = 1,5 и 0,4/0,8 = 0,5) класс точности канала будет назначен в виде 1,5/0,5, что уже излишне оптимистично.

Таким образом, если предыдущий упрощенный метод занижил оценку погрешности примерно на 10%, то последний — уже на 20%. Поэтому дальнейшее упрощение в виде отказа от учета корреляционных связей и раздельного учета аддитивных и мультипликативных составляющих или переход к арифметическому суммированию составляющих вместо геометрического недопустимы при любой нехватке времени.

Так, если бы в нашем примере вместо анализа погрешностей было произведено арифметическое суммирование составляющих, то полученный для аналогового канала результат в виде $\gamma_{\Sigma} = 0,15 + 0,6 + 0,06 + 0,06 + 0,45 + 0,16 + 0,03 + 0,4 = 1,91 \approx 2\%$ завышал бы погрешность в конце диапазона измерений почти в 2 раза, а в начале — в 4 раза.

В этой связи уместно заметить, что нежелание вести обстоятельный анализ погрешностей нередко оправдывают тем, что арифметическое суммирование «обеспечивает полезный запас на незнание» и для убедительности приводят пример гранитных мостов, построенных в Ленинграде в петровскую эпоху, которые и теперь позволяют вести на парад любую современную военную технику. Но представьте себе современный авиалайнер, сконструированный с «запасами на незнание» петровской эпохи. Вряд ли такое сооружение сможет оторваться от взлетной полосы, даже без пассажиров.

3.6. РАСЧЕТ ПОГРЕШНОСТИ РЕЗУЛЬТАТОВ КОСВЕННЫХ ИЗМЕРЕНИЙ

Результат Z косвенного измерения определяется расчетом по измеренным значениям $x_1, x_2, \dots, x_k$ и заранее известной функции $Z = f(x_1, x_2, \dots, x_k)$. Так как каждое x_j , где $j = 1, \dots, k$, измерено с соответствующей погрешностью φ_j , то задача расчета погрешности φ_Z результата Z косвенного измерения сводится к суммированию всех k погрешностей измерения x_j , т. е. аналогична рассмотренной выше задаче расчета погрешности измерительного канала по известным составляющим. Но отличие состоит в том, что доля отдельных погрешностей φ_j в результирующей погрешности φ_Z может быть различной в зависимости от вида функции и соотношения между собой независимых переменных x_j .

Пусть $Z = x_1 + x_2$, но $x_1 \gg x_2$, например $x_1 \approx 100x_2$. В этом случае погрешность в 1%, допущенная при измерении x_2 , внесет в результат Z относительную погрешность всего в 0,01%, но такая же погрешность в 1%, допущенная при измерении x_1 , практически полностью войдет в погрешность результата Z . При функции $Z = f(x_1, x_2)$ вида $Z = x_1 \sqrt[5]{x_2}$ независимо от соотношения между собой x_1 и x_2 погрешность измерения x_1 полностью входит в погрешность Z , а погрешность измерения x_2 — только $1/5$ своей частью и т. д.

Так как возможные функции $Z = f(x_1, \dots, x_k)$ и соотношения x_j могут быть самыми разнообразными, то для определения чувствительности погрешности Z к изменению погрешностей x_j используют общий прием, заключающийся в определении частных производных:

$$\frac{\partial Z}{\partial x_j} = \frac{\partial [f(x_1, \dots, x_k)]}{\partial x_j}.$$

Полученные таким путем значения $\partial Z / \partial x_j$ при данном сочетании $x_1, \dots, x_k$ можно рассматривать как веса, с которыми в суммарную абсолютную погрешность ΔZ входят составляющие в виде абсолютных погрешностей измерения каждого из x_j . Отсюда составляющая абсолютной погрешности $\Delta_j(Z)$, возникающая от абсолютной

погрешности $\Delta(x_j)$, будет $\Delta_j(Z) = (\partial Z / \partial x_j) \Delta(x_j)$. Аналогично этому, если известны с. к. о. случайной абсолютной погрешности $\sigma(x_j)$ отдельных x_j , то с. к. о. соответствующих составляющих результирующей абсолютной погрешности ΔZ будут $\sigma_j(Z) = (\partial Z / \partial x_j) \sigma(x_j)$.

Далее методика вычисления суммарной погрешности ΔZ , т. е. суммирование всех составляющих $\sigma_j(Z)$, совершенно аналогична рассмотренной выше методике суммирования погрешностей измерительного канала. Так, например, для некоррелированных составляющих с. к. о. погрешности

$$\sigma(Z) = \sqrt{\sum_1^k \sigma_i^2(Z)} = \sqrt{\sum_1^k (\partial Z / \partial x_i)^2 \sigma^2(x_i)},$$

а для коррелированных составляющих $\Delta(x_a)$ и $\Delta(x_b)$ результирующая погрешность определяется как их алгебраическая сумма, но с учетом весов $\partial Z / \partial x_j$:

$$\Delta_{a+b} = \frac{\partial Z}{\partial x_a} \Delta(x_a) + \frac{\partial Z}{\partial x_b} \Delta(x_b).$$

Особенностью метода частных производных для расчета результирующей погрешности результата Z косвенных измерений является то, что он правомерен только для абсолютных погрешностей. Относительные их значения должны находиться соответствующим пересчетом.

Для простейших функций $Z = f(x_1, \dots, x_k)$ метод частных производных приводится к ряду простых соотношений, которые могут быть сформулированы в виде легко запоминающихся правил.

Так, для функции вида $Z = x_1 + x_2 + \dots + x_k$ все частные производные $\partial Z / \partial x_j$ равны единице и поэтому $\sigma(Z) = \sqrt{\sum \sigma^2(x_j)}$ и $\Delta_{a+b} = \Delta(x_a) + \Delta(x_b)$, т. е. абсолютная погрешность суммы просто равна сумме абсолютных погрешностей слагаемых. Однако относительная погрешность суммы $\gamma_Z = \sigma(Z) / (x_1 + x_2 + \dots + x_k)$, т. е. является сложной функцией, зависящей не только от погрешностей слагаемых, но и от сочетания значений самих измеренных слагаемых.

Для функции вида $Z = x_1 x_2 \dots x_k$, наоборот, относительная погрешность γ очень просто выражается через относительные погрешности аргументов γ_{x_j} . Действительно, если уравнение этой функции прологарифмировать, то получим

$$\ln Z = \ln x_1 + \ln x_2 + \dots + \ln x_k$$

и после дифференцирования

$$\frac{dZ}{Z} = \frac{dx_1}{x_1} + \frac{dx_2}{x_2} + \dots + \frac{dx_k}{x_k}.$$

Заменяя дифференциалы малыми конечными приращениями (чем погрешности по существу и являются), получим

$$\frac{\Delta Z}{Z} = \frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2} + \dots + \frac{\Delta x_k}{x_k}$$

или

$$\gamma_Z = \gamma_{x1} + \gamma_{x2} + \dots + \gamma_{xk}, \quad (3-7)$$

т. е. относительная погрешность произведения просто равна сумме относительных погрешностей сомножителей. Зато абсолютная погрешность в этом случае $\Delta Z = \gamma_Z Z = (\gamma_{x1} + \gamma_{x2} + \dots + \gamma_{xk}) x_1 x_2 \dots x_k$, т. е. зависит не только от значений γ_{xj} , но и от сочетания значений $x_1, \dots, x_k$.

Сформулированные простейшие правила определения погрешности результата Z косвенного измерения распространяются не только на сумму и произведение x_j , но и на их разность и отношение. Но если при этом погрешности x_j рассматриваются как случайные, то получаемые при дифференцировании знаки производных не должны учитываться при суммировании составляющих, если только они не являются взаимно коррелированными.

Так, например, при функции вида $Z = x_1 + x_2 - x_3 - x_4$: $\partial Z / \partial x_1 = +1$, $\partial Z / \partial x_2 = +1$, но $\partial Z / \partial x_3 = -1$ и $\partial Z / \partial x_4 = -1$, отсюда $\Delta Z = \Delta(x_1) + \Delta(x_2) - \Delta(x_3) - \Delta(x_4)$, но для некоррелированных погрешностей $\sigma^2(Z) = 1^2 \sigma^2(x_1) + 1^2 \sigma^2(x_2) + 1^2 \sigma^2(x_3) + 1^2 \sigma^2(x_4)$. Таким образом, дисперсия равности двух случайных величин равна не разности, а сумме их дисперсий. Поэтому несмотря на отрицательные знаки частных производных, погрешности составляющих должны не вычитаться, а складываться. То же самое должно производиться и для соотношения (3-7) при расчете относительной погрешности не произведения, а частного. Хотя, например, для функции $Z = x_1 x_2 / (x_3 x_4)$ после логарифмирования $\ln Z = \ln x_1 + \ln x_2 - \ln x_3 - \ln x_4$ относительные погрешности должны всегда складываться, если только они жестко не коррелированы.

Использованный выше прием логарифмирования и последующего дифференцирования функции $Z = f(x_1, \dots, x_k)$ с целью вывода формулы для $\gamma_Z = F(x_1, \dots, x_k, \gamma_{x1}, \dots, \gamma_{xk})$ удобно использовать во всех случаях, когда это оказывается возможным.

Например, для приведенной в начале этого параграфа функции $Z = x_1^5 \sqrt{x_2}$ $\ln Z = \ln x_1 + 0,2 \ln x_2$ и с. к. о. относительной погрешности $\sigma(Z) = \sqrt{\sigma^2(x_1) + (0,2)^2 \sigma^2(x_2)}$, где $\sigma(x_1)$ и $\sigma(x_2)$ — относительные значения с. к. о. погрешностей x_1 и x_2 . Если бы эта функция имела вид $Z = x_1 x_2^5$, то $\ln Z = \ln x_1 + 5 \ln x_2$ и с. к. о. относительной погрешности было бы $\sigma(Z) = \sqrt{\sigma^2(x_1) + 5^2 \sigma^2(x_2)}$ и т. д.

Итог проведенного рассмотрения можно подвести следующим образом. Расчет погрешности результатов косвенных измерений

складывается из двух этапов. Первый этап — это вывод формулы для абсолютной или относительной погрешности результата косвенного измерения исходя из вида функции $Z = f(x_1, \dots, x_n)$. Второй этап — расчет погрешности Z в соответствии с полученной формулой путем суммирования ее составляющих по правилам суммирования случайных погрешностей с учетом корреляционных связей и их законов распределения, как это было рассмотрено применительно к расчету результирующей погрешности измерительного канала.

Проблема нормирования погрешности косвенных измерений осложняется тем, что одно и то же значение результата косвенного измерения Z может быть получено при самых различных сочетаниях между собой непосредственно измеряемых величин x_j .

Так, например, если измеряемой величиной является мощность, вычисляемая по измеренным значениям тока I и напряжения U как $P = IU$, то $P = 100$ Вт получится и при $I = 1$ А и $U = 100$ В, и при $I = 10$ А и $U = 10$ В. Погрешности же измерений I и U будут при этом существенно различными, следовательно, и погрешности результатов косвенного измерения мощности также будут разными. Еще большее различие в погрешностях может быть, если в функцию Z входит операция деления. Значение $R = 10$ Ом может быть получено в результате деления $U = 10$ В на $I = 1$ А и $U = 10$ мВ на $I = 1$ мА, и погрешности этих результатов могут различаться в тысячу раз.

Этот вопрос в последние годы приобретает очень серьезное значение в связи с проблемой нормирования погрешностей косвенных измерений. До тех пор пока косвенные измерения и обработка их результатов выполнялись человеком-экспериментатором, этой проблемы не возникало. Заметив, что после установления нового режима в испытуемой системе один из приборов показывает, например, всего 2—3 первых деления, а указатель другого прибора вышел за шкалу, экспериментатор изменит их чувствительность, с тем чтобы погрешности измерений оставались в допускаемых пределах.

Положение оказывается совершенно иным, когда в современных условиях измерение всех x_j производится k каналами ИВК: его процессор выдает на регистрацию уже рассчитанные значения Z и получатель этой информации не имеет представления, из какого сочетания x_j получен этот результат Z . Рассчитанные ИВК результаты косвенных измерений используются в дальнейшем для принятия весьма ответственных решений (управления ходом сложного технологического процесса, управления курсом корабля или ракеты и т. п.). Поэтому оценка погрешности таких результатов чрезвычайно важна. В этой связи выдвигаются требования об аттестации ИВК путем нормирования погрешности получаемых от них результатов аналогично тому, как нормируются погрешности всех остальных средств измерений. В ГОСТ 16263—70 записано: «Средство измерений — техническое средство, исполь-

взуемое при измерениях и имеющее нормированные метрологические свойства». На этом основании сторонники нормирования нередко утверждают, что если погрешность косвенных измерений, выполняемых ИВК, не нормирована, то ИВК не может считаться средством измерений.

Тем не менее, погрешности результатов косвенных измерений, выполняемых ИВК, не могут быть **з а р а н е е** нормированы.

Пределы допускаемых погрешностей всех остальных средств измерений при выполнении ими прямых измерений можно было **з а р а н е е** нормировать только благодаря тому, что их погрешности являются **о д н о з н а ч н о й** функцией получаемого отсчета. Эта однозначная функция и описывается одночленной, двучленной или трехчленной формулой нормирования погрешности СИ. При косвенных измерениях такой однозначной функциональной зависимости между получаемым результатом Z косвенного измерения и его погрешностью γ_Z не существует и поэтому подобное нормирование невозможно.

Поэтому в ИИС или ИВК могут и должны нормироваться лишь погрешности их измерительных каналов, выполняющих прямые измерения x_j . Такое нормирование погрешностей и осуществляется обычно с использованием двучленных формул погрешности. Это и исчерпывает метрологическую аттестацию этих систем. Вопрос же нужно ставить не об ужесточении аттестации ИИС и ИВК, а об **а т т е с т а ц и и** каждого получаемого от них **р е з у л ь т а т а** косвенного измерения.

Решение проблемы аттестации результатов косвенных измерений состоит в индивидуальном расчете с помощью процессора ИВК погрешности результата каждого косвенного измерения и вывода этой оценки погрешности на регистрацию одновременно с самим результатом косвенного измерения.

Действительно, для расчета результата Z косвенного измерения по формуле $Z = f(x_1, \dots, x_k)$ процессор ИВК должен располагать значениями всех x_j . Эти же известные в данный момент значения x_j нужны и для расчета погрешности данного Z по формуле $\gamma_Z = F(x_1, \dots, x_k, \gamma_{x1}, \dots, \gamma_{xk})$. После завершения расчета Z и γ_Z данные о значениях x_j могут быть стерты в памяти, так как при следующем измерении их значения чаще всего будут уже иными. Необходимые же для расчета γ_Z значения $\gamma_{x1}, \dots, \gamma_{xk}$ должны рассчитываться по известным в этот момент времени x_j и метрологическим характеристикам соответствующих измерительных каналов (коэффициентам двучленных формул), которые должны постоянно храниться в памяти ИВК. Иными словами, для того чтобы каждый результат косвенного измерения был аттестован указанием при регистрации его погрешности, в процессор ИВК должны быть введены **д в е п р о г р а м м ы** для одновременного вычисления как самого значения Z , так и его погрешности γ_Z .

В этой связи ГОСТ 26.203—81 указывает, что «ИВК должны обеспечивать оценку точности измерений и представление резуль-

татов измерений в формах, установленных ГОСТ 8.011—72», а затем уточняет, что программы обработки сигналов в ИВК должны сопровождаться подпрограммой оценки точности результатов в формах, установленных ГОСТ 8.011—72.

Реализация такой индивидуальной аттестации каждого результата косвенных измерений на практике осложняется тем, что программа вычисления γ_Z оказывается существенно сложнее, чем программа расчета самого Z . Это обусловлено тем, что даже при использовании самых простейших методов суммирования случайных составляющих погрешности (например, соотношения $\Delta_{0,9} = 1,6 \sigma$, см. § 3-4) программа вычисления $\gamma_Z = F(x_1, \dots, x_k, \gamma_{x1}, \dots, \gamma_{xk})$ должна оперировать с $2k$ переменными, тогда как программа вычисления $Z = f(x_1, \dots, x_k)$ — только с k переменными. Однако выигрыш оказывается весьма существенным и состоит в обеспечении получателя информации указанием интервала неопределенности каждого выдаваемого ему результата, в возможности осуществления рационального автоматического округления выдаваемых результатов в соответствии с их погрешностью по правилам, изложенным в § 1-4, а следовательно, в экономии времени, затрачиваемого на печать результатов, а также автоматической отбраковке получаемых результатов по их погрешности и запрете вывода результатов, погрешность которых оказывается выше некоторого (например, функционально заданного) предела (см. выводы § 3-8).

3-7. ПРИМЕР РАСЧЕТА РЕЗУЛЬТИРУЮЩЕЙ ПОГРЕШНОСТИ КОСВЕННОГО ИЗМЕРЕНИЯ

Постановка задачи конкретного примера. Для проведения тепловых расчетов при проектировании поршневых машин (двигателей внутреннего сгорания, поршневых компрессоров и др.) одним из исходных расчетных коэффициентов является коэффициент теплоотдачи ξ поверхности внутренней стенки цилиндра. При работе, например, поршневого компрессора воздух, всасываемый в цилиндр компрессора из атмосферы, имеет температуру 20°C . При сжатии его в цилиндре компрессора температура воздуха повышается, например, до 120°C . Поэтому при непрерывной работе компрессора изменение температуры Θ_1 воздуха внутри цилиндра происходит во времени по кривой $\Theta_1(t)$ (рис. 3-6) в пределах от 20 до 120°C . В установившемся режиме средняя температура тела цилиндра Θ_s сохраняется постоянной на уровне, например, 70°C .

При частоте вращения коленчатого вала компрессора $n = 600$ об/мин $= 10$ об/с период кривой $\Theta_1(t)$ составляет $T = 100$ мс. Поэтому в течение примерно 50 мс температура газа в полости цилиндра превышает температуру тела цилиндра и тепловой поток Φ идет от более теплого газа к более холодному телу цилиндра. За это время температура внутренней поверхности цилиндра несколько повышается (например, на $2,5$ К). В течение

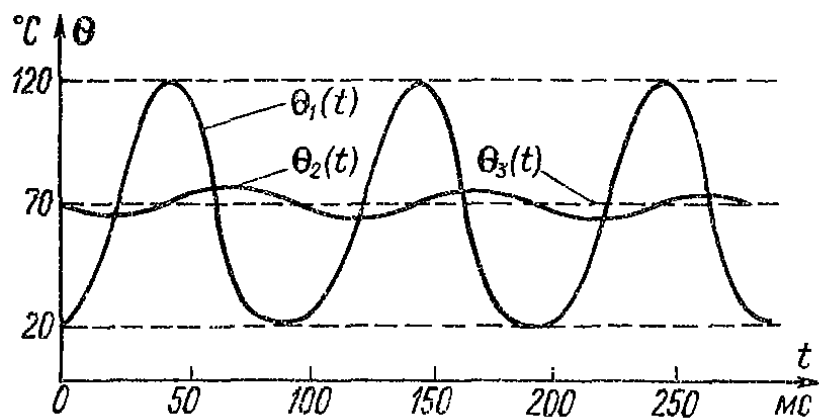


Рис. 3-6

же следующих 50 мс температура тела цилиндра превышает температуру поступившего в цилиндр холодного воздуха и тепловой поток Φ идет, наоборот, от тела цилиндра к воздуху. В результате этого поверхность цилиндра охлаждается на те же 2,5 К. В итоге этого процесса изменение температуры Θ_2 внутренней поверхности цилиндра происходит по кривой $\Theta_2(t)$ (рис. 3-6).

Интенсивность этого непрерывного теплообмена и определяется коэффициентом теплоотдачи поверхности цилиндра ξ . Поэтому измерение фактического значения этого коэффициента представляет существенный интерес для разработчиков подобных машин. Расчет погрешности экспериментального определения этого коэффициента на работающем компрессоре и рассмотрим в качестве практического примера расчета погрешности косвенных измерений.

Коэффициент теплоотдачи ξ представляет собой отношение теплового потока через единицу поверхности к перепаду температуры между этой поверхностью и воздухом. Так как тепловой поток измеряется в ваттах, то удельный тепловой поток Φ через единицу поверхности будет измеряться в ваттах на квадратный метр. Отсюда коэффициент теплоотдачи (в ваттах на квадратный метр-кельвин) $\xi = \Phi / (\Delta\Theta)$.

Пусть косвенное измерение осуществляется путем использования нескольких каналов ИИС с последующей обработкой данных прямых измерений на ЭВМ или с использованием нескольких каналов ИВК и его процессора. Для ввода данных в ИИС или ИВК на компрессоре должны быть установлены соответствующие датчики.

Схема размещения датчиков (рис. 3-7). Здесь 1 — датчик термометра из тонкой платиновой проволоки, воспринимающий температуру Θ_1 воздуха внутри полости цилиндра, 2 и 3 — пленочные датчики термометров, напыленные на тонком изоля-

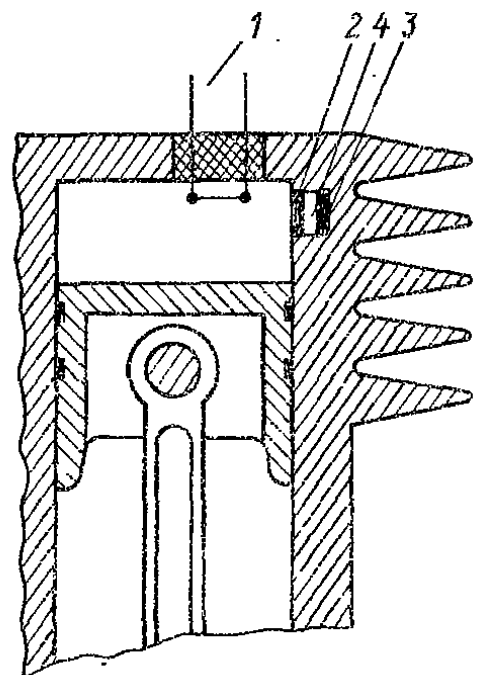


Рис. 3-7

ционном покрытии на вставке 4 из металла, близкого по теплопроводности к материалу тела цилиндра, и воспринимающие температуры Θ_2 и Θ_3 .

Данные о температуре от датчиков 1 и 3 введены в соответствующие измерительные каналы ИВК, тщательно проградуированные до установки датчиков в головку цилиндра компрессора путем помещения датчиков в термостат с известной температурой. Измерение же малой разности температур $\Theta_2 - \Theta_3$, колеблющейся в пределах всего $\pm 2,5$ К, путем вычитания данных, полученных по двум каналам ИВК, привело бы к большим погрешностям измерения величины $\Theta_2 - \Theta_3$. Поэтому для непосредственного измерения величины $\Theta_2 - \Theta_3$ датчики 2 и 3 включены в два смежных плеча моста, который тщательно уравнивается по постоянной составляющей выходного сигнала. В результате этого выходное напряжение моста пропорционально лишь отклонениям температуры Θ_2 от ее среднего значения Θ_3 , и, поданное в канал ИВК, оно несет данные лишь об этих отклонениях. Таким образом, в общей сложности для косвенного измерения ξ используются три канала ИВК.

Тепловой поток, проходящий через вставку 4, $\Phi = (\Theta_2 - \Theta_3)/R_T = G_T (\Theta_2 - \Theta_3)$, где R_T и G_T — тепловое сопротивление и тепловая проводимость вставки. Эта постоянная вставки тщательно определяется в специальной измерительной установке до размещения вставки 4 в стенке цилиндра компрессора. Погрешность установления значения R_T или G_T можно оценить значением примерно $\pm 0,4\%$. Как будет видно из дальнейшего расчета, она значительно меньше, чем другие составляющие погрешности, и поэтому этой составляющей погрешности можно пренебречь.

Вывод формулы погрешности косвенного измерения. Коэффициент теплоотдачи

$$\xi = \Phi/(\Theta_1 - \Theta_2) = G_T (\Theta_2 - \Theta_3)/(\Theta_1 - \Theta_2). \quad (3-3)$$

Уравнение вида $\xi = f(\Theta_1, \Theta_2, \Theta_3)$ и является уравнением данного косвенного измерения. Полный дифференциал от него

$$\begin{aligned} d\xi &= G_T \frac{d(\Theta_2 - \Theta_3)(\Theta_1 - \Theta_2) - d(\Theta_1 - \Theta_2)(\Theta_2 - \Theta_3)}{(\Theta_1 - \Theta_2)^2} = \\ &= G_T \frac{d(\Theta_2 - \Theta_3)}{\Theta_1 - \Theta_2} - G_T d(\Theta_1 - \Theta_2) \frac{\Theta_2 - \Theta_3}{(\Theta_1 - \Theta_2)^2}. \end{aligned}$$

Отсюда абсолютная погрешность определения ξ по измеряемым Θ_1 , Θ_2 и $(\Theta_2 - \Theta_3)$

$$\Delta \xi = G_T \frac{\Delta(\Theta_2 - \Theta_3)}{\Theta_1 - \Theta_2} - G_T \frac{\Delta(\Theta_1 - \Theta_2)}{\Theta_1 - \Theta_2} \frac{\Theta_2 - \Theta_3}{\Theta_1 - \Theta_2}.$$

Относительная погрешность определения ξ без учета погрешности определения постоянной G

$$\begin{aligned}\varphi(\xi) &= \frac{\Delta\xi}{\xi} = \frac{\Delta\xi}{\frac{G_T(\Theta_2 - \Theta_3)}{\Theta_1 - \Theta_2}} = \frac{\Delta(\Theta_2 - \Theta_3)}{\Theta_2 - \Theta_3} - \frac{\Delta(\Theta_1 - \Theta_2)}{\Theta_1 - \Theta_2} = \\ &= \varphi_{\Theta_2 - \Theta_3} - \varphi_{\Theta_1 - \Theta_2},\end{aligned}$$

т. е. равна сумме относительных погрешностей числителя и знаменателя уравнения (3-8), так как при суммировании квадратов этих случайных погрешностей знак минус будет опущен. Поэтому приведенного вывода формулы погрешности можно было и не делать, а сразу воспользоваться правилом о погрешности дроби.

Однако, определяя относительные погрешности числителя ($\gamma_{\text{чис}}$) и знаменателя ($\gamma_{\text{зн}}$) в выражении (3-8), необходимо учесть, что величина Θ_2 не измеряется, а вычисляется через значения Θ_1 и Θ_3 — $\Theta_2 = \Delta\Theta$ как $\Theta_2 = \Theta_3 + \Delta\Theta$. Это приводит к следующему изменению уравнения (3-8):

$$\xi = G_T \frac{\Theta_2 - \Theta_3}{\Theta_1 - \Theta_3} = G_T \frac{\Delta\Theta}{\Theta_1 - \Theta_3 - \Delta\Theta}, \quad (3-9)$$

т. е. абсолютная погрешность числителя $\Delta_{\text{чис}}$ просто равна абсолютной погрешности измерения величины $\Delta\Theta = \Theta_2 - \Theta_3$, а абсолютная погрешность знаменателя $\Delta_{\text{зн}}$ складывается из погрешностей трех измеряемых величин Θ_1 , Θ_3 и $\Delta\Theta$:

$$\Delta_{\text{зн}}^2 = \Delta^2(\Theta_1) + \Delta^2(\Theta_3) + \Delta^2(\Theta_2 - \Theta_3). \quad (3-10)$$

Относительная погрешность определения числителя в выражении (3-9) $\gamma_{\text{чис}} = \Delta_{\text{чис}}/(\Theta_2 - \Theta_3)$, а относительная погрешность знаменателя $\gamma_{\text{зн}} = \Delta_{\text{зн}}/(\Theta_1 - \Theta_3)$. Поэтому относительная погрешность результата косвенного измерения может быть выражена как

$$\varphi(\xi) = \sqrt{\gamma_{\text{чис}}^2 + \gamma_{\text{зн}}^2} = \sqrt{\left(\frac{\Delta_{\text{чис}}}{\Theta_2 - \Theta_3}\right)^2 + \left(\frac{\Delta_{\text{зн}}}{\Theta_1 - \Theta_3}\right)^2}. \quad (3-11)$$

Численный расчет погрешности результатов косвенного измерения требует знания метрологических характеристик каналов ИИС или ИВК, которыми производится измерение величин Θ_1 , Θ_3 и $\Delta\Theta = \Theta_2 - \Theta_3$. Пусть каналы измерения температур Θ_1 и Θ_3 аттестованы классом точности 1,0/0,5, т. е. обеспечивают значения приведенной погрешности в начале диапазона $\gamma_n(\Theta_1) = \gamma_n(\Theta_3) = 0,5\%$ и в конце диапазона $\gamma_k(\Theta_1) = \gamma_k(\Theta_3) = 1\%$, а предел измерений этих каналов $X_K = 200^\circ\text{C}$. Канал же измерения разности температур $\Theta_2 - \Theta_3 = \Delta\Theta$ аттестован постоянной приведенной погрешностью $\gamma_0(\Delta\Theta) = 2\%$ при $X_K = 2,5\text{ K}$. Тогда, используя одночленную формулу погрешности (см. § 1-2), получаем, что измерение любого значения величины $\Delta\Theta$ в пределах от 0 до 2,5 K характеризуется постоянным значением абсолютной погрешности $\Delta(\Delta\Theta) = \gamma_0(\Delta\Theta) X_K = 0,02 \cdot 2,5 = 0,05\text{ K}$.

Таблица 3-1

$\Theta_1, ^\circ\text{C}$	$\Delta(\Theta_1), \text{K}$	$\Theta_1 - \Theta_3, \text{K}$	$\Theta_2 - \Theta_3, \text{K}$	$\Delta_{\text{чис}}, \text{K}$	$\Delta_{\text{зн}}, \text{K}$	$\gamma(\xi), \%$
120	1,60	50	2,5	0,05	2,90	4,6
118	1,59	48	2,4	0,05	2,08	4,8
110	1,55	40	2,0	0,05	2,05	5,7
100	1,50	30	1,5	0,05	2,02	7,5
86	1,43	16	0,8	0,05	1,97	14
78	1,39	8	0,4	0,05	1,94	27
72	1,36	2	0,1	0,05	1,91	108
70	1,35	0	0	0,05	1,90	∞

Относительное и абсолютное значения погрешности каналов измерения температур Θ_1 и Θ_3 изменяются в зависимости от текущего значения этих температур и описываются двучленными формулами (см. § 1-2). Относительная погрешность измерения описывается в соответствии с ГОСТ 8.401—80 формулой (1-5) вида $\gamma(x) = \gamma_{\text{н}} + \gamma_{\text{н}}(X_{\text{н}}/x - 1)$, что при $\gamma_{\text{н}} = 1\%$, $\gamma_{\text{н}} = 0,5\%$ и $X_{\text{н}} = 200^\circ\text{C}$ дает (в процентах) $\gamma(\Theta_1) = \gamma(\Theta_3) = 1 + 0,5 \times \times (200/\Theta - 1) = 1 - 0,5 + 0,5 \cdot 200/\Theta = 0,5 + 0,5 \cdot 200/\Theta$. Отсюда абсолютная погрешность измерения текущего значения температур Θ_1 и Θ_3 (в кельвинах) $\Delta(\Theta_1) = \Delta(\Theta_3) = \gamma(x) \cdot x/100 = = 0,5\Theta/100 + 0,5 \cdot 200/100 = \Theta/200 + 1$.

Дальнейший расчет погрешности косвенного измерения коэффициента теплоотдачи ξ , согласно формулам (3-10) и (3-11) для различных значений температур Θ_1 , $\Theta_1 - \Theta_3$ и $\Theta_2 - \Theta_3$, удобнее представить в виде значений табл. 3-1. Погрешности, соответствующие положительному и отрицательному полупериодам кривой $\Theta_1(t)$ (рис. 3-6), почти одинаковы (чуть больше погрешности для положительного полупериода в соответствии с двучленной формулой), поэтому в табл. 3-1 даны их значения лишь для положительного полупериода, приведены значения $\Theta_1 - \Theta_3$ и $\Theta_2 - \Theta_3$, нужные для подстановки в формулу (3-11), значения абсолютной погрешности канала измерения разности температур $\Delta(\Theta_2 - \Theta_3) = = \Delta(\Delta\Theta) = \Delta_{\text{чис}}$, а также значения абсолютной погрешности $\Delta_{\text{зн}}$ знаменателя (3-9), вычисленные по формуле (3-10).

Выбранные для расчета значения температур были взяты на основании следующих рассуждений. Выше было указано, что период изменения измеряемых температур при частоте вращения коленчатого вала компрессора $n = 600$ об/мин составляет $T = = 100$ мс. Для того чтобы погрешность восстановления была при этом примерно 1%, нужно (как это будет показано ниже), чтобы на каждый период измеряемой величины приходилось не менее 20 отсчетов, т. е. они должны производиться ИИС с периодом дискретизации $t_0 = 5$ мс. Тогда на каждую четверть периода будет приходиться по 5 отсчетов. Так как отсчеты равномерны во вре-

мени, а кривая изменения температуры в грубом приближении может быть принята близкой к синусоидальной, то при амплитуде синусоиды, равной 50 единиц (табл. 3-1), эти отсчеты будут равны 50, 48, 40, 30, 16 и 0 К.

Однако если частота дискретизации отсчетов ИИС не синхронизирована с частотой вращения компрессора, то момент отсчета в области нуля синусоиды может не совпадать точно с моментом прохождения ее через нуль. Поэтому была оценена погрешность измерений в точках 0, 2 и 8 К.

Максимальное значение динамической погрешности γ_m при периоде дискретизации $t_0 = 5$ мс, обеспечивающем получение $n = 20$ отсчетов на период измеряемой величины, может быть подсчитано по формуле (1-9): $\gamma_m \leq \pi^2/(2n^2) = \pi^2/800 = 0,012 = 1,2\%$.

Фазовая погрешность исключается следующим путем. Из кривых рис. 3-6 видно, что кривая $\Theta_2(t)$ является интегралом от кривой $\Theta_1(t)$, т. е. отстает от нее по фазе на 90° . Поэтому если в формулу (3-9) для вычисления значений коэффициента теплоотдачи ξ подставлять значения температур, полученные одновременно, то ξ получится также в виде периодической функции времени, что противоречит физическому смыслу этой постоянной.

Для исключения этого при вычислении ξ в формулу (3-9) подставляются отсчеты Θ_1 , задержанные во времени на четверть периода, т. е. на 5 отсчетов. Эти пять отсчетов значений Θ_1 хранятся в памяти ЭВМ до получения соответствующего значения температуры Θ_2 . В результате рассчитанное значение ξ получается действительно постоянным, так как, например, максимальное значение $\Theta_1 - \Theta_3$ делится на максимальное значение $\Theta_2 - \Theta_3$, т. е. $50/2,5 = 20$. Затем следующее значение $\Theta_1 - \Theta_3$ делится на следующее значение $\Theta_2 - \Theta_3$, т. е. $48/2,4 = 20$, и т. д.

Результат расчета погрешностей косвенного измерения по формуле (3-11) приведен в последнем столбце табл. 3-1. Он показывает, что в то время как погрешность двух из использованных измерительных каналов не превосходит 1%, а у третьего канала составляет 2%, погрешность получаемых результатов даже без учета динамической погрешности на большей части периода составляет 4—7%, т. е. именно с такой погрешностью может быть определено путем косвенного измерения в этих условиях значение фактического коэффициента теплоотдачи на работающем компрессоре.

Но в небольшой части (примерно $1/5$ периода) в области прохождения кривыми $\Theta_1 - \Theta_3 = f(t)$ и $\Theta_2 - \Theta_3 = f(t)$ через нуль погрешность результатов, получаемых путем косвенных измерений, достигает 14, 27, 108% и даже может быть равна бесконечности. Это та самая особенность косвенных измерений, на которую указывалось в предыдущем параграфе, когда отмечалось, что погрешность результата косвенных измерений существенно зависи

от сочетания между собой значений непосредственно измеряемых величин. Это подчеркивает важность расчета погрешностей косвенных измерений.

3-8. ОСОБЕННОСТИ ВЕРОЯТНОСТНОГО ОПИСАНИЯ ПОГРЕШНОСТИ РЕЗУЛЬТАТА КОСВЕННОГО ИЗМЕРЕНИЯ И НЕКОТОРЫЕ ВЫВОДЫ

Важность обсуждения результатов рассмотренного конкретного примера расчета погрешностей косвенного измерения обусловлена тем, что изложенные в § 3-6 доводы и заключение о принципиальной невозможности нормирования получаемых от ИВК результатов косвенных измерений не являются общепризнанными.

Для «убедительного» опровержения этих доводов приводят пример того, что результат измерения сопротивления четырехплечим мостом всегда рассчитывается как $R_1 = R_2 R_3 / R_4$, тем не менее это не создает каких-либо препятствий для нормирования погрешностей таких мостов.

Таким образом, рассматриваемый вопрос не является таким простым, как это может показаться с первого взгляда. Его сложность заключается в следующем. В рассматриваемом выше примере, казалось бы, были обеспечены все условия для достаточно точного измерения коэффициента теплоотдачи ξ — использовался ИВК с погрешностью измерительных каналов 1—2%. Между тем мы в лучшем случае, т. е. при наиболее благоприятном соотношении непосредственно измеряемых величин, получили (см. табл. 3-1) результат косвенного измерения с погрешностью $\varphi(\xi) = 4,6\%$, а при других соотношениях непосредственно измеряемых величин эта погрешность могла оказаться равной и 108%, и бесконечности. При этом пользователь, получая от ИВК уже готовые результаты расчета косвенно измеряемой величины, не знает, из какого сочетания исходных данных получены эти результаты.

Если в условиях предыдущего примера составить для ИВК программу вывода на графопостроитель всех получаемых во время работы компрессора результатов измерения коэффициента теплоотдачи ξ в виде печати точек на графике с координатами ξ_i и t_i , причем время t на графике ограничить, например, тремя циклами работы цилиндра компрессора, а все последующие циклы вновь и вновь накладывать на этот же график, то получим картину, изображенную на рис. 3-8. Вся внутренняя площадь изображенной здесь фигуры будет при достаточно продолжительной работе компрессора заполнена точками, отражающими значения ξ_i , полученные в результате многократно повторенных косвенных измерений величины ξ .

Из рисунка видно, что при лучших сочетаниях измеряемых температур получаемые результаты косвенного измерения находятся в полосе шириной $\pm 5\%$, т. е. выдаваемые ИВК значения ξ_i

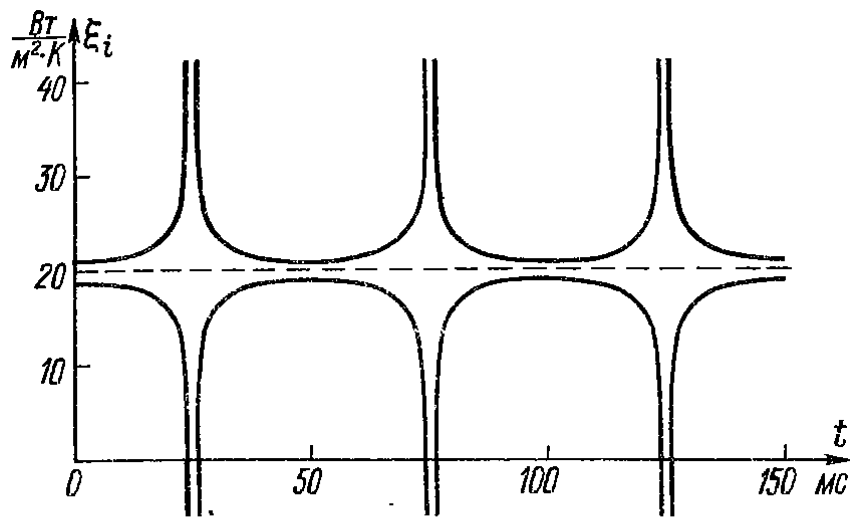


Рис. 3-8

колеблются в пределах от 19 до 21 Вт/(м²·К). Но при других сочетаниях температур (см. табл. 3-1) эта полоса расширяется. При ширине этой полосы $\pm 100\%$ выводимые на печать значения ξ_i будут находиться в интервале ξ от 0 до 40 Вт/(м²·К), а при еще менее благоприятных сочетаниях измеряемых температур, когда $\Theta_1 - \Theta_3$ и $\Theta_2 - \Theta_3$ близки к нулю, возможно получение даже отрицательных значений, тогда как коэффициент теплоотдачи ξ — величина положительная.

Встретившись на практике с подобной ситуацией, экспериментатор попытается усреднить разброс этих данных обычным путем вычисления их среднего арифметического и будет поражен тем, что их среднее арифметическое имеет разброс того же порядка, что и сами эти данные. Это вызвано тем, что рассматриваемая ситуация не тривиальна даже с позиций теории вероятностей.

Почему же ничего подобного не наблюдается при измерении неизвестного сопротивления четырехплечим мостом и вычислении результата по формуле $R_1 = R_2 R_3 / R_4$? Да потому, что соотношение сопротивлений плеч моста R_3 / R_4 является отношением не случайных, а заранее известных постоянных величин. В данном же случае результат измерения в виде неслучайной величины ξ находится как отношение случайных величин $\Theta_2 - \Theta_3$ и $\Theta_1 - \Theta_2$.

Как отмечалось в § 2-4, отношение двух центрированных нормально распределенных случайных величин распределено по закону распределения Коши. Особенностью распределения Коши является то, что для него центр распределения не может быть найден как среднее арифметическое полученных отсчетов, так как распределение Коши не имеет моментов — определяющие их выражения равны бесконечности. Поэтому и рассеяние оценки координаты центра в виде среднего арифметического наблюдаемых отсчетов также стремится к бесконечности (см. § 4-3).

В данном случае и числитель $\Theta_2 - \Theta_3$ и знаменатель $\Theta_1 - \Theta_2$ дроби (3-8), вычислением которой определяется искомая величина ξ , изменяются во времени в первом приближении синусоидально и одновременно содержат в себе случайные погрешности, которые

мы считаем центрированными и вычисляем по соотношениям (3-10). До тех пор пока детерминированная часть этих величин (изменяющаяся во времени приблизительно синусоидально) остается существенно большей, чем размах центрированной случайной погрешности, можно считать, что эти погрешности суммируются и их результирующее распределение является композицией законов распределения суммируемых составляющих. Однако в области приближения детерминированной составляющей к нулю такая трактовка уже неправомерна.

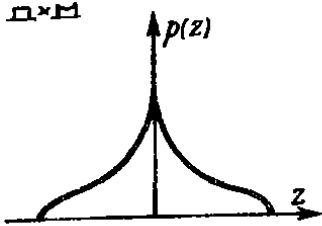
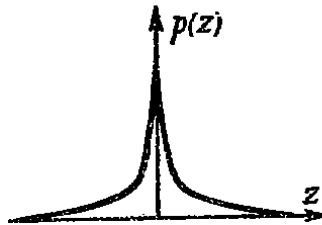
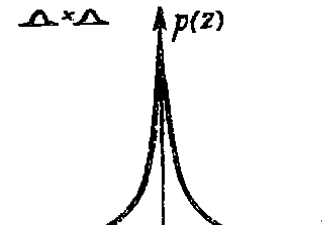
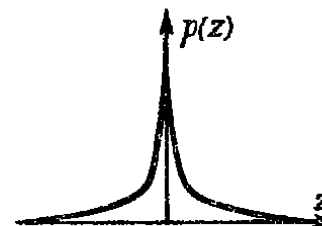
При отсутствии детерминированной составляющей и числитель и знаменатель дроби (3-8) суть центрированные случайные величины и распределение получаемых результатов есть распределение отношения этих центрированных случайных величин, которое весьма существенно отличается по своим свойствам от широко известных распределений и их композиций. Так как в курсах по теории вероятностей этот вопрос, как правило, не рассматривается, то представляется целесообразным коснуться его хотя бы самым кратким образом.

Примеры законов распределения произведений и отношений некоторых центрированных независимых случайных величин. В п. 1 табл. 3-2 приведены график и аналитическое выражение плотности распределения случайной величины $Z = XY$, являющейся произведением двух независимых центрированных случайных величин. При этом случайная величина X распределена по закону арксинуса на интервале $[-a, a]$, а величина Y распределена равномерно на интервале $[-b, b]$.

В п. 2 приведено распределение плотности произведения двух случайных величин, равномерно распределенных на интервалах $[-a, a]$ и $[-b, b]$, в п. 3 — двух нормально распределенных, а в п. 4 — распределенных по законам Лапласа с с. к. о. σ_1 и σ_2 соответственно.

Общим свойством всех этих распределений является то, что контрэксцесс распределения произведения приблизительно равен произведению контрэксцессов сомножителей, т. е. $\kappa \approx \kappa_1 \kappa_2$. Поэтому если, например, для арксинусоидального распределения (п. 1 табл. 3-2) $\kappa_1 = \sqrt{1/1,5} = \sqrt{2/3}$ и для равномерного распределения $\kappa_2 = \sqrt{1/1,8} = \sqrt{5/9}$, то контрэксцесс их произведения $\kappa \approx \sqrt{2/3} \cdot \sqrt{5/9} = 0,61$, т. е. существенно меньше, чем контрэксцесс исходных распределений. Подобное же понижение значений контрэксцесса имеет место и при других законах распределения: с 0,745 до 0,55 — для п. 2, с 0,577 до 0,33 — для п. 3 и с 0,408 до 0,16 — для п. 4.

Это может иметь существенное практическое значение при расчете погрешности результатов косвенных измерений. Как будет показано в § 4-3 и 4-4, чем меньше значение контрэксцесса распределения, тем труднее (с большими погрешностями) определяются по экспериментальным данным его параметры (оценка координаты

Но- мер пп.	Вид распределения	Параметры распределения
1		$p(z) = \begin{cases} \frac{1}{2\pi ab} \ln \frac{1 + \sqrt{1 - z/(ab) ^2}}{1 - \sqrt{1 - z/(ab) ^2}} \\ \text{при } z \leq ab; \\ 0 \text{ при } z > ab; \end{cases}$ $\kappa = \kappa_1 \kappa_2 = \sqrt{\frac{2}{3}} \sqrt{\frac{5}{9}} \approx 0,61$
2		$p(x) = \begin{cases} \frac{1}{2ab} \ln(ab/ z) \text{ при } z \leq ab; \\ 0 \text{ при } z > ab; \end{cases}$ $\kappa = \frac{1}{\sqrt{1,8}} \frac{1}{\sqrt{1,8}} \approx 0,55;$ $\Delta_3 = 3\sigma_1\sigma_2 e^{C-1}, \text{ где } C = 5,57721 \dots \text{ — постоянная Эйлера};$ $k_3 = 3e^{C-1} \approx 1,966$
3		$p(z) = \frac{1}{\sqrt{4\pi c z }} e^{C-1}, \text{ где } c = \sigma_1\sigma_2;$ $\kappa = \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \approx 0,33$
4		$p(z) = \frac{1}{c} \sqrt[4]{\frac{\pi^2 c}{2 z }} e^{\sqrt{8 z /c}}, \text{ где } c = \sigma_1\sigma_2;$ $\kappa = \frac{1}{\sqrt{6}} \frac{1}{\sqrt{6}} \approx 0,16$

центра распределения, оценка ширины распределения, т. е. погрешности, и т. д. — см. рис. 4-2 и 4-4).

При прямых измерениях это не имеет большого значения, так как практически контрэксцесс распределения погрешностей не бывает меньше 0,4. Но при косвенных измерениях искомый результат Z может быть произведением большого числа сомножителей.

Как видно из табл. 3-2, при перемножении лишь двух равномерно распределенных центрированных случайных величин ($\kappa_1 = \kappa_2 = 0,745$) контрэксцесс распределения их произведения ($\kappa =$

$= 0,55$) становится уже меньше, чем у нормального. При перемножении двух нормально распределенных или четырех равномерно распределенных величин контрэксцесс произведения становится равным $0,33$, т. е. таким малым, какого никогда не бывает у погрешности прямых измерений. При перемножении четырех нормальных $\kappa = 0,11$, а для произведения трех распределенных по закону Лапласа центрированных величин — $\kappa = 0,07$ и т. д.

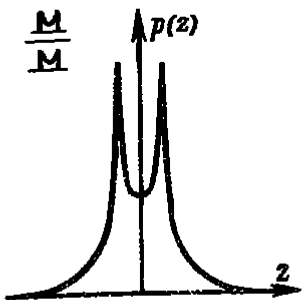
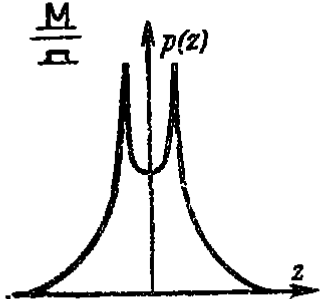
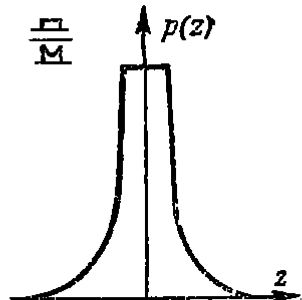
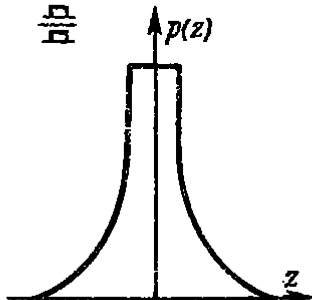
Уменьшение разброса результатов путем вычисления среднего арифметического при таких значениях контрэксцесса распределений крайне неэффективно (см. § 4-3). Уменьшение разброса может быть достигнуто только при использовании медианных методов усреднений. Как будет показано на рис. 4-2, а, эффективность оценки координаты центра медианой для нормального распределения примерно в 2 раза больше, чем для равномерного, для распределения Лапласа — в 4 раза больше, чем для равномерного, а для распределений с κ , равным $0,2$ и $0,1$, соответственно больше в 12 и 24 раза. Это полезно иметь в виду при организации косвенных измерений и оценке их погрешностей.

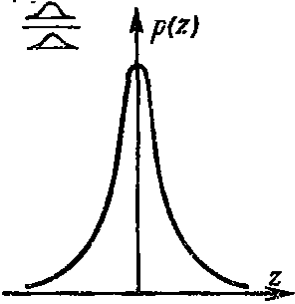
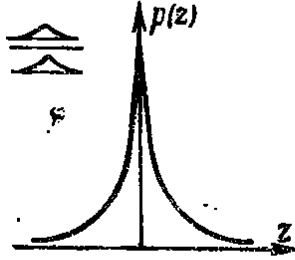
Еще большими особенностями по сравнению с законами распределения погрешностей прямых измерений обладают законы распределения частного от деления друг на друга независимых центрированных случайных величин; некоторые из них приведены в табл. 3-3. В п. 1 табл. 3-3 приведен закон распределения случайной величины $Z = X/Y$, являющейся частным от деления друг на друга независимых центрированных случайных величин X и Y , распределенных по закону арксинуса соответственно, на интервалах $[-a, a]$ и $[-b, b]$; в п. 2 — закон распределения частного, когда величина X распределена на интервале $[-a, a]$ арксинусоидально, а величина Y на интервале $[-b, b]$ — равномерно; в п. 3 — когда, наоборот, X на интервале $[-a, a]$ распределена равномерно, а Y на интервале $[-b, b]$ — по закону арксинуса; в п. 4 — когда числитель и знаменатель дроби Z распределены равномерно, соответственно в интервалах $[-a, a]$ и $[-b, b]$. В п. 5 приведено распределение частного $Z = X/Y$ двух нормально распределенных независимых центрированных величин со с. к. о., равным соответственно σ_x и σ_y . Это единственно общеизвестный закон распределения Коши с параметром ширины распределения $c = \sigma_x/\sigma_y$.

И наконец, в п. 6 табл. 3-3 приведен закон распределения частного от деления друг на друга двух независимых случайных величин, распределенных по закону Лапласа со с. к. о., равными соответственно σ_x и σ_y .

Общей особенностью этих распределений является отсутствие моментов выше нулевого порядка.

Взаимосвязь существования моментов распределения и возможностей оценки его параметров. Выше уже подчеркивалось, что кривой плотности распределения может быть лишь такая кривая, площадь под которой может быть приравнена единице, т. е.

Но- мер пп.	Вид распределения	Параметры распределения
1		$p(z) = \frac{1}{\pi^2 z } \ln \left \frac{b+az}{b-az} \right ;$ $\Delta_{\vartheta} = \frac{\pi^2 b}{2a} \approx 4,93 \sigma_x / \sigma_y$
2		$p(z) = \begin{cases} \frac{a}{b\pi} \frac{1}{1 + \sqrt{1 - (az/b)^2}} \\ \text{при } z \leq \frac{b}{a}; \\ \frac{b}{a\pi z^2} \text{ при } z > \frac{b}{a}; \end{cases}$ $\Delta_{\vartheta} = \frac{b\pi}{4a} e^{\frac{4}{\pi} G + 1} \approx 6,85 \frac{b}{a} =$ $= 5,59 \sigma_x / \sigma_y, \text{ где } G = 0,91596 \dots \text{ — постоянная Каталана}$
3		$p(z) = \begin{cases} \frac{a}{\pi b} \text{ при } z \leq \frac{b}{a}; \\ \frac{b}{\pi a z^2} \frac{1}{1 + \sqrt{1 - b/(az) ^2}} \text{ при } \\ z > \frac{b}{a}; \end{cases}$ $\Delta_{\vartheta} = \frac{b}{a} \pi^{\frac{4}{\pi} G - 1} \approx 3,71 \frac{b}{a} = 4,54 \sigma_x / \sigma_y$
4		$p(z) = \begin{cases} \frac{a}{4b} \text{ при } z \leq \frac{b}{a}; \\ \frac{b}{4a z^2} \text{ при } z > \frac{b}{a}; \end{cases}$ $\Delta_{\vartheta} = 2e \frac{b}{a} \approx 5,437 \frac{b}{a}$

Номер пп.	Вид распределения	Параметры распределения
5		$p(z) = \frac{1}{c\pi [1 + (z/c)^2]}, \text{ где } c = \sigma_x/\sigma_y;$ $\Delta_\theta = 2\pi c \approx 6,28\sigma_x/\sigma_y$
6		$p(z) = \frac{1}{2c (1 + z /c)^2}, \text{ где } c = \sigma_x/\sigma_y;$ $\Delta_\theta = e^2 c \approx 7,39\sigma_x/\sigma_y$

выполняется условие нормирования в виде равенства единице момента нулевого порядка $\int_{-\infty}^{+\infty} x^0 p(x) dx = 1$. Подобное равенство может существовать только тогда, когда кривая $p(x)$ спадает при $x \rightarrow \infty$ круче, чем $1/x^{1+\alpha}$, где α — сколь угодно малая положительная величина.

Для существования у распределения математического ожидания, определяемого первым моментом $M[X] = \int_{-\infty}^{+\infty} xp(x) dx$, кривая распределения $p(x)$ при $x \rightarrow \infty$ должна спадать круче, чем $1/x^{2+\alpha}$. Если же этот спад более пологий, чем $1/x^2$, то интеграл, определяющий первый момент, не имеет конечного значения и, следовательно, распределение с такой кривой плотности не имеет математического ожидания.

Для существования дисперсии, а следовательно, и для существования с. к. о., кривая распределения при $x \rightarrow \infty$ должна спадать круче, чем $1/x^{3+\alpha}$, где α — сколь угодно малая величина. Если это условие не выполняется, то с. к. о. и дисперсии у такого распределения не существуют (они равны бесконечности). Это одновременно означает, что с. к. о. оценки математического ожидания, равное $\sigma_x = \sigma_{xi}/\sqrt{n}$, при любом конечном объеме выборки n наблюдаемых отсчетов также будет равно бесконечности, т. е. сам факт существования математического ожидания распределения еще не означает, что его значение может быть определено при конечном n . Если у распределения $D \rightarrow \infty$ и $\sigma \rightarrow \infty$, так как не существует второго момента, то и значение математического ожи-

дания может быть определено лишь с бесконечно большим разбросом, т. е. практически неопределимо.

Эти чисто математические ограничения не имели бы большого значения, если бы подобные распределения не встречались на практике. Однако примеры распределений, полученные в табл. 3-3, убеждают в обратном, так как операция деления для получения результатов косвенных измерений используется весьма часто, это влечет за собой ряд совсем не очевидных, с первого взгляда, следствий.

Так, например, любой человек, рассматривая кривые плотности распределения, приведенные в табл. 3-3, без каких-либо затруднений укажет центры этих распределений. Однако кривые всех этих распределений при $x \rightarrow \infty$ спадают положе, чем $1/x^2$, и не имеют математических ожиданий. Поэтому попытка определить координаты их центров как среднее арифметическое на практике приведет к очень большому разбросу. Тем более все эти распределения не имеют дисперсии (у всех $\sigma = \infty$), поэтому попытка сравнить их ширину путем вычислений с. к. о. — бессмысленна. Как же это понимать?

Для достаточно круто спадающих кривых распределений (нормального, экспоненциального, трапецеидального и т. п.) теория вероятностей использует очень удобный метод для оценки их параметров — метод моментов. Следствием этого метода и явились очень удобные понятия «математического ожидания», «среднего квадратического отклонения» и т. д. Но метод моментов и основанные на нем понятия правомерны лишь до тех пор, пока существуют моменты, и не правомерны в тех случаях, когда у распределения этих моментов не существует.

Возможно, что подготовленный в области теории вероятностей читатель еще в начале гл. 2 недоумевал по поводу того, зачем авторам этой книги потребовалось использовать понятие «центр распределения» вместо общепризнанного и привычного понятия «математическое ожидание»? Но после приведенных примеров ясно, что понятие «центр распределения» является более общим, а следовательно, и более фундаментальным, чем понятие «математическое ожидание», так как центр имеют все распределения, а математическое ожидание — лишь некоторые из них.

Естественно, что координаты центра и ширина распределения могут быть определены как математическое ожидание и с. к. о. лишь у тех распределений, которые имеют конечные первый и второй моменты. Определение параметров более полого спадающих распределений должно опираться на моменты **б о л е е н и з к и х п о р я д к о в**, которые существуют у любого распределения.

Координата центра может быть определена как 50%-ная квантиль, т. е. медиана, или как центр сгибов (т. е. полусумма 25%- и 75%-ной квантили), так как существование квантилей требует наличия лишь условия нормирования, т. е. существования момента **н у л е в о г о п о р я д к а**.

Оценка ширины таких распределений может быть произведена на основе шенноновской энтропии и вытекающего из нее понятия энтропийного интервала неопределенности. Действительно, для энтропии непрерывной случайной величины K . Шеннон ввел определение в виде своеобразного л о г а р и ф м и ч е с к о г о м о м е н т а

$$H = - \int_{-\infty}^{+\infty} [\ln p(x)] p(x) dx$$

(см. § 2-3). Этот момент существует для всех распределений, кривая плотности $p(x)$ которых спадает при $x \rightarrow \infty$ менее полого, чем $1/(x \ln^3 x)$, т. е. существует для всех распределений, приведенных в табл. 3-3, поэтому в ней указаны энтропийные значения Δ_0 погрешности для всех приведенных распределений.

Достоинство энтропийной оценки ширины распределения по сравнению со с. к. о. состоит не только в том, что она существует для более пологих распределений, для которых с. к. о. не существует, но и в том, что энтропия H и энтропийное значение Δ_0 не требуют для своего определения знания координаты центра распределения, в то время как с. к. о. есть корень из второго центрального момента и не может быть найдено без предварительного определения координаты центра, а поэтому сохраняет в себе ошибку определения этой координаты.

На основании изложенного можно сделать о б щ и е в ы в о д ы относительно вероятностного описания погрешностей результата косвенного измерения.

1. Если при расчете погрешности прямых измерений необходимо учитывать деформации законов распределения при с у м м и р о в а н и и случайных величин в виде образования их композиций, то при расчете погрешностей косвенных измерений необходимо учитывать более сложные деформации законов распределения при п е р е м н о ж е н и и и д е л е н и и случайных величин.

2. Если при суммировании случайных величин основной тенденцией при деформации законов распределения является стремление закона распределения суммы к нормальному распределению, то при перемножении или делении случайных величин эта тенденция совершенно иная. При п е р е м н о ж е н и и центрированных независимых случайных величин контрэксцесс распределения произведения примерно равен произведению контрэксцессов сомножителей. Поэтому контрэксцесс распределения произведения всегда только м е н ь ш е контрэксцесса сомножителей и тенденция его изменения заключается в стремлении $\kappa \rightarrow 0$. Распределение ч а с т н о г о от деления двух независимых центрированных случайных величин всегда имеет $\kappa = 0$, т. е. не имеет моментов, определяющих математическое ожидание и дисперсию.

3. Следствием предыдущего является то, что оценки центра

и ширины распределения отношений могут быть определены только медианными, квантильными и энтропийными методами.

4. В общем случае случайные величины, над которыми производятся операции умножения и деления при вычислении результата косвенного измерения, не являются центрированными, а могут содержать значительную систематическую составляющую. Так, например, если $X = \bar{X} + \dot{X}$ и $Y = \bar{Y} + \dot{Y}$, где $\bar{X}$ и $\bar{Y}$ — средние значения, а $\dot{X}$ и $\dot{Y}$ — центрированные случайные составляющие, то $Z = XY = \bar{X}\bar{Y} + \bar{X}\dot{Y} + \dot{X}\bar{Y} + \dot{X}\dot{Y} = \bar{X}\bar{Y} [1 + \dot{X}/\bar{X} + \dot{Y}/\bar{Y} + \dot{X}\dot{Y}/(\bar{X}\bar{Y})]$. Если $\dot{X} \ll \bar{X}$ и $\dot{Y} \ll \bar{Y}$, то член $\dot{X}\dot{Y}/(\bar{X}\bar{Y})$ есть величина второго порядка малости и ею можно пренебречь. Тогда мы полагаем, что $Z = \bar{X}\bar{Y} (1 + \gamma_x + \gamma_y)$, т. е. $Z = \bar{X}\bar{Y}$ с относительной погрешностью $\gamma_z = \gamma_x + \gamma_y$, равной сумме относительных погрешностей x и y . Если же $\bar{X} \rightarrow 0$ и $\bar{Y} \rightarrow 0$, то $Z = \dot{X}\dot{Y}$. В промежуточном же случае: $Z = (\bar{X} + \dot{X})(\bar{Y} + \dot{Y})$, нужно решение с учетом доли указанного влияния. Аналогично при $Z = (\bar{X} + \dot{X})/(\bar{Y} + \dot{Y})$ такое влияние нельзя не учитывать.

Из рассмотрения изложенных особенностей оценки погрешностей косвенных измерений можно сделать следующие практические выводы.

1. При прямых измерениях погрешность результата измерения при использовании данного СИ изменяется (см. рис. 1-4) от некоторого минимального значения и до бесконечности. Но это изменение является функцией лишь одного аргумента — самой измеряемой величины, и поэтому такая зависимость может быть выражена в явном виде и использована для нормирования этой погрешности в виде одночленных, двучленных и трехчленных формул.

При косвенном измерении погрешность результата является функцией многих аргументов $(x_1, \dots, x_k, \gamma_{x1}, \dots, \gamma_{xk})$, часть из которых $(x_1, \dots, x_k)$ при автоматическом проведении косвенных измерений остается неизвестной получателю информации. Что же касается результата Z косвенного измерения, который выдается ИВК получателю информации, то он (см. формулу 3-10) не входит в число аргументов, определяющих его погрешность. Поэтому нормирование погрешности $\gamma(Z)$ в функции Z , как это делается для всех СИ, для результатов косвенных измерений принципиально невозможно.

2. В этих условиях приобретает особую важность анализ и расчет погрешностей результатов косвенных измерений. Однако строгий расчет погрешности $\gamma(Z)$ затруднен тем, что в этом случае погрешности результатов измерения могут быть распределены по таким законам распределения вероятностей, которые не рассматриваются в учебниках и справочниках по теории вероятностей. Рассмотренный в § 3-5 пример расчета погрешности измеритель-

ного канала свелся к рассмотрению ряда законов распределения, показанных на рис. 3-5, параметры которых общеизвестны. Однако в примере расчета погрешности косвенного измерения не случайно пришлось воспользоваться самым приближенным методом суммирования составляющих погрешностей, так как описание закона распределения отношения двух центрированных случайных величин слишком сложно.

3. Однако даже самый приближенный расчет погрешности результатов косвенного измерения (подобный приведенному выше) для нескольких возможных сочетаний аргументов ($x_1, \dots, x_k$), если число аргументов не так велико, позволяет выявить особо опасные их сочетания.

4. Если представляется возможным указать зоны наиболее опасных сочетаний аргументов в функции от какого-либо аргумента (не обязательно из числа $x_1, \dots, x_k$), в нашем примере — от угла поворота коленчатого вала, который не входил в число аргументов (температур Θ_1, Θ_2 и Θ_3 , используемых для вычисления ξ), то результаты измерений, проводимых в этих зонах, просто должны исключаться из рассмотрения как недостоверные. Так, если в рассмотренном примере исключить из рассмотрения все результаты, полученные при $|\Theta_1 - \Theta_3| < 30 \text{ K}$, то погрешность оставшихся результатов (см. табл. 3-1) будет распределена по симметрично усеченному закону в пределах $\pm 7,5\%$ и никаких трудностей в усреднении этих оставшихся результатов уже не будет. Этот прием совершенно аналогичен за исключением пользоваться результатами прямых измерений, если они получены в начальной части шкалы аналогового прибора.

5. Если же аргументов ($x_1, \dots, x_k$) в формуле результата косвенного измерения достаточно много или они входят в таких сочетаниях, что анализ положения опасных зон оказывается затруднительным, то единственным практически возможным путем обеспечения достоверности косвенных измерений остается описанный в § 3-6 метод одновременного использования в процессоре ИВК двух параллельно действующих программ: одной — для вычисления $Z = f(x_1, \dots, x_k)$ и второй — для вычисления $\gamma(Z) = (x_1, \dots, x_k, \gamma_{x1}, \dots, \gamma_{xk})$. Получатель информации, располагая данными о погрешности каждого из представленных ему результатов, сам отберет из них те, которые сочтет достаточно достоверными. При желании этот процесс отбора достоверных результатов может быть автоматизирован, и на выход из ИВК будут выдаваться лишь те из результатов, погрешность которых оказалась меньше некоторого, например, функционально заданного значения.

6. При крайней необходимости использование тех результатов косвенного измерения, погрешность которых оказалась за пределами заданного значения (от заданного и до бесконечности), усреднение должно производиться медианными методами (см.

§ 4-3, 6-2, а также § 5-1), а оценка разброса — энтропийным значением, так как определение последнего по гистограмме в отличие от с. к. о. не требует знания координаты центра распределения.

В заключение следует отметить, что § 3-6 — 3-8 написаны по материалам работы [8], а табл. 3-2 и 3-3 разработаны И. А. Назаровым.

ГЛАВА ЧЕТВЕРТАЯ

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МНОГОКРАТНЫХ ОТСЧЕТОВ

4-1. ЗАДАЧИ, РЕШАЕМЫЕ ПУТЕМ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МНОГОКРАТНЫХ ОТСЧЕТОВ

Определение погрешности результата измерения по паспортным данным СИ, по существу, есть использование результатов поверки, проведенной полгода, год тому назад, к тому же в условиях, отличных от условий данного эксперимента. Поэтому, естественно, предпочтительнее путем статистической обработки проводимого эксперимента определить случайную погрешность в данный момент и в данных условиях, чем опираться на старые сведения.

При этом фактические погрешности, возникающие при эксперименте, могут быть как меньше, так и больше рассчитанных по паспортным данным СИ. Меньше потому, что для СИ нормируются пределы допускаемой погрешности, которые содержат запас на старение. Поэтому погрешность нового или только что прошедшего ремонт и регулировку прибора, как уже указывалось, может быть от 0,8 до 0,4 от нормируемого предела. Погрешности измерительного канала или косвенного измерения могут быть меньше рассчитанных и потому, что отдельные составляющие при расчете практически всегда тем или иным образом суммируются, а в действительности они могут вычитаться и взаимно компенсироваться.

Фактические погрешности могут оказаться и больше расчетных прежде всего потому, что погрешность СИ — это лишь обязательно присутствующая часть погрешности экспериментальных данных, к которой добавляются методические погрешности постановки эксперимента, погрешности, вызванные невоспроизводимостью (диффузностью) самого объекта исследования и особенно точностью задания варьируемых и стабилизируемых величин, и т. п. Поэтому при появлении возможности определения фактической погрешности экспериментальных данных она всякий раз безусловно должна быть использована.

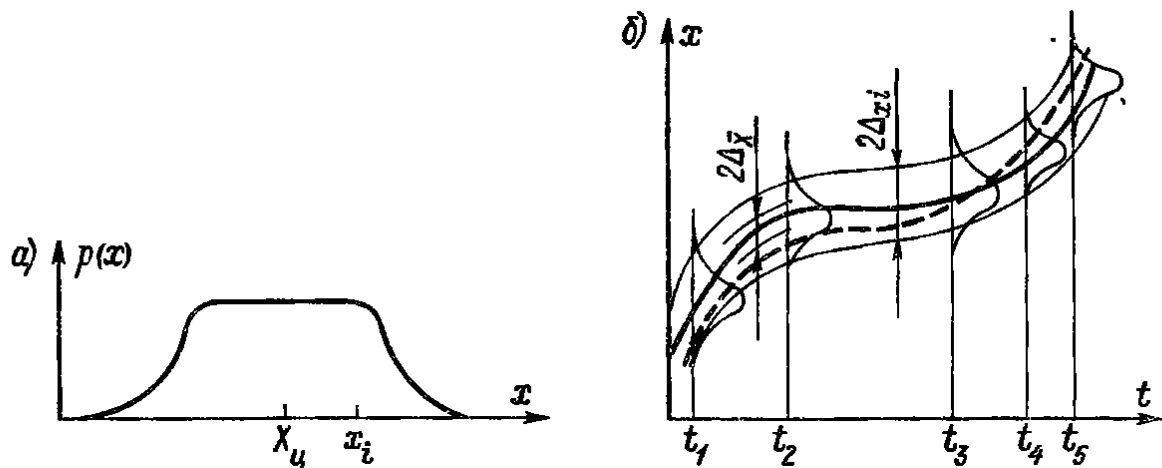


Рис. 4-1

Кроме определения случайной погрешности исходных данных, статистическая обработка позволяет их усреднить и найти как более точный усредненный результат, так и его погрешность. Если эксперимент состоит в многократном измерении одного и того же значения измеряемой величины, то усредненный результат — это центр распределения всех полученных отсчетов (рис. 4-1, а). Если эксперимент состоит в определении функции одной или многих переменных, то усредненный результат — это искомая функция, а погрешность результата — это область неопределенности этой функции, в то время как случайная погрешность исходных данных проявляется в разбросе отдельных отсчетов вокруг найденной функции. При определении функции $x = f(t)$ переменной t этот разброс описывается набором распределений (рис. 4-1, б) при $t = t_1$, $t = t_2$ и т. д. Если они близки друг к другу по ширине и форме, т. е. погрешность аддитивна и имеет близкие законы распределения при разных значениях t , то эти распределения можно объединить и представить аналогично рис. 4-1, а.

Таким образом, путем статистической обработки многократных отсчетов решаются три задачи:

оценивание случайной погрешности, т. е. области неопределенности исходных экспериментальных данных;

нахождение более точного усредненного результата исследования;

оценивание погрешности этого усредненного результата, т. е. более узкой его области неопределенности.

Методы статистической обработки многократных отсчетов (при допущении о неизменности их закона распределения во всех точках модели исследуемого явления) оказываются сходными как в простейшем однофакторном, так и в сложных многофакторных экспериментах и сводятся к определению числовых оценок параметров соответствующих законов распределения (координаты центра, оценок ширины и формы).

Поэтому для практического выполнения статистической обработки многократных отсчетов необходимо знание методов опреде-

ления по экспериментальным данным числовых характеристик распределений, которые и излагаются ниже. При этом приходится отметить, что во многих пособиях и руководствах по обработке экспериментальных данных некоторые положения математической статистики даются без должного пояснения исходных допущений, что нередко является источником ошибочных ограничений для неправильного их использования. Поэтому некоторым, казавшимся бы, общепринятым положениям ниже будет дана несколько иная трактовка.

4.2. РАССЕЯНИЕ ОЦЕНКИ КООРДИНАТЫ ЦЕНТРА РАСПРЕДЕЛЕНИЯ. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Основной смысл усреднения многократных отсчетов заключается в том, что найденная усредненная оценка координаты их центра имеет меньшую случайную погрешность, чем отдельные отсчеты, по которым она находится. Так, например, если бы искомая кривая $x = f(t)$ на рис. 4-1, б определялась путем лишь одного отсчета при каждом из t , то она могла бы случайным образом занять любое положение внутри полосы погрешностей (например, штриховая линия). При проведении n отсчетов при каждом t , сами отсчеты по-прежнему будут располагаться случайным образом внутри той же полосы, однако линия их центров будет более устойчива.

Тем не менее, если все исследование (снимая по n отсчетов при каждом из t_j) провести еще и еще раз, то получаемые линии центров не совпадут между собой, а хотя и меньше, но будут случайным образом отличаться друг от друга. Таким образом, усреднение не устраняет полностью случайный характер усредненного результата, а лишь уменьшает в какое-то число раз ширину полосы его неопределенности.

Наиболее широко распространенным методом определения координаты $X_{\text{ц}}$ центра распределения (рис. 4-1, а) является ее оценка в виде *среднего арифметического* всех отсчетов, т. е. в виде

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (4-1)$$

Преимущественное использование этой оценки среди теоретиков, а вслед за этим и практиков, объясняется отнюдь не тем, что это «самая лучшая» или, как говорят математики, *эффективная* оценка центра, а тем, что это единственная оценка, которую можно выразить аналитически, т. е. формулой, и подставлять в таком виде в другие соотношения, анализировать их и т. д. Среднее квадратическое отклонение рассеяния этой оценки зависит от с. к. о. σ_{xi} разброса усредняемых отсчетов и их числа n (при независимости отсчетов друг от друга) как

$$\sigma_{\bar{x}} = \sigma_{xi} / \sqrt{n}. \quad (4-2)$$

Таким образом, с. к. о. случайной погрешности усредненного результата убывает по сравнению со с. к. о. самих усредняемых отсчетов в $\sqrt{n}$ раз. На этом и основан принцип повышения точности в результате усреднения многократных отсчетов (см. также § 8-9).

Положительной стороной метода статистического усреднения является то, что при усреднении одновременно уменьшаются все случайные погрешности вне зависимости от их происхождения (диффузность, т. е. невоспроизводимость самого объекта исследования, случайные погрешности всех используемых средств измерений, случайные погрешности округления при вычислении отдельных наблюдений и т. д.). При этом соотношение (4-2) справедливо при любом законе распределения исходных данных (с конечным вторым моментом) и любом их числе, но при условии их **н е з а в и с и м о с т и**.

Однако, используя этот метод, следует помнить, что если во всех результатах наблюдений присутствует одна и та же систематическая погрешность, то согласно формуле (4-1) она никак не усредняется, т. е. систематические погрешности при этом не устраняются (см. § 4-7, 8-3).

Другой особенностью метода усреднения является то, что он возможен в том случае, если исходные данные имеют разброс, т. е. в них первоначально сохраняются последние неустойчивые десятичные знаки.

Таким образом, оперируя оценками σ_{xi} и $\sigma_{\bar{x}}$, необходимо четко различать их между собой и помнить, что они характеризуют лишь случайную составляющую погрешности. Оценка с. к. о. σ_{xi} характеризует ширину полосы неопределенности самих исходных данных (на рис. 4-1, б — это ширина полосы $2\Delta_{xi}$). Оценка же $\sigma_{\bar{x}}$ характеризует в $\sqrt{n}$ раз более узкую полосу неопределенности найденной усредненной зависимости, обозначенную на рис. 4-1, б через $2\Delta_{\bar{x}}$.

Закон распределения $\bar{X}$ при $n \geq 30$ близок к нормальному при любом законе распределения исходных данных с $\kappa \neq 0$. Поэтому переход от оценки $\sigma_{\bar{x}}$ к квантильной оценке погрешности с заданной доверительной вероятностью P производится в этом случае как

$$\Delta_P = t_n \sigma_{\bar{x}} = t_n \frac{\sigma_{xi}}{\sqrt{n}}, \quad (4-3)$$

где t_n — нормированная квантиль нормального распределения для заданной вероятности P .

Значения нормированных квантилей нормального распределения для ряда уровней значимости $q = 1 - P$, где P — двусторонняя вероятность, приведены ниже:

$q = 1 - P$	0,2	0,1	0,05	0,02	0,01	0,005	0,002
P	0,8	0,9	0,95	0,98	0,99	0,995	0,998
t_n	1,28	1,64	1,96	2,33	2,58	2,81	3,09

Распределение Стьюдента используется для квантильной оценки Δ_P погрешности $\bar{X}$ при малом числе наблюдений ($n < 30$). Для расчета оценки погрешности Δ_P с заданной доверительной вероятностью согласно выражению (4-3) необходимо знание оценки σ_{xi} разброса самих исходных данных. При $n \geq 30$ эта оценка может быть найдена путем обработки этих данных. Однако если мы располагаем малым числом наблюдений ($2 < n < 30$), то формально найденная оценка σ_{xi} (обозначим ее в этом случае через S_{xi}) согласно соотношению

$$S_{xi} = \sqrt{\sum_1^n (x_i - \bar{X})^2 / (n - 1)} \quad (4-4)$$

будет сама иметь большой разброс и найденная квантильная оценка разброса среднего может иметь большую ошибку. Поэтому пользоваться для оценки рассеяния $\bar{x}$ квантилями t_{α} нормального распределения можно только в тех случаях, когда оценка σ_{xi} была определена по выборке $n \geq 30$ наблюдений. Затем ее можно использовать в выражении (4-3) и при малом n , если измерения продолжают тем же методом, в тех же условиях и на той же аппаратуре, т. е. есть основания полагать, что σ_{xi} остается неизменной.

Если выполнить серию из 30 измерений невозможно или нецелесообразно, то использование соотношения (4-3) неправомерно. Так, если $\bar{X}$ находится всего по двум наблюдениям x_1 и x_2 , как $\bar{X} = (x_1 + x_2)/2$, то оценка с. к. о. x_i в виде

$$\begin{aligned} S_{xi} &= \sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 / (n - 1)} = \\ &= \sqrt{2 \left(\frac{x_1 - x_2}{2} \right)^2 / (2 - 1)} = \frac{|x_1 - x_2|}{\sqrt{2}} \end{aligned}$$

из-за своей случайности несет мало информации о действительном значении σ_{xi} .

С подобным положением мы будем встречаться и при определении оценок других параметров распределений (экссесса, контрэксесса, энтропийного коэффициента), когда оценка, определенная по малой выборке, может существенно отличаться от значения этого параметра для генеральной совокупности. Для оценки таких расхождений придется довольствоваться весьма приближенными методами. Для рассматриваемого же случая квантильной оценки отклонений среднего строгое решение было найдено еще в 1908 г. англичанином В. С. Госсетом (1876—1937). Он показал, что распределение величины $t = (\bar{X} - \bar{X}_{\infty}) / (S_{xi} / \sqrt{n}) =$

$$= (\bar{X} - \bar{X}_{\infty}) / \sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 / [n(n - 1)]} \text{ при нормально распре-}$$

деленной генеральной совокупности x_i подчиняется соотношению (2-13). При публикации этого соотношения Госсет решил подписать работу псевдонимом Student, поэтому распределение (2-13) и получило в литературе название «распределение Стьюдента».

Квантили распределения Стьюдента для двустороннего симметричного доверительного интервала приведены ниже:

n	2	3	4	5	7	10	15	20	30	∞
$t_{0,9}$	6,31	2,92	2,35	2,13	1,94	1,83	1,76	1,73	1,70	1,64
$t_{0,95}$	12,7	4,30	3,18	2,78	2,45	2,26	2,14	2,09	2,04	1,96

Пользование ими состоит в том, что при $n < 30$ доверительное значение погрешности оценки $\bar{X}$ находится как

$$\Delta_{0,9\bar{x}} = t_{0,9} \frac{S_{xi}}{\sqrt{n}} \text{ или } \Delta_{0,95\bar{x}} = t_{0,95} \frac{S_{xi}}{\sqrt{n}}. \quad (4-5)$$

При уменьшении объема выборки n , по которой находится оценка S_{xi} для σ_{xi} , значения t Стьюдента резко возрастают: при $n = 3$ $\Delta_{0,95} = 4,3S_{xi}$, а при $n = 2$ $\Delta_{0,95} = 12,7S_{xi}$. Однако при $n \geq 8$ отличие квантилей распределения Стьюдента от квантилей нормального распределения ($n = \infty$) составляет уже менее 20%.

Для распределений x_i , отличных от нормального, использование распределения Стьюдента некорректно.

4.3. СРАВНИТЕЛЬНАЯ ЭФФЕКТИВНОСТЬ РАЗЛИЧНЫХ МЕТОДОВ ОПРЕДЕЛЕНИЯ КООРДИНАТЫ ЦЕНТРА РАСПРЕДЕЛЕНИЯ

Известен целый ряд оценок координаты центра симметрии распределения: среднее арифметическое, медиана, центр размаха и др. При статистической обработке экспериментальных данных важно использовать наиболее эффективную оценку, так как погрешность в определении координаты центра распределения влечет за собой неправильную оценку с. к. о., границ доверительного интервала, эксцесса и контрэксцесса, вида распределения и т. п., т. е. всех последующих оценок, кроме энтропийных. При этом большая затрата времени и других ресурсов для получения многократных отсчетов имеет целью прежде всего уточнение именно координаты центра распределения, и если для этого используется неэффективная оценка, то в соответствующее число раз непроизводительно возрастают затраты на достижение желаемого результата.

Если считать, что затрата времени на проведение измерений пропорциональна числу n осуществленных отсчетов, то целесообразно сравнивать различные оценки именно по числу n , необходимому для достижения одинаковой дисперсии оценки центра распределения.

Проведем такое сравнение для оценки координаты центра распределения в виде медианы X_m , т. е. 50%-ной квантили эмпирического распределения. Дисперсия определенной по выборке с объе-

мом n координаты любой квантили, соответствующей вероятности P [17, с. 404],

$$D[X_P] = P(1 - P)/\{n[p(x_P)]^2\}, \quad (4-6)$$

где $p(x_P)$ — плотность распределения в точке x_P ; x_P — координата квантили.

Для медианы $P = 0,5$ и ее дисперсия $D(X_M) = 1/\{4n[p \times \times (X_M)]^2\}$. Например для распределения Лапласа вида $p(x) = A \exp(-|z|) = \exp(-|x\sqrt{2}/\sigma|)/(\sigma\sqrt{2})$ плотность в центре распределения, т. е. при $x = 0$, составляет $p(X_M) = 1/(\sigma\sqrt{2})$. Отсюда дисперсия рассеяния оценки медианы для распределения Лапласа

$$D[X_M] = \sigma^2/(2n), \quad (4-7)$$

где $\sigma = \sigma_{xt}$.

Сопоставляя дисперсию оценки координаты центра в виде медианы X_M согласно выражению (4-7) и в виде среднего арифметического $\bar{X}$ согласно (4-2) видим, что для распределения Лапласа $D[X_M]$ в два раза меньше, чем $D[\bar{X}]$, т. е. определение координаты центра медианой в два раза эффективнее, а следовательно, для достижений той же погрешности требуется в 2 раза меньший объем выборки исходных данных.

В общем случае для класса экспоненциальных распределений $p(x) = A \exp(-|x|^\alpha)$ можно показать, что относительная эффективность оценки координаты центра медианой по сравнению со средним арифметическим, оцениваемая как отношение дисперсий при равных объемах выборки $E_M = D(\bar{X})/D(X_M)$,

$$E_M = \alpha^2 \Gamma(3/\alpha) / [\Gamma^2(1/\alpha)],$$

где α — показатель степени экспоненты. Это точное соотношение для распределений с эксцессом от $\varepsilon = 3$ до $\varepsilon = 8$ ($0,577 > \varepsilon > 0,35$) может быть аппроксимировано (с погрешностью 5—10%) функцией вида $E_M \approx 0,12\varepsilon^{1,6}$.

Для иллюстрации изложенного кривая изменения эффективности медианы E_M в функции от контрэксцесса нанесена на графике рис. 4-2, а и помечена знаком X_{M1} . Эта кривая пересекается с горизонтальной прямой $E = 1$, соответствующей эффективности среднего арифметического, при $\varepsilon = 0,515$. Для островершинных распределений с $\varepsilon < 0,515$ оценка координаты центра медианой эффективнее, чем оценка в виде $\bar{X}$. Таким образом, оценка центра в виде среднего арифметического отнюдь не является эффективной для всех распределений.

Оценка в виде медианы эффективна только для островершинных распределений. Для плосковершинных и двухмодальных распределений эффективность медианы падает вплоть до нуля. Это иллюстрируется рис. 4-3, где одновременно с формой распределений засечками схематически показано расположение эксперимен-

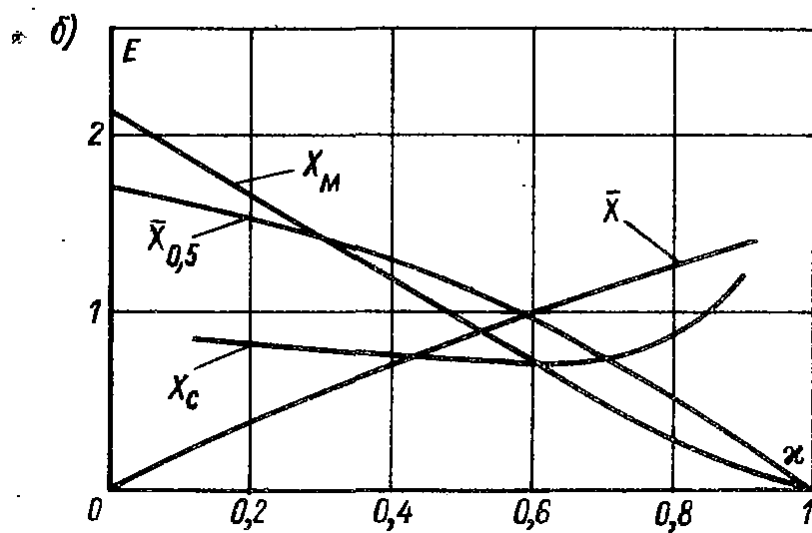
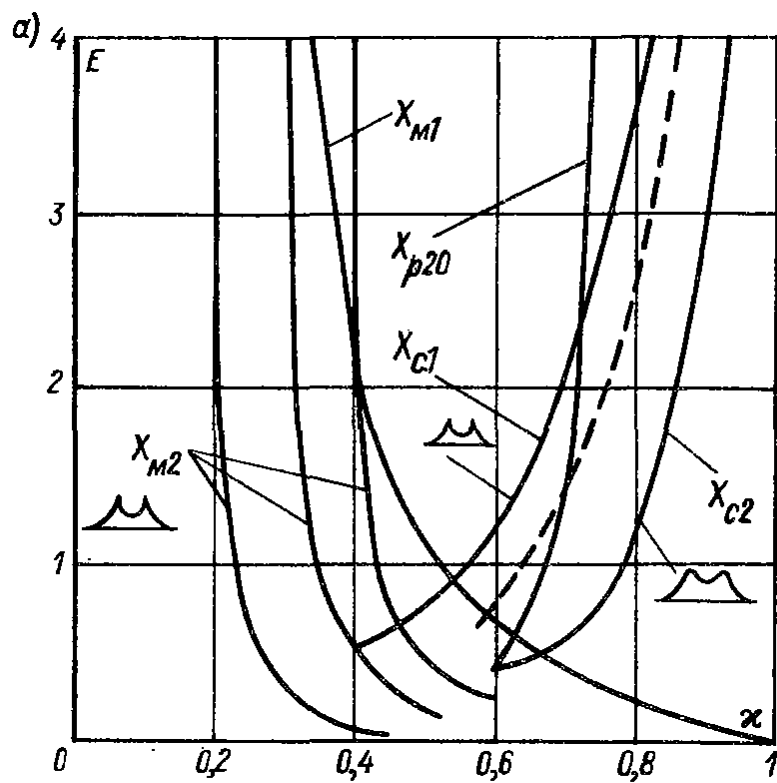


Рис. 4-2

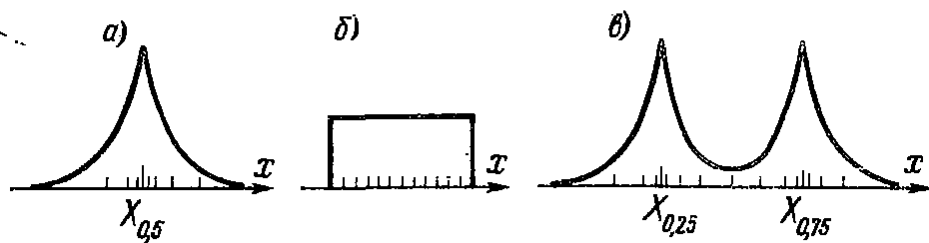


Рис. 4-3

тальных точек на оси x . При островершинных распределениях (рис. 4-3, а) экспериментальные точки в центре распределения располагаются плотно и медиана определяется более точно. При плосковершинных (рис. 4-3, б), а тем более при двухмодальных распределениях (рис. 4-3, в) экспериментальные точки в центре располагаются редко и неопределенность медианы резко возрастает, что показано кривыми X_{m2} на рис. 4-2, а для композиций экспоненциальных (с показателем степени $\alpha = 1/2, 2/3$ и 1) и двузначного дискретного распределений. При симметричных двухмодальных распределениях (рис. 4-3, в) экспериментальные точки группируются наиболее плотно в области 25%- и 75%-ной квантилей, которые Дж. Тьюки [43] предложил именовать левым и правым сгибам и распределения. В результате этого при двухмодальном распределении с наименьшей дисперсией определяются именно координаты сгибов. Поэтому для таких распределений эффективна оценка координаты центра, определяемая как центр сгибов в виде $X_c = (X_{0,25} + X_{0,75})/2$, где $X_{0,25}$ и $X_{0,75}$ — соответственно 25%- и 75%-ная квантили, т. е. сгибы эмпирического распределения.

Изменение эффективности оценки координаты центра распределения как центра сгибов X_c показано на рис. 4-2 кривой X_{c1} для композиций распределения Лапласа и дискретного двузначного распределения, т. е. островершинных, и кривой X_{c2} для композиций нормального и дискретного двузначного, т. е. кругловершинных. Кривая X_{c2} пересекает уровень $E = 1$ при $\kappa \approx 0,79$, а кривая X_{c1} — при $\kappa = 0,57$.

Так как при относительно малом объеме выборки экспериментальных данных трудно установить, является ли исследуемое двухмодальное распределение островершинным или кругловершинным, то на рис. 4-2, а штриховой линией нанесена некоторая средняя кривая, которая описывается простейшим соотношением $E_c = 1,4/(e - 1)$ и поэтому легко может быть использована на практике. Она пересекает уровень $E = 1$ при $\kappa = 0,645$.

Для ограниченных распределений (равномерного, трапецеидального, арксинусоидального и т. д.) оказывается эффективной простейшая оценка координаты центра распределения в виде центра размаха экспериментальных данных, т. е. в виде полусуммы крайних членов вариационного ряда $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ $X_p = (x_{(1)} + x_{(n)})/2$. Дисперсия и эффективность такой оценки зависят от вида распределения и объема выборки и для некоторых распределений и $n = 20$ приведены в табл. 4-1. Соответствующая кривая зависимости эффективности $E_p = f(\kappa)$ помечена на рис. 4-2, а X_{p20} . Она пересекает уровень $E = 1$ при $\kappa = 0,675$.

Таким образом, кривые рис. 4-2, а показывают, что оценка центра распределения в виде $\bar{X}$, безусловно, эффективна лишь для одномодальных распределений, близких к нормальному с контрэксцессом κ от 0,515 до 0,645 (с v от 2,4 до 3,8). Для полого спа-

Таблица 4-1

Вид распределения	κ	$D X_p $	$E = \frac{D \bar{X} }{D X_p }$	E (при $n = 20$)
Нормальное	0,577	$\frac{\pi^2 \sigma^2}{24 \ln n}$	$\frac{24 \ln n}{\pi^2 n}$	0,37
Треугольное	0,65	$\frac{3(4 - \pi) \sigma^2}{2n}$	$\frac{2}{3(4 - \pi)}$	0,70
Равномерное	0,745	$\frac{6\sigma^2}{(n-1)(n-2)}$	$\frac{(n-1)(n-2)}{6n}$	3,85
Арсинусоидальное	0,82	$\frac{5\pi^4 \sigma^2}{n^4}$	$\frac{n^3}{5\pi^4}$	16,4

дающих распределений с $\kappa > 3,8$ более эффективна оценка центра медианой, для двухмодальных распределений — центром сгибов, а для ограниченных — центром размаха.

Особый интерес для практики представляет высокая эффективность оценки в виде центра размаха. Такая оценка широко используется чисто интуитивно (отмечаются максимальные и минимальные значения и за центр принимается середина этого интервала). Оказывается, что для строго ограниченных распределений (равномерного, арксинусоидального, но не треугольного) она много эффективнее $\bar{X}$. Так, при равномерном распределении дисперсия X_p при $n = 80$ равна дисперсии $\bar{X}$ при $n = 1200$, т. е. переход от $\bar{X}$ к X_p при сохранении той же точности позволяет в 15 раз сократить объем наблюдений. Этот эффект убывает при уменьшении n ($E = 1$ при $n \approx 9$), но даже при $n = 20$ дисперсия X_p равна дисперсии $\bar{X}$ при $n = 30$, т. е. сохраняется полуторакратный выигрыш в необходимом объеме выборки.

Еще более резко это различие проявляется при арксинусоидальном распределении. В этом случае $E = 1$ достигается при $n = 8$, а при $n = 20$ дисперсия X_p равна дисперсии $\bar{X}$ при $n = 328$, т. е. $E = 16,4$ раза.

Отсюда следуют прямые рекомендации по построению цифровых усредняющих устройств или выбору программ для усреднения на ЭВМ. Следуя традициям, усреднение осуществляют путем вычисления $\bar{X}$, в то время как переход к определению X_p мог бы дать выигрыш в десятки раз.

Чувствительность различных оценок координаты центра распределения к наличию промахов в выборе экспериментальных данных. Сопоставляя различные оценки определения центра распределения, необходимо, кроме их эффективности (размера дисперсии), учитывать чувствительность к наличию в выборке про-

махов, т. е. резко выделяющихся наблюдений, принадлежащих к другой генеральной совокупности. С этой точки зрения оценка в виде центра размаха $\bar{X}_p$ исключительно чувствительна к наличию промахов, так как она определяется только по наиболее удаленным от центра наблюдениям, каковыми промахи и являются. Оценка центра распределения в виде $\bar{X}$ также слабо защищена от влияния промахов, — влияние промаха ослабляется лишь в n раз, в то время как его возможный размер ничем не ограничен.

Защищенными от влияния размера промахов являются лишь квантильные оценки, т. е. медиана и центр сгибов, так как они не зависят от координат промахов. Однако явлением, резко понижающим точность квантильных оценок, является естественное группирование отсчетов, обусловленное дискретностью возможных отсчетов случайной величины. Пусть, например, она имеет разброс от 608 до 612 единиц и измеряется цифровым прибором с диапазоном от 0 до 999 квантов. Хотя сама величина в диапазоне от 608 до 612 единиц может принимать любые дробные значения, на выходе прибора будут фиксироваться лишь отсчеты 608—609—610—611—612 единиц. Увеличивая объем выборки и вычисляя $\bar{X}$, можно установить, что центр распределения находится, например, при $\bar{X} = 610,53$. Однако сгибы такого распределения при сколь угодно большом объеме выборки будут оставаться равными 609 и 611, а медиана и центр сгибов неизменно равными 610, что не позволяет получить какого-либо уточнения координаты центра при любом увеличении объема выборки и является основным недостатком квантильных оценок.

Попытка сочетать защищенность квантильных оценок от промахов и возможность уточнения с ростом выборки среднего арифметического приводит к смешанному методу определения $\bar{X}$ не по всем n , а лишь после отбрасывания с каждого из концов вариационного ряда равной их доли (например, по 10 или 25 %). В последнем случае сумма при определении $\bar{X}_{0,5}$ из 50 % центральных наблюдений берется не от 1 до n , а от $0,25n$ до $0,75n$, т. е.

$$\bar{X}_{0,5} = \frac{2}{n} \sum_{|n/4|}^{|3n/4|} x_i.$$

В приведенном примере такая методика вычисления $\bar{X}_{0,5}$ приводит к тому, что из суммы $x_{(i)}$ будут опущены все $x_{(i)} = 608$ и $x_{(i)} = 612$, а также часть $x_{(i)} = 609$ и $x_{(i)} = 611$, дополняющая отброшенные с каждого конца до 25 %. Оценка $\bar{X}_{0,5}$ будет вычисляться из оставшихся $x_{(i)} = 609$, $x_{(i)} = 611$ и всех $x_{(i)} = 610$, и если их достаточно много, то она может быть заметно уточнена.

Медиана из пяти разных оценок центра (обозначим ее X_5). В условиях, когда еще нет данных о виде распределения и нет возможности выбрать наиболее эффективную оценку согласно

кривым рис. 4-2, а, рационально вычислить все пять оценок (X_m , X_c , X_p , $\bar{X}_{0,5}$, $\bar{X}$) и за оценку центра X_c для дальнейших вычислений принять оценку X_5 , занявшую медианное положение в их вариационном ряду. Достоинство такой оценки состоит в том, что она полностью защищена от наличия промахов и поэтому не требует их предварительного удаления.

В литературе, однако, нет исследования эффективности такой оценки центра. Поэтому ее эффективность была определена путем моделирования на ЭВМ случайных выборок для следующих законов распределения: Коши ($\kappa = 0$), Лапласа ($\kappa = 0,408$), Симпсона ($\kappa = 0,645$), равномерного ($\kappa = 0,745$) и арксинусоидального ($\kappa = 0,816$). Для каждого распределения на ЭВМ формировалось по 100 выборок с объемом от $n = 9$ через $\Delta n = 8$ (т. е. 17, 25, 33, 41, 49 и т. д.) до $n = 97$. Для распределения Коши из-за большой затраты машинного времени максимальный объем выборок был ограничен $n = 49$. Для каждой группы из 100 выборок данного объема находилось 100 значений оценок X_m , X_c , X_p , $\bar{X}_{0,5}$, $\bar{X}$ и X_5 и по ним вычислялась оценка с. к. о. их рассеяния. Для оценки X_p наблюдалась резкая зависимость с. к. о. от объема выборки. Для остальных же оценок был характерен лишь случайный разброс с. к. о. для различных объемов выборок, поэтому было произведено усреднение их дисперсий и найденные таким образом средние значения с. к. о. для выборок с n от 9 до 97 использовались для вычисления относительной эффективности различных оценок центра X_i относительно оценки X_5 в виде $E_m = \sigma_{X_5}^2 / \sigma_{X_m}^2$, $E_c = \sigma_{X_5}^2 / \sigma_{X_c}^2$, $E_{\bar{X}_{0,5}} = \sigma_{X_5}^2 / \sigma_{\bar{X}_{0,5}}^2$ и $E_{\bar{X}} = \sigma_{X_5}^2 / \sigma_{\bar{X}}^2$; они нанесены на график рис. 4-2, б и приведены ниже:

κ	0	0,408	0,645	0,745	0,816
E_c	0,88	0,77	0,73	0,78	0,93
E_m	2,17	1,14	0,65	0,41	0,25
$E_{\bar{X}_{0,5}}$	1,72	1,28	0,80	0,69	0,50
$E_{\bar{X}}$	0	0,75	1,07	1,20	1,28

Из этих кривых видно, что оценка центра $X_c = X_5$ для распределений от Коши до нормального ($0 < \kappa < 0,58$) более эффективна, чем $\bar{X}$ и X_c , а для распределений с $0,58 < \kappa < 0,83$ более эффективна, чем X_m , $\bar{X}_{0,5}$. Таким образом, $X_c = X_5$ позволяет определять координату центра до удаления промахов во всем диапазоне реальных распределений погрешностей с κ от 0,4 до 0,83 примерно с такой же эффективностью, как X_m в области распределения Лапласа, или $\bar{X}$ и $\bar{X}_{0,5}$ — в области нормального.

Окончательное округление оценки X_c производится по правилам, изложенным в § 1-4, в соответствии с оценкой погрешности этой оценки. Это может быть сделано (см. § 5-3) в соответствии с формулами (4-3) и (4-5).

4-4. РАССЕЯНИЕ ОЦЕНОК С. К. О., КОНТРЭКСЦЕССА, ЭНТРОПИЙНОГО ЗНАЧЕНИЯ И ЭНТРОПИЙНОГО КОЭФФИЦИЕНТА В ЗАВИСИМОСТИ ОТ ОБЪЕМА ВЫБОРКИ И ЭКСЦЕССА РАСПРЕДЕЛЕНИЯ

Расcеяние оценки с. к. о., определяемой по формуле (4-4), зависит от объема выборки n и эксцесса распределения ε . Дисперсия выборочной дисперсии согласно работе [39, с. 320] $D[D^*] = \mu_4/(n-1) - \sigma^4(n-3)/[n(n-1)]$, где σ^2 и μ_4 — второй и четвертый центральные моменты генеральной совокупности, т. е. для оценки неопределенности оценки дисперсии нужно знать второй и четвертый моменты. При $n > 20$ можно с погрешностью 10% полагать $n-3 \approx n-1$, что приводит к приближенным соотношениям [17, стр. 387]:

$$D[D^*] = (\mu_4 - \sigma^4)/n \text{ и } \sigma(\sigma^*) = \sqrt{\mu_4 - \sigma^4}/(2\sigma\sqrt{n}).$$

Последнее равенство можно преобразовать следующим образом: $\sigma(\sigma^*) = \sqrt{(\mu_4/\sigma^4 - 1)\sigma^2}/(2\sqrt{n}) = \sigma\sqrt{\varepsilon - 1}/(2\sqrt{n})$, где $\varepsilon = \mu_4/\sigma^4$ — эксцесс распределения. Отсюда относительная средняя квадратическая погрешность $\delta(\sigma^*)$ определения оценки σ^* зависит только от объема выборки n и эксцесса распределения и не зависит от σ , т. е. от точности, с которой производятся измерения:

$$\delta(\sigma^*) = \sigma(\sigma^*)/\sigma = \sqrt{\varepsilon - 1}/(2\sqrt{n}). \quad (4-8)$$

Для данного вида распределения она может быть уменьшена только путем увеличения объема выборки.

Расcеяние оценки контрэксцесса κ для различных законов распределения выражается разными формулами. Можно показать, что относительная средняя квадратическая погрешность $\delta(\kappa) = \sqrt{D(\kappa)}/\kappa$ [17, стр. 384—388] для равномерного распределения ($\varepsilon = 1,8$, $\kappa = 0,74$) $\delta(\kappa) = \sqrt{32/(315n)}$, для нормального распределения ($\varepsilon = 3$, $\kappa = 0,577$) $\delta(\kappa) = \sqrt{2/(3n)}$, для распределения Лапласа ($\alpha = 1$, $\varepsilon = 6$, $\kappa = 0,4$) $\delta(\kappa) = \sqrt{33/(4n)}$, а для экспоненциального распределения с показателем степени $\alpha = 0,5$ ($\varepsilon = 25,2$, $\kappa = 0,199$) $\delta(\kappa) = \sqrt{557/n}$. Эти соотношения с погрешностью, не превосходящей 8—10%, можно приближенно аппроксимировать единой формулой:

$$\delta(\kappa) = \sigma(\kappa^*)/\kappa \approx \sqrt[4]{(\varepsilon^2 - 1)^3}/\sqrt{29n}. \quad (4-9)$$

Расcеяние оценок энтропийного коэффициента и энтропийного значения погрешности. Для приближенного определения выражений для расчета этих оценок на ЭВМ были сформированы выборки случайных чисел, подчиненных различным законам распределения (арксинусоидальному, равномерному, трапецеидальному, треугольному, нормальному и распределению Лапласа), с объемами выборок от 30 до 200. Путем многократных повторных формирова-

Таблица 4-2

Вид распределения	κ	k	$\sigma(k)$ при объеме выборки n , равном				
			30	50	80	100	200
Лапласа	0,408	1,92	0,16	0,09	0,08	0,08	—
Гаусса	0,577	2,066	0,09	0,07	0,05	0,05	0,045
Симпсона	0,645	2,02	—	0,07	—	0,05	0,035
Трапецеидальное	0,73	1,89	—	0,05	—	0,045	0,030
Равномерное	0,745	1,73	0,10	0,10	0,08	0,08	—
Арсинусоидальное	0,816	1,11	0,17	0,13	0,12	0,10	—

ний выборки и расчета их параметров были определены с. к. о. получаемых оценок энтропийного коэффициента и энтропийного значения случайной величины. Результаты такого моделирования при малых объемах выборок ($n = 30 \div 200$) имели очень большой разброс. Для иллюстрации этого в табл. 4-2 приведены полученные оценки с. к. о., оценки энтропийного коэффициента, вычисленные для разных законов и разных объемов выборок. Вследствие большого разброса этих данных они были аппроксимированы методом наименьших квадратов. Полученные результаты могут рассматриваться лишь как весьма приближенные. Тем не менее их анализ позволяет получить достаточно полную качественную картину рассматриваемого вопроса.

Аппроксимация этих данных может быть произведена следующими выражениями (погрешность адекватности неизвестна):

$$\sigma(k) = \frac{1}{k \sqrt{kn} \sqrt{\kappa^3}}; \quad \delta(k) = \frac{\sigma(k)}{k} = \frac{1}{k^2 \sqrt{kn} \sqrt{\kappa^3}};$$

$$\delta(\Delta_\theta) = \frac{\sigma(\Delta_\theta)}{\Delta_\theta} = \sqrt{\frac{9,15 \cdot 10^{-4}}{n(1-\kappa)^3} + \frac{5,1(1-\kappa)^3}{n}}. \quad (4-10)$$

Из выражений (4-8) — (4-10) необходимый объем выборки экспериментальных данных для получения соответствующей статистической оценки с заданной относительной средней квадратической погрешностью δ соответственно равен:

$$n_\sigma = (\varepsilon - 1)/(4\delta^2); \quad n_\kappa = \sqrt{(\varepsilon^2 - 1)^3}/(29\delta^2);$$

$$n_k = 1/(k^5 \delta^2 \sqrt{\kappa^3});$$

$$n_{\Delta_\theta} = [5,1(1-\kappa)^3 + 9,15 \cdot 10^{-4}/(1-\kappa)^3]/\delta^2.$$

Рассчитанные по этим соотношениям объемы выборок n_σ , n_κ , n_k и n_{Δ_θ} для получения оценок σ , κ , k и Δ_θ с относительной средней квадратической погрешностью $\delta = 0,05 = 5\%$, изображены в виде кривых на рис. 4-4 в функции от значений контрэксцесса κ . Для значений $\kappa < 0,577$ эти кривые рассчитаны для экспоненциальных распределений, а для $\kappa > 0,577$ — для двухмодальных

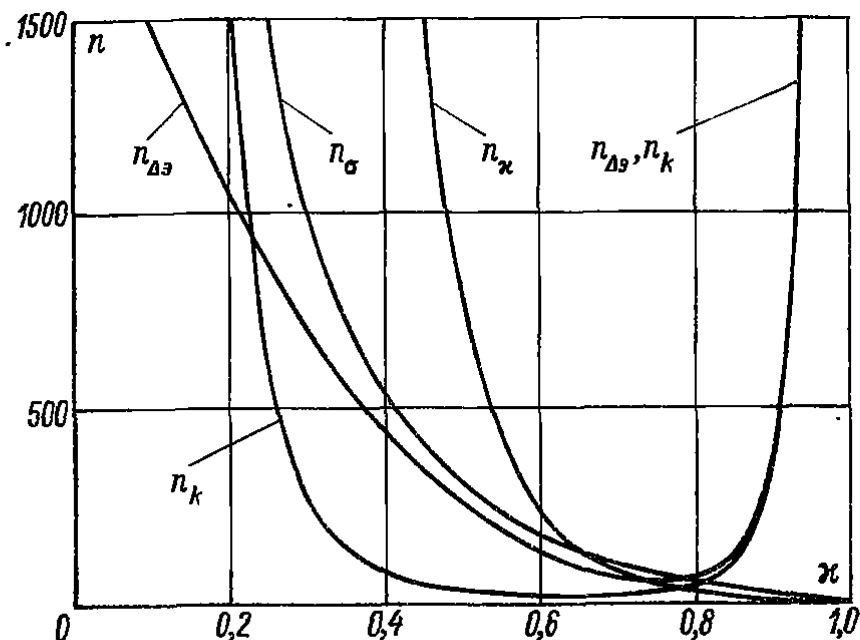


Рис. 4-4

кругловершинных распределений в виде композиций нормального и дискретного двузначного распределений.

Из сопоставления кривых рис. 4-4 следуют достаточно важные для практики выводы. Для распределений с контрэксцессом $k = 0,6 \div 0,8$ рассматриваемые оценки могут быть достаточно точно ($\delta = 5\%$) найдены по относительно малым выборкам ($n = 50 \div 200$). Необходимый объем выборки существенно зависит от вида распределения. Так, если для определения оценки σ с $\delta_{\sigma} = 10\%$ при равномерном распределении необходимо $n = 20$, то при нормальном распределении необходимо $n = 50$, а при распределении Лапласа требуется уже $n = 130$. Еще более резко проявляется подобная зависимость при определении оценки контрэксцесса k (для $\delta_k = 10\%$ требуется при равномерном распределении $n = 10$, при нормальном — $n = 70$, а при распределении Лапласа — $n = 750$).

Для распределений с $k = 0,4 \div 0,8$ рассеяние оценки Δ_3 имеет примерно то же значение, что и рассеяние оценки σ (кривые n_{σ} и n_{Δ_3} идут рядом). Однако при $k < 0,2$ положение существенно меняется, так как рассеяние оценки σ при $k \rightarrow 0$ стремится к ∞ , в то время как оценка Δ_3 имеет конечное рассеяние и при $k = 0$. Но при $k \rightarrow 1$, наоборот, рассеяние оценок σ и k падает до нуля, а рассеяние оценки Δ_3 резко возрастает.

Отмеченные свойства приводят к тому, что в диапазоне значений k от 0,4 до 0,8, т. е. для наиболее часто встречающихся законов распределения погрешностей, с наименьшим разбросом определяется оценка энтропийного коэффициента k . Как видно из кривой n_k на рис. 4-4, для $\delta_k = 0,05 = 5\%$ в этом диапазоне k достаточно выборки с объемом всего $n = 25 \div 50$ отсчетов, в то время как для получения с той же погрешностью оценок σ и Δ_3 , нужны выборки в 200—300 отсчетов.

Странность этого положения подчеркивается еще тем, что оценка k находится путем деления друг на друга оценок Δ_3 и σ .

Казалось бы, что при использовании оценки Δ_σ , найденной с погрешностью $\delta_\Delta = 12\%$, и оценки σ , найденной с погрешностью $\delta_\sigma = 13\%$, рассеяние оценки k должно составлять в лучшем случае $\delta_k = \sqrt{13^2 + 12^2} = 18\%$. В действительности же это рассеяние составляет $\delta_k = 0,05 = 5\%$. Это могло бы быть объяснено, например, тем, что найденные по одной и той же выборке оценки Δ_σ и σ оказываются сильно взаимно коррелированными. Вследствие этого рассеяние оценки k определяется как $\delta_k = \sqrt{\delta_\sigma^2 - \delta_\Delta^2} = \sqrt{13^2 - 12^2} = 5\%$. Прямой расчет коэффициента корреляции $\rho_{\Delta\sigma}$ показывает, что он остается близким к единице при изменении контрэксцесса распределений от 0,3 до 0,7, а затем симметрично падает до $\rho_{\Delta\sigma} = 0$ при $\kappa \rightarrow 0$ и $\kappa \rightarrow 1$. Именно этим и обусловлен U-образный характер кривой n_k (рис. 4-4) и очень малое рассеяние оценки k при $0,4 < \kappa < 0,8$.

Практический вывод проведенного сопоставления рассеяния статистических оценок и интервалов их неопределенности с применением эксцесса состоит в том, что все распределения по рассеянию их оценок можно разделить на две части.

Распределения с κ от 0,6 до 1 (с эксцессом ε от 1 до 3), т. е. крутоспадающие распределения (арксинусоидальное, равномерное, трапецеидальные, треугольное и до нормального), оцениваются относительно легко, так как кривые n_σ и n_k на рис. 4-4 в области изменения κ от 1 до 0,6 идут очень полого и для обеспечения $\delta_\sigma = \delta_k = 10\%$ достаточно выборки исходных данных всего в 50—70 отсчетов. Параметры же распределений с $\kappa < 0,6$, т. е. с эксцессом более 3, имеющих длинные медленно спадающие «хвосты», оцениваются очень трудно, так как кривые рис. 4-4 в этой области круто устремляются вверх. Для определения параметров этих распределений с достаточной точностью нужны большие выборки исходных данных и тем больше, чем больше их эксцесс. Расположение кривых для n_σ и n_k на рис. 4-4 (для $\delta = 5\%$) показывает, что увеличение объемов выборки даже до 500, 1000, 1500 отсчетов не позволяет достоверно оценить значение κ при $\kappa < 0,4$ ($\varepsilon > 6$), а с. к. о. при $\kappa < 0,2$ ($\varepsilon > 24$).

Эти объективные свойства распределений вызвали в последние годы интенсивное развитие специальных методов обработки, достаточно «дюжих» (по-английски — робастных) к росту эксцесса, чтобы обеспечить приемлемую точность получаемых оценок и для распределений с «утяжеленными хвостами», как принято выражаться в работах по «робастному» оцениванию. С этими идеями можно ознакомиться по переведенным на русский язык книгам П. Хьюбера, Р. Лонера и Г. Уилкинсона [44, 47] и др.

Некоторые методы определения оценок, устойчивых к «утяжеленным хвостам», были изложены выше (определение координаты центра медианой, центром сгибов, медианой из пяти оценок). а некоторые будут описаны в дальнейшем (защита от далеко отстоящих отсчетов и др.). Однако наиболее действенным приемом

является идентификация вида обрабатываемого распределения и использование тех оценок, которые наиболее эффективны для данного распределения. Поэтому методам идентификации формы распределений посвящена гл. 5.

Изложенные выше далеко не очевидные зависимости свойств статистических оценок параметров распределений от вида закона распределения еще раз подчеркивают, что требования, которые были введены в ГОСТ 8.011—72, о необходимости определения и указания вида закона распределения погрешности являются реальной потребностью повседневной практики и его отмена в 1984 г. ради механического сокращения общего числа стандартов вряд ли была научно обоснованна.

4.5. ПРОМАХИ И МЕТОДЫ ИХ ИСКЛЮЧЕНИЯ

Одним из условий правомерности статистической обработки выборки является требование ее однородности, т. е. принадлежности всех ее членов к одной и той же генеральной совокупности. Однако на практике это требование очень часто нарушается. Ясно, что совместно обрабатывать данные, принадлежащие двум совершенно различным генеральным совокупностям, бессмысленно. Однако дать формальное определение «чужим» отсчетам, т. е. перечислить объективные признаки, по которым их можно было бы достоверно отличать от «нужных» отсчетов, практически невозможно.

Если измерения и их последующая обработка проводятся одним и тем же человеком, то для исключения из выборки «неподходящих» отсчетов он может воспользоваться своими воспоминаниями о каких-либо нарушениях условий эксперимента в момент получения этих отсчетов, положиться на свою интуицию и т. п. Но в тех случаях, когда обработка, а, возможно, и сам эксперимент проводятся ИВК без участия оператора, формальные методы исключения «чужих» для данной выборки отсчетов приобретают первостепенное значение.

Загрязнения и промахи. Отсчеты, принадлежащие другой генеральной совокупности, по своим значениям могут не отличаться существенно от значений интересующей нас генеральной совокупности. Обнаружить одновременное присутствие двух таких групп отсчетов в анализируемой выборке (при достаточно большом ее объеме) можно по виду кривой плотности распределения (см. рис. 5-9, а и б), получающейся в результате такого смешения. По виду подобной кривой плотности распределения ясно, что выборка состоит из смеси отсчетов двух генеральных совокупностей с разными законами распределения. Наличие таких аномальных отсчетов принято называть **з а г р я з н е н и е м** выборки, однако выделить члены выборки, принадлежащие к каждой из генеральных совокупностей, в этом случае практически невозможно.

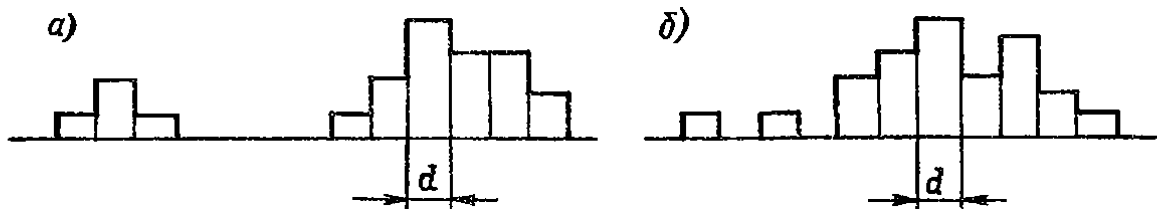


Рис. 4-5

Другим видом аномальных отсчетов являются отсчеты, принадлежащие к разным генеральным совокупностям и различающиеся по своим значениям. Если это различие настолько существенно (рис. 4-5, а), что выборки на гистограмме не соприкасаются и не смешиваются между собой, то визуальное разделение таких двух групп отсчетов не представляет труда. Отсчеты, резко отклоняющиеся по своим значениям от большинства других отсчетов выборки, принято называть *промахами* и исключать из выборки.

Но особую неприятность при обработке экспериментальных данных доставляют отсчеты, которые, как это показано на рис. 4-5, б, хотя и не входят в компактную группу основной массы отсчетов выборки, но и не удалены от нее на значительное расстояние. Такой отсчет будем называть *предполагаемым промахом*.

Таким образом, единственным формальным признаком «чужого» наблюдения является его аномально большое удаление от центра распределения. Поэтому в экспериментальной практике исследователи издавна стали просто отбрасывать крайние, «слишком» удаленные от центра наблюдения. Этот способ получил название *цензурирования выборки*. Однако для принятия решения об исключении предполагаемого промаха необходимы какие-либо формальные критерии.

Методы назначения границ цензурирования выборки для удаления промахов. Простейший из таких методов заключается в использовании «правила 3σ », когда по выборке с удаленными отсчетами, похожими на промахи, вычисляется оценка σ и граница цензурирования назначается в виде $|X_{гр}| = 3\sigma$, а все $|x_i| \geq 3\sigma$ признаются промахами и удаляются из дальнейших расчетов.

Г. А. Агекян [1] отмечает, что «правило 3σ » во многих случаях может оказаться слишком «жестким» и рекомендует оценки $X_{ц}$ и σ определять без использования отсчетов, предполагаемых промахами, а границу цензурирования назначать в зависимости от объема выборки n :

$$\begin{aligned} \text{при } 6 < n \leq 100 & \quad |X_{гр}| = 4\sigma; \\ \text{при } 100 < n \leq 1000 & \quad |X_{гр}| = 4,5\sigma; \\ \text{при } 1000 < n \leq 10\,000 & \quad |X_{гр}| = 5\sigma. \end{aligned}$$

Еще более квалифицированно (с использованием таблиц вероятности крайних членов вариационного ряда для нормального распределения) это было регламентировано в ГОСТ 11.002—73.

Однако для распределений, отличающихся от нормального, использование подобных таблиц лишено смысла. Действительно, если для нормального распределения при $n = 100$ появление $|x_i| \geq 3\sigma$ можно считать промахом, то для равномерного распределения (рис. 4-6) промахом является уже $|x_i| = 1,8\sigma$, в то время как для экспоненциального распределения Лапласа $|x_i| = 3\sigma$ есть, безусловно, отсчет, принадлежащий данной выборке.

Таким образом, границы $t\sigma$ цензурирования выборки должны зависеть не только от объема n выборки, но и от вида распределения.

Назначая ту или иную границу цензурирования, необходимо оценить уровень значимости $q = 1 - P$, т. е. вероятность отсеять какую-то часть отсчетов, принадлежащих обрабатываемой выборке. Этот уровень значимости может быть выбран на основании следующих рассуждений. Как было показано в § 2-2, n экспериментальных отсчетов разбивают ось x в пределах от $-\infty$ до $+\infty$ на $n + 1$ интервалов и между крайними членами вариационного ряда заключено $n - 1$ интервалов. Если назначить границы цензурирования так, чтобы расстояние между ними совпадало с расстоянием между крайними точками, то малейшая ошибка в определении координаты центра распределения будет приводить к тому, что одна из крайних точек будет оказываться за границей цензурирования, т. е. назначение границ с уровнем значимости $q = 1 - P$, где $P = (n - 1)/(n + 1)$, обеспечивает всегда отсечение какой-то одной точки выборки (слева или справа).

Если же поставить условие, что границы цензурирования должны в среднем отсекалть менее одной точки выборки, то $P = n/(n + 1)$ и $q = 1 - P = 1/(n + 1)$. Это соотношение и определяет выбор границ цензурирования в функции от объема выборки n .

Однако зависимость $P = f(t)$ существенно различна для разных законов распределения. Рассматривая этот вопрос применительно к погрешностям результатов измерений, ограничим разнообразие законов распределения лишь теми четырьмя классами симметричных распределений, которые использовались выше, а имен-

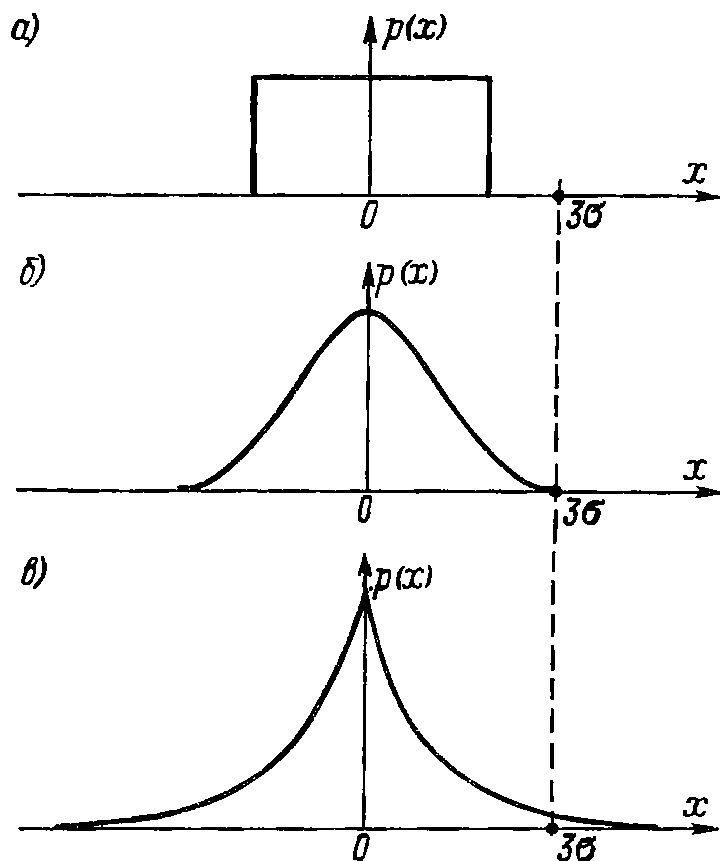


Рис. 4-6

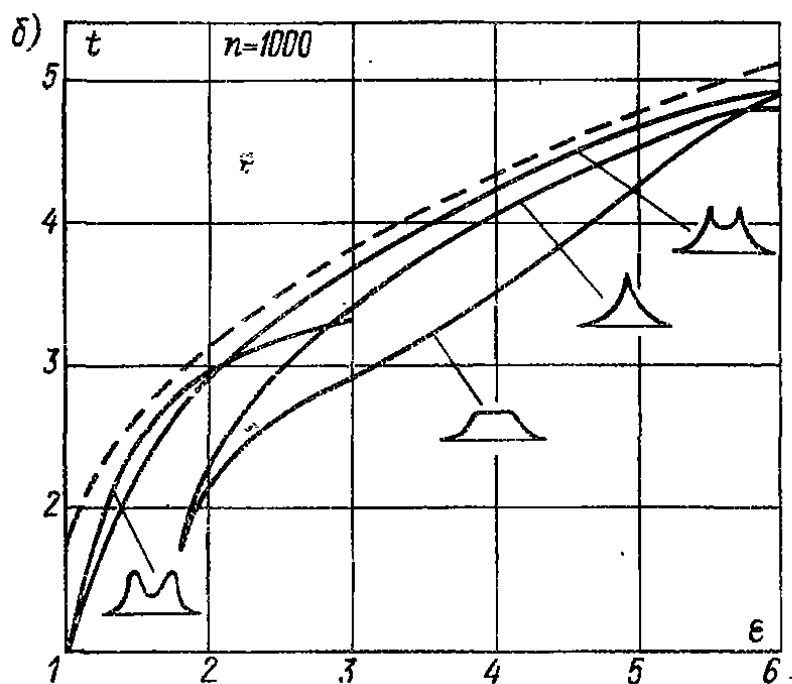
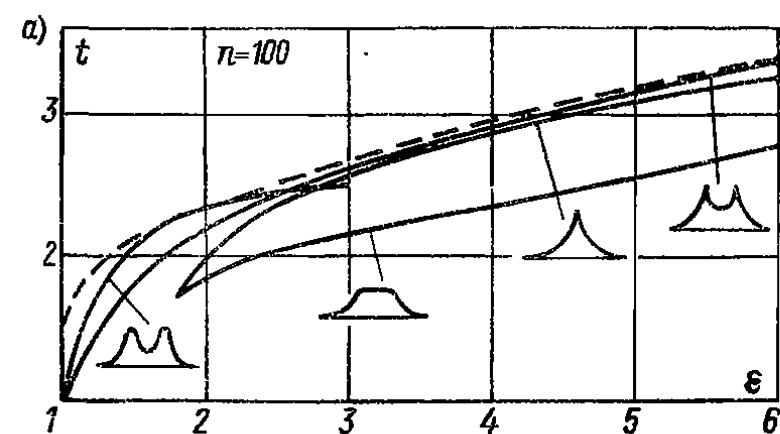


Рис. 4-7

но: классом двухмодальных кругловершинных композиций нормального и дискретного двузначного распределения в пределах изменения эксцесса от $\varepsilon = 1,5$ до $\varepsilon = 3$, классом островершинных двухмодальных композиций дискретного двузначного распределения и распределения Лапласа в пределах изменения ε от 1,5 до 6, классом композиций равномерного распределения с экспоненциальным распределением с показателем степени $\alpha = 1/2$ в пределах изменения эксцесса ε от 1,8 до 6 и классом экспоненциальных распределений также в пределах изменения эксцесса ε от 1,8 до 6.

Зависимости квантильного множителя t перечисленных распределений от ε для $P = n/(n+1)$ при $n = 100$ и $n = 1000$ пред-

ставлены в виде кривых на рис. 4-7. Для выполнения условия $q < 1/(n+1)$ значения $t_{гр}$ выбираемых границ цензурирования должны быть при всех сочетаниях ε и n больше $t = f(\varepsilon)$ для всего представленного на рис. 4-7 жгута кривых, т. е., например, соответствовать огибающим этого жгута, нанесенным на рис. 4-7 штриховыми линиями. Огибающие могут быть аппроксимированы более простыми выражениями (без использования переменных или дробных показателей степени в отличие от кривых $P = f(\varepsilon, t)$, обсуждавшихся в § 2-6), например выражением

$$t_{гр} = 1,55 + 0,8 \sqrt{\varepsilon - 1} \lg(n/10), \quad (4-11)$$

согласно которому и нанесены штриховые кривые на рис. 4-7.

Семейство кривых $t_{гр} = f(\varepsilon, n)$ согласно формуле (4-11) для n в пределах от 20 до 10 000 и ε от 1,5 до 6 представлено на рис. 4-8, из которого видно, как зависит граница цензурирования от эксцесса распределения и объема выборки.

При использовании для удаления промахов соотношения (4-11)

следует иметь в виду, что $X_{гр} = t_{гр}\sigma$ рассчитывается с существенными погрешностями вследствие статистического разброса найденных по выборке оценок σ и ϵ . Относительная погрешность определения оценки контрэксцесса κ согласно (4-9), например, для $n = 100$ и ϵ от 1,8 до 6 составляет от 3 до 26%. Однако такой разброс κ приводит к изменению $t_{гр}$ менее чем на 1—6%. Поэтому основная доля рассеяния $X_{гр}$ определяется не рассеянием оценки контрэксцесса κ , а рассеянием оценки с. к. о. σ .

Рассеяние оценки с. к. о. σ при $n = 100$ и ϵ от 1,8 до 6 согласно формуле (4-8) составляет от 4 до 11%. Если бы рассеяния оценок κ и σ , определяемых по одной и той же выборке, были бы независимыми, то погрешность определения $X_{гр}$ по соотношению (4-11) находилась бы (при $n = 100$) в пределах от $\sqrt{1^2 + 4^2} = 4,1\%$ до $\sqrt{6^2 + 11^2} = 12,5\%$, т. е. практически была бы близка к погрешности определения σ . Однако анализ результатов обработки экспериментальных данных показывает, что отклонения оценок σ и $t_{гр}$, определенных по одной и той же выборке, оказываются отрицательно коррелированными, т. е. возрастание σ сопровождается уменьшением $t_{гр}$ и их погрешности не складываются, а вычитаются. Это наблюдение заслуживает дальнейшего внимательного изучения. Однако уже из изложенного следует, что относительная погрешность границы отсечения промахов $X_{гр}$ прежде всего обусловлена относительной погрешностью $\delta(\sigma)$ оценки σ и имеет близкое к ней значение. Таким образом, для распределений с $\epsilon \leq 6$ определение $X_{гр}$ по соотношению (4-11) является достаточно точным и может широко использоваться на практике.

В ответственных случаях можно вычислить наибольшее возможное удаление границы промахов, подставив в выражении (4-11) вместо оценок σ и ϵ их наибольшие (с вероятностью 0,9) значения $\sigma_{0,9} = \sigma [1 + 1,6\delta(\sigma)]$ и $\epsilon_{0,9} = \epsilon [1 + 1,6\delta(\epsilon)]$. Учитывая, что $\delta(\epsilon) \approx 2\delta(\kappa)$, и используя формулы (4-8) и (4-9), получим:

$$\sigma_{0,9} = \sigma [1 + 0,8 \sqrt{(\epsilon - 1)/n}];$$

$$\epsilon_{0,9} = \epsilon [1 + 3,2 \sqrt[4]{(\epsilon^2 - 1)^3 / 29n}].$$

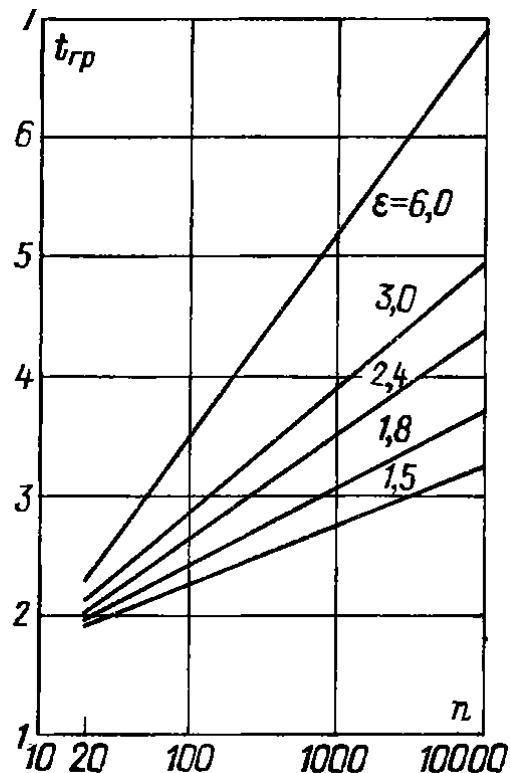


Рис. 4-8

В заключение описания аналитического, т. е. пригодного для использования в программах ЭВМ, метода исключения промахов необходимо заметить следующее. Если расчет оценок σ и κ про-

водится по выборке с неисключенным большим отклонением, то эти оценки существенно возрастают, увеличивая тем самым определяемую границу $X_{гр} = t_{гр}\sigma$. Поэтому вычисление оценок σ и κ должно производиться всегда после исключения далеко отстоящих отсчетов, похожих на промахи. После расчета границ промахов возвращаются в выборку лишь те из этих отсчетов, которые оказываются внутри вычисленных границ.

4.6. АЛГОРИТМ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МНОГОКРАТНЫХ ОТСЧЕТОВ НА ЭВМ БЕЗ УЧАСТИЯ ОПЕРАТОРА

Рассмотренные методы определения различных параметров выборки многократных отсчетов одной и той же измеряемой величины позволяют составить из них алгоритм обработки на ЭВМ.

Этот алгоритм прежде всего должен обеспечить удаление из выборки промахов, которые в противном случае могут существенно исказить все получаемые оценки. Однако воспользоваться для этого соотношением (4-11) можно лишь после того, как будут известны оценки $X_{ц}$, σ и ε . Чтобы разомкнуть этот круг, следует предусмотреть в программе нахождение, как это было проделано в § 4-3, не одной, а всех пяти рассмотренных там оценок центра (X_m , X_p , X_c , $\bar{X}$ и $\bar{X}_{0,5}$) и для предварительных расчетов принять ту из них, которая займет медианное положение в их упорядоченном ряду. Используя эту оценку координаты центра распределения, можно уже определить оценки σ , κ и ε и воспользоваться ими для вычисления границы $X_{гр}$ цензурирования выборки. Однако так как далеко отстоящие наблюдения могут существенно изменить оценки σ , κ и ε , то их вычисление и расчет границы цензурирования должны всегда производиться без учета тех отсчетов, которые наиболее удалены от центра.

Если после определения границ цензурирования $\pm X_{гр}$ изъятые из выборки наиболее удаленные от центра отсчеты окажутся внутри этих границ, то они должны быть возвращены в выборку и расчет всех оценок необходимо повторить заново. Если же кроме предварительно удаленных отсчетов найденные границы цензурирования $\pm X_{гр}$ отсекут от выборки еще какую-то часть отсчетов, то расчет также должен быть повторен после удаления этих отсчетов.

Этот итерационный процесс с использованием соотношения (4-11) обычно заканчивается после 2—3 повторных расчетов. При работе ЭВМ в режиме диалога с оператором по окончании каждого такого итерационного цикла полезно вывести на дисплей или цифропечатать сообщение «из выборки удалены промахи» и перечислить значения удаленных отсчетов.

После удаления промахов следует предусмотреть сопоставление полученной оценки контрэксцесса выборки с интервалами

$\kappa < 0,515$, $0,515 < \kappa < 0,63$ и $0,63 < \kappa$ и в зависимости от результата такого сравнения (в предположении, что при $\kappa < 0,515$ распределение одномодально) выбрать наиболее эффективную оценку центра. Относительно этой оценки центра вычисляются все оценки распределения и одновременно рассчитываются границы их интервалов неопределенности, в соответствии с которыми производится округление значений полученных оценок перед выводом их на цифропечать. Поэтому выводимый на печать протокол обработки имеет следующий вид:

«Из выборки удалены промахи 000, 000, 000, 000, 000.

Оставшийся объем выборки $= n$.

Центр распределения $= X_{\text{ц}} \pm \Delta_{0,9} (X_{\text{ц}})$.

С. к. о. $= \sigma \pm \Delta_{0,9} (\sigma)$.

Эксцесс $= \epsilon \pm \Delta_{0,9} (\epsilon)$.

Контрэксцесс $= \kappa \pm \Delta_{0,9} (\kappa)$ ».

При таком виде выходного документа получатель будет предупрежден об интервалах неопределенности всех полученных оценок.

4.7. УЧЕТ СИСТЕМАТИЧЕСКИХ ПОГРЕШНОСТЕЙ ПРИ ОЦЕНКЕ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МНОГОКРАТНЫХ ОТСЧЕТОВ

Оценивая погрешность результата, полученного при статистической обработке многократных отсчетов, нельзя забывать о том, что при усреднении уменьшаются лишь случайные погрешности, в то время как систематическая погрешность, присутствовавшая во всех усредняемых отсчетах, остается без изменения.

Эту особенность систематических погрешностей следует иметь в виду как при ручной статистической обработке экспериментальных данных, так и, особенно, при организации усреднения многократных отсчетов в ИИС и ИВК, когда их число может достигать сотен или даже тысяч. В этом случае случайная погрешность будет уменьшена в десятки или сотни раз и погрешность усредненного результата будет характеризоваться не этой ничтожной случайной погрешностью, а определяться не зависящим от числа усредняемых отсчетов значением систематической составляющей погрешности.

Вероятностное описание систематической и случайной составляющих погрешности. То или иное значение систематической составляющей погрешности, повторяющейся во всех отсчетах, а поэтому не усредняемой при статистической обработке, присутствует при любом измерении. На рис. 1-10 были приведены графики положения полосы погрешностей каналов ИИС внутри границ $+\gamma_{\text{кл}}$ и $-\gamma_{\text{кл}}$, определяемых их классом точности. Такое расположение полосы погрешностей характерно для любых СИ. При расположении полосы погрешностей, показанном на рис. 1-10, б или в, систематическая составляющая погрешности имеет место при всех

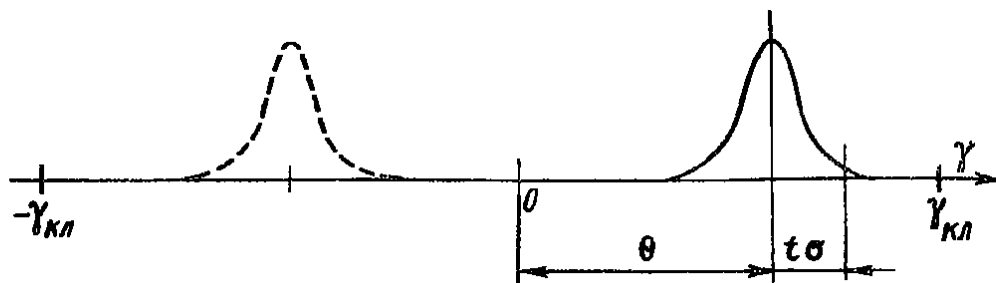


Рис. 4-9

значениях измеряемой величины x , за исключением лишь одной точки в середине диапазона на рис. 1-10, в, когда при малых x она положительна, а при больших — отрицательна.

Поэтому в общем случае положение кривой плотности вероятности погрешности в границах $\pm \gamma_{кл}$ может быть представлено графиком рис. 4-9, где сплошной линией показано положение кривой плотности вероятности погрешности при положительном значении θ систематической составляющей, а штриховой кривой — при отрицательном.

Особенности суммирования систематической и случайной составляющих погрешности. В книгах по метрологии нередко утверждается, что складывать между собой случайные и систематические составляющие погрешности «нельзя с принципиальных позиций, так как систематические и случайные погрешности имеют разную природу». Однако это утверждение вряд ли бесспорно хотя бы потому, что разделение погрешности на систематическую и случайную составляющие мы вводим сами для облегчения анализа. Но после проведения такого анализа правомерна постановка и обратной задачи — задачи суммирования этих составляющих.

Исходя из рис. 4-9 наглядно виден механизм такого суммирования. Если доверительная граница с вероятностью P_d (равной, например, 0,9) для случайной составляющей определяется как $\gamma_{0,9} = t\sigma$, то с учетом систематической составляющей она будет выражаться как $|\theta| + t\sigma$. Но при $|\theta| > 0,66\sigma$ выход погрешности за границы $\pm (|\theta| + t\sigma)$ будет происходить даже для распределения Лапласа практически только с одной стороны, т. е., например, при оценке случайной составляющей с $P_d = 0,9$ доверительная вероятность выхода результата за границы $\pm (|\theta| + t\sigma)$ будет иметь значение $P_d = 0,95$.

Таким образом, механизм суммирования систематической и случайной составляющих резко отличается от механизма суммирования случайных погрешностей. Во-первых, систематическая погрешность может суммироваться только с доверительным (или энтропийным) значением погрешности, а отнюдь не со с. к. о., во-вторых, это суммирование происходит арифметически с модулем систематической погрешности (без учета ее знака) и, в-третьих, результирующая погрешность, указываемая как $\gamma_{\Sigma} = (|\theta| + t\sigma)$ при $|\theta| > 0,66\sigma$, получается с уровнем значимости $q = (1 - P_d)/2$,

где P_d — доверительная вероятность, с которой была определена случайная составляющая погрешности.

Распределение погрешности усредненного результата многократных отсчетов также имеет вид кривой рис. 4-9. При этом систематическая составляющая погрешности θ остается без изменения, а ширина разброса случайной составляющей погрешности $\sigma_{xi}/\sqrt{n}$ уменьшается в $\sqrt{n}$ раз, где n — число усредненных отсчетов. Поэтому если n достаточно велико, то $\sigma_{\bar{x}} \ll \theta$ и результирующая погрешность усредненного результата определяется, по существу, только его систематической погрешностью.

В этой связи ГОСТ 8.207—76 устанавливает, что если $\theta < 0,8\sigma_{\bar{x}}$, то следует пренебречь систематической составляющей погрешности и учитывать только случайную погрешность усредненного результата в виде $\sigma_{\bar{x}}$. Если же $\theta > 0,8\sigma_{\bar{x}}$, то, наоборот, следует пренебречь случайной составляющей и усредненный результат характеризовать лишь его систематической погрешностью θ . «Магические числа» 8 и 0,8 в тексте ГОСТ 8.207—76 никак не обосновываются и не поясняются. Однако автора стандарта [37] можно понять, если учесть, что $\gamma_{0,9} = 1,6\sigma_{\bar{x}}$, отсюда следует, что условие $\theta < 0,8\sigma_{\bar{x}}$ эквивалентно условию $\theta < \gamma_{0,9}\sigma_{\bar{x}}/2$, а условие $\theta > 0,8\sigma_{\bar{x}}$ эквивалентно условию $\theta > 5\gamma_{0,9}\sigma_{\bar{x}}$; т. е. стандарт разрешает пренебрегать систематической погрешностью усредненного результата и учитывать лишь его случайную погрешность $\sigma_{\bar{x}} = \sigma_{xi}/\sqrt{n}$ только тогда, когда она больше удвоенного значения систематической составляющей. И, наоборот, когда случайная составляющая путем усреднения уменьшена до значения $\gamma_{0,9}\sigma_{\bar{x}} \leq \theta/5$, т. е. становится в 5 раз меньше систематической, то стандарт разрешает ее пренебречь и указывать только систематическую погрешность.

Ограниченность возможности повышения точности путем усреднения. Рассмотренные ограничения являются очень серьезными, так как наличие неисключенной систематической погрешности делает практически бессмысленным использование статистического усреднения.

Рассмотрим эти соотношения на конкретном примере. Описанное выше (§ 1-6) исследование погрешностей нескольких тысяч электроизмерительных приборов на ЛПЭО «Электросила» показало, что погрешность новых или только что отремонтированных приборов, когда их систематические погрешности путем тщательной регулировки сведены к минимуму, составляет не менее $0,4\gamma_{кл}$. Эту погрешность можно приближенно считать равной его случайной погрешности, т. е. считать $0,4\gamma_{кл} = 1,6\sigma_{xi}$, которая во время эксплуатации остается неизменной.

По мере старения прибора систематическая составляющая θ погрешности прибора возрастает и при достижении $\theta = 0,6\gamma_{кл}$, т. е. $\theta + 1,6\sigma_{xi} = 0,4\gamma_{кл} + 0,6\gamma_{кл} = \gamma_{кл}$, прибор бракуется при

очередной поверке и направляется в ремонт. Таким образом, систематические погрешности всего парка находящихся в эксплуатации приборов всегда находятся в пределах от $\theta = 0$ до $\theta = 0,6\gamma_{\text{кл}}$ и в среднем для парка составляют примерно $0,3\gamma_{\text{кл}}$.

Каковы же в среднем в этих условиях возможности повышения точности результатов измерений путем статистического усреднения многократных отсчетов?

Ясно, что при $2\theta = 0,6\gamma_{\text{кл}}$ и $\gamma_{0,9} = 0,4\gamma_{\text{кл}}$ условие $2\theta < \gamma_{0,9}$ не выполняется даже при однократных измерениях, не говоря уже о многократных, и пренебречь систематической составляющей погрешности нельзя при любом числе усредняемых отсчетов.

Условие $\gamma_{0,9\bar{x}} = \theta/5$ достигается при $\gamma_{0,9\bar{x}}/\sqrt{n} = \theta/5$, т. е. при $0,4\gamma_{\text{кл}}/\sqrt{n} = 0,3\gamma_{\text{кл}}/5$ или $n = (5 \cdot 0,4)^2 / 0,3^2 \approx 45$, т. е. при усреднении 45 отсчетов случайная погрешность усредненного результата $\gamma_{0,9\bar{x}}$ становится равной $\theta/5$ и в соответствии с рекомендацией ГОСТ 8.207—76 ею можно пренебречь, а погрешность усредненного результата оценивать значением систематической погрешности приборов, равной в среднем $\theta = 0,3\gamma_{\text{кл}}$. Поэтому дальнейшее увеличение числа усредняемых отсчетов ($n > 45$) и дополнительная трата времени становятся бессмысленными.

Таким образом, рассмотрение фактически существующего в среднем соотношения между случайной ($\gamma_{0,9\text{сл}} = 0,4\gamma_{\text{кл}}$) и систематической ($\theta = 0,3\gamma_{\text{кл}}$) составляющими погрешности электроизмерительных приборов показывает, что путем статистического усреднения погрешность результата может быть снижена с $\gamma_1 = 0,7\gamma_{\text{кл}}$ (при единичном измерении) до $\gamma_{45} = 0,3\gamma_{\text{кл}}$ (при $n = 45$), т. е. лишь в 2,3 раза при 45-кратном увеличении затрат времени, и что возможности снижения погрешностей путем статистического усреднения весьма ограничены.

Второй важный вывод из рассмотренного примера состоит в том, что при оценке погрешности результата статистического усреднения крайне важен всесторонний анализ и учет неисключенных систематических погрешностей, которые не уменьшаются при статистическом усреднении, о чем часто забывают, увлекшись изящностью методов статистической обработки.

И, наконец, третий вывод заключается в том, что практически реализовать все возможности статистического повышения точности можно лишь тогда, когда одновременно со статистическим усреднением случайных погрешностей производится достаточно полное и скрупулезное исключение систематических погрешностей.

Автоматическое исключение систематических погрешностей. При метрологических измерениях, проводимых в значительной своей части вручную, все систематические погрешности тщательно анализируются и исключаются путем введения соответствующих поправок. Подобное исключение систематических погрешностей, вообще говоря, могло быть произведено и при прове-

дении технических измерений. Так, например, если полоса погрешностей используемого СИ имеет вид, показанный на рис. 1-9, то при каждом конкретном значении измеряемой величины x может быть определена ордината центра полосы погрешностей и составлена соответствующая таблица поправок. Подобным же образом могут быть составлены таблицы поправок и на систематические погрешности от влияющих величин и т. п.

Однако такое тщательное исключение систематических погрешностей требует подробного изучения всех вызывающих их причин и очень трудоемко само по себе. Поэтому при проведении массовых измерений с помощью многоканальных ИИС или ИВК оно практически неосуществимо. В этом случае, как уже упоминалось в § 1-8, чаще прибегают к сквозной поверке каналов ИИС и ИВК с помощью образцовых мер или образцовых сигналов и осуществлению коррекции нуля и чувствительности канала, т. е. к устранению накопившихся к данному моменту прогрессирующих погрешностей без выяснения и анализа вызывающих их причин. В результате коррекции нуля и чувствительности начало и конец характеристики канала совмещаются со своими номинальными значениями и полоса погрешностей канала, казалось бы, должна занимать в границах $\pm \gamma_{кл}$ положение, показанное на рис. 1-10, а.

Однако и в этом случае нельзя считать, что систематические погрешности полностью исключаются, так как совмещение характеристики канала со своим номинальным положением происходит также с какой-то погрешностью. Поэтому положение полосы погрешности канала после проведения коррекции нуля и чувствительности не соответствует строго положению, показанному на рис. 1-10, а, а хотя и с меньшими отклонениями соответствует рис. 1-10, б или в.

Эти погрешности совмещения характеристики прибора или канала с ее номинальным положением после проведения коррекции нуля и чувствительности, по определению М. Ф. Маликова [23, с. 269], следует именовать *«неисключенными остатками систематических погрешностей»*. Эти «остатки» в зависимости от частоты проведения коррекции и длины серии усредняемых отсчетов могут иметь двоякий характер.

Если коррекция производится однажды, а затем проводится серия усредняемых отсчетов, то остаточная систематическая погрешность во всех этих отсчетах остается постоянной и не усредняется при их статистической обработке. Если же коррекция производится заново перед каждым измерением, то вследствие случайной погрешности прибора или канала значения остаточной систематической погрешности получают после каждой новой коррекции, т. е. в каждом новом отсчете, случайные независимые значения и будут усредняться при усреднении включающих их отсчетов. Рассмотрим оба эти случая.

Величина неисключенных остатков систематической погрешности при осуществлении коррекции не так уж мала. В § 1-6

приводились данные экспериментального определения погрешности, вносимой при коррекции. Оказалось, что дисперсия случайной погрешности установки указателя на нулевую отметку шкалы прибора в точности равна дисперсии случайной погрешности его показаний, а общая дисперсия отсчетов удваивается.

Таким образом, при осуществлении коррекции нуля и чувствительности прибора или канала ИВК (как ручной, так и автоматической) систематическая погрешность отнюдь не устраняется полностью. Происходит лишь замена ранее имевшей место систематической погрешности θ (см. рис. 4-9) ее новым постоянным значением, получающим случайным образом одно из значений в пределах $\pm t\sigma$.

Среднее медианное значение неисключенного остатка систематической погрешности равно средневероятному по модулю значению случайной погрешности корректируемого прибора или канала. В зависимости от вида закона распределения случайной погрешности средневероятное значение составляет для равномерного распределения $\gamma_{\text{ср}} = 0,86\sigma$, для распределения Симпсона $\gamma_{\text{ср}} = 0,82\sigma$, для распределения Лапласа $\gamma_{\text{ср}} = 0,69\sigma$ и для распределения Гаусса $\gamma_{\text{ср}} = 0,68\sigma$. Поэтому приблизительно неисключенные остатки систематической погрешности при однократной коррекции можно оценить величиной $\theta = (0,77 \pm 0,1) \sigma$. И следовательно, при однократных измерениях по нормам ГОСТ 8.207—76 ($\theta < 0,8\sigma$) этим остатком неисключенной систематической погрешности можно пренебречь. Но при усреднении хотя бы двух отсчетов этим остатком систематической погрешности пренебрегать уже нельзя, так как имевшийся запас будет уже исчерпан.

Приведенные выше средние соотношения ($\theta = 0,3\gamma_{\text{кл}}$ и $\gamma_{0,9 \pm 1} = 0,4\gamma_{\text{кл}}$) можно использовать не только для определения погрешности электроизмерительных приборов, но и для приближенной оценки соотношения систематической и случайной составляющих погрешности измерительных каналов ИИС и ИВК, так как они обусловлены лишь тем, что при выпуске из ремонта или производства погрешность канала не бывает меньше, чем $0,4\gamma_{\text{кл}}$, а при достижении погрешностью значения $\gamma_{\text{кл}}$, соответствующего классу точности, канал изымается из эксплуатации. Поэтому при условии, что погрешность каналов ИИС или ИВК периодически контролируется и каналы, погрешность которых достигла нормируемого предела в виде $\gamma_{\text{кл}}$, из эксплуатации исключаются, эти соотношения могут использоваться и для оценки систематических и случайных погрешностей каналов ИИС и ИВК. (Конечно, если ИИС или ИВК эксплуатируется без проведения периодических проверок, то погрешности каналов могут быть и значительно больше нормированных при выпуске значений $\gamma_{\text{кл}}$, и эти соотношения будут неправомерны.)

Полагая при указанном ограничении, что для канала ИИС или ИВК систематическая погрешность без проведения коррекции

нуля и чувствительности составляла $\theta = 0,3\gamma_{\text{кл}}$, а случайная погрешность была $\gamma_{0,9\text{кл}} = 0,4\gamma_{\text{кл}}$, при проведении однократной коррекции нуля и чувствительности ограничение ГОСТ 8.207—76 в виде $\gamma_{0,9\bar{x}} < \theta/5$ принимает вид $0,4\gamma_{\text{кл}}/\sqrt{n} < 0,77\sigma/5 = 0,77 \frac{0,4}{1,6} \gamma_{\text{кл}}/5$ или $n > \left(\frac{5 \cdot 0,4 \cdot 1,6}{0,77 \cdot 0,4} \right)^2 \approx 100$, т. е. введение коррекции по однократно определенной (а следовательно, случайной) поправке позволяет уже перейти от усреднения 45 отсчетов к усреднению 100 отсчетов и понизить тем самым погрешность получаемых результатов с $\gamma_1 = 0,7\gamma_{\text{кл}}$ до $\gamma_{100} = \theta_1/5$, где $\theta_1 = 0,77\sigma = 0,77 \cdot 0,4\gamma_{\text{кл}}/1,6 = 0,2\gamma_{\text{кл}}$, т. е. $(0,7 \cdot 5)/0,2 \approx 18$ раз.

Дальнейшее понижение погрешности усредненного результата возможно лишь тогда, когда для коррекции нуля и чувствительности будут использоваться не случайные значения поправок, получаемых при однократных поверках по образцовым мерам или сигналам, а их усредненные значения по ряду последовательных поверок. Но в этом случае ИВК должен располагать дополнительной емкостью памяти для хранения полученных поправок, прежде чем они будут усреднены и использованы для введения усредненной поправки, что создает дополнительные трудности.

В предельном случае поверка канала может производиться перед каждым измерением. Тогда вводимая поправка будет усредняться по такому же числу n поверок, что и усредняемый результат многократных отсчетов. Такой режим коррекции удобен тем, что не требует дополнительной емкости памяти для хранения поправок. В результат каждого отдельного отсчета входит своя случайная (из области $\pm \theta\sigma$, см. рис. 4-9) поправка, а при усреднении этих отсчетов одновременно усредняются как результат измерения, так и вводимые поправки. При этом совокупность всех конкретных реализаций поправок образует центрированную случайную величину с нулевым средним и законом распределения, повторяющим закон распределения случайной погрешности данного прибора или канала. Именно поэтому результирующая дисперсия отдельных отсчетов удваивается, а закон их распределения представляет собой композицию двух одинаковых законов распределения случайной погрешности канала или прибора.

Эксцесс образующейся при этом композиции может быть рассчитан по формуле (3-5), но так как в этом случае относительный вес дисперсии каждого из слагаемых $p = 1/2$, то соотношение (3-5) приводится к виду $\epsilon_2 = \epsilon \cdot 1/4 + 6 \cdot 1/4 + \epsilon \cdot 1/4 = \epsilon/2 + 1,5$, где ϵ — эксцесс распределения случайной погрешности канала или прибора. Поэтому, например, если распределение случайной погрешности подчиняется распределению Симпсона ($\epsilon = 2,4$), то композиция будет иметь эксцесс $\epsilon = 2,4/2 + 1,5 = 2,7$, т. е. более близкий к нормальному, но еще далеко не равный 3.

Достигаемый при этом эффект снижения погрешности усредненного результата может быть весьма значительным. Так, например, если без усреднения и коррекции погрешность результатов состав-

ляла $0,7\gamma_{\text{нл}}$ ($\theta = 0,3\gamma_{\text{нл}}$ и $t\sigma = 0,4\gamma_{\text{нл}}$), то при проведении коррекции нуля и чувствительности перед каждым отсчетом систематическая погрешность ($\theta = 0,3\gamma_{\text{нл}}$) будет устранена почти полностью, но суммарная дисперсия отсчетов удвоится, т. е. σ_x возрастет в $\sqrt{2}$ раз, и при усреднении, например, $n = 1000$ отсчетов погрешность с доверительной вероятностью $P_d = 0,9$ составит $\gamma_{1000} = 1,6 \cdot \sqrt{2} \cdot \sigma / \sqrt{1000} \approx 0,018\gamma_{\text{нл}}$, т. е. будет уменьшена в $0,7/0,018 = 40$ раз.

Предел возможного уменьшения погрешности при одновременном введении и коррекции систематических погрешностей и статистическом усреднении определяется уже другими, не учитываемыми до сих пор причинами, как-то: погрешностью, используемой при поверках образцовой меры, рассмотренными в § 1-5, 3-5, 3-7 динамическими погрешностями и особенно отсутствием независимости, т. е. наличием хотя бы слабой коррелированности усредняемых погрешностей (см. § 8-8 и 8-9). Все эти причины должны в этом случае внимательно анализироваться (см. в § 8-8 пример такого анализа).

Некоторые выводы. 1. Суммирование случайных и систематических составляющих погрешности при оценке погрешности усредненного результата не только правомерно, но и необходимо.

2. Правила такого суммирования зависят от того, остается ли неисключенный остаток систематических погрешностей постоянным во всех усредняемых отсчетах.

3. Если знак неисключенной систематической погрешности нам неизвестен (см., например, рис. 1-10, в), но известно, что она остается постоянной во всех усредняемых отсчетах, то ее модуль должен суммироваться с доверительным значением случайной погрешности а р и ф м е т и ч е с к и. Суммирование ее со случайной погрешностью по правилам суммирования дисперсий неправомерно, хотя подобный некорректный метод суммирования с рекомендацией эвристических формул, существенно занижающих получаемые оценки, и включен в текст ГОСТ 8.207—76.

4. Если неисключенная систематическая погрешность в каждом из усредненных отсчетов принимает случайное значение, то она приобретает все свойства случайной величины и суммирование ее с другими случайными погрешностями должно производиться по правилам, описанным в гл. 3, для суммирования случайных погрешностей.

МЕТОДЫ ИДЕНТИФИКАЦИИ ФОРМЫ ЗАКОНА РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

5.1. ПОТРЕБНОСТЬ В ИДЕНТИФИКАЦИИ РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ

Пока было общепризнанным, что погрешности на основании центральной предельной теоремы всегда должны быть распределены нормально, задача определения формы распределения погрешностей, естественно, не ставилась. Однако сомнение во всеобщности нормального распределения существовало давно. Так, еще М. Ф. Маликов в книге [22, с. 97] приводил слова физика Липпмана: «Все верят в закон погрешностей, ибо экспериментаторы думают, что этот закон — математическая теорема, а математики считают, что он установлен экспериментальным путем».

Тем не менее заблуждение о правомерности лишь нормального распределения погрешностей продолжает существовать до сих пор. Для подтверждения его бесспорности, например, в [21] (учебное пособие для вузов, 1982 г.) приводится цитата из книги Ю. В. Линника [20]: «Как оказывается, в широком классе несмещенных оценок оценки, найденные по методу наименьших квадратов, могут быть совместно эффективными, лишь если вектор погрешностей Δ нормален. Таким образом, наличие оптимальных свойств у метода наименьших квадратов тесно связано с нормальностью вектора погрешностей». И на основании этого в книге [21] делается вывод, что при распределениях, отличных от нормального, использование методов статистической обработки (среднее арифметическое, метод наименьших квадратов и т. п.) неправомерно.

В § 4-3 мы уже обсуждали значение математического термина «эффективный», что означает «самый лучший по минимуму дисперсии». В этой связи представляет интерес рассмотреть результаты, приведенные в книге [26], по сравнительной эффективности метода наименьших квадратов (МНК) и метода наименьших модулей (МНМ). Эти данные сведены в табл. 5-1 и графически представлены в виде кривых на рис. 5-1 в функции от значений контрэксцесса κ . Из этих кривых видно, что МНК достаточно эффективен для широкого класса распределений (по крайней мере, с κ от 0,5 до 0,7), что аналогично обсуждаемому выше применительно к среднему арифметическому (§ 4-3). Поэтому цитата Ю. В. Линника означает лишь то, что для распределений с $\kappa \leq 0,5$ (или с $\kappa > 0,645$, см. рис. 4-2) могут быть использованы другие методы, которые будут несколько лучше, чем МНК и среднее арифметическое. Но в ней отнюдь не утверждается, что для всех распределений, отличных от

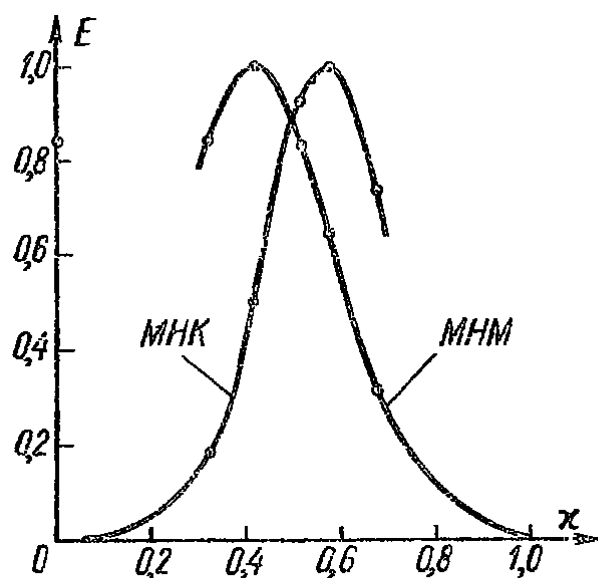


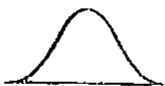
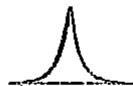
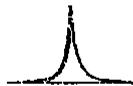
Рис. 5-1

нормального, нельзя использовать МНК или среднее арифметическое (конечно, кроме распределений с $\kappa = 0$, подобных распределениям, приведенным в табл. 3-3). Обоснованное решение подобных задач и требует идентификации формы распределения погрешностей.

Идентификация распределений нужна и потому, что рассеяние всех оценок $[\sigma(\sigma), \sigma(\kappa), \sigma(\epsilon), \sigma(k)]$, как было показано в § 4-4, зависит от формы закона распределения. Для обеспечения одной и той же погрешности в определении оценки при одном законе

можно ограничиться достаточно малой выборкой, тогда как при другом — выборка исходных данных должна быть значительно больше. Совершенно необходимо знание вида закона распределения при определении одних параметров закона распределения по другим его параметрам. Так, например, квантили, т. е. значения погрешности с заданной доверительной вероятностью, без знания вида закона распределения не могут быть выражены через с. к. о. И, наконец, изменение вида закона распределения погрешностей нередко может служить признаком какого-либо резкого изменения условий проведения измерений. Например, объединение нескольких серий наблюдений, в одной из которых возникла заметная систематическая погрешность, приводит к образованию двухмодального распределения, в то время как в обычных условиях оно было одномодальное.

Таблица 5-1

Вид рас- пределения						
α	4	2	1,5	1	0,75	Распре- деление Коши
κ	0,68	0,577	0,51	0,408	0,32	0
ϵ	2,16	3	3,85	6	10	∞
Относитель- ная эффек- тивность: МНК	0,73	1	0,91	0,5	0,175	0
	МНМ	0,31	0,64	0,83	1	0,83

Экспериментальные данные о разнообразии форм распределения погрешностей измерений в период 1965—1975 гг. накоплены в достаточном количестве. Так, в книге 1968 г. [32] приводились результаты исследования законов распределения погрешностей электромеханических приборов на кернах, электронных приборов для измерения температур и усилий, цифровых приборов с ручным уравниванием. Объем выборок экспериментальных данных для каждого экземпляра составлял 100—400 отсчетов. Только у одного из 47 распределений погрешностей оценки контрэксцесса и энтропийного коэффициента оказались близкими к нормальному ($k = 2,05$ и $\kappa = 0,58$). Все же остальные распределения отличались от нормального, имея энтропийный коэффициент $k < 2,02$. Оценки контрэксцесса κ исследованных распределений погрешностей приборов на кернах в основном оставались в пределах от 0,57 до 0,73, т. е. от нормального до равномерного. Погрешности же термоакустических термометров, динамометров с датчиками ТПА, приборов с цифровым отсчетом (ОКВ-2, КЛ-48, Р306) имели в основном распределения с контрэксцессом от 0,25 до 0,54, т. е. от острых экспоненциальных до нормального.

В работе 1971 г. [7] В. Я. Галочкиной (Дальневосточный политехнический ин-т) была подробно исследована форма распределения погрешностей у 25 экземпляров цифровых вольтметров типа Щ-1411 в 10 точках диапазона измерений. Все 250 распределений оказались двухмодальными, имеющими $\kappa \approx 0,68$ и $k \approx 1,65$, т. е. по своим параметрам очень далекими от нормального. В 1971 г. З. Таушанов (г. Варна, НРБ) исследовал законы распределений погрешностей магнитоэлектрических приборов НРБ, ЧССР и ГДР. Полученные им 34 распределения были сходны между собой, но существенно отличны от нормального. Значение энтропийного коэффициента этих распределений колебалось от 1,79 до 1,97, и автор рекомендовал рассматривать их как трапецеидальные. В 1973 г. сотрудником ВНИИМ им. Д. И. Менделеева К. А. Резником была исследована форма распределений погрешностей средств измерений 16 типов (по 100—200 экземпляров каждого типа). Гипотеза о нормальном распределении для половины приборов оказалась неприемлемой. Эксцесс исследованных распределений находился в пределах от 1,8 до 6, т. е. от равномерного до распределения Лапласа. В работе [3] И. У. Алексеевой были сопоставлены 219 распределений погрешностей, показано их широкое разнообразие и дана их топографическая классификация по значениям k и κ .

В итоге работ этих лет факт разнообразия законов распределения погрешностей был признан законодательно — с 1 января 1974 г. был введен в действие ГОСТ 8.011—72, устанавливающий, что при сообщении размера погрешности результата измерения целесообразно указывать вид распределения, и были стандартизованы модели равномерного, трапецеидального, треугольного, нормального и двухмодальных распределений.

Возможность идентификации формы распределения экспериментальных данных ограничена прежде всего малостью объема выборки. При большом объеме выборки, например в несколько тысяч наблюдений, построение гистограммы часто позволяет получить достаточно плавную кривую, отражающую все характерные особенности наблюдаемого закона. При малом объеме выборки ($n = 20 \div 200$) особенности распределения оказываются замаскированными случайностью самой выборки. Таким образом, малость выборки укрупняет возможные классы идентификации и сокращает их число, а следовательно, более тонкая идентификация формы распределения возможна лишь при соответствующем увеличении объема выборки экспериментальных данных.

Следовательно, приобретает большое значение накопление данных и составление каталога распределений погрешностей различных средств измерений (датчиков, приборов), различных широко применяемых методов измерений, с тем чтобы этими данными можно было затем пользоваться и при наличии малых серий измерений.

5-2. ОПТИМАЛЬНОЕ ЧИСЛО ИНТЕРВАЛОВ ГРУППИРОВАНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Для определения оценок математического ожидания, с. к. о., эксцесса или контрэксцесса не требуется какого-либо группирования экспериментальных данных. Эти оценки могут быть найдены непосредственно по исходной неупорядоченной выборке.

Для определения медианы, сгибов, других квантилей, использования критерия согласия Колмогорова—Смирнова или для обнаружения промахов экспериментальные данные необходимо расположить в порядке возрастания, т. е. построить вариационный ряд (упорядоченную выборку).

Для определения формы распределения, для использования критериев согласия Пирсона и др., для сопоставления гипотез о форме распределения или для вычисления оценок энтропии и энтропийного коэффициента простого упорядочения выборки уже недостаточно, а выборка должна быть представлена в виде гистограммы, состоящей из m столбцов с определенной протяженностью d соответствующих им интервалов.

Общепринято делать эти интервалы одинаковыми. Так, М. Кендалл и А. Стьюарт в [15] неоднократно подчеркивают, «что всякий раз, когда это возможно, класс-интервалы должны быть одинаковыми» [15, с. 19] или «настолько, насколько возможно, интервалы должны быть одинаковыми, с тем чтобы количества попаданий в различные интервалы были сравнимы» [15, с. 17].

А. Хальд в своей книге [45] затрачивает несколько страниц на то, чтобы показать, что существует оптимальное число интервалов группирования, когда ступенчатая огибающая гистограммы наиболее близка к плавной кривой распределения гене-

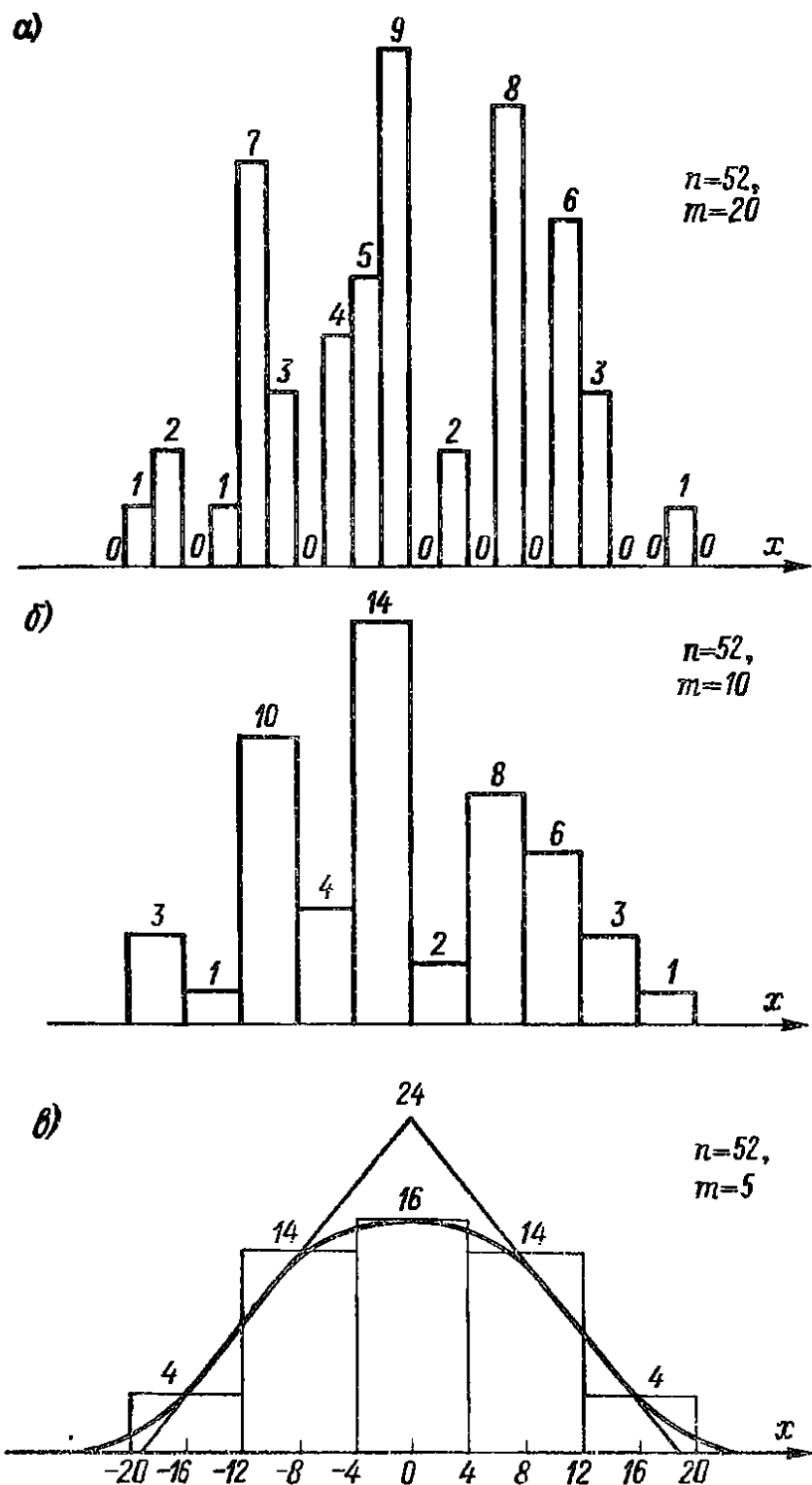


Рис. 5-2

ральной совокупности. Можно сформулировать ряд аналитических критериев такой близости, используя показатели в виде энтропийного коэффициента, эксцесса, критерия χ^2 и т. д. Различные критерии могут давать несколько различающиеся значения оптимального числа m интервалов группирования. Однако сам факт существования оптимума не зависит от выбора критерия близости, так как при группировании данных в слишком большое число мелких интервалов некоторые из них окажутся пустыми или мало заполненными. Гистограмма будет отличаться от плавной кривой распределения вследствие изрезанности многими всплесками и провалами, т. е. будет иметь «гребенчатый» вид (рис. 5-2, а и б).

При слишком малом числе m интервалов гистограмма будет отличаться от действительной кривой распределения вследствие слишком крупной ступенчатости, из-за чего характерные особенности будут просто потеряны. Например, если интервал группирования брать равным размаху экспериментальных данных, то любое распределение будет сведено к равномерному, а если делить размах на три равных интервала, то любое колоколообразное распределение будет сведено к треугольному. Этих чисто качественных оценок вполне достаточно для доказательства самого факта существования оптимального числа m интервалов, обеспечивающего наибольшую близость ступенчатой гистограммы к плавной кривой распределения.

Исходя из предположения, что генеральная совокупность, из которой взята исследуемая выборка, имеет гладкую кривую распределения, естественно считать, что появляющиеся при группировании провалы и выбросы являются случайным «шумом», порождаемым случайностью попадания тех или иных значений x_i в малую выборку. Укрупнение интервалов группирования является методом фильтрации этого случайного «шума». Однако при слишком протяженных интервалах начинает «фильтроваться» уже не «шум», а сам «сигнал», т. е. начинают сглаживаться особенности искомого закона распределения.

Таким образом, задача выбора оптимального числа интервалов при построении гистограммы — это задача *оптимальной фильтрации*, а оптимальным числом m интервалов является такое, когда максимальное возможное сглаживание случайных флуктуаций данных сочетается с минимальным искажением от сглаживания самой кривой искомого распределения.

Для распределений погрешностей одним из практических признаков приближения к оптимуму может служить исчезновение в гистограмме провалов и близким к оптимальному может считаться наибольшее m , при котором гистограмма еще сохраняет плавный характер. Иногда это требование выражают по-другому, рекомендуя, чтобы столбцы имели такую ширину, при которой в наименьшем столбце гистограммы содержалось бы не менее 10 наблюдений. Эту рекомендацию можно выполнить лишь при $n > 200$, поэтому В. Кокрен [55] замечает, что на практике в крайних столбцах гистограммы можно допустить и менее 5 наблюдений.

В литературе по статистической обработке экспериментальных данных приводятся и более конкретные рекомендации относительно выбора числа m интервалов группирования, которые, однако, существенно различаются между собой. Эти рекомендации можно разделить на две группы: рекомендации, приводимые без использования каких-либо формальных критериев, и рекомендации, получаемые на основе использования различных критериев близости между ступенчатой гистограммой и плавной кривой распределения.

К первой группе следует отнести рекомендации, приводимые

в подавляющем большинстве пособий по математической статистике (например, [51, с. 92]), где говорится, что группировать данные следует так, чтобы получилось не меньше 6 и не более 20 интервалов. Иногда эта рекомендация дается еще более категорично; так, в работе [25, с. 20] сказано: «число рядов обычно принимается равным 12, причем допускаются отклонения от этого числа на 2—3 единицы в ту или другую сторону». Во многих, как классических [60], так и изданных в последние годы [6, 35], руководствах по статистике для определения оптимального числа интервалов рекомендуется формула Старджеса:

$$m = \log_2 n + 1 = 3,3 \lg n + 1. \quad (5-1)$$

Р. Шторм [51] и другие авторы рекомендуют для определения оптимального числа интервалов формулу Брукса и Каррузера (без ссылки на первоисточник) в виде

$$m = 5 \lg n. \quad (5-2)$$

В книге 1964 г. И. Хайнхольд и К. Гаеде [57] рекомендуют для той же цели соотношение

$$m = \sqrt{n}. \quad (5-3)$$

В изданной ВНИИМ им. Д. И. Менделеева в 1972 г. «Рекомендации по методам обработки результатов наблюдений», имеющей силу директивного документа, также без каких-либо обоснований приводится таблица, согласно которой, в зависимости от объема выборки от 40 до 10 000, число m назначается от 7 до 22. Для сопоставления между собой перечисленных рекомендаций на рис. 5-3 представлены зависимости, соответствующие рекомендации Старджеса (прямая 1), Брукса (прямая 2) и Хайнхольда (кривая 3). Рекомендация ВНИИМ показана заштрихованными прямоугольниками. В области значений $n \approx 100$ эти рекомендации близки между собой, однако при n , равном 1000 или 10 000, — они существенно различны.

При этом следует отметить, что хотя посылки, на основании которых получены эти рекомендации, не сообщаются, почти все авторы, приводя их, указывают, что они определяют именно «оптимальное» значение m , а следовательно, отклонение от него влечет за собой уменьшение эффективности обработки данных. В этом отношении показательна «кочующая» из книги в книгу формула Старджеса. Если же обратиться к первоисточнику, а это заметка

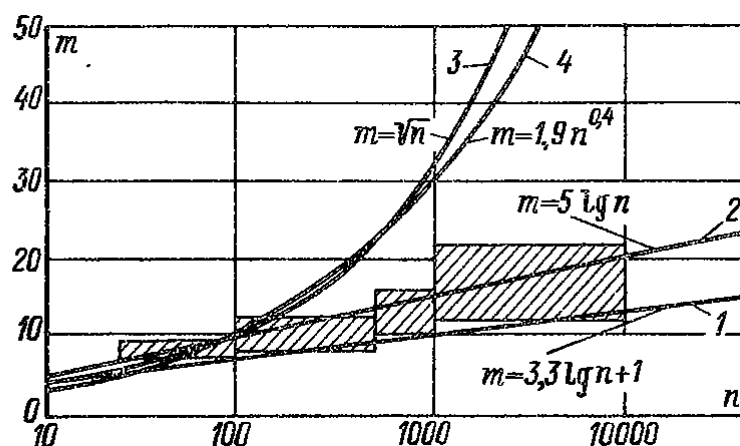


Рис. 5-3

[61] на полстраницы в журнале Американской ассоциации статистиков за 1926 год, то оказывается, что автор выдвигал чисто эвристическое предположение. Старджес отмечает, что свойством $m = \log_2 n + 1$ обладает число m биномиальных коэффициентов, в то время как их сумма равна n . Например, пишет он, $1 + 4 + 6 + 4 + 1 = 16$, поэтому вариационный ряд из 16 членов наиболее удобно разбить на 5 интервалов, а из 64 членов — соответственно на 7 интервалов и т. д. На этом «основании» автор предполагает, что таким путем устанавливается «оптимальный интервал группирования для расчета среднего арифметического, дисперсии, асимметрии и т. д.».

Ко второй группе рекомендаций прежде всего относятся фундаментальные теоретические исследования по использованию критерия согласия χ^2 . Как известно, этот критерий не может быть использован без разбиения выборки на интервалы, в которых производится вычисление частных разностей между принятой моделью и сравниваемой выборкой. Однако применение критерия χ^2 для интервалов постоянной длины d , используемых обычно для построения гистограммы, неэффективно. Поэтому исходной посылкой всех работ по эффективности критерия χ^2 является рассмотрение не интервалов с равной длиной, а интервалов с *равной вероятностью* в соответствии с принимаемой моделью. Следует заметить, что число m интервалов равной длины и число K интервалов равной вероятности различаются в большое число раз. Так, например, если исследуется выборка с $n = 600$, а в качестве модели принимается нормальное распределение, то для того чтобы в крайние интервалы попадало приблизительно по 10 наблюдений, необходимо взять $m = 7$. В центральные столбцы гистограммы при этом попадет более чем по 100 наблюдений. Число же равновероятных интервалов с $n_j = 10$ при $n = 600$, естественно, $K = 60$.

В 1942 г. Г. Манном и А. Вальдом [58] было установлено, что при $n \rightarrow \infty$ оптимальное число K равновероятных интервалов при использовании критерия χ^2 имеет порядок

$$K \sim 4 \sqrt[5]{2} (n/t)^{0.4}, \quad (5-4)$$

где t — безразмерная квантиль нормального распределения, соответствующая заданной вероятности $P = 1 - q$, где q — принятый уровень значимости.

В 1950 г. К. Уильямс в работе [63] показал, что в формуле (5-4) коэффициент 4 можно заменить на 2 без какой-либо потери эффективности, так как оптимум очень пологий, и дал конкретную таблицу значений K , которая приводится, например, в [4, с. 163]. Соотношение (5-4) варьируется различными авторами. Так, в работе [15] оно усложняется до

$$K \leq b [\sqrt[5]{2} (n-1)/(t_1 + t_2)]^{0.4}, \quad (5-5)$$

где t_1 и t_2 — некоторые задаваемые квантили, а b — принимается между 2 и 4. В работе [56] оно упрощается до $K = 4 [0,75 (n -$

— 1)^а 10⁻², что получается из (5-5) при $b = 4$ и $t_1 + t_2 = 1,65$. Если же в (5-4) в соответствии с [63] коэффициент 4 заменить на 2 и принять $t = 1,65$, что соответствует $P = 0,9$ и $q = 0,1$, то оно получит вид

$$K = 1,9n^{0.4}. \quad (5-6)$$

Это соотношение представлено на рис. 5-3 кривой 4, близкой к кривой 3, соответствующей рекомендации (5-3) Хайнхольда и Гаеде.

Относительно этих рекомендаций необходимо также заметить, что их «миграция» из книги в книгу приводит к тому, что забываются те условия, которые были положены первоначально в их основу. Так, Манн и Вальд четко указывали, что их соотношение (5-4) соответствует наибольшей эффективности критерия χ^2 при использовании равновероятных интервалов. Тем не менее в [4] оно «для простоты» используется Бендатом и Пирсолом для построения гистограммы с интервалами постоянной длины. Так как $K \gg m$, то огибающая гистограммы в приводимом авторами примере [4, с. 165] «оптимальной» обработки данных при $n = 200$ имеет несглаженные провалы и всплески, т. е. принятое $m = 16$ существенно больше оптимального.

К этой же группе рекомендаций относятся работы 1973—1981 годов, основанные на использовании критерия близости в виде энтропийного коэффициента k_e . Понятие энтропийного коэффициента как числовой характеристики формы распределения было предложено в работе [30]. По гистограмме эта оценка вычисляется [52] как

$$k_e = \frac{dn}{2\sigma} 10^{-\frac{1}{n} \sum_{j=1}^m n_j \lg n_j}, \quad (5-7)$$

где d — ширина столбца гистограммы; n — объем выборки; σ — с. к. о.; m — число столбцов гистограммы; n_j — число наблюдений в j -м столбце ($j = \overline{1, m}$).

Исследования состояли в генерировании выборок разного объема из одной и той же генеральной совокупности, для чего с одного или близких объектов снимались, например, 40 серий по 25 или 100 наблюдений. Объединяя их между собой, получали представительную выборку с $n = 1000 \div 4000$ наблюдений. Такая выборка дает плавную гистограмму и при достаточно большом числе интервалов группирования. При этом особенно важно то, что изменение числа интервалов при такой выборке не меняет общего вида гистограммы и оценки энтропийного коэффициента k_e , найденные при разных значениях m , равных, например, 19, 23 и 27, достаточно близко совпадают между собой. Эта оценка k_e и принималась за значение k_e «генеральной совокупности». Затем определялись оценки k_e для малых выборок при разных значе-

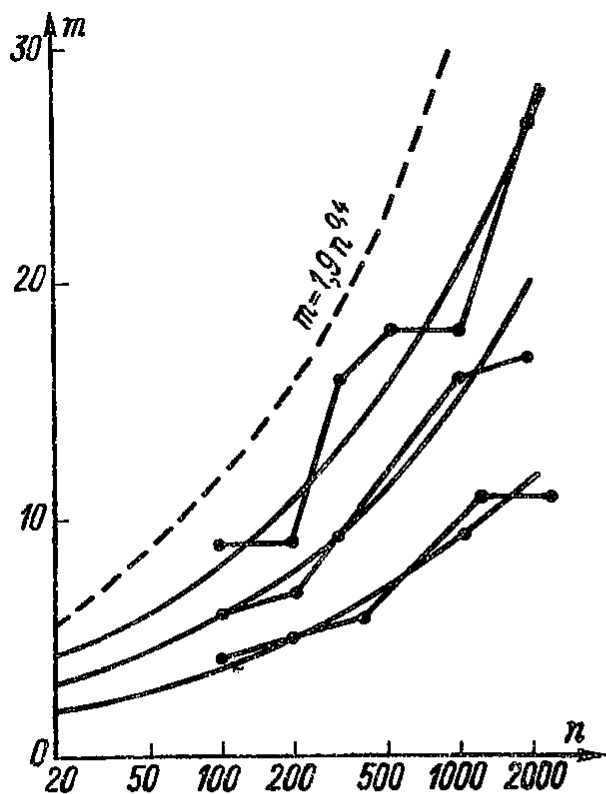


Рис. 5-4

ниях m . При больших m в гистограммах появлялись провалы и оценка k_0 резко падала. При уменьшении m оценка k_0 возрастала. При некотором промежуточном m оценка k_0 равна найденной для данной «генеральной совокупности» и это значение m принималось за оптимальное.

Результаты первых из этих работ, выполненных коллективом болгарских авторов (ВМЭИ, Варна) под руководством З. Таушанова, были опубликованы в 1973 г. [41], где излагался метод и приводился полученный результат в виде

$$m = 4 \lg n, \quad (5-8)$$

что очень близко к рекомендации (5-2) Брукса и Каррузера. Эта

работа была продолжена Е. Тоневой [42]. Исследование 72 разных вариантов выборок из одних и тех же экспериментальных статистик, где n варьировалось от 25 до 500 наблюдений, привело к соотношению

$$m = 5 \lg n - 5 = 5 \lg (n/10). \quad (5-9)$$

С помощью описанного метода в работе [3] И. У. Алексеевой (Башкирия, Уфимский авиационный институт) на основе других экспериментальных статистик были найдены зависимости оптимального m для трех существенно различных распределений — близкого к трапецеидальному ($\kappa = 0,7$, $\varepsilon = 2$), близкого к нормальному ($\kappa = 0,53$, $\varepsilon = 3,56$) и близкого к распределению Лапласа ($\kappa = 0,4$, $\varepsilon = 6,25$). Выборки содержали по 2000—2400 наблюдений и делились на серии с n от 100 до 1000—1200 наблюдений. Полученные зависимости оптимального m от n и $1/\kappa$ были аппроксимированы аналогично выражениям (5-1), (5-2), (5-8), (5-9) веером прямых линий, в результате чего для оптимального m было получено соотношение

$$m = \frac{4}{\kappa} \lg \frac{n}{10}. \quad (5-10)$$

Это соотношение открывает совершенно новый подход к рассматриваемому вопросу. Все исследователи, начиная со Старджеса и кончая Тоневой, рассматривали m как функцию только объема выборки n и расходились между собой лишь в оценке вида этой функции. Никто из них не предполагал, что m зависит от вида закона распределения. Из формулы же (5-10) следует, что опти-

мальное число столбцов m существенно зависит от значений контрэксцесса κ или эксцесса ε . Полученные в работе [3] (с использованием критерия близости в виде энтропийного коэффициента) оптимальные значения числа интервалов m для законов распределения с оценками контрэксцесса κ , равными 0,4, 0,53 и 0,7, и разными объемами выборки после усреднения по всем исследованным распределениям нанесены на график рис. 5-4 и приведены ниже:

n	100	200	300	400	500	600	1000	1200	2000	2400
κ , равная:										
0,7	4	5	—	6	—	8	—	11	—	11
0,53	6	7	9	—	12	—	16	—	17	—
0,4	9	9	16	—	18	—	18	—	27	—

Поскольку характер ломаных линий на рис. 5-4 более близок к штриховой кривой $m = 1,9n^{0.4}$ (5-6), чем к вееру прямых линий, аппроксимируем его функцией вида $m = A(\varepsilon) n^\alpha$, где $A(\varepsilon)$ — некоторая функция от ε , а α — показатель степени при объеме выборки n . В координатах $\lg m = f(\lg n)$ приведенные данные близки к параллельным прямым с угловым коэффициентом $\alpha = 0,4$. Таким образом, найденный Манном и Вальдом показатель степени при n для равновероятных интервалов и использовании критерия близости в виде χ^2 остается также равным 0,4 и для равновеликих интервалов и критерия близости в виде энтропийного коэффициента. Однако при равновероятных интервалах зависимость их числа от формы кривой распределения автоматически устранялась тем, что интервалы в этом случае располагались строго пропорционально изменению плотности модели. При равновеликих интервалах функция $A(\varepsilon)$ должна быть введена в явном виде. Полученный И. У. Алексеевой сдвиг кривых $m = f(n)$ в зависимости от эксцесса ε распределений может быть аппроксимирован выражениями вида $A(\varepsilon) = (\varepsilon + 1,5)/6$ или $A(\varepsilon) = \varepsilon^{0.8}/3$, что приводит к соотношениям для расчета оптимального значения

$$m = \frac{\varepsilon + 1,5}{6} n^{0.4} \quad (5-11)$$

или

$$m = \frac{1}{3} \sqrt[5]{\varepsilon^4 n^3} = \frac{1}{3} \sqrt[5]{\frac{n^3}{\kappa^8}}. \quad (5-12)$$

Формула (5-11) дает несколько лучшее приближение к данным рис. 5-4, а формула (5-12) более удобна для использования в аналитических выражениях и наглядно показывает, что оптимальное число m интервалов существенно сильнее зависит от формы распределения (ε^4 или κ^8), чем от объема выборки (n^3).

Трудность практического использования соотношений (5-10), (5-11) или (5-12) состоит в том, что число интервалов группирования m нужно выбрать прежде, чем будут найдены оценки $X_{ц}$.

σ , μ_4 и ε . Обойти эту трудность можно, воспользовавшись тем, что наиболее часто при обработке погрешностей встречаются распределения с ε от 1,8 до 6, т. е. от равномерного до распределения Лапласа. Для этих граничных значений из формулы (5-11) получаем

$$m_{\min} = 0,55n^{0.4} \text{ и } m_{\max} = 1,25n^{0.4} \quad (5-13)$$

и искомое значение m может быть выбрано близким к этому интервалу. При этом предпочтительно выбирать m нечетным, так как при четном m и островершинном или двухмодальном симметричном распределении в центре гистограммы оказываются два равных по высоте столбца и середина кривой распределения принудительно уплощается.

В итоге проведенного сопоставления различных предложений по выбору числа интервалов группирования экспериментальных данных можно сформулировать следующие выводы.

1. Задача выбора числа интервалов группирования экспериментальных данных при их статистической обработке есть задача оптимальной фильтрации случайных отклонений гистограммы от плавной кривой плотности распределения, соответствующей генеральной совокупности.

2. С этой точки зрения, безусловно, существует оптимальное число m интервалов группирования, обеспечивающее при данной форме закона распределения и данном объеме выборки наибольшую близость ступенчатой гистограммы (или ее полигона) к плавной кривой плотности распределения генеральной совокупности.

3. Использование интервалов неравной длины, а равной вероятности, для которых существует рекомендация (5-4) Манна и Вальда, позволяет автоматически учитывать форму закона распределения, но на практике это неудобно, чем и объясняются приведенные выше настоячивые рекомендации Кендалла и Стьюарта о предпочтительности равновеликих интервалов.

4. При использовании интервалов постоянной длины их оптимальное число сильнее зависит от эксцесса распределения ε , чем от объема выборки n .

5. Зависимость m от объема выборки, как теоретически [см. (5-4)], так и экспериментально [см. (5-11)], имеет вид $m = An^{0.4}$ и в дальнейшем уточнении, по-видимому, не нуждается. Что же касается аппроксимации зависимости m от ε в виде $A(\varepsilon) = (\varepsilon + 1,5)/6$ [см. выражение (5-11)] или $A(\varepsilon) = \varepsilon^{0.8}/3$ [см. (5-12)], то, естественно, желательно их аналитическое уточнение.

6. Для практического назначения числа интервалов группирования вполне достаточно определение его оптимального интервала по соотношениям (5-13), тем более, что при назначении его нечетным выбор возможных его значений оказывается крайне ограниченным.

7. При практическом выборе интервалов группирования полезно иметь в виду следующее замечание М. Кендалла, А. Стью-

арта [15, с. 20]: «Несмотря на важность равенства класс-интервалов, бывает полезно рассмотреть более мелкие класс-интервалы на тех участках, где частоты меняются очень быстро». Так, при островершинном распределении центральный столбец шириной d при построении полигона полезно рассматривать состоящим из двух столбцов шириной $d/2$, а самые крайние столбцы, в развитие рекомендации В. Кокрена, при необходимости можно делать шириной $2d$.

8. В тех случаях, когда гистограмма или полигон распределения погрешностей оказываются явно двухмодальными, число столбцов гистограммы может быть увеличено в 1,5—2 раза, чтобы на каждый из двух «колоколов» приходилось примерно по m интервалов.

5-3. ПРИМЕРЫ ПОСТРОЕНИЯ ГИСТОГРАММ И ПОЛИГОНОВ РАСПРЕДЕЛЕНИЙ

Главным фактором, затрудняющим идентификацию формы кривой распределения экспериментальных данных при относительно малой выборке, является случайность (т. е. неповторимость от выборки к выборке) появления различных значений случайной величины. Надежным путем преодоления этого разброса является увеличение объема экспериментальных данных. Однако это сопряжено с резким ростом затрат на проведение измерений, а часто невозможно по самой сути эксперимента. Поэтому все усилия экспериментаторов направлены на то, чтобы отгадать форму кривой распределения генеральной совокупности, имея из нее лишь слабopедставительную малую случайную выборку. При этом опираются на максимальное использование априорной информации о виде распределения погрешности, заключающейся в том, что кривая плотности распределения предполагается п л а в н о й и с и м м е т р и ч н о й. Аксиома плавности устанавливается исходя из того, что сама измеряемая величина является непрерывной. Аксиома симметрии базируется на относительной малости размера погрешностей. Поэтому если функция преобразования прибора даже существенно нелинейна, то на малом ее участке, соответствующем размаху погрешности, изменение ее крутизны столь незначительно, что не может вызвать заметной скошенности кривой распределения. Исходя из этого появление скошенности или чередование всплесков и провалов эмпирического распределения погрешности приписывается исключительно случайности малой выборки, в то время как распределение генеральной совокупности считается плавным и симметричным. Эти аксиомы подтверждаются практикой: при увеличении объемов выборки до тысяч и десятков тысяч наблюдений эмпирические распределения погрешностей чаще всего становятся плавными, симметричными и хорошо воспроизводимыми.

Методику построения гистограммы распределения экспери-

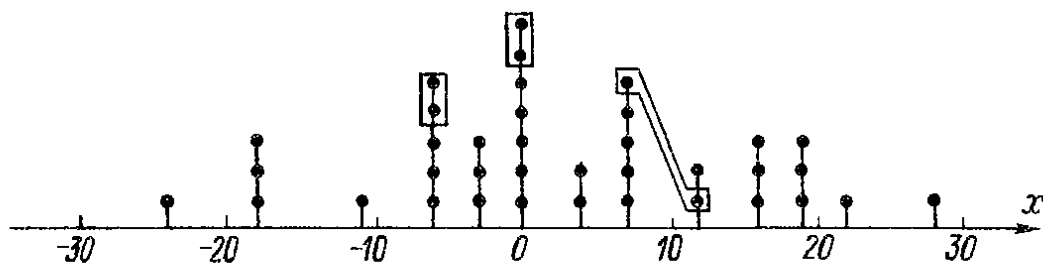


Рис. 5-5

ментальных данных можно проиллюстрировать конкретным примером. Пусть в результате эксперимента было зарегистрировано $n = 38$ отсчетов погрешностей x_i . Прежде всего их нужно расставить в порядке возрастания, т. е. образовать из них *вариационный ряд*. Проще всего это сделать путем нанесения этих значений вдоль заранее начерченной оси x , отмечая каждое из них точкой (рис. 5-5). Такое построение называется *линейчатой диаграммой* вариационного ряда. После нанесения всех наблюдавшихся значений на линейчатую диаграмму по ней легко составить таблицу вариационного ряда:

x_i	-24	-18	-11	-6	-3	0	+4
n_i	1	3	1	5	3	8	2
x_i	+7	+12	+16	+19	+22		+28
n_i	5	2	3	3	1		1

Определение координаты центра распределения. Для этого предварительно на линейчатой диаграмме полезно отметить положение точек или промежутков между ними, соответствующих медиане и сгибам распределения. Точки, между которыми располагаются эти оценки, обведены на рис. 5-5 прямоугольными рамками. По этим данным находятся значения пяти оценок центра распределения, описанных в § 4-3.

Оценка координаты центра в виде медианы, т. е. 50%-ной квантили, при $n = 38$ равна средней абсциссе 19-й и 20-й точек вариационного ряда. В данной выборке обе эти точки лежат в одном и том же столбце линейчатой диаграммы при $x_i = 0$, поэтому получаем $X_m = 0$.

Оценки сгибов, т. е. 25%- и 75%-ной квантилей (средняя абсцисса 9-й и 10-й точек и, соответственно, 28-й и 29-й точек), в рассматриваемом примере равны: $X_{0,25} = -6$ и $X_{0,75} = (7 + 12)/2 = +9,5$. Отсюда оценка центра сгибов (округленно, в соответствии с § 1-4) $X_c = (-6 + 9,5)/2 = +1,75 \approx +1,8$. Оценка центра размаха равна полусумме крайних точек вариационного ряда и составляет $X_p = (-24 + 28)/2 = +2,0$. Среднее арифметическое выборки $\bar{X} = \sum x_i n_i / n = +0,125 \approx +0,1$. Для определения среднего арифметического 50% центральных наблюдений сумма должна быть рассчитана только для точек, находящихся между сгибами, и поделена на $n/2$. Это дает $\bar{X}_{0,5} = +2,5$.

Расстановка полученных оценок центра в порядке возрастания дает вариационный ряд: 0; +0,1; +1,8; +2,0; +2,5. Медиану

этого ряда рекомендуется принять за оценку координаты центра распределения $X_{\text{ц}} = +1,8$, как достаточно эффективную инвариантно к форме распределения и полностью защищенную от влияния промахов (см. § 4-3). После того как будет определен вид закона распределения, для центра может быть выбрана та из оценок, которая более эффективна в соответствии с графиком рис. 4-2, но при предварительной обработке это невозможно.

Интервал неопределенности оценки центра. При ручном расчете оценка центра должна быть сразу же округлена в соответствии с ее абсолютной погрешностью (см. § 1-4). Но для такого округления необходимо знать хотя бы весьма приближенную оценку σ_{xi} с. к. о. анализируемых отсчетов. Для этого достаточно воспользоваться следующими простыми соотношениями. Для большинства законов распределения погрешностей в границах $\pm \Delta_{0,9} = \pm 1,6\sigma_{xi}$ содержится в среднем $P = 0,9$ всех отсчетов. Найти эти границы на линейчатой диаграмме рис. 5-5 можно, воспользовавшись соотношением (§ 2-2) в виде $P = (n - 2n_{\text{отб}})/(n + 1)$. Отсюда при $P = 0,9$ число точек, находящихся за границами интервала $\pm 1,6\sigma_{xi}$, $2n_{\text{отб}} = n/10 - 0,9 \approx n/10 - 1$. В нашем случае это дает общее число отбрасываемых с обеих сторон точек $2n_{\text{отб}} = 38/10 - 1 = 2,8 \approx 3$. Без трех наиболее удаленных от центра точек границы оставшихся на рис. 5-5 точек есть -18 и $+19$. Следовательно, размах оставшихся точек $R_{0,9} = 2\Delta_{0,9} = 18 + 19 = 37$ и $\sigma_{xi} \approx \Delta_{0,9}/1,6 = R_{0,9}/3,2 = 37/3,2 = 11,6$ и $\sigma_x \approx \sigma_{xi}/\sqrt{n} = 11,6/\sqrt{38} = 1,88$, а $\Delta_{0,9x} \approx 1,6\sigma_x = 1,6 \times 1,88 = 3$. Отсюда интервал неопределенности оценки центра с вероятностью 0,9 ограничен значениями $\pm \Delta_{0,9x} = \pm 3$. В нашем примере все пять полученных выше оценок центра лежат внутри этого интервала и речь может идти не об их уточнении, а лишь об округлении центральной из них до целого значения, т. е. $X_{\text{ц}} = +1,8 \approx +2$.

Построение гистограммы. Для наименьшего искажения кривой плотности в области центра распределения число столбцов m , как отмечалось в § 5-2, следует принимать нечетным, располагая центральный столбец симметрично относительно принятого центра распределения. Определяя предельные значения m_{min} и m_{max} согласно (5-13), имеем $m_{\text{min}} = 0,55n^{0,4} = 0,55 \cdot 38^{0,4} = 2,35$ и $m_{\text{max}} = 1,25n^{0,4} = 1,25 \cdot 38^{0,4} = 5,36$. Отсюда, учитывая, что $m = 3$ не дает информации о форме распределения, принимаем $m = 5$.

Так как крайние точки экспериментальных данных могут располагаться несимметрично относительно центра, то ширина d столбца гистограммы определяется по отклонению от центра ΔX_m наиболее удаленной точки: $d = 2\Delta X_m/m = 2(28 - 2)/5 = 10,4$. Округлять значение интервала группирования d нужно всегда в большую сторону, иначе крайняя точка окажется за пределами крайнего столбца. При этом удобно, чтобы d легко делилось на 2 (для определения координат центров столбцов). Отсюда

Таблица 5-2

x_i	n_i	$x_i - 2$	$(x_i - 2)_{гр}$	x_j	n_j	$\bar{n}_j$
—	—	—	—30	—34	0	0
—24	1	—26		—22	4	3
—18	3	—20	—18	—10	6	7
—11	1	—13		—	—	—
—6	5	—8	—6	—	—	—
—3	3	—5		—	—	—
0	8	—2	+6	+2	18	18
+4	2	+2		—	—	—
+7	5	+5	+6	—	—	—
—	—	—		—	—	—
+12	2	+10	+18	+14	8	7
+16	3	+14		—	—	—
+19	3	+17	+18	—	—	—
—	—	—		—	—	—
+22	1	+20	+30	+26	2	3
+28	1	+26		—	—	—
—	—	—	—	+36	0	0

принимая $d = 12$. Тогда центры столбцов получают значения $-22, -10, +2, +14, +26$, а границы интервалов, соответственно, $-28, -16, -4, +8, +20$ и $+32$. Дальнейшие вычисления удобно расположить в виде табл. 5-2, где x_i и n_i — исходный вариационный ряд; $x_i - 2$ — центрированные отклонения; $(x_i - 2)_{гр}$ — границы интервалов группирования; x_j и n_j — центры полученных интервалов в координатах исходных данных и число попадающих в них наблюдений.

Расчетное симметрирование гистограммы состоит в переносе некоторого числа отсчетов из данного столбца в симметричный с ним столбец. Так, в нашем примере высота столбцов, ближайших

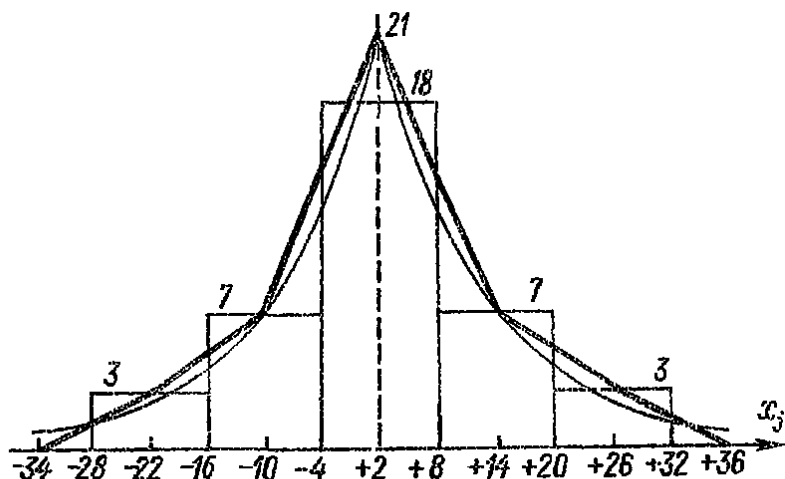


Рис. 5-6

к центральному, определяется как $(6 + 8)/2 = 7$, а следующих за ними, — как $(4 + 2)/2 = 3$. Отсимметрированные значения $\tilde{h}_i$ приведены в табл. 5-2, а полученная гистограмма показана на рис. 5-6.

Метод графического дифференцирования сглаженной и отсимметрированной интегральной функции распределения. Симметрирование производится путем переноса всех линий правой половины линейчатой диаграммы (рис. 5-5) на ее левую половину симметрично относительно принятого центра. Проще всего это выполнить, пользуясь значениями n_i и x_i — 2 табл. 5-2. Полученная таким образом левая половина отсимметрированной линейчатой диаграммы приведена ниже:

$- x_i - 2 $	. . .	-26	-20	-17	-14	-13	-10	-8	-5	-2
n_i		2	4	3	3	1	2	5	8	10
Σn_i		2	6	9	12	13	15	20	28	38

Затем подсчитывают набегавшую сумму Σn_i и строят кумулятивную ступенчатую кривую (рис. 5-7, а). Сглаживание этой кривой производят любым глазомерным способом, например «методом контура» (см. рис. 6-1).

Построение гистограммы. Гладкую и симметричную интегральную кривую графически дифференцируют, т. е. определяют ее прирост в границах столбцов будущей гистограммы, которые показаны на рис. 5-7 (при $x = -28, -16, -4, +8, +20$ и $+32$) вертикальными штриховыми линиями. Точки их пересечения со сглаженной интегральной кривой разбивают Σn_i на части n_i , равные (с округлением до целых чисел и с учетом того, что макси-

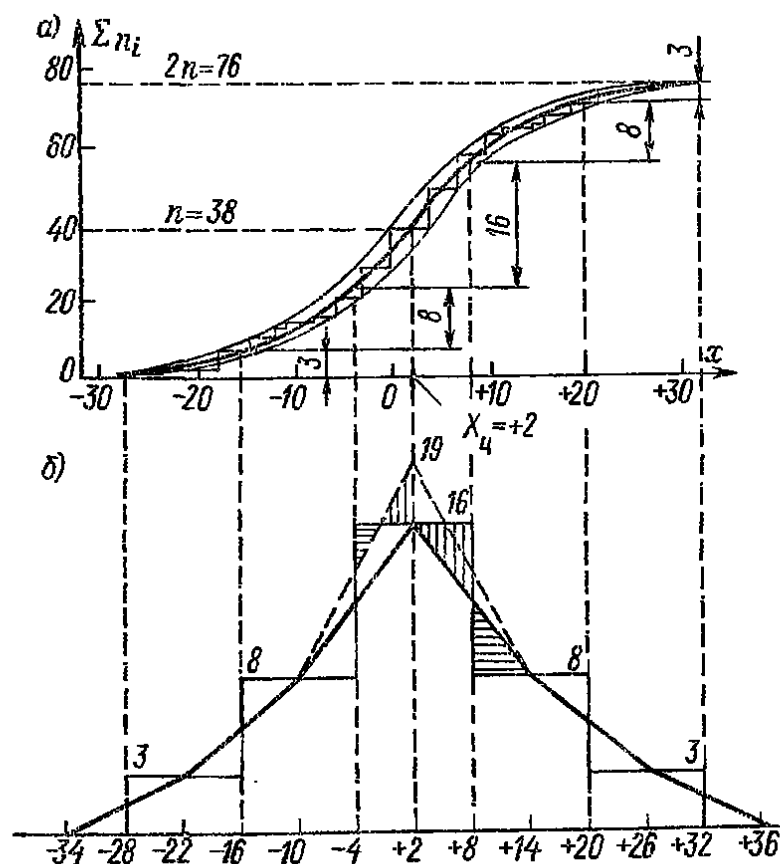


Рис. 5-7

мум интегральной кривой равен $2n = 76$) 3—8—16—8—3. По этим значениям n_j и строится симметричная гистограмма рис. 5-7, б.

Построение полигона, который более наглядно, чем гистограмма, отражает форму распределения, производится путем соединения прямыми середин верхних оснований каждого столбца гистограммы. За пределами гистограммы, как слева, так и справа следуют пустые интервалы, в которых точки, соответствующие их серединам, лежат на оси абсцисс. Все эти точки при построении полигона и соединяются между собой отрезками прямых линий, образуя совместно с осью x замкнутую фигуру, площадь которой в соответствии с правилом нормирования должна быть равна единице. В этой связи необходимо заметить, что во многих пособиях по теории вероятностей полигон распределения изображают в виде ломаной линии с незамкнутыми на ось x концами, что принципиально ошибочно.

При замене кривой плотности в виде гистограммы кривой в виде полигона правило нормирования выполняется автоматически, так как от каждого большого столбца гистограммы (см. рис. 5-7, б) отсекается и отбрасывается часть площади в виде треугольников, заштрихованных на рис. 5-7, б вертикально, а к каждому меньшему столбцу добавляются такие же площади треугольников, заштрихованных горизонтально. В итоге общая площадь под кривой полигона остается равной площади исходной гистограммы.

Однако это «перемещение площадей» при переходе от гистограммы к полигону происходит всегда в одном направлении, а именно «вниз по склону». В результате этого площадь центрального столбца в полигоне оказывается меньше площади центрального столбца гистограммы, т. е. форма распределения уплощается. Для устранения этого явления (с некоторым нарушением правила нормирования) в соответствии с п. 7 выводов § 5-2 для островершинных распределений следует считать центральный столбец гистограммы состоящим из двух равных столбцов с шириной $d/2$ и точки середин верхних оснований для построения полигона брать для этих более узких столбцов. В этом случае вершина кривой полигона будет выше верхнего основания центрального столбца, как это показано штриховой линией на рис. 5-7, б, но площадь под кривой полигона в пределах центрального столбца будет равна площади центрального столбца гистограммы. Подобным же образом построен полигон распределения и на рис. 5-6.

5.4. АППРОКСИМАЦИЯ ПОЛИГОНА РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНОЙ АНАЛИТИЧЕСКОЙ ФУНКЦИЕЙ

Случайная величина не имеет более полного описания, чем аналитическая кривая плотности распределения. Поэтому идентификация формы распределения сводится к выбору аналитиче-

ской модели, которая не противоречит данной конкретной выборке экспериментальных данных.

Для подбора аппроксимирующей полигон кривой должно быть выдвинуто предположение о виде аппроксимирующей ее функции. В § 2-4 было показано, что более 50% распределений погрешностей относятся к классу экспоненциальных:

$$p(x) = A \exp(-|x/(\lambda\sigma)|^\alpha), \quad (5-14)$$

поэтому прежде всего следует проверить, не описывается ли данное распределение этим соотношением. Такая задача сводится к нахождению параметров A , α и $\lambda\sigma = X_0$ этого выражения. Коэффициент A при любом $X_0 \neq 0$ просто равен значению функции $p(x)$ при $x = 0$. Для определения α удобен следующий прием. Логарифмируя (5-14), получаем

$$\ln p(x) = \ln A - \left| \frac{x}{X_0} \right|^\alpha \quad \text{или} \quad -\ln \frac{p(x)}{A} = \left| \frac{x}{X_0} \right|^\alpha, \quad (5-15)$$

логарифмируя еще раз, имеем $\ln \left(-\ln \frac{p(x)}{A} \right) = \alpha \ln |x| - \alpha \ln X_0$.

Так как значение масштабного коэффициента X_0 нам еще не известно, то значение α может быть найдено путем решения системы из двух уравнений, составленной для двух известных точек полигона. Для того чтобы аппроксимирующая кривая прошла через точки полигона $p(x_1)$ при x_1 и $p(x_2)$ при x_2 , необходимо удовлетворить системе уравнений:

$$\begin{cases} \ln \left[-\ln \frac{p(x_1)}{A} \right] = \alpha \ln |x_1| - \alpha \ln X_0; \\ \ln \left[-\ln \frac{p(x_2)}{A} \right] = \alpha \ln |x_2| - \alpha \ln X_0. \end{cases}$$

При вычитании этих уравнений друг из друга для определения α получается выражение

$$\alpha = \frac{\ln \left[-\ln \frac{p(x_2)}{A} \right] - \ln \left[-\ln \frac{p(x_1)}{A} \right]}{\ln x_2 - \ln x_1} = \frac{\Delta \ln \left[-\ln \frac{p(x)}{A} \right]}{\Delta \ln x}. \quad (5-16)$$

Так как по экспериментальным данным для каждой пары точек узлов полигона согласно формуле (5-16) будут получаться свои значения α , то рекомендуется нанести эти точки на график с координатами $\ln \{-\ln [p(x)/A]\} = F(\ln x)$ и на нем выбрать те точки, через которые рационально провести аппроксимирующую кривую. Если все точки такого графика лягут приблизительно на одной прямой, то тангенс угла наклона этой прямой и будет равен искомому значению α . Если же точки на этом графике образуют кривую, то должна быть принята другая модель.

Пример определения аппроксимирующих кривых распределений. Распределения, представленные на рис. 5-6 и 5-7, похожи на

Таблица 5-3

X_j	$\lg X_j$	n_j	$\frac{p(x)}{A}$	$\lg \left[-\lg \frac{p(x)}{A} \right]$	$\lg \left[-\ln \frac{p(x)}{A} \right]$
0	—	16	1	—	—
8	0,903	14	0,875	—1,237	—0,874
16	1,204	4	0,25	—2,209	+0,142
Δ	0,301	—	—	1,016	1,016

экспоненциальные с $\alpha \approx 1$, а на рис. 5-2, в — на экспоненциальное с $\alpha > 2$. Определим степени α этих трех распределений. В табл. 5-3 представлен расчет координат для построения указанного выше графика для полигона рис. 5-2, в. Так как α определяется отношением разностей логарифмов, то результат оказывается одинаковым при использовании как натуральных, так и десятичных логарифмов:

$$\alpha = \frac{\Delta \lg \left[-\lg \frac{p(x)}{A} \right]}{\Delta \lg x} = \frac{0,142 + 0,874}{1,204 - 0,903} = 3,3754 \approx 3,$$

т. е. распределение, приведенное на рис. 5-2, в, описывается формулой

$$p(x) = A \exp(-|x|^3).$$

В табл. 5-4 и 5-5 приведен аналогичный расчет для полигонов, изображенных на рис. 5-6 и 5-7. Из этих данных следует, что

Таблица 5-4

X_j	$\lg X_j$	n_j	$\frac{p(x)}{A}$	$\lg \left[-\lg \frac{p(x)}{A} \right]$	$\lg \left[-\ln \frac{p(x)}{A} \right]$
0	—	21	—	—	—
12	1,079	7	0,3333	—0,320	0,041
24	1,380	3	0,1428	—0,073	0,289
Δ	0,301	—	—	0,248	0,248

Таблица 5-5

X_j	$\lg X_j$	n_j	$\frac{p(x)}{A}$	$\lg \left[-\lg \frac{p(x)}{A} \right]$	$\lg \left[-\ln \frac{p(x)}{A} \right]$
0	—	19	—	—	—
12	1,079	8	0,421	—0,425	—0,063
24	1,380	3	0,158	—0,096	+0,266
Δ	0,301	—	—	0,329	0,329

показатель степени для полигона на рис. 5-6 $\alpha = 0,248/0,301 = 0,82$, а для полигона на рис. 5-7 $\alpha = 0,329/0,301 = 1,09$. При округлении значений показателя степени в выражении плотности распределения (5-14) полезно иметь в виду, что интеграл от этой функции аналитически выражается только для α , равного 1, $1/2$, $1/3$, $1/4$. Поэтому эти значения являются предпочтительными. Для α , равного 1,5, 2 или 3 и т. д., решения могут быть получены только численно (необходимо вычислять специальные таблицы интегральной функции, как это и делается для нормального распределения с $\alpha = 2$). Вследствие этого, хотя различие в форме полигонов, представленных на рис. 5-6 ($\alpha = 0,82$) и на рис. 5-7 ($\alpha = 1,09$), является заметным, для их описания целесообразно (ради упрощения дальнейшего использования искомого аналитического описания) принять $\alpha = 1$.

Относительно используемого в расчете значения $n_{j \max}$, являющегося эквивалентом $p(x)_{\max} = A$, следует заметить, что для плосковершинных распределений (рис. 5-2, в) в качестве $n_{j \max}$ может быть взята просто высота центрального столбца гистограммы. Однако для островершинных распределений и нормального распределения в качестве $n_{j \max}$ должна приниматься ордината круглой или острой вершины, найденная графически (рис. 5-6 и 5-7).

После того как найдены значения A и α , остается определить из выражений (5-15) значение X_0 :

$$\ln \left(-\ln \frac{p(x)}{A} \right) = \left| \frac{x}{X_0} \right|^\alpha \quad \text{или} \quad \left[-\ln \frac{p(x)}{A} \right]^{1/\alpha} = x/X_0;$$

$$X_0 = \frac{x}{\left[-\ln \frac{p(x)}{A} \right]^{1/\alpha}}.$$

Для кривой рис. 5-2, пользуясь данными табл. 5-3, получаем $X_0 = 8/(-\ln 0,875)^{1/3} \approx 15$, а для кривой рис. 5-6 по данным табл. 5-4 (для $\alpha = 1$) получаем $X_0 = 12/(-\ln 0,3333) \approx 11$. Отсюда выражения, аппроксимирующие кривые для рис. 5-2, есть $n(x) = 16e^{-1x/15}$, и для рис. 5-6, — $n(x) = 21e^{-1x/11}$. Эти кривые и были нанесены на рис. 5-2 и 5-6.

Аппроксимация двухмодальных и уплощенных распределений класса шапо может быть выполнена только после разложения таких распределений на составляющие, что может быть выполнено графически следующим путем. Если полигон распределения получился двухмодальным, то его мысленно представляют как сумму двух экспоненциальных полураспределений 1 и 2 (штриховые кривые на рис. 5-8, а), сдвинутых от центра соответственно на $+a$ и $-a$, где a — полуразмах дискретного двузначного распределения. При этом, однако, возникает следующее практическое ограничение. Для установления вида выделенной составляющей ее гистограмма должна быть представлена не менее чем пятью столб-

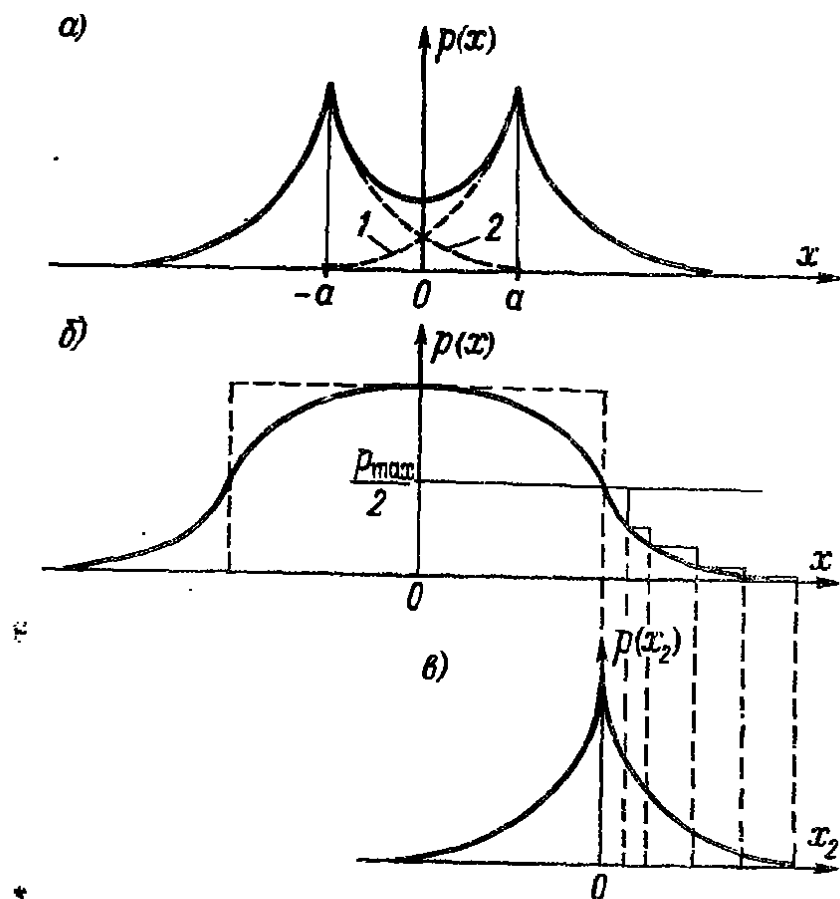


Рис. 5-8

цами, для чего исходное распределение должно состоять не менее чем из семи — девяти столбцов. Согласно соотношениям (5-13) для этого необходим объем выборки исходных данных $n = 200 \div \div 1000$ отсчетов. Таким образом, установление формы подобных сложных распределений требует существенно большего объема исходных данных, чем для простейших экспоненциальных распределений.

Если полигон распределения получается уплощенным (типа шапо), как это показано сплошной линией на рис. 5-8, б, то распределение такого вида следует попытаться разложить на равномерную и экспоненциальную составляющие. Равномерная составляющая, показанная на рис. 5-8, б штриховой линией, определяется довольно просто, так как в предположении симметрии экспоненциальной составляющей она должна пересекать кривую исходного распределения в точках, где $p(x) = p_{\max}/2$. Затем каждая половина кривой исходного распределения рассматривается как интегральная кривая экспоненциальной составляющей и путем графического дифференцирования (условно показанного на правом склоне кривой рис. 5-8, б) находятся ординаты кривой плотности (рис. 5-8, в). Если же выделенное экспоненциальное распределение оказывается резко несимметричным, т. е. верхняя и нижняя половины интегральной кривой этого распределения не являются двойным зеркальным отражением друг друга, исходное распределение не относится к классу шапо и его следует попробовать аппроксимировать экспоненциальным распределением

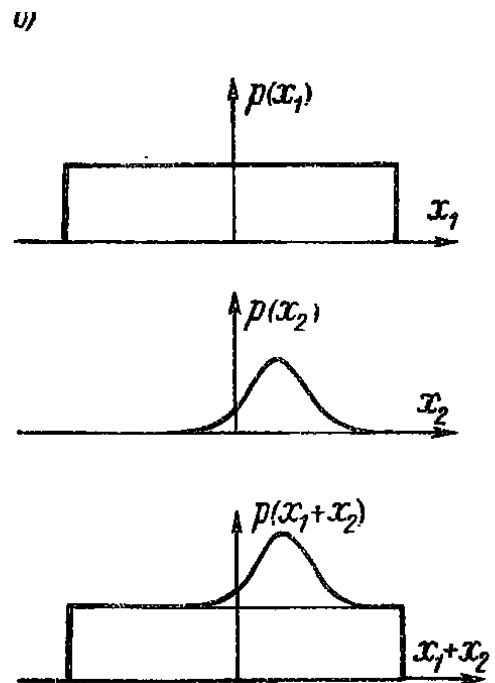
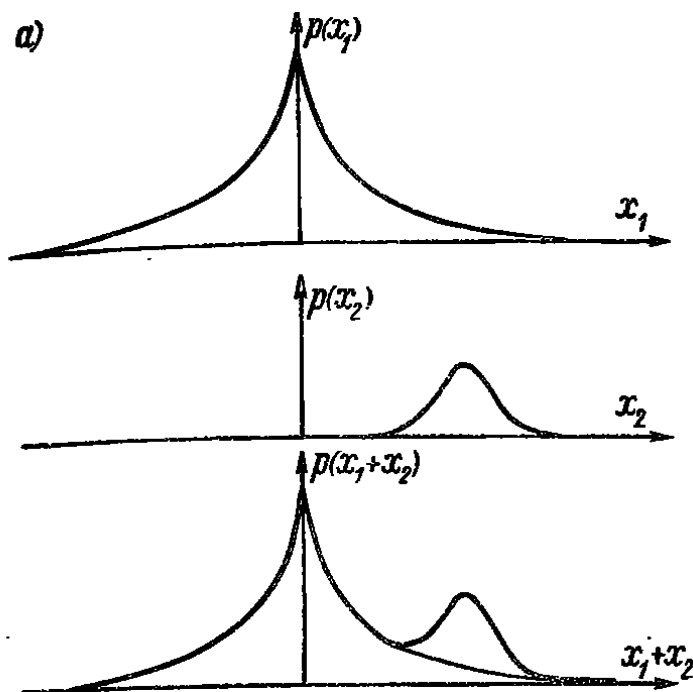


Рис. 5-9

с $\alpha > 2$. Для получения гистограммы экспоненциальной составляющей, состоящей хотя бы из пяти столбцов, исходное распределение должно состоять не менее чем из 9—11 столбцов, т. е. для достоверного установления формы таких композиций нужно $n = 500 \div 3000$ отсчетов.

Использование моделей в виде смеси распределений и ограниченных распределений. При анализе формы распределений следует иметь в виду, что далеко не все двухмодальные и уплощенные распределения могут быть разложены на экспоненциальную и дискретную или равномерную составляющие. Признаком возможности такого разложения является симметрия левого и правого спадов каждой из вершин двухмодального распределения или зеркальная симметрия верхней и нижней половин спада распределений типа шапо. А, например, арксинусоидальное распределение не может быть разложено на дискретную двухзначную и какую-либо еще составляющие, так как крутизна спадов его вершин резко несимметрична.

Некоторые из сложных несимметричных распределений могут быть представлены как смеси в каких-либо долях известных распределений с различными значениями координат центра. В качестве примеров таких моделей несимметричных распределений на рис. 5-9, а показана смесь экспоненциального распределения с небольшой долей нормального, а на рис. 5-9, б — смесь равномерно и нормально распределенных отсчетов.

Подобные распределения результатов измерений чаще всего являются сигналом о неисправности средств измерений или нарушении запланированной методики измерений. Поэтому обнаружение таких распределений является очень важным для практики экспериментальных исследований.

Распределения, круто обрывающиеся по краям, возникают

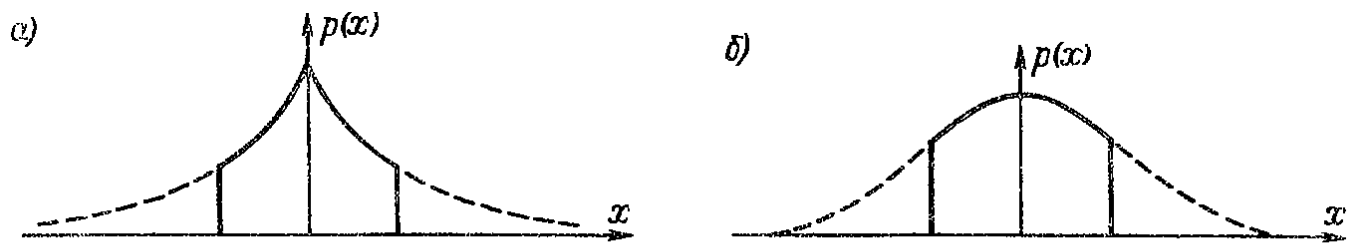


Рис. 5-10

в результате процессов ограничения сигналов, превышения сигналов пределов диапазона средства измерений, разбраковки сортируемых изделий, действия автоматических регуляторов и т. п. Примеры моделей, когда для описания наблюдаемого распределения могут быть использованы известные законы распределений, но усеченные в определенных пределах, приведены на рис. 5-10. Однако для достоверной идентификации формы таких распределений нужен очень большой объем исходных данных.

5-5. ИСПОЛЬЗОВАНИЕ КРИТЕРИЕВ СОГЛАСИЯ ПРИ ИДЕНТИФИКАЦИИ ФОРМЫ РАСПРЕДЕЛЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

В качестве ^{*}способа оценки близости распределения выборки экспериментальных данных к принятой аналитической модели закона распределения обычно рекомендуется использование так называемых критериев согласия. Известен целый ряд критериев согласия, предложенных разными авторами. Для более подробного знакомства с существом таких критериев можно рекомендовать монографию [10]. Для иллюстрации этих методов рассмотрим лишь два из них — критерий Колмогорова — Смирнова и критерий χ^2 Пирсона. С их помощью оценим близость экспериментальных выборок, представленных полигонами на рис. 5-2, 5-6 и 5-7, к полученным для этих распределений в § 5-4 аналитическим моделям $p(x) = 16e^{-|x/15|}$ — для рис. 5-2 и $p(x) = 21e^{-|x/11|}$ — для рис. 5-6.

Использование критерия согласия Пирсона заключается в вычислении величины χ^2 (хи-квадрат):

$$\chi^2 = \sum_{j=1}^m \frac{(E_j - M_j)^2}{M_j}, \quad (5-17)$$

где E_j — экспериментальные значения частот (n_j); M_j — значение частот в том же столбце, соответствующее выбранной модели; m — число столбцов гистограммы. Значения E_j , M_j и $E_j - M_j$ для различных X_j на рис. 5-2 приведены ниже:

X_j	...	-16	-8	0	+8	+16
E_j	...	4	14	16	14	4
M_j	...	4,3	13,7	16	13,7	4,3
$E_j - M_j$	...	-0,3	+0,3	0	+0,3	-0,3

Если бы выбранная модель в центрах всех m столбцов совпала с экспериментальными данными, то все m разностей $E_j - M_j$ были бы равны нулю, а следовательно, и значение критерия χ^2 также было бы равно нулю. Таким образом, χ^2 есть мера суммарного отклонения между моделью и экспериментальным распределением. Эта мера более наглядно работает, если рассматриваются две или более конкурирующие гипотезы. Например, применительно к гистограмме на рис. 5-2 могло быть выдвинуто конкурирующее предположение считать это распределение не экспоненциальным, а треугольным (рис. 5-2, в). Данные для расчета критерия для этого случая приведены ниже:

X_j	-16	-8	0	+8	+16
E_j	4	14	16	14	4
M_j	4	14	24	14	4
$E_j - M_j$	0	0	-8	0	0

Здесь разности $E_j - M_j$ во всех столбцах, кроме центрального, равны нулю и можно было бы предположить, что такая аппроксимация распределения более удачна. Однако расчет критерия χ^2 показывает, что в первом случае $\chi^2 = 0,3^2/4,3 + 0,3^2/13,7 + 0,3^2/13,7 + 0,3^2/4,3 = 0,15$, а во втором случае — $\chi^2 = 8^2/24 = 2,66$, т. е. много больше, следовательно, первая аппроксимация лучше второй.

Критерий χ^2 Пирсона позволяет произвести сравнение двух моделей и в том случае, когда для них используется разное число столбцов. Правда, критерий χ^2 не инвариантен к числу столбцов и существенно возрастает с увеличением их числа, но для использования его при различном числе столбцов составлены таблицы квантилей распределения χ^2 , входом в которые служит так называемое число степеней свободы $\nu = m - 1 - r$, где r — число определяемых по статистике параметров, необходимых для совмещения модели и гистограммы. Чтобы совместить модель с гистограммой, необходимо определить координату центра, а для того чтобы задать ширину модели, соответствующую ширине гистограммы, нужно определить оценку этой ширины, например с. к. о. Таким образом, число r определяемых по выборке параметров, необходимых для совмещения, равно 2. Отсюда число степеней свободы $\nu = m - 1 - r = m - 3$.

Часть таблицы квантилей распределения величины χ^2 приведена в табл. 5-6. С помощью этой таблицы от вычисленных значений χ^2 можно перейти к указанию вероятности P и уровня значимости $q = 1 - P$. Так, для модели рис. 5-2, в число столбцов $m = 5$, следовательно, $\nu = m - 3 = 2$ и нужно пользоваться первой строкой табл. 5-6. Отсюда значение $\chi_1^2 = 0,15$ соответствует $P_1 = 0,92$ и $q_1 = 0,08$, а значение $\chi_2^2 = 2,66$ соответствует $P_2 = 0,25$ и $q_2 = 0,75$.

Эти числа означают следующее. При использовании модели 1 вероятность получить χ_1^2 меньшим, чем 0,15, равна 0,92, т. е. такие значения χ^2 встретятся в 92 случаях из 100. Следовательно,

ν	χ^2 при P , равной					
	0,99	0,95	0,9	0,8	0,5	0,25
2	0,02	0,10	0,21	0,45	1,4	2,77
4	0,30	0,71	1,06	1,65	3,4	5,39
6	0,87	1,63	2,20	3,07	5,4	7,84
8	1,65	2,73	3,49	4,59	7,3	10,2
10	2,56	3,94	4,87	6,18	9,3	12,5
12	3,57	5,23	6,30	7,81	11,3	14,8
14	4,66	6,57	7,79	9,47	13,3	17,1
16	5,81	7,96	9,31	11,2	15,3	19,4

*

отвергая модель 1, мы при многократном повторении будем ошибаться в среднем 92 раза из 100, т. е. оснований для ее отклонения нет. При использовании модели 2 с $\chi^2 = 2,66$, $P_2 = 0,25$ и $q_2 = 0,75$, отвергая эту модель, мы будем в среднем ошибаться лишь 25 раз из 100, а в 75 случаях будем правы.

Однако первое утверждение, что «для отклонения модели 1 с $\chi^2 = 0,15$, $P_1 = 0,92$ и $q_1 = 0,08$ оснований нет», отнюдь не есть утверждение, что данная модель и есть искомое распределение. Дело заключается в том, что, кроме модели экспоненциального распределения с $\alpha = 3$, мы могли бы использовать модель вида трапеции, модель вида шапо и т. д. И во всех случаях использования подходящих моделей χ^2 получало бы малые значения, а вероятность оказывалась бы достаточно большой (равной, например, 0,9; 0,95 и т. п.). Поэтому при использовании критериев согласия положительный ответ нельзя рассматривать как утверждение о правильно выбранной модели. Определенным ответом является лишь отрицательный ответ.

Таким образом, диалог с критерием согласия следует понимать только следующим образом. Вы спрашиваете: «Не является ли это распределение экспоненциальным с $\alpha = 3$?» Ответ $\chi^2 = 0,15$, $P = 0,92$ означает: «Тоже может быть». Вы спрашиваете: «А может быть оно трапецеидальным?» Ответ $\chi^2 = 0,21$, $P = 0,9$ означает: «Тоже может быть». Вы спрашиваете: «А может быть, оно имеет форму шапо?» Ответ $\chi^2 = 0,10$, $P = 0,95$ означает: «Тоже может быть». И эти ответы равнозначны. Определенным является лишь отрицательный ответ. Вы спрашиваете: «А может быть, распределение треугольное?» Ответ $\chi^2 = 2,66$, $P = 0,25$, $q = 0,75$ означает: «Маловероятно, вероятность ошибки $q = 0,75$ ».

Подобным же образом с помощью критерия χ^2 может быть оценено различие между выбранной выше моделью экспоненциального распределения и гистограммами на рис. 5-6 и 5-7. Для гистограммы на рис. 5-6 E_j , M_j и $E_j - M_j$ для различных X_j приведены ниже:

X_j	...	...	...	...	...
E_j	...	...	...	...	...
M_j	...	...	...	...	...
$E_j - M_j$	...	...	...	...	...

Тогда $\chi^2 = 2 (0,63)^2 / 2,37 = 0,3$, $P_1 = 0,85$ и уровень значимости $q = 1 - P_1 = 0,15$, а для гистограммы на рис. 5-7 значения E_j , M_j , $E_j - M_j$ для различных X_j приведены ниже:

X_j	...	...	...	...	...
E_j	...	...	...	...	...
M_j	...	...	...	...	...
$E_j - M_j$	...	...	...	...	...

Отсюда $\chi^2 = 1,0$, $P_2 = 0,65$ и $q_2 = 1 - P_2 = 0,35$. По уровню значимости совпадение первой гистограммы с принятой моделью в $q_2/q_1 = 0,35/0,15 \approx 2$ раза лучше, чем второй. Однако эти оценки критерия χ^2 , вероятности P и уровня значимости q отнюдь не могут рассматриваться как показатели достоверности установления вида генеральной совокупности, из которой были получены эти выборки. Так, например, если бы при выборе вида модели для рис. 5-2 не было произведено округления показателя степени $\alpha \approx 3$, а было принято его значение в виде $\alpha = 3,3754$, то все $E_j - M_j$ для рис. 5-2 были бы близки к нулю и χ^2 было бы также близко к нулю, но это отнюдь не означало бы, что искомый вид распределения генеральной совокупности был бы определен нами с вероятностью 0,9999...

Критерий согласия Колмогорова — Смирнова [10] позволяет сравнить две независимые выборки и ответить на вопрос, относятся ли они к одной и той же генеральной совокупности без выбора для сравнения какой-либо предполагаемой модели. Критерий Колмогорова (1933 г.) и Смирнова (1939 г.) чувствителен к различию центров, величин рассеяния, асимметрии и эксцесса. В качестве статистики служит наибольшая (по модулю) разность между ординатами кривых обеих относительных накопленных частот. Сравнение может выполняться без построения гистограммы, так как интегральные функции распределений в виде набегающих сумм могут быть построены без какого-либо группирования. Если же используются гистограммы, то они должны быть построены при одинаковых для обеих выборок границах и числе интервалов группирования. Накопленные частоты F_1 и F_2 делятся на соответствующие объемы выборок n_1 и n_2 . Затем вычисляется разность $F_1/n_1 - F_2/n_2$. Максимум модуля этой разности и есть искомая статистика (для двустороннего критерия):

$$D = \max |F_1/n_1 - F_2/n_2|.$$

Распределение статистики было табулировано Смирновым (1948 г.). Для $n_1 + n_2 > 35$ граничное значение D с вероятностью P может быть приближенно найдено из аппроксимирующего выражения

$$D_{гр} = K_{гр} \sqrt{1/n_1 + 1/n_2} \quad (5-18)$$

где $K_{гр}$ — постоянная, зависящая от назначенной вероятности P или уровня значимости $q = 1 - P$. Наблюдаемое наибольшее значение разности D значимо на соответствующем уровне q , если оно достигает или превосходит $D_{гр}$. Для постоянной $K_{гр}$ Л. Закс [10, с. 302] дает аппроксимацию в виде $K_{гр} = \sqrt{-0,5 \ln(q/2)}$.

При сопоставлении данной выборки с объемом $n = n_1$ с выбранной аналитической моделью следует полагать $n_2 = \infty$. Тогда из выражения (5-18) следует

$$D_{гр} = K_{гр} / \sqrt{n} = \sqrt{-0,5 \ln(q/2) / n},$$

а D находится как

$$D = \max |F_1 - F_2| / n = \max (\Delta / n),$$

где $\Delta = |F_E - F_M|$. Отсюда искомая вероятность может быть найдена как

$$P \geq 2e^{-2nD^2} = 2e^{-2\Delta^2/n}. \quad (5-19)$$

Таким образом, достоинство критерия Колмогорова — Смирнова состоит в том, что при $n > 35$ он позволяет чисто аналитически по приближенной формуле (5-19) оценить вероятность ошибки $q = 1 - P$, не прибегая к использованию каких-либо таблиц, что резко ускоряет анализ и позволяет использовать ЭВМ.

Но допускаемая при этом приближенность должна полностью пониматься пользователем. При большом различии принятой модели и сравниваемой с ней выборки оценка вероятности, получаемая по критерию Колмогорова — Смирнова, достаточно близка к оценке вероятности по критерию χ^2 . Так, например, при сопоставлении гистограммы, показанной на рис. 5-2, в, и модели треугольного распределения по критерию χ^2 было получено $P = 0,25$. Максимальная разность $\Delta = |F_E - F_M|$ в этом случае равна 8, и при $n = 52$ по формуле (5-19) получаем

$$P = 2e^{-2\Delta^2/n} = 2e^{-2 \cdot 8 \cdot 8 / 52} = 0,17.$$

Но если бы Δ было равно 7, 6, 5 и т. д., то рассчитанные по формуле (5-19) при $n = 52$ вероятности P составили бы:

Δ	8	7	6	5	4	3	2	1
P	0,17	0,30	0,50	0,76	1,08	1,4	1,7	1,9

т. е. различие вероятностей 0,17 и 0,25 соответствует изменению Δ менее чем на одну точку, так как при $\Delta = 7$ вероятность $P = 0,30$.

Однако при $\Delta \leq 4$ рассчитанные вероятности получаются больше единицы и должны считаться равными единице. Это будет происходить при всех $\Delta \leq \sqrt{0,5n \ln 2}$, когда критерий Колмогорова — Смирнова дает ответ: «Тоже может быть», т. е. уже не отвергает предположений о пригодности модели. Так будет при $\Delta < 9$ и $n = 233$, при $\Delta < 6$ и $n = 104$, при $\Delta < 5$ и $n = 72$,

при $\Delta < 4$ и $n = 46$. Поэтому использование критерия Колмогорова — Смирнова применительно к гистограммам на рис. 5-6 и 5-7, где $\Delta = 8 - 7 = 1$ и $\Delta = 21 - 19 = 2$, бессмысленно, так как согласно ему предложенная модель совпадает с обеими исследуемыми гистограммами с вероятностью $P = 1$.

Поэтому в работах по использованию критериев согласия обычно оговаривается, что их применение корректно лишь при достаточно представительных выборках исходных данных и в качестве приближенной ориентировки указывается $n > 200$.

Располагая формулой (5-11), это предупреждение можно раскрыть более наглядно. Действительно, в приведенных примерах для расчета критерия согласия использовалось очень малое число наблюдаемых разностей между гистограммой и моделью $\Delta = E_j - M_j$ — две в первом примере ($-0,3$ и $+0,3$), одна во втором (-8), одна в третьем ($+0,63$) и три в четвертом ($0,63$, $1,0$, $-2,0$). Естественно, статистика отклонений, состоящая из 1—3 разностей, слишком мала и основывать на ней статистические заключения некорректно.

Чтобы число этих отклонений было больше, гистограмма должна состоять не из 5 столбцов, а хотя бы из 7—9—11 столбцов. Но согласно формуле (5-11) для того, чтобы гистограмма имела m столбцов, выборка должна содержать $n = [6m/(\epsilon + 1,5)]^{2,5}$ отсчетов. При нормальном распределении с $\epsilon = 3$ для наличия в гистограмме $m = 7—11$ столбцов необходимо $n = 170—800$ отсчетов, а при равномерном распределении с $\epsilon = 1,8$ для $m = 7—11$ столбцов нужно $n = 600—2700$ отсчетов. Отсюда и вытекает в качестве приближенной оптимистической ориентировки указание $n > 200$, однако для $m = 9$ для тех же распределений нужно, соответственно, 500 и 1080 отсчетов. Корректность использования как критерия Пирсона, так и критерия Колмогорова — Смирнова при таких выборках будет достаточно обеспечена. Но беда в том, что такими выборками по 500—2500 отсчетов экспериментатор практически никогда не обладает, так как для их получения необходима непомерно большая затрата времени и средств.

5-6. ВЫЧИСЛЕНИЕ ПО ГИСТОГРАММЕ ОЦЕНОК ШИРИНЫ И ФОРМЫ КРИВОЙ ИССЛЕДУЕМОГО РАСПРЕДЕЛЕНИЯ

Определение оценки координаты центра распределения, как было показано выше, производится еще до симметрирования гистограммы, ибо симметрирование осуществляется именно относительно уже выбранного центра. Знание оценки центра необходимо и для расчета оценки ширины распределения в виде с. к. о. и оценки формы распределения в виде эксцесса, так как обе эти оценки определяются через центральные моменты распределения.

Определение оценки ширины распределения в виде энтропийного интервала неопределенности (или его половины как энтро-

Таблица 5-7

x_j	n_j	$n_j \lg n_j$	$2x_j^2 n_j$	$2x_j^4 n_j$
0	16	19,266	0	0
8	2×14	$2 \times 16,046$	1792	144 688
16	2×4	$2 \times 2,408$	2048	524 288
Σ	52	56,174	3840	638 976

нийного значения погрешности) не нуждается в предварительном определении координаты центра распределения. Поэтому если ошибка в определении координаты центра влечет за собой существенный сдвиг оценок с. к. о., эксцесса и контрэксцесса, то оценка энтропийного значения погрешности свободна от этой зависимости. В этом состоит одно из ее основных достоинств, но она в отличие от с. к. о. не может быть найдена без группирования членов вариационного ряда по соответствующим интервалам.

Однако после построения симметризованной гистограммы настолько резко упрощается расчет с. к. о., энтропийного коэффициента и эксцесса, что к построению симметризованной гистограммы часто прибегают только для облегчения таких расчетов.

Действительно, при расчете с. к. о. нужно найти сумму $\sum_1^n (x_i - X_{\text{ц}})^2$, при расчете эксцесса, кроме того, — сумму $\sum_1^n (x_i - X_{\text{ц}})^4$, а для расчета энтропийного значения погрешности и энтропийного коэффициента — сумму $\sum_1^m n_j \lg n_j$. При центрированной гистограмме, середина центрального столбца которой совмещена с центром распределения, т. е. $X_{\text{ц}} = 0$, все расчеты существенно упрощаются. В этом случае $\sum_1^n (x_i - X_{\text{ц}})^2 = \sum_1^m X_j^2 n_j$; $\sum_1^m (x_i - X_{\text{ц}})^4 = \sum_1^m X_j^4 n_j$, где X_j — координаты центров столбцов гистограммы; n_j — число отсчетов, попадающих в j -й столбец; m — число столбцов. Вместо возведения в квадрат или в четвертую степень n разностей при симметричной гистограмме с $m = 2k + 1$ столбцами возводить в квадрат или четвертую степень нужно всего k чисел. Таким образом, расчет всех необходимых оценок становится предельно простым и выполнять его удобно путем заполнения таблицы.

Для иллюстрации сказанного проведем расчет перечисленных выше оценок для трех распределений, гистограммы которых были представлены на рис. 5-2, 3, 5-6 и 5-7. В табл. 5-7 приведены

Таблица 5-8

K_j	n_j	$n_j \lg n_j$	$2X_j^2 n_j$	$2X_j^4 n_j$
0	16	19,266	0	0
12	2×8	$2 \times 7,225$	2304	331 776
24	2×3	$2 \times 1,431$	3456	1 990 656
Σ	38	36,576	5760	2 322 432

данные распределения в соответствии с гистограммой рис. 5-2, в, причем центральный столбец гистограммы представлен верхней строкой таблицы. В каждой строке производится вычисление величин $n_j \lg n_j$, $X_j^2 n_j$, $X_j^4 n_j$ и затем определяются суммы каждого из столбцов таблицы. При вычислении этих сумм необходимо только не забывать, что центральный столбец гистограммы (верхняя строка таблицы) — один, а остальных симметричных столбцов гистограммы — по два, поэтому при суммировании соответствующие им числа должны удваиваться. Для этого запись чисел в столбцах таблицы n_j и $n_j \lg n_j$ удобно вести так, как это сделано в табл. 5-7 ($n_1 = 16$, $n_2 = 2 \times 14$, $n_3 = 2 \times 4$ и т. д.). По полученным в ней суммам находим для гистограммы рис. 5-2, в

$$\sigma = \sqrt{\sum X_j^2 n_j / (n - 1)} = \sqrt{3840/51} = 8,68;$$

$$k = \frac{dn}{2\sigma} 10^{-\frac{1}{n} \sum_1^m n_j \lg n_j} = \frac{8 \cdot 52}{2 \cdot 8,68} 10^{-56,174/52} = 1,99;$$

$$\varepsilon = \mu_4 / \sigma^4 = \sum_1^m X_j^4 n_j / (n \sigma^4) = 638 976 / (52 \cdot 8,68^4) = 2,17;$$

$$\kappa = 1/\sqrt{\varepsilon} = 1/\sqrt{2,17} = 0,68.$$

Таким образом, расчет оценок σ , k , ε и κ по симметрированной гистограмме занимает всего несколько строк.

В табл. 5-8 и 5-9 приведен аналогичный расчет сумм, необходимых для определения оценок распределений, представленных

Таблица 5-9

X_j	n_j	$n_j \lg n_j$	$2X_j^2 n_j$	$2X_j^4 n_j$
0	18	22,594	0	0
12	2×7	$2 \times 5,916$	2016	290 304
24	2×3	$2 \times 1,431$	3456	1 990 656
Σ	38	37,289	5472	2 280 960

гистограммами на рис. 5-6 и 5-7. Используя суммы табл. 5-8, получаем для гистограммы на рис. 5-7

$$\sigma_1 = \sqrt{5760/37} = 12,5; \quad k_1 = \frac{12 \cdot 38}{2 \cdot 12,5} 10^{-36,58/38} = 1,99;$$

$$\kappa_1 = \sigma_1^2 \left| \sqrt{\Sigma X_i^4 n_i / n} \right| = 0,63 \text{ и } \varepsilon_1 = 2,52.$$

Используя суммы табл. 5-9, получаем для гистограммы на рис. 5-6

$$\sigma_2 = \sqrt{5472/37} = 12,2;$$

$$k_2 = \frac{12 \cdot 38}{2 \cdot 12,2} 10^{-37,29/38} = 1,95; \quad \kappa_2 = 147,9/245 = 0,60 \text{ и } \varepsilon_2 = 2,77.$$

Поправки на группирование и малый объем выборки. Наряду с резким упрощением вычисления оценок ширины и формы распределения с использованием симметризованных гистограмм эти оценки имеют систематические погрешности, которые могут быть устранены введением соответствующих поправок. Поправка для устранения влияния описанного в конце § 5-4 «смещения площадей вниз по склону» при столбцах гистограммы шириной d на оценку с. к. о. носит название поправки Шеппарда на группирование. Она заключается в том, что исправленное значение с. к. о. σ для одномодальных распределений находится как

$$\sigma = \tilde{\sigma} \sqrt{1 - \frac{1}{12} \left(\frac{d}{\tilde{\sigma}} \right)^2}, \quad (5-20)$$

а для арксинусоидального распределения эта поправка имеет обратный знак: $\sigma = \tilde{\sigma} \sqrt{1 + \frac{1}{2} \left(\frac{d}{\tilde{\sigma}} \right)^2}$, где $\tilde{\sigma}$ — оценка с. к. о., найденная без поправки на группирование. Оценка четвертого момента μ_4 одномодальных распределений с поправкой на группирование выражается через неисправленное значение $\tilde{\mu}_4$, оценку σ и значение d :

$$\mu_4 = \tilde{\mu}_4 - \frac{\sigma^2 d^2}{2} + \frac{7d^4}{240}, \quad (5-21)$$

для равномерного распределения $\mu_4 = \tilde{\mu}_4 + \sigma^2 d^2/2 - 7d^4/240$ и для арксинусоидального распределения $\mu_4 = \tilde{\mu}_4 + \sigma^3 \left(d - \frac{5}{40} \sqrt{\sigma d} \right)$.

Оценка четвертого момента испытывает смещение также и в зависимости от объема выборки n . Для введения этой поправки обычно рекомендуется формула

$$\mu_4 = \frac{\tilde{\mu}_4 n (n^2 - 2n + 3) - \sigma^4 3n (2n - 3)}{(n-1)(n-2)(n-3)},$$

однако при $n \gg 6$ это выражение по правилам приближенных вычислений значительно упрощается, получая вид

$$\mu_4 = \bar{\mu}_4 (1 + 4/n) - 6\sigma^4/n. \quad (5-22)$$

Обе эти поправки должны вводиться последовательно, т. е. в предыдущие формулы в качестве $\bar{\mu}_4$ должно подставляться значение μ_4 , уже исправленное в соответствии с формулой (5-22). Пример использования этих поправок будет приведен в § 5-9.

Неоднозначность суждений, получаемых при анализе малых выборок. Неоднозначность, уже упоминавшаяся при обсуждении критериев согласия, существует и при аналитическом расчете оценок ширины и формы распределения по его гистограмме. Так, в рассмотренном примере при графическом сглаживании и симметрировании мы получили гистограмму (рис. 5-7) с $\sigma_1 = 12,5$, а при расчетном симметрировании — гистограмму (рис. 5-6) с $\sigma_2 = 12,2$. Относительное различие этих оценок составляет $\gamma(\sigma) = (12,2 - 12,5)/12,2 = -0,025 = -2,5\%$. Относительное различие оценок контрэксцесса и энтропийного коэффициента этих гистограмм составляет, соответственно, $\gamma(\kappa) = (0,60 - 0,63)/0,60 = -0,05 = -5\%$ и $\gamma(k) = (1,95 - 1,99)/1,95 = -0,02 = -2\%$. Однако среднее квадратическое значение возможных отклонений, обусловленных не методом симметрирования гистограммы, а самим фактом малости выборки ($n = 38$) и формой закона распределения, согласно формулам (4-8)—(4-10), составляет (при $\varepsilon = 2,52$ и $k = 1,99$):

$$\begin{aligned} \delta(\sigma^*) &= \sigma(\sigma^*)/\sigma^* = \sqrt{\varepsilon - 1}/(2\sqrt{n}) = \\ &= \sqrt{2,52 - 1}/(2\sqrt{38}) = 0,10 = 10\%; \\ \delta(\kappa^*) &= \sigma(\kappa^*)/\kappa^* = \sqrt[4]{(\varepsilon^2 - 1)^3}/\sqrt{29n} = \\ &= \sqrt[4]{(2,52^2 - 1)^3}/\sqrt{29 \cdot 38} = 0,10 = 10\%; \\ \delta(k^*) &= \sigma(k^*)/k^* = 1/k \sqrt{kn \sqrt{\kappa^3}} = \\ &= 1/1,99^2 \sqrt{1,99 \cdot 38 \sqrt{0,63^3}} = 0,04 = 4\%. \end{aligned}$$

Таким образом, неоднозначность оценок графического или расчетного симметрирования гистограмм существенно меньше неопределенности, обусловленной случайностью малых выборок исходных данных и видом закона распределения.

5-7. ПРИБЛИЖЕННАЯ ИДЕНТИФИКАЦИЯ ФОРМЫ РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ ПО СОЧЕТАНИЮ ОЦЕНОК КОНТРЭКССЕССА И ЭНТРОПИЙНОГО КОЭФФИЦИЕНТА С ИСПОЛЬЗОВАНИЕМ ТОПОГРАФИЧЕСКОЙ КЛАССИФИКАЦИИ

Как было показано в предыдущем параграфе, числовые оценки формы распределения в виде контрэксцесса κ и энтропийного коэффициента k даже при малом объеме выборки эксперименталь-

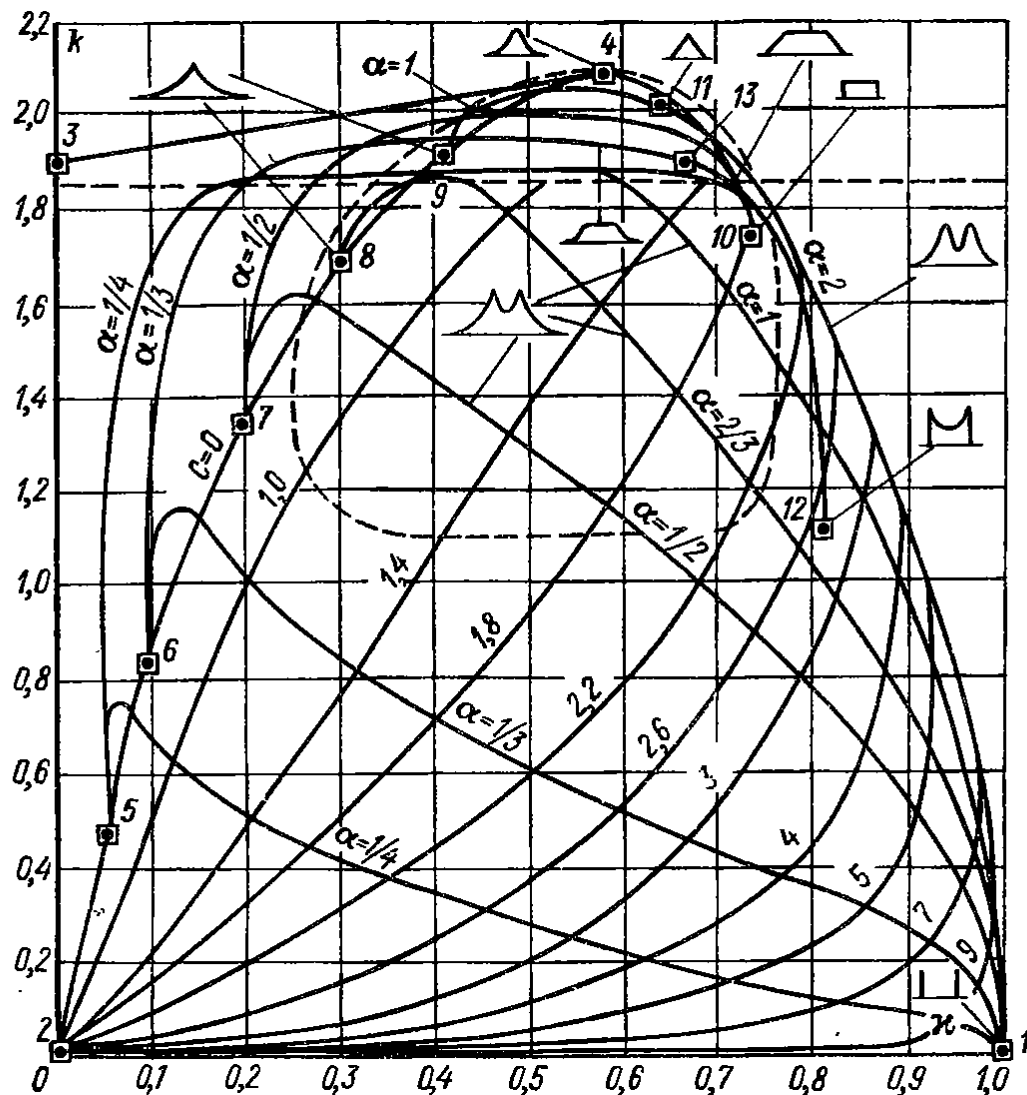


Рис. 5-11

ных данных ($n \approx 40$) определяются уже с достаточной точностью ($\gamma = 5 \div 10\%$). Этим можно воспользоваться для вынесения суждения о виде формы кривой исследуемого распределения с помощью топографической классификации математических моделей распределений в координатах k и η , которая была показана на рис. 2-10. Действительно, вычислив оценки k и η и нанеся точку с этими координатами на плоскость $k - \eta$, можно, кроме визуального впечатления от вида полигона, получить еще одно независимое формальное указание на возможную форму кривой распределения.

Для практического осуществления метода приближенной идентификации необходимо иметь более подробную топографическую классификацию законов распределения в плоскости $k - \eta$, чем та, что была приведена на рис. 2-10. Такая классификация показана на рис. 5-11. На ней линия, соединяющая точки 2—3—4, как и на рис. 2-10, есть геометрическое место точек, соответствующих семейству распределений Стьюдента с числом степеней свободы ν от 1 до ∞ . Линия, соединяющая точки 2—5—6—7—8—9—4—10, есть геометрическое место точек, соответствующих классу экспоненциальных распределений с показателем степени α от 0 до ∞ . Точка 2 соответствует распределению с $\alpha \rightarrow 0$, точка 5 — с $\alpha =$

$= 1/4$, точки 6—8 соответствуют значениям α , равным $1/3$, $1/2$, $2/3$, точка 9 соответствует распределению Лапласа с $\alpha = 1$, точка 4 (с $\alpha = 2$) — нормальному распределению Гаусса и точка 10 (с $\alpha \rightarrow \infty$) — равномерному распределению.

Для описания уплотненных высокоэнтропийных распределений (с $k \geq 1,87$) использованы распределения класса шапо, т. е. композиции равномерного распределения (точка 10) с различными экспоненциальными распределениями. Геометрические места точек таких композиций расположены на рис. 5-11 на линиях, соединяющих точку 10 с точками соответствующих экспоненциальных распределений, начиная от линии 5—10 (помеченной значением $\alpha = 1/4$), линии 6—10 ($\alpha = 1/3$) и т. д. до линии 4—10, являющейся геометрическим местом точек, соответствующих композициям нормального (точка 4) и равномерного (точка 10) распределений.

Для описания низкоэнтропийных распределений ($k < 1,87$) использованы композиции экспоненциальных распределений с дискретным двузначным распределением вида $p(x) = 0,5 [\delta(x - a) + \delta(x + a)]$ (точка 1 на рис. 5-11). Геометрические места точек таких распределений расположены на рис. 5-11 по линиям, соединяющим точки 5, 6, 7, 8, 9 и 4 с точкой 1 и помеченным значениями $\alpha = 2$, $\alpha = 1$, $\alpha = 2/3$ и т. д. до $\alpha = 1/4$. Относительное содержание дискретной составляющей в таких распределениях удобно характеризовать отношением $C = a/\sigma$, где a — полуразмах дискретного распределения, а σ — с. к. о. экспоненциального. Линии равных C также нанесены на рис. 5-11.

Кроме того, на рис. 5-11 нанесены геометрические места точек трапецеидальных распределений в виде линии, соединяющей точку 10, соответствующую равномерному распределению, и точку 11, соответствующую треугольному распределению, а также композиций двух арксинусоидальных распределений — линия, соединяющая точку 12 (арксинусоидальное распределение) и точку 13 (композиция двух равных арксинусоидальных распределений).

Метод использования топографической классификации распределений (рис. 5-11) для вынесения суждения о возможной форме распределения генеральной совокупности, из которой была получена исследуемая выборка, сводится к следующему. Так, например, для полигона распределения на рис. 5-6 в § 5-6 были получены оценки $\gamma_1 = 0,63$ с относительной погрешностью от малой выборки ($n = 38$) $\gamma_n = 10\%$ и $k_1 = 1,99$ с $\gamma_k = 4\%$. Таким образом, контрэксцесс этого распределения лежит в пределах от 0,57 до 0,69, а энтропийный коэффициент — в пределах от 1,91 до 2,07. Прямоугольник с этими границами наносится на график рис. 5-11 и очерчивает область распределений, к которым может принадлежать исследуемая выборка. На этом основании утверждается, что исследуемое распределение не может принадлежать к распределениям, находящимся вне этого прямоугольника,

и, наоборот, распределения, находящиеся в этих границах, не противоречат исследуемой выборке.

Этот метод приближенной идентификации формы распределений погрешностей был использован в работах, упоминавшихся в § 5-1, а именно в работе [8] для идентификации формы более 100 распределений погрешностей ИИС и в работе [3] для идентификации распределений погрешностей более 200 приборов для измерения различных электрических и неэлектрических величин. В результате этих и других работ выяснилось, что формы распределений погрешностей достаточно разнообразны и соответствующие им точки располагаются в пределах области, ограниченной штриховым овалом (рис. 5-11), охватывающей нормальное, уплощенные, равномерное и двухмодальные распределения.

Для распределений, точки которых в координатах рис. 5-11 попадают в область линии 4—10, оказывается возможным одновременно использовать целый ряд равноправных моделей в виде трапеций, в виде композиций класса шапо или в виде $p(x) = Ae^{-|x|^{\alpha}}$, где $\alpha = 2 \div \infty$. Выбор одного из этих трех видов аналитического описания должен производиться исходя из практического удобства дальнейшего использования получаемых выражений (см. § 5-4).

Основным недостатком, обуславливающим приближенность описанной систематизации и классификации распределений по их форме, является ее не однозначность. Параметры k и κ данного аналитического распределения определяются однозначно. Так, например, если распределение описывается симметричной трапецией с отношением оснований 2 : 3 (п. 2 табл. 2-2), то ее параметры однозначно равны: $\kappa = 0,728$ и $k = 1,83$. Однако обратный переход уже неоднозначен. Экспериментальное распределение с $\kappa = 0,728$ и $k = 1,83$ совсем необязательно будет трапецеидальным. Через точку топографической классификации (рис. 5-11) с этими координатами проходит целый пучок кривых, соответствующих и трапециям, и уплощенным экспоненциальным распределениям, и композициям класса шапо, и даже линия 12—13, соответствующая двухмодальным арксинусоидальным распределениям. Подобная же неоднозначность обратного перехода характерна и для линии класса экспоненциальных распределений (в области точек 5—6—7—8—9—4 на рис. 5-11), которая пересекается линиями, соответствующими островершинным двухмодальным распределениям.

5-8. МЕТОД ПРИБЛИЖЕННОЙ ИДЕНТИФИКАЦИИ ФОРМЫ РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ, ПРИГОДНЫЙ ДЛЯ РЕАЛИЗАЦИИ НА ЭВМ БЕЗ УЧАСТИЯ ОПЕРАТОРА

Описанные в предыдущем параграфе операции, необходимые для идентификации формы распределений, могут быть переданы ЭВМ, если вместо графического представления кривых (рис. 5-11)

иметь их аналитическое описание. Для этого необходимо найти аналитические формулы перехода, например, при определении параметров двухмодальных распределений, от двух параметров (k и κ), заданных в сетке прямоугольных координат, к двум параметрам (C и α) формы двухмодального распределения, заданных в сложной сетке криволинейных координат (рис. 5-11).

Такая методика приближенной машинной идентификации формы распределений была разработана авторами совместно с инженерами Т. Ф. Шевчук, В. В. Орловской, О. К. Фирсовой и Н. А. Масленниковой и состоит в следующем. Прежде всего производится проверка того, не близко ли исследуемое распределение к чисто экспоненциальным, т. е. не располагается ли отображающая его точка на рис. 5-11 вблизи линии экспоненциальных распределений 2—5—6—7—8—9—4. Эта линия на участке от точки 7 до точки 4 с погрешностью менее 5% может быть представлена аппроксимирующим выражением

$$k = \sqrt{\pi e/2} - 5,2 (1/\sqrt{3} - \kappa)^2, \quad (5-23)$$

а на участке от точки 4 до точки 10 с погрешностью $\pm 0,5\%$ выражением

$$k = \sqrt{\pi e/2} - 69,4 (\kappa - 1/\sqrt{3})^3.$$

Распределение признается принадлежащим к классу экспоненциальных, если отклонение отображающей точки от кривой (5-23) не превосходит разброса оценок k и κ , определяемого малостью выборки. Согласно выражениям (4-9) и (4-10)

$$\Delta_{0,9}(\kappa) = 1,6\sigma(\kappa) = 1,6\kappa \sqrt{(\varepsilon^2 - 1)^3/29n}; \quad (5-24)$$

$$\Delta_{0,9}(k) = 1,6 \sqrt{k \sqrt{kn \sqrt{\kappa^3}}}. \quad (5-25)$$

По формуле (5-23) вычисляются значения $k_{\max}$ и $k_{\min}$ путем подстановки вместо κ значений $\kappa + \Delta\kappa$ и $\kappa - \Delta\kappa$, где κ — оценка, полученная по выборке, и распределение признается экспоненциальным, если найденная по выборке оценка удовлетворяет условию

$$(k_{\min} - \Delta k) < k < (k_{\max} + \Delta k). \quad (5-26)$$

Для класса экспоненциальных распределений зависимость эксцесса от показателя экспоненты α теоретически выражается (см. с. 67) соотношением $\varepsilon = \Gamma(1/\alpha) \Gamma(5/\alpha) / [\Gamma(3/\alpha)]^2$, где $\Gamma(\cdot)$ — гамма-функция. Обратного решения этого трансцендентного уравнения относительно α в элементарных функциях не существует. Оно может быть представлено лишь тем или иным аппроксимирующим выражением. Для этой цели авторами этой книги в 1975 г. была предложена приближенная формула

$$\alpha = (1,355 - \kappa^2) / \ln [5/(9\kappa^2)], \quad (5-27)$$

где $\kappa = 1/\sqrt{\varepsilon}$, которая при $\varepsilon = 1,8 = 5/9$ для равномерного распределения дает $\alpha = \infty$. При изменении ε от 3 до 100, т. е. для α от 2 до $1/3$, она описывает искомую зависимость с погрешностью не более 1,3%, для $\alpha = 1/4$ и $\varepsilon = 458$ — с погрешностью 2,3%, а для $\alpha = 4$ и $\varepsilon = 2,188$ — с погрешностью 15%.

В 1985 г. И. В. Джунь (в статье в журнале «Кинематика и физика небесных тел», 1985, т. 1, № 4, с. 65) предложил более точное аппроксимирующее соотношение

$$\alpha = 1,4555 / [\ln(\varepsilon - 2/9) - 0,2939],$$

которое в диапазоне изменения ε от 3 до 458 (α от 2 до 0,25) обеспечивает погрешность аппроксимации не более 0,45% и лишь при $\alpha = 4$ дает погрешность до 5%, при $\alpha = 7$ — до 20% и при $\alpha = 20$ — до 60%.

Однако при $\varepsilon = 1,8$ для равномерного распределения эта формула вместо $\alpha = \infty$ дает $\alpha = 9$. Такое распределение (см. рис. 2-7, г) значительно отличается от равномерного. Но если в эту формулу ввести еще один уточняющий член $(-10,7/\varepsilon^7)$, то такая формула

$$\alpha = 1,46 / [\ln(\varepsilon - 2/9 - 10,7/\varepsilon^7) - 0,289] \quad (5-27a)$$

обеспечивает в диапазоне α от 4 до 0,25 (ε от 2,188 до 458) погрешность не более $\pm 0,3\%$, для $\alpha = 5 \div 7$ — около 1%, для $\alpha = 10 \div 20$ ($\varepsilon = 1,824 \div 1,884$) — до $\pm 6\%$, а при $\varepsilon = 1,8$ вместо $\alpha = \infty$ дает значение степени экспоненты $\alpha = 29$. Экспоненциальное распределение такой степени уже мало отличается от равномерного (по значению эксцесса на 2,2% и по значению энтропийного коэффициента 1,3%). Поэтому полученная формула может найти широкое практическое применение.

Границы $\alpha_{\min}$ и $\alpha_{\max}$ интервала неопределенности показателя степени α определяются по приведенным формулам путем подстановки в них в качестве значения ε пределов интервала его неопределенности в виде $\varepsilon - \Delta\varepsilon$ и $\varepsilon + \Delta\varepsilon$, где относительная погрешность значений эксцесса $\Delta\varepsilon/\varepsilon = 2\Delta\kappa/\kappa$, а $\Delta\kappa$ находится по формуле (5-24). Если условие (5-26) не выполняется, то распределение признается или двухмодальным, если $k < 1,87$, или уплощенным, если $k > 1,87$.

При $k > 1,87$ и признании распределения уплощенным оно рассматривается как композиция равномерного распределения со с. к. о. $\sigma_{\text{равн}}$ и экспоненциального распределения с $\sigma_{\text{эксп}}$. Относительное содержание той и другой составляющих численно определяется показателем содержания равномерной составляющей $C_p = \sigma_{\text{равн}}/\sigma_{\text{эксп}}$, изменяющимся от 0 до 5, так как при $C_p \geq 5$ отображающая точка на рис. 5-11, приближаясь к точке 10, опускается ниже границы $k = 1,87$, где кривые, соединяющие на рис. 5-11 точки 5—10, 6—10, ..., 9—10 и 4—10, сливаются между собой. В области $k > 1,87$ значение C_p может быть определено по соотношению, аппроксимирующему зависимость $C_p =$

$= f(k, \kappa)$ в области $0,2 < \kappa < 0,745$, т. е. от точки 7 до точки 10 с погрешностью менее 5% по формуле

$$C_p = 0,77 + 35,6 [0,8 (k_n - k)^2 + (k_n - k) (\kappa - 0,288) + 0,025\kappa^2], \quad (5-28)$$

где $k_n = \sqrt{\pi e/2}$. Показатель степени экспоненциальной составляющей композиции определяется по соотношению, аппроксимирующему в той же области зависимость $\alpha = f(k, \kappa)$ с погрешностью до 12%:

$$\alpha = \left[6,87 - 8,71 \frac{(1,57 - 1,9\kappa) (k - 1,87)}{(0,7332 - \kappa) \sqrt[3]{\kappa}} \right]^{-2/3}. \quad (5-29)$$

И, наконец, при невыполнении условия (5-26) при $k < 1,87$ исследуемое распределение признается двухмодальным и рассматривается как композиция дискретного двузначного распределения (точка 1 на рис. 5-11) и различных экспоненциальных распределений с α от 2 до $1/2$. Относительное содержание дискретной составляющей характеризуется показателем $C_d = \sigma_{\text{дискр}}/\sigma_{\text{эксп}} = a/\sigma_{\text{эксп}}$ и с погрешностью менее 5% может быть найдено по аппроксимирующему соотношению

$$C_d = 0,9 (\sqrt{\pi e/2} - k) + (\kappa - 0,06)/(0,775 - 0,46\kappa), \quad (5-30)$$

а показатель степени α определяется по аппроксимирующему (с погрешностью 6%) выражению

$$\alpha = 0,31 + \frac{2,12\kappa^2 - 2,8\kappa + 0,65}{8,83\kappa^2 - 9,34\kappa + 0,13 + k}. \quad (5-31)$$

При использовании соотношений (5-28)—(5-31) должно учитываться, как было показано применительно к соотношению (5-27), что используемые в них оценки k и κ были определены по выборке с погрешностями $\pm \Delta(k)$ и $\pm \Delta(\kappa)$. Эти погрешности вычисляются по соотношениям (5-24) и (5-25). С учетом интервалов неопределенности используемых оценок k и κ ответ ЭВМ будет: «Распределение представляет собой композицию равномерного и экспоненциального распределений с содержанием равномерной составляющей от $C_{p \min}$ до $C_{p \max}$ и экспоненциальной составляющей с показателем степени от $\alpha_{\min}$ до $\alpha_{\max}$ ». В зависимости от объема анализируемой выборки эти интервалы могут быть существенно различными и четко характеризуют область неопределенности полученного результата.

5.9. О ДОСТОВЕРНОСТИ ОПРЕДЕЛЕНИЯ ФОРМЫ РАСПРЕДЕЛЕНИЯ ПО МАЛОЙ ВЫБОРКЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Методы идентификации формы распределения экспериментальных данных представляются столь совершенными, что могут ввести в заблуждение относительно действительных возможно-

стей такой идентификации. Это относится особенно к методам аналитической аппроксимации полигонов (§ 5-4) и проверке выдвинутых гипотез критериями согласия (§ 5-5).

Применение аппроксимации полигонов распределений (рис. 5-2, 5-6 и 5-7) и проверка этих гипотез с помощью критериев согласия были лишь иллюстрацией, призванной помочь читателю в освоении этих методов. Одновременно отмечалось, что малость критериев согласия отнюдь не является показателем того, что по малой выборке (с $n = 40 \div 50$ отсчетов) достоверно определена функция распределения генеральной совокупности. Это обусловлено прежде всего тем, что аппроксимация и критерии согласия применяются к данной конкретной выборке экспериментальных данных. Если эксперимент повторить несколько раз, то каждый раз полигоны, а следовательно, и аппроксимирующие функции будут получаться разными. Вследствие этого достоверность идентификации вида распределений при $n < 200$ не может быть надежно оценена (см. конец § 5-5) критериями согласия. Ее можно оценить расчетом интервалов показателей формы распределения по соотношениям (5-24) и (5-25).

Пример оценки границ интервала неопределенности показателей формы кривой распределения. Гистограмма этого распределения была приведена на рис. 5-2. Параметры этого распределения, рассчитанные без поправок на группирование и объем выборки (см. табл. 5-7 в § 5-6) при объеме выборки $n = 52$, числе столбцов $m = 5$, ширине столбца $d = 8$, были получены равными $\sigma^* = 8,68$, $\epsilon^* = 2,17$ и $\kappa^* = 0,68$. С учетом поправки Шеппарда на группирование, согласно формуле (5-20), с. к. о. будет $\sigma = 8,68 \sqrt{1 - (8/8,68)^2/12} = 8,68 \cdot 0,964 = 8,37$. Четвертый момент, вычисленный без поправок, составляет $\mu_4^* = \sum X_j^4 n_j / n = 638\,976/52 = 12\,288$, а с поправкой на группирование согласно формуле (5-21) получает значение $\bar{\mu}_4 = 12\,288 - (8,37^2 \cdot 8^2)/2 + (7 \cdot 8^4)/240 = 10\,167$ (поправка существенна и составляет -17%), а дополнительно с поправкой на объем выборки согласно выражению (5.22) $\mu_4 = \bar{\mu}_4 (1 + 4/n) - 6\sigma^4/n = 10\,167 (1 + 4/52) - (6 \cdot 8,37^4)/52 = 10\,949 - 566 = 10\,383$, т. е. эта поправка незначительна и составляет всего $+2\%$. Отсюда значение эксцесса будет $\epsilon = \mu_4/\sigma^4 = 10\,383/8,37^4 = 2,12$ и контрэксцесса $\kappa = 1/\sqrt{\epsilon} = 1/\sqrt{2,12} = 0,69$. Согласно выражению (5-24) неопределенность оценки контрэксцесса вследствие малости выборки

$$\Delta_{0,9}(\kappa) = 1,6\sigma(\kappa) = 1,6\kappa \sqrt[4]{(\epsilon^2 - 1)^3/\sqrt{29n}} = \\ = 1,6 \cdot 0,69 \sqrt[4]{(2,12^2 - 1)^3/\sqrt{29 \cdot 52}} = 0,07.$$

Таким образом, интервал неопределенности оценки контрэксцесса этого распределения вследствие малости выборки с доверительной вероятностью $P = 0,9$ ограничен пределами: $\kappa_{\min} = 0,69 - 0,07 = 0,62$ и $\kappa_{\max} = 0,69 + 0,07 = 0,76$. (Отсюда,

кстати, видно, что поправки незначительно влияют на результат по сравнению с интервалом неопределенности и в неответственных случаях могут и не использоваться.)

При сохранении в силе гипотезы, что данное распределение описывается экспонентой с показателем степени $2 < \alpha < \infty$, полученные значения $\kappa_{\min} = 0,62$ и $\kappa_{\max} = 0,76$ должны быть интерпретированы следующим образом. Согласно соотношению (5-27) для экспоненциального распределения с $\kappa_{\min} = 0,62$ показатель степени экспоненты $\alpha = (1,355 - 0,62^2)/\ln [5/(9 \cdot 0,62^2)] = 2,6$, а для $\kappa_{\max} = 0,76$ значение α не существует, так как уже при $\kappa = \sqrt{5/9} = 0,745$ значение $\alpha = \infty$, ибо распределение становится равномерным. Таким образом, располагая выборкой $n = 52$, мы вправе лишь утверждать, что α лежит в пределах от 2,6 до ∞ и что значение κ генеральной совокупности лежит в пределах $0,62 < \kappa < 0,76$, т. е. распределение генеральной совокупности не является нормальным ($\kappa = 0,577$), а принадлежит к классу распределений, близких к трапецеидальным. Однако каков вид этой трапеции, установить при $n = 52$ нельзя, так как интервал возможных значений контрэксцесса ($0,62 < \kappa < 0,76$) шире, чем диапазон κ трапецеидальных распределений: от $\kappa = 0,645$ и $\varepsilon = 2,4$ (для треугольного) до $\kappa = 0,745$ и $\varepsilon = 1,8$ (для равномерного).

Чтобы интервал значений κ , соответствующий трапецеидальным распределениям, мог быть разделен хотя бы на три части ($0,645 \pm 0,025$ — распределения, близкие к треугольному, $0,695 \pm 0,025$ — трапецеидальные и $0,745 \pm 0,025$ — близкие к равномерному), необходимый объем выборки может быть найден из выражения (5-24):

$$n \geq 1,6^2 \kappa^2 \sqrt{(\varepsilon^2 - 1)^3} / [29 (\Delta \kappa)^2]. \quad (5-32)$$

Отсюда для рассматриваемой выборки с $\kappa = 0,69$, $\varepsilon = 2,12$ будет обеспечено $\Delta_{0,9}(\kappa) < 0,025$ при

$$n \geq 2,56 \cdot 0,69^2 \sqrt{(2,12^2 - 1)^3} / [29 (0,025)^2] \approx 440.$$

Подобным путем необходимый объем выборки может быть найден по формуле (5-32) в любом конкретном случае. Однако необходимо отметить, что увеличение объема выборки исходных данных до $n = 400$ или $n = 4000$ далеко не всегда обеспечит возможность надежной идентификации формы распределения экспериментальных данных. Следует иметь в виду, что все приведенные формулы были получены в предположении о непрерывности закона распределения и независимости каждого из получаемых отсчетов от положения на числовой оси уже полученных ранее. На практике эта непрерывность и независимость очень часто нарушаются, а в теории вероятностей нет критерия для оценки степени такого несоответствия.

Это свойство различных случайных величин на практике проявляется в следующем. При малой выборке исходных данных линейчатая диаграмма распределения (см. рис. 5-5) или его гистограмма при достаточно большом числе столбцов (см. рис. 5-2, а) оказываются сильно изрезанными, т. е. имеют много беспорядочно расположенных «пустых мест», не заполненных отсчетами. Иногда при увеличении объема выборки, скажем, с $n = 50$ до $n = 200$, эти провалы быстро и дружно заполняются, гистограмма получает достаточно плавный вид и без принудительного сглаживания и симметрирования. Но часто случайная величина имеет не такой «мягкий характер», а проявляет исключительное «упрямство», состоящее в том, что все последующие отсчеты (даже при увеличении выборки до 1000 или нескольких тысяч) ложатся точно на предыдущие. В итоге число отдельных линий линейчатой диаграммы остается тем же. Все новые отсчеты лишь добавляются в уже имеющиеся линии, а провалы между ними так и остаются незаполненными. Увеличение объема выборки хоть до 10 000 ничего не меняет, распределение остается таким же «гребенчатым», каким оно было при $n = 50$.

Ясно, что в этом случае распределение отсчетов не является непрерывным, и к ним не применимы понятия и формулы теории вероятностей, правомерные лишь для непрерывно распределенных отсчетов. В этих условиях неприменимы и формулы (5-24) и (5-25) для оценки погрешностей определения оценок, так как вычисленные по ним оценки погрешности (при $n = 10\,000$) будут лишь создавать иллюзию полного благополучия, тогда как в действительности положение может быть катастрофическим.

Это особенно характерно для измерительных каналов сложных ИВК, в которых в цепи многих преобразований может быть несколько квантующих устройств, а результат выдается с большим числом знаков.

Так, например, при одном из летных испытаний авиадвигателя ИВК выдавал на регистрацию частоту вращения роторов двигателя с шестью десятичными знаками. Испытатели были уверены, что частота вращения, измеряемая цифровым частотомером с погрешностью 10^{-6} , регистрируется идеально. Но при анализе 10 000 отсчетов оказалось, что 1000 из них имеют отклонение от номинала, равное $-200 \cdot 10^{-6}$, 5000 отсчетов — отклонение -100×10^{-6} , а остальные 4000 — отклонение $+200 \cdot 10^{-6}$, т. е. гистограмма состоит всего из трех (!) линий и, несмотря на то, что $n = 10\,000$, никаких заключений о виде закона распределения сделать было нельзя, а погрешность измерений составляла отнюдь не 10^{-6} , а всего $\pm 2 \cdot 10^{-4}$.

МЕТОДЫ ОБРАБОТКИ И ОЦЕНКИ ПОГРЕШНОСТЕЙ ПРИ ОДНОФАКТОРНОМ ЭКСПЕРИМЕНТЕ

6-1. ЦЕЛЬ И ОСОБЕННОСТИ ЭКСПЕРИМЕНТА ПО ОПРЕДЕЛЕНИЮ ФУНКЦИОНАЛЬНОЙ ЗАВИСИМОСТИ

На практике сама необходимость измерений большинства величин вызывается именно тем, что они не остаются постоянными, а изменяются в функции от изменения других величин. В этом случае целью измерения является установление вида функциональной зависимости $y = f(x)$. Для этого должны одновременно определяться как значения x , так и соответствующие им значения y , а задачей эксперимента является, как принято теперь говорить, установление *математической модели* исследуемой зависимости.

Определение математической модели включает в себя указание вида модели и определение значений ее параметров (коэффициентов, показателей степени и т. д.). Искомая функция может быть как функцией одной независимой переменной, так и функцией многих переменных. В современной теории эксперимента независимые переменные принято называть *факторами*, а зависимую переменную y — *откликом* (ГОСТ 24026—80). В соответствии с этим стандартом эксперимент по определению функции вида $y = f(x)$ принято именовать *однофакторным*, а эксперимент по определению функций вида $y = F(x_1, \dots, x_k)$ — *многофакторным*. В дальнейшем будем использовать терминологию ГОСТ 24026—80, а не ГОСТ 16263—70, подразделяющего такие измерения на *совместные* и *совокупные*. Вопросы, относящиеся к однофакторному эксперименту, будут рассмотрены в данной главе, а обработке и оценке погрешностей многофакторного эксперимента будет посвящена гл. 7.

Искомая математическая модель функциональной зависимости $y = f(x)$ может быть найдена лишь в результате совместной обработки всех полученных значений x и y . На рис. 6-1 это кривая, проходящая по центру полосы экспериментальных точек, которые могут и не лежать на искомой кривой $y = f(x)$, а занимают некоторую полосу вокруг нее. Эти отклонения вызваны погрешностями измерений, неполнотой модели и учитываемых факторов, случайным характером самих исследуемых процессов и другими причинами. Разделить погрешности, вызванные неточностью

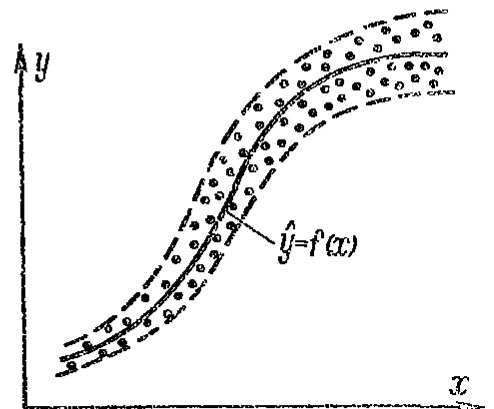


Рис. 6-1

измерения x и неточностью измерения y , невозможно, так как смещение точки на рис. 6-1, например, выше кривой могло быть вызвано как положительной погрешностью при измерении y , так и отрицательной погрешностью при измерении x . Поэтому описанием погрешности исходных данных может быть лишь указание ширины полосы их разброса вокруг найденной кривой зависимости $y = f(x)$. При этом полоса разброса экспериментальных данных необязательно будет иметь постоянную ширину по всей своей длине. Она может быть узкой вначале и расширяться в конце или, например, иметь узкий перешеек в средней части и расширяться по концам и т. п. Поэтому форма полосы погрешностей должна анализироваться в каждом отдельном случае.

6.2. ВЫБОР ВИДА МАТЕМАТИЧЕСКОЙ МОДЕЛИ И ОСТАТОЧНАЯ ПОГРЕШНОСТЬ АДЕКВАТНОСТИ

Задача выбора вида функциональной зависимости — задача неформализуемая, так как одна и та же кривая на данном участке примерно с одинаковой точностью может быть описана самыми различными аналитическими выражениями. Так, например, U-образная кривая может быть описана участком параболы, гиперболы, эллипса или синусоиды. Рациональный выбор того или иного аналитического описания может быть обоснован лишь при учете определенного перечня требований.

Главное требование к математической модели — это удобство ее последующего использования. Основное, что обеспечивает удобство математического выражения, — его компактность. Например, известно, что любую функцию $y = f(x)$ можно описать многочленом $y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k$. Но если же оказывается возможным с приемлемой точностью описать ее одночленом вида $y = a \ln(x/b)$, $y = ae^{bx}$, $y = a \sin bx$ и т. п., то ясно, что такое компактное представление много лучше. Таким образом, компактность модели достигается удачным выбором элементарных функций, обеспечивающих хорошее приближение при малом их числе.

Другое весьма желательное (но иногда трудно достижимое) требование — это содержательность, иначе говоря, интерпретируемость предлагаемого аналитического описания. Как правило, это достигается путем придания определенного смысла константам или функциям, входящим в найденную математическую модель.

Отсюда следует важный практический вывод. Даже в наш век широкого использования ЭВМ в научных исследованиях принятие решения о выборе той или иной математической модели остается за человеком-исследователем и не может быть передано ЭВМ. Только человек, а не ЭВМ, знает, для чего будет в дальнейшем использоваться эта модель, на основе каких понятий будут интерпретированы ее параметры и т. д.



Рис. 6-2

Быстрые методы установления графического вида однофакторных зависимостей. Основной помехой для установления вида исследуемой зависимости является случайный разброс экспериментальных данных.

Если случайный разброс координат x и y почти отсутствует или, как иногда говорят, диффузность исходных данных очень мала, то привлечение статистических методов для их обработки излишне и кривую можно просто провести через эти точки. Однако даже в этом случае не следует соединять нанесенные на график экспериментальные точки отрезками прямых линий, а провести через них плавную кривую. При проведении такой кривой (рис. 6-2, а) может оказаться, что одна или две точки все-таки не лежат на этой кривой и их следует рассматривать как возможные выбросы или промахи.

Если диффузность исходных данных значительна, т. е. вследствие случайного разброса отсчетов x и y точки на графике имеют существенный случайный разброс, то соединение их между собой отрезками прямых линий (рис. 6-2, б) просто бессмысленно и для обработки таких данных надо применять простейшие или более сложные статистические методы.

Одним из таких простейших экспресс-методов статистической обработки является метод обведения контура плавных границ полосы рассеяния экспериментальных точек. Если при этом для сохранения плавности этих границ какие-то из точек приходится оставить вне контура (рис. 6-2, в), то их следует рассматривать как возможные промахи или аномально большие случайные отклонения. Форма обведенной контуром полосы рассеяния экспериментальных точек чаще всего уже позволяет вынести суждение о характере функциональной зависимости $y = f(x)$. Для однозначного указания вида этой зависимости необходимо провести на глаз осевую линию этого контура.

Несмотря на исключительную простоту метода контура, он позволяет быстро указать желаемое положение и форму искомой кривой и провести ее не через какие-то отдельные точки, а сообразуясь с положением на графике всех экспериментальных точек в целом. Однако при большом рассеянии результатов эксперимента форма контура может иметь бессмысленные, случайные очертания. В этих условиях приходится ограничиваться установлением лишь уровня и наклона искомой зависимости, полагая

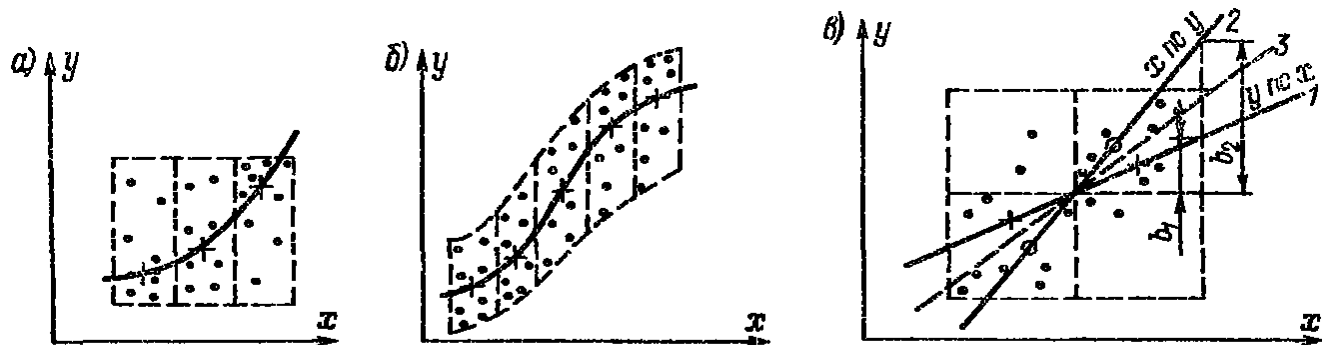


Рис. 6-3

ее прямой линией (рис. 6-2, г), проходящей по центру обведенной контуром полосы точек.

При очень большой диффузности экспериментальных данных, когда использование метода контура не дает ответа, может оказаться полезным *метод медианных центров*. Сущность этого метода поясняет рис. 6-3, а. Обведенное контуром поле точек делят на несколько частей, и в каждой из них находят медианный центр, т. е. пересечение вертикали и горизонтали слева и справа, и выше и ниже которых оказывается равное число точек. Затем через эти медианные центры проводят плавную кривую. Так как общее число отсчетов, как правило, не очень велико, то не следует стремиться к разделению поля точек на излишне большое число областей. Так, например, если поле точек на рис. 6-2, г решено описать прямой линией, для определения положения которой достаточно двух медианных центров, то и поле точек нужно разделить только на две равноценные области. Положение и форма кривых на рис. 6-3, а и б определяется соответственно тремя и пятью точками. Поэтому и поля точек должны быть разбиты не более чем на три и пять областей.

Быстрая оценка коэффициента корреляции и погрешности исходных данных также может быть произведена методом медианных центров. Для этого обведенное прямоугольным контуром поле исходных данных (рис. 6-3, в) вначале разбивается вертикальной границей на две равные по числу точек области, в каждой из которых находятся медианные центры, отмеченные на рисунке крестиками. Через них проводится прямая 1, являющаяся линией регрессии y по x .

Затем поле точек разбивается горизонтальной границей на две части с равным числом точек, и в каждой из них также находятся медианные центры, помеченные на рисунке незалитыми точками. Через эти центры проводится прямая 2, являющаяся линией регрессии x по y . (Эти понятия рассматриваются в пояснении рис. 6-10.)

Прямые 1 и 2 совпадают между собой и с прямой 3 лишь при коэффициенте корреляции между x и y $\rho_{xy} = 1$, когда все экспериментальные точки лежат строго на прямой 3, т. е. погрешность разброса исходных данных $\varphi = 0$. При $\varphi \neq 0$ и $\rho \neq 1$ прямая 1

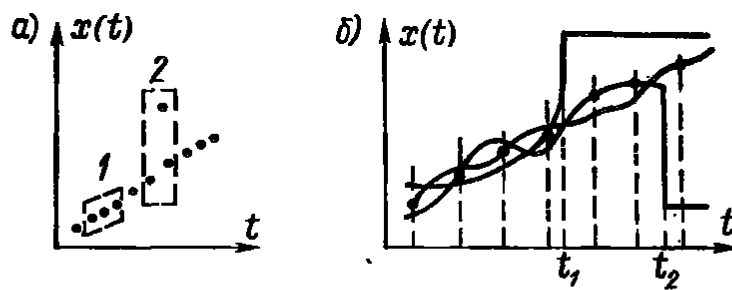


Рис. 6-4

идет положе прямой 3 на величину коэффициента корреляции ρ , а прямая 2 — во столько же раз круче. Поэтому если прямая 1 имеет уравнение $y = a_1 + b_1x$, а прямая 2 — уравнение $y = a_2 + b_2x$, то всегда $b_2 > b_1$ и $\rho = \sqrt{b_1/b_2}$, а относительная (приведенная к диапазону изменения y) погрешность исходных данных, как было отмечено в § 3-1, будет $\gamma = \sqrt{(1 - \rho^2)/4}$.

Так как для расчета ρ_{xy} нужны не сами значения угловых коэффициентов b_1 и b_2 , а лишь их отношение, то для расчета можно использовать просто отношение ординат b_1 и b_2 , отмеченных на рис. 6-3, в. У прямой 1 $b_1 = 9$, а у прямой 2 $b_2 = 25$, отсюда коэффициент корреляции поля точек на рис. 6-3, в $\rho_{xy} = \sqrt{9/25} = 0,6$, а разброс исходных данных характеризуется погрешностью $\gamma = \sqrt{(1 - 0,6^2)/4} = 0,4 = 40\%$.

Вряд ли возможно указать другой, более быстрый способ оценки этих параметров.

Влияние промахов, присутствующих в экспериментальных данных, также полностью определяется свойствами оценок центра распределения, положенных в основу используемых экспресс-методов. Так, оценка центра распределения в виде центра размаха исключительно чувствительна к наличию промахов. Поэтому при использовании метода контура промахи должны исключаться из исходных данных предварительно, как это и было сделано на рис. 6-2, в. Оценка же центра в виде медианы нечувствительна к промахам. Поэтому при использовании метода медианных центров удаления промахов производить не требуется.

Медианные методы хотя и не являются аналитическими, но легко алгоритмизируются и могут широко использоваться при машинной обработке данных. Для иллюстрации этого ограничимся двумя примерами.

Для устранения промахов в последовательно поступающих во времени данных с задержкой передачи всего на один отсчет очень эффективен метод использования медианы трех соседних отсчетов. При монотонно изменяющихся отсчетах (группа 1 на рис. 6-4, а) передается отсчет, занявший в группе медианное положение, т. е. средний по времени, а в группе с выбросом (группа 2) он заменяется последующим или предыдущим отсчетом. Алгоритм не срабатывает только при двух выбросах, следующих подряд.

Второй пример (рис. 6-4, б) — это выбор наиболее достоверного из трех параллельно действующих измерительных каналов. Выбранные по правилу медианы отсчеты помечены на рис. 6-4, б точками. Их достоверность гарантирована сверху и снизу отсчетами, полученными по двум другим каналам. Алгоритм продолжает работать, даже если в момент времени t_1 один из каналов «зашкалил», а другой — в момент времени t_2 — замкнулся.

Метод «выделения остатка». Этот метод оказывается одинаково полезным как при аналитическом, так и при графическом анализе исходных данных с целью определения вида подходящей для их описания математической модели. Он состоит в том, что рассматриваемую функцию $y = f(x)$ полагают состоящей из двух слагаемых: $y = f_0(x) + f_1(x)$. Например, для начального участка кривой на рис. 6-1 можно положить $y = y_0 + f_1(x)$. Затем выделить $f_1(x)$ как $f_1(x) = y - y_0$ и вновь положить этот остаток равным $f_1(x) = f_2(x) + f_3(x)$. Для кривой на рис. 6-1 $f_2(x) = a_1x^2$. Тогда можно построить остаток $f_3(x) = y - y_0 - a_1x^2$ и найти его форму и т. д. до тех пор, пока не будет найдено описание для последнего из таких остатков.

Погрешность адекватности модели. Это погрешность в описании данного явления, возникающая вследствие недостаточного соответствия аппроксимирующей функции всем особенностям формы экспериментальной кривой.

Для достижения все лучшего и лучшего соответствия модели описываемому явлению обычно приходится усложнять модель. Это приводит к противоречию между компактностью модели и точностью описания экспериментальных данных. Рациональное решение состоит в том, чтобы прекратить усложнение модели, когда она еще относительно проста, примирившись с приемлемой остаточной погрешностью адекватности.

Особое значение при этом имеет достоверность и достаточный объем исходных экспериментальных данных, позволяющий надежно установить степень случайности этих данных. Наиболее распространенной ошибкой начинающих исследователей являются попытки детерминированного описания, т. е. включение в математическую модель тех наблюдавшихся особенностей, которые в действительности являются случайными. Ориентиром при решении этого сложного вопроса можно считать примерное равенство остаточной погрешности адекватности принятой модели и ширины полосы ее неопределенности (см. § 6-8) вследствие случайного разброса исходных данных.

Более простым ориентиром для прекращения дальнейшего уточнения модели следует считать условие, что «принимаемая модель не противоречит данному полю экспериментальных данных». Эта формулировка «не противоречит» является наиболее правильной при таких заключениях, и поэтому именно ею необходимо пользоваться во всех подобных случаях.

6-3. ПОДБОР АППРОКСИМИРУЮЩИХ ФУНКЦИЙ

После того как каким-либо из описанных выше методов установлен графический вид искомой функциональной зависимости, задача состоит в выборе ее аналитического описания, т. е. в подборе аппроксимирующей функции. Для этого полезно иметь перед собой каталог графиков различных функций. Подборка графиков функций приводится в справочниках по математике. Поэтому ограничимся обсуждением лишь трех самых употребительных классов элементарных функций: степенных, показательных и дробно-рациональных.

Графики простейших степенных функций приведены на рис. 6-5. Если линия на графике не имеет кривизны, то это прямая $y = ax$, если кривизна имеет постоянный знак вдоль всей кривой, то это парабола четной степени $y = ax^2$, симметричная относительно оси y , или $y = \pm a\sqrt{x}$, симметричная относительно оси x . Для нечетных функций ($y = ax^3$ или $y = a\sqrt[3]{x}$) характерна S-образная форма с точкой перегиба в начале координат. По этим простейшим признакам и можно отобрать подходящую функцию.

Естественно, кривая может быть сдвинута от начала координат или повернута, например, так, как это показано на рис. 6-6, а и б. В этом случае надо так преобразовать переменные x и y , чтобы начало системы координат переместилось в вершину кривой, т. е. в точку x_0, y_0 . Для нахождения вершины кривой можно использовать простейший графический прием, показанный на рис. 6-6, б. Кривая пересекается несколькими параллельными прямыми, находятся центры отрезков AB, CD, EF и т. д. и на продолжении траектории перемещения этих центров находится вершина кривой.

Для проверки того, является ли данная функция именно степенной функцией вида $y = ax^n$, следует прологарифмировать ее вместе с аргументом, что возможно сделать лишь после выделения членов вида y_0 и $a(x - x_0)$. Получаемое после логарифмирования уравнение $\lg y = \lg a + n \lg x$ есть уравнение прямой в координатах $\lg y, \lg x$. По-

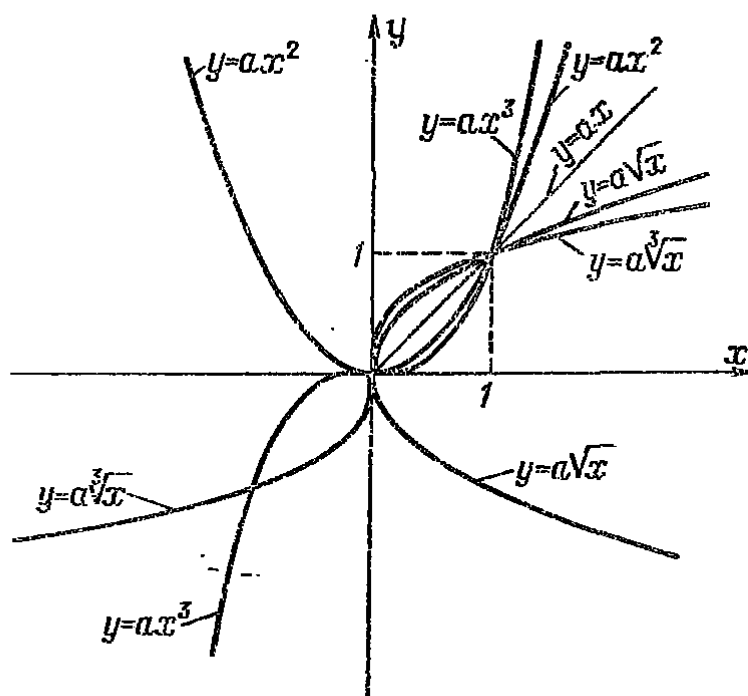


Рис. 6-5

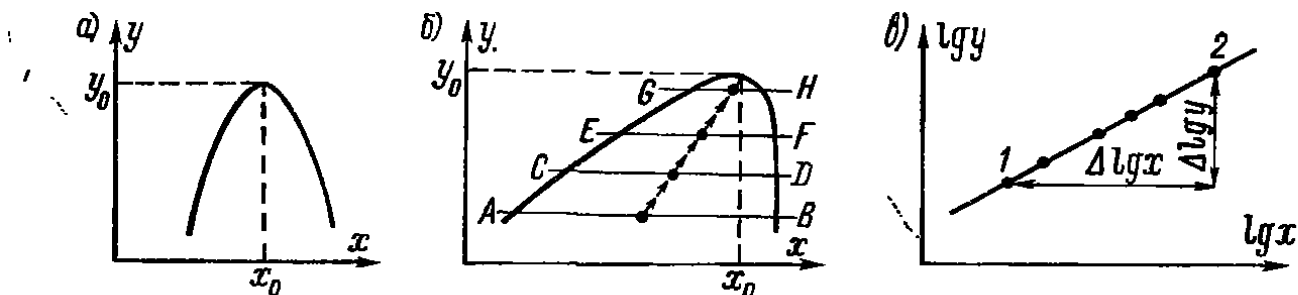


Рис. 6-6

этому проверка правильности выбора такой модели состоит в логарифмировании экспериментальных значений x и y и построении графика в осях $\lg x$ и $\lg y$ (рис. 6-6, в). Если поле экспериментальных точек на этом графике удовлетворительно группируется относительно прямой линии, т. е. модель в виде прямой не противоречит данному полю точек, то выбор данной функции может быть принят окончательно.

Построение графика 6-6, в позволяет кроме проверки годности принятой аппроксимации одновременно найти и входящие в нее параметры a и n . Для этого на графике достаточно выбрать любые две точки 1 и 2, через которые должна проходить аппроксимирующая кривая, т. е. должны одновременно удовлетворяться два уравнения:

$$\begin{cases} \lg y_1 = \lg a + n \lg x_1; \\ \lg y_2 = \lg a + n \lg x_2. \end{cases}$$

Отсюда

$$n = \frac{\Delta \lg y}{\Delta \lg x} = \frac{\lg y_2 - \lg y_1}{\lg x_2 - \lg x_1};$$

$$a = y_1/x_1^n \text{ или } a = y_2/x_2^n.$$

Графики показательных функций вида $y = ae^{bx}$ при разных значениях b приведены на рис. 6-7, а. Их характерной особенностью является постоянный знак кривизны как при $b > 0$, так и при $b < 0$. Для проверки соответствия показательной функции экспериментальным данным удобно построение графиков в полулогарифмическом масштабе, так как в координатах $\lg y$, x они образуют пучок прямых (рис. 6-7, б).

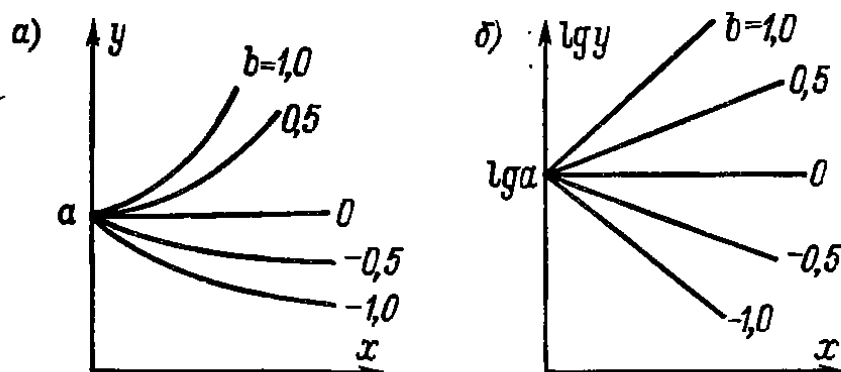


Рис. 6-7

Очень удобными при аппроксимации оказываются дробно-рациональные функции. Их практическая особенность состоит в том, что чаще всего исследователь не подозревает и не может догадаться, что интересующая его зависимость аппроксимируется именно этими функциями. Мы легко узнаем только простейшую равнобокую гиперболу, асимптотами которой служат оси координат (кривая 1 на рис. 6-8, а). Но если эта гипербола сдвинута (кривая 2) или перевернута (кривая 3), то мы уже ее не узнаем, а попытка аппроксимировать ее многочленом приводит к громоздкому выражению. По-

этому целесообразно взять себе за правило, прежде чем использовать аппроксимацию многочленом или экспонентами, обязательно проверить, не является ли искомая кривая простейшей, сдвинутой от начала координат гиперболой.

Проверка того, является ли данная кривая гиперболой, состоит в построении графика $1/y = f(x)$ (рис. 6-8, б). Если является, то экспериментальные точки ложатся на прямую, а ее продолжение до пересечения с осями x и y позволяет графически определить неизвестные коэффициенты a , b или c .

При этом замена координаты y на $1/y$ (и соответственно x на $1/x$) допустима лишь в том случае, если сдвиг по этой координате отсутствует, т. е. другая ось является асимптотой такой гиперболы. Это вызвано тем, что при сдвиге, например, вдоль оси x аналитическое описание гиперболы имеет вид $y = a/(b - x)$. Обращение оси y приводит к зависимости вида $1/y = (b - x)/a = b/a - x/a$, т. е. прямой в координатах $1/y$ и x . При сдвиге же вдоль оси y уравнение гиперболы имеет вид $y = c + a/x$, а следовательно, $1/y = 1/(c + a/x) = x/(a + cx)$ не есть прямая линия.

Особенно сложен для проверки случай, когда гипербола сдвинута одновременно по обеим осям (на величину c по оси y и на величину b по оси x), т. е. имеет уравнение $y = c + a/(x - b)$ или в неявном виде $(x - b)(y - c) = a$. В этом случае приходится двигаться методом последовательных приближений: задавать ряд возможных значений b , вычислять значения $1/(x - b)$ и остановиться на том значении b , когда $y = c + a/(x - b)$ в координатах y и $1/(x - b)$ даст расположение точек, наиболее близкое к прямой линии. Или же методом наименьших квадратов одновременно найти коэффициенты a , b , c в уравнении $(x - b)(y - c) = a$.

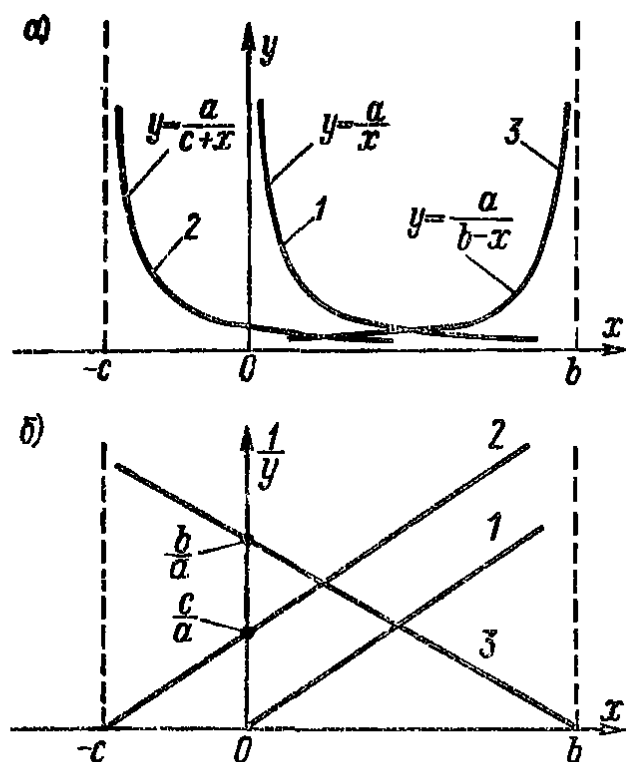


Рис. 6-8

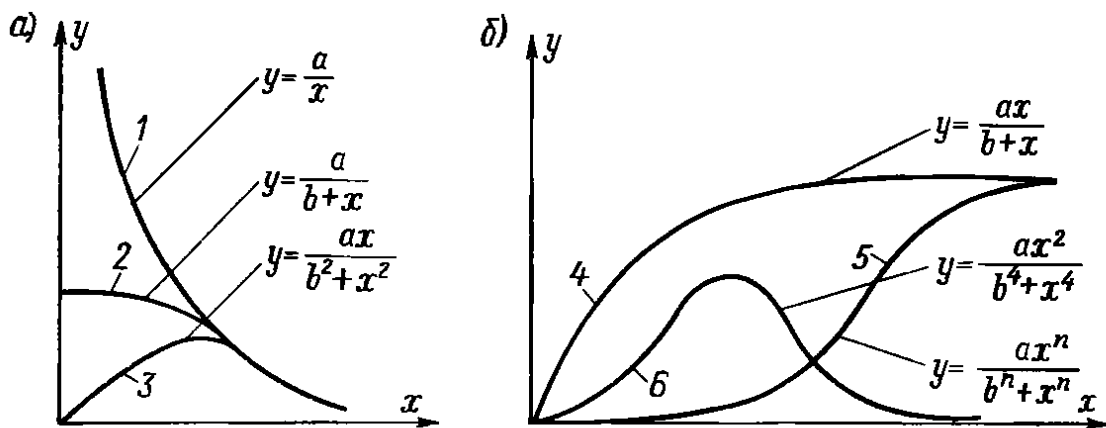


Рис. 6-9

В качестве примера дробно-рациональной функции рассмотрим функцию вида

$$y = \frac{ax^m}{b^n + x^n}.$$

(Задавать постоянное слагаемое в знаменателе в виде b^n удобнее для последующей интерпретации, так как в этом случае постоянная b имеет размерность x .) При $m = 0$, $b = 0$ и $n = 1$ эта функция — простая гипербола вида $y = a/x$ (кривая 1 на рис. 6-9, а); при $m = 0$ и $n = 1$ — кривая 2 вида $y = a/(b + x)$, при $m = 1$ и $n = 2$ — кривая 3 на рис. 6-9, а; при $m = 1$ и $n = 1$ это кривая 4 на рис. 6-9, б вида $y = ax/(b + x)$, вначале возрастающая линейно с увеличением x , а затем стремящаяся к постоянному значению, равному a , при $n = m$ — кривая 5, переходящая с уровня $y = 0$ на уровень $y = a$ с крутизной, зависящей от значения $n = m$, и, наконец, при $m = 2$ и $n = 4$ — колоколообразная кривая 6, сначала возрастающая пропорционально x^2 , а затем спадающая обратно пропорционально x^2 . Таким образом, функции этого вида могут оказаться полезными в самых разнообразных случаях аппроксимации экспериментальных кривых.

6.4. РАСЧЕТ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ ПАРАМЕТРОВ ВЫБРАННОЙ АППРОКСИМИРУЮЩЕЙ ФУНКЦИИ

Если подбор вида аппроксимирующей функции, как это подчеркивалось выше, процесс неформальный и не может быть полностью передан ЭВМ, то расчет параметров аппроксимирующей функции выбранного вида — операция чисто формальная и ее следует осуществлять на ЭВМ. Более того, это трудный и утомительный расчет, в котором человек не застрахован от ошибок, а ЭВМ выполнит его быстрее и качественнее.

В общем случае этот расчет состоит в решении системы нелинейных уравнений. В частных случаях это может быть: система уравнений, линейных относительно искомых параметров, система

Используя подобные преобразования и подстановки, необходимо, правда, иметь в виду следующее. При графической аппроксимации или аппроксимации методом наименьших квадратов мы стремимся минимизировать (уравнять положительные и отрицательные значения) абсолютные погрешности. При этом относительные погрешности в начале и конце диапазона изменения переменных оказываются существенно различными.

Если после этого производят подстановки вида $X = 1/x$ или $Y = 1/y$, то начало и конец диапазона меняются местами, а следовательно, меняются местами и погрешности. Эти преобразования погрешностей следует контролировать, в противном случае они могут привести к большим неточностям.

Покажем это на конкретном примере. Пусть рассчитывалась аппроксимация кривой вида $y = a/x$ и в ее нижнем конце текущее значение y имело порядок 0,1 единицы, а в ее верхней части достигало значения 10 единиц. После преобразования $Y = 1/y$ прямая $Y = x/a$ в верхней части получит значения $Y = 1/0,1 = 10$, а в нижней части — значения $Y = 0,1$. Пусть мы аппроксимировали ее с абсолютной погрешностью, как в верхней, так и в нижней части, равной 0,02. Но в верхней части прямой эта ошибка соответствует относительной погрешности $0,02/10 = 0,002 = 0,2\%$, а в нижней — $0,02/0,1 = 0,2 = 20\%$. Поэтому после возвращения в координаты $y = a/x$ нижняя часть кривой будет аппроксимирована с абсолютной погрешностью, равной $0,1 \cdot 0,2\% = 0,02/100 = 0,0002$ единицы, в то время как верхняя часть кривой будет аппроксимирована с погрешностью $10 \cdot 20\% = 200/100 = 2$ единицы, т. е. в 10 000 раз большей.

Обратив на это внимание, положение легко исправить, но если подобное явление останется без соответствующего анализа получающихся погрешностей, то может быть причиной серьезных недоразумений.

Графоаналитические методы определения параметров выбранной аппроксимирующей функции благодаря своей простоте и, особенно, наглядности не потеряли своего значения при анализе однофакторных зависимостей даже при наличии больших ЭВМ и карманных программируемых калькуляторов. Они позволяют очень быстро получить приближенные значения параметров, а тем самым за малое время проверить несколько вариантов, и если принятая модель оказывается неподходящей, то графическое построение наглядно показывает, каким образом следует ее изменить или дополнить.

На графике однозначно можно опознать только прямую линию. Только прямую линию при помощи линейки можно продлить на достаточно большое расстояние. Другие кривые такими свойствами не обладают. Поэтому при построении экспериментальных данных координатные оси модели следует преобразовывать до тех пор, пока не будет получено уравнение прямой линии.

Рассмотрим простейший пример. Пусть экспериментально по-

лученные значения x и y располагаются на графике так, что нами было принято решение описать их в среднем моделью вида $y = ax/(1 - bx)$.

Вычислить значения постоянных a и b можно по любым двум из экспериментальных точек. Но этого делать не следует, так как из-за случайного разброса каждая пара экспериментальных точек будет давать свои значения для a и b . Полезнее преобразовать это уравнение в уравнение прямой, например, так: $y - bxy = ax$ и $z = y/x = a + by$. Тогда, вычислив для каждой точки значение $z = y/x$ и нанеся эти точки в координатах (z, y) , мы будем иметь возможность провести через полосу этих точек (на глаз или методом медианных центров) среднюю прямую, которая отсечет на оси z значения a , а по ее наклону можно будет рассчитать значение b . Особенно эффективно использование графоаналитических методов определения искомых параметров тогда, когда обратное преобразование в элементарных функциях не выражается и аналитическое решение невозможно.

Аналитические методы. Как видно из примеров систем уравнений [см., например, систему (6-1)], число независимых уравнений системы равно числу n поставленных опытов.

С другой стороны, для определения k коэффициентов необходимо не менее k независимых уравнений. Но если число n поставленных опытов и число независимых уравнений равно числу искомых коэффициентов, то решение системы е д и н с т в е н н о, а следовательно, случайно, так как точно соответствует случайным значениям исходных данных.

При числе опытов n , большем, чем число k искомых коэффициентов, число независимых уравнений системы избыточно. Из этих уравнений в разных комбинациях можно составить несколько систем уравнений, каждая из которых в отдельности даст свое решение. Но между собой они будут н е с о в м е с т н ы. Каждое решение будет соответствовать своей аппроксимирующей функции. Если все их построить на графике, то получим целый пучок аппроксимирующих кривых.

Это открывает при $n > k$ совершенно новые возможности. Во-первых, этот пучок кривых показывает форму и ширину области неопределенности проведенного эксперимента. Во-вторых, может быть произведено усреднение всех найденных кривых. Полученная усредненная кривая будет гораздо точнее и достовернее описывать исследованное явление, так как она в значительной степени освобождена от случайных погрешностей, приводивших к разбросу отдельных экспериментальных точек.

Таким образом, проведение большего числа опытов n , чем минимально необходимое для расчета коэффициентов $n = k$, позволяет определить область неопределенности по фактическим данным эксперимента и одновременно существенно уменьшить случайную погрешность окончательного результата путем усреднения.

6-5. РЕГРЕССИОННЫЙ АНАЛИЗ. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Усреднение несовместных решений избыточной системы уравнений может быть произведено различными способами (на глаз, методом медианных центров и т. п.). Наиболее мощный метод был разработан в 1795—1805 гг. Лежандром и Гауссом и получил название регрессионного анализа, или метода наименьших квадратов (МНК). Таким образом, МНК — это «новинка» 180-летней давности. Но теперь благодаря возможности широкого доступа исследователей к ЭВМ этот метод получил, по существу, второе рождение. Это произошло потому, что вычисления по МНК чрезвычайно громоздки.

С появлением легкого доступа к ЭВМ положение коренным образом изменилось. Программа обработки данных МНК содержится в математическом обеспечении любой ЭВМ. Поэтому практически надо только ввести свои данные в соответствии с этой программой и дождаться готового ответа. Но чтобы представлять себе, что делать с нашими данными ЭВМ по программе МНК, рассмотрим сущность МНК на примере простейшей задачи.

Сущность МНК. Пусть после предварительного анализа, описанного в предыдущем параграфе, была выбрана модель вида $\hat{y} = a_0 + a_1x + a_2x^2$. Теперь задача состоит лишь в том, чтобы найти наилучшие значения параметров модели a_0 , a_1 и a_2 . Значения же x_i и y_i , наоборот, нам известны. Это не переменные, а конкретные числа, полученные в наших опытах. Поэтому любая функция от x (будь то x^2 , $\sqrt{x}$, $\ln x$, $\sin x$ и т. д.) при известном x — это тоже определенное число. Отсюда, введя обозначения $x = x_1$ и $x^2 = x_2$, можно сделать подстановку, после чего наша модель получит вид $\hat{y} = a_0 + a_1x_1 + a_2x_2$.

Между рассчитанными по модели значениями $\hat{y}_i$ и экспериментальными отсчетами y_i будут наблюдаться отклонения. Введем для них обозначение $\Delta_{yi} = y_i - \hat{y}_i$. МНК позволяет найти такие значения искомых параметров моделей a_0 , a_1 и a_2 , при которых

сумма по всем n точкам $\sum_{i=1}^n \Delta_{yi}^2 \rightarrow \min$.

Если это описать формально, т. е. взять поочередно частные производные по a_0 , a_1 , a_2 и приравнять их нулю, то получим систему из трех уравнений, решением которой и будут искомые a_0 , a_1 и a_2 . Итак,

$$\Delta_{yi}^2 = (y_i - \hat{y})^2 = (y_i - a_0 - a_1x_{1i} - a_2x_{2i})^2;$$
$$\sum_{i=1}^n \Delta_{yi}^2 = \sum_{i=1}^n (y_i - a_0 - a_1x_{1i} - a_2x_{2i})^2,$$

производная по a_0

$$\partial \left(\sum_{i=1}^n \Delta_{yi}^2 \right) \bigg| \partial a_0 = -2 \sum_{i=1}^n (y_i - a_0 - a_1x_{1i} - a_2x_{2i}) = 0,$$

и первое уравнение системы есть

$$a_0 n + a_1 \sum_{i=1}^n x_{1i} + a_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i.$$

Производная по a_1

$$\partial \left(\sum_{i=1}^n \Delta_{yi}^2 \right) / \partial a_1 = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) x_{1i} = 0,$$

и второе уравнение системы

$$a_0 \sum_{i=1}^n x_{1i} + a_1 \sum_{i=1}^n x_{1i}^2 + a_2 \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n x_{1i} y_i$$

и т. д.

Полная система так называемых нормальных уравнений для расчета параметров модели будет:

$$\begin{cases} a_0 n + a_1 \sum x_{1i} + a_2 \sum x_{2i} = \sum y_i; \\ a_0 \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i} x_{2i} = \sum x_{1i} y_i; \\ a_0 \sum x_{2i} + a_1 \sum x_{1i} x_{2i} + a_2 \sum x_{2i}^2 = \sum x_{2i} y_i. \end{cases} \quad (6-4)$$

Для решения этой системы ЭВМ по введенным экспериментальным данным: $x_1, \dots, x_n$ и $y_1, \dots, y_n$ с учетом введенных подстановок, должна подготовить все коэффициенты этой системы, т. е. подсчитать n , $\sum x_{1i}$, $\sum x_{2i}$, $\sum x_{1i}^2$, $\sum x_{2i}^2$, $\sum x_{1i} x_{2i}$, $\sum y_i$, $\sum x_{1i} y_i$ и $\sum x_{2i} y_i$, а затем произвести решение этой системы уравнений и вывести на печать полученные значения параметров модели a_0 , a_1 и a_2 . (В программах современного математического обеспечения ЭВМ это делается с использованием специальных методов, не требующих решения системы.)

Линия, которая определяется МНК, называется *линией регрессии*, а коэффициенты a_1 , a_2 и т. д. называются *коэффициентами регрессии* y по x_1 , y по x_2 и т. д.

Свойства МНК. Для наглядного уяснения свойств МНК полезно рассмотреть еще более простую линейную модель, когда $y = a_0 + a_1 x$.

Для нее система нормальных уравнений будет иметь вид:

$$\begin{cases} a_0 n + a_1 \sum_1^n x_i = \sum_1^n y_i; & a_0 \sum_1^n x_i + a_1 \sum_1^n x_i^2 = \sum_1^n x_i y_i, \end{cases} \quad (6-5)$$

а ее решение

$$a_0 = \left[\sum_1^n x_i^2 \sum_1^n y_i - \sum_1^n x_i \sum_1^n x_i y_i \right] / \left[n \sum_1^n x_i^2 - \left(\sum_1^n x_i \right)^2 \right];$$

$$a_1 = \left[n \sum_1^n x_i y_i - \sum_1^n x_i \sum_1^n y_i \right] / \left[n \sum_1^n x_i^2 - \left(\sum_1^n x_i \right)^2 \right].$$

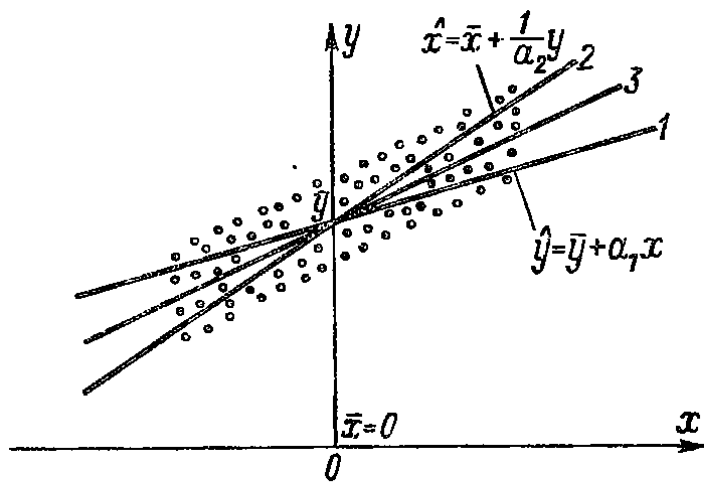


Рис. 6-10

Если первое уравнение этой системы разделить на n , то получим

$$a_0 + a_1 \sum_1^n x_i/n = \sum_1^n y_i/n,$$

$$\text{т. е. } a_0 + a_1 \bar{x} = \bar{y}. \quad (6-6)$$

Таким образом, первое уравнение системы уравнений МНК требует, чтобы линия регрессии проходила через точку с координатами

$(\bar{x} \text{ и } \bar{y})$, т. е. через центр тяжести поля экспериментальных точек (рис. 6-10). Это свойство МНК еще более наглядно выступает в том случае, если значение $\bar{x}$ принять за нуль, т. е. перенести начало координат в точку $\bar{x}$. При $\bar{x} = 0$ уравнение (6-6) дает $a_0 = \bar{y}$.

Следовательно, можно утверждать, что первое уравнение системы МНК определяет средний уровень искомой линии регрессии. *

Второе уравнение системы (6-5) при $\bar{x} = \sum_1^n x_i/n = 0$ дает

$$a_1 \sum_1^n x_i^2 = \sum_1^n x_i y_i, \text{ откуда}$$

$$a_1 = \frac{\sum_1^n x_i y_i}{\sum_1^n x_i^2}. \quad (6-7)$$

Таким образом, второе уравнение системы (6-5) определяет коэффициент a_1 наклона искомой прямой относительно оси x . При большем числе независимых переменных $x_1, x_2, \dots, x_l$ решение системы (6-4) нормальных уравнений МНК есть плоскость в многомерном пространстве. Первое уравнение этой системы определяет, как и выше, совмещение этой плоскости с центром тяжести поля экспериментальных точек, а остальные уравнения системы совместно определяют коэффициенты регрессии, т. е. коэффициенты наклона этой плоскости к осям $x_1, x_2, \dots, x_l$.

Если до проведения решения МНК все исходные данные центрировать, т. е. найти $\bar{x}$ и $\bar{y}$ и перенести начало системы координат в центр тяжести поля экспериментальных точек (в точку $\bar{x}, \bar{y}$ на рис. 6-10), то в этих новых координатах будет $\bar{x} = 0$ и $\bar{y} = 0$. В этом случае из полной системы нормальных уравнений МНК (6-4) опускаются как первое уравнение (оно уже оказывается удовлетворительным, т. е. обращается в тождество),

так и первый столбец коэффициентов (благодаря $a_0 = 0$). В этом случае система нормальных уравнений получает вид:

[illegible]

Такой вид системы нормальных уравнений МНК, во-первых, является более общим, так как система вида (6-4) получается из (6-8), если положить все $x_{1i} = 1$. Во-вторых, если $\bar{y}_i = a_0$ и $\bar{x}_{ji} = \bar{x}$ находятся заранее, то число определяемых коэффициентов уменьшается на единицу. Это имеет существенное практическое значение, так как решение МНК для трех искоемых коэффициентов можно запрограммировать на настольных микрокалькуляторах (типов БЗ-34, МК-52 и др.), а для четырех и более коэффициентов — только на ЭВМ, обладающих большой емкостью памяти. И, в-третьих, система вида (6-8) удобна во всех тех случаях, когда заранее известно, что поверхность отклика должна проходить через какую-то определенную точку, например начало координат.

Особенность МНК состоит в том, что полученные этим методом решения необратимы. Это обусловлено тем, что, например, коэффициент a_1 регрессии y по x , входящий в модель $\hat{y} = y_0 + a_1x$, определяется соотношением (6-7). Если же вычисляется обратная регрессия x по y , то коэффициент $1/a_2$ в модели $\hat{x} = x_0 + y/a_2$ вычисляется согласно соотношению $1/a_2 = \sum_1^n x_i y_i / \sum_1^n y_i y_i$, а это уже другое, отличное от (6-7), соотношение.

Естественно, что $a_1 \neq a_2$, поэтому произведение $a_1 (1/a_2) \neq 1$, а равно величине

$$a_1 \frac{1}{a_2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^2} = \rho^2, \quad (6-9)$$

корень квадратный из которой и называется коэффициентом взаимной корреляции значений x_i и y_i . Следствием этого является то, что линии регрессии y по x , т. е. $\hat{y} = y_0 + a_1x$ и x по y , т. е. $\hat{x} = x_0 + y/a_2$, не совпадают с осью эллипса рассеяния экспериментальных точек (см. рис. 6-10). Линия регрессии y по x проходит более полого (ближе к оси x), чем ось эллипса, а линия регрессии x по y проходит более круто (ближе к оси y) на ту же величину. Поэтому угловой коэффициент оси симметрии эллипса

может быть выражен как a_1/ρ или как $a_2\rho$, а отношение этих величин равно единице: $\frac{a_1}{\rho} \frac{1}{a_2\rho} = 1$, откуда $a_1/a_2 = \rho^2$.

При коэффициенте корреляции ρ , близком к 1, т. е. при малом рассеянии экспериментальных точек и большой протяженности поля точек, обе линии регрессии близки к оси эллипса и их различием можно пренебречь. При малом же коэффициенте корреляции ($\rho \ll 0,96$) это различие становится весьма существенным, а сам МНК — неэффективным. В этом случае целесообразно переходить от МНК к методу ортогональной регрессии (см. § 7-4).

И, наконец, в качестве существенной особенности МНК следует отметить, что МНК минимизирует абсолютные значения погрешностей, так как в его основу было положено условие $\sum \Delta^2 y_i \rightarrow \min$. Следствия этого обсуждались в § 6-4.

6-6. ЧУВСТВИТЕЛЬНОСТЬ МНК

К НЕОДНОРОДНОСТИ СТАТИСТИКИ

И ОБУСЛОВЛЕННОСТЬ ПОЛУЧАЕМЫХ РЕШЕНИЙ

Метод наименьших квадратов — формальный, так сказать, «слепой». В ЭВМ вводятся десятки или сотни произведенных отсчетов, она их в течение какого-то времени обрабатывает и выдает значения параметров заданной модели. Эти коэффициенты чаще всего определяют функцию, наилучшим образом описывающую экспериментальные данные, но иногда дают просто абсурдные решения, ставящие экспериментатора в тупик. Эти особенности МНК необходимо рассмотреть более подробно.

Неоднородность статистики экспериментальных данных. Это наиболее часто встречающаяся на практике причина возникновения абсурдных решений при использовании МНК. Рассмотрим эти явления на простом примере. Пусть исследовалась зависимость от температуры Θ электрического сопротивления $R(\Theta)$ медной обмотки какого-либо устройства. Предполагалось, что $R(\Theta) = R_0(1 + \alpha\Theta)$, где R_0 — сопротивление при $\Theta = 0$, а α — температурный коэффициент. Были произведены многократные измерения $R(\Theta)$ при температуре Θ , равной 20, 40, 60 и 80 °C. При правильном решении эти данные должны выглядеть так, как это показано на рис. 6-11, а, и решение МНК должно было дать $R_0 = 100$ Ом и $\alpha = +0,4\%/K$.

Однако ЭВМ вывела на печать ответ в виде $R_0 = 141,015$ Ом и $\alpha = -0,410791\%/K$. Ясно, что это абсурдный ответ, так как медная обмотка не может иметь отрицательный температурный коэффициент сопротивления. Для того чтобы понять причину такого решения, целесообразно вывести на печать не только значения искомых R_0 и α , но и исходные данные, с которыми фактически оперировала ЭВМ, а затем вручную или даже с помощью графопостроителя ЭВМ нанести эти данные на график.

После построения такого графика чаще всего причина неверного решения становится ясной. Пусть эти данные на графике расположились так, как это показано на рис. 6-11, б, т. е. данные, относящиеся к Θ , равной 60 и 80 °С, были восприняты ЭВМ, как относящиеся к Θ , равной 20 и 40 °С, и наоборот. Причина подобного явления может быть самой различной (например, неверно помечены и введены при перфорации протоколы измерений или присутствует систематическая ошибка при проведении части измерений и т. п.), но итог состоит в том, что для решения МНК была представлена неоднородная статистика. Как видно из рис. 6-11, б, она состоит из двух групп данных, помеченных цифрами 1 и 2. Если проводить решения МНК для этих групп отдельно, то для каждой из них будет получен $\alpha = +0,4\%/K$, так как линии 1 и 2 приблизительно параллельны линии 4 правильного решения.

Совместное же решение МНК этих двух групп данных дает линию регрессии с $R_0 \approx 140$ Ом и $\alpha \approx -0,4\%/K$, помеченную цифрой 3 и проходящую через центры тяжести этих групп.

Причиной подобного решения могла быть и другая ошибка. Пусть, например, при вводе данных вместо $\Theta = 80$ было введено $\Theta = 8,0$. Тогда на графике такие исходные данные будут выглядеть так, как это показано на рис. 6-11, в, т. е. будут состоять из однородной статистики для температур 20, 40 и 60 °С, помеченной цифрой 2, и массового промаха, помеченного цифрой 1. Ясно, что рассчитанная по таким данным линия регрессии 4 будет проходить через центры тяжести этих двух групп данных и соответствовать абсурдному решению.

Ситуация может оказаться близкой к описанной даже в том случае, если в области 1 (рис. 6-11, в) в виде

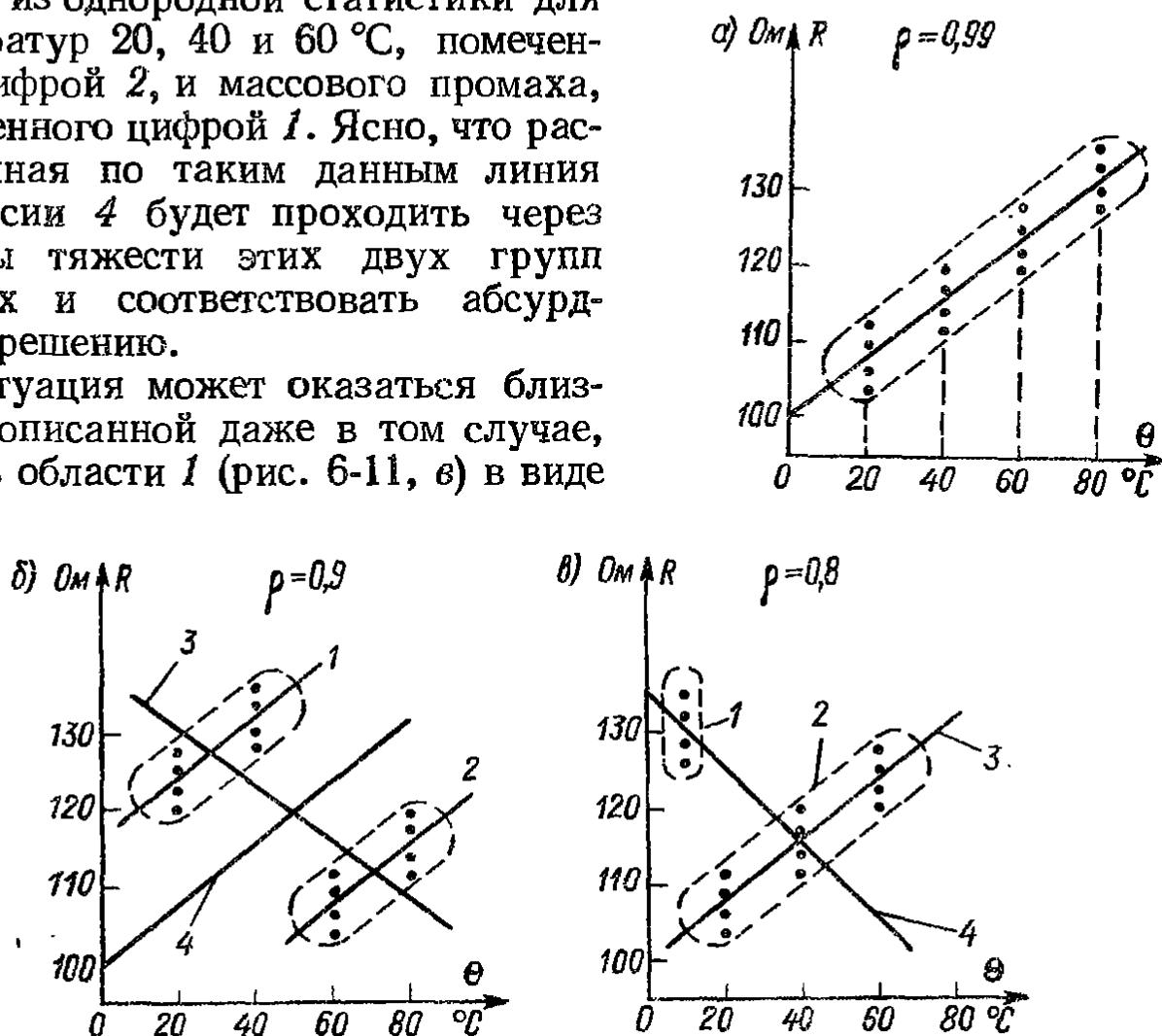


Рис. 6-11

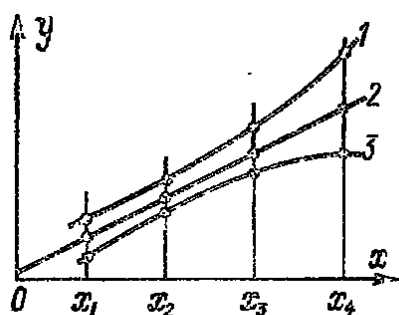


Рис. 6-12

промаха будет всего одна точка. На основании рассмотренного следует заключить, что МНК очень чувствителен к неоднородности статистики исходных данных и может дать абсурдное решение даже в том случае, если эта неоднородность вызвана наличием всего одного, далеко отстоящего промаха. Эту особенность МНК надо всегда иметь в виду. Практически это означает, что окончательным должно

приниматься решение МНК по однородной, очищенной от промахов статистике.

Для того чтобы судить об абсурдности или приемлемости данного ответа, весьма целесообразно еще до решения МНК узнать любым приближенным, графическим методом (или из априорных предположений), каков должен быть порядок искомых величин. При получении решения, которое вызывает недоумение, целесообразно построить (вручную или на графопостроителе ЭВМ) графики, подобные рис. 6-11. Если получим график, подобный изображенному на рис. 6-11, б или в, необходимо принять решение, какую же из групп данных (1 или 2) мы предполагаем обрабатывать. При графике рис. 6-11, в эта задача решается проще, так как при отбрасывании данных области 1 обработка данных области 2 сразу дает правильное решение в виде линии регрессии 3. С графиком, изображенным на рис. 6-11, б, дело обстоит сложнее, так как обе группы данных (и 1, и 2) являются ошибочными. В подобном случае полезно разделить отдельно обработать каждую из них: они дадут совпадающие коэффициенты регрессии $\alpha = +0,4\%/K$, но разные R_0 (область 1 даст $R_0 = 115 \text{ Ом}$, а область 2 даст $R_0 = 95 \text{ Ом}$). Это и укажет на то, где искать ошибку.

Обусловленность (устойчивость) решений, получаемых МНК. Решение задачи, получаемое МНК по экспериментальным данным, содержащим случайные ошибки, естественно, само также случайно. Благодаря усреднению результатов многократных отсчетов оно становится более определенным, более устойчивым.

Так, например, линия регрессии на рис. 6-11, а была получена путем усреднения координат 16 точек (по четыре точки в четырех сечениях). Спрашивается, как изменится положение линии регрессии, если ее определять не по 16, а по 15 точкам, опустив одно наблюдение в сечении $\Theta = 40^\circ\text{C}$? В данном случае положение линии регрессии останется почти тем же самым. Это и доказывает устойчивость (обусловленность) данного решения. Но может быть и по-другому. Представим себе, что обрабатываемые исходные данные располагаются так, как это показано на рис. 6-12, а определению подлежат коэффициенты a_0 , a_1 и a_2 модели $y = a_0 + a_1x + a_2x^2$.

Пусть в каждом из четырех сечений было определено по три экспериментальные точки и при обработке по МНК всех 12 точек

были получены какие-то значения a_0 и a_1 , а значение коэффициента a_2 оказалось равным нулю, т. е. точки 1 и 3 в сечении x_2 лежат симметрично относительно точки 2. Если же теперь отбросить точку 1, то обработка оставшихся 11 точек даст примерно те же значения a_0 и a_1 , но a_2 будет отрицательно, а если, наоборот, отбросить точку 3, то a_2 примет положительное значение. Таким образом, оказывается, что расчетное значение коэффициента a_2 неустойчиво, а значения коэффициентов a_0 и a_1 устойчивы. Подобная плохая обусловленность решения указывает, как правило, границу переусложнения модели.

Абсурдные решения могут возникать и по другим причинам. В МНК на определенном этапе производится решение системы уравнений, в ходе которого всегда производится вычитание чисел, которые могут оказаться большими и близкими между собой. Тогда из-за малейшей ошибки в исходных данных их разность может легко измениться в несколько раз и решение будет плохо обусловлено. Покажем это на примере.

Пусть используется модель вида $y = a_1x_1 + a_2x_2$. Для того чтобы вычислить a_1 и a_2 , надо задать один раз какие-то x_1 и x_2 и измерить y и второй раз задать другие значения x_1 и x_2 и снова измерить y . Пусть в первом опыте были заданы $x_1 = 1$ и $x_2 = 10$ и получено $y = 11$, а во втором опыте $x_1 = 10$ и $x_2 = 101$ и получено $y = 111$. Тогда система уравнений для определения a_1 и a_2 будет

$$\begin{cases} a_1 \cdot 1 + a_2 \cdot 10 = 11; \\ a_1 \cdot 10 + a_2 \cdot 101 = 111. \end{cases}$$

Решение этой системы есть $a_1 = 1$ и $a_2 = 1$, что можно проверить подстановкой. Но представим себе, что $y = 11$ было измерено с погрешностью и зафиксировано как $y = 11,1$. Тогда система уравнений будет

$$\begin{cases} a_1 \cdot 1 + a_2 \cdot 10 = 11,1; \\ a_1 \cdot 10 + a_2 \cdot 101 = 111. \end{cases}$$

А решение этой системы уравнений есть: $a_1 = 11,1$ и $a_2 = 0$. Вот к какой разнице может привести ошибка измерения y всего в $0,1/11 \approx 0,9\%$. Это значит, что данная система уравнений плохо обусловлена. Иначе говоря, плоскости, определяемые ее уравнениями, имеют между собой очень малый угол.

Не останавливаясь на теории этого вопроса, укажем, что практический метод проверки устойчивости решений состоит в поочередном отбрасывании некоторых экспериментальных точек и оценке возникающих вследствие этого изменений получаемого решения. Этот метод мы уже рассмотрели на приведенных выше примерах. Заметим, что наибольшее влияние оказывает, как правило, исключение точек, наиболее удаленных от центра тяжести поля точек, в то время как исключение центральных точек поля

даже в том случае, если они явно выпадают из полосы остальных точек, мало влияет на получаемое решение. Поэтому проверку устойчивости следует проводить путем поочередного отбрасывания крайних, наиболее удаленных от центра точек поля.

Для исключения абсурдных решений, получаемых МНК вследствие неоднородности статистики, наличия промахов или других причин неустойчивости, в последние годы интенсивно разрабатываются [44, 47] так называемые «робастные» (термин *робастность* введен в статистику Дж. Боксом), т. е. устойчивые к неоднородности данных, статистические методы, когда коэффициенты регрессии уточняются путем ряда последовательных итераций, а «мешающие» отсчеты отбрасываются. Программа для вычисления на ЭВМ с одновременным определением степени устойчивости проведена, например, в книге [38]. Программа работает в режиме итераций, а при расхождении итераций сообщает об этом оператору.

Оригинальная робастная программа обработки данных. Эта программа, устойчивая к присутствию промахов, предложена И. Л. Мисюченко (ЛПИ, Ленинград) и основана на использовании медианы пяти оценок центра X_5 , описанной в § 3-4.

Алгоритм этого метода обработки состоит в следующем. Все поле экспериментальных данных зависимости $y = f(x)$, состоящее из n точек, разбивается на m групп, как это было показано на рис. 6-3, б. В каждой группе точек находится ее центр, но не просто в виде медианы, а как медиана пяти оценок центра X_5 по абсциссам и ординатам точек этой группы. И затем по найденным таким способом центрам МНК определяется аппроксимирующая функция.

Такая программа работает в десятки раз быстрее, чем расчет МНК по всему массиву точек, так как время обработки по МНК уменьшается пропорционально квадрату числа обрабатываемых отсчетов и квадрату числа определяемых коэффициентов модели. В результате такая программа не только устраняет влияние промахов, но за то же время позволяет просчитать несколько вариантов и выбрать наилучший.

Затруднение создает необходимость предварительного выбора числа m групп, на которые делится исходное поле данных. При известном заранее виде модели это решается по правилам, изложенным в § 8-6. Если же вид модели заранее не известен, то приходится прибегать к перебору вариантов, чему и способствует малое время выполнения программы.

Если заранее известны наклоны кривой на ее концах, что часто бывает на практике, то вместо МНК более эффективно воспользоваться аппроксимацией сплайн-функцией, позволяющей задать значения производных на границах аппроксимируемого участка, что МНК не позволяет. Если же производные исследуемой зависимости на ее концах заранее не известны, то приближенное значение этих производных можно оценить по взаимному

положению медианных центров X_5 двух крайних групп точек с каждого из концов зависимости, но такой прием может оказаться часто и рискованным.

Предложенный метод полезен при обработке диффузных исходных данных, содержащих значительное число промахов. Благодаря резкому сокращению числа отсчетов, обрабатываемых по МНК (m центров вместо n исходных точек), метод особенно эффективен при обработке диффузных данных на микропроцессорах и микро-ЭВМ. В своей основе метод схож с методом простых медианных центров, однако обеспечивает достаточно эффективную обработку при любом законе распределения погрешностей исходных данных.

Самоадаптирующийся робастный метод обработки данных, являющийся существенным дальнейшим развитием МНК, разработан И. С. Кирилловой и В. Я. Крейновичем (ВНИИЭП, Ленинград) и опубликован в 1983—1985 гг. [18]. В основе метода лежит использование предложенного в 1965 г. И. А. Назаровым (ЛЭТИ, Ленинград) единого математического описания [27] всего разнообразия класса экспоненциальных распределений формулой (2-8) вида $p(x) = A(\alpha) \exp(-|x|^\alpha)$, которая при изменении показателя степени α от ∞ до 0 последовательно описывает все распределения от равномерного с эксцессом $\varepsilon = 1,8$ и $\alpha = \infty$, через нормальное ($\varepsilon = 3$, $\alpha = 2$), распределение Лапласа ($\varepsilon = 6$, $\alpha = 1$) и до распределений с бесконечно большим эксцессом ($\varepsilon \rightarrow \infty$, $\alpha \rightarrow 0$).

Идея самоадаптирующегося алгоритма обработки заключается в следующем. Согласно табл. 5-1 и кривым на рис. 5-1, данные для которых были заимствованы из работы В. И. Мудрова и В. Л. Кушко (1976 г.) [26] по сравнительной эффективности МНК и метода наименьших модулей (МНМ), для нормального распределения, имеющего показатель степени $\alpha = 2$, наиболее эффективен метод наименьших квадратов и оценка центра математическим ожиданием. Для распределения Лапласа с $\alpha = 1$ наиболее эффективен метод наименьших модулей и оценка центра медианой. Вообще-то говоря, это следует из метода максимального правдоподобия (ММП). Так, например, для распределений с $\alpha = 2$ ММП приводит к МНК, для распределений же с $\alpha = 4$ (см. табл. 5-1) ММП приводит к использованию не минимума суммы квадратов, а минимума суммы четвертых степеней отклонений $\sum \Delta_i^4$, для распределений с $\alpha = 1,5$ — минимума $\sum |\Delta_i|^{1,5}$, для распределений с $\alpha = 0,75$ — минимума $\sum |\Delta_i|^{0,75}$ и т. д.

Для практического использования этого метода авторами [18] разработан и программно реализован алгоритм обработки данных, аналогичный алгоритму МНК, но для произвольно задаваемой как целой, так и дробной степени α . При $\alpha = 2$ он дает ответ, соответствующий МНК, т. е. среднему арифметическому, при $\alpha = 1$ — соответствующий МНМ, т. е. медиане, а при любом другом заданном значении α (от ∞ до 0) — обеспечивает получе-

ние оценок, эффективных именно для данного распределения погрешностей исходных данных.

Расчет ведется в виде итерационного процесса. При первом прогоне программы задается, например, $\alpha = 2$. После расчета параметров модели вычисляются отклонения Δ_i каждой экспериментальной точки от полученной модели и по ним находятся оценки формы распределения погрешностей в виде эксцесса ε и контрэксцесса κ . Далее по известному значению ε надо найти степень α фактического распределения погрешностей исходных данных, т. е. необходимо решить трансцендентное уравнение (2-10) $\varepsilon = \Gamma(1/\alpha) \cdot \Gamma(5/\alpha) / [\Gamma(3/\alpha)]^2$ относительно α при известном ε . Для этого могут быть использованы аппроксимирующие формулы (5-27а).

Это соотношение позволяет по выборке фактических значений погрешностей Δ_i , рассчитав оценку контрэксцесса κ , найти степень α кривой плотности экспоненциального распределения в формуле Назарова (2-8), соответствующего данной выборке, и задать это значение для следующего прогона программы.

После нескольких прогонов программы получаемое значение α в результате последовательных итераций стабилизируется на некотором уровне, указывающем вид закона распределения погрешностей исходных данных, а следовательно, и вид получаемого решения.

Теперь может быть произведено цензурирование исходных экспериментальных данных, т. е. согласно формуле (4-11) удалены промахи, и итерационный процесс продолжен для получения окончательного решения уже на очищенных от промахов исходных данных.

Все перечисленные операции чисто формальны, и поэтому они могут быть заранее предусмотрены в тексте программы. Благодаря этому такая программа не будет требовать какого-либо вмешательства оператора на протяжении всего итерационного решения, а может просто информировать его о ходе решения: сообщать получаемые после каждой итерации значения ε и κ распределения погрешностей, сообщать о типе получаемого решения путем указания назначаемых значений α , перечислять значения удаляемых промахов и приводить параметры качества исходных данных в виде коэффициента множественной корреляции ρ и приведенной погрешности γ , а также указывать область неопределенности, т. е. погрешность, получаемых решений.

6-7. РАСЧЕТ ПАРАМЕТРОВ ПОЛОСЫ НЕОПРЕДЕЛЕННОСТИ ИСХОДНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Как уже указывалось, погрешности смещения экспериментальной точки вдоль оси x и вдоль оси y не могут быть разделены между собой. Поэтому общепринято условно считать, что опре-

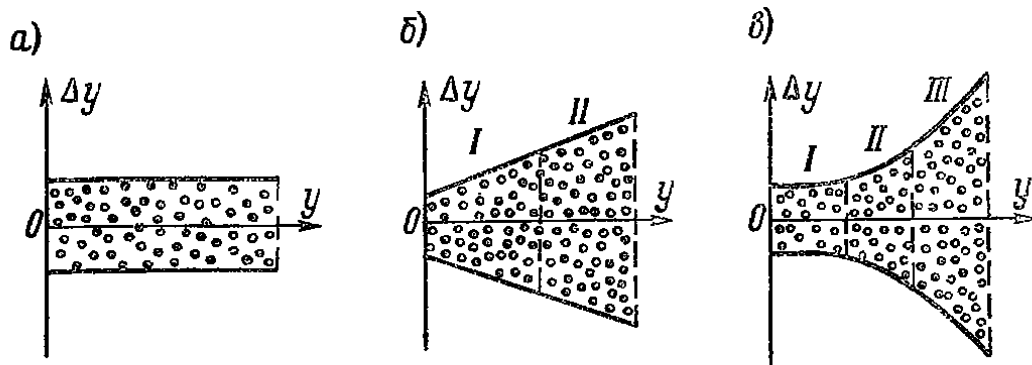


Рис. 6-13

деление значений x_i производится без погрешности, а всю случайную погрешность относить к координате y_i , а также считать случайными погрешностями исходных экспериментальных данных разности между координатой y данной экспериментальной точки и координатой $\hat{y}$, найденной по усредненной модели исследуемого явления (при одном и том же значении независимой переменной x), т. е. $\Delta_i = y_i - \hat{y}$, где $\hat{y} = f(x)$.

Поэтому значение этих случайных погрешностей Δ_i может быть найдено лишь после того, как будет найдена усредненная модель $\hat{y} = f(x)$.

Полоса рассеяния экспериментальных точек вокруг найденной усредненной зависимости $\hat{y} = f(x)$ может иметь самую разнообразную форму (см. рис. 6-1, 6-2, в и г, 6-3 или 6-12). Поэтому наиболее полным описанием формы этой полосы по аналогии с ГОСТ 8.401—80 является описание ее границ формулами в функции значений $\hat{y}$. Такое аналитическое описание позволяет определить погрешность для каждого значения y и тем самым является наиболее *полной метрологической аттестацией* получаемых результатов измерений. Однако определение такой метрологической характеристики метода измерений относительно трудоемко и требует наличия большого объема исходных экспериментальных данных.

Для получения этой характеристики после определения окончательного вида искомой усредненной зависимости $\hat{y} = f(x)$ надо найти частные значения погрешности Δ_i для каждой экспериментальной точки и построить их в координатах $\Delta_i = f(\hat{y})$. В результате вместо полосы погрешностей, располагающейся по обе стороны кривой $\hat{y} = f(x)$ (например, на рис. 6-1, 6-2, в и г, 6-12), мы получим полосу погрешностей, располагающуюся по обе стороны оси $\hat{y}$ (рис. 6-13). В зависимости от формы этой полосы (рис. 6-13, а, б или в) необходимо принять волевое решение (см. § 7-1) о том, какого вида формулой (одночленной, двучленной или трехчленной) целесообразно описать границы этой полосы.

Для получения такого описания все поле точек должно быть разделено на участки (как это показано на рис. 6-13) и в каждом

из них должно быть вычислено значение с. к. о., а также определен квантильный t или энтропийный k коэффициенты. Тогда может быть достаточно обоснованно по формулам (2-6) или (2-19)—(2-22) указана ширина доверительного или энтропийного интервала неопределенности на каждом из участков. Полагая, что эти значения соответствуют серединам участков *I*, *II* и *III*, могут быть найдены соответствующие двучленные или трехчленные формулы границ полосы неопределенности по всему диапазону измерений.

Невозможность определения такой наиболее полной метрологической аттестации полосы неопределенности результатов единичных измерений чаще всего обусловлена малостью общего числа проведенных измерений. Действительно, для определения оценки контрэксцесса распределения экспериментальных данных с погрешностью $\pm 10\%$ (см. § 4-4) в зависимости от вида закона распределения на каждом участке рис. 6-13 необходимо иметь от 10 до 750 отсчетов. Таким образом, ориентируясь на среднее число примерно 70 измерений, для установления доверительных границ по одночленной формуле необходимо 70 измерений, по двучленной формуле — 140 измерений, а для трехчленной — 210 измерений.

С этой точки зрения представляется весьма заманчивым перейти от доверительного интервала неопределенности к указанию энтропийного интервала неопределенности, так как оценка энтропийного коэффициента (см. § 4-4) определяется с $\gamma_k = 5\%$ уже при объеме выборки всего в 25—50 отсчетов, а прямое определение оценки энтропийной погрешности для распределений с $\kappa = 0,6 \div 0,7$ требует в 1,5—2 раза меньшей выборки, чем определение σ при том же рассеянии (см. рис. 4-4).

И, наконец, всегда остается возможность указания границ полосы погрешностей с доверительной вероятностью $P_d = 0,9$, когда вообще нет необходимости определять значения квантильного коэффициента. Поэтому всякий раз, как описание границ полосы погрешностей по аналогии с ГОСТ 8.401—80 оказывается возможным, наиболее полная метрологическая аттестация результатов измерений должна проводиться.

При отсутствии этой возможности приходится упрощать метрологическое описание, переходя от трехчленной формулы к двучленной и даже — к одночленной. В последнем случае (рис. 6-13, а) все наблюдавшиеся погрешности объединяются в единую выборку и по ним вычисляются оценки с. к. о., эксцесса или энтропийного интервала неопределенности. Далее в достаточной степени произвольно принимается тот или иной вид распределения и вычисляется одна усредненная оценка абсолютной, относительной или приведенной погрешности результатов эксперимента.

Оценка рассеяния экспериментальных данных значением коэффициента корреляции. Если относительная приведенная погрешность есть показатель разброса экспериментальных данных, то

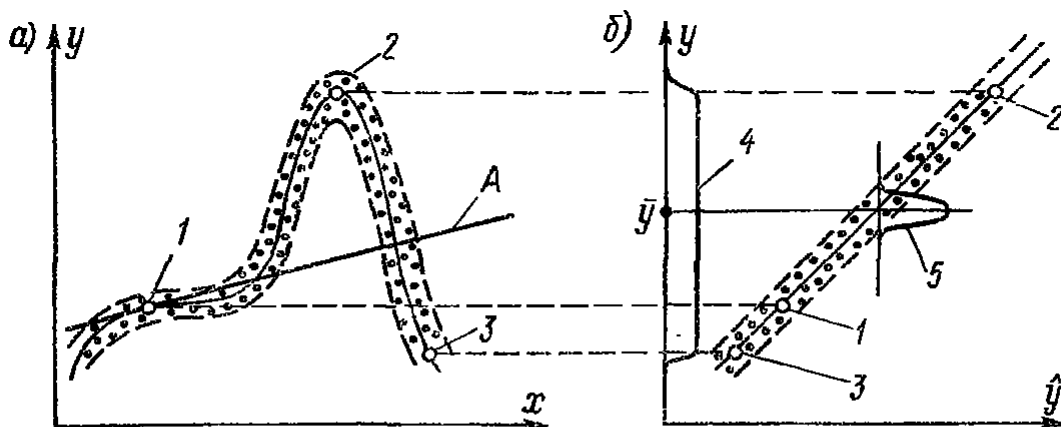


Рис. 6-14

коэффициент корреляции, наоборот, есть показатель тесноты их группирования относительно принятой модели. Эти оценки характеризуют одно и то же свойство этих данных и поэтому взаимосвязаны между собой.

Однако коэффициент корреляции r_{xy} значений x_i и y_i применительно к однофакторной зависимости $y = f(x)$ характеризует тесноту группирования данных лишь относительно некоторой прямой (например прямой 3 на рис. 6-10). При более сложной зависимости (рис. 6-14, а) коэффициент корреляции r_{xy} будет оценивать тесноту экспериментальных точек относительно некоторой прямой, обозначенной буквой А, что, естественно, несет мало сведений о тесноте их группирования относительно искомой кривой 1—2—3.

Поэтому оценка разброса экспериментальных данных значением коэффициента корреляции r_{xy} может быть использована лишь только для линейной однофакторной модели вида $\hat{y} = a_0 + a_1x$.

Однако существует понятие *коэффициента множественной корреляции*, расчетная оценка которого правомерна для любых многофакторных зависимостей, в том числе и для сложных нелинейных однофакторных зависимостей. Пояснить смысл этого понятия, не прибегая к строгим математическим выкладкам, можно следующим образом.

Если используемая в качестве модели функция $\hat{y} = f(x)$ есть однозначная функция x , т. е. любому заданному x соответствует лишь одно определенное значение $\hat{y}$, то при отсутствии погрешностей экспериментальных данных тождество $y \equiv \hat{y}$ изобразится в координатах $y = f(\hat{y})$ (рис. 6-14, б) в виде биссектрисы прямого угла, как бы ни была сложна используемая модель. Если же погрешность экспериментальных данных не равна нулю, т. е. экспериментальные точки на рис. 6-14, а располагаются не на кривой 1—2—3, а в некоторой полосе вдоль этой кривой, то на рис. 6-14, б экспериментальные значения y_i расположатся в некоторой полосе вокруг прямой $y = f(\hat{y})$. Это обусловлено тем, что каждому значению x соответствует (при однозначной

модели) лишь одно значение $\hat{y}$, а экспериментальные значения y_i при заданном x могут быть как больше $\hat{y}$, так и меньше.

Коэффициент корреляции ρ_{yy} между экспериментальными значениями y_i и полученными по модели $\hat{y}$ при разных x и принято называть коэффициентом множественной корреляции. Его использование одинаково правомерно как для действительно многофакторных моделей вида $y = f(x_1, \dots, x_l)$, так и для сложных однофакторных моделей, когда $x_1, \dots, x_l$ суть просто разные функции ($x, x^2, x^3, \sin x, \ln x$ и т. д.) одной переменной x .

Построение графиков экспериментальных данных в виде $y = f(\hat{y})$, подобных рис. 6-14, б, очень полезно для окончательной наглядной оценки полученной модели. Для того чтобы представить себе, как деформируется полоса погрешностей (рис. 6-14, а), три произвольные точки (1, 2, 3) помечены на обоих рисунках одинаковыми цифрами.

К графику рис. 6-14, б, где все экспериментальные точки расположены вдоль прямой $y = f(\hat{y})$, понятие коэффициента корреляции применимо без каких-либо оговорок и, как уже упоминалось в § 3-1, он составляет

$$\rho_{yy} = 1/\sqrt{1 + (\sigma_\Delta/\sigma_y)^2} \approx \sqrt{1 - (\sigma_\Delta/\sigma_y)^2}, \quad (6-10)$$

где σ_Δ — с. к. о. экспериментальных точек от прямой $y = \hat{y}$, а σ_y — с. к. о. тех же точек от горизонтальной прямой на уровне $\hat{y}$. Так как в дальнейшем под ρ будет всегда пониматься значение ρ_{yy} , то индексы далее будут опущены.

Оценка σ_Δ есть абсолютная оценка ширины полосы разброса, а оценка σ_y есть оценка ширины диапазона изменения значений y_i . Отношение этих величин есть некоторое подобие относительной приведенной погрешности, употребляемой обычно для характеристики погрешности Δy приборов с учетом длины (Y_k) их диапазона измерений. Но, определяя приведенную относительную погрешность как $\gamma = \Delta y/Y_k$, обычно под Δy понимают не всю ширину полосы неопределенности, а лишь ее половину, в то время как под Y_k — не отклонения y от $\hat{y}$, а весь диапазон изменения от $y_{\min}$ до $y_{\max}$. Отсюда аналогом относительной приведенной погрешности в первом приближении следует считать $\gamma \approx \sigma_\Delta/(2\sigma_y)$. С учетом этого отношение σ_Δ/σ_y , входящее в выражение (6-10), можно считать равным удвоенной относительной приведенной погрешности, т. е. $(\sigma_\Delta/\sigma_y) \approx 2\gamma$. Тогда соотношение (6-10) примет вид

$$\rho \approx \sqrt{1 - (2\gamma)^2}, \quad (6-11)$$

или, наоборот,

$$\gamma \approx \frac{1}{2} \sqrt{1 - \rho^2}. \quad (6-12)$$

Дальнейшее уточнение соотношения между ρ и γ может быть введено исходя из следующих рассуждений. Под приведенной

погрешностью γ обычно понимают не с. к. о. δ , а какую-то интервальную (доверительную или энтропийную) оценку. Соотношение между ними определяется, естественно, видом закона распределения (см. гл. 3 и 5). В данном же случае форма распределений, характеризующих σ_Δ и σ_y , резко различна. Распределение, имеющее с. к. о. σ_y , определяется размещением экспериментальных точек по оси y (на рис. 6-14, б — это кривая 4). Оно более близко к равномерному. Поэтому для него энтропийное значение можно принять $\Delta_{\text{э1}} \approx \sqrt{3}\sigma_y$. Распределение же, характеризующее с. к. о. σ_Δ , помечено на рис. 6-14 цифрой 5 и определяется погрешностями экспериментальных точек. Оно чаще всего бывает колокообразным. Полагая его в среднем в какой-то степени близким к нормальному, его энтропийное значение примем $\Delta_{\text{э2}} = 2,066\sigma_\Delta$. Отсюда энтропийное значение относительной приведенной погрешности экспериментальных данных будет

$$\gamma_{\text{э}} = \frac{\Delta_{\text{э2}}}{2\Delta_{\text{э1}}} = \frac{2,066}{2\sqrt{3}} \frac{\sigma_\Delta}{\sigma_y} \approx \frac{\sigma_\Delta}{\sigma_y \sqrt{3}},$$

т. е. $\sigma_\Delta/\sigma_y = \sqrt{3}\gamma_{\text{э}}$. С учетом этого вместо выражений (6-11) и (6-12) получаем более уточненные соотношения:

$$\rho = \sqrt{1 - 3\gamma_{\text{э}}^2}; \quad (6-13)$$

$$\gamma_{\text{э}} = \sqrt{(1 - \rho^2)/3}. \quad (6-14)$$

Сопоставляя соотношения (6-12) и (6-14), следует заметить, что (6-12) более справедливо при $\rho < 0,9$, а (6-14) — при $\rho > 0,9$ (см. табл. 7-1).

Эти соотношения дают возможность очень простой приближенной оценки коэффициента корреляции ρ_{yy} по формуле (6-13). Для этого достаточно полуширину полосы погрешностей поделить на диапазон изменения y , а далее по (6-13) рассчитать ρ_{yy} .

И, наоборот, располагая в результате обработки на ЭВМ более точным значением ρ , найденным по формуле (6-9), можно по (6-12) или (6-14) рассчитать более точное значение энтропийной оценки погрешности экспериментальных данных, основанной на средней квадратической оценке отклонений всех экспериментальных точек от кривой регрессии.

Здесь необходимо обратить внимание читателя на два обстоятельства. Во-первых, расчет оценки ρ по формуле (6-9) удобно программно совместить с обработкой по МНК, так как в (6-9) входят те же $\sum x_i^2$, $\sum x_i y_i$, что вычисляются и для МНК. Для этого в программе МНК необходимо лишь дополнительно предусмотреть вычисление $\sum y_i^2$, которая нужна для вычисления ρ и не нужна в программе МНК. (Необходимые готовые программы см., например, в работе [38].)

Во-вторых, получаемая оценка относительной приведенной погрешности разброса исходных экспериментальных данных в виде

ρ , φ или φ_0 есть некоторая средняя (опирающаяся на с. к. о.) оценка ширины полосы погрешностей одним числом без учета формы этой полосы. При аддитивной полосе погрешностей (рис. 6-13, а или 6-14, б) ничего большего и не нужно, но при более сложной форме полосы погрешностей (рис. 6-1, а; 6-3, б; 6-12; 6-13 б или в) такая оценка есть нечто среднее, слабо характеризующее максимальную и минимальную ширину этой полосы, особенно если эти размеры существенно различаются между собой.

Поэтому более полной характеристикой полосы разброса экспериментальных данных при определении функциональных зависимостей следует считать не указание одного числа в виде φ , σ или ρ , а сообщение о форме полосы погрешностей в соответствии с рис. 6-13 и описание ее соответствующей одночленной, двухчленной или трехчленной формулой.

В простейшем случае — это указание погрешности в начале и конце диапазона, т. е. при $y = 0$ и $y = y_{\max}$ (для полосы на рис. 6-13, б), или в трех точках диапазона (для рис. 6-13, в). Этот простейший способ указания погрешности результатов эксперимента и должен быть, по-видимому, рекомендован как основной способ нормирования.

6-8. РАСЧЕТ ПАРАМЕТРОВ ПОЛОСЫ НЕОПРЕДЕЛЕННОСТИ УСРЕДНЕННОЙ ОДНОФАКТОРНОЙ МОДЕЛИ

Весь смысл большой дополнительной затраты времени на проведение многократных отсчетов состоит в том, что модель исследуемой зависимости, полученная путем усреднения всех этих данных, будет в значительной степени свободна от случайных погрешностей, присущих используемому методу измерений.

Однако путем усреднения случайные погрешности не устраняются полностью, а лишь уменьшаются в определенное число раз. Поэтому полученная в результате статистической обработки (как глазомерными, или медианными способами, так и МНК) усредненная модель также имеет свою полосу неопределенности, хотя и более узкую, чем полоса разброса исходных экспериментальных данных.

Рассмотрим это на примере простейшей модели вида $\hat{y} = \bar{y} + ax$, которая была представлена на рис. 6-10. При $x = 0$ произведение ax также равно нулю и неопределенность центральной точки модели определяется только погрешностью определения $\bar{y}$. При использовании МНК с. к. о. $\sigma_{\bar{y}} = \sigma_{\Delta}/\sqrt{n-l}$, т. е. дисперсия $\sigma_{\bar{y}}^2$ меньше дисперсии σ_{Δ}^2 разброса исходных данных относительно линии регрессии в $n-l$ раз, где n — число усредняемых экспериментальных точек, а l — число определяемых коэффициентов.

Однако при отступлении от $x = 0$, кроме ошибки $\sigma_{\bar{y}}$ будет сказываться и ошибка Δa в определении коэффициента регрес-

сии a . Этот вопрос исследовал еще Бартлет и нашел, что квантильная оценка этого разброса

$$\Delta a = t \frac{\sigma_y \sqrt{1 - \rho^2}}{\sigma_x \sqrt{n - l}},$$

где t — квантиль распределения Стьюдента с $v = n - l$ степенями свободы; l — число определяемых МНК коэффициентов (в данном случае равное 2), а ρ — коэффициент корреляции экспериментальных точек. А так как определяемое по МНК само значение $a = \rho \sigma_y / \sigma_x$, то

$$\frac{\sigma_y}{\sigma_x} = \frac{a}{\rho} \quad \text{и} \quad \sigma_a = \frac{\Delta a}{t} = \frac{a}{\rho} \sqrt{\frac{1 - \rho^2}{n - l}}.$$

Если погрешность, создаваемая неопределенностью $\bar{y}$, является аддитивной, т. е. дает постоянную по ширине полосу возможных положений линии регрессии, то изменение коэффициента регрессии a на величину Δa изменяет наклон линии регрессии и создает мультипликативную составляющую погрешности модели. Среднее квадратическое отклонение величины ax

$$\sigma_{ax} = \sigma_a x = \frac{ax}{\rho} \sqrt{\frac{1 - \rho^2}{n - l}}, \quad (6-15)$$

т. е. линейно возрастает с увеличением x по мере удаления от центра тяжести поля экспериментальных точек.

При расположении начала координат в точке $\bar{x}$ и $\bar{y}$ эти погрешности независимы и

$$\sigma_\Sigma = \sqrt{\sigma_{\bar{y}}^2 + \sigma_{ax}^2} = \sqrt{\frac{\sigma_\Delta^2}{n - l} + \frac{x^2 a^2 (1 - \rho^2)}{(n - l) \rho^2}}. \quad (6-16)$$

ГЛАВА СЕДЬМАЯ

МЕТОДЫ ОБРАБОТКИ И ОЦЕНКИ ПОГРЕШНОСТЕЙ ПРИ МНОГОФАКТОРНОМ ЭКСПЕРИМЕНТЕ

7-1. МНОГОФАКТОРНЫЙ ЭКСПЕРИМЕНТ И ЕГО ОСОБЕННОСТИ

Как уже указывалось выше, измерение отдельно взятых величин в практике измерений встречается довольно редко. Однако и определение функциональных зависимостей вида $y = f(x)$ также

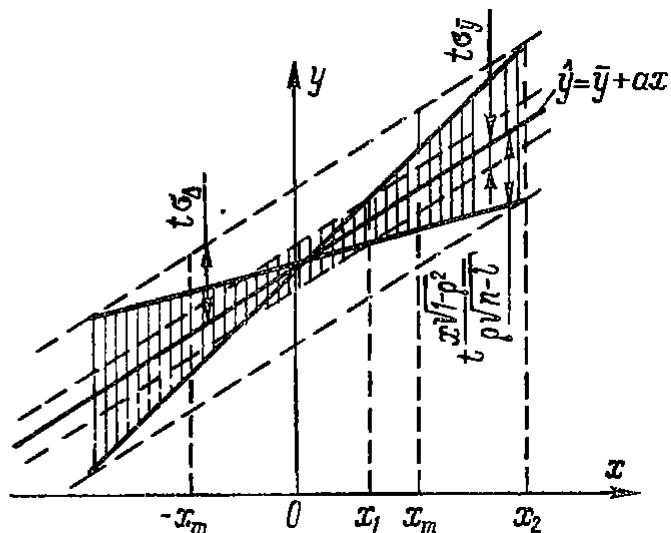


Рис. 6-15

является идеализацией, так как в природе очень трудно указать такую величину y , которая зависела бы лишь только от одного единственного фактора x , а к изменению всех других факторов была бы абсолютно нечувствительна. В действительности как раз наоборот: наиболее часто встречающейся на практике задачей является установление вида функции многих независимых переменных

$$y = f(x_1, x_2, \dots, x_k),$$

т. е. отыскание математической модели исследуемого явления, описывающей характер взаимосвязей между включенными в нее переменными $y, x_1, x_2, \dots, x_k$.

Осуществление многофакторного эксперимента существенно более трудоемко, чем однофакторного. Однако развитие всех наук (от механики до электроники или биофизики) убеждает в том, что это единственный путь научных исследований. Так, закон равноускоренного движения теперь мы записываем в виде формулы $S = gt^2/2$. Однако на установление вида этой функциональной зависимости Галилей затратил более 25 лет своей жизни. Он начал свои опыты в возрасте 19 лет (в 1583 г.) с наблюдения раскачиваний лампы, подвешенной на длинном подвесе в Пизанском соборе, которую кто-нибудь нечаянно толкнул. При этом, следуя совету Кардана, период колебаний он измерял по биению собственного сердца. Затем (с 1590 г.) продолжал эти исследования путем бросания тяжелых и легких тел с Пизанской башни. Этими опытами он опроверг бытовавшее ранее мнение, что скорость падающих тел пропорциональна их весу. Однако в 1606 г., т. е. после 23 лет исследований, Галилей еще считал, что скорость падения возрастает пропорционально пройденному пути. Наконец, в период между 1606 и 1609 гг. Галилеем был поставлен уже двухфакторный эксперимент. Различные значения высоты S и ускорения g моделировались с помощью наклонной плоскости, а измерялся отклик в виде времени t . Для повышения точности эксперимента в длинных (12 локтей) досках был прорезан прямой желоб, покрытый для уменьшения трения гладким пергаментом. Шар из твердой бронзы был хорошо отполирован и отпускался с разных высот, а точность измерения времени была существенно повышена путем взвешивания на весах количества воды, натекавшей через тонкую трубочку из ведра в подставленный бокал.

В итоге было доказано, что скорость падения пропорциональна не пройденному пути, а истекшему времени, т. е. $v = gt$ и $S = gt^2/2$. При этом замечательно то, что Галилей уже отчетливо понимал роль математики, он хотел, чтобы на обложке собрания его сочинений были написаны слова: «Отсюда станет понятным на бесчисленных примерах, сколь полезна математика в заключениях, касающихся того, что предлагает нам природа».

Этот исторический пример наглядно показывает, сколь важен

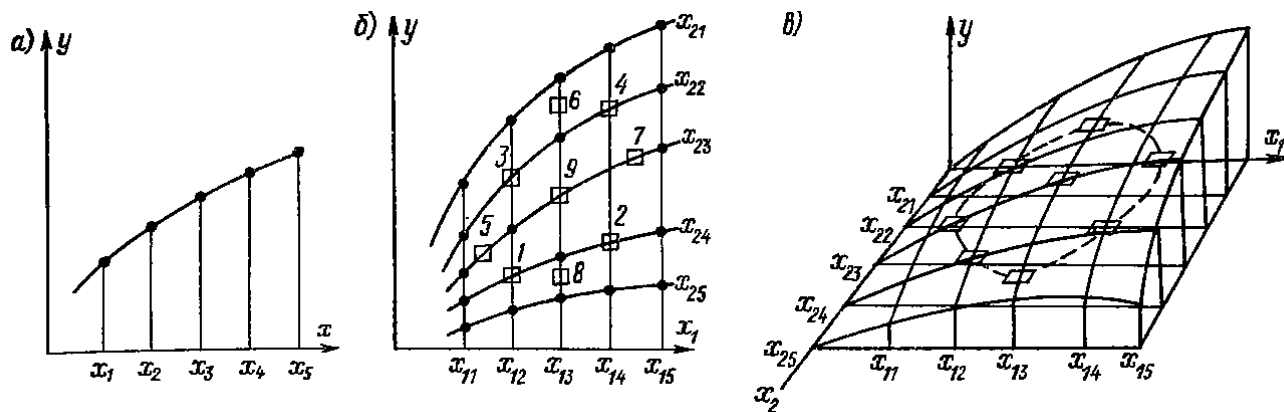


Рис. 7-1

переход от однофакторного эксперимента к многофакторному и сколь трудным оказывается самый первый этап исследования, состоящий в обнаружении всего перечня влияющих факторов.

Достаточно большой раздел современной электротехники сводится, по существу, к одной математической модели, состоящей из двух соотношений:

$$r = \frac{\rho l}{S} = \frac{4\rho l}{\pi d^2} \quad \text{и} \quad x = \omega L = 2\pi f \frac{\omega^2}{R_M},$$

и их композиции в виде

$$z = \sqrt{r^2 + x^2} = \sqrt{\left(\frac{4\rho l}{\pi d^2}\right)^2 + \left(\frac{2\pi f \omega^2}{R_M}\right)^2}.$$

Но на отыскание этой математической модели всего шести факторов $z = f(\rho, l, d, f, \omega, R_M)$ электротехника затратила не одно десятилетие. Отсюда можно заключить, что история развития точных наук (а возможно, и не только их) состоит в нахождении перечня факторов и создании многофакторных математических моделей исследуемых явлений и их интерпретации. Важной ступенью в решении любой научной проблемы является интерпретация постоянных, входящих в математические модели, т. е. введение новых, не известных до этого понятий: ускорение земного притяжения — g , удельное электрическое сопротивление — ρ , угловая частота — ω , индуктивность — L , магнитное сопротивление — R_M и т. д.

Таким образом, обработка результатов измерений, получаемых при многофакторном эксперименте, имеет целью получение основополагающих научных данных в виде неизвестных ранее математических моделей и их интерпретацию, а отнюдь не сводится к простому вычислению $\bar{X}$, σ_{xi} или $\sigma_{\bar{X}}$.

Однако постановка многофакторных исследований чрезвычайно трудоемка. Поэтому не случайно, что исследования Галилея заняли более 25 лет, а электротехнике для получения упомянутой модели из шести факторов потребовалось несколько десятилетий.

Для определения однофакторной зависимости $y = f(x)$ (рис. 7-1, а) нам достаточно поставить всего, например, пять опы-

тов. Но если это двухфакторная зависимость $y = f(x_1, x_2)$, то, желая исследовать по каждому из факторов x_1 и x_2 те же пять уровней, мы уже вынуждены поставить (см. рис. 7-1, б) $n = 5^2 = 25$ опытов. В этом случае число опытов n , число уравнений варьирования переменных q и число исследуемых факторов k связаны соотношением $n = q^k$. Для рассмотренной электротехнической задачи при $k = 6$ и построении каждой частной кривой всего по пяти точкам ($q = 5$) общее число опытов $n = 5^6 = 15\,625$, т. е. в 625 раз больше, чем в задаче Галилея.

Это катастрофическое возрастание числа n необходимых опытов с ростом числа k учитываемых факторов при многофакторном исследовании получило в теории эксперимента образное наименование «проклятия размерности». Единственный принципиально возможный проход сквозь «проклятие размерности» был указан в 1925—1929 гг. английским статистиком Рональдом Фишером. Он показал, что «проклятие размерности» наложено на нас только до тех пор, пока мы находимся в плену традиций однофакторного эксперимента, т. е. пытаемся по очереди снимать частные зависимости $y = f_1(x_1)$, $y = f_2(x_2)$, $y = f_3(x_3)$ и т. д., а все другие влияющие факторы, кроме исследуемого, стараемся стабилизировать. (Отсюда и происходит классическое правило: «никогда не меняй два фактора x_i и x_j одновременно, так как не поймешь, от чего же изменился y ».)

Р. Фишер указал, что проход сквозь «проклятие размерности» обеспечивается при отказе от этого классического правила, при отказе от многих однофакторных экспериментов в пользу единого многофакторного эксперимента. Этот принцип состоит в одновременном варьировании всех k переменных. При этом, правда, нельзя логически заключить, от какого сочетания x_j произошло изменение отклика y , однако можно составить систему уравнений и, решив ее, сразу получить коэффициенты влияния для всех k факторов.

Математические модели многофакторных зависимостей. Эти модели, естественно, разнообразнее однофакторных моделей. Поэтому и подбор наиболее удобной аппроксимирующей функции для многофакторной зависимости является существенно более сложным. Наиболее действенным методом, облегчающим подбор и интерпретацию многофакторных моделей, является удачное расчленение общей задачи на ряд частных, решение каждой из них в отдельности и, наконец, объединение, композиция полученных частных функций в общую математическую модель.

Именно таким путем происходил, например, поиск математической модели в электротехнике. Прямое решение общей задачи, т. е. поиск функции сразу шести факторов, практически невозможен, так как выделить и опознать вид частных функций, входящих в модель из шести факторов, чрезвычайно трудно. Поэтому исторически решение шло другим путем. Были опознаны частные функции: модель активного сопротивления $r = f_1(x_1, x_2, x_3)$

реактивного сопротивления $x = f_2(x_4, x_5, x_6)$, и затем решалась задача их композиции в общую модель. Этот метод декомпозиции сложной задачи и есть, по-видимому, общий метод решения в трудных случаях исследования.

Другим методом, часто значительно облегчающим классификацию и интерпретацию сложных моделей, является их геометрическое представление. Для этого все направления изменений факторов $x_1, x_2, \dots, x_k$ трактуются как ортогональные оси некоего многомерного пространства факторов. При однофакторной зависимости $y = f(x_1)$ все значения этого фактора размещены лишь на оси x_1 . При двухфакторной зависимости $y = f(x_1, x_2)$ пространство факторов представляет собой уже плоскость с осями x_1 и x_2 , а любое сочетание факторов характеризуется точкой на этой плоскости с координатами x_1, x_2 . При трехфакторной зависимости данное сочетание факторов характеризуется точкой в трехмерном пространстве с координатами x_1, x_2, x_3 , при четырехфакторном или в общем случае k -факторном эксперименте — координатами точки в k -мерном пространстве факторов.

Так как пространство модели y кроме осей $x_1, \dots, x_k$ содержит еще и ось y , то размерность пространства модели равна $k + 1$, т. е. всегда на единицу больше размерности пространства факторов.

К сожалению, графически (в аксонометрии) невозможно изобразить более чем трехмерное пространство, поэтому приходится ограничиваться графическим изображением лишь одно-, двух- и трехмерных пространств, на них выяснять правила перехода при повышении размерности и распространять их на пространства большей размерности, которые приходится описывать лишь словесно или аналитически.

Так, на рис. 7-1, а пространство фактора x — одномерно, а пространство модели $y = f(x)$ — двумерно, — это плоскость x, y . На рис. 7-1, в пространство факторов — это плоскость x_1, x_2 , а модель $y = f(x_1, x_2)$ расположена в трехмерном пространстве x_1, x_2, y .

Сократить размерность пространства на единицу можно путем его проекции вдоль какой-то оси. Так, на рис. 7-1, б изображена проекция трехмерной функции $y = f(x_1, x_2)$ рис. 7-1, в вдоль оси x_2 на плоскость x_1, y . Этим приемом часто пользуются, изображая функцию многих переменных в виде семейства кривых. Координаты же по оси x_2 , вдоль которой производилась проекция, приходится помечать числами или знаками у каждой кривой семейства, понимая их как параметр данной кривой.

Функция $y = f(x_1, x_2)$ геометрически интерпретируется как некоторая поверхность в трехмерном пространстве (плоскость, часть сферы и т. д.). Для наименования аналогичной поверхности в k -мерном пространстве (где $k > 3$) используется, как правило, тот же термин, но с приставкой «гипер». Так, например, геометри-

ческое представление функции $y = a_0 + a_1x_1$ — это прямая на плоскости x_1, y , функции $y = a_0 + a_1x_1 + a_2x_2$ — это плоскость в трехмерном пространстве x_1, x_2, y . Функция $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$ — это гиперплоскость в четырехмерном пространстве, а $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$ — гиперплоскость в $(k + 1)$ -мерном пространстве.

Геометрическое представление уравнения $x_1^2 + x_2^2 = a^2$ — это окружность радиуса a на плоскости x_1, x_2 , уравнения $x_1^2 + x_2^2 + x_3^2 = a^2$ — это сфера радиуса a в трехмерном пространстве x_1, x_2, x_3 , а уравнения $x_1^2 + x_2^2 + x_3^2 + \dots + x_k^2 = a^2$ — это гиперсфера радиуса a в k -мерном пространстве $x_1, \dots, x_k$ и т. д.

Подобно тому, как в окружность на плоскости можно вписать квадрат, а в сферу в трехмерном пространстве можно вписать трехмерный куб, так и в гиперсферу в k -мерном пространстве можно вписать гиперкуб той же размерности. Таким образом, для описания функций в k -мерном пространстве пользуются понятиями гиперплоскости, гиперсферы, гиперкуба и т. п.

Принципиальное отличие многофакторной математической модели от однофакторной состоит в невозможности ее графического изображения. При однофакторной зависимости можно нанести на один и тот же график (см. рис. 6-14, а) как экспериментальные точки, так и кривую предлагаемой модели и наглядно увидеть, где и в какую сторону они отклоняются друг от друга. При многофакторной зависимости это возможно лишь тогда, когда ее удастся разложить на ряд однофакторных. Если же сделать этого не удастся, то проследить изменение погрешности адекватности можно путем построения графика $y_i = f(\hat{y})$, который был показан на рис. 6-14, б. Такой график может показать, что при одних значениях y принятая модель хороша, а при других — неудовлетворительна. Но вопрос о том, как же нужно ее изменить, остается без ответа, и практическим путем подбора адекватной многофакторной модели остается лишь метод проб и ошибок, так как выбор того или иного решения приходится производить, не имея нужной информации.

До разработки теории планирования эксперимента решения, принимаемые без какого-либо обоснования, называли просто необоснованными. Но при планировании эксперимента принятие таких решений неизбежно, поэтому в современной литературе для них утвердились специальные термины.

Если для принятия решения нет никаких данных и его приходится принимать на собственный страх и риск, то такое решение именуют *волевым*. Если же для принятия решения есть какие-то, хотя бы отдаленные, аналогии или ассоциации, косвенные соображения, наводящие на мысль о возможности такого решения, то его именуют *эвристическим*. Этими терминами и будем далее пользоваться.

Вследствие указанной трудности отыскания решения, обоснованного путем последовательной аппроксимации, для описания

результатов многофакторного эксперимента на практике часто прибегают к прямо противоположному приему, а именно чисто волевому назначению вида математической модели и лишь последующей проверке ее адекватности.

Так, например, из опыта различных наук (физики, механики, радиотехники и т. д.) известно, что большая часть уже полученных формул, описывающих самые разные явления, имеет вид произведения величин в различных степенях:

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_k^{a_k}, \quad (7-1)$$

где $a_1, \dots, a_k$ могут быть как целыми, так и дробными, как положительными, так и отрицательными.

Исходя из этого может быть выдвинуто предположение о том, не описывается ли данная исследуемая многофакторная зависимость подобной математической моделью? Для проверки такого предположения удобен следующий прием. Если выражение принятой модели прологарифмировать, то получим $\lg y = \lg a_0 + a_1 \lg x_1 + a_2 \lg x_2 + \dots + a_k \lg x_k$. Если теперь произвести подстановки вида $\lg y = Y$, $\lg a_0 = A_0$, $\lg x_1 = X_1$, $\dots$, $\lg x_k = X_k$, то уравнение модели примет вид

$$Y = A_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k, \quad (7-2)$$

следовательно, в логарифмических координатах (по всем осям) такая модель представляет собой *гиперплоскость*.

Учитывая, что стандартные программы для обработки МНК экспериментальных данных на ЭВМ составлены именно применительно к уравнению (7-2), т. е. к уравнению гиперплоскости, это представляет несомненное удобство.

Модель (7-1) в виде произведения всех факторов, как в положительных, так и в отрицательных степенях, называется в теории эксперимента *мультипликативной моделью* и замечательна тем, что с ее использованием были решены очень многие задачи гидродинамики, аэродинамики и теплотехники.

Другой, часто назначаемой волевым образом моделью является *аддитивная модель*, т. е. модель гиперплоскости (7-2). Это оправдано тем, что как бы ни была сложна поверхность отклика многофакторного эксперимента, ее малые участки, не содержащие экстремальных точек, с достаточным приближением могут быть заменены участками плоскости, касательными к кривой поверхности в центре этого участка. Это так называемое «чешуйчатое», или «граненое», описание кривой поверхности, которое позволяет получить ценные сведения о такой поверхности. В частности, по этой модели можно узнать «высоты» участков поверхности, углы и направления склонов, определить положение и направления изолиний и т. п., т. е. иногда такое «граненое» описание кривой поверхности часто оказывается очень удобным.

Основным недостатком аддитивной модели является то, что

для каждой соседней грани все коэффициенты $A_0, a_1, \dots, a_k$ модели (7-2) должны определяться заново, так как при переходе от грани к грани они получают новые значения.

Допустимая протяженность грани при аддитивной модели определяется нарастающей на краях грани погрешностью адекватности, а выбор ее аналогичен выбору длин участков кусочно-линейной аппроксимации кривых. Первым условием для выбора допустимых границ изменения x_i может служить условие $\Delta x_i \ll x_i$.

Рассмотренные мультипликативные и аддитивная модели получили наиболее широкое распространение для описания многофакторных зависимостей, что объясняется прежде всего их предельной простотой: при k факторах они требуют определения всего $k + 1$ коэффициента.

Протяженность каждой грани аддитивной модели могла бы быть существенно увеличена, если бы эти грани были не плоскими, а искривленными. Этого можно достигнуть, конечно, переходя от аддитивных моделей первого порядка (включающих x_i только в первой степени) к аддитивным же моделям второго, третьего и более высоких порядков (включающих члены вида x_i^2, x_i^3 и т. д.). Но такие модели очень сложны. Однако есть один вид поверхности, который занимает промежуточное положение — это винтовая поверхность, сечения которой как вдоль оси x_1 , так и вдоль оси x_2 являются прямыми линиями, но наклон их линейно изменяется с изменением координаты. Это частный случай *многомерного гиперboloида*, описываемый, например, при трех факторах уравнением вида

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \\ + a_4x_1x_2 + a_5x_2x_3 + a_6x_1x_3 + a_7x_1x_2x_3. \quad (7-3)$$

При двух факторах — это гиперboloид в трехмерном пространстве с уравнением вида

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2. \quad (7-4)$$

Если на плоскости $y = a_0 + a_1x_1 + a_2x_2$ провести линии постоянных x_2 , а затем спроектировать эти линии на плоскость y, x_1 , т. е. вдоль оси x_2 , то прямые постоянных x_2 дадут в этой проекции пучок параллельных прямых (рис. 7-2, а). Если такую же операцию провести над гиперboloидом (7-4), то в проекции на плоскость x_1, y (рис. 7-2, б) получится веер расходящихся прямых. Этим и следует пользоваться на практике как признаком выбора между моделью в виде гиперплоскости (7-2) или же в виде гиперboloида (7-4).

Модель поверхности в виде многомерного гиперboloида (7-3) часто позволяет существенно уменьшить погрешность адекватности при описании изогнутых поверхностей отклика. Поэтому в современной теории планирования эксперимента эта модель заняла, по существу, доминирующее положение. Члены, содер-

жащие произведения x_1x_2 , x_2x_3 и т. д., называют членами, отражающими *попарное взаимодействие* факторов, члены вида $x_1x_2x_3$ — членами *тройного взаимодействия* и т. д. (см. [2, 9, 21, 26, 46]). Однако при увеличении числа k факторов модель в виде гипер-

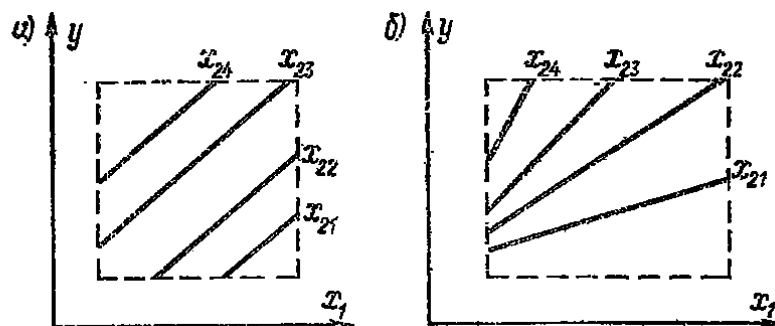


Рис. 7-2

болоида становится чрезвычайно громоздкой. Модель гиперплоскости (7-2) требует определения $k + 1$ коэффициента, для чего в теории планирования эксперимента разработаны предельно экономные планы в виде так называемых *дробных регулярных реплик*, в то время как в модели многомерного гиперболоида (7-3) содержится 2^k коэффициентов и для их определения требуется постановка такого же числа опытов, называемых в теории планирования *полным факторным планом*. Ясно, что $2^k \gg k + 1$, так, например, при $k = 6$ $k + 1 = 7$, а $2^6 = 64$.

7-2. МЕТОДЫ ОПРЕДЕЛЕНИЯ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ ВИДА И ПАРАМЕТРОВ МАТЕМАТИЧЕСКОЙ МОДЕЛИ МНОГОФАКТОРНОЙ ЗАВИСИМОСТИ

Для выбора вида многофакторной математической модели на практике используются лишь два метода: метод последовательной графоаналитической аппроксимации частных (базисных) функций отклика от отдельных групп факторов с последующей композицией их в общую многофакторную модель (см. § 6-2) и противоположный ему метод эвристического назначения какой-либо формальной модели. Выбор одного из этих методов прежде всего обусловлен наличием априорных сведений о возможном виде искомой математической модели исследуемого явления, а также степенью диффузности (разброса) исходных экспериментальных данных.

Важность использования априорных данных. Исследователю, впервые изучающему какое-либо новое явление, как правило, кажется, что все необходимые данные он должен извлечь из своего эксперимента, так как исследование для того и проводится. Однако при более внимательном анализе часто выясняется, что это не так, что очень многие данные могут быть установлены еще до постановки опыта и что они во многом обуславливают искомую модель, а опыт служит лишь для уточнения деталей этой модели.

Пусть, например, мы столкнулись с совершенно новой для нас задачей — составить тест, который за 24 ч мог бы численно характеризовать плодородие данного образца почвы инвариантно к ее температуре Θ и влажности φ (пример основан на фактическом материале, заимствованном из автореферата диссертации

Е. В. Соловьева «Исследование углекислотного режима корне-обитаемых сред»). Задача, действительно, является новой, и искать готовый ответ на нее в книгах по агротехнике, вероятно, бесполезно. Ни пшеница, ни тем более яблоня за 24 ч дать урожай не могут. Но для пробного посева могут быть использованы бактерии с коротким жизненным циклом, а в качестве количественного критерия интенсивности их жизнедеятельности может использоваться, например, количество q выделяемого ими CO_2 в единицу времени. Таким образом, математическая модель исследуемого явления есть $q = q_0 f(\Theta, \varphi)$, где q_0 и есть численный показатель качества почвы.

На первый взгляд, кажется, что у нас нет никаких данных о виде функции $f(\Theta, \varphi)$, описывающей влияние температуры Θ и влажности φ на жизнедеятельность почвенных бактерий, так как мы не имеем нужных агрономических знаний. Но это не так. Кто не знает, что при $\Theta = 0^\circ\text{C}$ вода замерзает и, следовательно, CO_2 выделяться не может, а при $\Theta = 80^\circ\text{C}$ производится пастеризация, т. е. большинство бактерий погибает. Априорных данных вполне достаточно, чтобы утверждать, что искомая функция имеет колоколообразный характер, близка к нулю при $\Theta = 0^\circ\text{C}$ и $\Theta = 80^\circ\text{C}$ и, следовательно, имеет максимум внутри этого интервала. Подобные же рассуждения о функции влияния влажности позволяют заключить, что она имеет максимум в области $\varphi \approx 20\%$ и спадает до нуля при $\varphi = 0\%$ и $\varphi = 40\%$ (жидкая грязь). Вид этой функции показан ниже на рис. 7-4, а. Таким образом, действительной задачей эксперимента является лишь уточнение характера спадов этого колокола в области $\Theta = 20 \div 30^\circ\text{C}$ и $\Theta = 50 \div 60^\circ\text{C}$, а также при $\varphi = 10 \div 15\%$ и $\varphi = 25 \div 30\%$ и более точное установление координат максимума.

Подобное положение обнаруживается при более внимательном анализе в большинстве других сложных исследовательских задач, поэтому максимальное использование априорных данных, резко сужающих круг возможных моделей, должно быть одним из основных правил при поиске модели сложных явлений.

Диффузность исходных экспериментальных данных является важным фактором, определяющим путь выбора модели. Если эти данные позволяют построить семейства графических зависимостей типа изображенных на рис. 7-1, б, т. е. их диффузность позволяет выделить не только уровень, но и наклон, и кривизну, а возможно, и изменения кривизны частных зависимостей, то в этом случае следует, безусловно, идти путем последовательной аппроксимации этих частных зависимостей, отыскивая затем вид их композиции между собой в единую математическую модель (рис. 7-1, в).

Если же, наоборот, попытки построения частных зависимостей $y = f_1(x_1)$, $y = f_2(x_2)$ и т. д. приводят к расположению экспериментальных точек в виде картины «звездного неба» (рис. 6-3, а), то наиболее рациональным путем поиска решения является зада-

Таблица 7-1

Число N различных градаций отклика	Относительная приве- денная погрешность γ , %	Коэффициент множе- ственной корреляции ρ_{yy}
1	50	0,0
2	25	0,9
3	16	0,96
4	12	0,98
6	8	$0,99 = 1 - 10^{-2}$
20	2,5	$0,999 = 1 - 10^{-3}$
60	0,8	$0,9999 = 1 - 10^{-4}$
200	0,25	$0,99999 = 1 - 10^{-5}$

ние функции отклика в виде описанных в § 7-1 аддитивных (гиперплоскости и гиперboloида) или мультипликативной моделей, так как при значительной диффузности исходных данных (рис. 6-3, а) и большом числе факторов ($k \geq 3$) какие-либо попытки «осмыслить» исходный экспериментальный материал путем построения и анализа частных зависимостей вида $y = f_1(x_1)$, $y = f_2(x_2)$ и т. д. чаще всего совершенно не продуктивны. Они лишь создают у исследователя видимость настойчивой деятельности и надежды, что решение «вот-вот будет найдено», в то время как в действительности отгадать искомую модель в таких условиях почти невозможно.

Использование коэффициента множественной корреляции для оценки качества получаемого решения. Шкала численных значений коэффициента корреляции ρ_{yy} весьма своеобразна. Для ее правильного восприятия полезно сопоставить значения ρ_{yy} , относительной приведенной погрешности γ и числа N различных градаций полосы разброса экспериментальных данных. Такое сопоставление в соответствии с формулами (6-13) и (6-14) сделано в табл. 7-1. При $\gamma = \pm 50\%$ интервал неопределенности составляет $2\gamma = 100\%$, т. е. разброс экспериментальных данных вокруг модели занимает весь диапазон изменения отклика и не позволяет тем самым установить даже угол наклона этой зависимости. Такому положению соответствует $\rho = 0$. При $\gamma = \pm 25\%$ интервал разброса составляет $2\gamma = 50\%$ от диапазона изменения отклика, что позволяет различить только две его градации (от 0 до 50% и от 50 до 100%). Этому соответствует значение $\rho = 0,9$, и поле точек может иметь вид рис. 6-11, б или в. При $\rho = 0,96$ может быть различено $N = 3$ градации отклика, при $\rho = 0,98$ — $N = 4$ и т. д. (см. рис. 6-11, а).

Если считать достаточным для приближенного определения функции отклика возможность различения в поле экспериментальных данных хотя бы 2—3 градаций, то это значит, что приемлемым значением коэффициента корреляции ρ является лишь ρ , большее 0,9—0,96. Весь же диапазон изменения ρ от 0 до 0,9

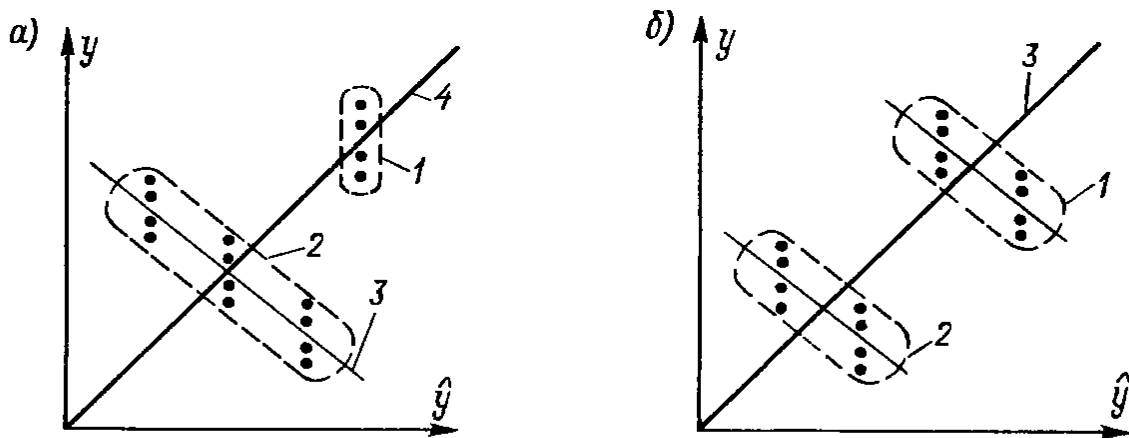


Рис. 7-3

соответствует тому, что в полосе экспериментальных данных невозможно достоверно различить даже двух градаций изменения искомой функции отклика. Какие-либо утверждения о виде функции отклика в этих условиях не могут быть обоснованными. Поэтому утверждения типа «...в результате этого коэффициент корреляции возрос с $\rho = 0,27$ до $\rho = 0,68$, т. е. более чем в 2,5 раза...», не имеют смысла, хотя нередко встречаются, например, для «научного подтверждения» существования телепатии.

Если в результате решения диффузной задачи по модели, назначенной волевым образом, получен коэффициент корреляции $\rho < 0,9$, то задачу нельзя считать окончательно решенной (см. рис. 6-11, б, где $\rho = 0,9$, или рис. 6-11, в, где $\rho = 0,8$).

Использование графических методов анализа. Часто существенную помощь в подобной ситуации оказывает нанесение экспериментальных данных в соответствии с полученной (плохой) моделью на график вида $y = f(\hat{y})$. Причиной неудачи часто является неоднородность исходных экспериментальных данных, вследствие чего график $y = f(\hat{y})$ получает вид рис. 7-3, а или б. Этот график построен на основании тех же данных, которые на рис. 6-11, б или в были представлены в координатах $y = f(x)$. На рис. 7-3 повторена та же нумерация полей и линий, что и на рис. 6-11. Но принципиальное различие этих графиков состоит в том, что поля точек на рис. 6-11 могут быть построены только для однофакторного эксперимента, а графики рис. 7-3 — для сколь угодно сложного многофакторного эксперимента.

Таким образом, если на графике $y = f(\hat{y})$ возникла ситуация, показанная на рис. 7-3, а, то, видимо, мы ввели в ЭВМ для обработки по МНК резко неоднородные исходные данные. Полученная модель в виде линии 4 не соответствует исследуемому явлению, так как полоса 2 экспериментальных точек лежит не вдоль, а поперек линии этой модели. Это вызвано наличием данных в области 1. Логическая часть анализа такой ситуации состоит в принятии решения о том, какую же часть такой статистики надо обрабатывать, а какую — следует отбросить. Можно отбросить крайние группы данных в области 2, тогда центр области 2 и область 1

дадут решение в виде линии 4 с хорошим коэффициентом корреляции ρ_4 . Или же можно опустить из рассмотрения группу данных области 1, тогда три оставшихся группы данных в области 2 дадут решение в виде линии 3 также с хорошим значением ρ_3 . Выбор между этими двумя решениями из чисто математических соображений может легко приводить к абсурду. Поэтому он должен производиться с точки зрения «здравого смысла», а не путем сравнения остаточной дисперсии или, что то же самое, путем сопоставления оценок ρ_3 и ρ_4 .

Так, возвращаясь к рис. 6-11 и вспомнив, что в данном случае y — это электрическое сопротивление медной обмотки, которое с ростом температуры Θ (или одного из $x_j = \Theta$ при многофакторной зависимости) должно возрастать, следует принять решение в виде линии 3, а не линии 4, даже в том случае, если $\rho_3 < \rho_4$.

Если построение данных в виде графика $y = f(\hat{y})$ дает картину рис. 7-3, б, то оно показывает, что решение в виде линии 3 имеет плохой коэффициент корреляции вследствие того, что обе полосы данных 1 и 2 лежат не вдоль, а поперек линии 3. В этом случае имеет смысл провести отдельные решения для полосы 1 и для полосы 2, а затем попытаться представить себе физический смысл всех трех решений. В данном случае мы получим, что решение 3 абсурдно, так как дает для меди отрицательное значение температурного коэффициента, а решения в виде линий 1 и 2 — не абсурдны, так как дают почти равные в обоих случаях положительные значения этого коэффициента. Таким образом, задача в значительной степени уже решена, но остается найти причину расхождения этих двух (правильных по темпу и знаку прироста) решений, различающихся лишь «постоянной составляющей».

Цензурирование данных с использованием «плохой» модели. Картины, подобные рис. 6-11 или 7-3, могут возникнуть как вследствие каких-либо устранимых ошибок (например, наличия промахов), так и в результате диффузности самих исходных экспериментальных данных. В гл. 4 были подробно рассмотрены методы обнаружения и удаления промахов. Но ими невозможно воспользоваться при многофакторном эксперименте, пока он не приведен к виду $y = f(\hat{y})$ с использованием сколь угодно «плохой» модели. Поэтому использование назначенной волевым образом модели позволяет решить эту задачу путем построения на ее основе графика $y = f(\hat{y})$ (рис. 6-14, б или 7-3) и удаления промахов из выбранной полосы.

Однако при решении вопроса о признании некоторых наблюдений отсчетами, относящимися к выборке из другой генеральной совокупности, следует в качестве главного критерия опираться на изложенные выше соображения «здравого смысла», т. е. на известные физические, экономические и т. п. закономерности. Так, например, получив график $y = f(\hat{y})$ (рис. 7-3, а)

по правилам исключения промахов, можно с совершенно равным правом исключить края области 2 и оставить ее центр и область 1 или, наоборот, исключить область 1 и оставить все три группы в области 2. Поэтому в трудных случаях приходится параллельно использовать оба варианта, а выбор производить уже между полученными моделями 3 и 4 по их физическому смыслу.

Подбор аппроксимирующих базисных функций методом выделения остатка. После использования назначенной модели и удаления с ее помощью наблюдений, принадлежащих другой генеральной совокупности, уже может открываться возможность дальнейшего решения задачи методом выделения остатка (см. § 6-2) и подбора аппроксимирующих базисных функций для отдельных факторов или их групп с последующей их композицией в общую формулу модели. При этом, пользуясь методом выделения остатка, при отыскании вида базисных функций путем построения частных зависимостей необходимо каждый раз использовать и те точки (из области 1 на рис. 7-3, а), которые при более «грубой» модели вели себя как промахи. Очень часто при использовании уточненной модели, правильнее учитывающей влияние и взаимодействие всех факторов, многие из точек, казавшихся «промахами» при «грубой» модели, хорошо ложатся в полосу точек при более совершенной модели и даже помогают установить адекватный вид базисных функций.

7-3. МЕТОД ОРТОГОНАЛЬНОЙ РЕГРЕССИИ КАК ЧАСТНЫЙ СЛУЧАЙ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ (МГК)

Основным недостатком МНК является его несимметричность по отношению к аргументу и функции, о чем уже говорилось в § 6-5. Отсюда вытекает практическое ограничение области надежного использования МНК. Переменные при МНК неравноправны между собой, и перенос их из правой части уравнения в левую приводит к другому решению. Но при подготовке аппроксимирующих уравнений к решению МНК такие переносы часто необходимы, как мы уже видели выше на примерах уравнений (6-2) или (6-3) в § 6-4. Вследствие несимметрии МНК в таких случаях вместо решения в виде линии 1 (рис. 6-10) мы получаем решение в виде линии 2. Различие этих решений (см. § 6-5) определяется значением квадрата коэффициента множественной корреляции, так как

$$a_1 \frac{1}{a_2} = \frac{\sum x_i y_i}{\sum x_i^2} \frac{\sum x_i y_i}{\sum y_i^2} = r^2.$$

Поэтому если r близко к единице, то несимметрией МНК можно пренебрегать и переносить переменные из одной части уравнения в другую, перемножать их между собой, делить друг

на друга и т. д., полагая, что при $\rho \approx 1$ и $a_1 \approx a_2$. Если же ρ заметно меньше единицы, то такие действия приводят к существенным невязкам. Таким образом, МНК можно надежно использовать (см. табл. 7-1) лишь при $\rho \geq 0,96$. При этом следует учитывать, что даже при $\rho = 0,96$ $\rho^2 = 0,92$ и различие между a_1 и a_2 , которое может возникнуть из-за переноса переменных из правой части уравнения в левую, составляет 8%, хотя такое отклонение линии регрессии в поле экспериментальных точек для глаза и незаметно. Если при преобразовании уравнений образуются члены, например, вида $z = x^2 y$, то различие может даже утроиться, т. е. достигнуть 24%. Таким образом, необходимо иметь в виду, что при рассеянии исходных данных, характеризующихся $\rho_{y\theta} < 0,96$, решения МНК могут приводить к ошибочным заключениям.

В связи с этим в математической статистике разрабатывались другие методы, которые могли бы заменить МНК в подобных случаях. Один из таких методов, называемый *методом ортогональной регрессии*, описывается Ю. В. Линником [20]. Прямая ортогональной регрессии совпадает с осью симметрии эллипса рассеяния экспериментальных точек, т. е. соответствует на рис. 6-10 прямой 3, а не прямым 1 или 2, определяемым по МНК, т. е. несимметричность МНК устраняется.

В дальнейшем этот метод был развит применительно и к многофакторному эксперименту, когда экспериментальные точки образуют облако рассеяния в многомерном пространстве и задача состоит в нахождении уравнений всех плоскостей симметрии этого облака. В этом случае все переменные (как отклик y , так и все факторы x_i) являются равноправными, так как представляют собой оси координат $(k+1)$ -мерного пространства, в котором и расположено облако рассеяния экспериментальных точек с координатами $(y_1, x_{1i}, \dots, x_{ki})$ и которые в этом случае удобнее обозначать как $x_1, x_2, \dots, x_k, x_{k+1}$.

Для составления уравнений вида $a_1(x_1 - \bar{x}_1) + a_2(x_2 - \bar{x}_2) + \dots + a_{k+1}(x_{k+1} - \bar{x}_{k+1}) = 0$, определяющих положения гиперплоскости в $(k+1)$ -мерном пространстве, поступают так. Сумма квадратов расстояний от экспериментальных точек до этой гиперплоскости

$$M = \frac{\sum_1^n (a_1 x_{1i} + a_2 x_{2i} + \dots + a_{k+1} x_{k+1i})^2}{a_1^2 + a_2^2 + \dots + a_{k+1}^2}.$$

Если дополнительно положить, что $a_1^2 + a_2^2 + \dots + a_{k+1}^2 = 1$ (условие нормирования), то M может быть выражено как $M = a_1^2 D_{x_1} + a_2^2 D_{x_2} + \dots + a_{k+1}^2 D_{x_{k+1}} + 2a_1 a_2 B(x_1, x_2) + \dots + 2a_i a_j B(x_i, x_j) + \dots$, где D_{x_i} — дисперсия фактора x_j , а $B(x_i, x_j)$ — ковариация факторов x_i и x_j при условии $i \neq j$, когда $i = 1, (k+1)$ и $j = 1, (k+1)$.

Поэтому модель, выбираемая по минимуму остаточной дисперсии и по МНК, в большей мере отражает зависимость искомого отклика от шума, чем от исследуемого физического явления.

«Шум» в рассматриваемой задаче — это изменение отклика y под действием факторов, не включенных по тем или иным причинам в рассматриваемую модель. В примере, приведенном на рис. 6-11 или 7-3, «источником» такого «шума» явилось присутствие в части исходных данных систематической погрешности. В других случаях «источниками» такого «шума» могут быть самые разнообразные, но всегда неучитываемые факторы. Их воздействие и приводит к тому, что линия регрессии, имеющая минимальную остаточную дисперсию (например, линия 4 на рис. 6-11, в или 7-3, а), соответствует абсурдному с физической точки зрения решению. Линия же 3 на рис. 6-11, в или 7-3, а, соответствующая искомой физической закономерности, имеет бóльшую остаточную дисперсию и поэтому не может быть найдена МНК.

Описанная же разновидность МГК — это один из возможных способов генерирования целого ряда $n + 1$ возможных моделей (7-5), который применительно к рис. 6-11, в и 7-3, а одновременно дает уравнения как линии 4 с минимальной остаточной дисперсией, так и линии 3 с большей остаточной дисперсией. Поэтому выбор искомой модели из всех предлагаемых МГК моделей (7-5) должен производиться не по минимуму остаточной дисперсии, а по физическому, экономическому, геологическому и т. д. смыслу решаемой задачи.

7-4. МЕТОДЫ ОТБОРА

НАИБОЛЕЕ ЗНАЧИМЫХ ФАКТОРОВ

И НАИБОЛЕЕ ЗНАЧИМЫХ ЧЛЕНОВ АДДИТИВНОЙ МОДЕЛИ

В § 6-2 подчеркивалось, что основное требование к математической модели, обеспечивающее удобство ее последующего использования, — ее компактность. Обеспечить компактность аддитивной модели — это значит не вводить в нее ни лишних членов, ни тем более лишних факторов. Но это требование входит в прямое противоречие с требованием наибольшей точности описания исследуемого явления. Отбор для включения в модель только тех факторов и тех дополнительных членов, которые действительно повышают точность описания исследуемого явления, может быть осуществлен лишь на основе использования соответствующих методов.

Метод определения значимости членов модели по изменению коэффициента множественной корреляции. Он состоит в следующем. После того как вид модели уже выбран, из нее поочередно исключается каждый из членов, проводится новое решение МНК и определяются оценки ρ_{yy} коэффициента множественной корреляции. Рассмотрим этот метод на более конкретном примере. Пусть разыскивается модель от трех факторов $\hat{y} = f(x_1, x_2, x_3)$.

Таблица 7-2

При всех x_j	Без x_1	Без x_2	Без x_3	Без x_4	Без x_5	Без x_6
$\rho_{\Sigma} = 0,9992$	0,88	0,93	0,9995	0,992	0,99	0,99
$(\rho_{\Sigma}^2 - \rho_j^2) \cdot 100\%$	22	13	-0,06	1,4	1,8	1,8
Место по значимости	1	2	6	5	3—4	3—4

После ряда уточнений возникло предположение о целесообразности введения в нее следующих шести членов: $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_1x_2 + a_6\sqrt{x_2/x_1}$. Используя подстановку $x_1^2 = x_4$, $x_1x_2 = x_5$ и $\sqrt{x_2/x_1} = x_6$, приведем ее к виду

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6. \quad (7-6)$$

Результаты расчета значений коэффициентов множественной корреляции $\rho_{y\hat{y}}$ для моделей с поочередным отбрасыванием членов, содержащих $x_1, x_2, \dots, x_6$, дали значения ρ , приведенные в первой строке табл. 7-2.

Коэффициент множественной корреляции отражает размер относительной дисперсии точек модели σ_{Δ}^2 по сравнению с дисперсией значений отклика σ_y^2 , так как $\sigma_{\Delta}^2/\sigma_y^2 = 1 - \rho_{y\hat{y}}^2$. Отсюда для оценки значимости в модели каждого из ее членов удобно пользоваться показателем в виде $\rho_{\Sigma}^2 - \rho_j^2$. Но значения этой разности, как правило, составляют сотые доли единицы, поэтому их удобно умножать на 100, т. е. выражать в процентах (см. табл. 7-2). Сопоставляя эти оценки между собой, можно установить порядковые места по значимости каждого из членов модели.

Метод пошаговой регрессии является некоторым упрощением предыдущего метода с одновременной полной или частичной его автоматизацией. Программа пошаговой множественной регрессии [38, вып. 1, с. 210] состоит в следующем. В ЭВМ вводится массив данных о факторах $x_1, \dots, x_k$, участвующих в аддитивной модели вида $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$. Если вид некоторых членов модели известен заранее (например, известно, что $x_4 = x_1^2$, $x_5 = x_1x_2$ и $x_6 = \sqrt{x_2/x_1}$), то эти подстановки должны быть введены в предварительную программу, которая подготовит массив значений $x_1, \dots, x_k$.

Затем ЭВМ начинает поочередно искать решения МНК для моделей $y_0 = a_0$, $y_1 = a_0 + a_1x_1$, $y_2 = a_0 + a_1x_1 + a_2x_2$, $y_3 = a_0 + a_1x_1 + a_2x_2 + a_3x_3$, ..., определяя для каждой модели остаточные дисперсии $D_0, D_1, D_2, \dots$, а также их разности, опреде-

ляемые внесением в модель каждого последующего фактора, т. е. $\Delta_1 = D_0 - D_1$, $\Delta_2 = D_1 - D_2$, ..., $\Delta_k = D_{k-1} - D_k$.

В автоматическом режиме оператором заранее задается желаемое значение остаточной дисперсии $D_{\text{ост}}$ и программа останавливается и выводит на печать полученные значения коэффициентов регрессии последней i -й модели, у которой остаточная дисперсия $D_i \leq D_{\text{ост}}$.

В полуавтоматическом, диалоговом, режиме производится расчет всех моделей до y_k и на печать выводятся все полученные разности дисперсий от Δ_1 до Δ_k , которые рассматриваются как оценки значимости факторов $x_1, \dots, x_k$. Сопоставляя их между собой, оператор сам делает заключение об относительной значимости факторов.

Недостаток этого метода состоит в том, что значения Δ_i зависят от того, в каком порядке вводились в модель различные факторы. Поэтому в программе предусмотрено, что оператор может задать целесообразный, с его точки зрения, порядок введения в модель различных факторов, устанавливаемый из каких-то предварительных данных или из сопоставления результатов первого прогона программы.

Метод приближенного расчета коэффициентов значимости часто именуется в литературе методом β -коэффициентов.

После подстановок любое уравнение модели можно привести к виду уравнения (7-6). Поэтому задача формулируется как определение относительного веса членов уравнения (7-6) по их относительным вкладам в результирующее значение отклика y .

Перепишем эти равенства для измененной системы отсчета, когда начало координат перенесено в точку пространства с координатами $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_l$. Тогда

$$\hat{y}_i - \bar{y} = a_1(x_{i1} - \bar{x}_1) + a_2(x_{i2} - \bar{x}_2) + \dots,$$

где $i = \overline{1, n}$.

Проведем теперь операцию нормирования, разделив все члены этого равенства на σ_y , а члены правой части, кроме того, умножив и разделив на σ_{xi} . Тогда получим

$$\frac{\hat{y}_i - \bar{y}}{\sigma_y} = a_1 \frac{\sigma_{x1}}{\sigma_y} \frac{x_{i1} - \bar{x}_1}{\sigma_{x1}} + a_2 \frac{\sigma_{x2}}{\sigma_y} \frac{x_{i2} - \bar{x}_2}{\sigma_{x2}} + \dots,$$

где $i = \overline{1, n}$. Здесь $(\hat{y}_i - \bar{y})/\sigma_y$, $(x_{i1} - \bar{x}_1)/\sigma_{x1}$ и т. д. — центрированные и нормированные случайные величины, представленные выборками объема n , а $\beta_1 = a_1\sigma_{x1}/\sigma_y$, $\beta_2 = a_2\sigma_{x2}/\sigma_y$, ..., $\beta_l = a_l\sigma_{xl}/\sigma_y$ — по существу, суть веса их относительного вклада в значение отклика.

Удобство использования β -коэффициентов значимости членов аддитивной модели обусловлено простотой их расчета по формуле $\beta_j = a_j\sigma_{xj}/\sigma_y$. Однако при использовании этого метода необходимо иметь в виду следующие его особенности. Так как σ_{xj} и σ_y — величины положительные, то β получает знак коэффициента

регрессии a_j , который может быть как положительным, так и отрицательным. Поэтому не следует удивляться, что иногда получается, например, $\beta_j = -2$, но $\beta_2 = +2,5$, а $\beta_3 = -0,5$. Знаки при β -коэффициентах во внимание не принимаются, а значимость оценивается по соотношению их модулей.

Приближенность β -коэффициентов определяется прежде всего тем, что случайность малой выборки приводит к существенному разбросу оценок и σ_{xj} , и a_j , который зависит от объема выборки n , числа определяемых коэффициентов l , эксцесса распределений ε , и коэффициента множественной корреляции ρ_{yy} . Так, например, при $l = 4$, $n = 24$, $\rho = 0,9$ и $\varepsilon = 3$ относительное с. к. о. β_j составляет $\delta_\beta = 20\%$, а следовательно, $\gamma_{0,9} = 1,6 \delta_\beta = 32\%$.

Поэтому при использовании метода β -коэффициентов следует придерживаться следующего правила. Если β -коэффициенты двух членов модели различаются между собой менее чем в 1,3—1,5 раза в ту или другую сторону, то такие члены должны признаваться равнозначными. Различная значимость признается лишь тогда, когда β -коэффициенты различаются более чем в 1,3—1,5 раза.

7-5. ОЦЕНКА ПАРАМЕТРОВ ОБЛАСТИ НЕОПРЕДЕЛЕННОСТИ ИСХОДНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МНОГОФАКТОРНОГО ЭКСПЕРИМЕНТА

Вид области неопределенности исходных данных многофакторного эксперимента. Как было уже показано, при определении зависимости $y = f(x)$ область неопределенности экспериментальных данных — это полоса в двумерном пространстве, на плоскости (x, y) (см. рис. 6-1). При этом разделить погрешности, внесенные при измерении переменных x или y , уже невозможно (см. § 6-1). Поэтому неопределенность экспериментальных данных при измерении зависимости $y = f(x)$ характеризуется совокупно (и по x , и по y) указанием коэффициента корреляции или других параметров совокупной полосы неопределенности.

При многофакторном эксперименте для определения многомерной зависимости $y = f(x_1, \dots, x_k)$ это явление еще более усиливается, так как область неопределенности обусловлена в этом случае суммированием всех погрешностей измерений и y , и $x_1, \dots, x_k$.

Другой особенностью области неопределенности исходных экспериментальных данных является то, что эта область в данном случае представляет собой уже не интервал (рис. 4-1, а) и не полосу (рис. 4-1, б), а $(k + 1)$ -мерный слой, располагающийся по обе стороны $(k + 1)$ -мерной поверхности отклика.

Изобразить это графически можно только для двумерного пространства факторов, располагающегося на плоскости x_1, x_2 , и трехмерного пространства отклика $y = f(x_1, x_2)$ — рис. 7-4, а. Здесь изображена функция отклика $y = y_0 f(\Theta, \Phi)$ жизнедеятельности почвенных бактерий в зависимости от температуры Θ

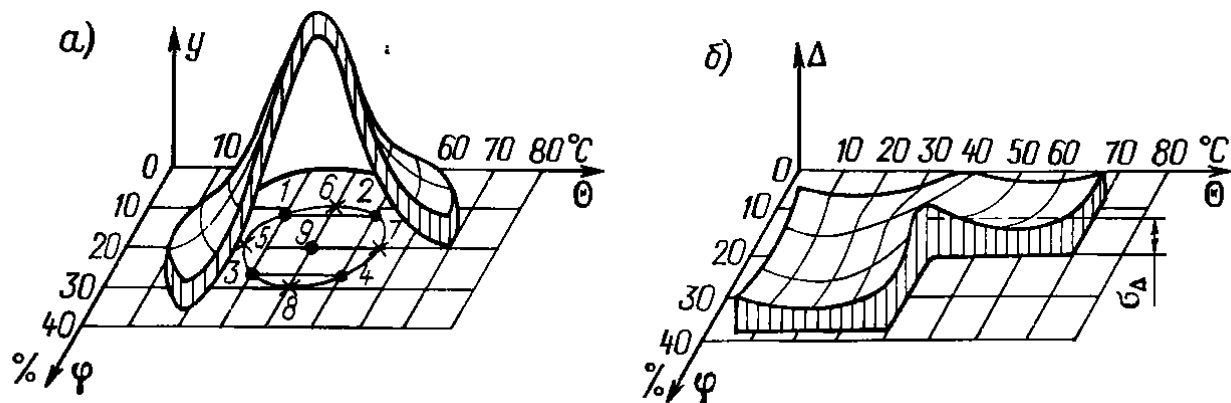


Рис. 7-4

и влажности φ почвы (см. § 7-2). В результате суммирования всех случайных погрешностей, допускаемых при измерении $x_1 = \Theta$, $x_2 = \varphi$ и y , экспериментальные точки в трехмерном пространстве x_1 , x_2 , y занимают также трехмерный слой с толщиной в направлении оси y , равной 2Δ .

Толщина 2Δ этого слоя отнюдь не обязана оставаться постоянной для любого сочетания k факторов, входящих в функцию отклика. Скорее как раз наоборот, она может существенно изменяться в зависимости от сочетания значений факторов, например, так, как это показано на рис. 7-4, б. Здесь при $\varphi = 20\%$ и $\Theta = 40^\circ\text{C}$ полутолщина слоя погрешности Δ достигает максимума, при отклонении температуры Θ и влажности φ от этих значений Δ сначала падает, а затем вновь возрастает.

Наиболее полное описание погрешностей, допускаемых при многофакторном эксперименте, состоит в указании также многофакторной зависимости для с. к. о. или доверительных границ погрешности в виде $\Delta = F(x_1, \dots, x_k)$, которая позволила бы вычислить погрешность Δ_i каждого из результатов измерений при любом сочетании факторов $x_1, \dots, x_k$. Найти такую зависимость не труднее, чем многофакторную функцию отклика $\hat{y} = f(x_1, \dots, x_k)$.

Располагая фактическими данными об изменении этой погрешности в функции от различных сочетаний факторов (см. рис. 7-4, б), можно найти коэффициенты этой модели в соответствии с фактическими данными. А в дальнейшем обе формулы должны использоваться совместно (совместно закладываться в программу ЭВМ), чтобы для каждого y_i , вычисленного как $\hat{y} = f(x_1, \dots, x_k)$, указывалась и его погрешность: $\Delta = F(x_1, \dots, x_k)$ (см. § 3-6).

Упрощенное описание погрешностей экспериментальных данных может состоять в сообщении значений погрешности, возникающей в характерных точках слоя (рис. 7-4, б), т. е. при некоторых определенных сочетаниях факторов. Такими точками на рис. 7-4, б могли бы быть: точка при $\varphi = 20\%$ и $\Theta = 40^\circ\text{C}$, соответствующая максимуму y и одновременно локальному максимуму Δ , точки при крайних значениях $\Theta = 0$, $\Theta = 80^\circ\text{C}$,

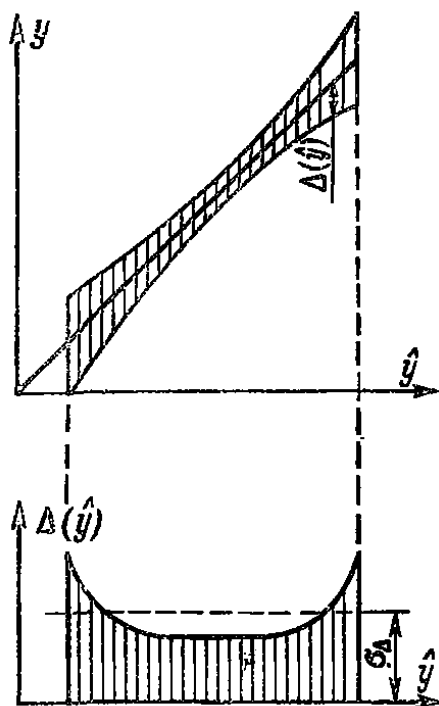


Рис. 7-5

$\varphi = 0$ и $\varphi = 40\%$, где погрешность достигает больших значений, а также промежуточные точки ($\varphi = 20\%$, $\Theta = 60^\circ\text{C}$, $\varphi = 30\%$, $\Theta = 40^\circ\text{C}$), где погрешность минимальна. Такой метод указания погрешности в некоторых произвольных точках и является сейчас наиболее распространенным.

Предельно упрощенное указание погрешности Δy как функции от значений $\hat{y}$. Если толщину слоя погрешностей представить не в функции $x_1 = \Theta$ и $x_2 = \varphi$, как это было сделано на рис. 7-4, б, а в виде полосы на плоскости с координатами y и $\hat{y}$, то вместо многомерного слоя сложной конфигурации получим относительно простое изображение полосы неопределенности (рис. 7-5, а). В этом случае зависи-

мость текущего значения погрешности $\Delta(\hat{y})$ в функции от значений $\hat{y}$ отклика может быть представлена графиком (рис. 7-5, б), т. е. соответствующей (одно-, двух-, трехчленной) формулой в функции от $\hat{y}$.

Однако такой однозначной зависимости погрешности Δ от значений результата $\hat{y}$, как на рис. 7-5, в действительности может и не существовать (см. § 3-6—3-8). Поэтому графики рис. 7-5 имеют смысл лишь для конкретной случайной выборки, т. е. отражают лишь частную реализацию полосы встретившихся в данной выборке погрешностей, соответствующих данному значению результата $\hat{y}$ при конкретных сочетаниях факторов x_j . Статистическая оценка ширины этой полосы соответствует некоторому усредненному значению фактических погрешностей. Тем не менее благодаря предельной простоте такое описание погрешностей может быть очень полезным на практике.

Совершенно аналогичный этому усредненный смысл имеет указание как значения коэффициента множественной корреляции $\rho_{y\hat{y}}$, так и вычисляемой через его оценку средней квадратической σ или энтропийной Δ_s погрешностей. Уровень такой средней квадратической погрешности, усредненной по всему пространству факторов, показан штриховой прямой на рис. 7-4, б и 7-5, б и помечен σ_Δ .

Естественно, что подобную усредненную оценку области неопределенности экспериментальных данных следует рассматривать лишь как самую примитивную, хотя она и приводится под внушающим уважение названием коэффициента множественной корреляции.

7-6. РАСЧЕТ ПАРАМЕТРОВ МНОГОМЕРНОГО СЛОЯ НЕОПРЕДЕЛЕННОСТИ УСРЕДНЕННОЙ МОДЕЛИ МНОГОФАКТОРНОЙ ЗАВИСИМОСТИ

Принципиальные соотношения между полосой неопределенности исходных экспериментальных данных и полосой неопределенности получаемой усредненной модели применительно к однофакторной функциональной зависимости были рассмотрены при обсуждении рис. 4-1, б. Они аналогичны и для модели многофакторной функциональной зависимости, но с той разницей, что если на рис. 4-1 они рассматривались на плоскости, то для многофакторной модели они должны рассматриваться в многомерном пространстве.

В качестве примера рассмотрим эти соотношения для зависимости вида $y = f(x_1, x_2)$ с двухмерным пространством факторов и трехмерным пространством отклика. Для упрощения задачи положим, что толщина слоя погрешности экспериментальных данных, по всем точкам пространства факторов постоянна, это аналогично аддитивной полосе неопределенности экспериментальных данных рис. 6-10. В этих условиях слой, в котором разбросаны значения Δ , случайных погрешностей экспериментальных данных, в отличие от рис. 7-4, б имеет форму параллелепипеда (рис. 7-6).

Слой неопределенности полученной модели благодаря усреднению существенно тоньше слоя разброса экспериментальных данных и его толщина определяется соотношениями, аналогичными (6-16). Поэтому сечения этого слоя параллельно каждой из k осей факторов $x_1, \dots, x_k$ имеют вид, который был показан на рис. 6-15. В связи с этим слой неопределенности усредненной модели в частном случае двухмерного пространства факторов представляет собой тело, имеющее форму двояковогнутой линзы, боковая поверхность которой на рис. 7-6 заштрихована. В общем случае k -мерного пространства факторов это не двояковогнутая линза, а k -вогнутое тело, которое в дальнейшем будем именовать *мениском погрешностей*.

Задача определения параметров области неопределенности усредненной модели сводится к расчету погрешности неопределенности модели, т. е. характерных размеров многомерного мениска погрешностей. Характерными размерами мениска погрешностей на рис. 7-6 следует считать его толщину

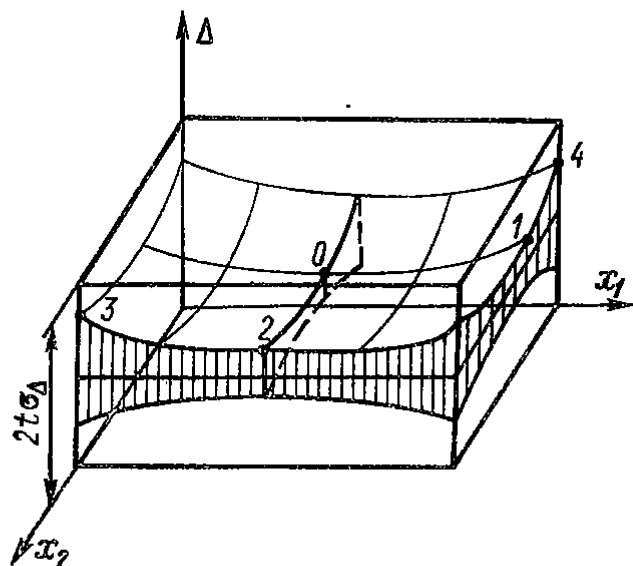


Рис. 7-6

рис. 7-4). Однако при выводе формул (7-7)—(7-9) мы исходили из того, что толщина слоя разброса экспериментальных данных всюду одинакова. Здесь следует различать два случая.

Если неравномерность толщины слоя разброса симметрична относительно центра описываемой моделью области пространства факторов (рис. 7-4), то использование вместо функциональной зависимости $\sigma = F(x_1, \dots, x_k)$ постоянной величины σ_Δ не приводит к какой-либо ошибке, так как все коэффициенты модели определяются путем усреднения по всему полю экспериментальных точек.

Но если неравномерность толщины слоя разброса экспериментальных данных несимметрична относительно центра описываемой области (рис. 6-12 или 6-13, б и в), то различие погрешностей в зонах I и III (рис. 6-13, в), например, в 10 или 100 раз приведет, естественно, к существенной несимметрии толщины слоя (рис. 7-6). В подобной ситуации бессмысленно заменять профиль части рельефа (рис. 7-4, б) каким-либо «средним» значением σ_Δ . (Что считать средним между 1 Ом и 1 МОм — 1 кОм или 500 кОм?) В этом случае целесообразно пространство факторов разбить на некоторые участки (подобно участкам I, II и III на рис. 6-13, в), вычислить параметры менисков погрешностей для каждого из участков, затем попытаться составить по этим данным общую картину, а возможно, найти и общую математическую модель для приближенного расчета параметров единого несимметричного мениска погрешностей.

Второе допущение состоит в том, что, говоря об экспериментально обследованной области пространства факторов, мы предполагали, что полученные экспериментальные точки равномерно заполняют всю исследованную область, например, так, как это показано на рис. 7-1, б или в.

Такое расположение экспериментальных точек, когда обеспечивается **равноточность** исследования по всем направлениям пространства факторов, в теории эксперимента принято называть *ротатабельным*. Так, например, классический ротатабельный план для поверхности второго порядка изображен на плоскости факторов x_1, x_2 (рис. 7-4, а). Он состоит из девяти опытов. Точки 1—8, координаты которых в плоскости x_1, x_2 соответствуют задаваемым сочетаниям факторов и располагаются на окружности, т. е. на равных расстояниях от точки 9, с координатами $\bar{x}_1, \bar{x}_2$ и равных расстояниях между собой. Результаты y_i этих опытов на поверхности отклика будут расположены на рис. 7-4, а также по окружности или эллипсу на склоне колокола (для точек 1—8) и на вершине колокола (для точки 9). При постановке опытов в соответствии с таким ротатабельным планом конфигурация области неопределенности полученной модели будет иметь вид центросимметричного мениска (рис. 7-6).

Однако если эксперимент выполняется не в соответствии с заранее разработанным ротатабельным планом (рис. 7-4, а), а про-

извольным образом, то мениск области неопределенности полученной модели будет несимметричным. Таким образом, при постановке активного эксперимента [2], согласно заранее разработанному ротатабельному плану, расчет погрешностей существенно упрощается и, наоборот, при пассивном эксперименте, т. е. при стихийной постановке исследования, — усложняется.

ГЛАВА ВОСЬМАЯ

ПУТИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ИЗМЕРИТЕЛЬНОГО ЭКСПЕРИМЕНТА

✱

8-1. ПРОБЛЕМА ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Как отмечалось во введении, экспериментальные исследования в различных областях науки, техники и производства (от сельского хозяйства до медицины и космоса) составляют весьма существенную часть всей человеческой деятельности, а следовательно, и национального продукта. Поэтому эффективность экспериментальных исследований имеет существенное практическое и экономическое значение и заслуживает всестороннего анализа.

Конечно, методы оценки погрешностей как непосредственно исходных экспериментальных данных, так и результатов их обработки, т. е. те задачи, которые рассматривались в предыдущих семи главах книги, безусловно, нужны и важны для оценки эффективности эксперимента. Но эффективность определяется не только качеством полученного результата, а сопоставлением этого качества с теми **з а т р а т а м и**, которые потребовались для его достижения.

Такого комплексного рассмотрения проблемы эффективности процессов измерений нет ни в отечественной, ни в зарубежной литературе. Предлагаемый вниманию читателя набросок возможного пути анализа этой проблемы, естественно, также весьма далек от ее исчерпывающего решения. Однако хотелось бы, чтобы тот перечень вопросов, которые далее будут затронуты в этой главе, явился бы толчком к разработке этого важного направления.

Сложившееся положение в какой-то степени можно объяснить тем, что лавинообразное развитие **с р е д с т в** измерений (электронные приборы, автоматические мосты и потенциометры, затем цифровые приборы, измерительные информационные системы и, наконец, микроэлектроника, ИВК и микропроцессоры) захлестнуло в последние десятилетия (1950—1990 гг.) всех работающих в области измерительной техники и почти полностью отвлекло их внимание от разработки **м е т о д о в** **и з м е р е н и й**. В результате этого в последние 30 лет прекратилось издание книг и монографий в области методов измерений. Последними книгами, которые еще содержали в своем названии слова «методы измерений», были, по-видимому, книги К. Б. Карандеева «Методы электрических измерений» (1953 г.) и «Специальные методы электрических измерений» (1963 г.).

Тем не менее, несмотря на длительное отсутствие внимания к этой проблеме как со стороны метрологов, так и теоретиков в области измерительной техники, она в эти годы получила серьезное развитие благодаря усилиям специалистов в области математической статистики и практиков, особенно в области химико-технологических измерений. Обсуждавшийся в § 7-1 феномен «проклятия размерности» особо резко проявляется именно при постановке химико-технологических измерений. Необходимость учитывать большое число компонентов состава реагирующих веществ, большое число изменяющихся в ходе процесса

концентраций, температур, давлений и т. п. приводит к тому, что в классической постановке такой эксперимент затягивается на долгие годы.

В подобной ситуации фактор **з а т р а т в р е м е н и** выступает сам собой на первый план и решающим становится не достигаемая точность, а именно эффективность эксперимента, как соотношение между получаемым результатом и необходимыми для этого затратами времени. Эта практическая потребность заставила вспомнить о выдвинутом еще в конце 20-х годов предложении Рональда Фишера по преодолению «проклятия размерности», но не нашедшем тогда практического применения. Теперь же, в 70-е годы, это привело к возникновению и успешному развитию нового научного направления, получившего известность под наименованием «теории планирования многофакторного эксперимента».

Основное направление этой теории состоит в разработке путей предельно возможной **э к о н о м и и з а т р а т** на проведение измерений путем использования предельно упрощенных математических моделей и предельного сокращения числа экспериментально определяемых точек. Поэтому использование теории планирования эксперимента может явиться одним из путей существенного повышения эффективности экспериментальных исследований.

Эта теория заслуживает серьезного изучения, например, по книгам [2, 9, 21, 26, 46]. Мы же ограничимся лишь самым кратким ознакомительным описанием методов планирования, разработанных в этой теории, и некоторыми практическими рекомендациями по ее использованию. Но проблема повышения эффективности измерительного эксперимента далеко не исчерпывается использованием только теории планирования, а связана с решением других задач, которые до сих пор не имеют не только известных решений, но часто не имеют даже четко сформулированной постановки задачи.

Так, в работах по теории планирования эксперимента подробно рассматривается лишь размещение опытов в пространстве факторов. При этом совершенно не затрагивается вопрос о выборе точности средств измерений для проведения данного исследования. По-видимому, выбор СИ также является одним из важных этапов планирования эксперимента и в существенной степени определяет его эффективность. Повышение точности может быть достигнуто не только путем увеличения числа усредняемых отсчетов, но и использованием более точной аппаратуры и более совершенных методик измерения. Но использование более точной методики измерений или более точной аппаратуры, как правило, также сопряжено с повышением затрат времени на единичное измерение, так как измерения с погрешностью, например, 5% могут быть проведены с много меньшими затратами времени и средств, чем измерения с погрешностью 0,01%.

Для решения проблемы повышения эффективности эксперимента крайне желательно знать, как повышается точность измерений с увеличением затрат времени на их проведение. В теории планирования оценка величины затрат времени на проведение эксперимента производится достаточно условно — только в виде числа поставленных опытов. Такой учет затрат весьма примитивен. Желателен комплексный учет как всех затрат времени, так и достигаемого эффекта. При этом необходимо сформулировать математически четкое определение самой эффективности эксперимента.

Ясно, что эффективность эксперимента — это его экономическая эффективность и должна математически выражаться как отношение достигнутого эффекта к потребовавшимся для этого затратам. Но затраты времени на проведение эксперимента складываются не только из затрат времени на проведение самих измерений, но включают в себя еще и время на подготовку эксперимента, регулировку аппаратуры, ожидание окончания переходных процессов при установке задаваемых значений факторов (например, температуры), затраты времени на обработку результатов и т. д.

В последние годы наряду с методами статистического усреднения случайных погрешностей широко используется автоматическое исключение систематических погрешностей путем применения образцовых сигналов — так называемые «тестовые методы» (см., например, [5] и др.). При этом приобретает исключительное значение совместный учет влияния как на точность получаемых результатов, так и на затраты времени наличия систематических и случайных погрешностей. Совместный учет нередко отсутствует в известных работах в этой области. При

анализе случайных погрешностей их полагают независимыми, т. е. без особой осторожности используют соотношение $\gamma_{\bar{x}} = \gamma_{xi}/\sqrt{n}$, а при анализе коррекции систематических погрешностей часто забывают учесть возрастание дисперсии случайных погрешностей. Часто вообще исключают сам факт существования систематических погрешностей и обнадеживают себя и читателя возможностями усреднения десятков тысяч отсчетов. В то время как введение поправок, как было отмечено в § 1-6, приводит к удвоению дисперсии случайных погрешностей, а в работе [8] обстоятельно показано, что тестовые методы могут привести к возрастанию случайной погрешности в 20, а иногда и в 200 раз. Поэтому совместный анализ всех этих вопросов совершенно необходим для оценки эффективности измерительного эксперимента.

8-2. ИСПОЛЬЗОВАНИЕ ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Основные понятия теории многофакторного эксперимента были изложены в § 7-1. Там было показано, что большая трудоемкость многофакторного эксперимента обусловлена катастрофическим возрастанием числа необходимых опытов с ростом числа учитываемых факторов. Поэтому все усилия теории планирования эксперимента направлены на повышение его эффективности путем предельного сокращения числа проводимых опытов. Это достигается жесткой экономией во всех возможных направлениях: использованием предельно упрощенных математических моделей, предельным сокращением числа определяемых точек (не более двух — для определения прямой, не более трех — для кривой второго порядка, $k + 1$ точек — для гиперплоскости, $(k + 1)(k + 2)/2$ — для модели второго порядка, где k — число варьируемых факторов).

Казалось бы, что при такой предельной экономии не может быть и речи о статистическом усреднении, требующем многих избыточных измерений. Однако предложенный в 1925—1929 гг. английским статистиком Рональдом Фишером принцип замены многих однофакторных экспериментов единым многофакторным позволяет использовать для перекрестного усреднения данные опытов, которые при однофакторных экспериментах принадлежали разным однофакторным зависимостям.

Но для этого размещение опытов в пространстве факторов (что, собственно, и называют в теории эксперимента его планированием) должно быть произведено особым образом, называемым «оптимальным планом эксперимента». Так, например, из 25 точек рис. 7-1, в при оптимальном плане второго порядка опыты должны быть поставлены только в девяти точках, обведенных на рис. 7-1, в квадратными рамками.

Замечательным свойством оптимальных планов является то, что они не только позволяют сократить число опытов, но одновременно обеспечивают лучшее усреднение, т. е. меньшую дисперсию разброса усредненного результата, чем то, которое достигалось при проведении отдельных однофакторных экспериментов.

Для иллюстрации этого воспользуемся примером из книги В. В. Налимова «Теория эксперимента» (М.: Наука, 1971, стр. 137—139). Пусть задача состоит во взвешивании на весах трех тел с массами m_A , m_B и m_C . Традиционно используют план эксперимента, который представлен в табл. 8-1, где знак «+» означает взвешивание, знак «—» — отсутствие взвешивания, а y_0 , y_A , y_B и y_C — полученные отсчеты. Опыт № 1 используется для определения знака и размера систематической погрешности весов y_0 , а опыты № 2, 3, 4 — для взвешивания тел А, В и С. Результаты определяются как $m_A = y_A - y_0$, $m_B = y_B - y_0$ и $m_C = y_C - y_0$, т. е. систематическая погрешность весов исключается.

Если веса обладают случайной погрешностью со с. к. о. σ , то дисперсия каждого из результатов будет $\sigma^2(m_A) = \sigma^2(y_A - y_0) = \sigma^2(y_A) + \sigma^2(y_0) = 2\sigma^2$, т. е. равна удвоенной дисперсии погрешности весов.

На удвоение дисперсии случайной погрешности при введении поправки

Таблица 8-1

Номер опыта	A	B	C	Отсчет
1	—	—	—	y_0
2	+	—	—	y_A
3	—	+	—	y_B
4	—	—	+	y_C

Таблица 8-2

Номер опыта	A	B	C	Отсчет
1	+	—	—	y_1
2	—	+	—	y_2
3	—	—	+	y_3
4	+	+	+	y_4

для исключения систематической погрешности необходимо обратить пристальное внимание, особенно при автоматическом исключении систематических погрешностей в ИИС и ИВК путем использования образцовых сигналов для определения поправок к результатам измерений.

Оптимальный план многофакторного эксперимента для взвешивания тел A, B и C приведен в табл. 8-2, где вместо проверки нуля весов для исключения их систематической погрешности (опыт № 1 в табл. 8-1) производится совместное взвешивание всех трех тел (опыт № 4 в табл. 8-2). Различие этих планов состоит в том, что план табл. 8-1 состоял из трех независимых однофакторных экспериментов, так как для определения m_A результаты опытов № 3 и № 4 не нужны, а для определения m_C не нужны результаты опытов № 2 и № 3, т. е. каждое из искомых m_A , m_B и m_C находится из своего однофакторного эксперимента.

В случае же плана табл. 8-2, т. е. общего многофакторного эксперимента, ни один из результатов не может быть найден иначе, как при использовании всех четырех отсчетов по формулам

$$m_A = (y_1 - y_2 - y_3 + y_4)/2; \quad m_B = (y_2 - y_1 - y_3 + y_4)/2;$$

$$m_C = (y_3 - y_1 - y_2 + y_4)/2.$$

Систематическая погрешность весов при этом также исключается, так как в каждую из формул она входит два раза со знаком плюс и два раза со знаком минус. Однако вследствие того, что $D[kx] = k^2 D[x]$, дисперсия результата будет $\sigma^2(m_A) = D[0,5(y_1 - y_2 - y_3 + y_4)] = 0,25(4\sigma^2) = \sigma^2$, т. е. в два раза меньше, чем при использовании неоптимального плана эксперимента табл. 8-1.

Это происходит от того, что при первом плане при каждом вычислении использовались лишь два отсчета, а два других опускались. При втором же плане усреднение каждый раз производится по всем четырем отсчетам, а каждое тело взвешивается не один, а два раза.

Таким образом, этот оптимальный план позволяет получить в два раза меньшую дисперсию усредненного результата при том же числе опытов ($n = 4$). При первом плане такая дисперсия могла быть получена при усреднении по восьми опытам (при двукратном повторении всех четырех опытов табл. 8-1), т. е. второй план позволяет в два раза сократить число опытов при той же дисперсии усредненного результата.

Уменьшение числа опытов, а следовательно, и затрат времени на проведение измерений, при использовании оптимальных планов оказывается тем значительней, чем более многомерным является пространство факторов, так как снижение дисперсии результата пропорционально числу одновременно варьируемых факторов.

• Таким образом, использование теории планирования эксперимента может явиться одним из путей существенного повышения эффективности многофакторных экспериментальных исследований. Поэтому далее приводится краткое изложение методики составления оптимальных планов эксперимента для моделей первого и второго порядка, обсуждается их эффективность и приводятся соображения по эвристическому составлению близких к оптимальным планов более высоких порядков.

8-3. ОПТИМАЛЬНЫЕ ПЛАНЫ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ ДЛЯ МОДЕЛЕЙ ПЕРВОГО ПОРЯДКА

План первого порядка для двух переменных. Пусть искомая функция двух переменных $y = f(x_1, x_2)$ описывается моделью в виде плоскости $y = a_0 + a_1x_1 + a_2x_2$ или в виде гиперболоида $y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$. Расположение этой модели в пространстве изображено жирным четырехугольником 1, 2, 3, 4 на рис. 8-1. Расположение необходимых для ее определения опытов показано точками 1, 2, 3, 4 на плоскости факторов x_1, x_2 . Опыты ставятся при значениях факторов x_1 и x_2 , равных $x_{1\min}$ и $x_{1\max}$, $x_{2\min}$ и $x_{2\max}$. В табл. 8-3 представлен план такого эксперимента, где знак «+» означает $x_{\max}$, а знак «—» означает $x_{\min}$.

Для модели в виде гиперболоида этот план является предельно экономным, или, как принято говорить в теории планирования эксперимента, *насыщенным*. Действительно, в модели гиперболоида должны быть найдены четыре коэффициента a_0, a_1, a_2 и a_3 . Это может быть сделано только путем решения системы четырех уравнений и, следовательно, требует постановки четырех опытов. Но решение такой системы единственно, т. е. поверхность гиперболоида точно пройдет через все четыре экспериментальных значения y_i , а следовательно, такой насыщенный (предельно экономный) эксперимент не позволяет усреднить случайные погрешности и не дает никаких сведений о их размере.

Для модели в виде плоскости этот план является *ненасыщенным* (избыточным), так как для определения трех искомых коэффициентов (a_0, a_1 и a_2) имеется четыре уравнения, соответствующих четырем опытам. Это уже позволяет произвести усреднение и оценить размер случайных погрешностей.

Другими словами, коэффициент a_1 наклона поверхности отклика к оси x_1 может быть найден (см. рис. 8.1) как из наклона прямой 1—2, так и из наклона прямой 3—4 на плоскости y, x_1 , а коэффициент a_2 при x_2 — из наклона прямых 1—3 и 2—4 на плоскости y, x_2 , и затем произведено их усреднение.

Это наглядно видно, если уравнение плоскости представить в виде $y = a_0 + a_1(x_1 - \bar{X}_1) + a_2(x_2 - \bar{X}_2)$, где $\bar{X}_1 = (x_{1\min} + x_{1\max})/2$ и $\bar{X}_2 = (x_{2\min} + x_{2\max})/2$, т. е. перенести начало координат в точку $\bar{X}_1, \bar{X}_2$. Тогда коэффициент a_0 , как высота центра плоскости 1, 2, 3, 4 на рис. 8-1, находится усреднением всех четырех значений y_i .

Следует заметить, что этот прием переноса начала координат в центр пространства факторов с координатами $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ очень полезен при обработке

данных любых экспериментов, описываемых моделью в виде гиперплоскости, так как позволяет получить более устойчивое усредненное решение для a_0 .

В результате усреднения найденная плоскость будет удовлетворять всем четырем значениям y_i лишь в среднем. В каждой из точек может быть найдена

Таблица 8-3

Номер опыта	x_1	x_2
1	—	—
2	+	—
3	—	+
4	+	+

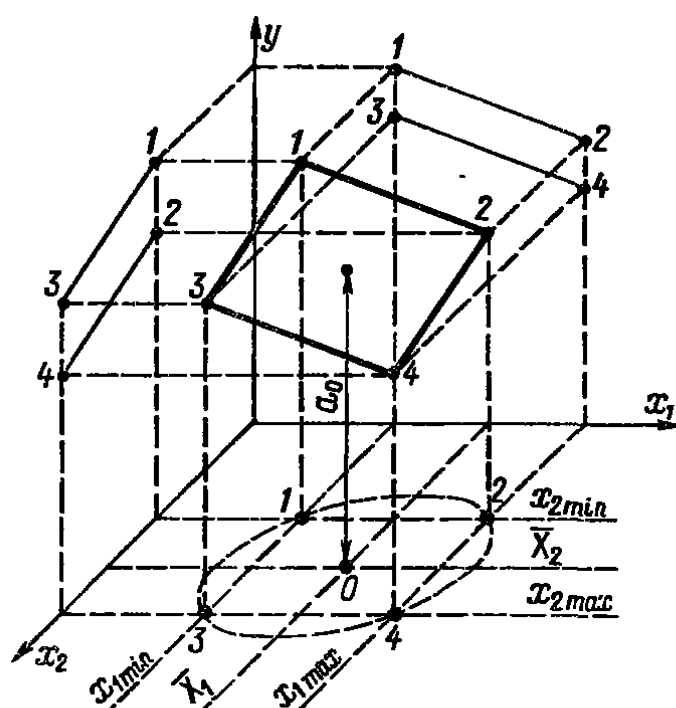


Рис. 8-1

погрешность отклонения экспериментальных данных относительно модели и по этим четырем отклонениям может быть вычислено их с. к. о.

Таким образом, наличие всего одного избыточного опыта оказывается очень полезным, так как позволяет (хотя и грубо, так как $n = 4$) оценить случайные погрешности и получить модель с их усреднением.

Этот избыточный опыт (например, точка 4 на рис. 8-1) может быть и исключен, но тогда план эксперимента станет *нерототабельным*, т. е. не равноточным по всем направлениям, так как в направлении 3—2 в плоскости факторов будет обеспечена большая точность, чем в направлении 1—0. Для восстановления рототабельности точки 1, 2, 3 опытов в плоскости факторов должны быть равноудалены как друг от друга, так и от центра 0, т. е. располагаться в вершинах равностороннего треугольника, вписанного в показанную на рис. 8-1 окружность. Но это может привести к возрастанию затрат времени при постановке опытов, так как по крайней мере один из факторов в этом случае должен варьироваться уже не на двух, а на трех уровнях.

Планы для модели гиперboloида при трех и более факторах. При трех переменных x_1 , x_2 и x_3 пространство факторов является трехмерным, а пространство отклика — четырехмерным. Поэтому представить в аксонометрии можно только пространство факторов x_1 , x_2 , x_3 , как это показано на рис. 8-2, а функцию отклика изобразить невозможно.

Для модели в виде четырехмерного гиперboloида $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1x_2 + a_5x_2x_3 + a_6x_1x_3 + a_7x_1x_2x_3$ искомыми являются восемь коэффициентов ($a_0, a_1, \dots, a_7$), для определения которых минимально необходимо иметь восемь опытов. Обеспечение рототабельности плана требует, чтобы они находились в пространстве факторов на равном удалении друг от друга и равных расстояниях от центра. Этим требованиям удовлетворяет размещение опытов в восьми вершинах куба, вписанного в сферу с центром, совпадающим с центром плана. Так как все сечения поверхности гиперboloида, параллельные осям x_1 , x_2 , x_3 , суть прямые линии, то для их определения достаточно двух точек, т. е. каждый из факторов должен варьироваться только на двух уровнях — $x_{\min}$ и $x_{\max}$. Размещение опытов в вершинах многомерного куба как раз и удовлетворяет всем этим требованиям.

Графически изобразить рототабельный план для большего, чем 3, числа факторов невозможно. Его можно только записать в виде таблицы, подобной табл. 8-3, обозначая по-прежнему $x_{\min}$ знаком «—» и $x_{\max}$ — знаком «+».

Особой заслугой теории планирования эксперимента является то, что она указала очень простое правило составления таких планов-таблиц для любого числа переменных, т. е. для координат всех вершин многомерного гиперкуба в k -мерном пространстве факторов любой размерности. Это правило гласит, что для модели в виде гиперboloида знаки «—» и «+» в столбцах плана-таблицы должны чередоваться по правилу чередования двоичных цифр в разрядах двоич-

Таблица 8-4

Номер опыта	x_1	x_2	x_3
1	—	—	—
2	+	—	—
3	—	+	—
4	+	+	—
5	—	—	+
6	+	—	+
7	—	+	+
8	+	+	+

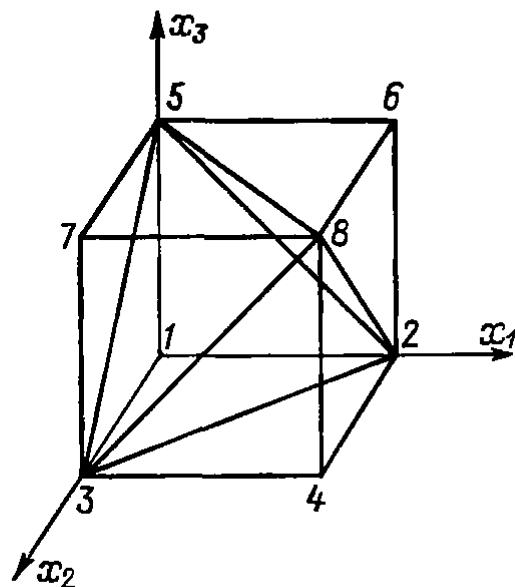


Рис. 8-2

ного кода, т. е. в столбце для x_1 — через один, в столбце для x_2 — через два, для x_3 — через четыре, для x_4 — через восемь, а для x_k — через 2^{k-1} .

Пример плана-таблицы для модели четырехмерного гиперboloида с тремя факторами приведен в табл. 8-4, а номера, присвоенные таким образом вершинам куба, помечены на рис. 8-2. Пользуясь этим правилом, можно составить планы-таблицы для модели гиперboloида для любого числа факторов.

Общим свойством этих планов является то, что они всегда насыщены, т. е. предельно экономны. Это обусловлено тем, что гиперкуб в k -мерном пространстве факторов имеет $n = 2^k$ вершин (квадрат в двухмерном пространстве — четыре вершины, куб в трехмерном — восемь вершин, гиперкуб в четырехмерном пространстве — 16 вершин и т. д.), а модель в виде гиперboloида содержит также $n = 2^k$ искоемых коэффициентов. Поэтому эти планы дают единственное решение, т. е. не позволяют усреднить данные и оценить размер случайных погрешностей получаемых по ним исходных данных, если не повторять все опыты плана по два или более раз. В теории планирования эти планы получили название *полных факторных планов*, или сокращенно — ПФП (ГОСТ 24026—80).

Планы для модели в виде гиперплоскости. В отличие от модели гиперboloида, требующей определения 2^k неизвестных коэффициентов, модель в виде гиперплоскости $y = a_0 + a_1x_1 + \dots + a_kx_k$ содержит всего $k + 1$ искомый коэффициент и, следовательно, требует не менее чем $n = k + 1$ опытов для их определения. Поэтому ПФП для гиперплоскости сильно избыточен. Для определения искоемых коэффициентов достаточно использовать лишь его некоторую часть, называемую в теории планирования *дробной репликой*, или в соответствии с ГОСТ 24026—80 — *дробным факторным планом*, сокращенно — ДФП.

Если дробление ПФП производится многократным последовательным делением числа его опытов на 2, то реплика называется *регулярной*. Число раз p такого последовательного деления называется *дробностью реплики*, а число опытов регулярного ДФП составляет $n = 2^{k-p}$. При $p = 1$ ДФП называют *полурепликой* (1/2-репликой), при $p = 2$ — *четвертьрепликой* (1/4-репликой), при $p = 3$ — *1/8-репликой*, при $p = 4$ — *1/16-репликой* и т. д.

Правило составления планов-таблиц для регулярных дробных реплик. Составление плана начинают с минимального уровня каждого фактора, обозначаемого знаком «—». Для столбцов первых факторов, пока обеспечивается чередование знаков, план составляется по изложенному выше правилу чередования цифр в разрядах двоичного кода. Начиная же с того столбца, где по правилу двоичного кода должны быть поставлены только знаки «—» и чередование знаков отсутствовало бы, знаки находятся по правилу перемножения знаков данной строки уже заполненных столбцов плана (x_1x_2 , x_2x_3 , x_1x_3 , $x_1x_2x_3$ и т. д.).

Рассмотрим это правило на примере составления плана для трех факторов x_1 , x_2 и x_3 . Полный факторный план для $k = 3$ содержал бы $2^3 = 8$ опытов, размещенных на поверхности сферы на равных расстояниях между собой и равных длин ребер куба, вписанного в эту сферу (см. рис. 8-2). Для модели гиперплоскости $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$ необходимо определить лишь 4 коэффициента, следовательно, достаточно полуреплики, 4 опытов из 8, указанных на рис. 8-2 и в табл. 8-4. Но для сохранения рототабельности плана они должны размещаться на поверхности сферы, описывающей куб на рис. 8-2, на равных расстояниях между собой.

Приведенное правило и позволяет найти такое их расположение. Для этого первые два столбца плана (табл. 8-5) заполняются по правилу двоичного кода, а третий столбец находится по правилу перемножения как $x_3 = x_1x_2$. Если теперь сравнить сочетание факторов в табл. 8-5 и 8-4, то можно заметить, что в табл. 8-5 вошли 5, 2, 3 и 8-й опыты из табл. 8-4, что отмечено в последнем столбце табл. 8-5. Таким образом, указанное правило из ПФП отбирает ДФП в виде регулярной реплики.

В данном случае это — полуреплика (4 опыта из 8), опыты которой размещены в вершинах правильного тетраэдра, который и был изображен на рис. 8-2. Оставшиеся же четыре вершины куба образуют вторую полуреплику, располагаясь в вершинах второго возможного тетраэдра.

Таблица 8-5

Номер опыта	x_1	x_2	x_3	Номер опыта по табл. 8-4
1	—	—	+	5
2	+	—	—	2
3	—	+	—	3
4	+	+	+	8

Простейшие правильные фигуры или тела (равносторонний треугольник на плоскости, тетраэдр в трехмерном пространстве и т. д.) в стереометрии называются *симплексами*. Поэтому полученные описанным путем ДФП, опыты которых располагаются в вершинах симплексов, называются в теории планирования *симплекс-планами*.

При определении гиперплоскости как отклика четырех факторов, когда ПФП содержит $2^4 = 16$ опытов, для определения 5 коэффициентов ее уравнения приходится брать симплекс-план в виде полуреплики, содержащей 8 опытов, так как четвертьреплика из 4 опытов, естественно, недостаточна для нахождения 5 коэффициентов. Такой план будет ненасыщенным, так как три опыта в нем избыточны. Но опустить их нельзя из-за нарушения ротатабельности, ибо оставшиеся опыты будут располагаться в пространстве факторов не симметрично относительно центра.

Параметры симплекс-планов для моделей в виде гиперплоскости с разным числом k факторов приведены в табл. 8-6, из которой видно, что симплекс-планы для модели в виде гиперплоскости могут быть как насыщенными (при $k = 3$, $k = 7$ и т. д.), так и содержать несколько избыточных опытов.

В последнем случае для уменьшения погрешности адекватности модели в нее могут быть дополнительно введены некоторые члены модели гиперболоида. Например, при $k = 4$ полуреплика содержит 8 опытов, модель же гиперплоскости требует определения только 5 коэффициентов. Оставшиеся в избытке 3 опыта позволяют ввести в модель еще 3 дополнительных члена в виде произведений факторов из оставшихся $16 - 5 = 11$ членов модели гиперболоида. Но для усреднения случайных погрешностей рационально оставить 1—2 избыточных опыта, т. е. ввести лишь одно или два наиболее значимых произведения факторов (см. § 7-5).

Если экспериментальные данные ограничены результатами 8 опытов ДФП, то для оценки значимости дополнительных членов последовательно рассчитывают по МНК несколько моделей, вводя в них по очереди различные двойки дополнительных членов. Но этот путь перебора весьма трудоемок и не очень надежен,

Таблица 8-6

Число факторов k	Число коэффициентов модели $k + 1$	Число опытов ПФП	Вид плана	Число опытов плана	Избыточность
2	3	4	ПФП	4	1
3	4	8	Полуреплика	4	0
4	5	16	Полуреплика	8	3
5	6	32	Четвертьреплика	8	2
6	7	64	1/8-реплика	8	1
7	8	128	1/16-реплика	8	0
8	9	256	1/16-реплика	16	7

так как один и тот же член может иметь малую значимость в сочетании с одними членами и большую — в сочетании с другими.

Более же обоснованный отбор будет в том случае, если есть возможность хотя бы однажды поставить и получить данные всех 16 опытов ПФП и рассчитать одновременно значимость всех членов полной модели гиперboloида, а в дальнейшем использовать лишь один или два наиболее значимых дополнительных члена, содержащих произведения факторов.

В работах по планированию эксперимента часто указывается, что в модели гиперboloида значимы лишь двойные произведения, а тройные и более произведения, как правило, малозначимы. Однако бывают и исключения. Например, решалась задача по подбору модели, описывающей действие прибора для определения текущего значения выработанного ресурса двигателей внутреннего сгорания. Радиорезонансным методом определялась динамика содержания шлама (мельчайших металлических опилок) в системе смазки двигателя. Рассматривая взвешенную в масле металлическую частицу как резонатор, поглощающий радиоизлучение, в модель кроме числа частиц x_1 был введен фактор x_2 в виде средней длины этих частиц как резонаторов. Адекватность модели была недостаточной, и для ее повышения кроме длины частиц было введено также произведение средней толщины x_3 и ширины x_4 этих частиц как оценка их проводимости (поперечного сечения). Но адекватность оставалась плохой. Тогда в используемую модель экспериментаторы ввели сначала все двойные произведения, затем все тройные произведения, а, отчаявшись, и четверное произведение $x_1x_2x_3x_4$. Значимость члена $x_1x_2x_3x_4$ оказалась неожиданно столь высока, что всеми остальными членами модели гиперboloида можно было уверенно пренебречь. Так, в итоге наращивания полной модели гиперboloида удалось «открыть», что поглощение определяется суммарной массой частиц, которая действительно пропорциональна произведению их числа, длины, толщины и ширины, а остальные факторы дают лишь небольшие поправки.

Существенное улучшение адекватности модели, как показывает практика, чаще достигается не перебором возможных взаимных произведений факторов, а переходом к моделям более высоких порядков, т. е. введением членов, содержащих x^2 , x^3 и т. д. Это становится понятным, если вспомнить, что введение членов, содержащих взаимные произведения первых степеней, может описать только расходящийся веер прямых линий, показанный на рис. 7-2, б, а для описания веера кривых, показанного на рис. 7-1, б, необходимы произведения более высоких степеней.

Оценка и усреднение случайных погрешностей экспериментальных данных при насыщенных, предельно экономных планах эксперимента, как уже указывалось выше, полностью отсутствуют. Они не особенно надежны и тогда, когда для усреднения используются 1—2 избыточных опыта. Поэтому всякий раз, когда позволяют условия эксперимента, целесообразно идти на увеличение затрат времени на проведение измерений ради более полного усреднения случайной погрешности исходных данных и более надежной оценки ее размера.

При использовании насыщенных и ненасыщенных оптимальных планов любого вида (ПФП, ДФП) это производится повторением всех опытов в оптимального плана несколько раз (2—3 или более). При этом все положительные качества оптимальных планов сохраняются, но усреднение происходит, например, уже не по 8, а по 16, 24, 32 опытам.

8-4. ОПТИМАЛЬНЫЕ ПЛАНЫ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ ДЛЯ МОДЕЛЕЙ ВТОРОГО ПОРЯДКА

При двух факторах модель функции отклика $y = f(x_1, x_2)$ второго порядка представляет собой поверхность в виде цилиндра, конуса, эллипсоида, параболоида, гиперboloида, параболического цилиндра, параболического гиперboloида и т. д., описываемую в общем случае уравнением вида

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_1x_2 + a_5x_2^2.$$

Сечения этой поверхности, параллельные каждой из осей x_1 и x_2 , в общем случае представляют собой кривые также второго порядка (окружности, эллипсы,

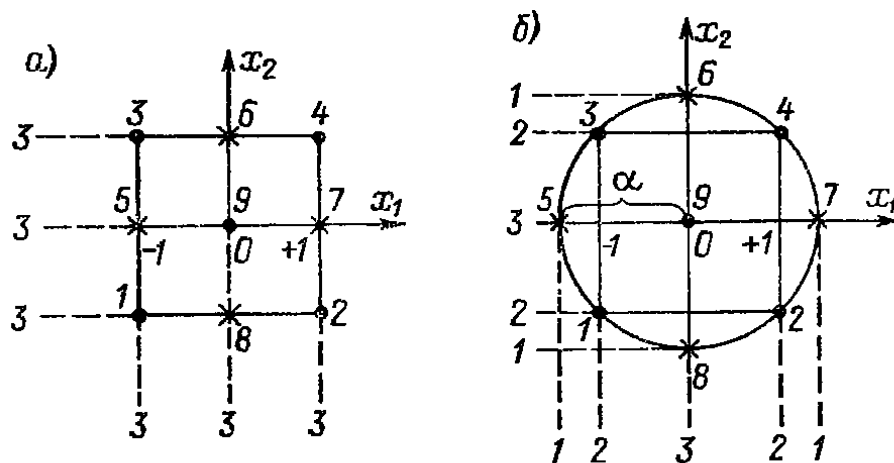


Рис. 8-3

параболы, гиперболы). Для определения такой кривой необходимо располагать координатами не менее трех ее точек, т. е. факторы x_1 и x_2 должны варьироваться не менее чем на трех уровнях.

Поэтому план эксперимента в плоскости факторов x_1 и x_2 на рис. 8-3, а не может состоять лишь из опытов 1, 2, 3, 4, располагающихся в вершинах квадрата, как это было для модели в виде плоскости, а к ним должны быть добавлены опыты 5, 6, 7, 8, располагающиеся в серединах сторон квадрата, и обязательно точка 9 в центре квадрата, чтобы по любому направлению (5—9—7, 1—9—4 и т. д.) располагались три точки, определяющие кривизну поверхности в этом направлении.

Представленный на рис. 8-3, а прямоугольный план эксперимента для модели второго порядка вполне работоспособен, хотя и несколько избыточен (9 опытов для определения 6 коэффициентов). Благодаря трем избыточным опытам он позволяет усреднить случайные погрешности и оценить их размер.

Но теория показывает, что такой план не рототабелен, так как удаление от центра точек 5, 6, 7, 8 в $\sqrt{2}$ раз меньше, чем удаление точек 1, 2, 3, 4. Для обеспечения рототабельности этого плана точки 5, 6, 7, 8 (их принято обозначать звездочками, как это сделано на рис. 8-3, б, и называть «звездными») нужно удалить от центра на расстояние α , называемое *звездным плечом* (ГОСТ 24026—80), в $\sqrt{2}$ раз большее, чем удаление точек 1, 2, 3, 4 от осей x_2 или x_1 . В результате этого все точки плана на рис. 8-3, б оказываются лежащими на окружности.

Размещение точек такого рототабельного плана второго порядка относительно поверхности отклика было показано на рис. 7-4, а, на котором видно, что при колоколообразной поверхности отклика значения y_i , соответствующие опытам 1—8, лежат на некоторой замкнутой кривой (окружности, эллипсе), опоясывающей колоколообразную функцию отклика, а значение y_i , соответствующее точке 9 центра плана, лежит в области вершины колокола.

Учитывая существенно большее влияние на функцию отклика случайной ошибки опыта в точке 9, если она лежит вблизи вершины колокола, рекомендуется ставить в этой точке плана не один, а n_0 дублирующих опытов для усреднения получаемых результатов. Если же поверхность отклика имеет вид, показанный на рис. 7-1, в, когда центральная точка плана ничем не выделяется среди других его точек, постановка нескольких дублирующих опытов в этой точке представляется неоправданной.

Общее правило построения рототабельных планов второго порядка для произвольного числа k факторов. За «ядро» плана принимают симплекс-план или ПФП линейной модели для k факторов. К нему добавляется $2k$ звездных точек (по две на оси каждого из факторов) на расстоянии α от центра плана. И, наконец, планируется один или n_0 опытов в центре плана.

Звездное плечо α (см. рис. 8-3, б) определяется по формуле

$$\alpha = 2^{\frac{k-p}{2}}, \quad (8-1)$$

Таблица 8-7

Число факторов k	Вид «ядра» плана	Число точек в «ядре»	Число звездных точек $2k$	Звездное плечо α	Рекомендуемое число n_0	Всего опытов n
2	ПФП	4	4	1,41	5	13
3	ПФП	8	6	1,68	6	20
4	ПФП	16	8	2,00	7	31
5	1/2-реплика	16	10	2,00	6	32
6	1/2-реплика	32	12	2,38	9	53
7	1/2-реплика	64	14	2,83	14	92

§

Таблица 8-8

Часть плана	Номер опыта	x_1	x_2	Часть плана	Номер опыта	x_1	x_2
«Ядро» плана	1	-1	-1	Звездные точки	5	-1,4	0
	2	+1	-1		6	+1,4	0
	3	-1	+1		7	0	+1,4
	4	+1	+1		8	0	-1,4
				Центр	9	0	0

Таблица 8-9

Часть плана	Номер опыта	x_1	x_2	x_3
«Ядро» плана	1	-1	-1	-1
	2	+1	-1	-1
	3	-1	+1	-1
	4	+1	+1	-1
	5	-1	-1	+1
	6	+1	-1	+1
	7	-1	+1	+1
	8	+1	+1	+1
Звездные точки	9	-1,7	0	0
	10	+1,7	0	0
	11	0	-1,7	0
	12	0	+1,7	0
	13	0	0	-1,7
	14	0	0	+1,7
Центр	15	0	0	0

где k — число факторов, а p — дробность реплики (для ПФП $p = 0$, для полу-реплики $p = 1$, для четвертьреплики $p = 2$, для $1/8$ -реплики $p = 3$ и т. д.).

В результате использования этих правил параметры рототабельных планов второго порядка принимают для различного числа факторов k значения, приведенные в табл. 8-7.

В ряде книг по планированию эксперимента значения звездного плеча α часто приводятся с указанием 5—6 десятичных знаков. Конечно, записать эти значения в виде $\alpha = \sqrt[4]{2} = 1,4142135$ или $\alpha = \sqrt[4]{2^3} = 1,6817927$ можно, но вряд ли что-либо подобное можно надеяться выдержать в реальном эксперименте. Чтобы приблизить это к практике измерений, по-видимому, необходимо округлить эти значения: $1,41 \approx 1,4$, $1,68 \approx 1,7$ и т. д. В заключение приведем таблицы рототабельных планов экспериментов второго порядка для двух (табл. 8-8) и трех (табл. 8-9) факторов.

8.5. ПРАКТИЧЕСКИЕ СООБРАЖЕНИЯ ПО ПЛАНИРОВАНИЮ ЭКСПЕРИМЕНТА

Литература по планированию эксперимента, к которой мы отослали читателя в § 8-1 для более глубокого изучения теории этого направления, отличается той особенностью, что все рекомендации этой теории поданы так, что производят впечатление непреложных истин, которым необходимо неукоснительно следовать. Однако общая теория не может, естественно, предусмотреть всех нюансов, которые встречаются в задачах практика-исследователя. Поэтому хотелось бы ориентировать читателя рассматривать разработанные в теории планирования эксперимента положения не как догму, а как руководство к действию, как аналог, пример возможных решений, приглашающий к размышлению и принятию собственных, возможно, волевых или эвристических решений в данной конкретной ситуации.

В этой связи прежде всего необходимо обратить внимание читателя на следующее. В § 6-3 и 6-4, а также в § 7-2 и 7-3 уже отмечалось принципиальное отличие неформализуемой задачи выбора в и д а н с к о м о й математической модели исследуемого явления и чисто формальной задачи нахождения по экспериментальным данным значений параметров в уже выбранной аппроксимирующей функции.

Теперь же, ознакомившись с основами теории планирования эксперимента, следует твердо запомнить, что планирование эксперимента согласно изложенным принципам предназначено для решения второй, формальной задачи и может быть осуществлено только после установления вида математической модели исследуемого явления.

Действительно, выбор того или иного плана эксперимента возможен лишь после того, как стало известно, что исследуемое явление описывается моделью в виде или гиперплоскости, или гиперболоида, или поверхности второго порядка и т. д.

Отсюда основная область применения планирования эксперимента — это планирование массовых однотипных испытаний, когда вид модели известен заранее, а задача состоит в выборе наиболее экономного плана испытаний для установления численных значений параметров этой модели. Наиболее типичный пример — это испытание серийно изготавливаемых изделий с целью определения заносимых в паспорт индивидуальных для каждого экземпляра значений их параметров. Например, значений чувствительности, нелинейности, погрешностей от колебания температуры и питающего напряжения серийно изготавливаемых датчиков.

Тем не менее идеи и основные положения теории планирования эксперимента оказываются весьма полезными и при постановке исследовательских экспериментов, целью которых является первая из указанных выше неформализуемая задача поиска вида математической модели впервые исследуемого явления.

В этом случае, естественно, до получения ряда результатов исследования не может быть составлен окончательный, исчерпывающий план проведения эксперимента. В теории эксперимента такие исследования принято именовать (с замет-

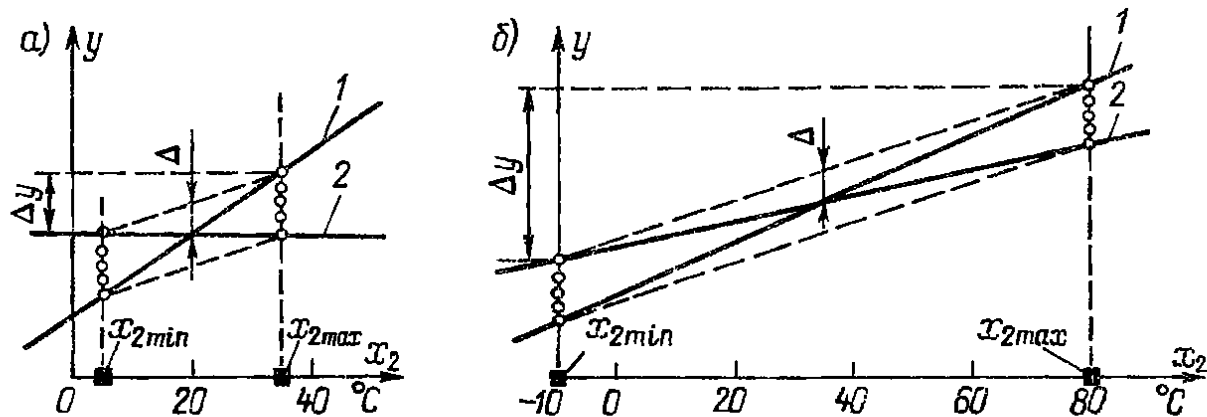


Рис. 8-4

ной долей пренебрежения) как «пристрелочные», или «предварительные». Но при серьезных научных исследованиях новых явлений, новых образцов техники или новых технологических процессов именно эти «пристрелочные» эксперименты и составляют само существо поискового исследования.

План такого эксперимента не может быть составлен по какой-то заранее разработанной классической схеме, он будет творчески изменяться экспериментатором в ходе проводимого исследования по мере выяснения все новых и новых данных о свойствах исследуемого явления. Но при разработке и модификациях этих промежуточных планов проведения поискового исследования (и даже при планировании и проведении соответствующих расчетов) использование идей и положений теории планирования эксперимента позволяет экспериментатору существенно повысить эффективность своего труда. Но при этом не следует бездумно, как некое везде и всюду пригодное решение, использовать конкретные рекомендации теории планирования эксперимента.

Пусть, например, планируется двухфакторный эксперимент по определению температурной погрешности прибора в разных точках его шкалы от 0 до 100 делений, предназначенного по паспорту для работы в диапазоне температур от $+5$ до $+35$ °C. Модель для описания зависимости температурной погрешности y от этих двух факторов x_1 и x_2 выбрана в виде плоскости, т. е. задача в точности соответствует рис. 8-1.

Под обозначениями -1 и $+1$ в теории планирования понимаются границы $x_{\min}$ и $x_{\max}$ диапазона варьирования факторов. Следует ли планировать эксперимент, строго полагая $x_{1\min} = 0$ делений, $x_{1\max} = 100$ делений, $x_{2\min} = +5$ °C и $x_{2\max} = +35$ °C? Нет, не следует.

Знак температурной погрешности заранее не известен, и она может оказаться как положительной, так и отрицательной. При выборе $x_{1\min} = 0$ и отрицательной погрешности указатель исследуемого прибора уйдет на неградуированную часть шкалы и отсчет показаний будет неопределенным. То же самое может произойти и при $x_{1\max} = 100$ и положительной погрешности. Поэтому рациональнее назначить $x_{1\min} = 5$ делений, а $x_{1\max} = 95$ делений.

Размер искомой температурной погрешности прибора достаточно мал: он имеет тот же порядок, что и основная погрешность прибора Δ . Поэтому, назначив $x_{2\min} = 5$ °C и $x_{2\max} = 35$ °C при разбросе получаемых отсчетов в интервале $\pm\Delta$ (см. обозначения на рис. 8-4, а), мы можем столкнуться с тем, что изменение отклика Δy при переходе от 5 °C к 35 °C окажется того же порядка, что и разброс $\pm\Delta$ экспериментальных данных вследствие основной погрешности прибора. В результате оценка температурного коэффициента прибора будет малодостоверной (линии 1 и 2 на рис. 8-4, а).

Выход из подобного положения состоит в возможно большем увеличении диапазона варьирования фактора x_2 , как это показано на рис. 8-4, б, когда при том же абсолютном значении случайной погрешности Δ исходных данных значительно больший прирост отклика Δy существенно уменьшает диапазон неопределенности наклона линий 1 и 2 на рис. 8-4, б, т. е. искомого температурного коэффициента.

Максимально допустимые пределы варьирования факторов определяются границами области существования данного исследуемого явления, пока оно не

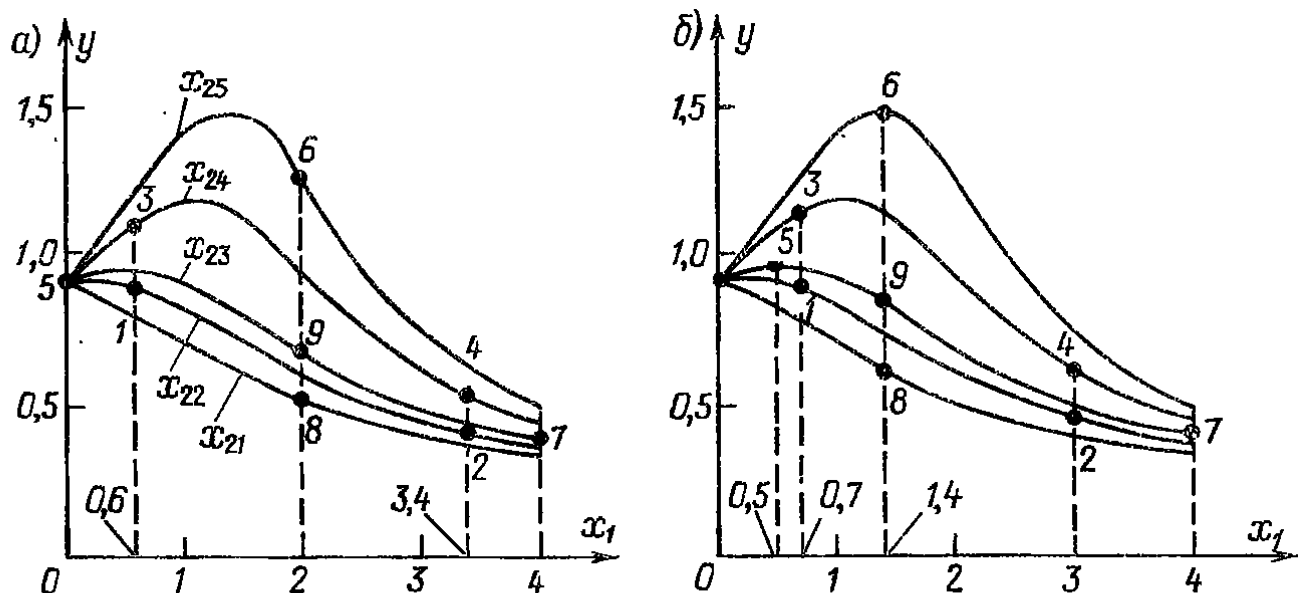


Рис. 8-5

сменяется другими явлениями. Так, например, детали подвижной части электроизмерительных приборов часто приклеиваются шеллаком, который плавится при $90-95^\circ\text{C}$. Поэтому в данном случае диапазон варьирования можно расширить до $x_{2\min} = -10^\circ\text{C}$ и $x_{2\max} = +80^\circ\text{C}$, а для проверки справедливости предположения о линейной зависимости Δy от x_2 может быть дополнительно проверена точка при $x_2 = 20^\circ\text{C}$.

В рототабельных планах второго порядка теория планирования рекомендует относительные значения уровней варьирования факторов назначать равными $-1,41, -1, 0, +1, +1,41$. Выше уже обсуждалась возможность округления значений этих уровней. Но как быть, если практически удобнее принять значение уровня не 1,4, а, скажем, 1,5? Теоретически план станет нерототабельным, но как это повлияет на точность получаемого решения? Если оно находится по МНК, то практически никак (теоретически погрешность будет в 1,02 раза выше потенциально возможной). Но, строго говоря, такой план уже нельзя именовать «оптимальным», а следует пользоваться термином «план, близкий к оптимальному».

После составления оптимального или близкого к оптимальному плана второго или более высокого порядка очень полезно (если это возможно) нанести его точки на графическое изображение поверхности отклика и оценить, действительно ли оптимально они располагаются на этой поверхности? Если в результате этого выяснится, что положение соответствует изображенному на рис. 7-4, а, то может быть принято решение о проведении в центре плана n_0 опытов в соответствии с табл. 8-7. Если же ситуация соответствует рис. 7-1, в, то, видимо, можно принять $n_0 = 4$ или даже $n_0 = 1$.

Но при такой проверке может оказаться, что точки составленного по всем правилам оптимального плана на данной конкретной поверхности отклика располагаются отнюдь не оптимальным образом. В качестве примера такой ситуации на рис. 8-5, а изображена поверхность отклика $y = f(x_1, x_2)$ в виде семейства кривых $y = f(x_1)$ при некоторых постоянных значениях фактора x_2 , обозначенных у каждой кривой в виде ее параметра x_{21}, x_{22} и т. д.

Семейство имеет вид резонансных кривых, поэтому можно предположить, что поверхность отклика является поверхностью второго порядка, и воспользоваться готовым оптимальным планом, который был представлен на рис. 8-3, б.

Так как фактор x_1 изменяется в пределах от $x_{1\min} = 0$ до $x_{1\max} = 4$, то крайние точки плана (точки 5 и 7 на рис. 8-3, б) естественно расположить на средней кривой ($x_2 = x_{23}$) при $x_1 = 0$ и $x_1 = 4$, как это показано на рис. 8-5, а. Тогда три точки (6, 9, 8) центрального сечения плана расположатся при $x_1 = 2$. И, наконец, точки 1 и 3 и точки 2 и 4 должны быть отнесены от центрального сечения плана на $2/\sqrt{2} = 1,4$, т. е. будут располагаться соответственно при $x_1 = 0,6$

и $x_1 = 3,4$. Такое расположение точек постановки опытов и показано на рис. 8-5, а.

Однако нанесение точек составленного плана на фактические кривые семейства поверхности отклика показывает, что они располагаются плохо, так как характерные для исследуемого явления максимумы кривых с x_{24} и x_{25} при таком плане эксперимента будут пропущены. В результате модель, полученная по данным такого, казалось бы, строго спланированного эксперимента, может оказаться достаточно далекой от фактической функции отклика.

Причина состоит в том, что, во-первых, при планировании не позаботились о совмещении центрального сечения плана с максимумом функции отклика и, во-вторых, план на рис. 8-3, б, предназначенный для эллипсоида, параболоида и т. д., предполагал симметрию функции отклика, а резонансные кривые при исследовании их в широком диапазоне частот и линейном масштабе оси x_1 существенно несимметричны. Но известно, что они становятся более симметричными при нанесении оси x_1 в логарифмическом масштабе. Поэтому выход из положения состоит в применении готового оптимального плана, представленного на рис. 8-3, б, к этим данным, перестроенным в логарифмическом масштабе по оси x_1 . Однако в логарифмическом масштабе точка $x_1 = 0$ уходит в минус бесконечность и диапазон варьирования x_1 от 0 и до 4 не может быть использован. Поэтому приходится принять волевое решение — ограничить его снизу, скажем, значением $x_{1\min} = 0,5$.

Тогда интервал варьирования переменной $\lg x_1$ будет простираться от $\lg 0,5$ до $\lg 4$. Центром этого интервала будет точка $\lg x_1 = \lg 1,4$, что почти совпадает с абсциссой максимума кривой при $x_2 = x_{25}$. Полагая округленно значение звездного плеча $\alpha = 1,5$, получаем уровни варьирования переменной $\lg x_1$, соответствующие уровням варьирования плана на рис. 8-3, б — $-1,5, -1, 0, +1, +1,5$, равными $\lg 0,5, \lg 0,7, \lg 1,4, \lg 3$ и $\lg 4$, а следовательно, уровни варьирования переменной x_1 равными $0,5, 0,7, 1,4, 3$ и 4 .

Преобразованный таким образом план эксперимента изображен на рис. 8-5, б, и расположение его точек на кривых отклика подтверждает, что модель, полученная по таким данным, будет лучше отражать форму исследуемой многомерной функции.

Подобный анализ при планировании эксперимента может оказаться полезным во многих случаях практики. Так, например, в § 7-3 и 7-6 обсуждалась функция отклика, изображенная на рис. 7-4, которая описывает интенсивность жизнедеятельности почвенных бактерий и является функцией двух переменных: температуры почвы θ и ее влажности φ . Она описывается экспоненциальной моделью вида

$$y = y_0 e^{-\left(\frac{\theta - \theta_0}{a}\right)^2 - \left(\frac{\varphi - \varphi_0}{b}\right)^2},$$

аналогичной двумерному нормальному распределению, а не моделью в виде параболоида. Поэтому формально план двухфакторного эксперимента второго порядка (рис. 8-3, б) для нее неприменим. Но план для модели в виде двумерного нормального распределения в теории планирования также не разработан. Однако если эту функцию прологарифмировать, то получим

$$\ln y = \ln y_0 - \left(\frac{\theta - \theta_0}{a}\right)^2 - \left(\frac{\varphi - \varphi_0}{b}\right)^2,$$

т. е. достаточно перейти к логарифмическому масштабу лишь по оси y , как эта функция отклика получает вид параболоида, для которого и разработан план на рис. 8-3, б. Но для эффективного применения этого плана к данной задаче необходимо, чтобы центр плана в плоскости факторов был бы достаточно близок к координатам максимума этой функции, как это показано на рис. 7-4, а.

Если же к планированию подходить чисто формально и не заботиться о совмещении центра плана на рис. 8-3, б с максимумом функции на рис. 7-4, а, то точки плана могут расположиться на поверхности отклика подобно точкам на рис. 8-5, а

и решение, полученное по таким исходным данным, может быть далеким от истинного.

Приведенные примеры призваны показать, что формальное использование теории планирования эксперимента приводит только к недоразумениям и разочарованию. Эта теория требует обдуманного, творческого ее использования, и только в этом случае она может послужить существенному повышению эффективности экспериментальных исследований для широкого круга практических задач.

8.6. ЭВРИСТИЧЕСКИЕ ПЛАНЫ ЭКСПЕРИМЕНТА ДЛЯ МОДЕЛЕЙ ВЫСОКИХ ПОРЯДКОВ

Планирование многофакторного эксперимента, как это было показано выше, разработано в теории планирования лишь для простейших моделей в виде гиперплоскости (симплекс-планы), гиперболоида (ПФП) и для моделей в виде поверхности второго порядка типа параболоида, эллипсоида и т. п. Причем описание плана второго порядка приводится далеко не во всех книгах по теории планирования эксперимента: из приведенных в списке литературы — лишь в книге [46] и кратко в [21].

В практике же экспериментальных исследований очень часто сталкиваются с более сложными зависимостями. Найти готовый план для таких задач невозможно, и его приходится разрабатывать самому исследователю. В этом случае могут оказаться полезными следующие рассуждения.

Естественно, что теоретически строгий рототабельный план лучше, чем нерототабельный, насыщенный план экономичнее, чем не насыщенный, и т. д. Но если несколько худшее решение можно получить сегодня, а оптимальное — неизвестно через какой срок, то, видимо, целесообразнее исходить из того, что «лучшее — враг хорошего», и рассматривать разработанные разделы теории планирования эксперимента не как единственно возможные решения, а как примеры, достойные подражания для принятия эвристических решений в ситуации, оптимальное решение которой еще не известно.

В этой связи уместно заметить, что современная наука не отвергает с порога возможности принимать такие научно не обоснованные решения. Несколько десятилетий тому назад такие решения презрительно называли бы «среднепотолочными». Сейчас же для них введены вполне пристойные научные термины — *волевое решение* и *эвристическое решение* (см. § 7-1).

Так, например, ограничив в § 8-5 диапазон изменения x_2 на рис. 8-5, б значением $x_{2\min} = 0,5$, мы приняли чисто волевое решение, так как никаких доводов в пользу значения именно 0,5 против, например, 0,4, мы привести не могли. А решение о замене линейно равномерного размещения уровней варьирования (рис. 8-5, а) на логарифмически равномерное (рис. 8-5, б) — есть эвристическое решение. («А что, если попробовать разместить их в логарифмическом масштабе, ведь в этом масштабе резонансная кривая более симметрична?»)

При эвристическом составлении планов эксперимента для моделей выше второго порядка прежде всего следует обеспечить работоспособность плана. Для этого число уровней варьирования данного фактора должно быть не меньше, чем число членов многочлена, содержащих этот фактор в разных степенях, включая его нулевую степень при свободном члене $a_0 = a_0x^0$.

Так, для многочленов $y = a_0 + a_1x$, $y = a_0 + a_1x^0$, $y = a_0 + a_1/x$ или $y = a_0 + a_1\sqrt{x}$ необходимо для x иметь не менее двух уровней варьирования, для многочленов $y = a_0 + a_1x + a_2x^2$, $y = a_0 + a_1/x + a_2 \ln x$ — не менее трех уровней, для многочленов $y = a_0 + a_1x + a_2x^2 + a_3x^3$, $y = a_0 + a_1 \lg x + a_2\sqrt[3]{x} + a_3e^x$ — не менее четырех уровней и т. д.

Поэтому при построении эвристического плана (рис. 8-3, а) для поверхности второго порядка вида $y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_1x_2 + a_5x_2^2$ мы выбрали по три уровня для каждого из факторов x_1 и x_2 , так как в члены a_1x_1 и $a_4x_1x_2$ фактор x_1 входит не в разных, а в одной и той же степени.

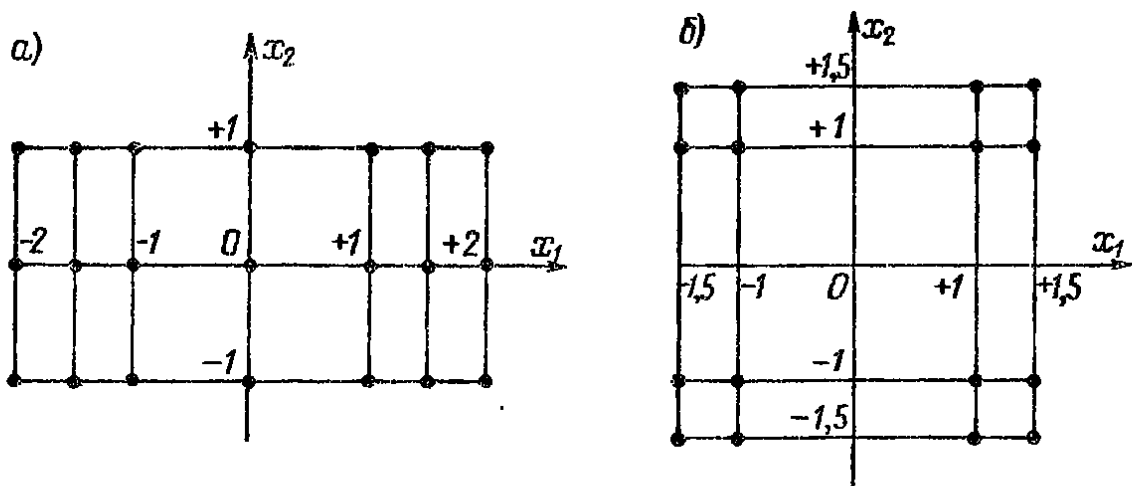


Рис. 8-6

Для размещения более трех уровней можно по ассоциации установить следующее приближенное правило. В теоретически обоснованных оптимальных планах (см. табл. 8-7) дополнительные уровни звездных точек отнесены на звездное плечо $\alpha = 1,4$ при $k = 2$, на $\alpha = 1,7$ при $k = 3$ и на $\alpha = 2$ при $k = 4$. Примем приближенное среднее значение $\alpha = 1,5$, и в эвристическом плане последующие уровни, кроме трех (или двух) центральных, будем разносить между собой на $\Delta\alpha = 0,5$.

Используя эти два положения, попробуем составить эвристический план для модели $y = f(x_1, x_2)$, в которой x_1 входит в многочлен (включая нулевую степень) в семи различных степенях, а x_2 — в трех. Вид такого плана с семью уровнями для x_1 и тремя — для x_2 показан на рис. 8-6, а, а на рис. 8-6, б представлен эвристический план для модели, в которую как x_1 , так и x_2 входят (включая нулевую степень) в четырех различных степенях. Такие планы будем называть *прямоугольными*.

Дальнейшее совершенствование эвристических прямоугольных планов может вестись в двух направлениях: в приближении их к рототабельным или в приближении к насыщенным. Рассмотрим оба эти направления.

Приближение эвристических планов к рототабельным. Сравнивая между собой прямоугольный нерототабельный план (рис. 8-3, а) и рототабельный план (рис. 8-3, б), можно заметить их следующие внешние признаки. Число опытов в обоих планах остается одним и тем же (9 и 9). Число уровней варьирования факторов по каждой из осей в рототабельном плане равно 5, а в прямоугольном — трем, т. е. к прямоугольному плану со всех сторон добавляется по одному уровню, на которые переносятся точки, нарушающие «округлость» плана. Внешним, чисто формальным итогом такого переноса является изменение размещения опытов по отдельным уровням плана: в прямоугольном плане оно было 3—3—3, а в рототабельном стало 1—2—3—2—1. Геометрически это можно представить себе и по-другому — как поворот исходного квадрата точек 1, 2, 3, 4 на 45° в положение 5, 6, 7, 8.

Отсюда может быть выдвинуто следующее эвристическое предположение: если внесение таких изменений превратило прямоугольный план второго порядка (рис. 8-3, а) в полностью рототабельный план (рис. 8-3, б), то внесение подобных же изменений в прямоугольные планы более высоких порядков позволит приблизить их в какой-то степени к рототабельным. Однако степень такой близости остается, естественно, неизвестной, поэтому полученные подобным путем планы следует именовать лишь «близкими к рототабельным».

Улучшенные таким способом планы, представленные на рис. 8-6, изображены на рис. 8-7. В плане на рис. 8-7, а на каждой из осей добавлено по два уровня для звездных точек и на них перенесены внешние угловые точки исходного плана рис. 8-6, а. В плане на рис. 8-7, б также добавлены дополнительные уровни для звездных точек и на них перенесены угловые точки внешнего квадрата. Кроме того, перенесены на оси и обозначены как звездные и четыре угловые точки внутреннего квадрата. Все перемещенные точки обозначены на новых планах как

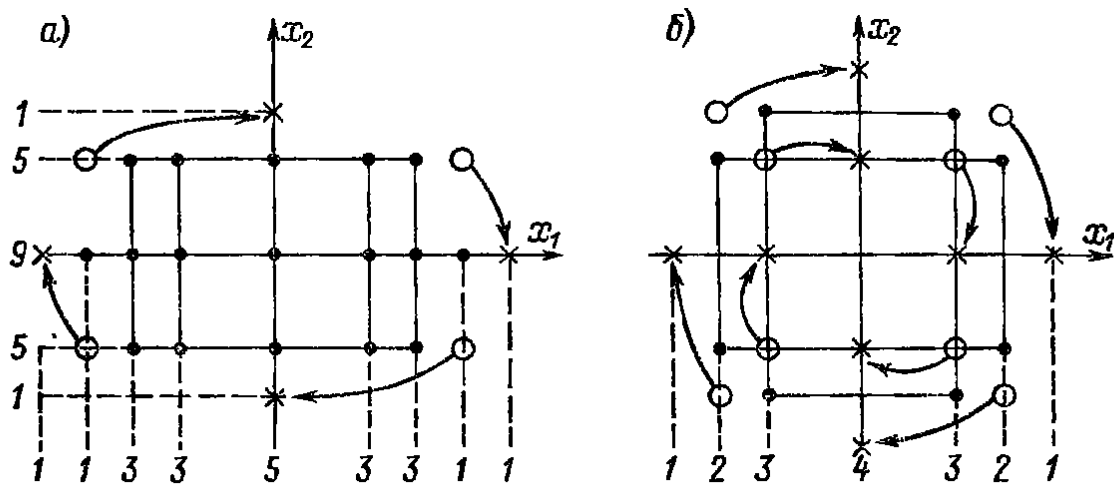


Рис. 8-7

звездные, а места их бывшего размещения (в прямоугольном плане) показаны кружками.

В результате этого распределение точек по уровням переменных x_1 и x_2 на рис. 8-7, б вместо 4—4—4—4 стало 1—2—3—4—3—2—1, а на рис. 8-7, а распределение точек по уровням фактора x_2 вместо 7—7—7 стало 1—5—9—5—1, а по уровням фактора x_1 вместо 3—3—3—3—3—3—3 стало 1—1—3—3—5—3—3—1—1, т. е. измененные таким образом планы существенно приблизились к рототабельным.

Однако приближение плана эксперимента к рототабельному — не самоцель, а средство повышения точности получаемого решения. Так, например, то, что в рототабельном плане на рис. 8-3, б почти половина точек (точки 5, 6, 7, 8) разнесена на расстояние, в $\sqrt{2} = 1,41$ раза большее, чем в плане на рис. 8-3, а, обеспечивает (см. рис. 8-4) в среднем по всему эксперименту приблизительно в $\sqrt[4]{2} = 1,19$ раза меньшую ширину полосы неопределенности получаемого решения при том же разбросе Δ исходных экспериментальных данных.

Но это повышение точности получаемого решения от перехода к рототабельному плану эксперимента далеко не всегда приводит к такому же повышению эффективности эксперимента. Дело в том, что при переходе к рототабельному плану число уровней варьирования факторов возросло с трех до пяти. Если одним из этих факторов является, например, температура и для достижения ее установившегося значения необходимо время, равное, например, двум часам, то общая затрата времени возрастет с четырех до восьми часов, так как кроме исходной температуры, которая не требует выдержки времени при ее установке, при первом плане нужно устанавливать два уровня температуры, а при втором — четыре.

Если бы такое же увеличение затрат времени на проведение измерений было использовано не для перехода на рототабельный план (рис. 8-3, б), а для удвоения выборки исходных данных при плане на рис. 8-3, а, то точность увеличилась бы в $\sqrt{2} = 1,41$ раза, в то время как переход к рототабельному плану дает повышение точности результата лишь в 1,19 раза, т. е. в подобных условиях рототабельный план с пятью уровнями варьирования оказывается существенно менее эффективным, чем прямоугольный план с тремя уровнями варьирования.

Если же в числе варьируемых факторов отсутствуют такие медленно устанавливающиеся факторы, как температура, вакуум, концентрация и т. п., а все факторы являются быстро устанавливающимися (например, механические силы, электрические напряжения и т. д.), то рототабельные планы будут, безусловно, более эффективными. К сожалению, такие вопросы существующая теория планирования не рассматривает и преимущество рототабельных планов преподносится читателю как абсолютная истина.

Приближение эвристических планов к насыщенным. В связи с изложенным часто возникают задачи приближения эвристических прямоугольных планов не к рототабельным, а к предельно экономным, насыщенным планам. Для этого

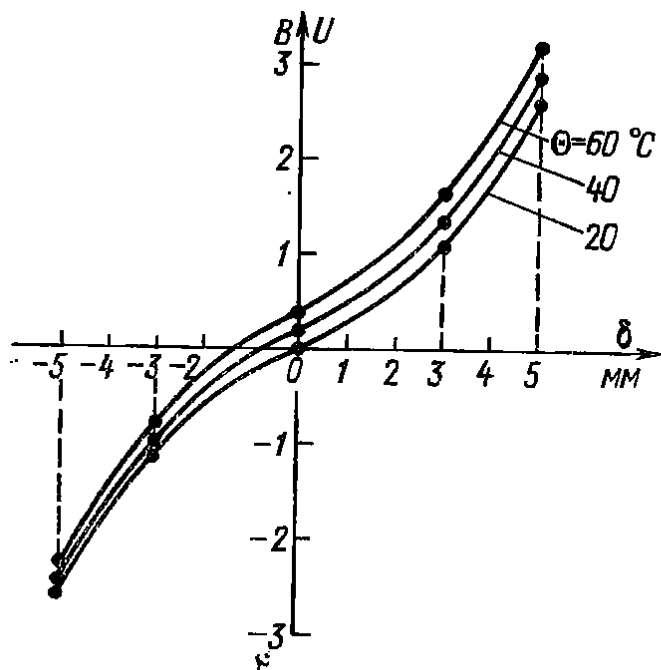


Рис. 8-8

выходной сигнал датчика, U , δ — измеряемое перемещение, мм, Θ — температура датчика, $^{\circ}\text{C}$.

При выборе вида модели, описывающей это семейство кривых, старались учесть все их особенности. Кривые $U = f(\delta)$ изогнуты на концах в разные стороны, поэтому в линейную модель $U = a_0 + a_1\delta$ был введен член $a_3\delta^3$, а так как эти искривления не всегда симметричны, то в качестве поправки был введен член $a_2\delta^2$. Учитывая, что пучок кривых при изменении температуры имеет расходящийся характер (ср. с рис. 7-2, б), для описания температурной погрешности была использована модель гиперboloида, т. е. кроме члена $a_4\Theta$ был дополнительно введен член $a_5\delta\Theta$, отражающий изменение чувствительности к температуре с изменением измеряемого перемещения δ . Таким образом, полная модель семейства характеристик датчика (рис. 8-8) получила вид

$$U = a_0 + a_1\delta + a_2\delta^2 + a_3\delta^3 + a_4\Theta + a_5\delta\Theta. \quad (8-2)$$

Теперь задача состоит в том, чтобы для этой модели спланировать эксперимент массового испытания серийно изготавливаемых датчиков, с тем чтобы были определены, внесены в паспорт, а при эксплуатации введены в память ИВК все индивидуальные для каждого экземпляра датчика коэффициенты $a_0, a_1, \dots, a_5$.

Приступая к составлению эвристического плана этого эксперимента, видим, что фактор δ входит в модель (8-2) в четырех различных степенях ($\delta^0, \delta, \delta^2, \delta^3$), т. е. требует варьирования, как минимум, на четырех уровнях, а фактор Θ входит в (8-2) в двух степенях (Θ^0 и Θ), т. е. требует варьирования не менее чем на двух уровнях.

Для приближения эвристического плана эксперимента к рототабельному число уровней варьирования должно быть увеличено, например, для фактора δ до пяти, а для фактора Θ — до трех. В качестве этих трех уровней варьирования для температуры Θ могут быть приняты уже использовавшиеся при предварительных испытаниях значения 20, 40 и 60 $^{\circ}\text{C}$. Выбор же уровней варьирования измеряемого перемещения δ требует дополнительного обсуждения.

Рассматривая кривые на рис. 8-8, замечаем, что в пределах изменения δ от -2 до $+2$ мм они остаются близкими к прямым линиям, поэтому для выявления нелинейности, т. е. уточнения коэффициентов a_2 и a_3 при δ^2 и δ^3 , размещение опытов на этом участке бесполезно. Кривизна характеристик по существу проявляется на участках изменения δ от -5 до -3 и от $+3$ до $+5$ мм, т. е. прежде всего должны быть уточнены координаты точек именно этих участков. Поэтому вопреки положениям теории планирования формуле (8-1) и данным табл. 8-7 мы назначаем звездное плечо α крайних уровней равным не 1,41 и даже не 1,5,

выбор оптимального размещения уровней варьирования факторов удобно производить путем нанесения точек на семейство кривых поверхности отклика, подобно тому, как это было показано на рис. 8-5.

Рассмотрим это на конкретном примере. Пусть задача состоит в получении градуировочных характеристик большой партии индуктивных датчиков, которые будут использоваться для ввода данных измерений в ИВК. При этом ставится задача коррекции с помощью процессора ИВК как нелинейной градуировочной характеристики датчика, так и его температурной погрешности.

Для этого путем предварительных испытаний нескольких датчиков был получен графический вид этих характеристик датчика (рис. 8-8) в виде семейства кривых $U = f(\delta, \Theta)$, где U —

а равным $5/3 = 1,67$, и уровни варьирования фактора δ назначаем равными $-5, -3, 0, +3, +5$ мм.

В результате этого получаем эвристический прямоугольный план эксперимента, представленный на рис. 8-9, а и состоящий в определении выходного напряжения U датчика при трех значениях температуры в каждом из пяти сечений при указанных выше значениях δ , т. е. состоящий из 15 опытов.

Такой план является существенно ненасыщенным (15 опытов для определения 6 коэффициентов), поэтому его приближение по формальным признакам к рототабельному с одновременным приближением и к насыщенному может быть проведено просто отбрасыванием избыточных опытов, например, так, как это показано на рис. 8-9, б, что обеспечивает распределение опытов по уровням фактора δ в виде $1-2-3-2-1$. Это дает одновременно с приближением к рототабельности сокращение общего числа опытов с 15 до 9.

Однако, имея перед собой наглядную картину размещения опытов по кривым поверхности отклика (рис. 8-8), рациональнее опираться не на формальные признаки рототабельного плана, а конкретно проанализировать, какие именно из 15 точек плана обеспечивают усреднение каждого из коэффициентов $a_0, \dots, a_5$ модели (8-2).

Так, коэффициент a_0 вычисляется как среднее по всем 15 опытам. Значение коэффициента a_1 , т. е. наклона характеристик, определяется только 12 удаленными от центра точками, а три точки при $\delta = 0$ не вносят вклада в его определение. Коэффициент a_3 при δ^3 также определяется расположением 12 удаленных от центра точек и не зависит от трех центральных. Коэффициент a_5 при гиперболическом члене δ^5 определяется прежде всего разностью расстояний между кривыми в их крайних участках и также не зависит от положения центральных точек. И лишь только коэффициент a_2 при δ^2 требует наличия центральной точки при $\delta = 0$, но при зеркальной симметрии положительной и отрицательной ветвей характеристики он просто равен нулю, а при наличии некоторой несимметрии играет роль малой поправки, учитывающей эту несимметрию, и поэтому может быть определен весьма приближенно.

Таким образом, три опыта при $\delta = 0$ участвуют лишь в уточнении коэффициентов a_0 и a_2 , а на точность определения остальных коэффициентов влияют весьма слабо. Поэтому рототабельность плана, т. е. равноточность определения всех коэффициентов модели, нарушается лишь присутствием избыточных опытов при $\delta = 0$. Однако отказ от всех трех опытов при $\delta = 0$ может привести к тому, что вследствие случайного разброса точек при δ , равном ± 3 и ± 5 мм, коэффициент a_2 может получить такое значение, что кривые вообще не будут проходить через фактические точки при $\delta = 0$. Поэтому хотя бы одна из них должна быть сохранена. Как видим, результат конкретного анализа существенно разошелся с формальными положениями теории планирования, в чем и состоит его ценность для практики.

Подобный конкретный анализ целесообразен и для выбора уровней варьирования второго фактора — температуры Θ . Для сокращения затрат времени на проведение эксперимента целесообразно выбрать нижний уровень варьирования температуры равным не заранее назначенному значению $\Theta = 20^\circ\text{C}$, а тому значению температуры, которое фактически будет иметь место в цехе в момент проведения испытаний. Оно должно быть зафиксировано в протоколе (или памяти

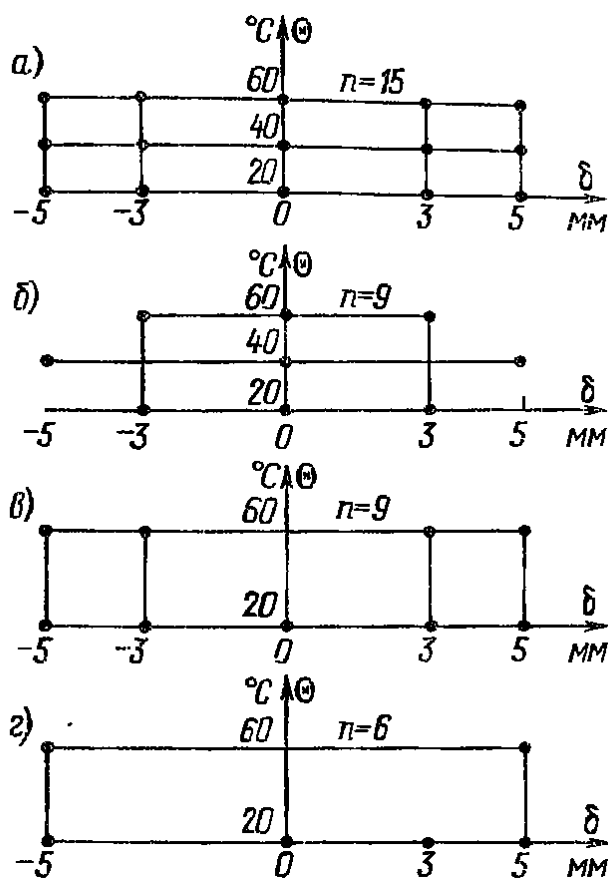


Рис. 8-9

ЭВМ) и использовано в дальнейшем при расчете значений коэффициентов. Тогда большая затрата времени потребуется на ожидание установившихся значений лишь двух других уровней температуры.

Но для модели гиперболоида в пределе достаточно не трех, а всего двух уровней факторов. При этом отказ от использования третьего, дополнительного уровня температуры, введенного лишь для приближения к рототабельности, дает почти двойное сокращение общих затрат времени на проведение испытаний. Никакие другие методы повышения эффективности эксперимента не дают подобной экономии.

Для повышения точности определения температурного коэффициента диапазон варьирования температуры целесообразно иметь наибольшим, т. е. оставить лишь уровень $\Theta = 60^\circ\text{C}$, а уровень $\Theta = 40^\circ\text{C}$ — исключить.

Полученный таким путем, хотя и не рототабельный, но зато оптимальный по эффективности и достаточно экономный план эксперимента показан на рис. 8-9, в. Он содержит, так же как и план на рис. 8-9, б, девять опытов и проводится в следующем порядке. Фиксируется значение температуры датчика, т. е. температуры помещения, где проводятся испытания, и проводятся измерения при δ , равной $+5$, $+3$, 0 , -3 и -5 мм. Затем датчик выдерживают в термостате для установления температуры $\Theta = 60^\circ\text{C}$ и проводятся измерения при δ , равной -5 , -3 , $+3$ и $+5$ мм.

План является ненасыщенным, так как для определения 6 коэффициентов используется 9 опытов, но наличие трех избыточных опытов обеспечивает возможность усреднения и оценки случайной составляющей погрешности датчика.

В целях предельной экономии времени можно перейти к насыщенному плану эксперимента из 6 опытов (рис. 8-9, г). Здесь при исходной температуре датчика проводятся 4 опыта при значениях δ , равных -5 , 0 , $+3$, $+5$ мм, однозначно, без усреднения определяющие коэффициенты a_0 , a_1 , a_2 и a_3 , а для определения коэффициентов a_4 и a_5 (также без усреднения) проводятся два дополнительных опыта при $\Theta = 60^\circ\text{C}$ и $\delta = \pm 5$ мм.

Однако при насыщенном плане, когда число искомых коэффициентов равно числу опытов, модель находится вообще без усреднения, т. е. с ошибками, равными случайным погрешностям исходных данных, а главное — размер этих погрешностей остается неизвестным, так как данные для их оценки отсутствуют. Поэтому переход от оптимального по эффективности плана с 9 опытами к насыщенному плану с 6 опытами допустим только при уверенности в пренебрежимо малом размере случайных погрешностей как используемой аппаратуры, так и исследуемого объекта (в нашем случае — датчика).

Рациональность же сокращения затрат времени на несколько минут для проведения трех дополнительных отсчетов при длительности всего эксперимента более двух часов также, естественно, весьма спорна.

Выявившееся в этом примере решающее влияние на эффективность эксперимента соотношения затрат времени на сами измерения (несколько минут на три отсчета) и на подготовку к измерениям (два часа на установление нужной температуры) будет подробнее рассмотрено в § 8-10.

Если же исследуемый индуктивный датчик предназначен не для измерения перемещений, а является, например, датчиком манометра или динамометра, то в рассмотренные планы должны быть внесены существенные коррективы. Дело в том, что в этом случае большая доля случайной погрешности датчика определяется механическим гистерезисом его упругого элемента. Для получения усредненной характеристики такого датчика порядок проведения опытов должен быть выбран с учетом гистерезисного цикла его упругого элемента.

При испытании таких датчиков обычно предусматривается 125%-ная перегрузка сверх номинального значения измеряемой величины. Поэтому уровни измеряемой величины (указанные условно в тех же единицах δ) будут устанавливаться в следующем порядке: положительная перегрузка (здесь, как и далее, без отсчета показаний) $+6,25$, затем $+5$, $+3$, 0 , -3 , -5 , $-6,25$ (отрицательная перегрузка), -5 , -3 , 0 , $+3$, $+5$. Аналогичным образом и при $\Theta = 60^\circ\text{C}$: $+6,25$, $+5$, $+3$, -3 , -5 , $-6,25$, -5 , -3 , $+3$, $+5$. Усреднение характеристики датчика вида (8-2) происходит по всем 18 отсчетам, а получаемая оценка с. к. о. погреш-

ности позволяет контролировать качество датчика в отношении его погрешности от гистерезиса.

Рассмотренные примеры позволяют показать, как могут использоваться основные положения теории планирования эксперимента для сознательной оценки значимости каждого опыта плана с целью повышения эффективности эксперимента в целом, хотя получаемые планы могут и существенно отличаться от формальных рекомендаций этой теории.

8-7. ОПТИМАЛЬНЫЙ ВЫБОР ТОЧНОСТИ СРЕДСТВ ИЗМЕРЕНИЙ В ЗАВИСИМОСТИ ОТ СООТНОШЕНИЯ ДИФфуЗИОННОСТИ ОБЪЕКТА И ПОГРЕШНОСТИ ИЗМЕРЕНИЙ

До сих пор, говоря об области разброса исходных экспериментальных данных, мы принимали во внимание лишь погрешности СИ и остаточную неадекватность принятой модели. Но кроме этих двух составляющих разброс данных вызывается еще и невоспроизводимостью от опыта к опыту, или *диффузностью*, самого исследуемого явления. Так, например, рост человека колеблется в такт с ударами сердца и частотой дыхания, длина металлического бруска — с микроколебаниями его температуры, максимумы давления в цилиндре двигателя внутреннего сгорания случайным образом существенно разнятся между собой и т. д.

Поэтому разброс исходных данных (обозначим его оценку $\Delta_{и.д.}$) всегда складывается из трех составляющих: Δ_0 — диффузности объекта измерений, Δ_m — погрешности адекватности модели и $\Delta_{с.и.}$ — погрешности средств измерений. Эти составляющие, как правило, можно считать некоррелированными,

когда $\Delta_{и.д.} = \sqrt{\Delta_0^2 + \Delta_m^2 + \Delta_{с.и.}^2}$ (более строгие методы суммирования и выражения погрешностей были рассмотрены в гл. 2, 3 и 5).

Для рационального выбора погрешности $\Delta_{с.и.}$ важно ее соотношение при том же допущении с суммарной погрешностью объекта и модели $\Delta_{0+m} = \sqrt{\Delta_0^2 + \Delta_m^2}$. Для упрощения последующих рассуждений предположим, что модель выбрана достаточно адекватной ($\Delta_m \ll \Delta_{с.и.}$ и $\Delta_m \ll \Delta_0$) и размером Δ_m можно пока пренебречь. При этом возможны три случая.

1. Обычно экспериментатор стремится использовать как можно более точную аппаратуру с $\Delta_{с.и.} \ll \Delta_0$. При этом результирующий разброс исходных данных будет $\Delta_{и.д.} = \sqrt{\Delta_0^2 + \Delta_{с.и.}^2} \approx \Delta_0$, т. е. будет определяться диффузностью объекта. Хорошо это или плохо? Безусловно, плохо. Чтобы усреднить этот разброс, необходимо провести большое число отсчетов, но излишне точная аппаратура требует, как правило, и больших затрат времени на каждое измерение. Если же в этих условиях уменьшать точность аппаратуры, то до тех пор, пока $\Delta_{с.и.} < \Delta_0/3$, погрешность измерений будет оставаться практически неизменной, а затраты времени будут существенно меньшими, а следовательно, эффективность эксперимента будет возрастать.

Таким образом, при $\Delta_{с.и.} \ll \Delta_0$ точность измерений не может быть заметно повышена использованием более точных СИ. Единственным путем повышения точности остается статистическая обработка многократных отсчетов, поэтому повышение эффективности эксперимента в этом случае может быть достигнуто путем снижения точности используемых СИ.

2. При $\Delta_{с.и.} \approx \Delta_0$ погрешность исходных данных составляет $\Delta_{и.д.} = \sqrt{\Delta_0^2 + \Delta_{с.и.}^2} \approx 1,4\Delta_0$, т. е. возрастает всего на 40% по сравнению с тем, когда $\Delta_{с.и.} \ll \Delta_0$. При проведении многократных отсчетов и их усреднении в $\sqrt{n}$ раз уменьшаются как влияние Δ_0 , так и влияние случайной составляющей $\Delta_{с.и.}$. В этом случае статистическая обработка весьма эффективна. Однако, стремясь к увеличению объема обрабатываемых выборок, нельзя забывать, что систематические погрешности при усреднении не уменьшаются. А систематическими являются как погрешность Δ_m адекватности модели, так и часть погрешности $\Delta_{с.и.}$.

3. При $\Delta_{с. и} \gg \Delta_0$ погрешность исходных данных полностью определяется погрешностью $\Delta_{с. и}$, так как при этом $\Delta_{и. д} = \sqrt{\Delta_{с. и}^2 + \Delta_0^2} \approx \Delta_{с. и}$. Если это значение $\Delta_{и. д}$ вполне устраивает экспериментатора, то нет нужды в организации многократных измерений и их статистическом усреднении. Если же возникает вопрос о необходимости снижения $\Delta_{и. д}$, то решение о целесообразности проведения многократных наблюдений с последующим усреднением или же замены СИ на более точные должен решаться путем специального исследования.

Сопоставляя между собой все три случая, следует заключить, что для обеспечения наибольшей эффективности эксперимента нет смысла уменьшать случайную погрешность аппаратуры больше, чем до $\Delta_{с. и} \leq \Delta_0/3$, а увеличивать объем выборки усредненных наблюдений имеет смысл только до тех пор, пока величина $\sqrt{(\Delta_{с. и}^2 + \Delta_0^2)/n}$ не будет сопоставима с погрешностью адекватности модели исследуемого явления или систематической составляющей погрешности СИ.

8-8. ПРИМЕР ЗАВИСИМОСТИ ДОСТИГАЕМОЙ ТОЧНОСТИ УСРЕДНЕННОГО РЕЗУЛЬТАТА ОТ ВРЕМЕНИ УСРЕДНЕНИЯ

Рассматривая вопрос о повышении точности путем статистического усреднения, обычно полагают, что точность с увеличением числа n усредняемых отсчетов возрастает как $\sqrt{n}$, ибо $\sigma_{\bar{x}} = \sigma_i/\sqrt{n}$. Это, однако, справедливо лишь при полном отсутствии систематических погрешностей и абсолютной независимости отсчетов между собой, т. е. полном отсутствии их взаимной корреляционной связи. Эти допущения используются настолько широко и часто, что нелишне рассмотреть на конкретном примере, к чему приводит присутствие хотя бы небольшой доли неисключенных систематических погрешностей.

Пусть для измерения суточных колебаний температуры воздуха вне помещения используется канал ИВК с датчиком в виде малоинерционного термометра сопротивления с временем установления выходного сигнала $t_i = 0,4$ с. Канал аттестован допусковым пределом приведенной погрешности $\gamma_{кл} = 1\%$, случайная погрешность (см. § 4-7) составляет $\gamma_i = 0,4\gamma_{кл} = 0,4\%$, а систематическая — $\theta = 0,3\gamma_{кл} = 0,3\%$. Период изменения измеряемой величины $T = 1$ сут = 86 400 с, а формы кривой изменения температуры близки к синусоидальной. Частота отсчетов выбрана в соответствии с временем установления сигнала датчика, т. е. $t_i = 0,4$ с. Число усредняемых отсчетов $n = t/t_i$ может быть любым в соответствии с общей затратой времени t на их получение.

В указанных условиях, кроме систематической погрешности канала, должны быть учтены динамические погрешности, которые также являются систематическими. Полагая, что термометр сопротивления является динамическим звеном первого порядка, его динамическая погрешность (см. § 1-5) отрицательна: $\gamma_d \approx -0,5\omega^2\tau^2$. Согласно действующим стандартам, время установления сигнала СИ определяется как время, по истечении которого приведенное значение динамической погрешности при ступенчатом входном воздействии равно 1%, независимо от класса точности. Отсюда в первом приближении приведенная динамическая погрешность датчика по истечении времени t , выраженная в процентах, $\gamma_d = (t_i/t)^2$. Для заданного t эта погрешность является систематической и должна суммироваться с систематической погрешностью θ измерительного канала.

Кроме того, необходимо учесть динамическую погрешность от усреднения мгновенных отсчетов синусоидального процесса за время t , сравнимое с периодом T этого процесса (рис. 8-10), состоящую в том, что вместо фактической кривой процесса $x(t) = X_m \sin \varphi$ (сплошная кривая на рис. 8-10) в результате усреднения за время t мы будем получать отсчеты, лежащие на штриховой кривой рис. 8-10 вида

$$\bar{x}(t) = X_m \sin \varphi \frac{\sin \Delta \varphi}{\Delta \varphi},$$

где $\Delta\varphi = \pi t/T$. Особенность этой погрешности состоит в том, что она возрастает, как показано в [32], с квадратом времени как

$$\gamma_{\text{уср}} \approx (t/T)^2 (\pi^2/6) \cdot 100\%.$$

Эта погрешность, как и динамическая погрешность датчика γ_d , всегда отрицательна, т. е. приводит к уменьшению амплитуды и поэтому должна суммироваться с γ_d арифметически. Однако систематическая погрешность θ измерительного канала ИВК (см. рис. 1-10, в) в одной части диапазона канала может быть положительной, а в другой — отрицательной. Поэтому их суммирование (§ 3-4) произведем по формуле

$$\theta_{\Sigma} = \sqrt{(\gamma_d + \gamma_{\text{уср}})^2 + \theta^2} = \sqrt{[(t_i/t)^2 + (10\pi)^2/6 (t/T)^2]^2 + \theta^2}.$$

Полагая, что при усреднении $n = t/t_i$ отсчетов случайная погрешность канала убывает как $\gamma_i/\sqrt{n} = \gamma_i \sqrt{t_i/t}$, для результирующей погрешности (в процентах) получим

$$\gamma_{\Sigma} = \sqrt{[(t_i/t)^2 + (10\pi t)^2/(6T^2)]^2 + \theta^2} + \gamma_i \sqrt{t_i/t}. \quad (8-3)$$

Характер кривой изменения достигаемой точности от затраты времени можно представить более наглядно, если воспользоваться численным выражением точности A как числа различных градаций измеряемой величины: $A = 1/(2\gamma_{\Sigma})$ или $A = 50/\gamma_{\Sigma}$, где γ_{Σ} — в процентах.

Рассмотрим в этих условиях четыре возможных режима измерений и проследим соответствующие им кривые изменения достигаемой точности A в функции от общей затраты времени на получение n усредняемых отсчетов.

Режим первый. Никаких дополнительных мер для повышения точности не принимается ($\theta = 0,3\%$, $\gamma_i = 0,4\%$), лишь только производится усреднение $n = t/t_i$ отсчетов. Подстановка этих данных в формулу (8-3) дает зависимость $A = f(t)$ (кривая 1 на рис. 8-11). Вначале точность растет как $A = ct^2$, но далее прирост точности резко замедляется. Скорость, соответствующая зависимости $A = c_1 \sqrt{t}$, наблюдается лишь в области $n = 2$, а далее происходит еще большее

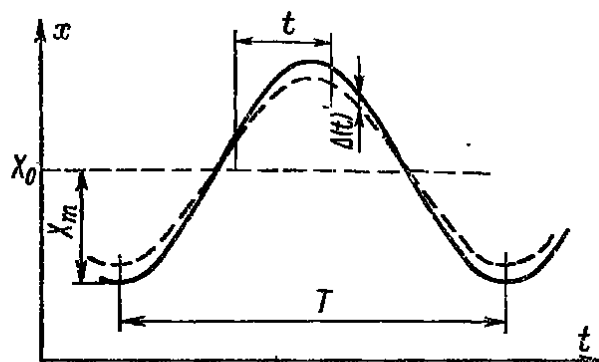


Рис. 8-10

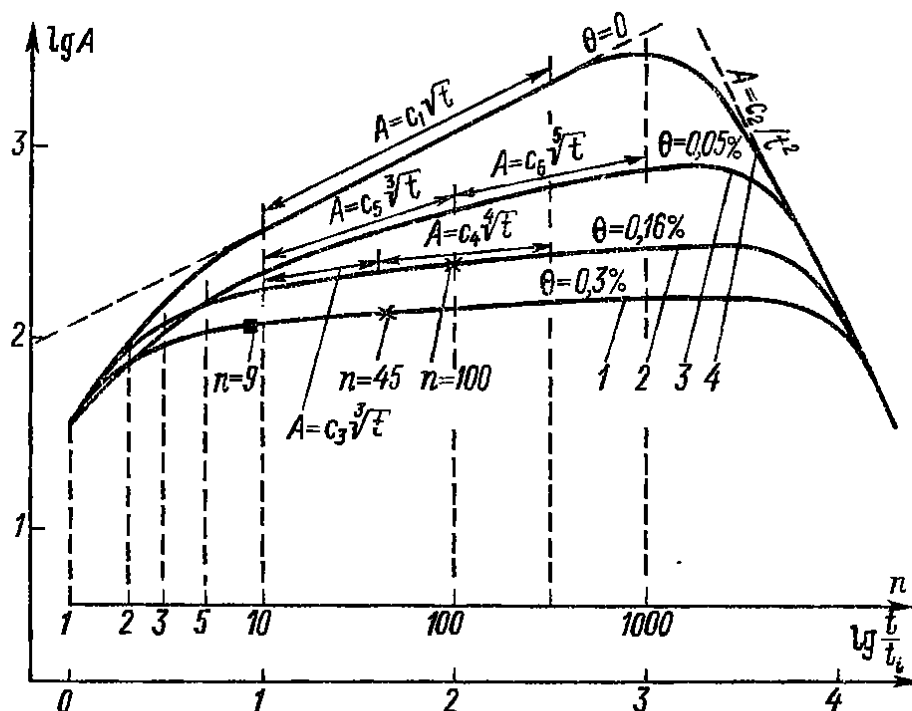


Рис. 8-11

замедление прироста точности. Максимум точности ($\gamma_{\Sigma} = \theta = 0,3\%$) достигается при усреднении $n \approx 2000$ отсчетов, а при дальнейшем увеличении числа усредняемых отсчетов вместо прироста начинается резкое падение точности, стремясь к $A = c_2/t^2$.

Это наиболее широко применяемый на практике режим использования измерительных каналов ИИС и ИВК. Спрашивается, в каких же пределах числа n усредняемых измерений правомерно в этом случае применение оценки в виде $\gamma_{\bar{x}} = \gamma_i/\sqrt{n}$? Приблизительно только в пределах от $n = 2$ до $n = 3$.

Режим второй. С целью повышения точности перед каждой серией усредняемых отсчетов производится проверка канала по образцовой мере и вводится постоянная поправка. Такой режим использования СИ был рассмотрен в § 4-7, где было показано, что таким путем систематическая погрешность может быть снижена в среднем до $0,77\sigma_i$ случайной погрешности. Полагая закон распределения случайной погрешности нормальным и что нормируемая погрешность была указана с доверительной вероятностью $P_d = 0,95$ ($t_{\text{норм}} = 1,95$), для оценки оставшейся систематической погрешности имеем: $\theta = 0,77 \sigma_i = 0,77 \gamma_i/1,95 = 0,77 \cdot 0,4/1,95 = 0,16\%$. Зависимость $A = f(t)$ для $\theta = 0,16\%$ и $\gamma_i = 0,4\%$, вычисленная по формуле (8-3), приведена на рис. 8-11 в виде кривой 2. Она проходит несколько выше кривой 1, но имеет аналогичную форму. Максимум точности ($\gamma_{\Sigma} = \theta = 0,16\%$) достигает также при $n \approx 2000$, а соотношение $\gamma_{\bar{x}} = \gamma_i/\sqrt{n}$ справедливо в пределах n от 3 до 5. В пределах n от 10 до 40 точность возрастает пропорционально $\sqrt[3]{t}$, а в пределах n от 40 до 300 — пропорционально $\sqrt[4]{t}$.

На кривых 1 и 2 крестиками отмечены точки, соответствующие усреднению $n = 45$ и $n = 100$ отсчетов согласно ГОСТ 8.207—76 (см. § 4-7), когда достигаются погрешности $\gamma_{\Sigma} = 0,36\%$ (вместо $\gamma_{\Sigma} = \theta = 0,3\%$) и $\gamma_{\Sigma} = 0,2\%$ (вместо $\gamma_{\Sigma} = \theta = 0,16\%$). Такая затрата времени с целью повышения точности усредненного результата вряд ли целесообразна. Так, например, при пятикратном сокращении затрат времени (точка $n = 9$ на кривой 1) погрешность составляет $\gamma_{\Sigma} = 0,4\%$, т. е. пятикратное увеличение затрат времени снижает погрешность результата с $0,4\%$ до $0,36\%$ (всего на 1/10 часть).

Режим третий. В § 4-7 указывалось, что наибольшее повышение точности может быть достигнуто только при одновременном введении и автоматической коррекции систематической погрешности перед каждым измерением, и статистического усреднения всех получаемых скорректированных отсчетов. Естественно, что автоматическая сквозная поверка канала (включая чувствительность датчика) с периодом повторения $t_i = 0,4$ с может быть осуществлена лишь при полной автоматизации проверок, введения поправок и статистической обработки скорректированных отсчетов. Подобные методы разработаны, например, в работе [8], когда в термометр сопротивления, установленный на объекте, подается короткий импульс тока, нагревающий лишь его чувствительный элемент на заранее известный скачок температуры. В этом случае (см. § 4-7) пределом повышения точности является только систематическая погрешность образцовой меры и метода поверки. Пусть в данном примере она составляет $\theta = 0,05\%$.

Такому режиму измерений на рис. 8-11 соответствует кривая 3, имеющая максимум точности ($\gamma_{\Sigma} = \theta = 0,05\%$) при усреднении $n = 1600$ отсчетов. Однако и эта кривая соответствует соотношению $\gamma_{\bar{x}} = \gamma_i/\sqrt{n}$ ($A = c_1 \sqrt{t}$) лишь на участке $n = 5$ и $n = 10$. На участке n от 10 до 100 она идет как $A = c_5 \sqrt[3]{t}$, а на участке n от 100 до 1000 — как $A = c_6 \sqrt[5]{t}$. При таком режиме введения поправок (см. § 1-6 и 4-7) одновременно с устранением систематической погрешности измерительного канала происходит уменьшение дисперсии его случайной погрешности. Поэтому в формулу (8-3) при вычислении кривой 3 должно подставляться $\theta = 0,05\%$ и $\gamma_i^2 = 2 \cdot 0,4^2 = 0,32$. Вследствие этого кривая 3 при $n < 4$ проходит на рис. 8-11 ниже кривой 2.

Режим четвертый. Допустим, что путем сверхтщательной установки нуля и чувствительности канала полосу его погрешностей по всей длине диапазона

удалось расположить так, как это было показано на рис. 1-10, а, т. е. сделать в точности $\theta = 0$ и $\gamma_i = 0,4\%$. В этом случае зависимость $A = f(n)$ изобразится на рис. 8-11 кривой 4, когда при усреднении $n \approx 1000$ отсчетов достигается максимум A , соответствующий $\gamma_{\Sigma} \approx 0,016\%$. Но даже в этом идеальном случае соотношение $A = c_1 \sqrt{n}$ справедливо лишь на участке от $n = 10$ до $n = 300$.

8-9. КРИВАЯ ФАКТИЧЕСКОЙ ЗАВИСИМОСТИ ДОСТИГАЕМОЙ ТОЧНОСТИ УСРЕДНЕННОГО РЕЗУЛЬТАТА ОТ ЧИСЛА УСРЕДНЯЕМЫХ ОТСЧЕТОВ С УЧЕТОМ НЕИСКЛЮЧАЕМОЙ ПРИ ИХ УСРЕДНЕНИИ СИСТЕМАТИЧЕСКОЙ ПОГРЕШНОСТИ

Рассмотренный выше пример усреднения сотен отсчетов с одновременной автоматической коррекцией систематической погрешности убедительно свидетельствует о несостоятельности утверждения, широко пропагандируемого во всех учебниках по статистической обработке, о том, что погрешность усредненного результата неограниченно убывает с ростом числа n усредняемых отсчетов: $\gamma_{\bar{x}} = \gamma_{xi}/\sqrt{n}$. На практике это соотношение остается справедливым при усреднении без одиовременной коррекции систематической погрешности лишь 2—3 отсчетов, а при использовании коррекции перед каждой серией усредняемых отсчетов — для n от 3 до 5 и при автоматической коррекции перед каждым измерением — для n от 5 до 10 отсчетов.

В чем же причина этого серьезного конфликта теории с практикой? Дело в том, что теоретическое утверждение, что $\sigma_{\bar{x}} = \sigma_{xi}/\sqrt{n}$, введено лишь в предположении, что усредняемые отсчеты независимы, т. е. абсолютно некоррелированы между собой. Но на практике получить такие отсчеты абсолютно невозможно.

Отсчеты получают с помощью реальных СИ, которые всегда имеют заметную долю систематической погрешности. И не малую долю, а того же порядка, что и их случайная погрешность. В § 1-7 на основе фактических данных было показано, что если у СИ случайная погрешность составляет $0,4\gamma_{кл}$, то систематическая в конце межремонтного интервала достигает $0,6\gamma_{кл}$. Систематическими являются и все динамические погрешности, подавляющая часть методических погрешностей, погрешность адекватности модели и т. д.

Правда, относительная доля систематической погрешности может существенно уменьшаться за счет случайных отклонений отсчетов вследствие большой диффузности самого объекта измерения (см. § 8-7), но тем не менее она всегда остается существенной и требует внимательного анализа.

Как же в этих реальных условиях происходит усреднение? Очень просто. Случайные, независимые между собой погрешности γ_i при усреднении по n отсчетам действительно уменьшаются в $\sqrt{n}$ раз, а систематическая погрешность θ остается неизменной после усреднения по любому числу отсчетов, так как присутствует в каждом из них. В итоге результирующая погрешность усредненного результата

$$\gamma_{\bar{x}} = \theta + \frac{\gamma_i}{\sqrt{n}} = \frac{\gamma_i}{\sqrt{n}} \left(1 + \frac{\theta}{\gamma_i} \sqrt{n} \right) = \frac{\gamma_i}{\sqrt{n}} (1 + a \sqrt{n}),$$

где $a = \theta/\gamma_i$ — отношение систематической и случайной погрешностей исходных данных. Именно это отношение и определяет характер кривых рис. 8-11.

Следовательно, точность A усредненного результата возрастает с числом усредняемых отсчетов n не как $A = \sqrt{n}/(2\gamma_i)$, а как

$$A = \frac{1}{2\gamma_i} \frac{\sqrt{n}}{1 + a \sqrt{n}}. \quad (8-4)$$

n_1	n_2	$a = 0,02$	0,05	0,1	0,2	0,5	1,0
3	5	2,08	2,20	2,39	2,79	3,97	5,93
10	20	2,15	2,38	2,75	3,51	5,75	9,50
50	100	2,34	2,84	3,68	5,36	10,4	18,8
500	1000	3,06	4,66	7,31	12,6	28,5	54,9

Это дробное выражение не дает наглядного представления об изменении с ростом n кривизны обсуждаемой кривой. Для наглядности его можно аппроксимировать функцией в виде $A = c \sqrt[n]{n}$, где β — показатель степени корня из n . Значение β равно 2 только при абсолютном отсутствии корреляционных связей между погрешностями отдельных отсчетов, т. е. при $\theta = 0$. При $\theta \neq 0$ показатель β , как было показано на рис. 8-11, изменяется с ростом n от 2 до ∞ . Для аппроксимации участка кривой от n_1 до n_2 при $\theta \neq 0$ значения показателя корня β могут быть вычислены по формуле

$$\beta = \frac{2}{1 + 2 \frac{\lg [(1 + a \sqrt[n_1]) / (1 + a \sqrt[n_2])]}{\lg (n_2/n_1)}} \quad (8-5)$$

Так как предположение о справедливости соотношения $\gamma_{\bar{x}} = \gamma_{xi} / \sqrt{n}$ используется очень широко и в самых различных случаях, то полезно знать, когда оно верно хотя бы приблизительно, а когда становится ошибочным. Для этого в табл. 8-10 приведены значения показателя степени β , вычисленные по формуле (8-5) для различной доли систематической погрешности $a = \theta/\gamma_i$ и разных участков кривой $A = f(n)$ от n_1 до n_2 .

Из данных табл. 8-10 видно, что соотношение $\gamma_{\bar{x}} = \gamma_{xi} / \sqrt{n}$ на практике чаще всего неправомерно. Для выборки с n от 50 до 100 отсчетов оно хотя бы приближенно ($\beta = 2,34$) справедливо лишь тогда, когда θ в 50 раз меньше, чем γ_i . Даже при $\theta = \gamma_i/20$ оно приближенно ($\beta = 2,38$) справедливо лишь при усреднении 10—20 отсчетов, а при $\theta = \gamma_i/10$ — для $n = 3 \div 5$ отсчетов. Но такие соотношения систематической и случайной погрешностей возможны только или при очень большой диффузности исследуемых объектов, когда γ_i много больше систематической погрешности СИ, или при использовании СИ с весьма совершенной автоматической коррекцией систематических погрешностей.

Бездумное же усреднение тысяч и десятков тысяч отсчетов, чем нередко увлекаются ныне горячие поклонники больших возможностей современной вычислительной техники, является очень опасным самообманом.

Некоторые итоги. Как видно из данных табл. 8-10 и кривых на рис. 8-11, присутствие даже очень небольшой доли систематических погрешностей, в том числе малой динамической погрешности от инерционности датчика или динамической погрешности от усреднения отсчетов, располагающихся по кривой, приводит к тому, что повышение точности согласно соотношению $\gamma_{\bar{x}} = \gamma_i / \sqrt{n}$ наблюдается лишь в ограниченном диапазоне числа n усредняемых отсчетов. В этой связи автор работы [53] по определению орбит спутников П. Е. Эльясберг два параграфа в своей более популярной книге [54] озаглавил так: «§ 2-6. Конфликт теории с практикой» и «§ 2-7. Как можно испортить кашу маслом». Он пишет: «Математическую теорию решения задач рассматриваемого типа принято называть теорией оценивания. Особенно это относится к оценкам точности и надеж-

ности получаемых значений отыскиваемых параметров и выбору оптимальной стратегии решения задач оценивания. По всем этим вопросам имеется достаточно полная литература. Однако эта литература в большинстве носит специальный характер и остается в значительной мере недоступной для широкого круга читателей, связанных с вопросами применения методов этой теории к решению различных прикладных задач. Это приводит, с одной стороны, к заметному отставанию практики решения задач оценивания от результатов теории, а с другой — к отрыву теории от практики.

«Наиболее сильно это обстоятельство проявляется... при оценке точности получаемых результатов... и может иметь место из-за неучета влияния систематических ошибок, неучета корреляционных зависимостей между ошибками отдельных измерений...».

По поводу последней из этих причин П. Е. Эльясберг пишет: «К сожалению, на практике часто ограничиваются исследованием только гистограммы без достаточно глубокого анализа... независимости отдельных измерений. Из изложенного следует практическая бессмысленность столь широко разрекламированного свойства состоятельности (статистических оценок). Действительно, оно является асимптотическим, т. е. должно проявляться при достаточно большом числе измерений. Однако при этом определяющим становится влияние всегда существующих малых неучитываемых корреляционных связей, приводящих к нарушению этого свойства. Таким образом, состоятельность статистических оценок можно с полным основанием считать одним из „мифов XX века“».

Мы привели эти цитаты, естественно, не для того, чтобы «опровергнуть» статистическую теорию погрешностей, а для того, чтобы отослать заинтересованного читателя к соответствующей литературе. Эти вопросы рассмотрены в трудах В. П. Перова (1959), Л. Т. Кузина (1962), И. Б. Челпанова, В. Я. Катковника, Р. А. Полуэктова (1963—1965), С. Я. Виленькина (1967), И. Б. Челпанова и Е. П. Гильбо (1975), П. Е. Эльясберга (1978—1983) и других авторов.

8-10. ОПТИМАЛЬНАЯ ЭФФЕКТИВНОСТЬ ЭКСПЕРИМЕНТА С УЧЕТОМ ЗАТРАТ ВРЕМЕНИ НА ПОДГОТОВКУ К ИЗМЕРЕНИЯМ

Располагая видом функциональных зависимостей достигаемой точности от затрат времени (рис. 8-11), можно ставить вопрос о сознательном планировании оптимальной или рациональной затраты времени для наиболее эффективной постановки процесса измерений. Для этого необходимо сформулировать формальный критерий эффективности.

Подобные критерии широко используются в расчетах экономической эффективности самых различных отраслей техники и представляют собой чаще всего отношение затрат к размеру достигаемого эффекта, т. е. цену единицы получаемой продукции. Так, эффективность работы автомобиля характеризуется потреблением горючего на 100 км пробега, эффективность оборудования электростанции — ценой 1 кВт·ч выработанной электроэнергии, а эффективность использования электрооборудования — частью используемой в среднем мощности, приходящейся на 1 кВт его номинальной, т. е. установленной мощности, и т. д.

Основным требованием, предъявляемым к таким оценкам, является требование их линейности, т. е. чтобы установленная таким образом «цена» единицы продукции оставалась постоянной как для 1-го, так и для 101-го километра пробега автомобиля, как для 1-го, так и для 1001-го киловатт-часа выработанной энергии и т. д.

Оптимальная эффективность эксперимента. Если не учитывать затрат времени на подготовку эксперимента, то эффективность эксперимента при линейной шкале цены определяется отношением $E = A/t$ (градаций/ч). При фактической зависимости $A = f(t)$ (рис. 8-12, а) данное значение эффективности E достигается дважды: один раз при малой затрате времени t_1 , когда $E = A_1/t_1$, и другой раз при большой затрате времени t_2 , когда $E = A_2/t_2$. Отсюда ясно, что максимально возможное значение эффективности, оцениваемое как отношение числа различных градаций A полученного результата к затратам времени t , достигается при

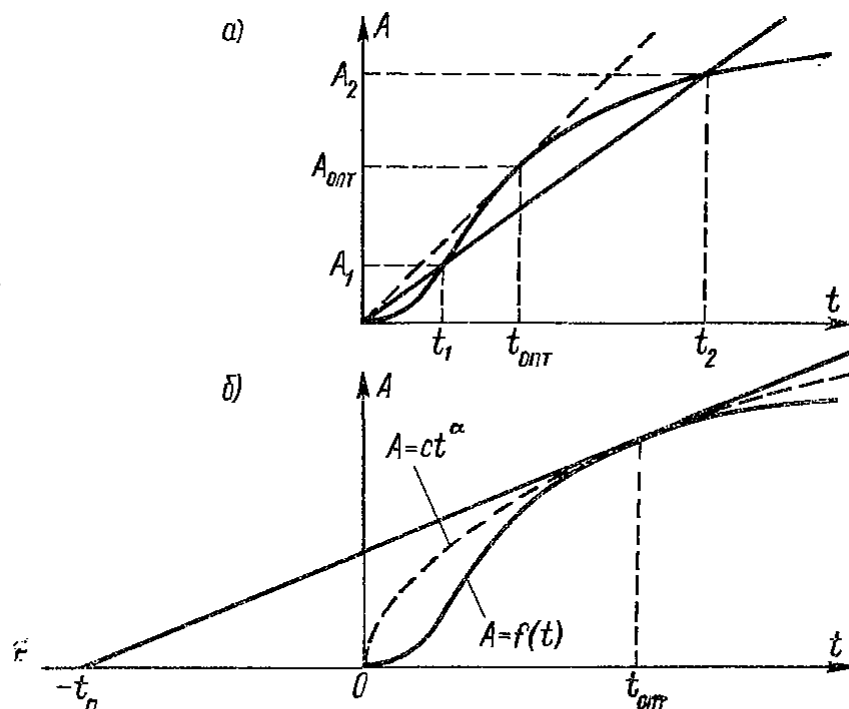


Рис. 8-12

$t = t_{\text{опт}}$, соответствующем точке касания штриховой прямой на рис. 8-12, а к кривой $A \approx f(t)$.

При $t < t_{\text{опт}}$ точность результата растет быстрее увеличения затрат времени и эффективность непрерывно повышается. При $t > t_{\text{опт}}$ точность результата продолжает возрастать, но растет медленнее увеличения затрат времени и эффективность падает.

Учет полных затрат времени на проведение эксперимента. В теории планирования эксперимента затраты учитываются лишь как число n проведенных опытов (полученных отсчетов). Если t_i — средние затраты времени на один опыт при данном сочетании факторов, то затраты на весь эксперимент возрастают как $t = nt_i$.

Но, кроме затрат, возрастающих пропорционально числу отсчетов, постановка и проведение эксперимента требуют еще и затрат времени, не зависящих от числа проводимых отсчетов. Это затраты времени на подготовку исследуемых образцов, подготовку и регулировку аппаратуры, ожидание окончания переходных процессов при смене уровней варьирования факторов и т. п. Практически не зависят от числа отсчетов затраты времени на последующую обработку на ЭВМ полученных результатов.

Все эти затраты времени необходимы и целесообразны, так как отказ от них влечет за собой повышение случайных, систематических или прогрессирующих погрешностей, т. е. ухудшает качество получаемых данных. Однако в сумме эти не зависящие от числа отсчетов затраты часто превышают затраты времени $t = nt_i$ на само проведение измерений и проведенный без их учета анализ часто (как было показано в § 8-6) полностью искажает действительное положение дел.

Сумму затрат времени на эксперимент в целом, не зависящих от числа отсчетов, условно назовем временем подготовки $t_{\text{п}}$. Результирующая эффективность эксперимента определяется, естественно, суммой обеих разновидностей затрат $t_{\Sigma} = t_{\text{п}} + t$ как $E = A/(t_{\text{п}} + t)$ градаций в час.

Графическое построение для определения времени $t_{\text{опт}}$, соответствующего максимуму эффективности эксперимента при затрате на подготовку эксперимента времени $t_{\text{п}}$, показано на рис. 8-12, б. Так как точка касания прямой в зависимости от времени подготовки $t_{\text{п}}$ может располагаться на любом участке кривой $A = f(t)$, то на рис. 8-12, б одновременно с фактической кривой $A = f(t)$ штриховой линией показана кривая $A = ct^{\alpha}$ (кривая вида $A = c\sqrt[t]{t}$), аппроксимирующая участок фактической кривой, содержащий точку касания. Приравнивая

n	$t(n)$ для $P_d = 0,95$	A при $\delta = 1\%$	n	$t(n)$ для $P_d = 0,95$	A при $\delta = 1\%$
1	—	0	21	2,09	112,8
2	12,71	5,5	31	2,04	139,4
3	4,30	20,1	61	2,00	193,7
4	3,18	31,4	121	1,98	277
5	2,78	40,1	∞	1,96	∞
11	2,23	74,6			

наклон касательной $A/(t_{\Pi} + t)$ производной $dA/dt = \alpha t^{\alpha-1}$, получаем $\alpha t^{\alpha}/(t_{\Pi} + t_{\text{опт}}) = \alpha t_{\text{опт}}^{\alpha-1}$ и $t_{\Pi} + t_{\text{опт}} = t_{\text{опт}}^{\alpha}/(\alpha t_{\text{опт}}^{\alpha-1}) = t_{\text{опт}}/\alpha$.

Отсюда $t_{\Pi} = t_{\text{опт}}(1/\alpha - 1)$ или $t_{\text{опт}} = t_{\Pi}/(1/\alpha - 1) = t_{\Pi}/(\beta - 1)$, где $\beta = 1/\alpha$ — показатель степени корня в выражении $A = c\sqrt[\beta]{t}$. Таким образом, при $\beta = 2$ $t_{\text{опт}} = t_{\Pi}$, при $\beta = 3$ $t_{\text{опт}} = t_{\Pi}/2$, при $\beta = 4$ $t_{\text{опт}} = t_{\Pi}/3$ и при $\beta = 5$ $t_{\text{опт}} = t_{\Pi}/4$.

Полученные соотношения для оптимального значения времени усреднения соответствуют лишь тому случаю, когда размер случайной погрешности γ_i известен заранее. Если же с. к. о. случайной составляющей погрешности заранее не известно и вычисляется по данным того же эксперимента, то доверительное значение погрешности усредненного результата $\gamma_{\bar{x}}$ должно вычисляться через квантиль $t(n)$ распределения Стьюдента, которая является табулированной функцией от числа n усредняемых отсчетов. Вследствие этого в формуле для определения погрешности $\gamma_{\bar{x}} = t(n) \delta_i / \sqrt{n}$, где δ_i — найденная по выборке оценка приведенного значения с. к. о. усредняемых отсчетов, оказываются зависящими от n как числитель $t(n)$, так и знаменатель $\sqrt{n}$. Это существенно изменяет форму кривой этой зависимости.

В этом случае зависимость точности A от $n = t/t_i$ получает вид $A = 50\sqrt{n}/[t(n) \delta_i]$. Значения этой функции для $P_d = 0,95$ и $\delta_i = 1\%$ приведены в табл. 8-11 и нанесены в виде кривой 1 на графике рис. 8-13, где одновременно штриховой кривой 2 показана зависимость $A = c_7 n^3$, совпадающая с начальным участком кривой 1, и кривой 3 — зависимость $A = c_8 \sqrt{n}$, совпадающая с кривой 1 при больших значениях n , а также прямая 4, проходящая через начало координат и касающаяся кривой 1 при $n = 5$.

Таким образом, в этом случае до $n \leq 3$ точность возрастает как $A \approx c_7 n^3$, в области $n \approx 5$ — растет приблизительно линейно с затратой времени, а при $n > 5$ — приближается к зависимости $A = c_8 \sqrt{n}$. Отсюда следует вывод: пространенное мнение, что усреднение 2, 3 или 5 отсчетов малоэффективно, является ошибочным. Наоборот, на этом участке происходит самое крутое возрастание точности, т. е. такая затрата времени на дополнительные отсчеты весьма эффективна.

Так как зависимость $A = f(t)$ (рис. 8-13) определяется табулированными значениями квантилей распределения Стьюдента, то аналитически взять от нее производную нельзя. Однако для $P_d = 0,95$ кривая рис. 8-13 или исходные данные табл. 8-11 с по-

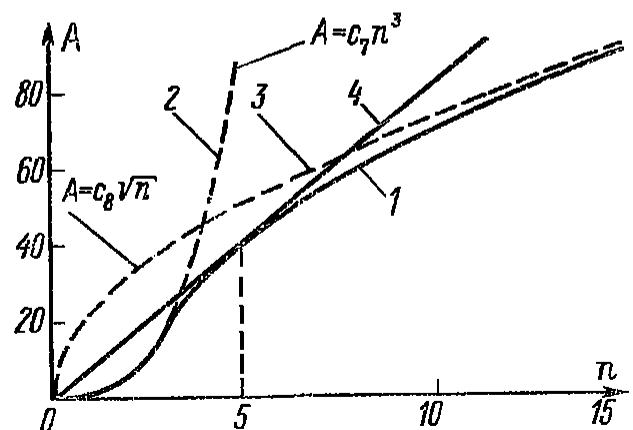


Рис. 8-13

грешностью примерно 1,5% аппроксимируются выражением

$$A = [0,8726 \ln(n-1) + 2,213]^3$$

или (для $n > 3$) с погрешностью примерно 0,5% — выражением

$$t(v) = \frac{1,96}{1 - 1,2/v + 0,354/v^3},$$

где $v = n - 1$ — число степеней свободы распределения Стьюдента. Тогда

$$A = \frac{\sqrt{n}}{2\delta_i t(n)} = \frac{\sqrt{v+1}}{2\delta_i \cdot 1,96} \left(1 - \frac{1,2}{v} + \frac{0,354}{v^3} \right).$$

Использование как первого, так и второго из этих аппроксимирующих выражений приводит к одному и тому же виду зависимости оптимальной затраты времени на проведение усредняемых отсчетов $t_{\text{опт}}$ от времени подготовки $t_{\text{п}}$ в виде $t_{\text{опт}} \approx t_{\text{п}} + 5t_i$ и оптимальному объему выборки усредняемых данных $n_{\text{опт}} = 5 + t_{\text{п}}/t_i$, т. е. при $P_d = 0,95$ и не известном заранее с. к. о. выборка, соответствующая $t_{\text{опт}}$, должна быть на 5 отсчетов больше, чем при заранее известном с. к. о. σ_i .

Практический расчет оптимального числа усредняемых отсчетов. Расчет $n_{\text{опт}}$ по формуле $n_{\text{опт}} = t_{\text{опт}}/t_i$, где $t_{\text{опт}} = t_{\text{п}}/(\beta - 1)$, а β рассчитывается по формуле (8-5), связан с необходимостью повторных итераций при назначении ряда пробных значений n . Однако, как показал И. А. Назаров, $t_{\text{опт}}$ в этой задаче определяется решением кубического уравнения вида $2\chi(1 + r\sqrt{\chi}) = 1 + \chi$, где $\chi = t_{\text{опт}}/t_{\text{п}}$;

$$r = \frac{\theta}{\gamma_i} \sqrt{\frac{t_{\text{п}}}{t_i}}. \quad (8-6)$$

Искомый корень χ этого кубического уравнения для r от 0,1 до 20 с погрешностью не более 1% может быть найден по аппроксимирующему соотношению

$$t_{\text{опт}} = \frac{t_{\text{п}}}{(2r)^{2/3} \left[1 + \left(\frac{1}{3r} \right)^{3/4} \right]}. \quad (8-7)$$

Таким образом, расчет оптимального числа усредняемых отсчетов $n_{\text{опт}}$ производится в следующем порядке. Определяются затраты времени $t_{\text{п}}$ на весь эксперимент в целом, не связанные с числом n усредняемых отсчетов (подготовка образцов и аппаратуры, ожидание окончания переходных процессов, время обработки данных на ЭВМ и т. п.). Оценивается время t_i , приходящееся в среднем на получение каждого из n усредняемых отсчетов. Оценивается возможное соотношение систематической θ и случайной γ_i погрешностей получаемых отсчетов: $a = \theta/\gamma_i$. По формуле (8-6) находится значение вспомогательной переменной r . По формуле (8-7) рассчитывается $t_{\text{опт}}$ и при известной заранее σ_i определяется $n_{\text{опт}} = t_{\text{опт}}/t_i$. Если же σ_i заранее не известно, а определяется из данных того же эксперимента, то $n_{\text{опт}} = t_{\text{опт}}/t_i + 5$.

Некоторые практические выводы. Приведенные соотношения, несмотря на предельно допустимые упрощения, положенные в основу их выводов, могут оказаться полезными при практическом планировании эксперимента. Тем более, что особой строгости при выборе оптимального объема n выборки исходных данных проявлять не следует, так как при использовании планов оптимального эксперимента речь может идти лишь о выборе числа повторения всех опытов этого плана. При этом следует иметь в виду, что при отклонении от $t_{\text{опт}}$ (см. рис. 8-12) в меньшую сторону эффективность падает резко, а при отклонении в большую сторону — снижается значительно медленнее. Поэтому предпочтение следует отдавать большему значению из ближайших к оптимальному.

Далее следует иметь в виду, что оптимальное по эффективности $t_{\text{опт}}$ (или $n_{\text{опт}}$) соответствует наиболее экономному плану проведения эксперимента, а достигаемая при этом точность усредненного результата (см. рис. 8-12) существенно меньше максимально возможной (см. рис. 8-11). Поэтому при необходи-

мости достижения большей точности, чем $A_{\text{опт}}$, можно сознательно идти на большие затраты времени, жертвуя эффективностью эксперимента.

Однако при этом следует помнить, что даже кривая 4 на рис. 8-11, соответствующая и д е а л ь н о отлаженному каналу измерения, не имеющему систематических погрешностей, не позволяет безгравично повышать точность путем простого усреднения. При неавтоматизированных измерениях получить для усреднения тысячи отсчетов было слишком трудно, поэтому подобный анализ не представлял практического интереса. При применении ИВК нередко используют усреднение 5—10 тысяч отсчетов. Поэтому анализ подобных задач приобретает сугубо практическое значение.

ЗАКЛЮЧЕНИЕ

Итак, авторы попытались в этой книге ознакомить читателя со всеми основными вопросами, возникающими при оценке погрешностей результатов измерений. Ответы на часть из них удалось дать в виде анализа имеющихся литературных источников. Но при этом оказалось, что многие практически необходимые задачи не имеют простых и удобных решений. В этих случаях были описаны методы, используемые авторами в своей практической деятельности. Однако некоторые из вопросов так и остались без ответа. Во всех этих случаях авторы считали своим долгом обратить внимание читателя на существование таких «белых пятен» в теории и практике оценки погрешностей. Таким образом, для развития этого направления имеется широкое поле деятельности.

Это и методические разработки, например: устранение разночтений термина «методическая погрешность» (см. § 1-1), введение в техническую документацию не только одночленной и двучленной формул нормируемой погрешности, но и трехчленной формулы для приборов с широким диапазоном измерений (см. § 1-2), разработка и нормативное закрепление прогрессивных правил округления погрешностей с переменным шагом в трех поддиапазонах каждой декады (0,1—0,2; 0,2—0,5; 0,5—1,0) (см. § 1-4), закрепление в методической документации удобной для практики системы моделей законов распределения погрешностей как для описания погрешностей СИ, так и погрешностей результатов прямых и, особенно, косвенных измерений (см. § 3-7, 3-8).

Очень широко среди практиков распространено мнение, что все затруднения с вероятностной оценкой погрешности объясняются лишь их слабой подготовкой в области математической статистики и теории вероятностей. Все необходимые для этого задачи, дескать, давно решены в теории вероятностей и теории случайных процессов. Стоит лишь как следует овладеть премудростью этих наук и все сложности разрешатся сами собой. Но это верно лишь отчасти. Выше мы несколько раз вынуждены были обращать на это внимание читателя. Очень многое применительно к нуждам оценки погрешностей еще ждет своей разработки.

Так, например, нельзя же ожидать, что для всего разнообразия законов распределения погрешностей математики дадут таб-

лицы квантилей. Такие таблицы заняли бы целый том. Нужно какое-то другое решение, например, в виде приближенных формул, а такие формулы нужно разработать. Приведенные в § 2-6 формулы, видимо, только начало такого пути. Подобное положение наблюдается и с методикой суммирования погрешностей. Строгое математическое решение в виде многомерного распределения для практики бесполезно. То же самое относится и к имитационному моделированию по методу Монте-Карло, так как оно не может дать общего решения, а численные решения всякий раз должны проводиться заново. Нужны упрощенные, практические методы, подобные описанным в § 3-2—3-5. Это особенно относится к расчету погрешности косвенных измерений (§ 3-6, 3-7), где авторы настоящей книги из-за математической сложности вынуждены были ограничиться самыми примитивными методами.

Не лучшее положение и со сравнительной эффективностью различных оценок центра (§ 4-3), рассеянием оценок контрэксцесса, энтропийного коэффициента и энтропийного значения (§ 4-4), исключением промахов при распределениях, отличных от нормального (§ 4-5). Полученные моделированием кривые рис. 4-2, 4-4, 4-7, естественно, не являются теоретически всесторонне обоснованным решением. Но ничего другого, нужного для практики, пока нет. Даже такой, казалось бы, классический вопрос математической статистики, как оптимальное число интервалов группирования экспериментальных данных для построения полигона или гистограммы, оказывается, имеет почти столько же «оптимальных» решений, сколько излагающих его авторов (§ 5-2). Всюду рекомендуемое использование критериев согласия для идентификации формы распределения практически не позволяет произвести желаемой идентификации при тех данных, которыми исследователь фактически располагает (§ 5-5).

Подобный перечень как теоретических, так и практических задач можно было бы дать по обработке однофакторных и многофакторных экспериментов. Здесь также большое количество нужных для практики задач в области разработки удобных методов описания параметров многомерного мениска погрешностей при многофакторном эксперименте и в использовании так называемых «робастных», т. е. не зависящих от вида закона распределения, устойчивых методов оценки параметров модели и исключения промахов, которые позволяют устранить неустойчивость при получении решений МНК для многомерных задач.

Многие практические методы для устойчивой оценки параметров распределений с «утяжеленными хвостами», как принято выражаться в теории «робастного» оценивания, были рассмотрены в настоящей книге: это аналитические модели распределений, отличных от нормального (с большим и бесконечным значениями эксцесса), примеры и классификация таких распределений, формулы для расчета квантилей распределений, отличных от нормального, сравнительная эффективность различных оце-

нок центра, единый критерий для цензурирования промахов для ряда классов распределений, метод медианных центров для предварительного выбора модели и т. д. Тем не менее дальнейшая разработка устойчивых, не зависящих от вида распределения методов, представляет собой одно из наиболее перспективных направлений развития методов обработки данных. Например, основе методов, изложенных в § 2-5, 2-6, 3-5, 3-6, 4-6, 5-8, 6-6, 7-3, 7-4, уже сейчас могут быть созданы удобные программы для обработки данных исследования на ЭВМ.

Особого внимания заслуживает анализ путей повышения эффективности измерительного эксперимента, Это прежде всего разработка шкалы затрат на подготовку, постановку и проведение эксперимента и шкалы достигаемого эффекта с учетом как параметров мениска погрешностей, так и протяженности варьирования факторов. Естественно, что оценка результата сложного многофакторного эксперимента одним числом в виде γ_{Σ} (как это сделано на рис. 8-11) — крайне примитивна. Здесь нужен системный, комплексный подход, своеобразная квалиметрия процесса измерения, в какой-то степени аналогичная квалиметрии СИ.

Одним словом, нерешенных вопросов в области оценки погрешностей результатов измерений вполне достаточно. Эти трудные и неблагодарные задачи еще ожидают энтузиастов для их разрешения.

ПРИЛОЖЕНИЕ

1. Оптимальное число столбцов гистограммы (с. 179):

$$m = \frac{\varepsilon + 1,5}{6} n^{0,4} \quad \text{или} \quad m = \frac{1}{3} \sqrt[5]{\varepsilon^4 n^2};$$

$$m_{\min} = 0,55n^{0,4}; \quad m_{\max} = 1,25n^{0,4}.$$

2. Объем выборки n для гистограммы из m столбцов (с. 197)

$$n \geq [6m/(\varepsilon + 1,5)]^{2,5}.$$

3. Экспесс распределения суммы двух случайных величин (с. 97)

$$\varepsilon_{\Sigma} = \varepsilon(x) p^2 + 6p(1-p) + \varepsilon(y)(1-p)^2.$$

4. Оценка энтропийного коэффициента по гистограмме (с. 177)

$$k_{\varepsilon} = \frac{dn}{2\sigma} 10^{-\frac{1}{n} \sum_1^m n_i \lg n_i}$$

5. Граница цензурирования промахов (с. 158)

$$t_{\text{гр}} = 1,55 + 0,8 \sqrt{\varepsilon - 1} \lg(n/10).$$

6. Таблица квантилей нормального распределения — с. 142.

7. Таблица квантилей распределения Стьюдента — с. 144.

8. Таблица квантилей χ^2 -распределения — с. 194.

9. Аппроксимация квантилей класса экспоненциальных и трапецеидальных распределений и распределений Стьюдента с $n > 8$ (с. 85)

$$t = 1,62 [3,8 (\varepsilon - 1,6)^{2/3}]^{\lg \lg [1/(1-P_d)]}.$$

СПИСОК ЛИТЕРАТУРЫ

1. Агебян Т. А. Основы теории ошибок для астрономов и физиков. — М.: Наука, 1972. — 172 с.
2. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1976. — 280 с.
3. Алексеева И. У. Теоретическое и экспериментальное исследование законов распределения погрешностей, их классификация и методы оценки их параметров: Автореф. дис. на соиск. учен. степени канд. техн. наук. — Л., 1975. — 20 с.
4. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов/Пер. с англ. — М.: Мир, 1971. — 408 с.
5. Бромберг Э. М., Куликовский К. Л. Тестовые методы повышения точности измерений. — М.: Энергия, 1978. — 176 с.
6. Венецкий И. Г., Венецкая В. И. Основные математико-статистические понятия и формулы. — М.: Статистика, 1979. — 447 с.
7. Галочкина В. Я. Исследование энтропийных оценок случайных погрешностей измерительных устройств: Автореф. дис. на соиск. учен. степени канд. техн. наук. — Л., 1971. — 18 с.
8. Гуткин Б. С. Исследование и разработка методов определения и повышения точности результатов экспериментов при автоматизации измерений: Автореф. дис. на соиск. учен. степени канд. техн. наук. — Л., 1981. — 18 с.
9. Зажигаев Л. С., Кишьян А. А., Романиков Ю. И. Методы планирования и обработки результатов физического эксперимента. — М.: Атомиздат, 1978. — 232 с.
10. Закс Л. Статистическое оценивание. — М.: Статистика, 1976. — 598 с.
11. Земельман М. А., Кньюфер А. П., Кузнецов В. П. О методах нормирования метрологических характеристик измерительных устройств//Измерительная техника. — 1969. — № 1. — С. 92—95; № 2. — С. 90—93; № 3. — С. 88—91.
12. Иванов В. Н. К вопросу о соотношении различных критериев точности средств измерений//Автометрия. — 1974. — № 5. — С. 89—91.
13. Кавалеров Г. И., Мандельштам С. М. Введение в информационную теорию измерений. — М.: Энергия, 1974. — 376 с.
14. Кендалл М., Стьюарт А. Теория распределений. — М.: Наука, 1966. — 588 с.
15. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973. — 542 с.
16. Колмогоров А. Н. Три подхода к определению понятия количества информации//Проблемы передачи информации. — Т. 1. — Вып. 1. — М.: Изд-во АН СССР, 1965. — С. 3—12.
17. Крамер Г. Математические методы статистики. — М.: Мир, 1976. — 648 с.
18. Крейнович В. Я., Кириллова И. С. Аппроксимация экспериментальных данных при различной степени засоренности зависимой и независимой переменных//Труды ВНИИЭП: Проектирование средств электроизмерительной техники. — 1985. — С. 101—107.

19. Лешинская Е. С., Новицкий П. В. Электрические измерения физических величин: Измерительные преобразователи: Учебное пособие для вузов. — Л.: Энергоатомиздат, 1983. — 320 с.

20. Линник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений. — М.: Физматгиз, 1958. — 334 с.

21. Львовский Е. Н. Статистические методы построения эмпирических формул: Учебное пособие для вузов. — М.: Высшая школа, 1982. — 224 с.

22. Маликов М. Ф. Основы метрологии. — М.: Стандартгиз, 1949. — 480 с.

23. Методика нормирования, оценки и контроля метрологической надежности электроизмерительных приборов//Метрология. — 1977. — № 2. — С. 9—25.

24. Методика установления вида математической модели распределения погрешности МИ 199-79. — М.: Изд-во стандартов, 1981. — 34 с.

25. Митропольский А. К. Техника статистических вычислений. — М.: Наука, 1971. — 576 с.

26. Мудров В. И., Кушко В. Л. Методы обработки измерений. — М.: Сов. радио, 1976. — 143 с.

27. Назаров И. А. К вопросу о предельных значениях энтропийного коэффициента//Известия ЛЭТИ. — Вып. 66. — Ч. 1. — 1965. — С. 21—23.

28. Николаев В. И. Контроль работы судовых энергетических установок. Ч. 1. Элементы теории. — Л.: Судостроение, 1965. — 238 с.

29. Новицкий П. В., Зограф И. А., Лабунец В. С. Динамика погрешности средств измерений. — Л.: Энергоатомиздат, 1990. — 192 с.

30. Новицкий П. В. Понятие энтропийного значения погрешности//Измерительная техника. — 1966. — № 7. — С. 11—14.

31. Новицкий П. В. Об особых свойствах 95%-ной квантили большого класса распределений и предпочтительных значениях доверительной вероятности при указании погрешностей приборов и измерений//Метрология. — 1979. — № 2. — С. 18—24.

32. Новицкий П. В. Основы информационной теории измерительных устройств. — Л.: Энергия, 1968. — 248 с.

33. Петров В. П., Рисный Ю. В. Оценка суммарной погрешности средств измерений//Измерительная техника. — 1977. — № 2. — С. 19—20.

34. Персии С. М. Количество информации при цифровом измерении//Измерительная техника. — 1964. — № 7. — С. 5—9.

35. Плюхинский А. Д. Алгоритмы биометрии. — М.: Изд-во МГУ, 1967. — 82 с.

36. Рабинович В. И., Цапенко М. П. О количестве измерительной информации//Измерительная техника. — 1963. — № 4. — С. 7—11; № 6. — С. 1—6, № 10. — С. 8—13.

37. Рабинович С. Г. Погрешности измерений. — Л.: Энергия, 1978. — 262 с.

38. Сборник научных программ на фортране. Вып. 1. Статистика. Вып. 2. Матричная алгебра и линейная алгебра/Пер. с англ.; Под ред. С. Я. Виленикина. — М.: Статистика, 1974. — 316 с.

39. Смирнов Н. В., Дуин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1965. — 556 с.

40. Сравнение оценок погрешности измерения по энтропийному среднеквадратическому и предельному значениям/П. В. Новицкий, И. А. Назаров, В. Я. Иванова, Г. А. Кондрашкова//Измерительная техника. — 1966. — № 9. — С. 20—24.

41. Таушанов З., Тонева Е., Пенова Р. Вычисление энтропийного коэффициента при малых выборках//Изобретательство, стандартизация и качество. — 1973. — № 5. — София.

42. Тонева Е. Аппроксимация распределений погрешности средств измерений//Измерительная техника. — 1978. — № 2. — С. 14—16; 1979. — № 7. — С. 17—18; 1981. — № 6. — С. 15—16.

43. Тьюки Дж. Анализ результатов наблюдений. — М.: Мир, 1981. — 693 с.

44. Устойчивые статистические методы оценки данных/Под ред. Р. Л. Лопера, Г. Н. Уилкинсона/Пер. с англ. — М.: Машиностроение, 1984. — 232 с.

45. Хальд А. Математическая статистика с техническими приложениями. — М.: Изд-во иностр. лит., 1956. — 642 с.

46. Хартман К., Лецкий Э., Шефер В. Планирование эксперимента в исследовании технологических процессов/Пер. с нем. — М.: Мир, 1977. — 552 с.
47. Хьюбер П. Робастность в статистике/Пер. с англ.; Под ред. И. Г. Журбенко. — М.: Мир, 1984. — 304 с.
48. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. — М.: Мир, 1968. — 280 с.
49. Цветков Э. И. Методические погрешности статистических измерений. — Л.: Энергоатомиздат, 1984. — 144 с.
50. Шенион К. Работы по теории информации и кибернетике. — М.: Изд-во иностр. лит., 1963. — 832 с.
51. Шторм Регина. Теория вероятностей: Математическая статистика. Статистический контроль качества. — М.: Мир, 1970. — 368 с.
52. Электрические измерения неэлектрических величин/Под ред. П. В. Новицкого. — 5-е изд. — Л.: Энергия, 1975. — 576 с.
53. Эльясберг П. Е. Об устойчивости оценок точности определения орбит по результатам измерений//Космические исследования. — 1978. — Т. 16. — Вып. 5. — С. 658—667.
54. Эльясберг П. Е. Измерительная информация. Сколько ее нужно, как ее обрабатывать? — М.: Наука, 1983. — 208 с.
55. Cochran W. G. The χ^2 -test of goodness of fit//Ann. Math. Statist. — Vol. 23. — 1952. — P. 315—316.
56. Hahn G. I., Shapiro S. Statistical models in engineering. — Presearch and development//Center General Electric Company. — New York; London; Sydney: John Wiley and Sons, 1967. — 396 p.
57. Heinhold I., Gaede K. W. Ingenieur statistic. — München; Wien, Springer Verlag, 1964. — 352 s.
58. Mann H. B., Wald A. On the choice of the number of intervals in the application of the chi-square test//Ann. Math. Statist. — Vol. 13. — 1942. — P. 478—479.
59. Martin-Löf. The definition of random sequences//Information and Control. — Vol. 9. — 1966. — P. 602—619.
60. Mills F. Statistical methods. — New York: Columbia University, 1955. — 304 p.
61. Sturges H. A. The choice of classic intervals//J. American Statist. — Ass. — march 1926. — 47 p.
62. Wiener N. Cybernetics or control and communication in the animal and the machine. — New York; London: Academic Press, 1961. — 320 p.
63. Williams C. A. On the choice of the number and midth of classes for chi-square test of goodness of fit//J. American Statist. — Ass. — Vol. 45. — 1950. — P. 77—78.

П.В. НОВИЦКИЙ
И.А. ЗОГРАФ

ОЦЕНКА ПОГРЕШНОСТЕЙ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ



ЭНЕРГОАТОМИЗДАТ