

**I.S.Gutman, I.J.Balaban**

**GEOSTATISTICS FOR GEOLOGICAL  
EXPLORATIONS**



**MOSCOW**  
Издательский центр  
РГУ нефти и газа имени И.М. Губкина  
2011

**И.С. Гутман, И.Ю. Балабан**

## **ГЕОСТАТИСТИКА В ПРОМЫСЛОВО- ГЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**

*«Допущено Учебно-методическим объединением вузов  
Российской Федерации по нефтегазовому образованию  
в качестве учебника для студентов высших учебных заведений,  
обучающихся по направлению подготовки магистров 131000  
«Нефтегазовое дело», по представлению Ученого совета  
Российского государственного университета нефти и газа  
имени И.М. Губкина».*



МОСКВА  
Издательский центр  
РГУ нефти и газа имени И.М. Губкина  
2011

Рецензенты:

Б.Е. Лухминский - профессор кафедры геофизики РГГРУ,  
доктор физ.-мат. наук,  
Билибин С.И. - заместитель генерального директора ОАО «ЦГЭ»,  
доктор технических наук, доцент.

**Гутман И.С., Балабан И.Ю.**

Геостатистика в промыслово-геологических исследованиях:  
Учебное пособие. – М.: Издательский центр РГУ нефти и газа  
имени И.М. Губкина, 2011. – 154 с.

ISBN 978-5-91961-053-3

В учебнике описаны математические методы решения промыслово-геологических задач при разработке нефти и газа, прежде всего оценки неоднородности продуктивного пласта и моделирования пространственного распространения неоднородностей.

Цель учебника – ознакомить читателя с геостатистическими основами методов обработки исходных данных по залежам УВ, вооружить их знаниями, необходимыми при геолого-промысловом обобщении результатов проводимых исследований и при разведке и разработке нефтяных и газовых месторождений; объяснить детали математической обработки данных и трехмерного моделирования залежей УВ при подсчете запасов и геологическом обосновании разработки.

Книга рассчитана на магистров, обучающихся по нефтегазовому делу и может использоваться студентами и аспирантами геологических специальностей, а также широким кругом специалистов, интересующихся математическими основами обработки геолого-промысловых данных и методик геологического моделирования.

Данное издание является собственностью РГУ нефти и газа имени И.М. Губкина и его репродуцирование (воспроизведение) любыми способами без согласия университета запрещается.

**ISBN 978-5-91961-053-3**

© И.С. Гутман, И.Ю. Балабан, 2011  
© РГУ нефти и газа имени И.М. Губкина,  
2011  
© Оформление серии Голубева М.С., 2011

# Введение

Нефтегазопромысловая геология является наукой о методах изучения подземного резервуара, подсчета запасов и о геологических факторах, определяющих рациональную разработку залежей нефти и газа.

Необходимость применения **геостатистики**, понимаемой как совокупность математических и статистических методов обработки геолого-промысловой информации, определяется следующим.

Любому продуктивному пласту, содержащему залежи нефти и газа, свойственна геологическая неоднородность. Геологической неоднородностью называют изменчивость условий залегания пород и их коллекторских свойств по площади и разрезу.

В зависимости от условий осадконакопления и диагенетических процессов, протекавших в недрах, пространственное распределение коллекторов в пласте и его коллекторские свойства могут резко изменяться даже на весьма малых площадях.

Неоднородность, выражающуюся в изменчивости пористости, проницаемости и других свойств коллекторов по площади и разрезу, называют **микронеоднородностью**. Микронеоднородность продуктивных пластов обуславливается различиями гранулометрического состава пород, формы частиц и их упаковкой, степени отсортированности и цементированности зерен породы, состава цементирующего материала, а также уплотнения осадков.

**Микронеоднородность** характеризуется количественным изменением свойств горных пород под действием множества случайных локальных факторов и, как правило, характеризуется однородным эмпирическим распределением (т.е. таким, под которое можно подобрать теоретическое, см. далее).

**Макронеоднородность** характеризует особенности пространственного распределения коллекторов в пласте. Она обусловлена значительными изменениями в условиях осадконакопления во времени и

пространстве, что влечет за собой чередование проницаемых и непроницаемых осадков в процессе их отложения.

В этом случае происходят качественные изменения свойств горных пород, появляется определяющая причина (условия осадконакопления), влияющая на распределение, геологических свойств и описываемая сложной, неоднородной функцией.

Опыт разведки и разработки большинства месторождений нефти и газа показывает, что продуктивные пласты с однородными коллекторскими свойствами представляют собой весьма редкое явление. Комплексное изучение неоднородности направлено на выявление в пределах каждой залежи зон распространения по площади и разрезу хорошо проницаемых, малопроницаемых и непроницаемых прослоев, на исследование гидродинамической связи между отдельными прослоями внутри продуктивного пласта и т.п. Прослеживание одноименных прослоев ведется с помощью детальной корреляции продуктивных толщ. Результаты геолого-промысловых исследований обобщают путем составления детальных геологических профилей продуктивного пласта, литофациальных и зональных карт, карт эффективных толщин, карт пористости и проницаемости коллекторов, песчанности и т.д. При этом микронеоднородность изучается с помощью карт и профилей проницаемости, пористости и других связанных с ними свойств коллекторов; макронеоднородность – с помощью детальных геологических профилей, карт толщины и т.п.

Однако неоднородность продуктивных пластов оказалась настолько сложным явлением, что для ее детального изучения потребовалось привлечение широкого аппарата математической статистики и теории вероятностей, а также теории информации. Математические методы позволяют более объективно анализировать фактический материал, количественно оценивать и сравнивать изучаемые объекты. Особенно сложной, до конца не решенной проблемой, является изучение и моделирование постседиментационной неоднородности, прежде всего, трещиноватости пород.

Результаты математической обработки используются при подсчете запасов, в процессе подготовки геологических данных при проектировании разработки и т.д.

На основе геолого-статистических методов изучение продуктивных пластов и содержащихся в них залежей нефти и газа осуществляется расчет параметров для подсчета запасов, выбор метода расчета эффективных объемов залежей.

Наряду с этим, при расчете эффективных объемов необходим учет кондиционных значений коллекторских свойств, определяемых на основе корреляционного анализа между геолого-промысловыми характеристиками.

На корреляционном анализе между предыдущими и последующими дебитами основан статистический метод подсчета запасов нефти. Все это обуславливает возможность широкого использования персональных компьютеров с программными пакетами нового поколения при подсчете запасов.

Изучение влияния геолого-физических и геолого-технологических факторов на нефтеотдачу диктует необходимость исследования корреляционных связей, имеющих многомерный характер, что требует большой затраты труда. К тому же ручная обработка большого объема геолого-промысловых данных малоэффективна.

В значительной мере повышению эффективности исследований при решении задач нефтегазопромысловой геологии способствует автоматизация обработки данных на основе применения рабочих станций и персональных компьютеров (ПК). Компьютерная обработка данных обеспечивает более глубокий детальный анализ геолого-физических данных, применение сложных математико-статистических моделей изучаемых объектов (пластов-коллекторов), рассмотрение различных вариантов связей между геолого-физическими параметрами и, наконец, уменьшение числа ошибок, связанных с недостаточной квалификацией исполнителей при ручной обработке данных.

Нефтегазопромысловая геология является, пожалуй, первой из отраслей нефтяной геологической науки, применивших статистические методы обработки и анализа промысловых данных.

Одной из первых следует признать работу старейшего русского геолога А.М. Коншина, в которой автор в начале 90-х годов XIX столетия осуществил подсчет запасов нефти для 4-х старых Бакинских площадей. В основу метода положена статистика добычи во времени. Более чем двумя десятилетиями позже американский геолог Паск опубликовал обобщающую работу, в которой дал общее название методам подсчета запасов нефти, основанным на применении математической статистики – методы кривых эксплуатации. Этими методами осуществил подсчет запасов нефти в 1916–1917 гг. по Ново-Грозненскому и Старо-Грозненскому районам известный русский геолог С.И. Чарноцкий, который в 1922 г. обобщил опыт подсчета запасов методами кривых эксплуатации у нас в стране и за рубежом. Впоследствии этот метод получил у нас название статистического, развитию которого в значительной мере способствовали работы В.В. Билибина (1930). В них дается теория применения методов математической статистики для подсчета запасов. В частности, В.В. Билибину принадлежит идея применения элементов статистики и для подсчета локализованных ресурсов. В 1934 г. В.В. Билибин совместно с М.А. Ждановым разработали методику выбора варианта размещения скважин на основе статистической обработки данных о количестве скважин, общей добыче на одну скважину и общего срока эксплуатации пласта  $C_2$  Апшеронского месторождения.

Дальнейшее развитие методов математической статистики в нефтегазопромысловой геологии связано с оценкой точности подсчета запасов нефти. Приоритет здесь принадлежит известному советскому геологу нефтянику М.В. Абрамовичу (1951).

Развитию математико-статистических методов в нефтегазопромысловой геологии для решения задач разведки, оценке точности подсчета запасов и разработки способствовали работы под руководством

Е.В. Фролова в лаборатории маркшейдерии и картографии ВНИИ-нефть.

В 50-х годах методы математической статистики начали использоваться для изучения геологической неоднородности и распределения параметров подсчета запасов. Первыми в этом направлении являются работы Е.И. Семина, Э.Л. Лейбина, Л.Ф. Дементьева, Е.А. Хитрова и др. Их исследования основываются на завоевавших популярность руководства по математической статистике и теории вероятностей Е.С. Вентцель, Н.В. Смирнова и И.В. Дунина-Барковского, Г.В. Гнеденко. Особое место здесь занимают монографии Л.Ф. Дементьева, явившиеся настольными книгами для многих промысловых геологов-нефтяников.

Широкие сведения о применении математических методов в различных отраслях геологии можно почерпнуть из работ А.Б. Вистелиуса, Д.А. Родионова, В.Н. Бондаренко, И.П. Шарапова, Ю.А. Воронина, С.В. Гольдина, В.П. Бухарцева, Р.И. Когана, а также из переводных работ Маттерона, У. Крумбейна, М. Кауфмена и Р. Мак-Кеммона. В настоящее время нет ни одного направления геологии, геофизики, разработки нефтяных и газовых месторождений, в котором изучение геологических объектов, проектирование и анализ разработки осуществлялись без применения персональных компьютеров и других более мощных быстродействующих устройств с построением постоянно действующих геолого-технологических моделей (ПДГТМ). Даже такая основополагающая в нефтегазопромысловой геологии и долго не поддававшаяся машинной обработке операция, как корреляция разрезов скважин, осуществляется теперь на ПК с помощью программного продукта AutoCorr, разработанного авторским коллективом сотрудников РГУ нефти и газа имени И.М. Губкина (И.С. Гутман, В.Е. Копылов, Г.П. Кузнецова), института прикладной математики им. М.В. Келдыша (И.Ю. Балабан), МГУ им. М.В. Ломоносова (В.М. Староверов), МНП (Н.Н. Лисовский).



# 1. ПРИМЕНЕНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ И ТЕОРИИ ВЕРОЯТНОСТИ В НЕФТЕГАЗОПРОМЫСЛОВОЙ ГЕОЛОГИИ

## 1.1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И СЛУЧАЙНЫЕ ЯВЛЕНИЯ В НЕФТЕГАЗОПРОМЫСЛОВОЙ ГЕОЛОГИИ

Величина, которая в результате наблюдения может принимать то или иное значение, причем, какое именно, заранее не известно, называется *случайной величиной*.

Например, нельзя заранее предсказать, сколько проницаемых прослоев (2, 3 или 4) будет установлено в скважине, которая должна вскрыть такой расчленённый продуктивный пласт, характеризующийся макронеоднородностью, или какова открытая пористость каждого из этих прослоев при микронеоднородности коллектора.

Число прослоев и значения открытой пористости будут случайными величинами, однако между ними есть существенное различие. Количество прослоев в продуктивном пласте, характеризуя его расчлененность, может изменяться от нуля до какого-либо числа, образуя тем самым определенный интервал целых чисел, отдельных и прерывными относительно друг друга. Такие случайные величины называются *прерывными* или *дискретными*.

В отличие от числа прослоев как параметра макронеоднородности возможные значения параметров микронеоднородности, например, открытой пористости, изменяются в некотором интервале чисел непрерывно. Теоретически интервал изменения открытой пористости ограничен возможным непрерывным рядом чисел от 0 до 0,40. Отдельные конкретные значения открытой пористости продуктивного пласта *непрерывно* заполняют этот интервал.

Случайные величины, возможные значения которых непрерывно заполняют некоторый интервал, называются *непрерывными*. К числу непрерывных случайных величин можно отнести, кроме пористости,

проницаемость, глинистость, содержание цемента в породе, толщину пласта или прослоя, нефтегазонасыщенность, выявленные размеры поднятий, к которым приурочены залежи нефти и газа, выявленные размеры запасов этих залежей и т.п.

Случайные величины всегда связаны со *случайным явлением*, наступление которого нельзя предсказать. Этой характеристике случайного явления, казалось бы, противоречит представление о том, что ни одно из геологических явлений нельзя относить к категории случайных в силу того, что все они подчиняются определенным геологическим закономерностям, то есть носят *причинно-следственный*, или детерминированный характер.

Действительно, в тех случаях, когда схему геологического процесса представить легко, а число рассматриваемых переменных незначительно, «причинно-следственный» подход к решению таких задач весьма эффективен. В таких случаях используется одно или несколько дифференциальных уравнений при заданных граничных условиях для выражения скорости изменения функции относительно изменения одного или нескольких аргументов. На основе наблюдаемых или теоретически рассчитанных данных получают различные характеристики зависимости  $Y = f(X_1, X_2, X_3...)$ , позволяющей точно предсказывать значение  $Y$  по заданным  $X_1, X_2...$

С увеличением числа рассматриваемых переменных сказывается влияние многочисленных усложняющих факторов, которые придают геологическому явлению случайный характер.

Сказанное выше Л.Ф. Дементьев [9] наглядно иллюстрирует анализом условий формирования открытой пористости и проницаемости терригенного коллектора.

Эти коллекторские свойства пород зависят от размера обломочных частиц, их окатанности, отсортированности, глинистости, от того, к какой группе относятся глинистые минералы, от минералогического состава пород. На фильтрационные свойства сцементированных пород существенное влияние оказывает состав цемента и его со-

держание в породе. Кроме того, свойства осадочной породы зависят от свойств первичных пород, послуживших источником образования обломочного материала, от однородности первичных пород с точки зрения их состава и сопротивляемости процессам выветривания и денудации.

Процесс формирования новых осадков протекает по известной схеме: с удалением от берега прибрежные пески сменяются алевритами, последние уступают место глинистым илам, еще глубже область развития известкового ила. Однако эта схема, справедливая в общем для любого бассейна, в каждом конкретном районе осложняется влиянием течений, рельефа дна, орографии берегов, а также наличием рек, волнениями моря, приливами и отливами. Все эти факторы вносят порой существенные локальные отклонения в общую схему регионального размещения осадочных пород. Если к перечисленному добавить различные условия, при которых происходит диагенез осадков, то становится очевидной сложность геологического процесса формирования коллекторов нефти и газа, обусловленная влиянием различных факторов, действующих на конкретные свойства пород. Естественно, все факторы, влияющие на коллекторские свойства пород, оказывают его по-разному на разных участках исследуемой площади.

Если в региональном плане закономерное изменение одного типа пород другим влечет закономерное изменение их коллекторских свойств, то локальные изменения гидродинамических и гидрохимических условий ведут к существенным изменениям коллекторских свойств на ограниченных участках. В связи с тем, что характер этих изменений на локальных участках геологу заранее не известен, он не может предсказать величину открытой пористости, проницаемости или какого-либо иного параметра в той или иной точке. В этом смысле коллекторские свойства горных пород, толщины пластов и прослоев, а также перечисленные выше характеристики их неоднородности в каждой отдельной точке являются величинами случайными.

Таким образом, все геологические явления, в сущности своей детерминированные, в силу большой локальной изменчивости пород под влиянием множества различных факторов и дискретности наблюдений носят более или менее случайный характер.

## 1.2. ВЫБОРКА И ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ, ВЕРОЯТНОСТНАЯ МОДЕЛЬ

Промысловый геолог создает представление о строении продуктивного пласта, скрытого от него мощной толщей осадочных пород, только по данным, полученным из скважин. Расстояния между эксплуатационными скважинами на нефтяных месторождениях измеряются несколькими сотнями (а на старых месторождениях десятками) метров. Расстояния между поисковыми и разведочными скважинами, в которых отбирается керн и которые несут наибольшую информацию о геологическом строении пласта и характере неоднородности его коллекторских свойств, измеряются несколькими километрами. Естественно, что данные, полученные из разведочных и даже эксплуатационных скважин, представляют собой *выборочную совокупность*  $n$  значений случайной величины  $X$  (например, пористости, проницаемости, толщины и т.п.).

Каждая из таких выборочных совокупностей или выборок может дать представление о генеральной совокупности  $N$  значений изучаемой величины  $X$ .

Генеральную совокупность значений случайной величины  $X$  изучаемого объекта образуют все ее возможные значения по всему его объему. Если, например, исследуется толщина продуктивного горизонта, то генеральную совокупность этого параметра в пределах залежи составляли бы значения толщин по скважинам, условно пробуренным вплотную друг к другу. Следовательно, генеральная совокупность  $N$  имеет бесконечный объем. Поскольку достижение такой плотности наблюдений нереально, о генеральной совокупности судят

по ограниченному объему данных, то есть по выборочным совокупностям.

Выборочная совокупность должна состоять из наблюдений по одному объекту, выполненных равномерно по площади и разрезу в одинаковых условиях опробования.

Генеральную совокупность характеризуют точные характеристики: функция распределения  $F(X)$ , математическое ожидание  $M[X]$ , дисперсия  $\sigma^2(X)$ , вероятность и т.д. Характеристиками выборочной совокупности, соответственно, являются: ряд распределения, среднее арифметическое значение  $\bar{x}$  как оценка математического ожидания, оценка дисперсии  $S^2(X)$ , частота  $\omega_i$  и т.д. Выборочные характеристики отличаются от соответствующих характеристик генеральной совокупности  $N$  из-за ограниченного объема  $n$  случайной выборки. Увеличение числа наблюдений, увеличение объема случайной выборки  $n$  способствует возрастанию ее точности, позволяет более полно охарактеризовать генеральную совокупность. При неограниченном увеличении случайной выборки все выборочные характеристики приближаются, или, как говорят, сходятся по вероятности к соответствующим характеристикам генеральной совокупности.

Это свойство является отражением **закона больших чисел**, которым называется совокупность теорем о вероятностях событий при большом числе наблюдений.

Так, теорема Я. Бернулли доказывает, что при большом числе наблюдений в большинстве случаев частота  $\omega_i$ , под которой понимается отношение частоты  $Z_i$  значений параметра внутри классового промежутка к общему объему  $n$  случайной выборки ( $\omega_i = \frac{Z_i}{n}$ ), должна быть близка к вероятности  $p_i$ . Это подтверждается и предельной теоремой Лапласа, согласно которой с вероятностью, близкой к 1, можно утверждать, что при большом числе наблюдений абсолютные значения отклонений  $\omega_i$  от  $p_i$  минимальны (в некоторых руководствах частоты называют статистическими или эмпирическими вероятностями).

Важное следствие вытекает из теоремы Чебышева: при достаточно большом числе наблюдений с вероятностью, близкой к 1, можно ожидать, что среднее из наблюдаемых значений случайной величины будет как угодно мало отличаться от математического ожидания.

При изучении пласта-коллектора, вмещающего залежь, следует добиваться равномерности сети наблюдений по всей площади. В противном случае при резкой неоднородности пласта-коллектора могут быть получены результаты, искажающие параметры генеральной совокупности. Равномерность наблюдений является важнейшим свойством выборочной совокупности. Если выборочная совокупность неравномерно покрывает изучаемый объект, то можно попробовать использовать одну из методик декластеризации, для получения более репрезентативной выборки.

**Декластеризация** – название комплекса методик для получения более репрезентативной выборки. Пример методик декластеризации – использование средних по равномерно распределённым классам, введение весов для точек выборки (там, где выборка плотнее, вес меньше – обратно пропорционально плотности).

Вместе с тем, стоит хоть немного изменить место заложения скважин при том же их числе, как выборочная совокупность окажется иной, но характеризующей генеральную совокупность в такой же степени, как и прежняя. Следовательно, у каждого значения случайной величины существует такая же объективная возможность оказаться в числе «выбранных», как и у любого другого значения генеральной совокупности.

Таким образом, выборочная совокупность представляет собой оценку реального геологического объекта как генеральной совокупности, являясь его *вероятностной моделью*.

Этим обуславливается возможность математического моделирования геологических объектов с помощью методов теории вероятностей.

### 1.3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ В ГЕОЛОГИИ

Применение математических методов при решении задач различных отраслей геологии основано на представлении реальных геологических объектов в виде математических моделей. Математическая модель должна учитывать специфические особенности геологических объектов, отражать их структуру. Все математические модели содержат то или иное численное представление изучаемых по имеющейся выборке геологических параметров.

Для опытного, эмпирического, распределения, полученного по выборочным данным каждого из моделируемых геологических параметров, можно подобрать теоретическую вероятностную модель, с помощью которой оцениваются основные черты его эмпирического распределения. Правильность выбора модели распределения проверяется путем проверки гипотезы о соответствии теоретического распределения эмпирическому.

Продуктивный пласт или нефтяную залежь в целом можно также представить в виде модели, охарактеризованной рядом параметров, являющихся, в силу сказанного выше, случайными величинами, для выборочного распределения которых подбирается та или иная теоретическая модель.

Геологическая модель продуктивного пласта залежи УВ, являющаяся самостоятельной или составной частью постоянно действующей геолого-технологической модели, создается с применением комплекса различного рода программных средств, охватывающих все многогранные характеристики изучаемых объектов.

С практической точки зрения при построении модели геологического объекта надо решить, насколько объёмная, детальная и сложная модель нам нужна для достижения целей моделирования.

**Сложность модели** отражает количество и качество взаимосвязей и типов неоднородностей, учитываемых в модели.

**Объём модели** определяется как количество данных, которые в неё загружены.

**Детальность модели** показывает, насколько подробную информацию гипотетически можно было бы получить в результате моделирования.

Например, наиболее простой моделью пласта может быть набор усреднённых значений параметров пласта, таких как среднее значение коэффициента открытой пористости, среднее значение эффективной толщины. Используя такую модель можно вычислить объём порового пространства.

Более сложной моделью является набор карт: общих толщин, эффективных толщин, ВНК, эффективных нефтенасыщенных толщин. Эта модель учитывает неоднородность пласта по простиранию.

Однако, неоднородности пласта по глубине в заданной точке карты не отражаются. Если в модели необходимо учесть и вертикальную неоднородность пласта коллектора, то лучше подходит 3D модель, в которой пласт представлен в виде набора кубов свойств. Кроме того, модели могут поддерживать (более сложные модели) или не поддерживать (более простые) наличие разломов (разрывных нарушений) определённых типов и другие виды неоднородности пластов.

Что касается объёма модели залежи, то одна, менее объёмная модель может моделировать залежь в одном пласте, а другая, более объёмная – может моделировать множество пластов, содержащихся в залежи.

Детальность модели можно проиллюстрировать на примере модели в виде набора карт. В этом случае детальность определяется шагом, с которым на картах проведены изолинии – чем он меньше, тем модель детальнее. Следует отметить, что уменьшение шага изолиний имеет свой предел и совсем не обязательно приводит тому, что можно будет обнаружить более мелкие детали.



## 2. ИЗУЧЕНИЕ ОСОБЕННОСТЕЙ РАСПРЕДЕЛЕНИЯ ПАРАМЕТРОВ НЕОДНОРОДНЫХ ПРОДУКТИВНЫХ ПЛАСТОВ С ПОМОЩЬЮ МАТЕМАТИЧЕСКИХ МЕТОДОВ

### 2.1. РЯДЫ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Большинство исследований в нефтегазопромысловой геологии связано с наблюдениями. Так, по геофизическим и керновым данным устанавливается толщина продуктивного пласта, число и толщина слагающих его проницаемых прослоев, их пористость, нефтенасыщенность и другие параметры. *Любой из видов геологических исследований, связанных с наблюдениями, можно обобщить в виде эмпирического (выборочного, статистического) ряда распределения, независимо от того, с какой случайной величиной имеет дело исследователь – дискретной или непрерывной.*

Простой формой выражения эмпирического ряда распределения является **таблица распределения**. Для ее составления весь ряд наблюдаемых значений случайной величины, например пористости, разбивается на ряд равновеликих интервалов (классов, разрядов), в каждом из которых определяется число попавших в него наблюдений (частота)  $Z_i$ , частость  $\omega_i$ , накопленная частость  $\Sigma\omega_i$ . Сумма всех частостей должна быть равна 1. Единице равна и накопленная частость в последнем интервале.

Для составления таблицы весь диапазон значений необходимо разделить на более мелкие равные промежутки, называемые *классами, или интервалами, разрядами*. Количество классов зависит от величины классовых промежутков ( $\Delta$ ). Величину  $\Delta$  класса рекомендуется определять по формуле

$$\Delta = \frac{R_b}{5 \lg n},$$

где  $n$  – объем выборки,  $R_b$  – размах выборки.

Пример. Таблица распределения для выборки из 100 определений.

Таблица 1

Интервалы (разряды, классы)	Среднее интервала	Частота $Z_i$	Частость $\omega_i$	Плотность $\frac{\omega_i}{\Delta}$ частности	Накопленная частость $\Sigma\omega_i$
2–4	3	2	0,019	0,0085	0,019
4–6	5	4	0,039	0,0195	0,058
6–8	7	7	0,068	0,034	0,126
8–10	9	15	0,145	0,0725	0,271
10–12	11	17	0,165	0,0825	0,436
12–14	13	19	0,185	0,0925	0,621
14–16	15	20	0,195	0,0975	0,816
16–18	17	10	0,097	0,0485	0,913
18–20	19	5	0,049	0,0245	0,962
20–22	21	2	0,019	0,0085	0,981
22–24	23	2	0,019	0,0085	1,000
		$\Sigma = 100$	$\Sigma = 1,000$		

Данные таблицы распределения графически изображаются в виде **гистограммы, полигона распределения и кумулятивной кривой**. Для построения гистограммы на оси абсцисс откладывают в выбранном масштабе интервалы, а на оси ординат – отношение частостей к величине классового промежутка, называемое плотностью частостей  $\left(\frac{\omega_i}{\Delta}\right)$ . График гистограммы является ступенчатым (рис. 1).

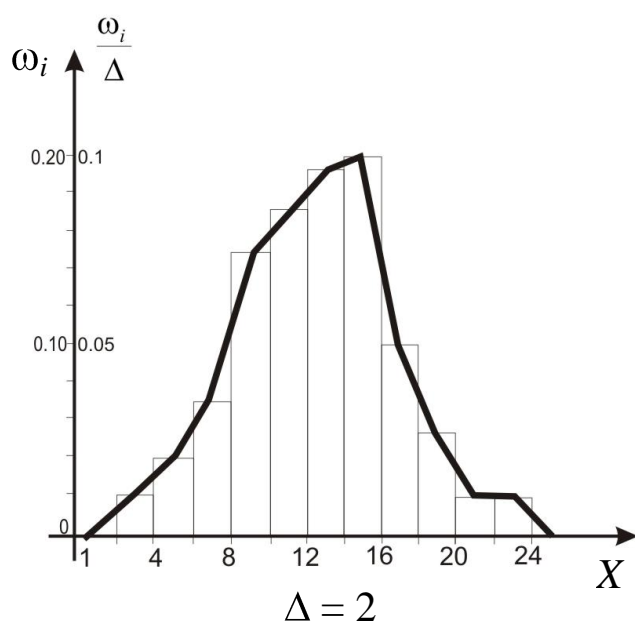


Рис. 1. Гистограмма и полигон распределения

Суммарная площадь прямоугольников гистограммы равна  $\Delta \cdot \sum \frac{\omega_i}{\Delta} = 1$ .

При построении полигонов распределения значения частот или частостей в каждом классе приурочивают к середине класса. Их графики представляют ломанные линии (рис. 1).

Кумулятивная кривая изображает графически накопленные частоты. Так как она является интегральной кривой, значения накопленных частостей откладываются на правом конце каждого интервала (рис. 2).

Анализ изучения распределения толщин продуктивных пластов и их прослоев, пористости, проницаемости и других параметров показывает, что гистограммы и полигоны распределения могут в каждом конкретном случае принимать различную форму. Они могут быть симметричными и асимметричными, островершинными и плосковершинными, одновершинными, двух- и трехвершинными.

Все эти различия обусловлены особенностями распределения параметров, различной степенью разбуренности разных частей залежи, неравномерным выносом керна, закономерностями распределения параметров исследуемых объектов как по площади залежи, так и по разрезу.

Ю.П. Борисовым, З.К. Рябининой и В.В. Воиновым [4] предлагается различать однородные и неоднородные эмпирические распределе-

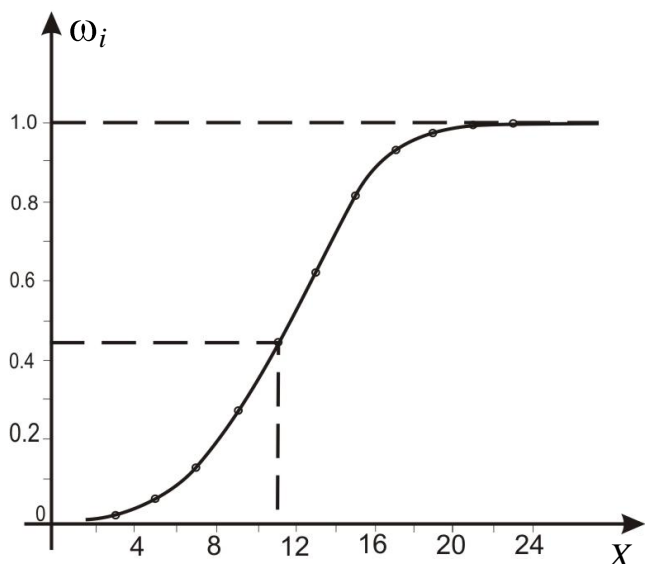


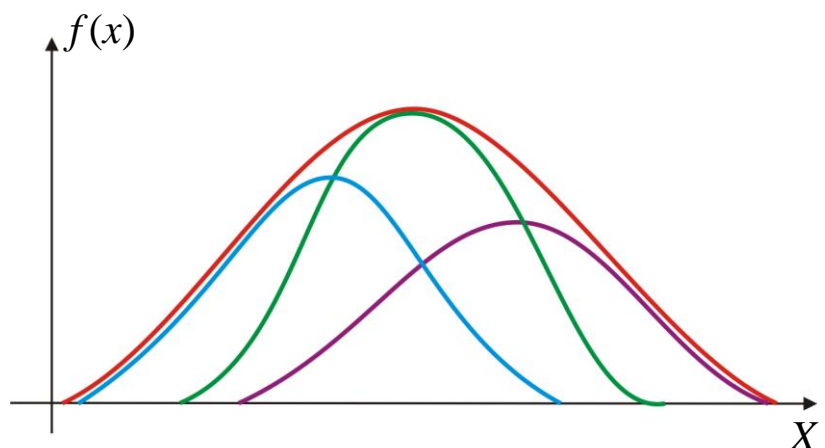
Рис. 2. Кумулятивная кривая

ния. Под однородным понимается распределение, для которого можно подобрать какую-либо из теоретических кривых распределения. Неоднородное, или сложное, распределение состоит, как правило, из нескольких однородных, входящих в общее различными долями (рис. 3).

Выбор теоретической модели, которая бы наилучшим образом описывала эмпирическое распределение, – важная задача геолого-промыслового анализа. Эта процедура выполняется путем проверки гипотезы о соответствии эмпирического распределения той или иной теоретической функции распределения с помощью критерия согласия.

Функция распределения – это полная характеристика случайной величины. Она полностью описывает как дискретные, так и непрерывные случайные величины с вероятностной точки зрения, то есть является одной из форм закона распределения.

Однако в целом ряде случаев невозможно или нецелесообразно охарактеризовать распределение изучаемого параметра с помощью функции распределения. Так, для неоднородного распределения нельзя, например, подобрать теоретическую кривую. В этом случае для описания эмпирического распределения параметров неоднородных пластов и сравнения его с распределением тех же параметров по другим объектам применяют точечные числовые характеристики, каждая из которых характеризует какие-либо основные черты эмпирического распределения.



**Рис. 3** Составные части неоднородного распределения

## 2.2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ РАСПРЕДЕЛЕНИЙ

Точечные числовые характеристики могут характеризовать положение случайной величины на числовой оси, степень ее рассеяния, асимметричность, крутость кривой распределения и т.д.

### Математическое ожидание

Математическое ожидание  $M X$ , или  $m_x$ , является важнейшей характеристикой положения случайной величины на числовой оси.

Математическим ожиданием случайной величины называется сумма произведений всех возможных ее значений на вероятности этих значений.

Для прерывной случайной величины  $X$  в соответствии с этим определением математическое ожидание вычисляется по формуле

$$M[X] = m_x = \sum_{i=1}^n x_i p_i, \quad (1)$$

где  $M[X]$ , или  $m_x$  – математическое ожидание,  $x_i$  – значение случайной величины,  $p_i$  – вероятности этих значений.

Для непрерывной случайной величины математическое ожидание выражается интегралом

$$M X = m_x = \int_{-\infty}^{\infty} x \cdot f(x) dx. \quad (2)$$

Эта формула соответствует формуле (1), если в последней  $x_i$  заменить на непрерывно изменяющийся параметр  $x$ ; соответствующие вероятности  $p_i$  – элементом вероятности  $f(x)dx$ , где  $f(x)$  – плотность распределения величины  $X$ ; конечную сумму – интегралом.

В механической интерпретации математическое ожидание представляет собой абсциссу центра тяжести. В случае прерывной случайной величины это будет абсцисса центра тяжести системы материальных точек, а в случае непрерывной – абсцисса центра тяжести, когда масса распределена по этой оси непрерывно с плотностью  $f(x)$ .

При рассмотрении выборочного распределения оценкой математического ожидания является среднее арифметическое наблюдаемых значений случайной величины, которое вычисляется по формуле

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (3)$$

где  $x_i$  – наблюдаемые значения случайной величины,  $n$  – число наблюдений.

При большом числе наблюдений среднее арифметическое наблюдаемых значений приближается (сходится по вероятности) к математическому ожиданию, что является содержанием одной из форм закона больших чисел.

### Начальные и центральные моменты

Математическое ожидание, характеризуя положение случайной величины на числовой оси, является одной из ряда характеристик, описывающих какое-либо из свойств распределения и называемых **моментами**. На практике наиболее применимы моменты двух видов: начальные и центральные.

Начальным моментом  $s$ -го порядка прерывной случайной величины  $X$  называется сумма вида:

$$\alpha_s[X] = \sum_{i=1}^n x_i^s p_i. \quad (4)$$

Для непрерывной случайной величины  $X$  начальным моментом  $s$ -го порядка называется интеграл.

$$\alpha_s[X] = \int_{-\infty}^{\infty} x^s f(x) dx. \quad (5)$$

Сравнивая формулы, соответственно (1) и (4), а также (2) и (5), нетрудно убедиться, что математическое ожидание – это первый начальный момент случайной величины  $X$ . Из сравнения этих формул также следует, что начальным моментом  $s$ -го порядка случайной ве-

личины  $X$  называется математическое ожидание в степени  $s$  этой случайной величины.

Центральные моменты, в отличие от начальных, являются моментами центрированной случайной величины, которой называется отклонение случайной величины  $X$  от ее математического ожидания

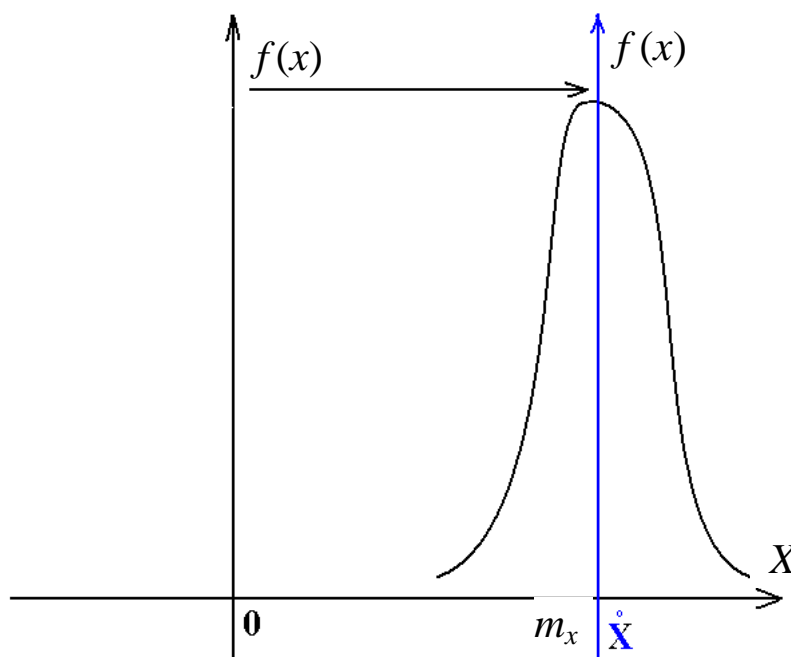
$$\overset{\circ}{X} = X - M[X], \text{ или } \overset{\circ}{X} = X - m_x. \quad (6)$$

Математическое ожидание центрированной случайной величины равно 0, откуда следует, что центрирование случайной величины равносильно переносу начала координат в точку, абсцисса которой равна математическому ожиданию  $m_x$  (см. рис. 4).

В общем виде центральным моментом  $s$ -го порядка случайной величины  $X$  называется математическое ожидание  $s$ -ой степени соответствующей центрированной случайной величины.

Для прерывной случайной величины  $s$ -й центральный момент выражается суммой

$$\mu_s[X] = \sum_{i=1}^n (x_i - m_i)^s, \quad (7)$$



**Рис. 4. Центрирование случайной величины**

а для непрерывной – интегралом

$$\mu_s[X] = \sum_{i=1}^n (x_i - m_i)^s. \quad (8)$$

В качестве важнейших характеристик случайной величины чаще всего применяются первый начальный момент (математическое ожидание), второй, третий и четвертый центральные моменты.

### Дисперсия

Дисперсия является вторым центральным моментом случайной величины  $\mu_2 = D[X]$ . Дисперсия является основной характеристикой рассеяния случайной величины около ее математического ожидания. В соответствии с формулой (6) дисперсией случайной величины  $X$  называется математическое ожидание квадрата соответствующей центрированной величины.

Дисперсия для дискретных случайных величин, представляющих генеральную совокупность, равна

$$D_x = D[X] = \sum_{i=1}^N (x - M[X])^2 p_i, \quad (9)$$

а для непрерывных

$$D_x = D[X] = \int_{-\infty}^{\infty} (x - M[X])^2 f(x) dx. \quad (10)$$

Оценка дисперсии, эмпирического ряда наблюдений вычисляется по формуле:

$$S_x^2 = S^2[X] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (11)$$

где  $\bar{x}$  – среднее арифметическое значение.

Дисперсия имеет размерность квадрата случайной величины.

В качестве меры рассеяния случайной величины более удобно пользоваться **средним квадратическим отклонением**, или стандартным отклонением  $\sigma[X]$ , равным корню квадратному из дисперсии:



$$\sigma_x = \sigma[X] = \sqrt{D[X]}. \quad (12)$$

В соответствии с формулой (11) оценка  $S_x$  среднего квадратического отклонения равна

$$S_x = s[X] = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (13)$$

Среднее квадратическое отклонение и его оценка вычисляются со знаком плюс и имеют размерность исследуемой случайной величины. Оно служит абсолютной мерой неоднородности. Чем больше величина оценки среднего квадратического отклонения, тем выше неоднородность пласта.

### Коэффициент вариации

Среди статистических мер неоднородности важное место отводится **коэффициенту вариации**  $W_x$ , равному отношению среднего квадратического отклонения параметра к математическому ожиданию:

$$W_x = \frac{\sigma_x}{m_x}, \quad (14)$$

где  $\sigma_x$  – среднее квадратическое отклонение параметра,  $m_x$  – математическое ожидание.

Для эмпирического распределения оценка коэффициента вариации, соответственно, равна:

$$\hat{W} = \frac{S_x}{\bar{x}}, \quad (15)$$

где  $S_x$  – оценка среднего квадратического отклонения,  $\bar{x}$  – среднее значение выборки.

Коэффициент вариации служит относительной мерой неоднородности продуктивного пласта, благодаря чему он широко используется при изучении неоднородности. Однако на возможность применения коэффициента вариации оказывает влияние коррелируе-

мость между собой среднего значения и среднего квадратического отклонения.

А.И. Орловым и Ю.В. Шуруборм установлено, что при наличии тесной корреляционной связи между этими статистическими характеристиками относительная мера неоднородности  $W$  более информативна, чем абсолютная –  $\sigma_x$ . Для условий яснополянского надгоризонта Пермского Прикамья этими исследователями показано, что средние значения и средние квадратические отклонения распределений толщин отдельного прослоя и числа прослоев в указанном горизонте тесно коррелируются между собой. Вместе с тем, почти полностью отсутствует корреляционная связь между  $\bar{x}$  и  $\sigma_x$  для таких параметров, как общая толщина яснополянского надгоризонта, эффективная толщина, пористость, проницаемость и др.

Следовательно, неоднородность по таким параметрам, как толщина нефтенасыщенных прослоев и их число в продуктивном горизонте, лучше оценивать с помощью коэффициента вариации, тогда как для оценки неоднородности коллекторских свойств продуктивных пластов и других указанных выше параметров более целесообразно пользоваться абсолютной мерой неоднородности – средним квадратическим отклонением.

### Коэффициент асимметрии

Третий центральный момент характеризует асимметрию распределения, коэффициент которой  $A_x$  для генеральной совокупности дискретных величин равен

$$A_x = \frac{1}{\sigma_x^3} \sum_{i=1}^n (x_i - m_x)^3 p_i, \quad (16)$$

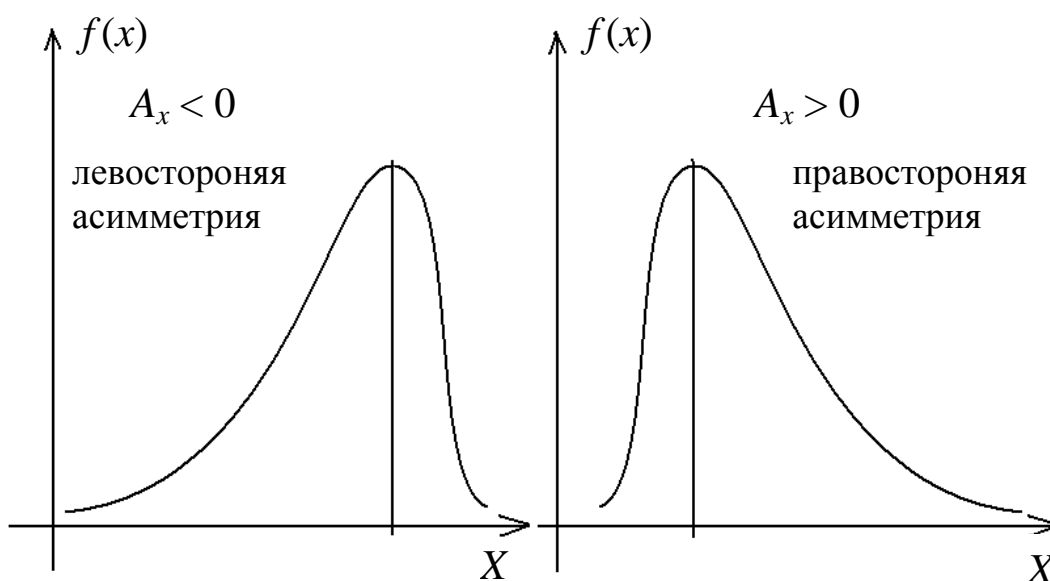
для непрерывных случайных величин

$$A_x = \frac{1}{\sigma_x^3} \int_{-\infty}^{\infty} (x_i - m_x)^3 f(x) dx. \quad (17)$$

Для эмпирического ряда наблюдений оценка коэффициента асимметрии вычисляется по формуле:

$$\hat{A}_x = \frac{1}{S_x^3} \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n}. \quad (18)$$

Если распределение симметрично относительно математического ожидания, то  $\hat{A}_x = 0$ . Это характерно для нормального распределения. При  $\hat{A}_x > 0$  распределение будет иметь положительную асимметрию, при  $\hat{A}_x < 0$  – отрицательную (см. рис. 5). Отрицательная асимметрия называется левосторонней, положительная – правосторонней. Эти названия объясняются тем, что если провести линию через вершину (моду) распределения, как это сделано на рис. 5, то у распределения с левосторонней асимметрией большая часть графика окажется слева от этой линии, а у распределения с правосторонней – справа.



**Рис. 5. Левосторонняя и правосторонняя асимметрия**

### Экцесс

Четвертый центральный момент характеризует крутость кривой распределения (см. рис. 6). Он имеет размерность четвертой степени случайной величины. Его формулы для непрерывной и дискретной

случайной величины аналогичны формулам (16) и (17) с той лишь разницей, что значения  $\sigma_x$ ,  $S_x$  и выражения в скобках берутся в четвертой степени.

С помощью четвертого центрального момента вычисляют эксцесс  $\hat{\epsilon}$ , характеризующий меру отклонения кривой эмпирического распределения от нормального по оси ординат. Оценку эксцесса вычисляют по следующей формуле:

$$\hat{\epsilon}_x = \frac{1}{S_x^4} \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{n} - 3. \quad (19)$$

Следует учесть, что для нормального распределения выражение под знаком суммы равно 3. Поэтому эксцесс для нормального распределения равен 0, при крутовершинном распределении  $\epsilon > 0$ , при плосковершинном  $\epsilon < 0$ .

Для непрерывных случайных величин:

$$\epsilon_x = \frac{1}{\sigma_x^4} \int_{-\infty}^{\infty} (x_i - m_x)^4 f(x) dx - 3.$$

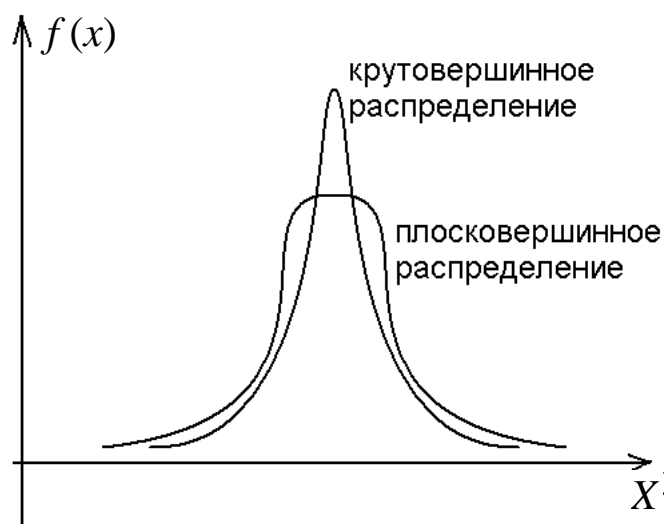


Рис. 6. Крутовершинное и плосковершинное распределения

## Эффект основания

Вышеприведённые рассуждения приводились в предположении, что измерения имеют точечный характер, хотя нередко оперируют значениями, являющимися результатом осреднения или суммирования реальных значений по некоторой области (основанию). Когда идёт речь о характеристиках распределения вычисленных по выборке осреднённых величин, необходимо иметь в виду **эффект основания**. Начиная со второго, все центральные моменты по выборке осреднённых величин будут меньше, чем в генеральной совокупности. Например, дисперсия среднесуточных по месяцам дебитов скважин, также как и коэффициент вариации среднемесячного дебита, меньше чем соответствующие величины, вычисленные по ежесуточным дебитам.

Перечисленные выше основные характеристики распределения довольно полно описывают эмпирические распределения значений различных параметров продуктивных пластов. При изучении, например, пористости в первую очередь по данным керна вычисляются ее среднее значение по отдельным скважинам в целом по залежи или отдельным ее частям. Из сравнения средних значений создается представление, какой из прослоев лучше, какой хуже, в какой части залежи лучше пористость, в какой хуже.

Из сравнения дисперсий и средних квадратических отклонений делается вывод о степени рассеяния значений открытой пористости около среднего значения. Чем выше значения дисперсии и среднего квадратического отклонения, тем сильнее неоднородность пласта.

В связи с тем, что продуктивный пласт зачастую представлен чередованием проницаемых и непроницаемых прослоев, геолог должен вводить в анализ только те значения открытой пористости, которые характеризуют коллектор. Однако и в этом случае картина распределения может быть резко искажена, что повлечет за собой увеличение дисперсии и среднего квадратического отклонения. Более того, рас-

пределение может оказаться двухвершинным, левая вершина которого будет характеризовать распределение значений открытой пористости для слабо проницаемых прослоев.

В сложно построенных коллекторах, к числу которых можно отнести карбонатные породы, дисперсия и среднее квадратическое отклонение могут возрастать из-за влияния каверновой составляющей, т.к. кавернозная пористость всегда больше межгранулярной. При этом не исключено появление плосковершинного распределения и даже двухвершинного. В этих случаях распределения будут иметь отрицательный эксцесс. Если число определений кавернозной пористости преобладает в их общем объеме, то распределение будет иметь правостороннюю асимметрию ( $A_x > 0$ ). Возникновение левосторонней асимметрии ( $A_x < 0$ ) может быть обусловлено неравномерным выносом керна.

Чем однороднее коллектор, тем меньше размах, дисперсия и среднее квадратическое отклонение. Для распределения открытой пористости однородного коллектора характерны симметричность кривой, поскольку подавляющее большинство определений имеют значения, близкие к среднему арифметическому.

### 3. ФУНКЦИИ РАСПРЕДЕЛЕНИЯ В НЕФТЕГАЗОПРОМЫСЛОВОЙ ГЕОЛОГИИ

#### 3.1. ПОНЯТИЕ О ФУНКЦИЯХ РАСПРЕДЕЛЕНИЯ

Выше отмечалось, что функция распределения является полной характеристикой случайной величины как дискретной, так и непрерывной.

Рассмотрим сначала функции распределения дискретных случайных величин на примере изменения числа прослоев в пласте.

Пусть продуктивный пласт содержит до 3 прослоев. Обозначим через  $X$  число прослоев в продуктивном пласте, а его возможные значения в отдельных скважинах через  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$ ,  $x_4 = 3$ .

$$X = x_1 = 0$$

$$X = x_2 = 1$$

$$X = x_3 = 2$$

$$X = x_4 = 3.$$

Для каждой скважины число этих прослоев возможно, но не достоверно. Величина  $X$  может принимать любое значение с некоторой вероятностью  $p_i$ .

Вероятности каждого из 4 указанных выше событий можно записать в виде:

$$P(X = x_1) = p_1; P(X = x_2) = p_2; P(X = x_3) = p_3; P(X = x_4) = p_4. \quad (20)$$

Естественно, что все эти события в каждой из скважин несовместны. В связи с этим, сумма их вероятностей равна 1, то есть

$$\sum_{i=1}^n p_i = 1.$$

Это равенство свидетельствует о том, что для данного конкретного случая суммарная вероятность определенным образом распределе-

на между отдельными значениями. Если будет указано, какой вероятностью характеризуется каждое из событий, то этим задается распределение.

По данным скважин вероятности каждого из несовместных событий распределились следующим образом: вероятность того, что в скважинах не обнаружено ни одного проницаемого прослоя, равна 0,216; вероятность обнаружения одного прослоя – 0,432; вероятность обнаружения двух прослоев – 0,288; вероятность обнаружения трех прослоев – 0,064.

Эти данные сводятся в таблицу распределения (табл. 2)

Таблица 2

Возможное значение $k_p = x_i$	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	$x_4 = 3$	
Вероятность $p_i$	0,216	0,432	0,288	0,064	$\Sigma p_i = 1,0$

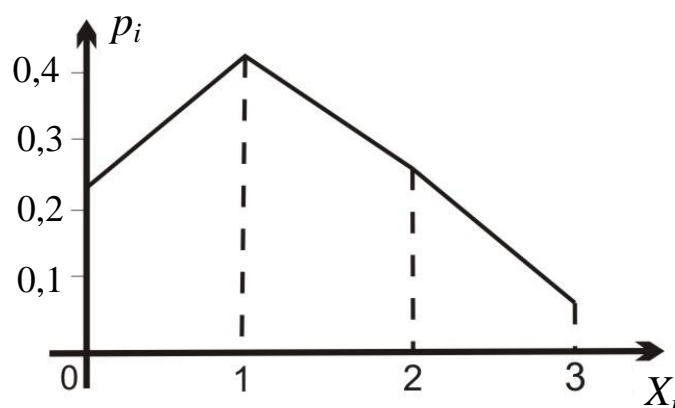


Рис. 7. Пример распределения вероятностей возможного количества прослоев в пласте

Графическим изображением ряда распределения является многоугольник распределения, который для условий табл. 2 будет иметь вид, представленный на рис. 7.

### 3.2. ИНТЕГРАЛЬНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Исходя из формулы (20), формой задания дискретной случайной величины служит функция распределения:



$$F(x) = \sum_{x_i < x} P(X=x_i), \quad (21)$$

где неравенство  $x_i < x$  справедливо для всех дискретных значений  $x_i$ , меньших некоторой текущей переменной  $x$ .

Графически функция распределения  $F(x)$  любой дискретной величины изображается в виде разрывной ступенчатой кумулятивной функции со скачками на уровне возможных значений случайной величины. Величины скачков равны вероятностям этих значений, а их сумма равна 1.

Для условий нашего примера функция  $F(x)$  будет иметь вид, изображенный на рис. 8.

По мере увеличения числа возможных значений дискретной случайной величины будет возрастать число скачков, но сами скачки станут меньше. Это приведет к сглаживанию кривой (рис. 9А). Дискретная величина будет постепенно приближаться к непрерывной, а ее функция распределения – к непрерывной функции (рис. 9Б).

Непрерывная случайная величина имеет бесчисленное множество значений, сплошь заполняющих некоторый интервал. В силу того, что вероятность любого отдельного значения непрерывной случайной величины равна 0, для ее количественной характеристики пользуются не вероятностью события  $X = x_i$ , а вероятностью события  $X < x$  в интервале от  $-\infty$  до  $x$ , где  $x$  – некоторая текущая переменная.

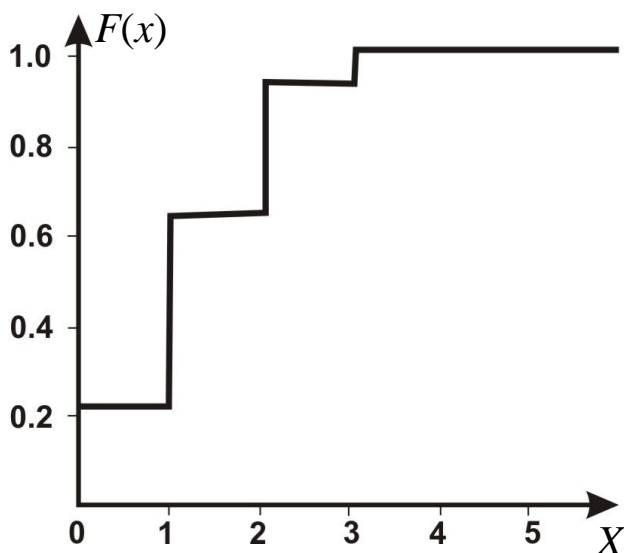


Рис. 8. Пример функции распределения дискретной величины

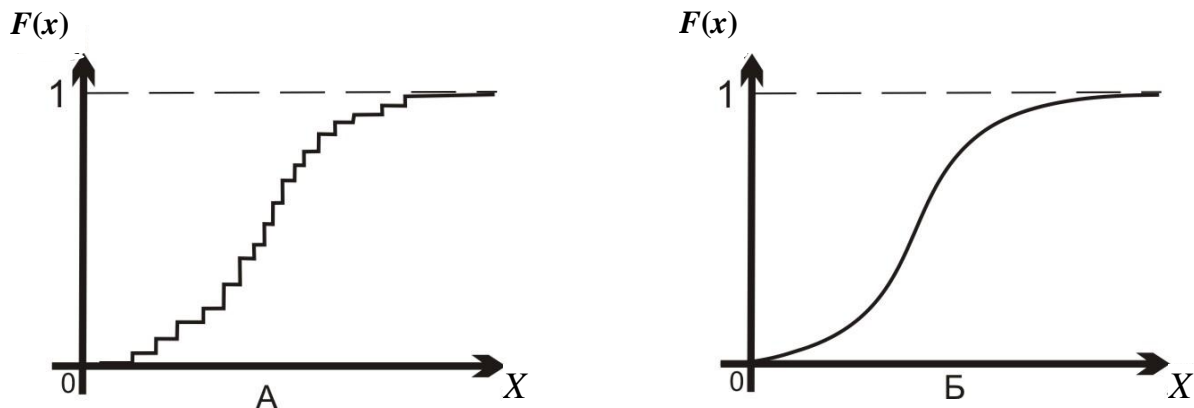


Рис. 9. Дискретная и непрерывная функции распределения

Вероятность этого события есть функция от  $x$ :

$$F(x) = P(X < x) = P(-\infty < X < x). \quad (22)$$

На рис. 8 и 9 видно, что кривые функции  $F(x)$  являются накопленными, или кумулятивными, в силу чего сама функция  $F(x)$  называется *интегральной функцией распределения*.

### 3.3. ДИФФЕРЕНЦИАЛЬНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Наряду с интегральной функцией  $F(x)$  для описания распределения непрерывных случайных величин широкое применение имеет **дифференциальная функция распределения  $f(x)$** , или **плотность вероятности**.

Рассмотрим рис. 10, на котором изображена кривая  $f(x)$ . Отложенному на оси  $x$  элементарному отрезку  $x$  соответствует вероятность  $\Delta F$  того, что наблюдаемое в результате опыта значение случайной величины будет принадлежать интервалу  $\Delta x$ .

Тогда отношение  $\frac{\Delta F}{\Delta x}$  представляет собой среднюю плотность вероятности в интервале  $\Delta x$ . Если  $\Delta x \rightarrow 0$ , то

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta F}{\Delta x} = \frac{d F(x)}{dx} = f(x), \quad (23)$$

т.е. функция  $f(x)$  является производной интегральной функции распределения  $F(x)$ . Она характеризует плотность, с которой значения

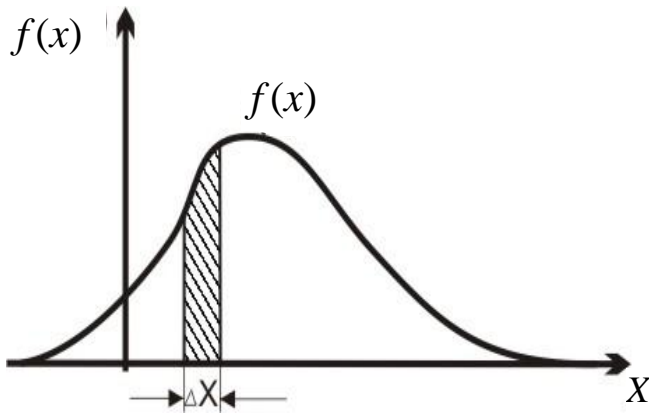


Рис. 10. Плотность вероятности

случайной величины распределяются в данном интервале. Поэтому функция  $f(x)$  называется *плотностью вероятности*.

Эмпирическим отображением кривой плотности вероятности является гистограмма распределения (рис. 11).

На этом рисунке по данным случайной выборки дано распределение непрерывной случайной величины коэффициента открытой пористости. Для этого размах выборки разделен через 0,02 на равновеликие классы. Например, на ряде залежей установлено, что группы классов по пористости соответствуют непродуктивным (при  $K_{оп} < 0,10$ ), малопродуктивным (при  $0,10 < K_{оп} < 0,16$ ) и высокопродуктивным породам (при  $K_{оп} > 0,16$ ). Следовательно, если условными знаками выделить эти группы классов на гистограмме, то соответствующие им частоты будут отражать долю площади, занятую каждой группой классов коэффициента открытой пористости от общей суммы долей площадей, равной 1.

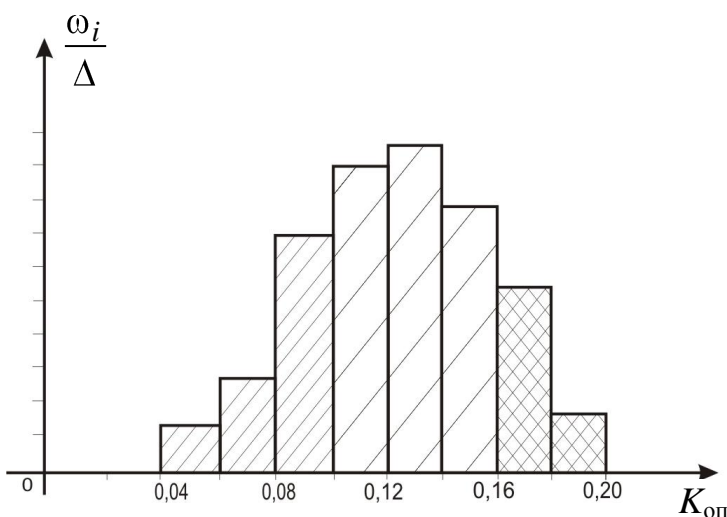
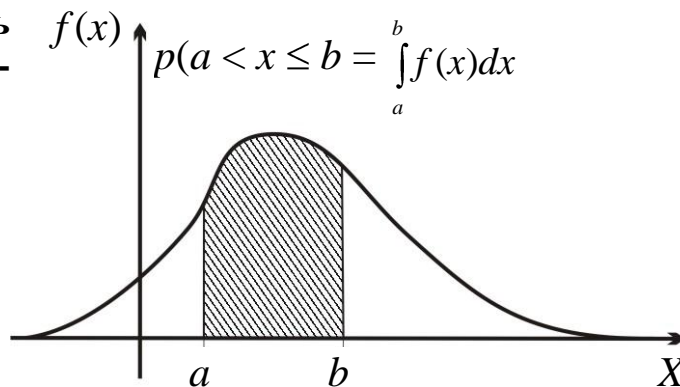


Рис. 11. Гистограмма распределения

**Рис. 12. Вероятность и площадь под графиком плотности вероятности**



На рис. 12 функция плотности вероятности отражает вероятность того, что одна из указанных выше трех групп классов примет значение открытой пористости  $x$  в интервале от  $a$  до  $b$  и будет определяться площадью, ограниченной кривой  $f(x)$  и ординатами в точках  $a$  и  $b$ .

Эта площадь определяется как интеграл  $\int_a^b f(x)dx$ . Следовательно,

$$p(a < x \leq b) = \int_a^b f(x)dx. \quad (24)$$

На основании формул (22) и (24) интегральную функцию распределения в общем виде можно выразить через плотность распределения:

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^x f(x)dx. \quad (25)$$

Этот интеграл характеризует площадь, ограниченную кривой  $f(x)$  левее координаты  $x$ .

Общая площадь, ограниченная кривой  $f(x)$ , равна 1 и ее принимают за 100 %, что соответствует полному объему выборки какого-либо параметра. В этом случае интегральная функция  $F(x)$  будет равна 1.

$$F(x) = P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1. \quad (26)$$

Естественно, гистограмма, ограничивающая площадь суммы многоугольников, равную 1, является оценкой функции плотности вероятности.

Законы распределения случайных величин в нефтегазопромысловой геологии представляют собой вероятностные модели реальных закономерностей в распределении различных параметров неоднородных продуктивных пластов. Имея дело с эмпирическими распределениями по данным случайных выборок, геолог старается подобрать к ним такую теоретическую модель, такую теоретическую функцию распределения, которая наилучшим образом описывала бы эмпирическое распределение, сглаживала влияние случайностей, связанных с недостаточным объемом выборок. Задача выравнивания сводится также к рациональному выбору таких значений параметров, при которых расхождение между теоретическим и эмпирическим распределениями оказываются минимальными. Например, расчет средних значений коллекторских свойств пласта должен основываться на характере распределения этих параметров в пределах изучаемого объекта.

Рассмотрим некоторые законы распределения дискретных и непрерывных случайных величин, наиболее часто встречающиеся в нефтегазопромысловой геологии.

### **3.4. ПРИМЕРЫ ДИСКРЕТНЫХ ТЕОРЕТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ**

Из числа известных распределений дискретных случайных величин остановимся на распределениях Пуассона и биномиальном.

#### **Распределение Пуассона**

Этот закон имеет довольно широкую область применения в статистике в случаях, связанных с числом попаданий в какую-либо плоскость или объем, если точки распределены на них более или менее равномерно и независимо друг от друга. Из практики нефтегазопромысловой геологии одним из таких случаев является распределение числа прослоев в неоднородном продуктивном пласте, вскрытом более или менее равномерной сеткой скважин.

Распределение коэффициента расчлененности не противоречит закону Пуассона, который описывается формулой

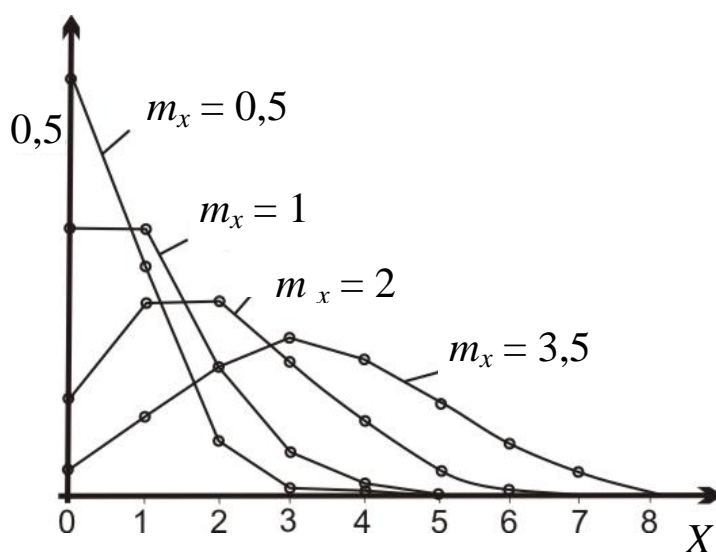
$$P(X = x_i) = \frac{m_x^{x_i}}{x_i!} e^{-m_x}, \quad (27)$$

где  $m_x$  – математическое ожидание величины  $X$ , в данном случае – средней расчлененности прослоев;  $x_i$  – число проницаемых прослоев в продуктивном пласте, пройденных скважинами,  $x_i = 1, 2, 3 \dots n$ ;  $P(X = x_i)$  – вероятность появления одного из значений  $x_i$ .

Особенностью распределения Пуассона является численное равенство математического ожидания  $m_x$  и дисперсии  $\sigma_x^2$ , вычисленных, соответственно, по формулам (1) и (9). Это свойство – важнейший критерий правдоподобия гипотезы о распределении исследуемой случайной величины по закону Пуассона. Если оценки значений математического ожидания и дисперсии близки между собой, то гипотезу правдоподобия можно принять.

Распределение Пуассона зависит от математического ожидания  $m_x$ . На рис. 13 видно, что с увеличением  $m_x$  уменьшается асимметричность распределения и оно приближается к нормальному.

Теоретические вероятности  $p$  в распределении Пуассона вычисляются по формуле (27) с учетом полученного среднего значения выборки для каждого заданного числа прослоев  $x_i$ .



**Рис. 13.** Распределение Пуассона при разных значениях математического ожидания

Например, при  $x_i = 2$ ,  $m_x = 0,692$ ,  $p = \frac{0,692^2 \cdot e^{-0,692}}{2!} = 0,120$ , где  $e^{-0,692} = 0,501$ .

Таким образом рассчитывают  $p$  для различных  $x_i$ .

### Биномиальное распределение

Этот закон распределения дискретных случайных величин описывается формулой:

$$P(X = x_i) = C_n^x p^x (1-p)^{n-x}, \quad (28)$$

где  $C_n^x$  – число сочетаний из  $n$  элементов по  $x$  и равно:

$$C_n^x = \frac{n!}{x!(n-x)!}, \quad (29)$$

где  $x = 0, 1, 2, 3, \dots, n$  – число, показывающее, сколько раз может наступить какое-либо событие при независимых друг от друга экспериментах;  $p$  – вероятность наступления события  $A$  при единичном эксперименте.

Основные характеристики биномиального распределения вычисляются по формулам: математическое ожидание при  $n$  экспериментах

$$m_x = np \quad (30)$$

и дисперсия

$$\sigma_x^2 = np(1-p). \quad (31)$$

Одним из первых применил биномиальное распределение при изучении макронеоднородности продуктивных пластов Л.Ф. Дементьев. Основываясь на фактических данных, он предложил предсчитывать вероятности попадания проектируемых скважин в зоны отсутствия или распространения коллекторов на пластах с резкой литофациальной изменчивостью [9].

**Пример.** К моменту заложения новых 5 скважин на горизонт  $D_{11}$  на Ромашкинском месторождении 22 скважины из 79 пробуренных

попали в зону отсутствия коллекторов. Необходимо было предварительно выяснить, какое наиболее вероятное число скважин из 5 проектируемых окажется в зоне отсутствия коллекторов. То есть задача сводилась к определению вероятностей попадания в зону отсутствия коллекторов числа скважин в сочетаниях 0 из 5, 1 из 5, 2 из 5...5 из 5.

Исходными данными в этом примере служат:

$$n = 5; x = 0, 1, 2, 3, 4, 5; p = 22/79 = 0,28; (1-p) = 0,72.$$

Тогда:

$$p_0 = \frac{5!}{0!(5-0)!} 0,28^0 \cdot 0,72^5 = 0,195,$$

$$p_1 = \frac{5!}{1!(5-1)!} 0,28^1 \cdot 0,72^4 = 0,376,$$

$$p_2 = \frac{5!}{2!(5-2)!} 0,28^2 \cdot 0,72^3 = 0,292,$$

$$p_3 = \frac{5!}{3!(5-3)!} 0,28^3 \cdot 0,72^2 = 0,144,$$

$$p_4 = \frac{5!}{4!(5-4)!} 0,28^4 \cdot 0,72^1 = 0,022,$$

$$p_5 = \frac{5!}{5!(5-5)!} 0,28^5 \cdot 0,72^0 = 0,001.$$

Таким образом, наиболее вероятным является попадание одной скважины в зону отсутствия коллекторов.

### 3.5. НЕПРЕРЫВНЫЕ РАСПРЕДЕЛЕНИЯ

Среди распределений непрерывных случайных величин в нефтегазопромысловой геологии наиболее часто имеют место нормальное распределение, логарифмически нормальное распределение, а также распределение Максвелла.



## Нормальный закон распределения

Этот закон занимает особое положение как наиболее часто встречающийся в практике.

Е.С. Венцель отмечает главную особенность нормального закона распределения в том, что он является предельным законом, к которому приближаются другие законы распределения при определенных условиях. Например, распределение Пуассона для дискретных случайных величин при увеличении математического ожидания  $m_x$  приближается к нормальному распределению. По данным Л.Ф. Деметьева [9], распределение Пуассона с удовлетворительной точностью может быть заменено нормальным при  $m_x \geq 9$ .

Установлено, что распределение независимых случайных величин *не противоречит* нормальному распределению, если на формирование генеральной совокупности какой-либо случайной величины оказывает влияние множество независимых друг от друга причин, и ни одна из них при этом не имеет преобладающего значения. Когда какая-нибудь одна из этого множества причин будет оказывать преобладающее влияние на формирование генеральной совокупности, может возникнуть распределение, отличающееся от нормального.

Если, к примеру, терригенные продуктивные пласты формировались в условиях мелководного бассейна, согласно залегая на нижележащих осадках, и ни одна из причин не оказывала преобладающего влияния на их формирование, то в таких случаях распределение толщин проницаемых прослоев не противоречит нормальному. Однако, если на образование продуктивного пласта местами оказывали влияние подводные русловые потоки или терригенные осадки местами заполняли неровности эрозионного рельефа, то влияние этих причин может повлечь отклонение распределения толщин от нормального.

Значения непрерывной случайной величины  $X$  распределены по нормальному закону, если функция плотности вероятности имеет вид

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{x - m_x}{2\sigma_x^2}}, \quad (32)$$

а соответствующая ей интегральная функция распределения:

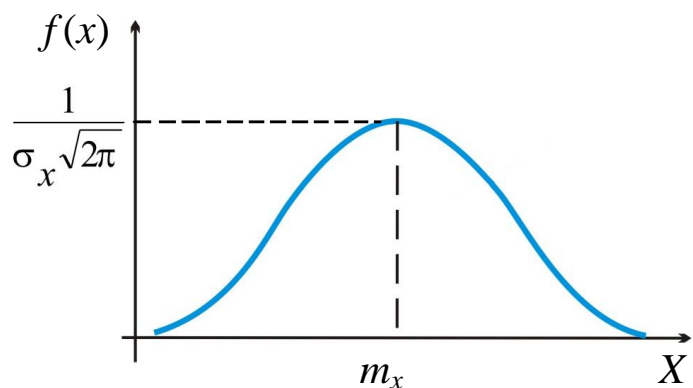
$$F(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x - m_x}{2\sigma_x^2}} \quad (33)$$

где  $m_x$  – математическое ожидание;  $\sigma_x$  – среднее квадратическое отклонение.

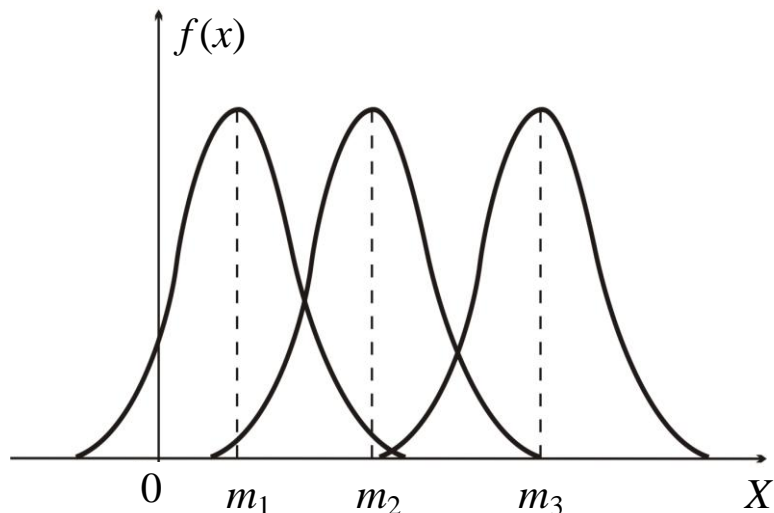
Графическим изображением функции  $f(x)$ , распределенной по нормальному закону, является рис. 14. Для нормального закона характерна симметричная кривая куполообразного вида. Ее максимальная ордината, соответствующая точке  $x = m_x$ , равна  $\frac{1}{\sigma_x \sqrt{2\pi}}$ . По мере

удаления в обе стороны от математического ожидания  $m_x$  кривая убывает, асимптотически приближаясь к оси абсцисс.

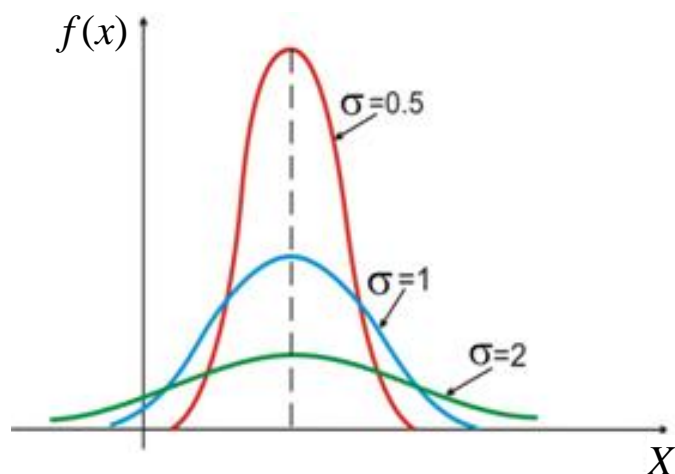
Положение центра распределения на оси абсцисс определяется величиной математического ожидания. При уменьшении  $m_x$  происходит смещение кривой влево, а при увеличении – вправо (рис.15). Параметр  $\sigma_x$  характеризует рассеяние случайной величины. В силу того, что площадь, ограниченная кривой плотности вероятности, равна 1, увеличение  $\sigma_x$  ведет к растягиванию кривой вдоль оси абсцисс и к снижению максимальной ординаты, а уменьшение  $\sigma_x$  – к сужению с боков и вытягиванию ее вверх (рис. 16).



**Рис. 14.** Характерный вид плотности вероятности нормального распределения



**Рис. 15. Влияние параметра  $m_x$  на расположение кривой плотности вероятности на относительно оси  $X$**



**Рис. 16. Влияние параметра  $\sigma$  кривизну кривой плотности вероятности**

За абсолютную погрешность при статистических вычислениях обычно принимается величина, равная оценке среднего квадратического отклонения  $S_x$ . Тогда относительная ошибка, соответствующая оценке коэффициента вариации, может быть вычислена по формуле:

$$W = \frac{S_x}{\bar{x}} = \frac{1}{\bar{x}} \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} . \quad (34)$$

Из симметричности нормального закона распределения следует, что коэффициент асимметрии  $A_s = 0$ .

Значение эксцесса, характеризующего крутость кривой плотности вероятности, равно  $0$ .

Так как непосредственное вычисление функции (32) затруднено, принято заменять переменную  $\frac{x - m_x}{\sigma_x}$  стандартизированной нормальной переменной  $t$  при  $m_t = 0$  и  $\sigma_t = 1$ . Тогда формула (32) примет вид:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \quad (35)$$

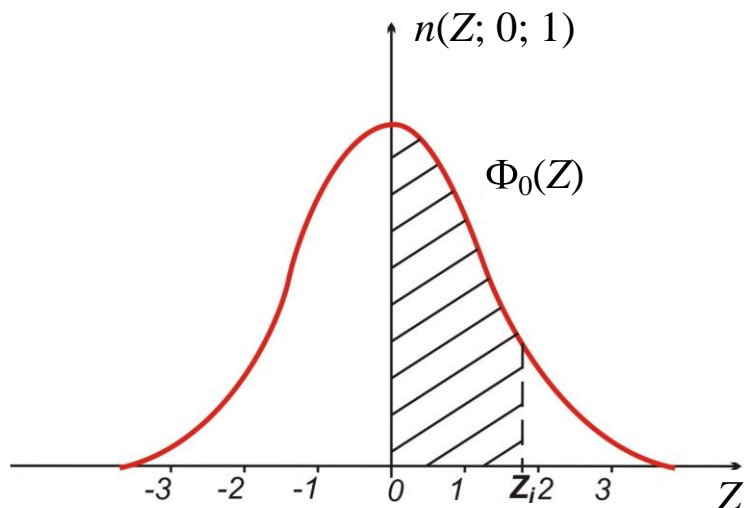
В соответствии с этой заменой переменной вероятность попадания случайной величины  $x$ , распределяющейся нормально, в определенный интервал от  $a$  до  $b$  вычисляется по формуле интегральной функции распределения:

$$F(t) = P(a < x < b) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a - m_x}{\sigma_x}}^{\frac{b - m_x}{\sigma_x}} e^{-\frac{t^2}{2}} dt. \quad (36)$$

Функцию  $F(t)$  можно упростить, если принять определенный интеграл с переменным верхним пределом вида:

$$\Phi_0(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz. \quad (37)$$

Этот интеграл определяет площадь, ограниченную кривой  $n(z; 0; 1)$  в интервале от 0 до  $z$  (рис. 17) и носит название нормированной функции Лапласа (далее функции Лапласа).



**Рис. 17. Связь нормального распределения и нормированной функции Лапласа**

Функции  $\Phi_0(z)$  присущи следующие свойства, вытекающие из рис. 17:

$$\Phi_0(-\infty) = -\frac{1}{2} \text{ и } \Phi_0(+\infty) = +\frac{1}{2}, \quad (38)$$

$$\Phi_0(-z) = -\Phi_0(z). \quad (39)$$

Выраженная через функцию Лапласа интегральная функция нормального распределения будет иметь следующий вид:

$$N(z; 0; 1) = \int_{-\infty}^0 n(z; 0; 1) dz + \int_0^z n(z; 0; 1) dz = 0,5 + \Phi_0.$$

Теперь формулу (36) можно выразить через функцию Лапласа, заменив предварительно значения  $a$  и  $b$ , соответственно, на  $x_1$  и  $x_2$ :

$$P(x_1 < x < x_2) = \Phi_0\left(\frac{x_2 - m_x}{\sigma_x}\right) - \Phi_0\left(\frac{x_1 - m_x}{\sigma_x}\right). \quad (40)$$

**Пример.** На основе статистического распределения толщин коллекторов продуктивного пласта  $i$ -ой залежи было установлено, что оно не противоречит нормальному закону распределения. Оценки параметров распределения соответственно равны  $\bar{X} = 18,04$ ;  $S_x = 5,5$ .

Необходимо определить вероятность вскрытия коллекторов толщиной в интервале от 10 до 20 м, то есть при  $x_1 = 10$  м;  $x_2 = 20$  м.

$$\begin{aligned} p(10 < x < 20) &= \Phi_0\left(\frac{20-18,04}{5,5}\right) - \Phi_0\left(\frac{10-18,04}{5,5}\right) = \\ &= \Phi_0\left(\frac{1,96}{5,5}\right) - \Phi_0\left(\frac{8,04}{5,5}\right) = \Phi_0 0,356 + \Phi_0 1,46. \end{aligned}$$

Используя табличный редактор, получим:

$$\Phi_0(0,356) = 0,139; \Phi_0(1,46) = 0,427,$$

$$p(10 < x < 20) = 0,566.$$

Следовательно, вероятность того, что при разбуривании иссле-

дуюемого пласта среди всех возможных значений толщин толщины в интервале от 10 до 20 м составят 0,566 или 56,6 %.

В табличных редакторах, например OpenOffice или Excel, для расчёта вероятности попадания нормально распределённой величины в заданный диапазон можно использовать функцию НОРМРАСП, установив у последнего параметра значение TRUE. Нормированная функция Лапласа  $\Phi_0(z)$  может быть рассчитана как  $\text{ФОШ}(z/\text{КОРЕНЬ}(2))/2$ .

### Уровни вероятности

Введем понятие об **уровнях вероятности**, под которыми принято понимать вероятность попадания случайной величины в интервал ширины  $2a$ , симметрично расположенный вокруг математического ожидания  $m_x$ .

Тогда:  $x_1 = m_x - a$ ;  $x_2 = m_x + a$ .

Учитывая замену  $\frac{x-m_x}{\sigma_x}$  на переменную  $t$  и произведя подстановку,

получим:

$$t_1 = \frac{m_x - a - m_x}{\sigma_x} = -\frac{a}{\sigma_x}; \quad t_2 = \frac{m_x + a - m_x}{\sigma_x} = \frac{a}{\sigma_x}.$$

В итоге уровень вероятности, соответствующий значению полуширины интервала  $a$ , находится по формуле

$$P(m_x - a < x < m_x + a) = 2\Phi_0(a/\sigma_x). \quad (41)$$

Значениям  $a$ , равным последовательно  $1\sigma$ ,  $2\sigma$ ,  $3\sigma$ , будут соответствовать значения  $t$ , равные 1,2 и 3. Соответствующие вероятности равны  $2\Phi_0(1)$ ;  $2\Phi_0(2)$ ;  $2\Phi_0(3)$ .

Пользуясь значениями  $\Phi_0(z)$ , получим уровни для  $\sigma$ ,  $2\sigma$  и  $3\sigma$ :

$$P_1(-\sigma < x < \sigma) = 0,683,$$

$$P_2(-2\sigma < x < 2\sigma) = 0,954,$$

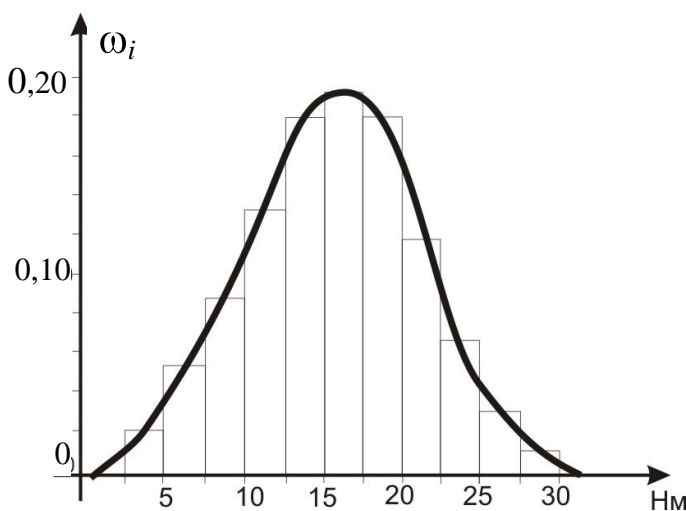
$$P_3(-3\sigma < x < 3\sigma) = 0,997.$$

Уровень вероятности 0,997 показывает, что при нормальном распределении почти все значения (99,7 %) отклоняются от  $m_x$  влево и вправо не более чем на  $3\sigma$ . Эти уровни вероятности называются **доверительными уровнями**.

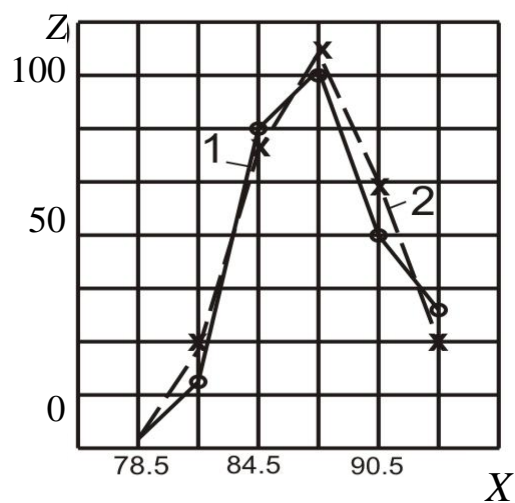
Применение машинной обработки позволяет значительно упростить процесс расчета вероятностей.

Среди параметров, характеризующих неоднородность продуктивных пластов, по нормальному закону нередко распределяются открытая пористость, эффективная и нефтенасыщенная толщины и т.п. На рис. 18 представлено распределение толщины горизонта  $D_{II}$  Туймазинского месторождения (гистограмма), которое не противоречит нормальному распределению (кривая линия). На рис. 19 приведены полигон (1) и теоретическая кривая (2) распределения нефтенасыщенности коллекторов.

Промысловому геологу в процессе подсчета запасов, при проектировании разработки залежей нефти необходимо знать, прежде чем приступить к расчетам, характер статистического распределения



**Рис. 18. Распределение толщины горизонта  $D_{II}$  Туймазинского месторождения**



**Рис. 19. Распределения нефтенасыщенности коллекторов горизонта  $D_{II}$  Туймазинского месторождения**

исследуемых параметров. В процессе подсчета запасов это важно при расчете средних значений, например, толщины и открытой пористости.

Если пласт однороден, то распределение этих параметров не противоречит нормальному закону и в качестве средней можно пользоваться даже средней арифметической величиной (для нефтенасыщенной толщины в пределах внутреннего контура). Для продуктивного пласта, характеризующегося резкой литологической неоднородностью, нередко статистическое распределение противоречит нормальному закону под влиянием различного рода причин: неравномерный вынос керна, разные условия осадконакопления в пределах изучаемой залежи, наличие нескольких типов поровых пространств в коллекторе и др.

Установив причину несоответствия статистического распределения теоретическому, например, пористости, геолог должен решить, каким видом среднего значения пользоваться при подсчете запасов.

Если аномальность статистического распределения зависит от неравномерного выноса керна, в общем объеме которого преобладают низкопоровые образцы, необходимо попытаться осуществить надежную привязку каменного материала к геофизическому и перейти на расчет средних по геофизическим данным. Одновременно нужно решить вопрос о дополнительных скважинах с отбором керна, чтобы обеспечить репрезентативность выборки.

В сложно построенных карбонатных коллекторах необходимо выделять составляющие пустотного пространства и расчет средних осуществлять для каждого типа коллекторов в пределах залежи.

Разные условия осадконакопления могут явиться причиной наличия в пределах залежей нескольких зон, резко отличающихся друг от друга толщиной, коллекторскими свойствами, нефтегазонасыщением. Нередки случаи, когда обнаруживают погребенные русла, коллекторские свойства которых также резко отличаются от смежных участков. Неравномерное заполнение неровностей эрозионного рельефа ведет к



резким локальным изменениям суммарных эффективных толщин продуктивных горизонтов. Пластовый характер карбонатных коллекторов может сочетаться с высокопроницаемыми биогермными породами.

Естественно, что в таких случаях использование средних арифметических значений по залежи в качестве параметров подсчета может привести к грубейшим ошибкам. Поэтому целесообразнее расчет средних значений и нефтенасыщенности при наличии в пределах залежи разных условий осадконакопления осуществлять путем взвешивания их значений по соответствующим объемам. Поскольку процедура эта трудоемкая, неоценимую помощь здесь может оказать применение машинной обработки.

### Логарифмически нормальное распределение

Академиком А.Н. Колмогоровым было доказано, что распределение размеров тел при дроблении подчиняются логарифмически нормальному закону. И. Лари и Д. Штефенеску установлено, что основным законом любого естественного явления, развертывающегося в массовой пропорции, является логарифмически нормальный закон.

Из коллекторских свойств пласта наиболее часто по логнормальному закону распределяется проницаемость. Этому же закону не противоречит распределение выявленных размеров запасов залежей нефти и газа в пределах региона, а также выявленные размеры поднятий, к которым они приурочены.

В логарифмически нормальных распределениях по нормальному закону распределяется не сама случайная величина  $X$ , а ее логарифм. Функция плотности вероятности логарифмически нормального распределения, соответственно, имеет следующий вид:

$$f(y) = \frac{1}{\sigma_{\lg} \sqrt{2\pi}} e^{-\frac{\lg x - M(\lg x)}{2\sigma_{\lg}^2}},$$

где  $M[\lg x]$  – математическое ожидание величины  $\lg(x)$ ,  $\sigma_{\lg}^2$  – дисперсия величины  $\lg(x)$ .

Характерно, что распределение самой непрерывной случайной величины  $X$ , значения логарифмов которой распределяются нормально, имеет ярко выраженную правостороннюю асимметрию.

Математическое ожидание и дисперсия случайной величины  $X$  определяются через  $M[\lg x]$  и  $\sigma_{\lg}$  по формулам:

$$M[X] = e^{M[\lg x] + \frac{1}{2} \sigma_{\lg}^2}, \quad (42)$$

$$D[X] = e^{2M[\lg x] + \sigma_{\lg}^2} (e^{\sigma_{\lg}^2} - 1). \quad (43)$$

Из формул (42) и (43) видна зависимость  $M[X]$  и  $D[X]$  от  $Mx[\lg x]$  и  $\sigma_{\lg}$ .

Вероятность попадания значений  $\lg x$  в интервал от  $a$  до  $b$  при логнормальном распределении случайной величины  $X$  определяется с помощью нормированной функции Лапласа:

$$\begin{aligned} F(y) &= P(\lg a < \lg x < \lg b) = \Phi_0 U_2 - \Phi_0 U_1 = \\ &= \Phi_0 \left( \frac{\lg b - M[\lg x]}{\sigma_{\lg}} \right) - \Phi_0 \left( \frac{\lg a - M[\lg x]}{\sigma_{\lg}} \right). \end{aligned} \quad (44)$$

Основные характеристики логнормального распределения рассчитываются по той же схеме, что и для нормального.

**Пример.** Расчет основных характеристик распределения 244 образцов проницаемости продуктивного пласта по способу произвольных моментов (см. табл. 3). В столбце 1 величина классового промежутка выбирается в логарифмах, которым соответствуют разряды в натуральных числах, увеличивающиеся от одного к другому в геометрической прогрессии. Так, например, классам в логарифмах 0–0,301; 0,301–0,602; 0,602–0,903; 0,903–1,204 и т.п. с классовым промежутком 0,301 соответствуют разряды в натуральных числах

1–2; 2–4; 4–8; 8–16 и т.п. Таблица для вычисления основных характеристик распределения проницаемости продуктивного пласта будет иметь следующий вид (табл. 3).

Таблица 3

Класс $k_{пр}$ ( $\Delta = 0,301$ ) в логарифмах	Класс $k_{пр}$ в натуральных числах	Частота $Z_i$	$\alpha$	$\alpha^2$	$Z_i\alpha$	$Z_i\alpha^2$
0–0,301	1–2	2	–4	16	–8	32
0,301–0,502	2–4	11	–3	9	–33	99
0,602–0,903	4–8	52	–2	4	–104	208
0,903–1,204	8–16	78	–1	1	–78	78
1,204–1,505	16–32	59	0	0	0	0
1,505–1,806	32–64	29	1	1	29	29
1,806–2,107	64–128	12	2	4	24	48
2,107–2,408	128–256	1	3	9	3	9
$n = \sum m_i = 244$					$\sum m\alpha = -167$	$\sum m\alpha^2 = 503$

$$\beta_1 = -\frac{167}{244} = -0,68; \quad \beta_2 = \frac{503}{244} = 2,06.$$

$$k_{пр} = k_{пр0} + \beta_1\Delta = 1,354 + (-0,68)0,301 = 1,149.$$

Значения  $k_{пр0} = 1,354$  и классового промежутка  $\Delta = 0,301$  приняты в логарифмах:

$$s_{y_{lg}}^2 = \Delta^2(\beta_2 - \beta_1)^2 = 2,06 - 0,462 = 0,144,$$

$$s_{y_{lg}}^2 = 0,38$$

и т.д.

Несмотря на то, что распределения таких параметров, как проницаемость (во многих случаях, но не всегда!), размеры запасов нефти и газа, площади локальных поднятий в пределах региона не противоречат логнормальному распределению, на практике, да и в печати, нередко ошибочно оперируют их средними арифметическими значениями, тогда как средними для них являются антилогарифмы средних из статистического распределения.

## Распределение Максвелла

Закон распределения Максвелла имеет функцию плотности вероятностей:

$$f(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{v^3} e^{-\frac{x^2}{2v^2}}, \quad (45)$$

где  $v$  – параметр распределения, связанный с математическим ожиданием и дисперсией случайной величины  $X$  следующими соотношениями:

$$v = 0,6267M[X]; D[X] = 0,454v^2.$$

Л.Ф. Дементьевым установлено [10], что проницаемость пласта Бавлинского месторождения, определенная по промыслово-геофизическим данным, имеет распределение, близкое к закону Максвелла при  $v \cong 0,359$ .

### 3.6. ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

В предыдущем разделе отмечалась важность определения законов распределения различных параметров, используемых в нефтегазовом промысловом геологии, для квалифицированного суждения об их средних значениях, возможности использования этих значений в последующих обобщениях и т. п. Так, при расчете показателей разработки неоднородных продуктивных пластов вычисления значительно упрощаются, если пользоваться параметрами теоретической кривой распределения случайной величины.

Во всяком статистическом распределении неизбежно присутствуют элементы случайности, зачастую связанные с тем, что наблюдений недостаточно. Нередко приходится иметь дело с проходкой скважин с ограниченным отбором керна по продуктивным пластам. Бывают случаи, когда отобранный керн полностью не обрабатывается

для определения коллекторских свойств продуктивного пласта. Отдельные части исследуемой залежи, по тем или иным причинам, могут быть не освещены бурением. Только при очень большом объеме наблюдений, равномерном по площади и разрезу, эти элементы случайности сглаживаются.

Поэтому при обработке эмпирических (статистических) распределений с ограниченным объемом выборки приходится решать вопрос, как подобрать для них теоретические кривые распределения, которые выражали бы лишь существенные черты эмпирического распределения и отбрасывали все второстепенное, связанное с недостаточным объемом фактических данных. Эта задача в статистике носит название **сглаживания**, или **выравнивания**, статистических распределений.

Однако выровненное распределение, несмотря на его внешнее сходство с эмпирическим, может не согласоваться с ним. В этом случае расхождения между распределениями будут существенными, не случайными.

Иногда расхождения между теоретическим и эмпирическим распределениями обусловлены неправильным выбором теоретической кривой. В таких случаях нужно повторить опыт с какой-либо другой кривой, описывающей иной теоретический закон.

Вопрос о согласованности эмпирического и теоретического распределений решается с помощью критериев согласия.

**Критерий согласия** – вычисляемая скалярная величина, рассчитываемая по выборке, служащая для проверки верности некоторой нулевой гипотезы  $H_0$ , против альтернативной гипотезы « $H_0$  не верна», называемой  $H_1$ .

Обычно нулевая гипотеза  $H_0$  имеет следующую форму: «наблюдаемое отклонение вызвано случайным стечением факторов», соответственно,  $H_1$ : «наблюдаемое отклонение не случайно и связано с некоторой закономерностью».

В этом разделе рассмотрим подробно статистические критерии

согласия, с помощью которых проверяется соответствие эмпирического распределения выбранному теоретическому. Однако эта задача не является единственной. В последующих разделах будет показано, что с помощью критериев согласия сравниваются также два эмпирических распределения, осуществляется сравнение средних и дисперсий.

В любом из этих случаев задача также сводится к проверке нулевой гипотезы  $H_0$ : «наблюдаемое расхождение сравниваемых распределений или характеристик вызвано случайными факторами, а сами они совпадают», т.е. отсутствует закономерное (систематическое) различие между сравниваемыми характеристиками или распределениями. Альтернативная гипотеза  $H_1$  выглядит в этой ситуации так: «наблюдаемое расхождение сравниваемых характеристик или распределений не случайно, а закономерно и связано с их различием».

### **Ошибки первого и второго рода при проверке нулевой гипотезы**

Исходя из изложенного выше, с проверкой нулевой гипотезы связано два рода ошибок.

**Ошибка первого рода:** отвергается нулевая гипотеза, в то время как она верна. Вероятность появления ошибки первого рода обозначают  $\alpha$  и называют уровнем значимости. Обычно уровень значимости принимается равным 0,05 или 0,01.

Ошибка первого рода, грубо говоря, это принятие случайного совпадения за закономерное.

**Ошибка второго рода:** принимается нулевая гипотеза, тогда как она в действительности неверна. Вероятность ошибки второго рода обозначают  $\beta$ .

Ошибка второго рода обычно связана с принятием какой-то закономерности в данных за случайное совпадение.

Для проверки нулевой гипотезы задаются уровнем значимости и по имеющейся выборке вычисляют значение критерия согласия. Если

вычисленное значение критерия при проверке нулевой гипотезы превысит допустимое теоретическое значение, соответствующее заданному уровню значимости, то нулевая гипотеза отвергается и принимается альтернатива.

Проиллюстрируем сущность ошибок первого и второго рода и соотношение между ними следующим образом (по В.Н. Бондаренко).

Пусть задана некоторая центрированная функция плотности вероятностей  $\varphi_0(t)$  случайной величины  $t$ , распределенной асимптотически нормально с математическим ожиданием  $M[t] = 0$  и дисперсией  $D[t] = 1$ . Такой величиной может быть коэффициент асимметрии  $A$ , который вместе с эксцессом нередко используется в качестве критерия согласия. Тогда  $\varphi_0(t) = \varphi_0(A)$  (рис. 18).

Следовательно, нулевая гипотеза будет иметь вид:

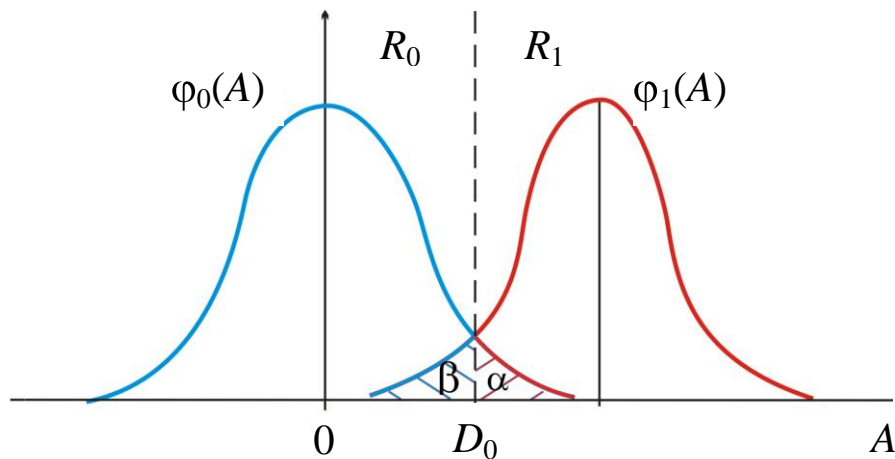
$$H_0 : M(A) = 0. \quad (46)$$

Ее альтернативой является:

$$H_1 : M(A) \neq 0. \quad (47)$$

Формула (47) характеризует двухсторонний набор альтернатив, охватывающий все возможные, отличные от нулевой гипотезы, предположения. Для простоты рассмотрим альтернативу  $H_1 : M(A) = 1$ . Тогда функцию плотности вероятностей в условиях этой альтернативы  $H$  обозначим  $\varphi_1(A)$  (рис. 20). Зная функцию плотности вероятности  $\varphi_0(A)$  в условиях нулевой гипотезы, можно на оси абсцисс указать такое значение  $A = D_0$ , чтобы появление событий, больших по величине, чем выбранное, происходило с заданной малой вероятностью  $\alpha$ . Таким образом, значение  $D_0$  устанавливается заранее. Оно соответствует заданному риску ошибочно отвергнуть верную нулевую гипотезу и тому самому уровню значимости, определяемому вероятностью появления ошибки первого рода  $\alpha$ .

Обратимся снова к рис. 20. Восстановим из точки  $D_0$  перпендикуляр к оси абсцисс, который разделит рисунок на области  $R_0$  и  $R_1$ . То-



**Рис. 20. Ошибки первого  $\alpha$  и второго  $\beta$  рода**

гда все эмпирические значения  $A < D_0$  попадут в область  $R_0$ , а  $A > D_0$  – в область  $R_1$ . В связи с этим, область  $R_0$  будет областью принятия нулевой гипотезы, а область  $R_1$  – областью отклонения нулевой гипотезы. Отсюда следует, что вероятность совершить ошибку первого рода  $\alpha$  отвечает площади, ограниченной кривой  $\varphi_0(A)$  в области  $R_1$ .

Вероятность совершить ошибку второго рода  $\beta$  будет определяться площадью, ограниченной кривой  $\varphi_1(A)$  в области  $R_0$ .

Если  $D_0$  приблизить к точке 0, то область, характеризующая ошибку первого рода  $\alpha$ , возрастет.

Тем самым увеличится уровень значимости и риск ошибочно отвергнуть верную нулевую гипотезу. Именно поэтому уровень значимости принимается минимальным и равным 0,05.

Если же принять другую альтернативу  $H_1^2$  с функцией плотности вероятностей  $\varphi_2(A)$ , сдвинутой по оси абсцисс в сторону  $D_0$  относительно первоначальной  $\varphi_1(A)$ , то возрастет ошибка второго рода  $\beta$  и риск ошибочно принять верную гипотезу, в то время как она не верна.

### 3.7. КРИТЕРИИ СОГЛАСИЯ

Для проверки соответствия эмпирического распределения тому или иному виду теоретической кривой применяются, как уже было сказано, различные скалярные величины, называемые критериями со-



гласия. Так, проверка соответствия эмпирического распределения кривой нормального и логнормального распределений может осуществляться путем совместной проверки равенства нулю коэффициента асимметрии и эксцесса.

Каждая из величин критериев согласия распределена по своему закону, но у многих критериев этот закон зависит ещё от одного или нескольких целочисленных параметров, называемых **степенями свободы**. Чтобы правильно пользоваться критериями согласия, важно знать формулы расчёта и уметь правильно рассчитывать не только саму величину критерия, но и определять количество степеней свободы.

Для проверки согласованности эмпирических распределений с нормальным и логнормальным в промысловой геологии наиболее часто применяются такие мощные критерии согласия, как критерий  $\lambda$  А.Н. Колмогорова и критерий  $\chi^2$  Пирсона.

### **Критерий $\lambda$ А.Н. Колмогорова**

Формула, по которой вычисляется критерий А.Н. Колмогорова, имеет следующий вид:

$$\lambda = D\sqrt{n}, \quad (48)$$

где  $D$  – максимальная разность между накопленной эмпирической частотой и соответствующей теоретической вероятностью  $p_i$ , которая находится из интегральной функции распределения;  $n$  – объем выборки, распределение которой сравнивается с теоретическим.

Из формулы (48) следует, что большим разностям  $D$  соответствуют большие значения  $\lambda$ . Проверка согласованности теоретического и эмпирического распределений осуществляется с помощью табл. 4, в которой возрастающим значениям  $\lambda$  соответствуют убывающие значения  $p(\lambda)$ . Распределения согласуются друг с другом, если вычис-

ленное  $\lambda_{\text{выч}}$  не превышает допустимого  $\lambda_{0,05}$  при заданном уровне значимости  $\alpha = 0,05$ .

Из табл. 4 следует, что при  $\lambda < 1,37$ , которому соответствует уровень значимости  $\alpha = 0,05$ , выборочные распределения не противоречат теоретическому.

**Пример.** Если максимальная разность  $D$  между эмпирической кумулятивной кривой и интегральной функцией распределения при изучении значений пористости составляет 0,076 при  $n = 100$ , то

$$\lambda = 0,076\sqrt{100} = 0,76$$

Так как вычисленное  $\lambda$  значение меньше теоретического при уровне значимости  $\alpha = 0,05$  ( $0,76 < 1,37$ ), то изучаемое эмпирическое распределение не противоречит выбранному теоретическому распределению.

Таблица 4

Значения  $\alpha$  при различных  $\lambda$

$\lambda$	$\alpha$	$\lambda$	$\alpha$
0,30	1,0000	1,10	0,1777
0,35	0,9997	1,20	0,1122
0,40	0,9972	1,30	0,0681
0,45	0,9874	1,37	0,0500
0,50	0,9639	1,40	0,0397
0,55	0,9228	1,50	0,0222
0,60	0,8643	1,60	0,0120
0,65	0,7920	1,70	0,0062
0,70	0,7112	1,80	0,0032
0,75	0,6272	1,90	0,0015
0,80	0,5441	2,00	0,0007
0,85	0,4653	2,10	0,0003
0,90	0,3927	2,20	0,0001
0,95	0,3276	2,30	0,0001
1,00	0,2700		

## Критерий $\chi^2$ Пирсона

Сущность проверки согласованности эмпирического и теоретического распределений с помощью критерия  $\chi^2$  Пирсона заключается в исследовании расхождения между числом наблюдений  $z_i$ , попавших в каждый класс ряда распределения, и числом наблюдений в этих же классах, вычисленным на основе теоретической функции распределения. Последнее получается путем перемножения общего числа наблюдений выборки ( $n$ ) на теоретическую вероятность ( $p_i$ ) попадания наблюдений в тот или иной класс. Мерой расхождения теоретического и эмпирического распределений служит:

$$\chi_{\text{выч}}^2 = \sum_{i=1}^r \frac{(z_i - np_i)^2}{np_i}, \quad (49)$$

где  $r$  – число классов в распределении.

Распределение  $\chi^2$  зависит от числа степеней свободы. Для случая соответствия нормальной модели число степеней свободы равно числу классов ( $r$ ) минус число независимых условий, наложенных на частоты попадания наблюдений в каждый класс. Такими независимыми условиями являются: число всех наблюдений  $n$ , число наблюдений в каждом классе  $z_i$  и сумма  $\sum_{i=1}^r \frac{z_i}{n}$ .

Если вычисленное по формуле (49) значение  $\chi_{\text{выч}}^2$  не превышает допустимого значения  $\chi^2$  для заданного уровня значимости  $\alpha = 0,05$  и  $r-3$  степеней свободы, то эмпирическое распределение не противоречит выбранной теоретической модели.

Значения  $\chi^2$  для разного числа степеней свободы  $k$  при уровне значимости  $\alpha = 0,05$  приводятся в табл. 5. В табличных редакторах имеется функция ХИ2ОБР, позволяющая вычислять теоретические значения  $\chi^2$ .

Значения  $\chi^2$  при  $\alpha = 0,05$ 

$k$	$\chi^2$	$k$	$\chi^2$	$k$	$\chi^2$
1	3,84	11	19,68	21	32,67
2	5,99	12	21,03	22	33,92
3	7,82	13	22,36	23	35,17
4	9,49	14	23,68	24	36,42
5	11,07	15	25,00	25	37,65
6	12,59	16	26,30	26	38,88
7	14,07	17	27,59	27	40,11
8	15,51	18	28,87	28	41,34
9	16,92	19	30,14	29	42,56
10	18,31	20	31,41	30	43,77

Сказанное выше имеет силу и для логарифмически нормальных распределений.

**Пример.** Проверка с помощью критерия  $\chi^2$ .

В процессе разведки продуктивного пласта, было сделано 183 определения проницаемости, которые образуют статистический ряд в логарифмической шкале согласно таблице распределения (табл. 6).

В результате расчета по способу произвольных моментов основные характеристики эмпирического распределения значений проницаемости оказались равными:

$$\lg(k_{\text{пр}}) = 1,17, \text{ или } 14,8 \cdot 10^{-3} \text{ мкм}^2;$$

$$S_{y_{\lg}} = 0,37.$$

Вместе с общим числом определений, равным 183, эти значения являются входящими параметрами для проверки согласованности эмпирического распределения теоретическому (табл. 7).

Значение вероятностей  $p_i$  в этой таблице рассчитаны по формуле нормированной функции Лапласа; число классов равно 8.

В результате расчетов значение критерия  $\chi^2$  оказалось равным 5,6. Согласно табл. 5, значение критерия  $\chi_{кр}^2$  при уровне значимости  $\alpha = 0,05$  и 5 степенях свободы равно 11,07.

Таблица 6

Номер классов	1	2	3	4	5	6	7	8	9
Границы классов в логарифмической шкале	0 0,301	0,301 0,602	0,602 0,903	0,903 1,204	1,204 1,505	1,505 1,806	1,806 2,107	2,107 2,408	2,408 2,709
Границы классов в обычной шкале, $10^{-3}$ мкм <sup>2</sup>	1–2	2–4	4–8	8–17	16–2	320– 64	64– 128	128– 256	256– 512
Число значений проницаемости.	1	9	32	54	61	17	7	1	1
Частоты, %	0,5	4,8	18,0	29,4	33,2	9,3	3,8	0,5	0,5
Накопленные частоты	0,5	5,3	23,3	52,7	85,9	95,2	99,0	99,5	100,0

Таблица 7

Номер классов	1	2	3	4	5	6	7	8
Классы в значениях логарифмов	0 0,301	0,301 0,602	0,602 0,903	0,903 1,204	1,204 1,505	1,505 1,806	1,806 2,107	2,107 2,709
Классы в значениях проницаемости, $10^{-3}$ мкм <sup>2</sup>	1–2	2–4	4–8	8–17	16–32	32–64	64–128	128– 512
$z_i$	1	9	32	54	61	17	7	2
$p_i$	0,0086	0,0526	0,1732	0,3023	0,2793	0,1339	0,0434	0,056
$np_i$	1,6	9,6	31,7	55,4	51	24,5	7,9	1,1
$z_i - np_i$	–0,6	–0,6	+0,3	–1,4	+10,0	–7,5	–0,9	+0,9
$(z_i - np_i)^2$	0,36	0,36	0,09	1,96	100	56,2	0,81	0,81
$\frac{(z_i - np_i)^2}{np_i}$	0,22	0,04	0	0,04	1,96	2,30	0,10	0,74

$$\chi^2 = \sum_{i=1}^{r=8} \frac{(z_i - np_i)^2}{np_i} = 5,6.$$

Так как  $(\chi^2 = 5,6) < (\chi_{кр}^2 = 11,7)$  при уровне значимости  $\alpha = 0,05$  и  $k = 5$ , то на основании анализа можно считать, что исследуемое распределение проницаемости продуктивного пласта не противоречит логарифмически нормальному распределению.

### 3.8. ОЦЕНКИ ПАРАМЕТРОВ И ИХ КАЧЕСТВО

Ранее неоднократно отмечалось, что генеральная совокупность оценивается по данным выборочной совокупности. Чем больше объем последней при соблюдении равномерности точек наблюдений (скважин) по площади и результатов анализов внутри скважин, тем точнее выборочная совокупность описывает, или оценивает, генеральную. Однако нередко геологу приходится, по целому ряду причин, иметь дело с незначительным объемом данных, которых явно недостаточно, чтобы установить заранее неизвестный закон распределения случайной величин.

На основе ограниченного объема данных некоторые сведения о случайной величине могут быть получены, если хотя бы приближенно оценить основные характеристики случайной величины – математическое ожидание и дисперсию. Нужно учесть, что любые, даже приближенные, значения параметров будут содержать в себе элементы случайности, в силу чего их принято называть **оценками параметров**. Поэтому в процессе рассмотрения основных характеристик распределения были приведены формулы, по которым рассчитываются их оценки. Например, оценкой математического ожидания является среднее арифметическое значение.

В силу того, что любая из оценок случайна, применение ее ведет к ошибке. Эта ошибка тем ощутимее, чем меньше число наблюдений. Естественно, оценки должны удовлетворять ряду требований, определяющих их качество.

Во-первых, оценки должны быть **состоятельными**. Для этого необходимо, чтобы с увеличением числа наблюдений оценки прибли-

жались к математическому ожиданию. Так, среднее арифметическое значение является состоятельной оценкой, поскольку с увеличением числа наблюдений соблюдается это условие. Состоятельной будет и оценка  $S_x^2$  дисперсии  $D$ .

Во-вторых, оценки должны быть **несмещенными**. Несмещенность оценки определяется отсутствием систематической ошибки в сторону завышения или занижения. Этому требованию соответствует следующее условие: математическое ожидание оценки  $\hat{a}$  какого-либо параметра  $a$  должно быть равно этому параметру:  $M(\hat{a}) = a$ .

Например, среднее арифметическое значение пористости является несмещенной и состоятельной оценкой этого параметра.

Иное дело дисперсия. Наиболее естественной ее оценкой пред-

ставляется  $S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ . Математическое ожидание оценки дисперсии определяется по формуле:

$$M(S_x^2) = \frac{n-1}{n} D_x, \quad (50)$$

откуда следует, что  $S_x^2$  не будет несмещенной оценкой дисперсии [5].

Дробь  $\frac{n-1}{n}$  указывает на то, что оценка дисперсии всегда меньше самой дисперсии, причем величина систематической ошибки тем больше, чем меньше объем наблюдений. Для устранения этой ошибки достаточно  $S_x^2$  умножить на  $\frac{n-1}{n}$ . Тогда несмещенная оценка:

$$\begin{aligned} S_x^2 &= \frac{n}{n-1} \times \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \\ &= \frac{1}{n(n-1)} \left( n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right), \end{aligned}$$

что и соответствует формуле (11).

При большом числе наблюдений  $n$  разницей между смещенной и несмещенной оценками дисперсии можно пренебречь.

В-третьих, выбранная несмещенная оценка должна обладать наименьшей дисперсией или наиболее быстро сходиться к истинному значению по сравнению с другими, то есть быть **эффективной**. Это зависит от вида закона распределения случайной величины.

### 3.9. ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ И ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

При небольшом числе наблюдений, когда математическое ожидание и дисперсия неизвестны, можно пользоваться их точечными оценками. Так как эти оценки носят случайный характер, замена ими неизвестных параметров может сопровождаться ошибками. Кроме того, точечная оценка параметров без указания точности и надежности носит неопределенный характер. Для того, чтобы получить представление о точности и надежности несмещенной оценки  $\hat{h}$  параметра  $h$  (например, эффективной толщины продуктивного пласта), пользуются **доверительными интервалами** и **доверительными вероятностями**. Сущность доверительного интервала сводится к следующему.

Если мы хотим оценить возможную ошибку оценки  $\hat{h}$  относительно истинного значения параметра  $h$ , зададимся достаточно большой вероятностью  $1-\alpha$ , позволяющей считать, что событие с такой вероятностью будет практически достоверным.

При этом диапазон практически возможных ошибок от замены истинного параметра  $h$  на несмещенную оценку  $\hat{h}$  будет  $\pm\delta$ . Тогда:

$$p(\hat{h}-\delta < h < \hat{h} + \delta) = 1-\alpha. \quad (51)$$

Из этой формулы видно, что ошибки, большие по абсолютной величине, чем  $\pm\delta$ , будут возможны только с весьма малой вероятностью



$\alpha = 0,05$ , а истинное значение параметра  $h$  будет заключено в интервал  $(\hat{h}-\delta; \hat{h} + \delta)$ . В связи с тем, что оценка  $\hat{h}$  случайна, ошибка  $\delta$  и сам интервал, равный  $2\delta$ , будут носить случайный характер. В силу этого вероятность  $1-\alpha$  определяет вероятность того, что случайный интервал  $(\hat{h}-\delta; \hat{h} + \delta)$  **накроет** точку, соответствующую положению истинного значения  $h$ .

Вероятность  $1-\alpha$  принято называть доверительной вероятностью, а интервал  $(\hat{h}-\delta; \hat{h} + \delta)$  – доверительным интервалом.

Рассмотрим наиболее часто встречающийся в практике случай построения доверительного интервала для центра распределения при неизвестной дисперсии. В этом случае математическое ожидание  $m_x$  и среднее квадратическое отклонение  $\sigma_x$  неизвестны и оцениваются по данным выборочной совокупности.

Задача сводится к отысканию доверительного интервала, отвечающего доверительной вероятности  $p = 1-\alpha$ , где, например,  $\alpha = 0,05$ .

При этом пользуются величиной  $t = \sqrt{n-1} \left( \frac{\hat{h} - h}{S_x} \right)$ , распределяющейся по закону Стьюдента с  $n-1$  степенями свободы. Эту величину называют критерием или статистикой Стьюдента.

В табл. 8 приведены зависимости критерия Стьюдента от числа степеней свободы при  $\alpha = 0,05$ . В табличных редакторах для нахождения этих величин следует использовать функцию СТЬЮДРАСПОБР. Пользуясь табл. 8 (или табличным редактором), можно найти соответствующие пределы  $\pm t_{0,05, n-1}$ , при которых и любом  $h$  вероятность  $p$  будет определяться выражением:

$$p \left\{ \hat{h} - t_{\alpha, n-1} \cdot \frac{S_x}{\sqrt{n-1}} < h < \hat{h} + t_{\alpha, n-1} \cdot \frac{S_x}{\sqrt{n-1}} \right\} = 1 - \alpha, \quad (52)$$

где  $S_x$  – оценка среднего квадратического отклонения,  $\hat{h}$  – оценка математического ожидания  $h$ .

Значения  $t$  для доверительной вероятности  $p = 1 - \alpha = 0,95$ 

$k = n-1$	$t_{0,05}$	$k = n-1$	$t_{0,05}$
1	12,706	12	2,179
2	4,303	14	2,145
3	3,182	16	2,120
4	2,776	18	2,101
5	2,571	20	2,086
6	2,447	22	2,074
7	2,365	24	2,064
8	2,306	26	2,056
9	2,262	28	2,048
10	2,228	30	2,042
		$\infty$	<b>1,960</b>

Отсюда доверительный интервал будет равен:

$$\left( \hat{h} - t_{\alpha, n-1} \cdot \frac{S_x}{\sqrt{n-1}}; \hat{h} + t_{\alpha, n-1} \cdot \frac{S_x}{\sqrt{n-1}} \right). \quad (53)$$

Если сравнить формулы (51) и (52), то можно увидеть, что  $\delta = t_{\alpha, n-1} \cdot \frac{S}{\sqrt{n-1}}$ . В интервале  $(\hat{h} - \delta; \hat{h} + \delta)$  вероятность будет равна  $1 - \alpha$  при любом  $h$ .

**Пример.** Приведем пример построения доверительного интервала.

По 10 скважинам среднее арифметическое эффективной толщины пласта оказалась равной 7,41 м. Оценка среднего квадратического отклонения  $S = 1,2$  м. Построить доверительный интервал для истинного значения эффективной толщины  $h$  при доверительной вероятности  $p = 1 - \alpha$  ( $\alpha = 0,05$ ).

По табл. 8 при  $k = 10 - 1 = 9$  степенях свободы  $t_{0,05} = 2,262$ .

Отсюда:

$$\delta = 2,262 \frac{1,2}{\sqrt{10-1}} = \frac{2,262 \cdot 1,2}{3} = 0,905.$$

С учетом этих данных доверительный интервал будет равен:

$$7,41 - 0,905 < h < 7,41 + 0,905,$$

$$6,505 < h < 8,315.$$

То есть истинное значение толщины будет иметь величину, большую 6,505 и меньшую 8,315.

Нетрудно заметить, что с увеличением числа наблюдений ошибка будет уменьшаться.

Изучение неоднородности продуктивных пластов с помощью мер теории информации

### **3.10. ИЗУЧЕНИЕ НЕОДНОРОДНОСТИ ПРОДУКТИВНЫХ ПЛАСТОВ С ПОМОЩЬЮ МЕР ТЕОРИИ ИНФОРМАЦИИ**

#### **Неоднородность продуктивных пластов как неопределённость системы**

Наряду с изучением неоднородности продуктивных пластов описанными выше методами математической статистики и теории вероятностей, в последнее время получили распространение методы изучения изменчивости неоднородных продуктивных пластов, основанные на использовании понятий и мер теории информации.

Последняя, являясь разделом теории вероятностей, дает возможность оперировать такими характеристиками статистических моделей, интерпретация которых не зависит от типа моделей. Как было показано выше, тип модели в полной мере описывается законом распределения случайной величины. С помощью информационных методов нельзя более полно описать исследуемый объект, чем это можно сделать с помощью теоретической функции распределения. Однако информационные методы значительно упрощают технику расчетов в процессе исследования, а также толкование полученных результатов.

Любое сообщение, с которым имеет дело теория информации, представляет собой совокупность сведений о некоторой физической системе. Такой физической системой можно представить и неоднородный продуктивный пласт, у которого наблюдаемые данные различных параметров характеризуют совокупность сведений о нем.

Одними из первых применили информационные методы для изучения неоднородности продуктивных пластов Л.Ф. Деменьтев, Е.А. Хитров и Ю.В. Шурубор [10].

Теория информации рассматривает физическую систему  $X$ , которой заранее присуща какая-то степень неопределенности, обусловленная случайностью оказаться в том или ином состоянии. Если физическую систему – продуктивный пласт – рассматривать как генеральную совокупность, то сведения по выборочной совокупности, какой является сеть пробуренных скважин, можно трактовать как случайное сообщение о генеральной совокупности, которое заранее не было известно. В этом смысле степень неопределенности будет иметь место и для продуктивного пласта. Чем сильнее изменчивость коллекторских свойств пласта по площади и разрезу, чем резче изменяется толщина, т.е. чем сильнее неоднородность, тем более неопределенными будут для геолога результаты в определении значений параметров пласта как по отдельным скважинам, так и по залежи в целом. В этом смысле степень неопределенности применительно к естествознанию, как показал У. Эшби, можно трактовать и как степень сложности, разнообразия системы.

Предположим, что продуктивный пласт повсеместно представлен двумя проницаемыми прослоями, а в некоторых местах залежи появляются еще два прослоя. Граница замещения каждого из этих прослоев неизвестна. Другой пласт содержит два проницаемых прослоя, причем границы распространения второго прослоя также неясны. Даже из простого сравнения двух пластов видно, что неопределенность первого пласта выше, чем второго, так как вскрывшие первый пласт скважины будут иметь значительно больше возможных состоя-

ний. Наряду с различным количеством возможных состояний, степень неопределенности обоих пластов будет отличаться и вероятностью обнаружения в каждой скважине числа проницаемых прослоев. Из этого примера видно, что степень неопределенности физической системы, под которой мы понимаем продуктивный пласт, определяется числом и вероятностями возможных состояний.

Для описания степени неопределенности системы важны только количество значений и их вероятность.

Следует учесть, что по каждому из пластов сумма вероятностей вскрытия скважинами разного числа прослоев должна быть равна 1, то есть  $\sum_{i=1}^n p_i = 1$ . Так, по первому пласту возможны три состояния системы:  $x_1 = 2$ ,  $x_2 = 3$ ,  $x_3 = 4$ . Им соответствуют вероятности  $p_1 = 0,25$ ,  $p_2 = 0,45$ ;  $p_3 = 0,30$ , сумма которых равна 1. Второй пласт будет иметь два состояниях  $x_1 = 1$  и  $x_2 = 2$  со своими вероятностями  $p_1 = 0,45$  и  $p_2 = 0,55$ , сумма которых также равна 1.

Если эти данные свести в таблицы, то они практически не будет отличаться от таблиц ряда распределения дискретной случайной величины (числа прослоев), с которой у физической системы  $X$  с конечным множеством состояний много общего:

I пласт	$x_i$	2	3	4
	$p_i$	0,25	0,45	0,30

II пласт	$x_i$	1	2
	$p_i$	0,45	0,55

Чтобы физическую систему свести к дискретной величине, достаточно приписать каждому состоянию какое-то числовое значение. Например, номер состояния приписать каждому состоянию, что и иллюстрируется двумя последними табличками. Отсюда видна справедливость трактовки степени неопределенности как степени ее сложности и разнообразия, а в данном случае – как степени макронеоднородности продуктивного пласта.

### 3.11. ЭНТРОПИЯ И ЕЁ СВОЙСТВА

В качестве меры неопределенности системы (или дискретной случайной величины) служит специальная характеристика, которую Шеннон назвал **энтропией**.

$$H[X] = -\sum_{i=1}^n p_i \log(p_i). \quad (54)$$

Энтропия характеризуется целым рядом свойств. Рассмотрим их.

**Свойство 1.** Если одно из состояний достоверно, то есть его вероятность  $p = 1$ , а другие состояния невозможны, то есть их вероятности равны 0, то энтропия такой системы равна 0:

$$H[X] = -\sum_{i=1}^n p_i \log(p_i) = -1 \log(1) = 0. \quad (55)$$

Например, однородный, нерасчлененный на прослой продуктивный пласт, макронеоднородность которого по расчлененности равна 0, будет характеризоваться нулевой энтропией.

На величине энтропии не отразится включение в число возможных состояний любого их количества с вероятностью, равной 0, так как в формулу (54) они войдут в качестве слагаемых, равных 0.

**Свойство 2.** Энтропия системы с равновероятными состояниями, вероятности которых равны  $p_1 = p_2 = p_3 \dots = p_n = \frac{1}{n}$  обращается в максимум и равна логарифму числа состояний.

$$H = -\sum_{i=1}^n \frac{1}{n} \log\left(\frac{1}{n}\right) = -n \frac{1}{n} \log\left(\frac{1}{n}\right) = -(\log(1) - \log(n)) = \log(n)$$

или

$$H = \log(n). \quad (56)$$

Отсюда следует, что энтропия будет увеличиваться с возрастанием числа возможных состояний. И действительно, чем больше содержится в пласте проницаемых прослоев, равновероятных во всех

скважинах, тем выше его расчлененность как мера степени неопределенности, разнообразия и неоднородности пласта – энтропия.

Отметим, если некоторая величина задана непрерывным распределением, то среди всех распределений, обладающих заданной дисперсией, максимальной энтропии соответствует нормальное распределение.

**Свойство 3.** Энтропия обладает аддитивностью. Это означает, что если несколько независимых равновероятных физических систем объединяются в одну, то их энтропии складываются.

Если система  $X$  имеет  $n$  возможных состояний  $(X_1, X_2, \dots, X_n)$ , а система  $Y$  имеет  $m$  возможных состояний  $(Y_1, Y_2, \dots, Y_m)$ , то число возможных состояний объединенной системы  $(X, Y)$  равно  $nm$ , так как она представляет все возможные комбинации состояний  $(X_i, Y_j)$  систем  $X$  и  $Y$ . Если каждая из этих систем принимает свои состояния независимо одна от другой, то вероятности их состояния можно выразить через  $P(X)$  и  $P(Y)$ . По теореме умножения вероятностей для независимых событий вероятность сложной системы  $P(X, Y)$  равна произведению вероятностей каждого независимого события, т.е.:

$$P(X, Y) = P(X) \cdot P(Y). \quad (57)$$

Прологарифмировав правую и левую части этой формулы, получим:

$$\log(P(X, Y)) = \log(P(X)) + \log(P(Y)). \quad (58)$$

Энтропию сложной системы можно представить в форме математического ожидания:

$$H(X, Y) = M[-\log(P(X, Y))]. \quad (59)$$

На основании (62) и (63) можно записать:

$$H(X, Y) = M[-\log(P(X)) - \log(P(Y))]$$

или

$$H(X, Y) = H(X) + H(Y). \quad (60)$$

Если продуктивный горизонт содержит два пласта, каждый из которых расчленен на несколько проницаемых прослоев, то энтропия этого продуктивного горизонта будет определяться суммой энтропий каждого пласта в отдельности.

При вычислении энтропии в качестве основания логарифмов принимают числа 2, 10 и основание натуральных логарифмов  $e$ . В соответствии с этим единицы измерения энтропии именуют **бит**, **хартли** и **нит**. Между ними наблюдаются следующие соотношения:

$$1 \text{ бит} \cong 0,693 \text{ нит} \cong 0,301 \text{ хартли};$$

$$1 \text{ нит} \cong 1,443 \text{ бит} \cong 0,434 \text{ хартли};$$

$$1 \text{ хартли} \cong 3,222 \text{ бит} \cong 2,302 \text{ нит}.$$

Л.Ф. Дементьев, Е.А. Хитров и К.А. Шурубор рекомендуют пользоваться единицей энтропии нит. В связи с этим, формула для вычисления энтропии примет следующий вид:

$$H(X) = -\sum_{i=1}^n p_i \ln(p_i). \quad (61)$$

Для облегчения расчетов в математических руководствах имеются таблицы значений функций  $-p \ln(p)$  и  $-p \log_2(p)$ , при пользовании которыми вычисление энтропии сводится к простому суммированию табличных значений. При использовании табличных редакторов эти величины вычисляются напрямую.

### Энтропия в нефтегазопромысловой геологии

Как показали Е.А. Хитров и И.П. Чоловский [15], с помощью энтропии можно изучать макронеоднородность продуктивных пластов по расчлененности не только по данным о вероятностях вскрытия разного числа прослоев в пласте. Если вместо вероятностей учитывать долю средней эффективной толщины каждого прослоя в средней эффективной толщине продуктивного пласта  $\frac{h_{\text{инр}}}{h_{\text{пл}}}$ , то никаких изме-



нений в принципиальной схеме расчета энтропии не произойдет. Расчет энтропии, в соответствии с такой заменой, будет осуществляться по формуле:

$$H = -\sum_{i=1}^n \frac{h_{i\text{пр}}}{h_{\text{пл}}} \ln \frac{h_{i\text{пр}}}{h_{\text{пл}}}, \quad (62)$$

где  $n$  – число прослоев в пласте.

В соответствии с расчетом  $\frac{h_{i\text{пр}}}{h_{\text{пл}}}$  будет соблюдаться условие:

$$\sum_{i=1}^n \frac{h_{i\text{пр}}}{h_{\text{пл}}} = 1.$$

Предполагается, что при равномерном расположении достаточно большого числа скважин должно выполняться равенство:

$$-\sum_{i=1}^n \frac{h_{i\text{пр}}}{h_{\text{пл}}} \ln \frac{h_{i\text{пр}}}{h_{\text{пл}}} \cong -\sum p_i \ln(p_i). \quad (63)$$

Это обусловит соблюдение и других свойств энтропии. Так, если продуктивный пласт не расчленен, т.е. однороден, то энтропия его в соответствии с формулой (55) равна 0.

Энтропия пласта, расчлененного на несколько проницаемых прослоев одинаковой толщины так же, как и по формуле (56), будет обращаться в максимум. На размер энтропии в этом случае не будет влиять величина толщины прослоев. Например, два пласта, содержащих по 5 прослоев толщиной по 1 м в первом пласте и по 2 м во втором, будут характеризоваться одинаковой энтропией  $H = 1,6095$  нит, так как в обоих случаях доля каждого прослоя в общей толщине будет составлять 0,2, чему соответствует  $-p_i \ln p_i = 0,3219$  нит. Вместе с тем, в соответствии с формулой (56) энтропия в обоих случаях равна  $\ln n = 1,6091$  нит.

Следует еще раз обратить внимание, что в соответствии с формулой (56) энтропия будет максимальной, а при увеличении числа равновеликих прослоев  $n$  она возрастет.

При содержании в пласте нескольких разновеликих по толщине

прослоев энтропия изменяется в зависимости от степени неоднородности и разнообразия системы, однако она во всех случаях будет меньше энтропии системы, содержащей равновеликие по толщине прослои. Например, при общей средней эффективной толщине продуктивного пласта 10 м средние эффективные толщины составляющих его отдельных прослоев равны 0,3 м, 0,7 м, 2 м, 3 м, 4 м. Соответственно,  $\frac{h_{инр}}{h_{пл}}$  будут равны 0,03; 0,07; 0,2; 0,3; 0,4. Энтропия такого пласта определится следующей суммой:

$$H = 0,1052 + 0,1861 + 0,3219 + 0,3612 + 0,3665 = 1,3409,$$

что меньше энтропии пласта, состоящего из 5 равновеликих прослоев ( $1,3409 < 1,6095$ ).

Как будет показано в следующем разделе, этой особенностью энтропии начинают широко пользоваться при сравнении макронеоднородности продуктивных пластов, содержащих в себе прослои, толщины которых изменяются в различных соотношениях относительно общей эффективной толщины пласта. В процессе сравнения двух или нескольких пластов энтропия может дать ответ на вопрос, при каких соотношениях толщин (при соблюдении прочих равных условий) разработка пласта будет эффективнее.

Энтропия является не только мерой неоднородности, но и мерой информативности имеющихся у нас данных, в частности, энтропия выраженная в битах показывает, какой объём информации по данному параметру имеется. Как отмечали Л.Ф. Дементьев, Е.А. Харитонов и Ю.В. Шуробур этот факт может быть использован для оценки достаточности имеющегося объёма наблюдений.

При изучении геологических объектов всегда есть дилемма между проведением дополнительных исследований и получением дополнительных данных об объекте и ценой такого исследования. Наряду с другими методами оценить разумный объём исследований можно, используя энтропию. Изучая некоторый параметр  $X$ , по эксперимен-

тальным выборкам объёма  $n_1, n_2, n_3, \dots$  можно построить график зависимости энтропии экспериментальной выборки от её объёма. С того момента, как этот график стабилизируется, т.е. энтропия по большей выборке будет незначительно отличаться от энтропии по меньшей, дальнейшего смысла в увеличении объёма выборки нет. Действительно энтропия, как было сказано, есть мера информации, полученной в выборке о параметре; если увеличение выборки не увеличивает энтропию, то значит мы не получаем дополнительной информации об объекте, проводя новые исследования.

#### **4. СРАВНЕНИЕ ДВУХ ИЛИ НЕСКОЛЬКИХ НЕОДНОРОДНЫХ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ**

Задача сравнения двух или нескольких геологических объектов, характеризующихся макро- и микронеоднородностью, в нефтегазопромысловой геологии приобретает крайне важное значение. На значительном числе выявленных и разрабатываемых месторождений продуктивная толща содержит два или более продуктивных горизонта, каждый из которых может вмещать самостоятельную залежь. Нередки случаи, когда продуктивная толща состоит из нескольких пластов, вмещающих единую залежь.

Когда залежь связана с несколькими продуктивными пластами, то естественно возникает вопрос, столь ли существенны различия между их коллекторскими свойствами, чтобы запасы считать отдельно по каждому пласту.

При сходстве геологического строения нескольких объектов бывает целесообразно объединять их для исследования, например, нижних пределов коллекторских свойств, исходных данных для которых по каждому объекту в отдельности зачастую бывает недостаточно. Задача в этом случае состоит в выборе объектов, которые можно вовлечь в такой комплексный анализ и которые нужно исключить, так как отличие их параметров от параметров остальных объектов носит неслучайный характер.

К сожалению, до последнего времени решение этих вопросов с помощью методов математической статистики не приняло массового характера и опирается, в основном, на интуицию исследователей, что нередко приводит к ошибкам. При решении этих вопросов методы математической статистики, базирующиеся на сравнении характеристик, имеющих числовое выражение, позволяют оценить, существенны ли различия в величинах изучаемого признака или они обусловле-

ны случайными причинами. Сравнение двух или нескольких геологических объектов может осуществляться путем сравнения рядов распределения, средних значений дисперсий, сравнения энтропии и других показателей.

Рассмотрим статистические методы сравнения двух и более геологических объектов, моделями которых являются одномерные компоненты.

#### **4.1. СРАВНЕНИЕ РЯДОВ РАСПРЕДЕЛЕНИЯ ЗНАЧЕНИЙ ПАРАМЕТРОВ ПРОДУКТИВНЫХ ПЛАСТОВ**

Этот метод сравнения является наиболее полным, поскольку охватывает не отдельные основные характеристики распределения, а весь ряд наблюдаемых значений исследуемого параметра.

С помощью такого анализа можно сравнивать распределение пористости, толщины, проницаемости двух или нескольких продуктивных пластов или одного пласта, но в разных частях залежи. Кроме того, нередко возникает задача сравнить значения пористости по одному пласту, определенные по керну, со значениями этого параметра, полученными по геофизическим данным. Иногда геологи, по-разному интерпретируя строение залежи, дают несколько вариантов карт изопахит. Сравнивая эти карты с помощью рядов распределения, можно установить, существенны ли различия между ними или они вызваны случайными причинами. При несущественных различиях разные варианты построения карт изопахит не окажут влияния на результат подсчета запасов. При существенных – нужны дополнительные исследования с целью доказательства правомерности одного из вариантов построения.

Наиболее простым способом сравнения рядов распределения является сравнение с помощью критерия согласия  $\lambda$  А.Н. Колмогорова, который в данном случае определяется по формуле:

$$\lambda = D_{\max} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \quad (64)$$

где  $D_{\max}$  – максимальная разность в каком-либо классе между накопленными частотами сравниваемых рядов распределения;  $n_1$  и  $n_2$  – объемы сравниваемых выборочных совокупностей.

Для расчетов по этой формуле следует пользоваться табл. 5 значений  $\lambda$ .

**Пример.** Два продуктивных пласта в пределах одной залежи охарактеризованы керном, на основании которого сделано 49 определений открытой пористости по одному пласту и 81 определение – по другому. По этим значениям построены ряды распределения, в которых максимальная разность между накопленными частотами, равная 0,15, установлена в классе со значениями от 16 до 20 %. Необходимо путем сравнения рядов распределений установить, существенны ли различия между исследуемыми пластами с целью вовлечения данных по ним в анализ зависимости между удельной продуктивностью скважин и  $\alpha_{\text{сп}}$ .

В соответствии с исходными данными будем иметь:

$$\lambda_{\text{выч}} = 0,15 \sqrt{\frac{49 \cdot 81}{49 + 81}} = \frac{0,15 \cdot 7 \cdot 9}{11,4} = 0,83.$$

Полученное значение  $\lambda_{\text{выч}} = 0,83$  значительно меньше допустимого 1,37 при заданном уровне значимости  $\alpha_{\text{сп}} = 0,05$ . Следовательно, различия между исследуемыми распределениями несущественны, и данные о пористости могут быть вовлечены в совместный анализ по залежи.

Сравнение рядов распределения значений параметра применимо при большом объеме наблюдений в выборочных совокупностях. Однако промысловому геологу нередко приходится оперировать малыми выборками с числом наблюдений, не превышающим 15–20.

Подобная ситуация может возникнуть при подсчете запасов нефти

и газа по небольшим залежам, пройденным ограниченным числом скважин. При этом предполагается, что формирование исследуемых объектов происходило в сходных геологических условиях. Если, например, геолог сможет доказать, что коллекторские свойства рассматриваемых залежей несущественно отличаются друг от друга, то объединение наблюдаемых значений в одну совокупность позволит более объективно оценить значение параметров подсчета. Также можно установить кондиционные пределы коллекторских свойств, вовлекая для этого в анализ все данные об удельной продуктивности скважин, число которых по каждой залежи зачастую бывает ограничено.

В подобных случаях сравнение двух или более неоднородных продуктивных объектов можно осуществлять путем сравнения средних и дисперсий.

#### 4.2. СРАВНЕНИЕ СРЕДНИХ ЗНАЧЕНИЙ ПАРАМЕТРОВ

Рассмотрим сначала процедуру сравнения средних двух продуктивных объектов, значения параметров которых, например, открытой пористости, распределяются нормально.

Сравнение средних значений открытой пористости двух разных продуктивных объектов осуществляется проверкой нулевой гипотезы  $H_0 : M(K_{оп1}) = M(K_{оп2})$  по эмпирическим данным, где  $M(K_{оп1})$  и  $M(K_{оп2})$  – математические ожидания значений пористости из двух выборок совокупностей. Естественно, истинные значения средних исследователю не известны.

Проверка гипотезы о равенстве двух неизвестных средних осуществляется с помощью критерия Стьюдента, сущность которого заключается в следующем. Если исследуемые случайные величины  $X_1$  и  $X_2$  двух объектов распределены нормально, то в случае примерного равенства средних, величина распределена по закону Стьюдента с  $(n_1 + n_2 - 2)$  степенями свободы:

$$t = \frac{X_1 - X_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (65)$$

где  $X_1$  и  $X_2$  – средние значения открытой пористости в исследуемых выборках;  $S_1^2$  и  $S_2^2$  – несмещенные оценки дисперсий;  $n_1$  и  $n_2$  – объемы выборок.

Критерием равенства средних служит критическое значение  $t$ , определенное по табл. 8 для уровня значимости 0,05 и  $n_1 + n_2 - 2$  степеней свободы.

**Пример.** Сравним средние значения открытой пористости по двум небольшим залежам. Число определений по первой залежи  $n_1 = 14$ ; средняя пористость  $K_{\text{оп}} = 18\%$ ; несмещенная оценка дисперсии  $S_1^2 = 3,6$ . По второй залежи:  $n_2 = 16$ ;  $K_{\text{оп}} = 16\%$ ;  $S_2^2 = 2,9$ . По формуле (69) вычислим  $t$ .

$$t_{\text{выч}} = \frac{18-16}{\sqrt{\frac{3,6}{14} + \frac{2,9}{16}}} = \frac{2}{\sqrt{0,257+0,181}} = 3,02.$$

Критическое значение  $t_{\alpha=0,05; k=28}$ , т.е. при  $\alpha = 0,05$  и 28 степенях свободы в соответствии с табл. 8, равно 2,048. Следовательно, полученное нами значение  $t_{\text{выч}} = 3,02$  превышает критическое  $t_{\alpha=0,05; k=28} = 2,048$ , в связи с чем нулевую гипотезу следует считать несостоятельной. Имеющиеся расхождения между средними значениями пористости обусловлены не случайными причинами, а отражают различия, связанные или с условиями формирования коллекторов, или с неравномерным выносом керна и выбором образцов на анализ пористости.

Если исследуются средние значения случайных величин, распределяющихся логнормально (например, проницаемости), то значение  $t$  вычисляется по формуле:



$$t_{\text{выч}} = \frac{\lg X_1 - \lg X_2}{\sqrt{\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2}}}, \quad (66)$$

где  $\lg X_1$  и  $\lg X_2$  – средние арифметические логарифмов значений про- ницаемости;  $S_1^2$  и  $S_2^2$  – несмещенные оценки дисперсий значений логнормально распределенных случайных величин.

### 4.3. СРАВНЕНИЕ ДИСПЕРСИЙ

При сравнении двух объектов сравниваются не только средние, но и дисперсии. Нулевая гипотеза о равенстве дисперсий  $H_0: \sigma_1^2 = \sigma_2^2$  про- веряется с помощью критерия Фишера ( $F$ ):

$$F = \frac{s_1^2}{s_2^2} \text{ при } s_1^2 > s_2^2, \quad F = \frac{s_2^2}{s_1^2} \text{ при } s_2^2 > s_1^2, \quad (67)$$

где  $S_1^2$  и  $S_2^2$  представляют собой несмещенные оценки дисперсий значений нормально распределенных случайных величин или дис- персий значений логнормально распределенных случайных величин. Полученное значение критерия  $F$  сравнивается с его критическим значением  $F_{\text{кр}}$  при установленном уровне значимости 0,05 и степенях свободы  $\left\{ \frac{k_1}{k_2} \right\}$ , где  $k_1 = n_1 - 1$  и  $k_2 = n_2 - 1$ . Критические значения  $F$ -кри- терия в табличном редакторе можно найти с помощью функции ФРАСПОБР.

Сравним дисперсии малых выборок рассмотренных выше двух объектов. По первому объекту  $S_1^2 = 3,6$ , а по второму  $S_2^2 = 2,9$ . Рас- считаем значение  $F_{\text{выч}}$ , разделив большую дисперсию на меньшую:

$$F_{\text{выч}} = \frac{3,6}{2,9} = 1,24.$$

По таблицам приложения  $F$  критическое при уровне значимости 0,05 и числе степеней свободы  $k_1 = 16 - 1 = 15$  и  $k_2 = 14 - 1 = 13$  рав-

но 2,53. Вычисленное значение  $F_{\text{выч}}=1,24$  оказалось меньше критического (2,53). Следовательно, нулевая гипотеза о равенстве дисперсий в этом случае не отвергается. Это свидетельствует о несущественной разнице в степени рассеяния значений открытой пористости около средних по обеим выборкам, что может в какой-то мере свидетельствовать о приблизительно одинаковой степени неоднородности коллекторов двух объектов относительно каждого среднего значения. Однако существенная разница в их средних значениях, обусловленная или различиями в условиях формирования, или в представительности кернового материала, не позволяет объединить эти малые выборки в одну выборочную совокупность.

#### 4.4. СРАВНЕНИЕ НЕСКОЛЬКИХ ПРОДУКТИВНЫХ ОБЪЕКТОВ

Постановка задачи и обоснование необходимости сравнения нескольких продуктивных объектов, разделенных друг от друга как по площади, так и по разрезу, похожа на задачу сравнения двух объектов. Однако для решения ее применяется более сложный, чем для сравнения двух объектов, математический аппарат.

**Сравнение ряда средних.** Рассмотрим, можно ли считать одинаковыми средние значения пористости четырех продуктивных пластов, приуроченных к единой продуктивной толще, каждый из которых содержит небольшую нефтяную залежь. Необходимость постановки такой задачи заключается в том, чего если пласты существенно не отличаются значениями пористости, то исследование зависимости между удельной продуктивностью скважин и пористостью можно было бы провести для всех четырех пластов совместно.

Пусть имеются данные о средних значениях пористости, числе определений, дисперсиях и средних квадратических отклонениях от средних по каждому из четырех пластов (малые выборки).

Нулевая гипотеза для этой задачи формулируется следующим образом:

$$H_0 : M(k_{\text{оп1}}) = M(k_{\text{оп2}}) = M(k_{\text{оп3}}) = M(k_{\text{оп4}}),$$

где  $M(K_{\text{оп1}}) \dots$  – математические ожидания значения пористости по каждой выборке.

Сравнение будем проводить методом **дисперсионного анализа**. Его основная идея в том, чтобы свести сравнение средних к сравнению дисперсий, а именно: если дисперсии внутри выборок значимо меньше (вариабельность внутри выборки), чем дисперсия средних между выборками (вариабельность между выборками), то нулевая гипотеза отвергается. Для сравнения дисперсий используется уже знакомый нам критерий Фишера.

Для дисперсии внутри выборки имеем формулу:

$$\bar{S}^2 = \frac{\sum_{i=1}^l (n_i - 1) \cdot S_i^2}{n - l}, \quad (68)$$

где  $n$  – общее число наблюдений,  $l$  – число выборок (в рассматриваемом примере 4),  $n_i$  и  $S_i$  – объём и несмещённая оценка среднеквадратического отклонения в выборке  $i$ . Оценка между выборками производится по формуле

$$\tilde{S}^2 = \frac{\sum_{i=1}^l n_i \cdot \bar{k}_{\text{оп}i}^2 - n \cdot \bar{k}_{\text{оп}}^2}{l - 1}, \quad (69)$$

где  $\bar{k}_{\text{оп}i}$  – среднее значение коэффициента открытой пористости в  $i$ -й выборке, а  $\bar{k}_{\text{оп}}$  – среднее значение коэффициента открытой пористости по всем выборкам, которое находится по формуле:

$$\bar{k}_{\text{оп}} = \frac{\sum_{i=1}^l n_i \cdot \bar{k}_{\text{оп}i}}{n}. \quad (70)$$

В итоге значение критерия Фишера находится по формуле

$$F = \frac{\tilde{S}^2}{\bar{S}^2} = \frac{n-l}{l-1} \frac{\sum_{i=1}^l n_i \cdot \bar{k}_{опi}^2 - n \cdot \bar{k}_{оп}^2}{\sum_{i=1}^l (n_i - 1) \cdot S_i^2}, \quad (71)$$

а это значит, что он распределён по закону Фишера с  $\left\{ \frac{l-1}{n-l} \right\}$  степенями свободы.

### Пример

Таблица 9

1 пласт	2 пласт	3 пласт	4 пласт
$n_1 = 18$	$n_2 = 16$	$n_3 = 20$	$n_4 = 19$
$K_{оп1} = 12$	$K_{оп2} = 10$	$K_{оп3} = 13$	$K_{оп4} = 14$
$S_1^2 = 1,4$	$S_2^2 = 2,1$	$S_3^2 = 1,6$	$S_4^2 = 1,9$
$S_1 = 1,18$	$S_2 = 1,45$	$S_3 = 1,27$	$S_4 = 1,38$

Подставляя фактические значения в формулы (68)–(71) получим:

$$\bar{k}_{оп} = \frac{1}{73}(18 \cdot 12 + 16 \cdot 10 + 20 \cdot 13 + 19 \cdot 14) = 12,356,$$

$$\bar{S}^2 = \frac{1}{69}(17 \cdot 1,4 + 15 \cdot 2,1 + 19 \cdot 1,6 + 18 \cdot 1,9) = 1,7377,$$

$$\tilde{S}^2 = \frac{1}{3}(18 \cdot 12^2 + 16 \cdot 10^2 + 20 \cdot 13^2 + 19 \cdot 14^2 - 73 \cdot 12,356^2) = 50,247,$$

$$F_{выч} = \frac{50,247}{1,7377} = 28,92.$$

Нулевая гипотеза верна, если вычисленные значения  $F_{выч}$  не превысят критическое при уровне значимости 0,05 и  $\left\{ \frac{3}{69} \right\}$  степенях свободы.

По табличным данным  $F_{крит}$  для перечисленных условий равно 2,737. В рассмотренном примере вычисленные значения критерия превышает критическое. Следовательно, нулевую гипотезу нужно от-

бросить и принять альтернативу: средние значения пористости по каждому из пластов существенно отличаются друг от друга и в совместный анализ для определения кондиционных пределов не могут быть вовлечены.

#### 4.5. СРАВНЕНИЕ ДВУХ И БОЛЕЕ НЕОДНОРОДНЫХ ПРОДУКТИВНЫХ ПЛАСТОВ С ПОМОЩЬЮ ЭНТРОПИИ

Рассматривая возможности применения энтропии для изучения макронеоднородности продуктивных пластов, мы показали, что эта характеристика выгодно отличается от такого показателя, как коэффициент расчлененности пластов. Если коэффициент расчлененности указывает лишь на число проницаемых прослоев в продуктивном пласте, на чем и основывается сравнение макронеоднородности по расчлененности разных объектов, то энтропия одним числом позволяет количественно оценить качество макронеоднородности.

**Пример.** Пусть двумя скважинами, вскрывшими продуктивный пласт, установлено по три продуктивных прослоя. Коэффициент расчлененности по таким скважинам равен трем. Энтропия, наряду с числом прослоев, может оценить и долю мощности каждого прослоя в общей эффективной толщине пласта. Если все прослои в каждой из скважин имеют одинаковую толщину, то энтропия такой системы обращается в максимум, характеризуя тем самым максимальную неоднородность физической системы – продуктивного пласта. В случае разновеликих пластов энтропия будет тем меньше, чем однороднее продуктивный пласт, хотя коэффициент расчлененности и в том, и в другом случае будет одинаков.

Если один и тот же пласт в разных скважинах содержит по 3 прослоя, толщина которых соответственно равна:

$$h'_1 = 1 \text{ м}; h'_2 = 1,5 \text{ м}; h'_3 = 7,5 \text{ м};$$

$$h''_1 = 2 \text{ м}; h''_2 = 3 \text{ м}; h''_3 = 5 \text{ м},$$

то суммарная эффективная толщина продуктивного пласта равна в обеих скважинах 10 м.

Спрашивается, какую из двух систем при соблюдении прочих равных условий считать более однородной?

Рассчитаем энтропии:

$$\begin{aligned} H_1 &= -(0,1\ln(0,1) + 0,15\ln(0,15) + 0,75\ln(0,75)) = \\ &= 0,2303 + 0,2846 + 0,2158 = 0,7307, \end{aligned}$$

$$\begin{aligned} H_2 &= -(0,2\ln(0,2) + 0,3\ln(0,3) + 0,5\ln(0,5)) = \\ &= 0,3219 + 0,3612 + 0,3466 = 1,0297. \end{aligned}$$

Следовательно, неоднородность, продуктивного пласта в первой скважине почти в полтора раза меньше неоднородности пласта во второй скважине. И это при одинаковом коэффициенте расчлененности. Значит, энтропия позволяет количественно оценить качество неоднородности пласта по расчлененности.

В рассмотренном примере пласт, содержащий два тонких прослоя (1,0 и 1,5 м) и один мощный (7,5 м), оказался более однороден, чем пласт, у которого толщины прослоев распределяются более равномерно. Нетрудно заметить, что с увеличением вероятностей  $p$  энтропия постепенно возрастает, достигая максимума при  $p = 0,37$ , а затем начинает постепенно убывать до 0 при  $p = 1$ , то есть, чем больше доля одного какого-либо прослоя в общей эффективной толщине пласта, тем меньше энтропия системы, тем более однороден разрез.

Изучая макронеоднородность по расчлененности, И.П. Чоловский выявил: чем больше доля мощных прослоев в общей эффективной толщине пласта, тем выше темп отбора нефти. Так, на некоторых площадях Ромашкинского месторождения, характеризующихся приблизительно одинаковой гидропроводностью, темп отбора увеличивается с увеличением однородности разреза. На Зай-Каратайской площади доля мощного пласта в общей эффективной толщине горизонта  $D_1$  составляет 53,3 %, а темп отбора равен 0,56 (за единицу

принят темп отбора на Миннибаевской площади). Лениногорская площадь более однородна. Здесь на мощный пласт приходится 57 % общей эффективной толщины горизонта  $D_1$ . Соответственно, выше по этой площади и темп отбора (0,67). На Алькеевской площади доля мощного пласта 62,8 %, а темп отбора – 0,81. Наиболее однороден горизонт  $D_1$  на Чишминской площади. Здесь доля мощного пласта составляет 81 % от общей эффективной мощности горизонта  $D_1$ . На этой площади достигнут наибольший темп отбора нефти.

Таким образом, характеристика неоднородности продуктивных пластов с помощью энтропии позволяет использовать ее в качестве меры неоднородности пластов при сравнении разных продуктивных объектов.

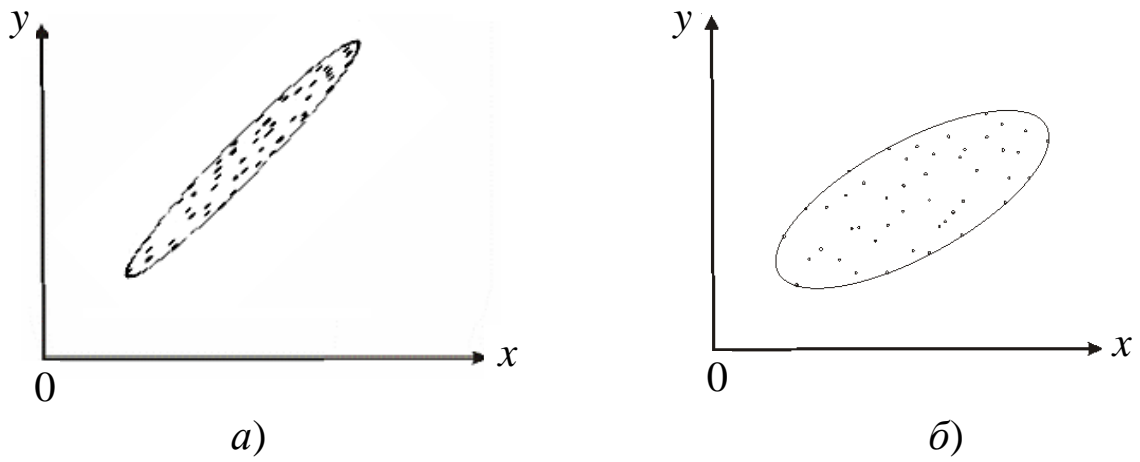
## 5. ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ МЕЖДУ ДВУМЯ ПАРАМЕТРАМИ

Изучение зависимости между различными параметрами продуктивных пластов находит в нефтегазопромысловой геологии самое широкое применение. На этом анализе основано определение кондиционных пределов коллекторских свойств пластов. Исследование зависимости между предыдущими и последующими дебитами положено в основу статистического метода подсчета запасов нефти. Для выяснения путей повышения нефтеотдачи и изучения влияния геолого-физических факторов на разработку исследуются зависимости между различными геолого-физическими параметрами: вязкостью флюидов, проницаемостью, песчаностью разреза и технологическими показателями: плотностью сетки скважин, темпом отбора и т.п.

Пожалуй, нет ни одного раздела в нефтегазопромысловой геологии, где бы не изучались зависимости между различными параметрами. Не менее широкое применение имеет изучение зависимостей между параметрами при геофизических исследованиях на установление петрофизических связей.

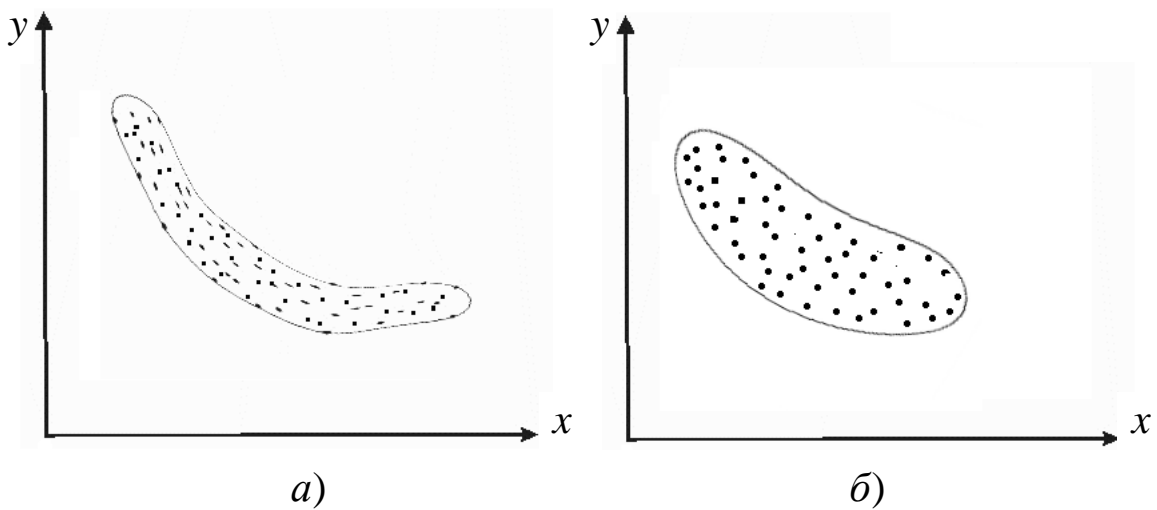
В отличие от функциональной зависимости, где какая-либо величина определяется как однозначная функция одной или нескольких величин, между случайными величинами может существовать связь, проявляющаяся лишь в тенденции к изменению одной величины с изменением другой. При этом связь между двумя случайными величинами изображается на диаграмме в виде множества точек, каждая из которых характеризует пару значений исследуемых величин  $x$  и  $y$ . Плотность точек, прямая или обратная направленность, а также линейность или криволинейность их множества свидетельствует о различных видах связи между двумя исследуемыми величинами. Такая связь называется стохастической или вероятностной, корреляционной. На рис. 21 показана прямая тесная (*а*) и слабая (*б*) линейные свя-



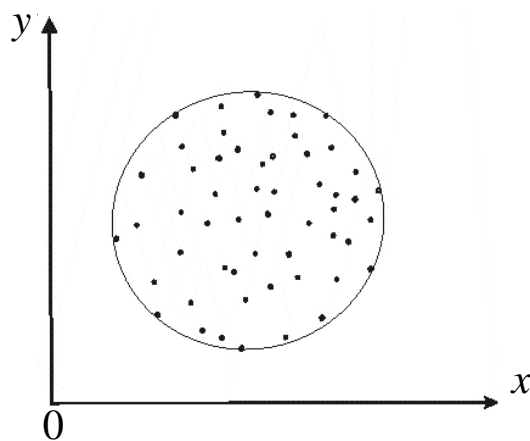


**Рис. 21. Прямая тесная (а) и слабая (б) линейные связи**

зи; на рис. 22 представлены обратные тесная (а) и слабая (б) криволинейные связи; на рис. 23 показано множество точек, взаимное расположение которых свидетельствует об отсутствии какой-либо связи.



**Рис. 22. обратные тесная (а) и слабая (б) криволинейные связи**



**Рис. 23. Отсутствие связи**

Особенность расположения точек на диаграмме в виде «облака рассеяния» обуславливает соответствие одному какому-либо фиксированному значению  $X = x_i$  ряда значений  $Y$  в диапазоне:  $Y = y_{i \max} - y_{i \min}$ . Вместе с тем, особенность множества точек такова, что одновременно одному фиксированному значению  $Y = y_j$  соответствует ряд значений в диапазоне:  $X = x_{j \max} - x_{j \min}$  (рис. 24).

В этом заключается принципиальное отличие стохастической, корреляционной связи от функциональной.

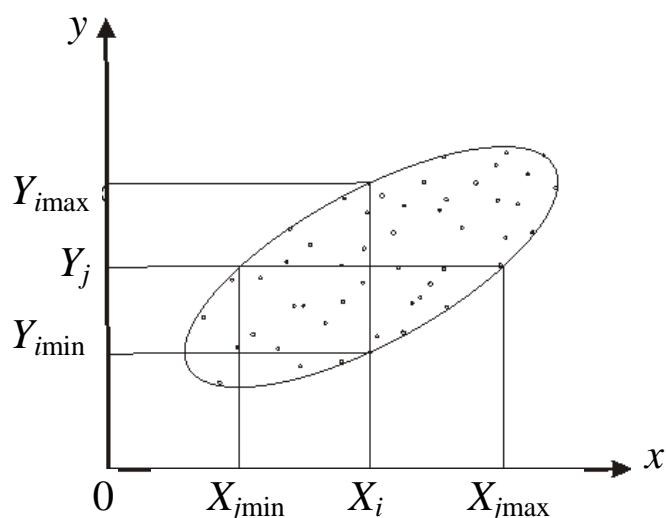


Рис. 24. Связь между параметрами  $x$  и  $y$

Следует обратить внимание на еще одну особенность корреляционной связи между двумя параметрами: любой вид связи множества точек описывается двумя уравнениями прямых или кривых, соответственно (рис. 25). При этом одна кривая описывает зависимость  $Y$  от  $X$ ,

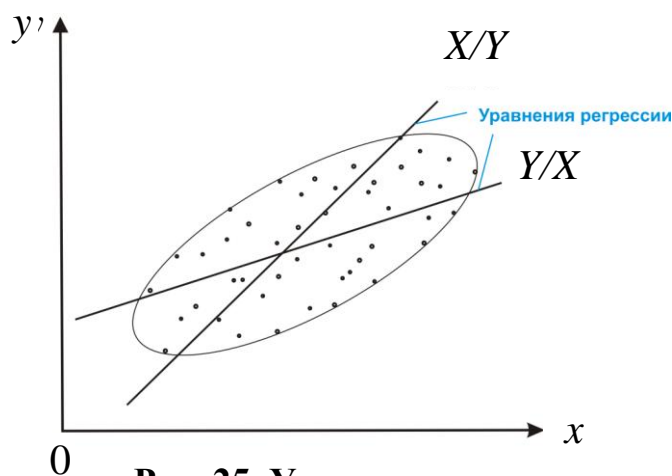


Рис. 25. Уравнения регрессии

а вторая  $X$  от  $Y$ . Чем теснее связь, тем угол между двумя прямыми меньше. На рис. 25 изображён случай прямолинейной зависимости. Координаты точки пересечения прямых соответствуют средним по выборке значениям величин  $X$  и  $Y$ .

### 5.1. ОПРЕДЕЛЕНИЕ КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ, БЛИЗКОЙ К ЛИНЕЙНОЙ, МЕЖДУ ДВУМЯ ГЕОЛОГО-ФИЗИЧЕСКИМИ ПАРАМЕТРАМИ

Если две случайных величины примерно одинаково меняются или, другими словами, варьируются, то степень согласованности варьирования показывает коэффициент ковариации (т.е. совместного варьирования). Коэффициент ковариации двух случайных величин есть математическое ожидание их центрированного произведения:

$$\text{cov}(X, Y) = M[\overset{\circ}{X} \overset{\circ}{Y}].$$

Оценка коэффициента ковариации по выборочным данным производится по формуле:

$$\widehat{\text{cov}}(x, y) = \frac{1}{n(n-1)} \left( n \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i \right).$$

Мерой степени зависимости между двумя случайными величинами, отражающей также существование между ними связи, близкой к линейной, является коэффициент корреляции  $r_{xy}$ , вычисляемый следующим образом:

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \cdot \sigma_y}, \quad (72)$$

где  $\sigma_x, \sigma_y$  – средние квадратические отклонения, соответственно, величин  $X$  и  $Y$ .

Оценка  $\hat{r}_{xy}$ , вычисляется по формуле:

$$\hat{r}_{xy} = \frac{\widehat{\text{cov}}(x, y)}{S_x \cdot S_y} = \frac{n \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right]} \cdot \sqrt{\left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}. \quad (73)$$

Согласно формуле (73) коэффициент корреляции является единым для обоих уравнений прямых, как  $y/x$ , так  $x/y$ , что отражает выражение в знаменателе, соответствующее произведению двух среднеквадратических отклонений для значений  $X$  и  $Y$ .

Коэффициент корреляции  $r_{xy}$  изменяется от  $-1$  до  $0$  – для обратной, и от  $0$  до  $+1$  – для прямой связей.

Если при расчете  $r_{xy}$  установлено наличие тесной корреляционной связи, что свидетельствует о наличии зависимости, близкой к линейной, то ее можно представить аналитически в виде уравнения  $y$  по  $x$  (или  $y/x$ ). Уравнение оценки величины  $y$  по  $x$  называется уравнением регрессии:

$$\hat{y}_i = a + b(x_i - \bar{x}). \quad (74)$$

В этом уравнении:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Графическим выражением уравнения (74) является прямая линия, построенная методом наименьших квадратов. Для построения прямой этим методом должно соблюдаться условие, что сумма квадратов отклонений всех наблюдаемых значений от этой прямой будет минимальной, то есть:

$$\sum (\hat{y}_i - y_i)^2 = \sum_{i=1}^n d_i^2 = \min.$$

Оценка коэффициента  $a$  определяется уравнением:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_i, \quad (75)$$

то есть оценка  $\hat{a}$  равна среднему арифметическому из наблюдаемых значений  $y$  и представляет собой ординату прямой в точке  $x = \bar{x}$ .

Угловым коэффициентом для уравнения  $y$  по  $x$   $b_{y/x}$ , называемый коэффициентом регрессии, характеризует угол наклона прямой. Его оценка определяется по формуле

$$\hat{b}_{y/x} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} = \hat{r}_{xy} \frac{S_y}{S_x}, \quad (76)$$

где числитель представляет собой оценку ковариации, а знаменатель оценку дисперсии  $S_x^2$ .

Таким образом, эмпирическое уравнение регрессии  $y$  по  $x$  будет иметь следующий вид:

$$\hat{y}_i = \bar{y} + b_{y/x} (x_i - \bar{x}). \quad (77)$$

Аналогичным образом можно написать и эмпирическое уравнение регрессии  $X$  по  $y$ :

$$\hat{x}_i = \bar{x} + b_{x/y} (y_i - \bar{y}). \quad (78)$$

Обе прямые проходят через центр тяжести облака рассеяния под некоторым углом относительно друг друга. Величина угла определяется разницей между  $b_{x/y}$  и  $b_{y/x}$ . Лишь в двух случаях, когда  $r = \pm 1$ , обе прямые совпадут.

В отличие от формулы (76), в знаменателе дроби, по которой вычисляется  $b_{x/y}$ , следует написать выражение:

$$n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2.$$

**Пример.** Определение кондиционного предела пористости по зависимости наблюдаемых значений этого параметра от проницаемости.

Исследованиями по  $i$ -й залежи установлен нижний предел проницаемости, равный 4 мД. Для более объективного подсчета запасов нефти необходимо определить нижний предел открытой пористости коллекторов для последующей отбраковки толщин, характеризующихся некондиционной пористостью.

Наблюдаемые значения проницаемости  $K_{пр}$  и соответствующие им значения открытой пористости  $K_{пi} = y_i$  сведены в табл. 10. Поскольку проницаемость в большинстве случаев распределяется логнормально, во втором столбце этой таблицы рассчитаны логарифмы значений этого параметра, то есть  $x_i = \lg K_{пi}$ .

Таблица 10

Расчетные данные для определения  $r$  и  $\eta$

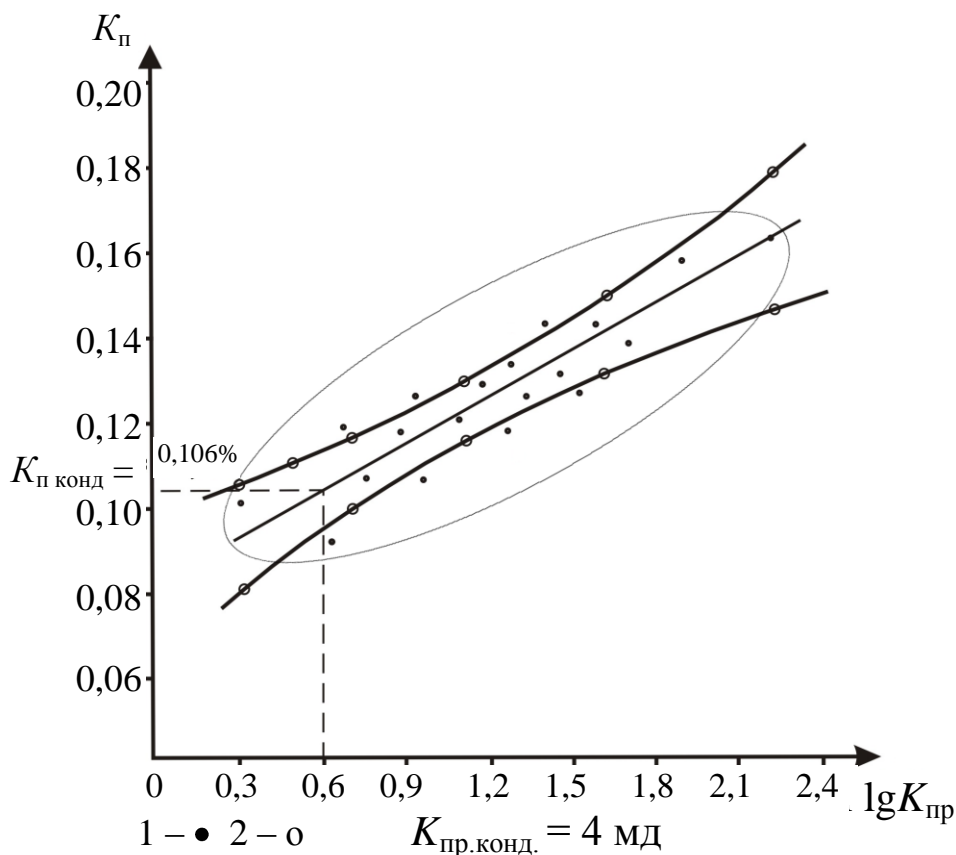
$k_{пр}$	$x_i = \lg K_{пi}$	$y_i = K_{пi}$	$x_i y_i$	$x_i^2$	$y_i^2$	$\hat{y}_i$	$d_i = y_i - \hat{y}_i$	$d_i^2$
2,0	0,302	10,2	3,08	0,091	104,04	9,414	+0,786	0,618
4,13	0,615	9,3	5,72	0,378	86,49	10,609	-1,309	1,713
5,56	0,745	10,8	8,05	0,555	116,64	11,106	-0,306	0,094
3,06	0,485	11,1	5,38	0,235	123,21	10,113	+0,987	0,974
4,67	0,670	12,0	8,04	0,449	144,00	10,819	+1,181	1,395
1,93	0,285	8,1	2,31	0,081	65,61	9,349	-1,249	1,560
12,0	1,079	13,3	14,35	1,164	176,89	12,382	+0,918	0,842
8,90	0,950	10,6	10,07	0,902	112,36	11,889	-1,289	1,662
8,32	0,920	12,9	11,87	0,846	166,41	11,774	+1,126	1,268
20,2	1,305	11,2	14,62	1,703	125,44	13,245	-2,045	4,182
18,6	1,270	13,7	17,40	1,613	187,69	13,111	+0,589	0,347
24,6	1,391	14,6	20,31	1,935	213,16	13,574	+1,026	1,053
32,0	1,505	12,9	19,41	2,265	166,41	14,009	-1,109	1,230
74,0	1,869	16,2	30,28	3,493	262,44	15,400	+0,800	0,640
166	2,200	14,6	32,12	4,840	213,16	16,664	-2,064	4,260
164	2,150	18,5	39,77	4,623	342,25	16,473	+2,027	4,109
сумма	17,741	200,0	242,78	25,173	2606,2	242,78		25,947

По данным табл. 10 на графике (рис. 26) нанесены точки, расположение которых свидетельствует о наличии довольно тесной связи между наблюдаемыми значениями проницаемости и пористости. Оценим ее количественно.

В первую очередь рассчитаем коэффициент корреляции, поскольку при отсутствии линейной связи вычислять уравнение регрессии для определения кондиционного значения пористости не имеет смысла.

Таким образом, зависимость между наблюдаемыми значениями проницаемости и пористости характеризуется достаточно высокой теснотой связи, что свидетельствует о наличии линейной зависимости.

$$\hat{r}_{xy} = \frac{16 \times 242,78 - 17,741 \times 200}{\sqrt{16 \times 25,73 - 314,743} \sqrt{(16 \times 2606,2 - 40000)}} = 0,87.$$



**Рис. 26. Определение кондиционных значений коэффициента пористости по кондиционному значению проницаемости:**  
1 – фактические точки, 2 – расчетные точки доверительного интервала

Теперь приступим к вычислению уравнения регрессии и построению прямой. По данным 2 и 3 столбцов определяем средние значения:

$$\bar{x} = \frac{17,741}{16} = 1,109; \quad \bar{y} = \frac{200,00}{16} = 12,5.$$

Следовательно, первый выборочный коэффициент  $\bar{a} = \bar{y} = 12,50$ .

Для расчета коэффициента  $\hat{b}$  используем данные столбцов 2, 3, 4 и 5. Подставив их в формулу (76), получим:

$$\hat{b} = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{16 \times 242,78 - 17,741 \times 200}{16 \times 25,173 - (17,741)^2} = 3,82.$$

Теперь можно рассчитать уравнение регрессии. Для этого значения  $\bar{y} = \hat{a}$  и  $\hat{b}$  подставим в формулу (77):

$$\hat{y} = \bar{a} + \hat{b} (x - \bar{x}) = 12,50 + 3,82(x - 1,109) = 3,82x + 8,26,$$

$$\hat{y} = 8,26 + 3,82x.$$

Рассчитав значения  $\hat{y}$  для максимального и минимального наблюдаемых значений  $x$ , получим координаты двух точек. Если эти точки нанести на график (рис. 26) и соединить их, то построенная таким образом прямая и будет удовлетворять уравнению регрессии  $\hat{y} = 8,26 + 3,82x$ .

Отложив по оси абсцисс (рис. 26) кондиционное значение проницаемости, равное  $\lg k_{\text{пр.конд}} = \lg 4 = 0,602$ , найдем соответствующее ему кондиционное значение открытой пористости, равное 10,6 %. Это значит, что при подсчете запасов нефти и газа по  $i$ -ой залежи мы имеем все основания отбраковать значения толщин, пористость которых меньше кондиционной, то есть  $< 10,6$  %.



## 5.2. ИЗУЧЕНИЕ ЗАВИСИМОСТИ МЕЖДУ ДВУМЯ ГЕОЛОГО-ФИЗИЧЕСКИМИ ПАРАМЕТРАМИ С ПОМОЩЬЮ КОРРЕЛЯЦИОННОГО ОТНОШЕНИЯ

Коэффициент корреляции можно рассматривать только как показатель линейности связи между двумя случайными величинами, при наличии которой коэффициент корреляции стремится к  $\pm 1$ . Только при наличии линейной связи коэффициент корреляции можно рассматривать как меру этой связи.

Однако при значениях этого коэффициента, близких к 0, нельзя сказать, что связь между исследуемыми случайными величинами отсутствует, поскольку может быть криволинейной.

Для оценки степени корреляционной связи в общем случае, в том числе и линейной, пользуются корреляционным отношением, определяемым равенством для  $Y$  по  $X$ :

$$\eta_{y/x} = \frac{\sigma_{\bar{y}_j}}{\sigma_y}, \quad (79)$$

и равенством

$$\eta_{x/y} = \frac{\sigma_{\bar{x}_j}}{\sigma_x} \quad (80)$$

для зависимости  $x$  по  $y$ .

В формуле (79)  $\sigma_y$  – общее среднее квадратическое отклонение выборочной совокупности значений  $y$ . Его оценка вычисляется по формуле

$$S_y = \sqrt{\frac{1}{n} \left[ \sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 \right]}. \quad (81)$$

Значение  $\sigma_{\bar{y}_j}$  представляет собой среднее квадратическое отклонение групповых средних  $\bar{y}_j$  от общего среднего  $\bar{y}$ . Оценка дисперсии групповых средних вычисляется по формуле:

$$S_{y_i} = \sqrt{\frac{1}{n} \sum n_i (\bar{y}_j - \bar{y})^2}. \quad (82)$$

По аналогичным формулам вычисляются и общая и групповая дисперсии для  $x$ .

Отметим, что похожая идея сравнения дисперсий была использована в дисперсионном анализе см. раздел про сравнение нескольких продуктивных объектов.

**Пример.** Приведем пример расчета корреляционного отношения для данных табл. 10.

Для расчета дисперсии групповых средних шкала значений  $x$  разбивается на ряд равновеликих интервалов, внутри которых определяются групповые средние арифметические для значений  $y$  (рис. 27). Все вычисления оценки  $S_{y_j}^2$  иллюстрирует табл. 11, составленная по данным табл. 10.

Подставляя результаты расчетов табл. 11 в формулу (82) при  $n = 16$  (из табл. 10), получим:

$$S_{y_j}^2 = \frac{78,378}{16} = 4,897.$$

Отсюда  $S_{y_j} = 2,21$ .

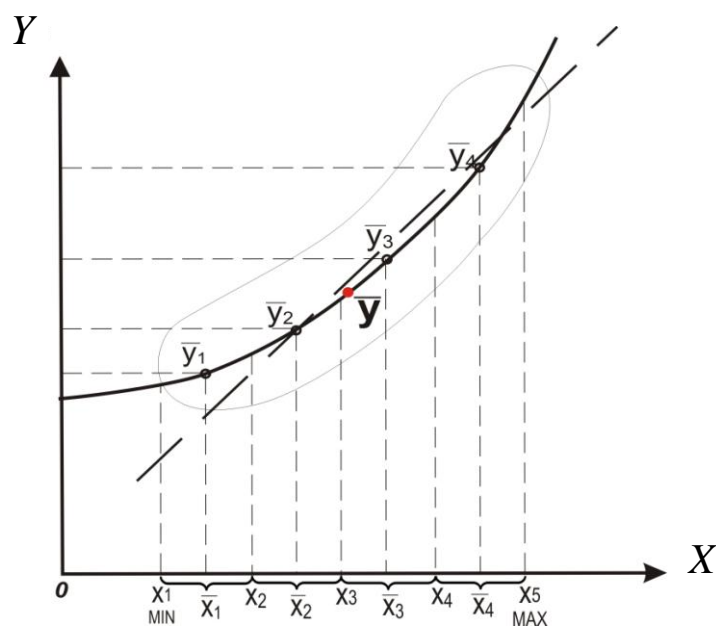


Рис. 27. К расчету корреляционного отношения

Таблица 11

Интервалы по $x$ в $\lg(\lg K_{\text{при}})$	Значения $y_i$ внутри интервала, в % ( $K_{\text{пи}}$ )	Число наблюдений внутри интервала $n_j$	Групповые средние $\bar{y}_j$	$\bar{y}_j - \bar{y}$	$(\bar{y}_j - \bar{y})^2$	$n_j (\bar{y}_j - \bar{y})^2$	$\bar{y}$
0,25–0,75	8,1; 10,2; 11,1; 9,3; 12,0; 10,8	6	10,250	2,25	5,062	30,376	12,5
0,75–1,25	12,9; 10,6; 13,3	3	12,267	0,233	0,054	0,162	из табл.10
1,25–1,75	13,7; 11,2; 14,6; 12,9	4	13,100	0,600	0,360	1,440	
1,75–2,25	16,2; 18,5; 14,6	3	16,433	3,933	15,468	46,404	
						$\Sigma = 78,376$	

Теперь рассчитаем общую дисперсию выборочной совокупности по формуле (81). Для этого воспользуемся данными табл. 10:

$$S_y^2 = \frac{1}{16} \left[ 2606,2 - \frac{1}{16} (40000) \right] = \frac{1}{16} 2606,2 - 2500 = \frac{106,2}{16} = 6,637.$$

Отсюда  $S_y = 2,576$ .

Определим корреляционное отношение:

$$\eta_{y/x} = \frac{S_{\bar{y}_j}}{S_y} = \frac{2,21}{2,576} = 0,86.$$

Корреляционное отношение оказалось почти равным коэффициенту корреляции, рассчитанному для уравнения регрессии:

$$\hat{Y} = 3,82x + 8,26.$$

Между  $r$  и  $\eta$  могут иметь место следующие соотношения:

$(|r| \approx \eta) \approx 0,75-1,0$ , если  $y$  и  $x$  связаны тесной линейной зависимостью;

$(|r| \approx \eta) \approx 0,5-0,75$ , если  $y$  и  $x$  связаны слабой линейной корреляционной зависимостью;

$(|r| \approx \eta) \approx 0,75-1,0$ , если  $y$  и  $x$  связаны тесной нелинейной статистической зависимостью;

$(|r| \approx \eta) \approx 0,5-0,75$ , если  $y$  и  $x$  связаны слабой нелинейной корреляционной зависимостью;

$|r| < 0,5$  и  $\eta < 0,5$ , если между  $y$  и  $x$  связь практически отсутствует.

*В тех случаях, когда корреляционное отношение выше, чем коэффициент корреляции, следует пользоваться параметрами криволинейной связи, в том числе и значениями корреляционного отношения.*

### 5.3. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА УРАВНЕНИЯ РЕГРЕССИИ

Так как уравнение регрессии строится по выборочной совокупности случайных величин, то и параметры его  $\bar{x}$ ,  $\bar{y}$ ,  $b_{y/x}$  являются случайными величинами и для них можно определить доверительные интервалы, так же как и для оценки величины  $y$  по  $x - Y(x_i)$ .

Н.В. Смирнов и И.В. Дунин-Барковский [13] рекомендуют оценивать точность и надежность уравнения регрессии следующим образом. Для любого произвольно взятого значения  $x = x_0$  ошибка значения прямой  $\hat{Y}$  в этой точке оценивается по формуле:

$$\sigma = \frac{t_q S_{\hat{y}}}{\sqrt{n-2}} \sqrt{1 + \frac{n^2 (x_0 - \bar{x})^2}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}}; \quad (83)$$

$t_q$  – распределяется по закону Стьюдента с  $n-2$  степенями свободы;  $S_{\hat{y}}$  – выборочное среднее квадратическое отклонение, вычисляемое по формуле:

$$S_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - Y(x_i))^2}{n}} = \sqrt{\frac{\sum d_i^2}{n}}. \quad (84)$$

**Пример.** Подставив данные примера на с. 97 в (88), получаем  $S_{\hat{y}} = \sqrt{\frac{25 \cdot 947}{16}} = 1,274$ . Для  $x_0 = 0,300$  при доверительной вероятности  $1-\alpha = 95\%$ , которой соответствует  $t_{0,05} = 2,145$  при  $k = n-2 = 14$  (табл. 8), получим:  $\hat{Y} = 8,26 + 3,82X 0,300 = 9,406$ . Тогда доверительный интервал будет:

$$9,406 - \frac{2,145 \cdot 1,274}{3,74} \sqrt{1 + \frac{16^2 \cdot 0,300 - 1,109^2}{16 \times 25,173 - 314,743}} <$$

$$< M(Y/x_0) < 9,406 + \frac{2,145 \cdot 1,274}{3,74} \sqrt{1 + \frac{16^2 \cdot 0,300 - 1,109^2}{16 \cdot 25,173 \cdot 314,743}}$$

или

$$8,164 < M(Y/x_0) < 10,648.$$

Рассчитав последовательно аналогичным образом значения  $\hat{Y}$  и ошибок  $\pm\delta$  для  $x_0 = 0,7; 1,109; 1,6; 2,2$ , укажем полученные доверительные интервалы:

$$\text{при } x_0 = 0,7; \quad \hat{y} = 10,934; \quad 10,043 < y < 11,825,$$

$$\text{при } x_0 = 1,109; \quad \hat{y} = 12,496; \quad 11,765 < y < 13,227,$$

$$\text{при } x_0 = 1,6; \quad \hat{y} = 14,372; \quad 13,418 < y < 15,281,$$

$$\text{при } x_0 = 2,2; \quad \hat{y} = 16,664; \quad 15,047 < y < 18,281.$$

Чтобы построить доверительные интервалы, достаточно на рис. 26 для соответствующих  $x_0$  отложить выше и ниже прямой вычисленные значения границ этих интервалов.

На указанном рисунке полученные таким образом точки соединены плавными линиями, которые и являются границами доверительного интервала уравнения регрессии:

$$\hat{y} = 8,26 + 3,82x.$$

В соответствии с этими границами доверительный интервал для

кондиционного значения пористости будет определяться величинами  $9,5 < m_{\text{конд}} < 11,7$ .

Расширение доверительного интервала к концам прямой свидетельствует об уменьшении точности оценки по мере удаления от среднего значения  $\bar{x}$ . Поэтому экстраполировать построенную прямую за наблюдаемые значения следует с очень большой осторожностью.

Для вычисления доверительного интервала коэффициента регрессии  $\hat{b}$  зададимся также 95%-ной доверительной вероятностью. Тогда

$$\hat{b} - t_{q_{0,05}; n-2} \cdot S_{\hat{b}} < b < \hat{b} + t_{q_{0,05}; n-2} \cdot S_{\hat{b}},$$

где  $t_{q_{0,05}}$  при  $k = 16 - 2$  степенях свободы по табл. 8 равно 2,145;  $S_{\hat{b}}$  выборочное среднее квадратическое отклонение для  $b$ , равное:

$$S_{\hat{b}} = \sqrt{\frac{1}{n-2} \cdot \frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}}. \quad (85)$$

**Пример.** Подставим данные примера на с. 97 в формулу (85), получим:

$$S_{\hat{b}} = \sqrt{\frac{1}{14} \cdot \frac{25,947}{25,173 - \frac{314,743}{16}}} = 0,583.$$

Теперь можно оценить доверительный интервал:

$$3,82 - 2,145 \cdot 0,583 < b < 3,82 + 2,145 \cdot 0,583,$$

$$2,569 < b < 5,071.$$

Следовательно, угол тангенс наклона прямой регрессии будет определяться доверительным интервалом от 2,569 до 5,071.

## 6. МАТЕМАТИЧЕСКОЕ ОБОСНОВАНИЕ ПРИЁМОВ ГЕОМЕТРИЗАЦИИ ПРИ ГЕОЛОГИЧЕСКОМ МОДЕЛИРОВАНИИ ПРОДУКТИВНЫХ ПЛАСТОВ

### 6.1. ПОНЯТИЕ О ГЕОЛОГИЧЕСКОМ МОДЕЛИРОВАНИИ

Человек изучает окружающий мир, моделируя его, т.е. создавая модели, подобные реальным объектам, с которыми ему удобнее работать. Построение геологической модели заключается в пошаговом применении тех или иных средств моделирования к имеющимся данным в соответствии с выбранной процедурой моделирования. Один шаг может быть простым добавлением имеющихся данных в модель, например, загрузкой траекторий скважин – **шаг загрузки**. Это могут быть шаги, связанные непосредственно с моделированием параметров и значений – **шаги моделирования**. И, наконец, могут вычисляться некоторые суммы и величины, например, объём нефтенасыщенного пространства при заданных величинах пористости и областях нефтенасыщенности – **вычислительный шаг**.

Ответственность за правильное выполнение шагов загрузки и вычислительных шагов несёт, прежде всего, разработчик программного средства, при помощи которого создаётся модель, хотя непосредственным исполнителем этих шагов является геолог. Особое внимание геолог должен уделять шагам моделирования.

Остановимся подробнее на типичных шагах моделирования. Это может быть шаг, связанный с изменением структуры модели, другой шаг – дискретизацией модели и представлением её в виде набора ячеек, третий – связанный с интерполяцией имеющихся экспериментальных данных с целью наполнения модели в областях с отсутствием экспериментальных данных, четвертый – с поиском и использованием зависимостей в экспериментальных данных с целью определения параметров и связей между ними в модели.

Предыдущая часть курса была, в основном, посвящена вопросам, возникающим на шагах моделирования, связанных с определением и использованием различных зависимостей в данных и их статистических характеристик. Теперь сосредоточимся на шагах интерполяции.

Если рассмотреть процесс геологического моделирования по существу, то его можно разделить на три последовательных этапа: структурное моделирование, литологическое моделирование и моделирование свойств.

### Структурное моделирование

На этапе структурного моделирования создаётся модель геометрической структуры залежи – строятся стратиграфические поверхности кровли и подошвы, разломов, газо-жидкостных контактов и др.; с использованием построенных поверхностей создаются трёхмерные сетки, разбивающие пространство залежи на элементарные ячейки, в рамках которых дальше будет вестись процесс моделирования.

Базисом для этого этапа является схема детальной корреляции разрезов скважин, поверхности сейсмоотражающих горизонтов и поверхности разломов по данным сейсмических исследований. На основе этих данных осуществляется построение стратиграфического каркаса – стратиграфических поверхностей кровли и подошвы пласта (и, возможно, опорных поверхностей внутри) для дальнейшего построения трёхмерной сетки. Эти построения ведут путём *интерполяции* одноимённых корреляционных границ на скважинах с учётом карты отражающего горизонта и поверхностей разломов. Далее производится разбиение стратиграфического каркаса на элементарные ячейки с учётом типа напластования. Главным результатом этого этапа является сеточная модель, отражающая геометрическую структуру залежи. Это – первый этап геометризации.



## Литологическое моделирование

Этап литологического моделирования связан с моделированием распределения в пространстве залежи разных литотипов. На этом этапе моделируется макронеоднородность. В наиболее простом и часто применяемом случае происходит расчленение пласта в разрезе на пропластки, а по площади их распространения – на участки коллектора и неколлектора. В более сложных случаях в каждом из пропластков может быть выделено несколько типов коллекторов. Количество типов коллекторов определяется требованием, чтобы каждый из типов обладал примерно однородным распределением фильтрационно-ёмкостных свойств.

Этап литологического моделирования заключается в специальной интерполяции на ячейки сетки данных о типах коллекторов со скважин. Важно отметить, что интерполяция должна осуществляться по сеточной модели *вдоль структуры*.

Это – второй этап геометризации.

### Моделирование свойств

Третий этап геометризации связан с моделированием значений фильтрационно-ёмкостных свойств. На этом этапе имеющаяся информация о фильтрационно-ёмкостных свойствах интерполируется на объём коллектора внутри залежи. Если на предыдущем этапе было выделено несколько типов коллекторов, то каждый из них моделируется по своей процедуре и со своими значениями параметров и данными. Так же, как и на этапе литологического моделирования, интерполяция осуществляется *вдоль пластов и пропластков*, но теперь при интерполяции параметров в ячейку учитываются только участки с одним типом коллектора.

## 6.2. ГЕОЛОГИЧЕСКИЕ ПОЛЯ

На всех этапах геометризации при геологическом моделировании возникает задача восстановления геологических полей, которая решается путём интерполяции. Геологические поля моделируются как набор значений параметра в узлах некоторой сетки. Сетка может быть двухмерной (карты и поверхности) или трёхмерной (кубы), соответственно, при двухмерном и трёхмерном моделировании.

Геологические поля классифицируют на **простые и сложные**.

**Простое поле** не прерывается и не имеет изломов, сложное поле – дискретно. Как правило, это является следствием того, что простое поле отражает изменение одного свойства, сложное – двух и более.

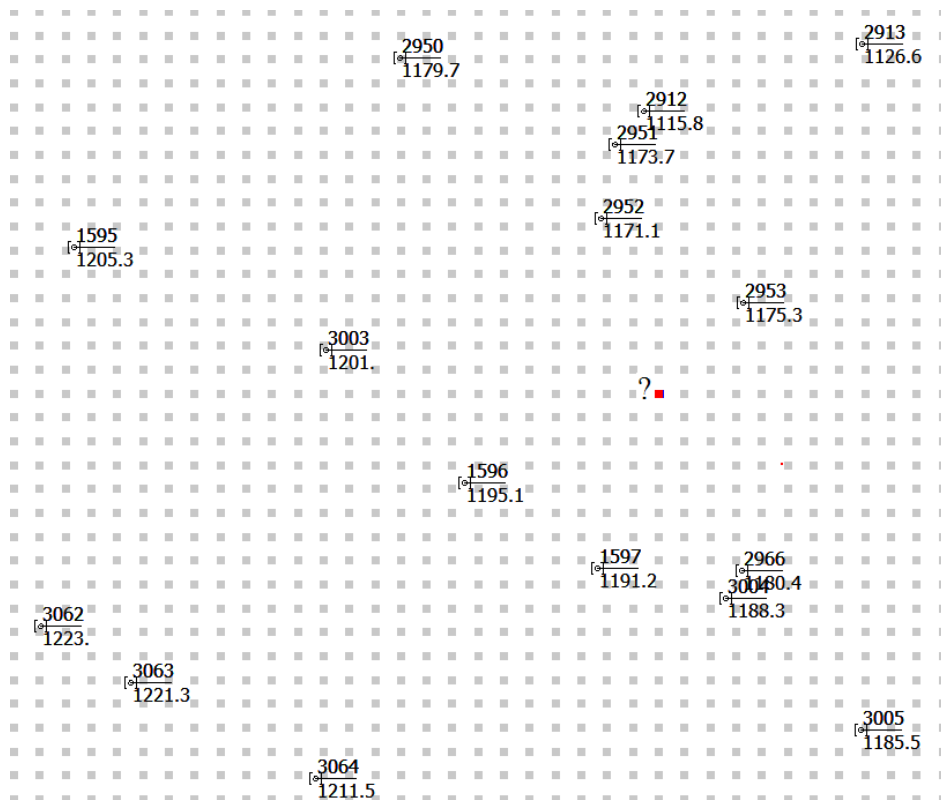
**Сложные поля** могут быть получены путем восстановления составляющих их простых полей с последующим синтезом сложных полей с помощью арифметических и логических операций над простыми полями.

Кроме этого, простые поля подразделяются на **арифметические (непрерывные)**, показывающие непрерывные изменения некоторой величины и **логические (дискретные)**, отражающие принадлежность той или иной точки к какой-то категории.

Рассмотрим пример на рис. 26. Здесь изображён ряд скважин с известными отбивками кровли пласта (они подписаны под номерами скважин). Задача восстановления поля значений абсолютной отметки кровли заключается в интерполяции этих значений в каждую из серых точек, расположенных по равномерной сетке на рисунке. С этой целью для каждой из точек (например, для помеченной) решается задача интерполяции.

Двухмерным случаем геологических полей являются поверхности топографического порядка.

**Поверхность топографического порядка** – это любая поверхность, характеризующая изменение параметра по изучаемой площади. К поверхностям топографического порядка относятся структур-



**Рис. 28. Восстановление кровли пласта по отбивкам в скважинах**

ные карты поверхностей, карты общих, эффективных, эффективных нефте(газо)насыщенных толщин, карты открытой пористости, проницаемости, нефте(газо)насыщенности и др.

Приведённое на рис. 28 представление основы для карты кровли пласта в виде сетки данных называется **растровым** представлением. Растровое представление удобно в тех случаях, когда нужно выполнять какие-то арифметические или логические операции. Когда речь идёт о поверхностях топографического порядка, для восприятия более удобно представление в изолиниях, называемое **векторным** представлением. Процедура перехода от растрового представления к векторному называется **векторизацией**, обратная – **растеризацией**.

Таким образом, построение изолиний поверхности топографического порядка по выборочным данным на ЭВМ проводится через векторизацию растрового представления этой поверхности. А для получения растрового представления необходима интерполяция данных выборки в каждый узел сетки раstra. При этом могут использоваться различные методы интерполяции.

### 6.3. МЕТОДЫ ИНТЕРПОЛЯЦИИ ДАННЫХ ПРИ ГЕОМЕТРИЗАЦИИ ПРОДУКТИВНЫХ ПЛАСТОВ

Геологов с первых шагов обучения знакомят с интерполяцией данных и даже с экстраполяцией. Однако этот, на первый взгляд, простой метод геометризации, базируется при геологических построениях, базируется на мощном математическом аппарате, который лежит в основе двух и трехмерного геологического моделирования.

Интерполяция применяется, когда имеются измерения некоторого параметра в одних точках пространства, а надо получить значения этого же параметра между точками, где его значения известны. Эта задача тем более важна для геометризации нефтяных и газовых залежей, поскольку расстояния между скважинами, в зависимости от стадии изученности и размеров месторождения, могут изменяться от нескольких сотен метров до нескольких километров.

#### Задача интерполяции

Допустим, что в интерполируемых точках двух- или трёхмерного пространства с векторными координатами  $q_1, \dots, q_n$  получены величины некоторого геологического параметра, который будем обозначать буквой  $z$ , –  $z_1, \dots, z_n$  (жирным шрифтом здесь и далее обозначены векторные величины, обычным – скалярные). В заданной точке с координатой  $q$  необходимо вычислить интерполированное значение параметра  $\hat{z}$ . При этом, если  $q$  совпадает с одной из интерполируемых точек, к примеру с  $q_i$ , то интерполированное значение  $\hat{z}$  должно совпасть со значением  $z_i$ .

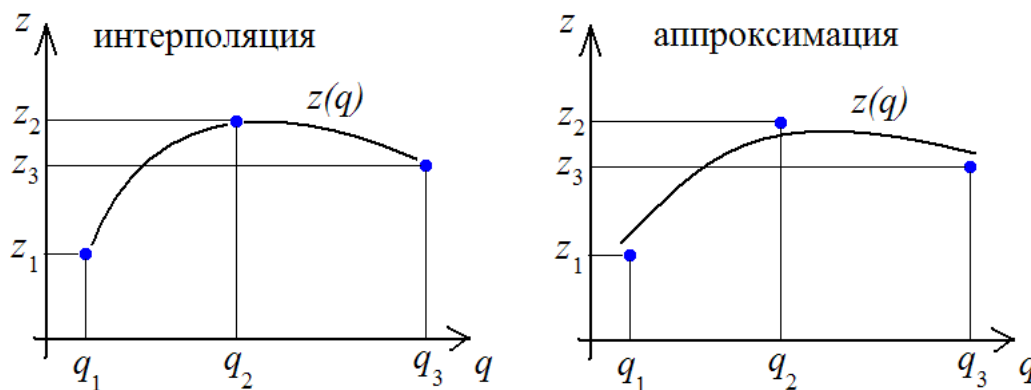
Другими словами, нужно так построить функцию  $\hat{z}(q) = \hat{Z}_{q_1, \dots, q_n}(q, z_1, \dots, z_n)$ , чтобы при этом выполнялось условие

$$\hat{z}(q_i) = z_i, \quad i = 1, \dots, n. \quad (86)$$

Иногда такое условие заменяют на более слабое:

$$\hat{z}(q_i) \approx z_i, i = 1, \dots, n, \quad (87)$$

в этом случае получаем аппроксимацию. Аппроксимация широко применяется при построении трендов.



**Рис. 29. Интерполяция и аппроксимация**

Ниже рассмотрим основные методы интерполяции, применяемые при геометризации.

### Скользящее среднее

Эта методика известна в статистике как метод статистического окна. Он применяется во всех программных продуктах геологического моделирования под названием скользящего среднего.

Чтобы вычислить значение в точке  $q$  берётся несколько ближайших точек и усредняется их значение. В качестве ближайших выбирается либо фиксированное количество ближайших точек, либо точки, оказавшиеся на расстоянии, меньшем заданного.

Обозначим количество ближайших точек через  $k$ , тогда  $\hat{z}$  находится по формуле

$$\hat{z} = \frac{1}{k} \sum_{i=1}^k z_{j(i)}, \quad (88)$$

где  $z_{j(1)}, \dots, z_{j(k)}$  —  $k$  ближайших точек к точке с координатами  $q$ .

Если количество ближайших точек равно нулю, то  $z$  устанавливается равным среднему  $z_1, \dots, z_n$ .

Рассмотрим пример на рис. 30. Для вычисления абсолютной отметки кровли пласта в красной точке выбрана область осреднения вокруг неё – круг радиусом шесть шагов сетки. В него попали 3 самые ближайшие скважины – 2952, 2951, 2953. В качестве интерполированного значения в красной точке берётся среднее арифметическое абсолютной отметки кровли в этих скважинах –  $(1171,1 + 1173,7 + 1175,3)/3 = 1173,37$  м.

Отметим, что на значение, полученное методом скользящего среднего в данной точке, одинаково сильно влияют все точки, взятые для осреднения, в том числе те, которые «загораживаются» другими. Например, скважина 2951 загорожена от целевой точки скважиной 2952, но влияет на результат интерполяции столь же сильно.

Метод скользящего среднего не обеспечивает выполнения условия формулы (86) и относится к методам аппроксимации, однако в комбинации с разнообразными приёмами «подтягивания» к значениям в интерполируемых точках очень широко используется на практике в силу быстроты, способности работать с трендами и ограниче-

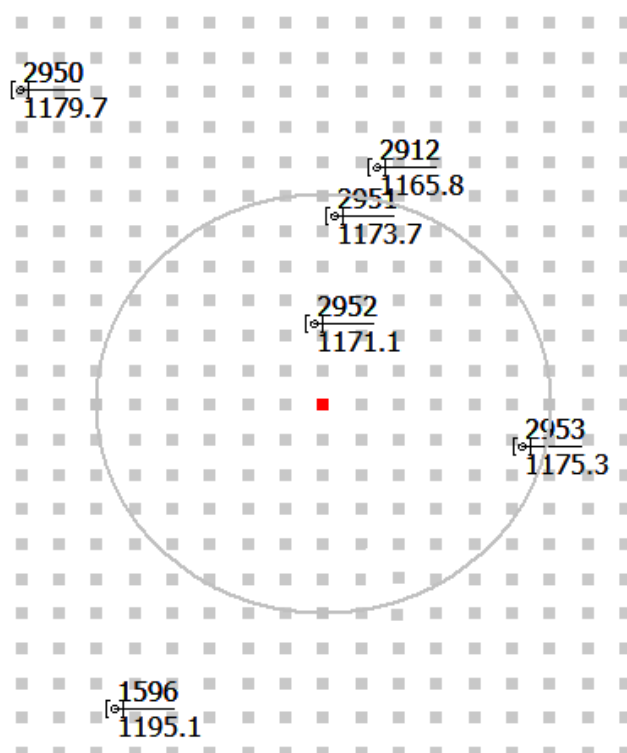


Рис. 30. Скользящее среднее

ниями, а также в силу простоты реализации. В чистом виде иногда используется для построения поверхности ВНК.

### Кусочно-линейная интерполяция

Сначала рассмотрим одномерный случай. Пусть точка, в которой надо получить интерполированное значение, имея координату  $x$ , находится между точками с координатами  $x_i$  и  $x_{i+1}$ . Выберем  $w_i = (x - x_{i+1}) / (x_{i+1} - x_i)$ ,  $w_{i+1} = (x_i - x) / (x_{i+1} - x_i)$ . Тогда функция окажется  $\hat{z}(x) = w_i(x)z_i + w_{i+1}(x)z_{i+1}$  отрезком прямой. Это и есть кусочно-линейная интерполяция. Её итогом будут кривые, состоящие из прямолинейных участков (рис. 31).

То есть для одномерного случая кусочно-линейная интерполяция заключается в том, что берутся две ближайшие к искомой точки и через них проводится прямая. В двух- или трёхмерном случае, чтобы вычислить значение в точке  $q$ , берутся ближайшие точки и через них строится гиперплоскость  $H$ , полагая что  $\hat{z}(q) = H(q)$ .

Для двухмерного случая плавная топографическая поверхность, полученная на основе кусочно-линейной интерполяции будет связанным набором плоских поверхностей.

Метод легко вычисляем для размерности пространства не выше 3. За пределами выпуклой оболочки не работает. Это единственный из рассматриваемых методов, для которого значения, загороженные от данной точки другими, никакого влияния на результат не оказывают.

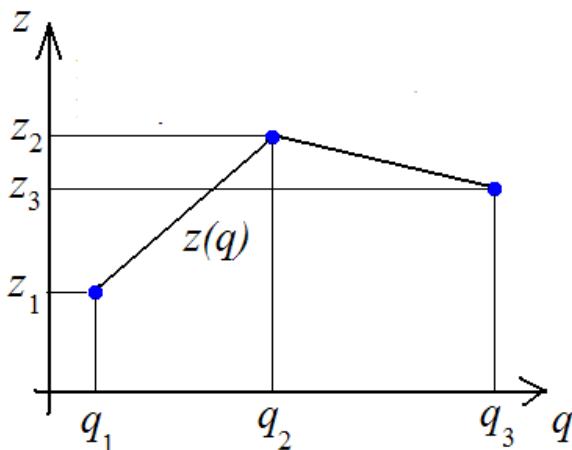


Рис. 31. Кусочно-линейная интерполяция

Важная область использования кусочно-линейных интерполяций – это геометризация разрывных нарушений, в частности, поверхностей разломов. Эта методика также широко применяется в различных пакетах геологического моделирования для восстановления карт по изолиниям.

### Разложение по базисным функциям

В этом случае выбирается набор базисных функций  $f_1, \dots, f_n$ , и значение в точке  $q$  находится по формуле:

$$\hat{z}(q) = \sum_{i=1}^n a_i f_i(q). \quad (89)$$

При этом коэффициент  $a_i$  находят, исходя из условия (86), решая соответствующую систему уравнений. Непрерывность и гладкость метода полностью определяется непрерывностью и гладкостью базисных функций.

В качестве базисных функций часто берут набор тригонометрических функций, в результате чего получается *интерполяция Фурье*. Она есть во всех пакетах геологического моделирования, например, AutoCorr, IRAP RMS, Petrel, DV-Geo и может применяться для построения структурных карт и иных поверхностей топографического порядка.

### Метод радиальных базисных функций

В этом методе выбирается **радиальная функция**  $c(l)$ , определённая на интервале  $l$  от 0 до  $+\infty$  и обычно стремящаяся в бесконечности к нулю. Функции  $f_i(r)$  записываются в виде:

$$f_i(q) = c(|q - q_i|) = c(l_i),$$

где  $l_i = |q - q_i|$  – расстояние от текущей точки до  $i$ -й точки выборки.

Формула (89) в этом методе приобретает вид



$$\hat{z}(q) = \sum_{i=1}^n a_i c(|q - q_i|), \quad (90)$$

а условия интерполяции (86) запишутся в виде:

$$\hat{z}(q) = \sum_{i=1}^n a_i c(|q - q_i|). \quad (91)$$

Решая последнее, находят коэффициенты  $a_i$  для (90).

На значения в интерполируемой точке  $q$  влияют все точки  $q_i$ , для которых  $c(|q - q_i|) > 0$ . Однако, если точка загоразивается другой, то её влияние существенно уменьшается. При детерминированном моделировании метод радиальных базисных функций эквивалентен методу кригинга и является основным методом интерполяции в пакете AutoCorr.

### Сплайн-интерполяция

Сплайн-интерполяция пришла из области компьютерной графики, где через заданные точки надо было провести гладкую поверхность. Для получения такой поверхности использовалась метафора гибкой упругой пластины: поверхность строилась так, чтобы формой повторяла упругую пластину, закреплённую в заданных точках (см. рис. 32).

Для получения формы такой сплайн-поверхности можно выписать уравнения равновесия упругой пластины. После ряда преобразований (см. [1]) для двумерного случая получится следующее решение:

$$\hat{z}(q) = \sum_{i=1}^n a_i c(|q - q_i|) + a_{n+1} + a_{n+2} \cdot x + a_{n+3} \cdot y,$$

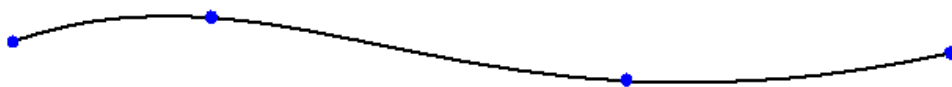


Рис. 32. Гибкая упругая линейка, закреплённая в синих точках, принимает форму сплайна

где  $c(l) = l^2 \ln(l)$ ,  $x$  и  $y$  – абсцисса и ордината вектора  $q$ . Значения коэффициентов  $a_i$  находят из системы  $n + 3$  линейных уравнений:

$$z_j = \sum_{i=1}^n a_i c(|q_j - q_i|) + a_{n+1} + a_{n+2} \cdot x_j + a_{n+3} \cdot y_j, \quad j = 1, \dots, n,$$

$$\sum_{i=1}^n a_i = \sum_{i=1}^n a_i \cdot x_i = \sum_{i=1}^n a_i \cdot y_i = 0.$$

Видно, что вычислительно метод имеет очень много общего с методом радиальных базисных функций, отличаясь от него в следующем:

1) к (90) прибавляется  $a_{n+1} + a_{n+2} \cdot x + a_{n+3} \cdot y$ ;

2) к системе (91) добавляются уравнения  $\sum_{i=1}^n a_i = \sum_{i=1}^n a_i \cdot x_i =$

$$= \sum_{i=1}^n a_i \cdot y_i = 0..$$

В остальном сплайн интерполяция является частным случаем метода радиальных базисных функций, когда в качестве радиальной функции используется функция  $l^2 \ln(l)$ .

Сплайн-интерполяция присутствует во многих пакетах геологического моделирования и используется для интерполяции структурных поверхностей.

### Метод обратных расстояний

Широкий класс методов посвящён определению значений в интерполируемой точке по формуле:

$$\hat{z}(q) = \sum_{i=1}^n w_i(q) z_i, \quad (92)$$

Назовём такие методы методами **взвешивания значений**.

Если значения взвешивающих коэффициентов  $w_i(q)$  вычислять по формуле

$$w_i(q) = |q - q_i|^{-k} / \sum_{j=1}^n |q - q_j|^{-k} = \frac{1/l_i^k}{\sum_{j=1}^n 1/l_j^k}, \quad (93)$$

где  $l_i = |q - q_i|$  – расстояние от текущей точки до  $i$ -й точки выборки,  $k$  – некоторое натуральное число (обычно 2), то легко доказать, что  $\hat{z}(q)$  удовлетворяет системе (86).

Интерполяция по формулам (92) и (93) называется **методом обратных расстояний**. Гладкость интерполяции методом обратных расстояний зависит от величины  $k$  в формуле (93). В частности, интерполяция будет гладкой при  $k = 2$  и негладкой при  $k = 1$ . Вдали от интерполируемых точек интерполяция по этому методу стремится к среднему значению по выборке. Стремление к среднему будет наблюдаться как при движении к границам изучаемой области, так и между данными редких разведочных скважин. Вот почему в неизученных зонах категории  $C_2$  модель разреза пласта может получиться однородной, тогда как рядом, в изученных зонах, по данным эксплуатационных скважин пласт оказывается расчленённым и прерывистым.

Метод обратных расстояний простой и быстрый в вычислительном смысле. При этом виде интерполяции значение в точке определяется влиянием остальных. Сила влияния обратно пропорциональна расстоянию в выбранной степени. Однако, недостаток этого метода в том, что «загороженная» точка влияет точно так же, как и «открытая», находящаяся на том же расстоянии. Так, при интерполяции данным методом (рис. 28) на значение в обведённой точке скважина 1 влияет так же, как скважина 5, хотя первая «загораживается» скважиной 3. Модифицированный метод обратных расстояний широко используется в пакете геологического моделирования DV-Geo.

**Пример.** Имеется 5 скважин (рис. 33), чьи условные координаты приведены в табл. 12. Там же приведены общие толщины исследуемого пласта в этих скважинах.

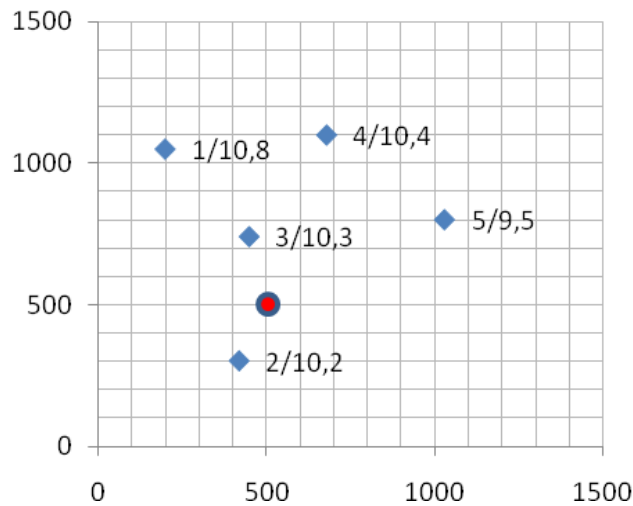


Рис. 33

Таблица 12

Номер скважины	$x$	$y$	$h, \text{ м}$
1	200	1050	10,8
2	420	300	10,2
3	450	740	10,3
4	680	1100	10,1
5	1030	800	9,5

Необходимо методом обратных расстояний ( $k = 2$ ) найти интерполированное значение толщины в точке  $x = 500, y = 500$  (обведена кружком). Составим следующую таблицу.

Таблица 13

Номер скважины	$(x_i - x)^2$	$(y_i - y)^2$	$1/l_i^2 = 1/((x_i - x)^2 + (y_i - y)^2)$	$w_i = \frac{1/l_i^2}{\sum_{j=1}^n 1/l_j^2}$	$w_i h_i$
1	90000	302500	0,0008	0,193355	2,088232
2	6400	40000	0,0013889	0,335685	3,423992
3	2500	57600	0,0008403	0,203104	2,091969
4	32400	360000	0,0005618	0,135783	1,371407
5	280900	90000	0,0005464	0,132073	1,254693
$\Sigma$			0,0041375		10,23029

В результате интерполированное значение общей толщины рассматриваемого пласта в точке (500×500) равно 10,23 м.

### Свойства методов интерполяции

Выше было перечислено пять часто применяемых методов интерполяции. На самом деле их больше и результаты, полученные разными методами, могут заметно отличаться. Поэтому при геометризации продуктивных пластов важно правильно подобрать тот метод интерполяции, свойства которого наиболее точно соответствуют решаемой задаче. Для изучения и сравнения свойств методов интерполяции введём следующие классификационные признаки.

1) Непрерывность зависимости  $\hat{z}(q)$ , гладкость полученной интерполяции.

2) Линейность отображения  $z_1, \dots, z_n \rightarrow z(q)$ , т.е. линейность  $\hat{Z}_{q_1, \dots, q_n}(q, z_1, \dots, z_n)$  по аргументам  $z_1, \dots, z_n$ .

Отображение  $\hat{Z}(q, z_1, \dots, z_n)$  является линейным по аргументам  $z_1, \dots, z_n$ , если соблюдаются следующие два условия:

а) для произвольного  $\alpha$ ,  $\hat{Z}(q, \alpha z_1, \dots, \alpha z_n) = \alpha \hat{Z}(q, z_1, \dots, z_n)$ ,

б)  $\hat{Z}(q, z_1 + z'_1, \dots, z_n + z'_n) = \hat{Z}(q, z_1, \dots, z_n) + \hat{Z}(q, z'_1, \dots, z'_n)$ .

Согласно а), если взять значения  $z$  в других единицах, то интерполированные значения пересчитаются в эти же единицы. Свойство б) означает, что карта от суммы значений равна сумме карт значений.

Метод интерполяции, обладающий свойством линейности, можно назвать **линейным методом интерполяции**, или **линейной интерполяцией**. (Не путать с кусочно-линейной интерполяцией: *кусочно-линейная интерполяция лишь один из видов линейной интерполяции*).

3) Выполнение принципа экстремальности интерполируемых значений:

$$\min(z_1, \dots, z_n) \leq \hat{z}(q) \leq \max(z_1, \dots, z_n).$$

Это свойство гарантирует, что в результате интерполяции не появится дополнительных плохо объяснимых минимумов или максимумов.

На рис. 34 слева изображена интерполяция трёх точек с помощью метода, обладающего свойством экстремальности интерполируемых значений. При этом кривая есть результат интерполяции. Все её точки лежат в области между минимальным и максимальным интерполируемым значением. Справа на этом рисунке показана интерполяция тех же трёх точек методом, не обладающим свойством экстремальности. В результате часть кривой, как результата интерполяции, выходит за максимальное значение в интерполируемых точках.

#### 4) Несмещённость.

Если значения во всех интерполируемых точках одинаковы, то и результат интерполяции будет с ними совпадать:

Если  $z_1 = \dots = z_n = z_0$ , то  $\hat{z}(q) = z_0$ .

#### 5) Возможность применения к данным с трендом.

#### 6) Легкость вычисления.

7) Поведение на границах области (экстраполирующее поведение).

Можно показать, что все перечисленные методы относятся к линейным. Скользящее среднее, кусочно-линейная интерполяция, сплайн интерполяция и метод обратных расстояний являются несмещёнными. Скользящее среднее, кусочно-линейная интерполяция и ме-

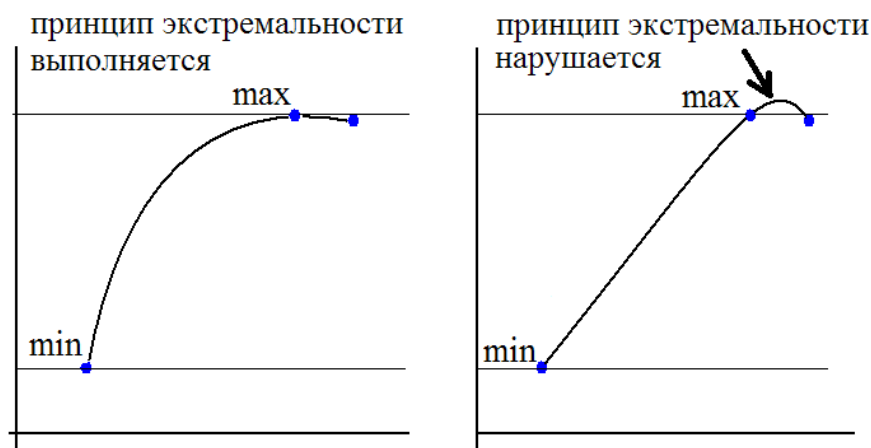


Рис. 34. Особенности принципа экстремальности

Метод	Непрерывный/ гладкий	Линейный	Экстремальность интерполируе- мых значений	Несмещённость	Работает с трен- дами	Быстрый	Экстраполяционное поведение
Скользящее средне	-/-	+	+	+	+	+	Зависит от реализации
Кусочно-линей- ный	+/-	+	+	+	+	+	Не работает
Интерполяция Фурье	+/+	+	-	-	-	-	Регулярное повторение
Метод радиаль- ных базисных функций	Зав. от функц.	+	-	-	-	-	Сходится к среднему
Сплайн интерпо- ляция	+/+	+	-	+	+	-	Продолжается по ли- нейному тренду
Метод обратных расстояний	$+/k = 2$	+	+	+	+	+	Сходится к среднему

Метод обратных расстояний обладают свойством экстремальности интерполируемых значений. В табл. 14 сведены основные данные по методам интерполяции.

### Закон распределения результатов интерполяции

Линейную интерполяцию можно применять только к случайным величинам, распределение которых не противоречит нормальному закону. Нельзя интерполировать линейными методами случайные величины, закон распределения которых далёк от нормального. Это следствие свойства **нормальности результатов интерполяции**:

*распределение значений, полученных в результате применения линейного метода интерполяции, стремится к нормальному закону распределения.*

Все применяющиеся на практике методы интерполяции являются

линейными. Следовательно, *если интерполируемые величины распределены по закону, отличному от нормального, надо предварительно преобразовывать распределение к нормальному.*

Например, при интерполяции значений проницаемости, имеющей логнормальное распределение, следует сначала взять логарифмы значений проницаемости, затем проинтерполировать их и лишь после этого вычислить экспоненту интерполированных значений. Этот пример – частный случай использования **нормализующего преобразования (Normal Score)**. Для любого однородного распределения подбирается преобразование, приводящее интерполируемые величины к нормальному распределению. Затем производится интерполяция нормализованных величин. Последним шагом является преобразование, обратное к нормализующему, применённому на первом шаге, переводящее нормальное распределение к исходному распределению интерполируемых величин.

При неоднородном распределении, обычно связанным с присутствием в интерполируемых данных разных литотипов, использование нормализующего преобразования приводит к геологически неверной картине пространственного распределения свойств. По этой причине, если распределение параметра в выборке неоднородно, то этот параметр нельзя интерполировать ни на прямую, ни с использованием нормализующего преобразования.

Чтобы решить проблему интерполяции неоднородного параметра, необходим этап литологического моделирования, на котором моделируемый объект разбивается на зоны литотипов, внутри каждой из которых распределение параметра будет однородным. Таким образом, литологическое моделирование требуется для процедуры интерполяции.

Описанные методы интерполяции не могут быть напрямую применимы при литологическом моделировании, поскольку их результатом являются поля, а поле распределения литотипов является полем логического типа.



Методика восстановления логического поля с помощью интерполяции называется **индикаторным моделированием**. Обычно оно заключается в том, что интерполируют непрерывную случайную величину – *вероятность* обнаружения того или иного литотипа в разных частях продуктивного пласта. В итоге рассчитанное *арифметическое поле вероятности* литотипов используется для приписывания того или иного литотипа разным ячейкам и создания логического поля литотипов.

## **7. ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКОЙ ЗАВИСИМОСТИ МЕЖДУ ЗНАЧЕНИЯМИ ОДНОГО ПАРАМЕТРА В РАЗНЫХ ТОЧКАХ ПРИ ГЕОМЕТРИЗАЦИИ ПРОДУКТИВНОГО ПЛАСТА**

### **7.1. ЗАВИСИМОСТЬ МЕЖДУ ЗНАЧЕНИЯМИ ОДНОГО ПАРАМЕТРА В РАЗНЫХ ТОЧКАХ**

Между значениями одного параметра в разных частях залежи может быть или отсутствовать статистическая связь. Эта связь может быть сильной или слабой, прямолинейной или криволинейной. Для характеристики этой связи могут быть использованы коэффициент ковариации, коэффициент корреляции, корреляционное отношение точно так же, как и для случая двух параметров. Однако, особенностью статистической связи между значениями одного и того же параметра, измеренными в разных точках, является её зависимость от расстояния между точками, в которых было проведено измерение. Качественный характер такой зависимости достаточно очевиден – между близкими точками высокая скоррелированность значений, между значениями в далёких точках статистическая зависимость отсутствует.

Этот раздел посвящен статистическим методам изучения зависимости между пространственно-разделёнными значениями одного геологического параметра (например, в разных скважинах), позволяющим количественно описать зависимость силы статистической связи между значениями одного параметра от расстояния. Эта зависимость очень важна для построения наиболее верной интерполяции и оценке её точности. При выполнении интерполяции вручную исследование подобной зависимости является очень трудоёмким делом, поэтому результаты, изложенные здесь, применяются только при моделировании месторождений углеводородов с использованием компьютерных средств.

Наличие статистической связи между значениями параметра  $z$  в одной точке и значениями в точке, отстоящей от неё на определённое расстояние, позволяет оценивать значения в неизвестных точках между ними. Построение и правильное использование таких пространственно-статистических зависимостей является важнейшим элементом, на котором основана интерполяция и на который опираются такие современные методики, как кригинг и стохастическое моделирование. В геостатистике для отражения подобной зависимости используется функция, называемая **вариограмма**.

## 7.2. ВАРИОГРАММА

Вариограмма параметра – функция, показывающая, как нарастает различие между значениями параметра с увеличением расстояния между точками, в которых эти значения измеряются (рис. 35).

Вариограмма  $\gamma(l)$  определяется по формуле:

$$\gamma(l) = \frac{1}{2}M[(z(q) - z(q+l))^2], \text{ где } |I| = l; M - \text{знак математического}$$

ожидания.

Кроме вариограммы иногда используется противоположенная ей функция – **функция пространственной ковариации** (рис. 35), которая находится по формуле:

$$c(l) = \gamma(\infty) - \gamma(l).$$



Рис. 35. Вариограмма и функция пространственной ковариации

Если вариограмма показывает, как растёт различие между значениями параметра с увеличением расстояния, то функция пространственной ковариации показывает, как падает сходство.

Свойства вариограммы и функции пространственной ковариации:

1) в нуле вариограмма равна нулю, а функция пространственной ковариации максимальна и равна дисперсии параметра  $z$ :

$$\gamma(0) = 0, \quad c(0) = D(z);$$

2) вариограмма – монотонно возрастающая функция, а функция пространственной ковариации – монотонно убывающая функция;

3) начиная с какого-то расстояния  $R$ , вариограмма выходит на порог  $s = \gamma(\infty) = c(0) = D(z)$ , с этого же расстояния функция пространственной ковариации становится равной нулю. Это расстояние называется **радиусом вариограммы**.

Радиус вариограммы – важнейшая характеристика, описывающая влияние пространственно разделённых значений параметра друг на друга. Значение в данной точке влияет только на значения в тех точках, которые к ней ближе, чем радиус вариограммы. Из этого следует: *для построения значения параметра в неизвестной точке можно использовать только те точки выборки, которые находятся к ней ближе, чем радиус вариограммы; если точки расположены дальше такого радиуса, то выборка не содержит информации, необходимой для построения искомого значения.*

Вариограмма достаточно просто строится по исходным данным. Для этого определяется размер класса по формуле:

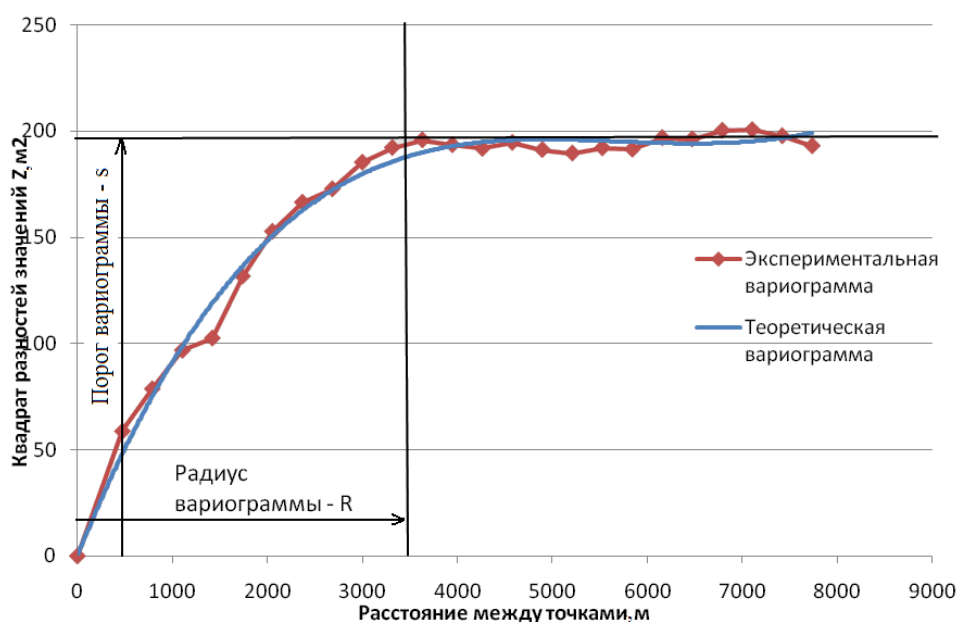
$$\Delta = \frac{R_{\max}}{10 \lg n},$$

где  $R_{\max}$  – максимальное расстояние между точками. Все расстояния между парами имеющихся точек разбиваются на классы (в первый класс попадают пары точек, оказавшиеся на расстоянии  $0-\Delta$ , во второй – на расстоянии  $\Delta-2\Delta$  и т.д.). Для каждого класса рассчитывается средний квадрат разности значений попавших в него пар.

Далее строится кривая зависимости половины среднего квадрата разности значений от расстояний между парами интерполируемых точек  $l$ . Она обычно имеет характер возрастающей кривой, выполаживающейся на некотором максимуме. Это и есть экспериментальная вариограмма (рис. 36).

Для расчетов используют не экспериментальную вариограмму, а её приближение при помощи одной из теоретических моделей или их комбинаций. Теоретические модели это простые для вычисления функции, зависящие от порога вариограммы  $s$  (который совпадает с величиной  $c(0)$ ) и соответствующего ему радиуса вариограммы  $R$ . В качестве теоретических моделей чаще всего используют ([16]) следующие модели (см. рис. 37–40).

Ещё раз напомним, что «рабочим», наиболее важным участком вариограммы (и функции пространственной ковариации) является диапазон от нуля до радиуса вариограммы. Особенно важно то, как вариограмма отходит от нуля. Следует обратить внимание, что на этом начальном участке Гауссова вариограмма растёт наиболее плавно, а экспоненциальная наиболее резко, сферическая вариограмма за-



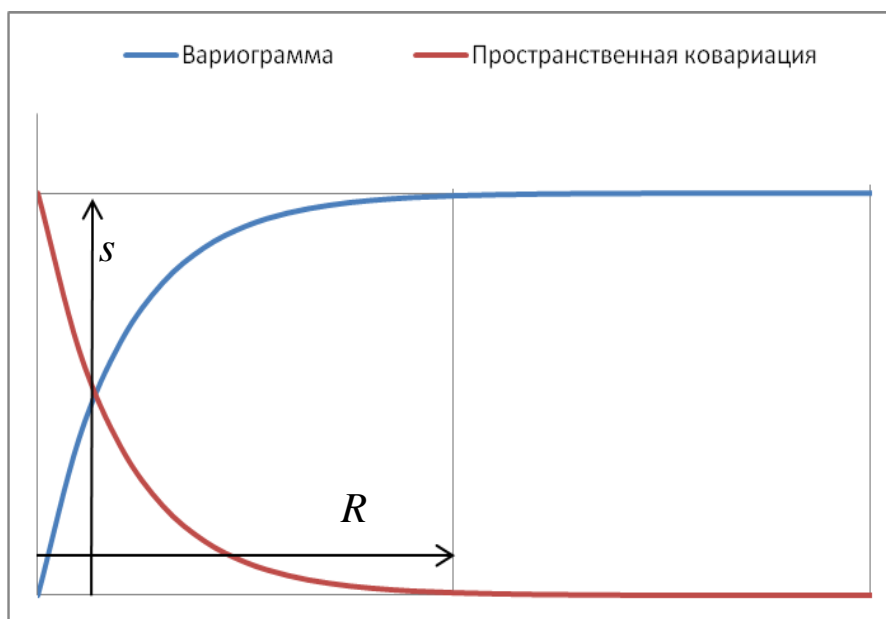
**Рис. 36. Пример экспериментальной вариограммы и её теоретического приближения, рассчитанных по 1300 стратиграфическим отметкам кровли продуктивного горизонта**

нимает промежуточное положение, что характеризуется углом наклона вариограммы.

Описанные модели относятся к нормальной «рабочей» ситуации. Есть, однако, модель, получение которой по экспериментальным данным должно сразу насторожить.

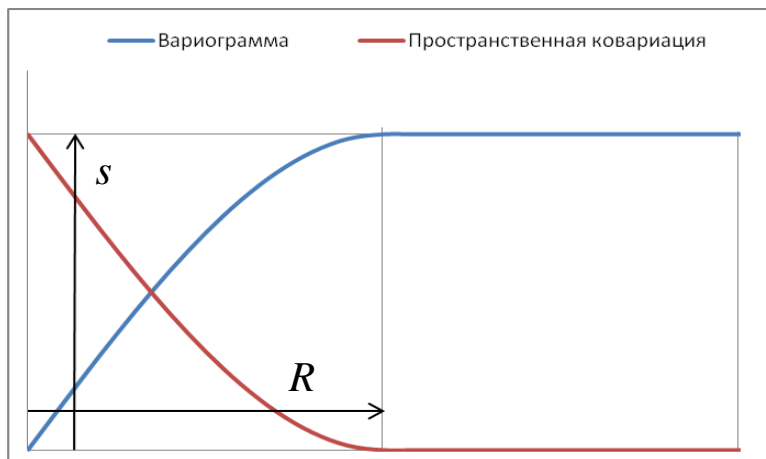
Если оказалось, что к построенной экспериментальной вариограмме ближе всего модель белого шума, то это означает, что изменчивость параметра настолько велика, что от одной точки выборки до другой он многократно меняется от больших до маленьких значений и обратно, а точки выборки слишком далеки друг от друга, чтобы с их помощью можно было выловить эти изменения. Именно такую картину мы видим на рис. 41. На нём изображены два параметра. Один, изображённый слева, быстро и хаотично меняется, другой, изображённый справа, меняется плавно и заметно более предсказуемо. Если построить вариограмму по точкам скважин (чёрные точки обведённые красным), в случае 41а получим вариограмму белого шума, а в случае 41б «нормальную» вариограмму.

Интерполяция данных по скважинам при вариограмме белого шума бессмысленна. Математика в данной ситуации бессильна.



**Рис. 37. Экспоненциальная модель:**

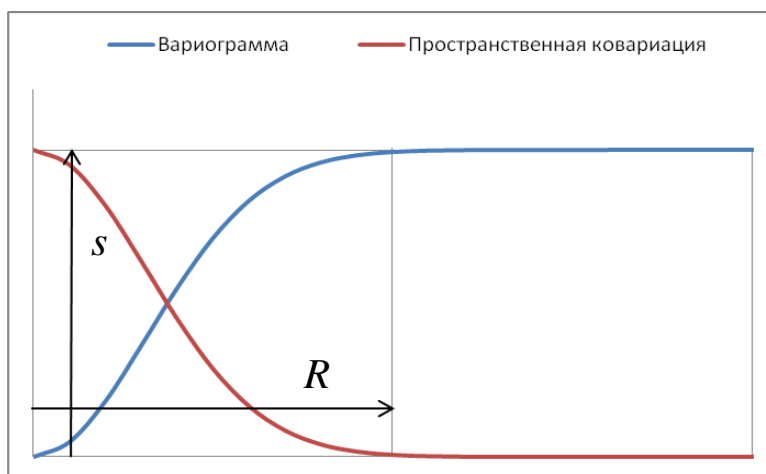
$$\gamma(l) = s(1 - e^{-3l/R}), \quad c(l) = se^{-3l/R}$$



**Рис. 38. Сферическая модель:**

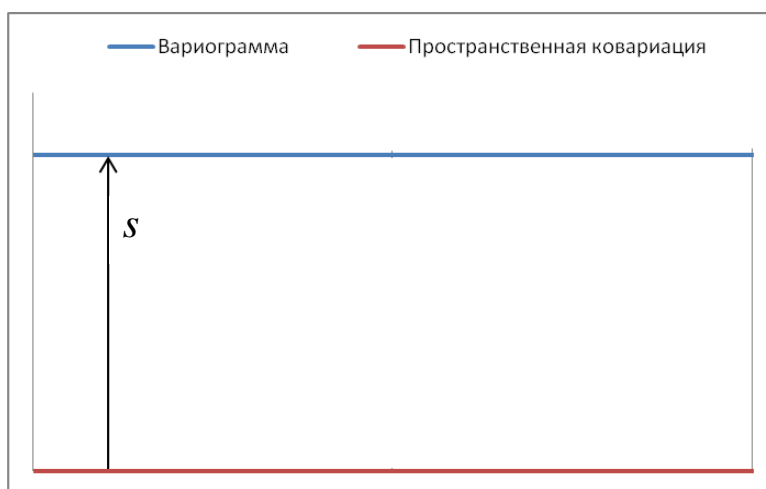
$$\gamma(l) = s \left( \frac{3l}{2R} - \frac{l^3}{2R^3} \right), \quad c(l) = s \left( 1 - \frac{3l}{2R} + \frac{l^3}{2R^3} \right), \quad \text{при } l < R$$

$$\gamma(l) = s, \quad c(l) = 0, \quad \text{при } l > R.$$



**Рис. 39. Гауссова модель:**

$$\gamma(l) = s(1 - e^{-3l^2/R^2}), \quad c(l) = se^{-3l^2/R^2}$$



**Рис. 40. Модель белого шума:**

$$\gamma(l) = 0, \quad c(l) = s, \quad \text{при } l = 0; \quad \gamma(l) = s, \quad c(l) = 0, \quad \text{при } l > 0.$$

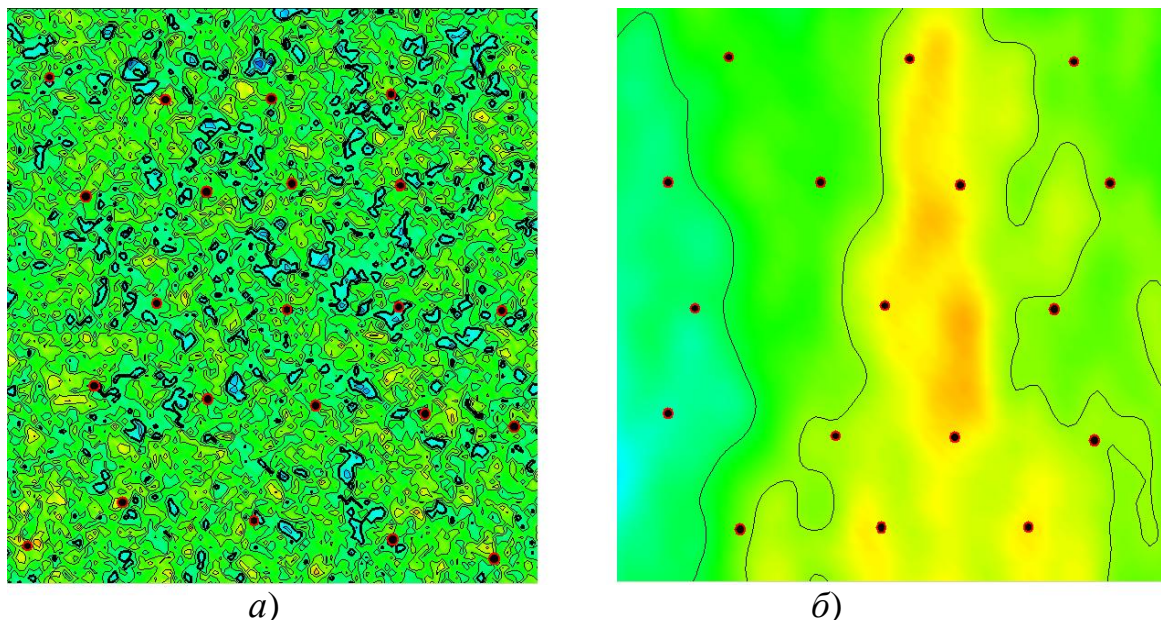


Рис. 41. Высокоизменчивый и низкоизменчивый параметры

### 7.3. КРИГИНГ

Кригинг – это название группы линейных методов интерполяции, которые используют вариограмму для построения интерполируемых значений. Кригинг (kriging) – интерполяция впервые предложена в 1951 году D.G. Krige. Процедура интерполяции методом кригинга устроена так, чтобы получающаяся среднеквадратическая ошибка интерполяции была минимальной среди линейных методов интерполяции. Таким образом, *из всех методов интерполяции, кригинг – интерполяция даёт наиболее точный результат при условии использования правильно построенной вариограммы.*

Кригинг интерполяция применима только к данным без тренда. Точнее говоря, интерполируемые данные должны отвечать условиям стационарности. Последнее означает, что если взять достаточно большие области в разных частях исследуемого объекта, то и средние значения интерполируемого параметра по этим областям и его дисперсии не должны иметь статистически значимых различий.

Кригинг делят на простой, обычный и универсальный. Интерполяцию **простым кригингом** (simple kriging) можно использовать



лишь в случае *нулевого математического ожидания* интерполируемой величины. В случае, когда *математическое ожидание интерполируемой величины известно*, можно использовать обычный кригинг (ordinary kriging). Метод **универсального кригинга** (universal kriging) применим к параметрам с неизвестным заранее математическим ожиданием и используется на практике наиболее часто.

Уравнения универсального кригинга записываются через вариограмму в виде:

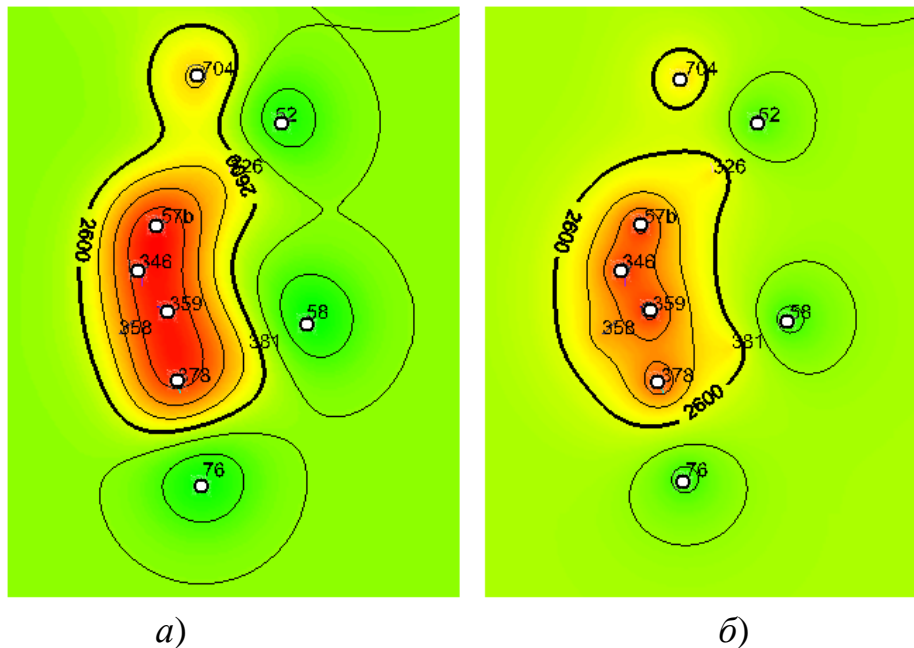
$$\begin{pmatrix} \gamma(|q_1 - q_1|) & \dots & \gamma(|q_1 - q_n|) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(|q_n - q_1|) & \dots & \gamma(|q_n - q_n|) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1(q) \\ \vdots \\ w_n(q) \\ \tau(q) \end{pmatrix} = \begin{pmatrix} \gamma(|q - q_1|) \\ \vdots \\ \gamma(|q - q_n|) \\ 1 \end{pmatrix} \quad (94)$$

Вычислив из линейной системы (100) величины  $w_1(q), \dots, w_n(q)$ , интерполированное значение находят по формуле

$$\hat{z}(q) = \sum_{i=1}^n w_i(q) z_i. \quad (95)$$

Результат интерполяции методом универсального кригинга зависит от того, какая теоретическая вариограмма использована. Всегда нужно использовать ту теоретическую вариограмму, которая наиболее точно воспроизводит экспериментальную, построенную по выборке. Если реальных данных немного и по ним трудно определиться с типом теоретической вариограммы, то следует помнить, что Гауссова модель даёт наиболее гладкую интерполяцию, экспоненциальная наиболее изменчивую, а сферическая занимает промежуточное положение между ними.

На рис. 42а изображена стратиграфическая поверхность кровли пласта, полученная путём кригинг-интерполяции отметок в скважинах с использованием Гауссовой вариограммы. На рис. 42б те же отметки проинтерполированы с помощью экспоненциальной вариограммы.



**Рис. 42. Использование а) гауссовой и б) экспоненциальной вариограммы**

Отметим, что не любую функцию можно использовать в качестве теоретической вариограммы. Поэтому при подборе типа вариограммы следует использовать только известные виды вариограмм и их комбинации.

Различные модификации кригинг-интерполяции позволяют дополнительно к интерполяции искомого параметра учитывать влияние на него других известных параметров. Так, **кригинг с внешним дрейфом** может быть использован при построении стратиграфической поверхности по отбивкам в скважинах с учётом данных близко расположенной сейсмоотражающей поверхности. **Кокригинг** используют, например, для интерполяции скважинных данных о фильтрационно-ёмкостных свойствах с учётом сейсмических атрибутов, если между ФЕС и сейсмическими атрибутами установлена корреляционная связь на основе проведения соответствующих исследований.

## 7.4. ТРЕНДЫ И ЭКСТРАПОЛЯЦИЯ

Трендом называется наличие в выборке выраженной тенденции к изменению тех или иных свойств по площади или разрезу. Например, тенденция к более высоким значениям коэффициента нефтенасыщенности в более высоко расположенных частях залежи.

Кригинг-интерполяция предназначена для интерполяции данных без трендов. Если в данных тренды есть, то кригинг, да и многие другие методы интерполяции, к таким данным не применимы.

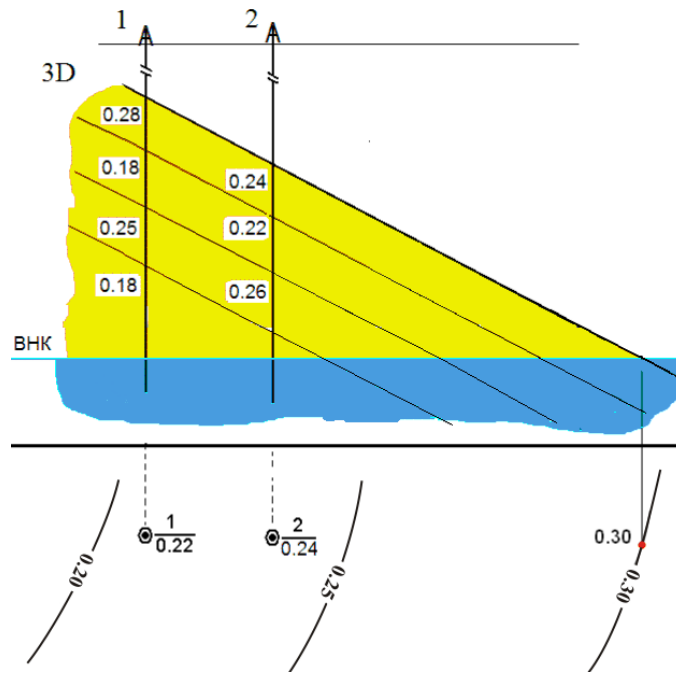
При построении экспериментальной вариограммы набора данных с трендом её вид будет отличаться от того, что показано на рисунках выше. Для таких данных вариограмма либо не выходит на порог, либо оказывается немонотонной функцией, поднимающейся до порогового значения, а потом резко падающей. Такие особенности вариограммы свидетельствует о наличии неслучайных регулярных особенностей в данных.

В этих ситуациях следует произвести предварительную обработку данных, представив их в виде суммы неслучайного тренда и отклонений от тренда, которые получаются вычитанием из исходных данных полученного тренда. По отклонениям строится вариограмма. По ней производится кригинг-интерполяция отклонений и, финально, результат интерполяции складывается с трендом. При построении тренда нужно опираться не столько на математические методы, сколько на геологические соображения.

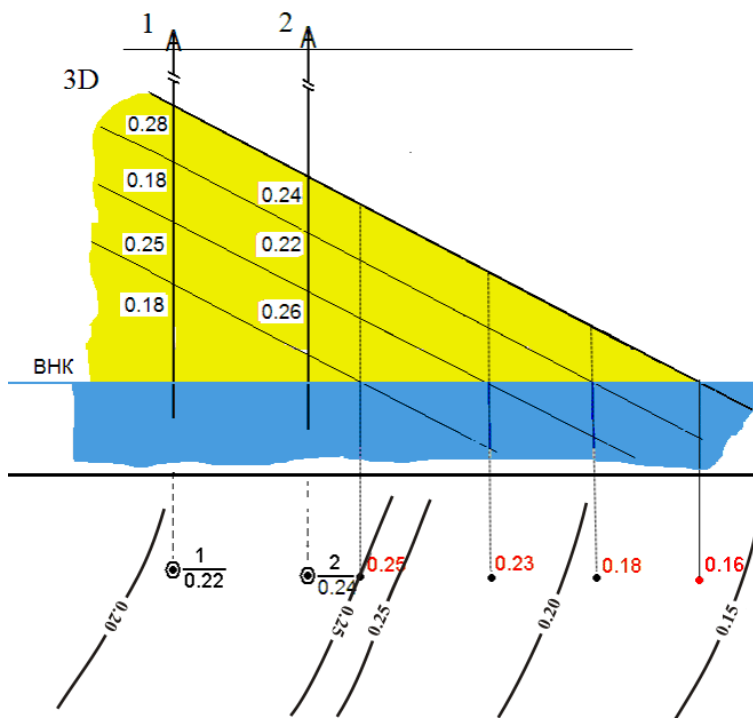
Использование трендов позволяет решить задачу экстраполяции. Разделение данных на тренд и отклонения от него дают возможность экстраполировать тренд за пределы той области, где есть данные.

Экстраполяцией следует пользоваться с особой осторожностью. В этом случае знание и понимание геологии особенно важно, поскольку неучёт особенностей строения продуктивного пласта может привести к ошибочным результатам.

Рассмотрим пример изображённый на рис. 43, 44, где показан



**Рис. 43. Пористость на контуре, полученная экстраполяцией тренда на карте**



**Рис. 44. Пористость на контуре полученная путём экстраполяции тренда вдоль стратиграфической кровли в 3D**

неоднородный по пористости пласт коллектора в зоне внешнего контура. На рисунках видно, что пористость верхнего слоя уменьшается от скважины 1 к скважине 2 до 0,16 на внешнем контуре. Пористость остальных слоев, наоборот, возрастает в направлении к контуру.

Вместе с тем, если не обращать внимание на значения параметра отдельных слоёв, а воспользоваться средними значениями по скважинам, то тренд будет однозначно показывать возрастание параметра пористости в сторону внешнего контура до 0,30, что иллюстрируется рис. 43. Чтобы получить более объективную картину, необходимо провести границы слоёв до пересечения с ВНК и на каждом пересечении определить среднее значение по трём, двум, одному прослою. В соответствии с трендом устанавливается значение пористости на внешнем контуре 0,16 (рис. 44).

### **Доверительные интервалы результатов интерполяции**

Допустим, что при планировании бурения новой скважины поставлена задача пробурить её в области максимальных общих толщин. Если выявлять области высоких значений общей толщины пласта по результатам интерполяции, то может оказаться, что они лежат весьма далеко от изученной части залежи и велик риск того, что на самом деле толщины там окажутся весьма незначительными. Заметно лучшей идеей является планирование бурения не только с учётом интерполированных значений общих толщин, но и с учётом доверительного интервала, в котором они могут лежать. Знание вариограммы (или функции пространственной ковариации) позволяет оценить доверительный интервал для результатов интерполяции.

Рассмотрим линейный метод интерполяции. Можно показать, что он всегда представим в виде некоторого метода взвешивания значений

$$\hat{z}(q) = \sum_{i=1}^n w_i(q) z_i,$$

где  $\hat{z}(q)$  – интерполированное значение (оценка)  $z(q)$  – истинного значения параметра в точке  $q$ . Обозначим через  $\Delta z(q) = z(q) - \hat{z}(q)$  ошибку интерполяции. Тогда среднеквадратическое значение ошибки интерполяции, которое обозначим через  $E(q)$ , оценивается как:

$$E(q) = \sqrt{M[\Delta z^2]} =$$

$$= \sqrt{D(z) - 2 \sum_{i=1}^n w_i(q) c(|q_i - q|) + \sum_{j=1}^n \sum_{i=1}^n w_j(q) w_i(q) c(|q_i - q_j|)}, \quad (96)$$

где  $M$  – математическое ожидание,  $D(z)$  – дисперсия параметра  $z$ . Следует отметить, что дисперсия интерполированных значений  $D(\hat{z}(q))$  всегда меньше или равна дисперсии интерполируемого параметра  $D(z)$ .

Для универсального кригинга среднеквадратическая ошибка интерполяции вычисляется по значительно более простой формуле:

$$E(q) = \sqrt{\sum_{i=1}^n w_i(q) \gamma(|q - q_i|) + \tau(q)}. \quad (97)$$

Эта формула может быть получена из (96) после ряда преобразований, учитывающих что  $c(l) = D(z) - \gamma(l)$  и что  $w_i(q)$  и  $\tau(q)$  являются решениями системы универсального кригинга (94).

Зная среднеквадратическую ошибку интерполяции, можно вычислить доверительный интервал значений  $z$  в любой точке. Так как процедура интерполяции должна применяться только к нормально распределённым случайным величинам, результат интерполяции также распределён нормально. Ошибка интерполяции  $\Delta z(q)$  также распределена по нормальному закону, поскольку является разностью двух случайных величин с нормальным законом распределения. Если применяемый метод интерполяции обладает свойством несмещённости, то математическое ожидание значения  $\Delta z(q)$  равно нулю. Более того,  $E(q)$  есть не что иное, как дисперсия величины  $\Delta z(q)$ . Следовательно, закон распределения ошибки интерполяции даётся формулой:

$$f(\Delta z) = \frac{1}{E\sqrt{2\pi}} e^{-\frac{\Delta z^2}{2E^2}}. \quad (98)$$

Вычислив значение  $E(q)$ , можно рассчитать, с какой вероятностью величина  $\Delta z$  попадёт в интервал ширины  $2a$ , симметрично расположенный относительно нуля. Согласно (41) искомая вероятность находится по формуле  $p = 2\Phi_0(a / E(q))$ . Исходя из этого, можно найти величину  $a$  при заданной вероятности  $p$  попадания в интервал:

$$a = \Phi_0^{-1}(p / 2)E(q),$$

где  $\Phi_0^{-1}$  – функция, обратная нормированной функции Лапласа, в частности, для  $p = 0,954$ ,  $\Phi_0^{-1}(p / 2) = 2$ . Из последней формулы следует, что для заданной доверительной вероятности  $p$ , доверительный интервал значений  $\Delta z(q)$  равен:

$$-\Phi_0^{-1}(p / 2)E(q) < \Delta z(q) < \Phi_0^{-1}(p / 2)E(q).$$

Зная доверительный интервал для ошибки интерполяции, легко подсчитать доверительный интервал для самой величины  $z$ :

$$\hat{z}(q) - \Phi_0^{-1}(p / 2) \cdot E(q) < z(q) < \hat{z}(q) + \Phi_0^{-1}(p / 2) \cdot E(q). \quad (99)$$

**Пример.** Определим доверительный интервал для примера на с. 78 с пятью скважинами (см. табл. 12). Составим таблицу квадратов расстояний  $-l^2$  от скважин, до интересующей нас точки (500, 500):

Таблица 14

Номер скважины	$l^2$
1	392500
2	46400
3	60100
4	392400
5	370900

А также таблицу квадратов расстояний между скважинами.

Таблица 15

Номер скважины	1	2	3	4	5
1	0	610900	158600	232900	751400
2	610900	0	194500	707600	622100
3	158600	194500	0	182500	340000
4	232900	707600	182500	0	212500
5	751400	622100	340000	212500	0

По результатам исследований принята следующая функция пространственной ковариации

$$c(l) = se^{-3l^2/R^2}, \quad (100)$$

где  $s = 0,25$ ,  $R = 2000$  м.

Выпишем, опираясь на эту функцию и табл. 14, коэффициенты ковариации между общими толщинами в скважинах и данными в целевой точке (500,500). В табл. 16 приведены значения функции  $c(l)$ , полученные подстановкой в формулу (100) значений  $l^2$  из табл. 15.

Таблица 15

Номер скважины	$c(l)$
1	0,18624927
2	0,24144964
3	0,23898145
4	0,18626324
5	0,18929108

Также выпишем коэффициенты ковариации между скважинами, используя формулу (100) и табл. 16.

Таблица 16

Номер скважины	1	2	3	4	5
1	0,25	0,158109	0,221963	0,209932	0,142296
2	0,158109	0,25	0,216066	0,147048	0,156787
3	0,221963	0,216066	0,25	0,21802	0,193729
4	0,209932	0,147048	0,21802	0,25	0,213169
5	0,142296	0,156787	0,193729	0,213169	0,25



Используя столбец  $w_i$  из табл. 13, данные табл. 15 и 16, получим по формуле (96):

$$E(q = (500, 500)) = 0,135 \text{ м.}$$

Отсюда по формуле (99) вычисляется доверительный интервал для общей толщины пласта в точке (500, 500), подсчитанный методом обратных расстояний и соответствующий 95,4 % доверительной вероятности:

$$10,23 - 2 \cdot 0,135 < h < 10,23 + 2 \cdot 0,135;$$

$$9,96 \text{ м} < h < 10,5 \text{ м.}$$

При оценке истинного значения в результате интерполяции получена величина 10,23 м. Эта величина принимается в качестве значения общей толщины пласта в точке (500, 500). Истинное значение толщины лежит где-то в интервале от 9,96 до 10,5 м.

## 8. СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

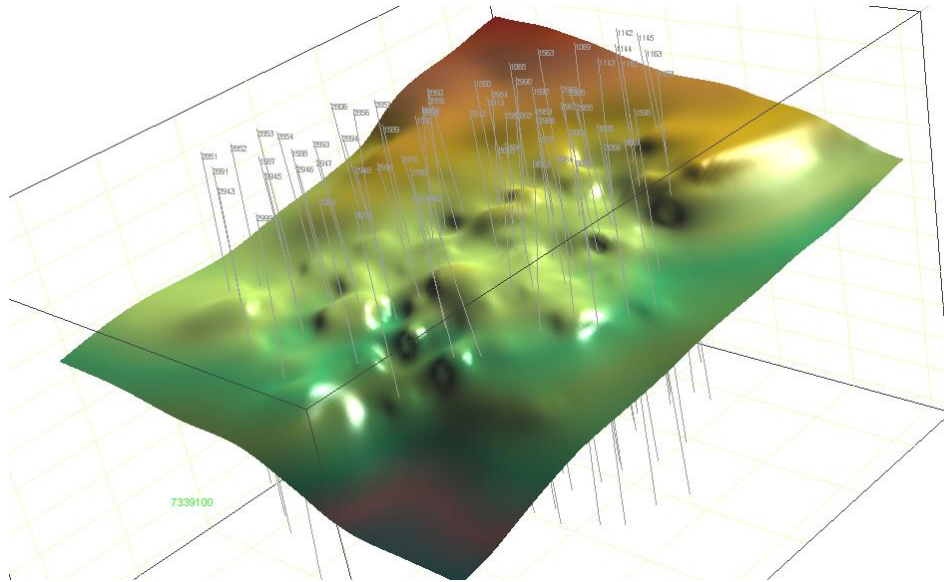
### 8.1. ДЕТЕРМИНИРОВАННАЯ ИНТЕРПОЛЯЦИЯ И СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Всё что говорилось ранее, относится к детерминированной интерполяции. В частности, кригинг-интерполяция даёт некоторое «математическое ожидание» значения величины в точке. При этом поведение интерполированных значений может очень сильно отличаться от характера поведения реальных данных. В частности, далеко от интерполируемых значений результаты интерполяции сходятся к среднему значению. То есть в результате интерполяции данных «вдалеке» от них получится «стол», выровненный по среднему значению, а в тех областях, где расположены интерполируемые точки, изменчивость результатов интерполяции будет высокой.

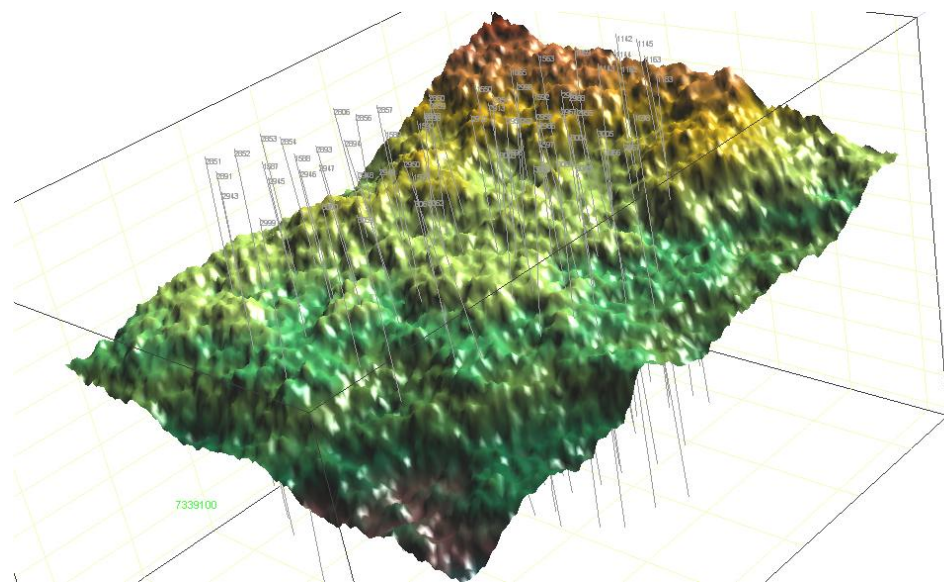
Так, при детерминированной интерполяции отметок кровли пласта в разбуренной части получаем карту кровли, на которой будет много небольших осложнений в виде куполов, мульд, флексур, прогибов и др., а в слабо разбуренной результатом интерполяции будет гладкая ровная поверхность. То есть проинтерполированная детерминированным методом кровля пласта будет изменчивой в области скважин и ровной в слабо разбуренной части (рис. 45). Характерный размер особенностей видимых на детерминированной модели имеет порядок расстояния между скважинами.

Стохастическое моделирование – методика получения набора стохастических реализаций. **Стохастическая реализация** – это вариант интерполяции, который везде ведёт себя с изменчивостью, соответствующей модели изменчивости разбуренной части залежи. Модель изменчивости определяется вариограммой и задаёт амплитуду деталей того или иного размера.

Если вместо детерминированной интерполяции построить стохастические реализации поверхности кровли, то они окажутся одинаково



**Рис. 45. Детерминированная интерполяция поверхности кровли по скважинным данным (предоставлено Е.В. Ковалевским)**



**Рис. 46. Стохастическая реализация поверхности кровли по скважинным данным (предоставлено Е.В. Ковалевским)**

изменчивы как в сильно разбуренной, так и в слабо разбуренной частях залежи. При стохастическом моделировании сохраняется характер изменчивости везде, но сами реализации будут дальше от наиболее вероятных значений, чем результаты детерминированной интерполяции (см. рис. 46). На стохастической реализации появляются детали меньшие, чем расстояние между скважинами. Эти детали явля-

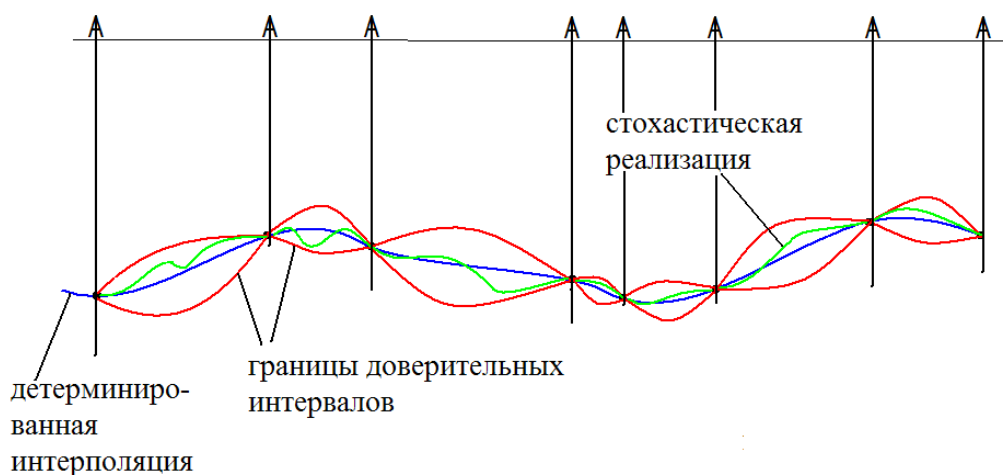
ются результатом симуляции изменчивости в мелких масштабах и не основаны на исходных данных.

*Задача стохастического моделирования – создать набор реализаций, каждая из которых сохраняет все стохастические связи, обнаруженные при исследовании выборки. В частности, вариограмма каждого из моделируемых параметров стохастической реализации должна совпадать с теоретической вариограммой, подобранной по исходным данным. В случае использования детерминированных методов интерполяции вариограммы интерполированных параметров никогда не совпадают с вариограммами исходных данных. Первые всегда более пологи и имеют значительно бóльший радиус вариограммы.*

Таким образом, результаты детерминированной интерполяции не могут соответствовать стохастической модели, содержащей пространственные статистические зависимости, т.е. вариограммы.

Стохастическое моделирование и детерминированная интерполяция связаны следующим образом:

*при правильном стохастическом моделировании пространственное распределение усреднённого по множеству реализаций значения каждого моделируемого параметра должно быть близко к результату детерминированной интерполяции.*



**Рис. 47. Построение стратиграфической поверхности по отбивкам в скважинах детерминированным и стохастическим методами**

Геометрически это выражается в том, что стохастические реализации меняются внутри доверительного интервала вокруг значений, полученных детерминированной интерполяцией (рис. 47). Другая трактовка этого свойства: если построить карты распределения какого-то параметра для большого количества стохастических реализаций и из них создать осреднённую по реализациям карту, то она должна совпадать с результатом детерминированной интерполяции.

## 8.2. ПОСЛЕДОВАТЕЛЬНОЕ ГАУССОВО СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

В большинстве программных продуктов для геологического моделирования в качестве метода стохастического моделирования используется последовательное Гауссово стохастическое моделирование, которое тесно связано с кригинг-интерполяцией.

Пусть имеются точки, в которых значения нам известны (известные точки) и множество точек, значение в которых необходимо смоделировать (моделируемые точки). Тогда основной шаг метода выглядит следующим образом:

а) случайно выбрать одну из моделируемых точек –  $q$ ;

б) используя кригинг-интерполяцию, рассчитать интерполированное значение  $\hat{z}(q)$  в этой точке и среднеквадратическую ошибку интерполяции  $E(q)$ ;

в) установить  $z(q)$  равным значению нормально распределённой с параметрами  $m = \hat{z}(q)$  и  $\sigma = E(q)$  случайной величины и полученному при помощи генератора случайных чисел;

г) добавить точку  $q$  к списку известных точек и, убрав из списка моделируемых, вернуться к первому пункту.

Гауссово стохастическое моделирование, как и другие подобного рода методы, приводит к нормально распределённым результатам моделирования. Следовательно, напрямую стохастическое моделиро-

вание применимо только к нормально распределённым моделируемым значениям, например, значения открытой пористости. Если параметр не имеет нормального распределения, но распределён однородно, то стохастическое моделирование применяется в комбинации с нормализующим преобразованием. Если параметр распределён неоднородно, то необходим предварительный этап литологического моделирования, который может быть проведён с использованием как детерминированных, так и стохастических индикаторных методов.

В случае применения стохастических методов для литологического моделирования, появляется возможность воспользоваться **объектным моделированием**. Объектное моделирование – это метод стохастического литологического моделирования, заключающийся в заполнении модели участками литотипов, геометрия которых напоминает геометрию естественных геологических объектов, например, русел рек или прибрежных баров. Заполнение ведётся при помощи стохастических алгоритмов, но так, чтобы результат соответствовал скважинным данным. Реализации, полученные методом объектного моделирования, должны совпадать в окрестности скважин и различаться в межскважинном пространстве. Анализ метода объектного моделирования с точки зрения математических свойств весьма затруднен, поэтому применяя объектное моделирование следует полагаться, прежде всего, на знание геологии и здравый смысл.

### 8.3. МЕТОД МОНТЕ-КАРЛО

Метод Монте-Карло представляет собой метод определения параметров сложных стохастических систем на основе подобранной статистической модели по данным выборочного распределения, полученного по семейству реализаций [14]. В итоге семейство реализаций должно быть сформировано таким образом, чтобы его вероятностные характеристики совпадали с аналогичными величинами решаемой задачи, т.е. на основе стохастической модели.

Метод Монте-Карло применяется для оценки геологических запасов. В этом случае основой статистической модели служат выявленные статистические связи и вариограммы моделируемых параметров. Семейство стохастических реализаций следует строить с помощью методов стохастического моделирования, в частности, Гауссова.

При подсчёте запасов детерминированными методами, как правило, получают одно значение, которое, вроде бы, отражает реальное количество углеводородов в пласте. При оценке стохастическими методами и методом Монте-Карло получают гистограмму распределения запасов и доверительный интервал, в котором лежат цифры запасов с заданной доверительной вероятностью. Использование метода Монте-Карло тесно связано с оценкой рисков при разработке месторождения. Под **риском** подразумевается возможность получения результатов деятельности, отличных от ожидаемых или планируемых, как в меньшую, так и большую сторону.

Итак, сущность метода Монте-Карло состоит в:

- подборке правильной статистической модели, корректно описывающей распределение запасов как случайной величины с учётом всех вариограмм, связей и корреляций между параметрами;
- генерации реализаций в соответствии со статистической моделью;
- подсчёте запасов для каждой реализации;
- анализе полученного распределения запасов и определения их среднего значения и границ изменения.

Проиллюстрируем работу метода Монте-Карло на простом примере. Допустим, необходимо оценить геологические запасы нефти чисто нефтяной зоны некоторого пласта. Геологические запасы подсчитываются объёмным методом по формуле (см. [7]):

$$Q = Fh_{\text{н}}k_{\text{п}}k_{\text{н}}\theta\rho, \quad (101)$$

где  $F$  – площадь проекции залежи или её части на горизонтальную плоскость,  $\text{м}^2$ ;  $h_{\text{н}}$  – вертикальная проекция средней эффективной неф-

тенасыщенной толщины пласта, м;  $k_{\text{п}}$  – средний коэффициент пористости, доли единиц;  $k_{\text{н}}$  – средний коэффициент нефтенасыщенности, доли единиц;  $\theta$  – пересчётный коэффициент, доли единиц;  $\rho$  – плотность нефти в стандартных условиях, кг/м<sup>3</sup>.

В данном примере  $F = 20000000$  м<sup>2</sup>,  $\theta = 0,896$ ,  $\rho = 850$  кг/м<sup>3</sup>.

### Подборка статистической модели для метода Монте-Карло

Статистическая модель определяет, какие величины можно считать случайными, как они распределены и как связаны между собой. При трёхмерном геологическом моделировании наиболее важной частью статистической модели являются вариограммы моделируемых параметров, однако в рассматриваемом упрощённом примере вариограммы не нужны.

В связи с ограниченным числом наблюдений таких параметров, как плотность нефти и пересчётный коэффициент, воспользуемся их величинами, указанными выше. Также постоянным примем значение площади залежи  $F$ . Случайные величины  $h_{\text{эн}}$ ,  $k_{\text{по}}$  и  $k_{\text{н}}$  непрерывно меняются по площади залежи и оцениваются по выборке из скважин. В результате изучения параметров пласта было установлено, что наиболее адекватной моделью, отражающей распределение величин  $h_{\text{эн}}$ ,  $k_{\text{по}}$  и  $k_{\text{н}}$  является нормальное распределение с параметрами указанными в таблице.

Кроме этого, учтём, что указанные величины не являются независимыми – между ними есть положительная корреляция: большим зна-

Таблица 17

Величина	$\bar{x}$ , м	$S$ , м
$h_{\text{н}}$	$\bar{h}_{\text{н}} = 10$	$S_{h_{\text{н}}} = 0,5$
$k_{\text{п}}$	$\bar{k}_{\text{п}} = 0,22$	$S_{k_{\text{п}}} = 0,02$
$k_{\text{н}}$	$\bar{k}_{\text{н}} = 0,65$	$S_{k_{\text{н}}} = 0,02$



чениям эффективной толщины обычно соответствуют более высокое значение пористости, а её лучшим значениям – более высокие коэффициенты нефтенасыщенности. Для упрощённой учебной модели попарные коэффициенты корреляции  $r$  между параметрами имеют значения порядка 0,8. Отметим, что значения, записанные во втором столбце табл. 17 это как раз те величины, которые должны быть использованы при подсчёте запасов по формуле (101) детерминированным методом.

### Генерация реализаций в соответствии с выбранной моделью

Чтобы оценить запасы методом Монте-Карло, сгенерируем в соответствии с описанной выше моделью 30 реализаций случайных произведений троек  $h_n, k_n, k_n$  и для каждой из них подсчитаем геологические запасы по формуле (101).

Для генерации  $h_n, k_n, k_n$  распределённых в соответствии с описанной выше моделью, поступим следующим образом. Сгенерируем три независимые случайные переменные  $x_1, x_2, x_3$  распределённые по нормальному закону с нулевым математическим ожиданием и дисперсией, равной 1. Для этого используем комбинацию функций табличного редактора НОРМСТОБР(СЛЧИС()). В стохастической модели величины  $h_n, k_n, k_n$  скоррелированы с коэффициентом  $r$ . Чтобы реализации были также скоррелированы, необходимо иметь скоррелированные случайные величины, тогда как  $x_1, x_2, x_3$  независимые. Чтобы получить скоррелированные случайные величины, преобразуем  $x_1, x_2, x_3$  в величины  $y_1, y_2, y_3$  с помощью следующих линейных преобразований:

$$y_1 = x_1; \quad y_2 = r \cdot x_1 + \sqrt{1-r^2} \cdot x_2; \quad y_3 = r \cdot x_1 + r \sqrt{\frac{1-r}{1+r}} \cdot x_2 + \sqrt{1-\frac{2r^2}{1+r}} \cdot x_3.$$

Величины  $y_1, y_2, y_3$ , так же как и  $x_1, x_2, x_3$ , распределены по нормальному закону с нулевым математическим ожиданием и единичной

дисперсией, но не являются независимыми, а попарно скоррелированы с коэффициентом корреляции  $r$ . Отметим, что для получения соответствующих линейных преобразований в общем случае, надо найти разложение Холецкого ковариационной матрицы, но эти вопросы выходят далеко за пределы данного курса.

Таблица 18

Номер реал.	$x_1$	$x_2$	$x_3$	$y_1$	$y_2$	$y_3$	$h_n$ , м	$k_n$	$k_n$	$Q$ , тыс. т
	-0,2763	0,5708	0,1354	-0,2763	0,1214	0,0040	9,86	0,2224	0,6501	21721
	-0,3539	-0,0082	-0,3956	-0,3539	-0,2880	-0,4980	9,82	0,2142	0,6400	20517
	1,0173	1,2372	-0,6546	1,0173	1,5562	0,7919	10,51	0,2511	0,6658	26765
	0,8565	-0,9101	1,3257	0,8565	0,1392	1,1551	10,43	0,2228	0,6731	23820
	0,6822	0,6847	1,1836	0,6822	0,9565	1,3645	10,34	0,2391	0,6773	25511
	1,4292	-1,7013	-0,8887	1,4292	0,1225	0,2120	10,71	0,2225	0,6542	23752
	0,1643	-0,3172	-1,1666	0,1643	-0,0589	-0,5802	10,08	0,2188	0,6384	21453
	0,6101	0,8989	0,1543	0,6101	1,0274	0,8107	10,31	0,2405	0,6662	25155
	0,4016	-1,4552	-1,2840	0,4016	-0,5518	-0,7569	10,20	0,2090	0,6349	20613
	0,8646	-1,2606	0,8929	0,8646	-0,0647	0,8355	10,43	0,2187	0,6667	23170
	-1,8419	1,5279	-0,4699	-1,8419	-0,5568	-1,3187	9,08	0,2089	0,6236	18013
	0,3981	0,6611	0,7266	0,3981	0,7151	0,8853	10,20	0,2343	0,6677	24304
	-1,8535	0,7905	1,1139	-1,8535	-1,0085	-0,6733	9,07	0,1998	0,6365	17579
	1,4490	0,6992	0,5140	1,4490	1,5787	1,6219	10,72	0,2516	0,6824	28045
	1,0040	-0,6967	-0,9145	1,0040	0,3852	0,1259	10,50	0,2277	0,6525	23768
	1,1707	-0,0350	1,3970	1,1707	0,9155	1,6781	10,59	0,2383	0,6836	26265
	-0,2906	0,3451	1,8734	-0,2906	-0,0254	0,8665	9,85	0,2195	0,6673	21987
	-0,9951	-0,2472	-1,3276	-0,9951	-0,9444	-1,5756	9,50	0,2011	0,6185	18004
	-0,3545	-0,4485	-0,0537	-0,3545	-0,5527	-0,4320	9,82	0,2089	0,6414	20051
	-1,2788	-0,1314	-0,2446	-1,2788	-1,1019	-1,1895	9,36	0,1980	0,6262	17675
	0,2183	2,6438	0,3039	0,2183	1,7609	1,0430	10,11	0,2552	0,6709	26364
	-0,7397	-0,8520	1,2593	-0,7397	-1,1030	-0,1422	9,63	0,1979	0,6472	18790
	0,4133	0,5834	-0,0982	0,4133	0,6807	0,4334	10,21	0,2336	0,6587	23922
	-1,3525	-1,3159	0,0971	-1,3525	-1,8716	-1,3808	9,32	0,1826	0,6224	16137
	-0,0358	-0,8988	0,5537	-0,0358	-0,5679	0,0293	9,98	0,2086	0,6506	20639
	-0,6041	-0,2949	-1,2852	-0,6041	-0,6602	-1,2527	9,70	0,2068	0,6249	19091
	-1,8738	-0,3242	-0,7916	-1,8738	-1,6936	-2,0110	9,06	0,1861	0,6098	15668
	-0,6898	-1,9115	0,4522	-0,6898	-1,6988	-0,8186	9,66	0,1860	0,6336	17335
	-1,0178	0,3005	0,1207	-1,0178	-0,6340	-0,7990	9,49	0,2073	0,6340	19003
	-0,3717	1,7112	-1,3760	-0,3717	0,7294	-0,5805	9,81	0,2346	0,6384	22387
Среднее по 30-и реализациям							9,95	0,2182	0,6486	21583

На основе величин  $y_1, y_2, y_3$  подсчитаем  $h_{ЭН}, k_{ПО}$  и  $k_H$  по формулам:

$$h_H = S_{h_H} y_1 + \bar{h}_H,$$

$$k_{II} = S_{k_{II}} y_2 + \bar{k}_{II},$$

$$k_H = S_{k_H} y_3 + \bar{k}_H.$$

и запишем результаты в табл. 18.

Средние значения по 30 реализациям, вычисленные в последней строке таблицы, с точностью до 0,5 % совпадают со средними в табл. 18, что свидетельствует о корректности приведённого стохастического ансамбля реализаций.

### Анализ результатов подсчёта запасов методом Монте-Карло

По данным табл. 18 построим гистограмму распределения запасов (рис. 48) и кумулятивную кривую (рис. 49).

Теперь можно оценить не только «среднее» значение запасов – 21500 тыс.т, но и возможный разброс значений. Из рис. 49 видно, как определить границы доверительного интервала, в который запасы попадают с 80%-ной вероятностью: минимальную – 17000 тыс.т и максимальную 26000 тыс. т. Заметим, что среднее значение, полученное методом Монте-Карло, 21500 тыс.т. отличается от детерминированной оценки – 21782 тыс. т. менее чем на 2 %.

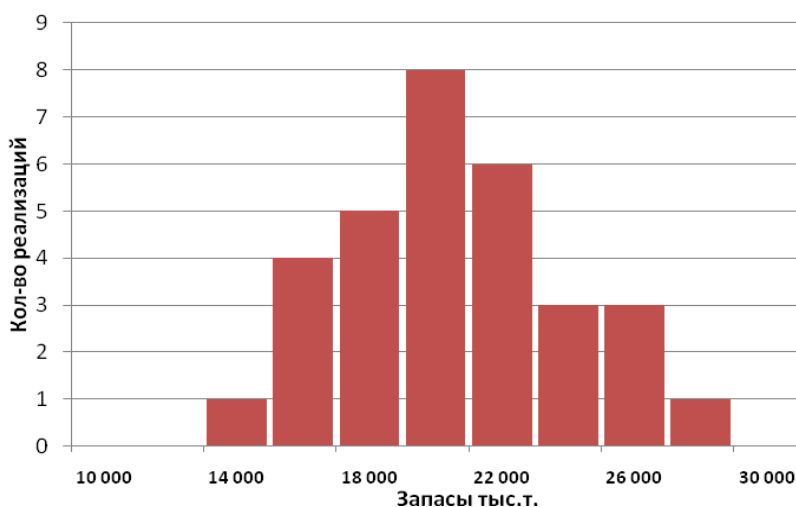
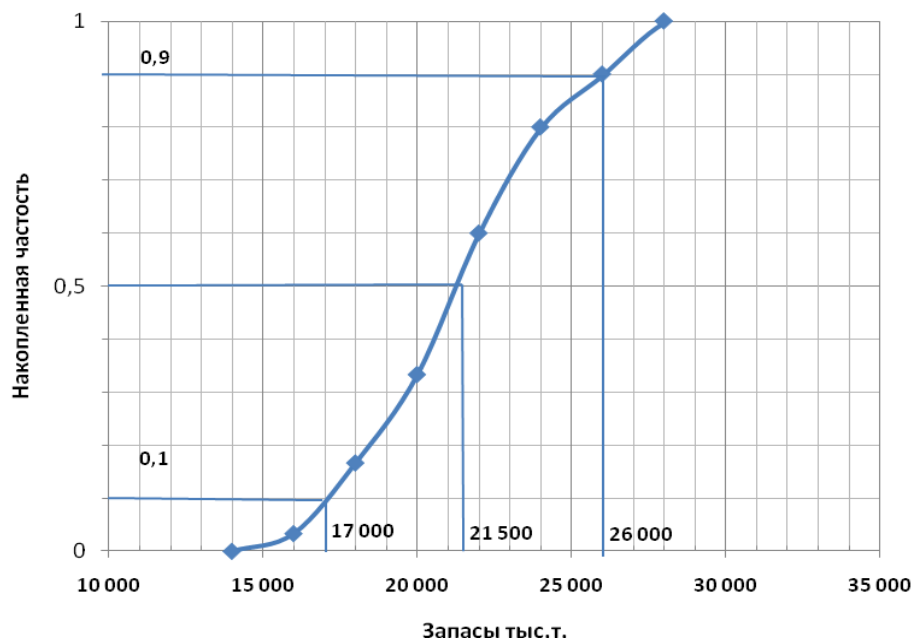


Рис. 48. Гистограмма распределения геологических запасов



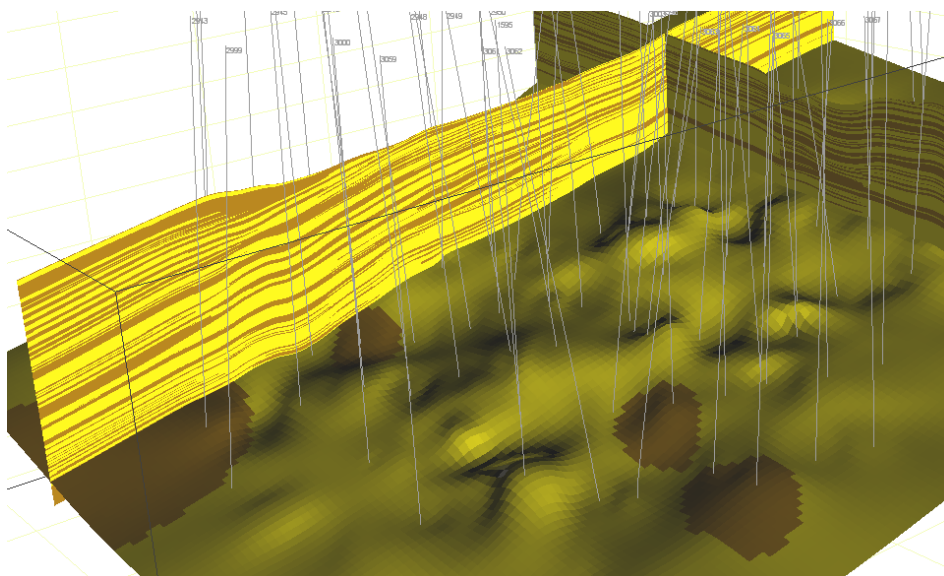
**Рис. 49. Кумулятивная кривая распределения геологических запасов**

Таким образом, проделав в 30 раз больше работы, мы получаем практически тот же результат. *Такая ситуация типична для подсчёта геологических запасов методом Монте-Карло – огромный объём работы приводит практически к тем же средним результатам, что и намного менее трудоёмкий подсчёт детерминированными методами, но в отличие от последних, позволяет определить границы, в которых лежат запасы с заданной доверительной вероятностью.*

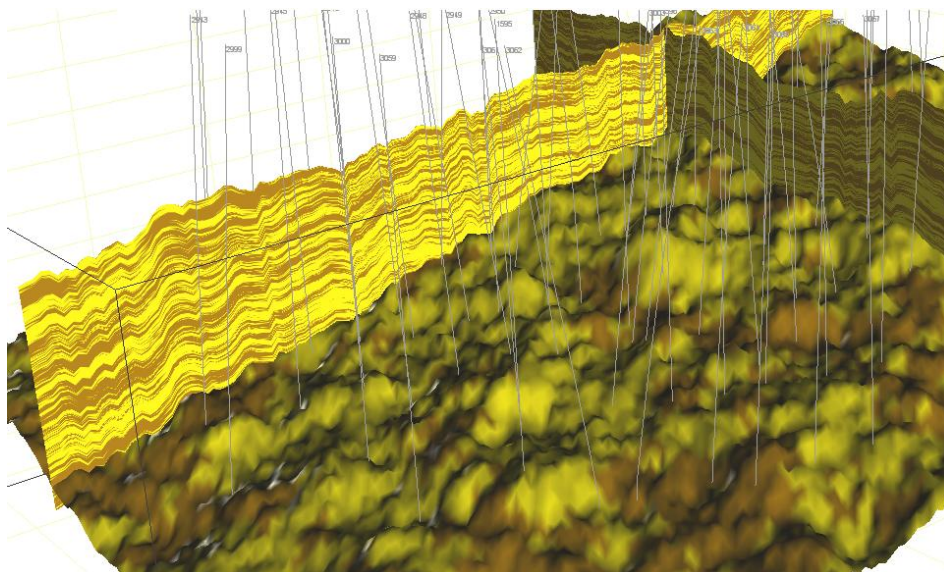
Метод Монте-Карло при моделировании геологических объектов и оценке запасов действительно полезен, если с его помощью строятся гидродинамические модели, с помощью которых рассчитываются проектные извлекаемые запасы. Для гидродинамических моделей стохастическое моделирование по набору реализаций методом Монте-Карло может дать величину проектных извлекаемых запасов, заметно отличную от определённой по одной детерминированной модели. Это связано с тем, что гидродинамические модели и результаты подсчётов по ним очень чувствительны к таким параметрам геологического объекта, как расчленённость и прерывистость пласта коллектора, непрерывная изменчивость фильтрационно-ёмкостных свойств по площади и разрезу, тогда как геологические запасы от этих пара-

метров практически не зависят. Хотя в среднем, стохастические реализации и должны совпадать с детерминированными моделями, изменчивость, прерывистость, расчленённость стохастических реализаций существенно выше.

По детерминированной модели фильтрационно-ёмкостные свойства в высоко разбуренной и хорошо изученной части залежи будут



**Рис. 50а.** Литологическая модель пласта (коллектор – светлый, неколлектор – тёмный), построенная с применением детерминированной интерполяции (предоставлено Е.В.Ковалевским)



**Рис. 50б.** Одна из реализаций литологической модели (коллектор – светлый, неколлектор – тёмный), построенная с применением стохастического моделирования (предоставлено Е.В. Ковалевским)

обладать значительной изменчивостью, тогда как слабо разбуренная и слабо изученная часть будет весьма однородна (см. рис. 50*a*). В результате гидродинамического моделирования, например, процесса заводнения, по такой модели, окажется, что процесс вытеснения в слабо изученных частях залежи заметно эффективнее процесса вытеснения в высоко разбуренной части в силу кажущейся большей однородности первых. В результате КИН в слабо изученных частях окажется выше, а в хорошо изученных – ниже. Но большая однородность слабо изученной части это результат отсутствия информации и детерминированной методики интерполяции. Во всех стохастических реализациях слабо изученная часть будет столь же изменчива и неоднородна, как и хорошо изученная (рис. 50*б*). В итоге КИН, полученный в результате усреднения этих реализаций, будет меньше полученного детерминированным методом.

## **Заключение**

Решение о применении той или иной методики анализа имеющейся геолого-промысловой информации должно быть основано на целях анализа, понимании его геологической сущности и знании математических свойств и особенностей различных методик такого анализа. Данный учебник раскрывает математические основы тех методов обработки данных, без которых немыслима работа современного специалиста-геолога. Авторы надеются, что содержащаяся в нём информация поможет читателям грамотно и профессионально пользоваться современным программным обеспечением для геологического моделирования, правильно выбирать методики анализа данных и устанавливать их параметры, избегая субъективности и произвольности в этих вопросах.

## Литература

1. *Ашкеназы В.О.* Сплайн-поверхности: Основы теории и вычислительные алгоритмы: Учебное пособие. – Тверь: Тверской гос. ун-т, 2003. – 82 с.
2. *Богацкий В.В.* Математический анализ разведочной сети. – М.: Госгеолтехиздат, 1963.
3. *Бондаренко В.Н.* Статистические решения некоторых задач геологии. – М.: Недра, 1970.
4. *Борисов Ю.П., Рябинина З.К., Воинов В.В.* Особенности проектирования разработки нефтяных месторождений с учетом и неоднородности. – М.: Недра, 1976.
5. *Вентцель Е.С.* Теория вероятностей. – М.: Высшая школа, 2003.
6. *Гавришин А.И.* Гидрогеохимические последования с применением математической статистики и ЭВМ. – М.: Недра, 1974.
7. *Гутман И.С.* Методы подсчета запасов нефти и газа: Учебник для вузов. – М.: Недра, 1985.
8. *Гутман И.С.* Применение математических методов и ЭВМ в нефтегазопромышленной геологии: Учебное пособие. – М.: МИНХиГП, 1976.
9. *Дементьев Л.Ф.* Статистические методы обработки и анализа промыслово-геологических данных. – М.: Недра, 1966.
10. *Дементьев Л.Ф., Хитров Е.А., Шурубор Ю.В.* Применение информационных мер в нефтепромышленной геологии / Труды ПермНИПИнефть, вып. 10. – Пермь: Пермское книжное издательство, 1974.
11. *Дементьев Л.Ф., Жданов М.А., Кирсанов А.Н.* Применение математической статистики в нефтегазопромышленной геологии. – М.: Недра, 1977.
12. *Дюбрул О.* Использование геостатистики для включения в геологическую модель сейсмических данных. – SEG, EAGE, 2002.
13. *Смирнов Н.В., Дунин-Борковский И.В.* Краткий курс математической статистики для технических приложений. – М.: Физматгиз, 1959.
14. *Соболь И.М.* Численные методы Монте-Карло. – М.: Наука, 1973.
15. *Чоловский И.П.* Геолого-промысловый анализ при разработке, нефтяных месторождений. – М.: Недра, 1977.
16. *Kelkar M., Perez G.* Applied Geostatistics for Reservoir Characterization. SPE 2002.



## Содержание

Введение .....	3
<b>1. Применение методов математической статистики и теории вероятности в нефтегазопромысловой геологии .....</b>	<b>8</b>
1.1. Случайные величины и случайные явления в нефтегазопромысловой геологии.....	8
1.2. Выборка и генеральная совокупность, вероятностная модель.....	11
1.3. Математическая модель в геологии .....	14
<b>2. Изучение особенности распределения параметров неоднородных продуктивных пластов с помощью математических методов.....</b>	<b>16</b>
2.1. Ряды эмпирических распределений .....	16
2.2. Основные характеристики распределений .....	20
Математическое ожидание.....	20
Начальные и центральные моменты .....	21
Дисперсия.....	23
Коэффициент вариации .....	24
Коэффициент асимметрии .....	25
Экцесс .....	26
Эффект основания .....	28
<b>3. Функции распределения в нефтегазопромысловой геологии .....</b>	<b>30</b>
3.1. Понятие о функциях распределения .....	30
3.2. Интегральная функция распределения .....	31
3.3. Дифференциальная функция распределения .....	33
3.4. Примеры дискретных теоретических распределений .....	36
Распределение Пуассона. ....	36
Биномиальное распределение. ....	38
3.5. Непрерывные распределения .....	39
Нормальный закон распределения. ....	40
Уровни вероятности .....	45
Логарифмически нормальное распределение. ....	48
Распределение Максвелла. ....	51
3.6. Проверка гипотезы о виде статистического распределения.....	51
Ошибки первого и второго рода при проверке нулевой гипотезы .....	53
3.7. Критерии согласия .....	55
Критерий $\lambda$ А.Н. Колмогорова.....	56
Критерий $\lambda^2$ Пирсона. ....	58
3.8. Оценки параметров и их качество .....	61
3.9. Доверительная вероятность и доверительный интервал.....	63

3.10. Изучение неоднородности продуктивных пластов с помощью мер теории информации.....	66
Неоднородность продуктивных пластов как неопределённость системы.....	66
3.11. Энтропия и её свойства.....	69
Энтропия в нефтегазопромысловой геологии.....	71
<b>4. Сравнение двух или нескольких неоднородных геологических объектов</b> .....	75
4.1. Сравнение рядов распределения значений параметров продуктивных пластов.....	76
4.2. Сравнение средних значений параметров. ....	78
4.3. Сравнение дисперсий.....	80
4.4. Сравнение нескольких продуктивных объектов.....	81
4.5. Сравнение двух и более неоднородных продуктивных пластов с помощью энтропии .....	84
<b>5. Исследование зависимости между двумя параметрами</b> .....	87
5.1. Определение корреляционной зависимости, близкой к линейной, между двумя геолого-физическими параметрами .....	90
5.2 Изучение зависимости между двумя геолого-физическими параметрами с помощью корреляционного отношения .....	96
5.3. Построение доверительного интервала уравнения регрессии .....	99
<b>6. Математическое обоснование приёмов геометризации при геологическом моделировании продуктивных пластов</b> .....	102
6.1. Понятие о геологическом моделировании. ....	102
Структурное моделирование.....	103
Литологическое моделирование .....	104
Моделирование свойств.....	104
6.2. Геологические поля.....	105
6.3. Методы интерполяции данных при геометризации продуктивных пластов.....	107
Задача интерполяции.....	107
Скользящее среднее.....	108
Кусочно-линейная интерполяция .....	110
Разложение по базисным функциям.....	111
Метод радиальных базисных функций. ....	111
Сплайн-интерполяция .....	112
Метод обратных расстояний. ....	113
Свойства методов интерполяции .....	116
Закон распределения результатов интерполяции .....	118
<b>7. Исследование статистической зависимости между значениями одного параметра в разных точках при геометризации продуктивного пласта</b> .....	121
7.1. Зависимость между значениями одного параметра в разных точках .....	121

7.2. Вариограмма .....	122
7.3. Кригинг.....	127
7.4. Тренды и экстраполяция.....	130
Доверительные интервалы результатов интерполяции.....	132
<b>8. Стохастическое моделирование .....</b>	<b>137</b>
8.1. Детерминированная интерполяция и стохастическое моделирование.....	137
8.2. Последовательное Гауссово стохастическое моделирование .....	140
8.3. Метод Монте-Карло.....	141
Подборка статистической модели для метода Монте-Карло .....	143
Генерация реализаций в соответствии с выбранной моделью .....	144
Анализ результатов подсчёта запасов методом Монте-Карло.....	146
Заключение.....	150
Литература .....	151

*Учебное пособие*

ISBN 978-5-91961-053-3



**Гутман И.С., Балабан И.Ю.**

**ГЕОСТАТИСТИКА  
В ПРОМЫСЛОВО-ГЕОЛОГИЧЕСКИХ  
ИССЛЕДОВАНИЯХ**

Редактор *Л.А. Суаридзе*  
Компьютерная верстка *И.В. Севалкина*

Подписано в печать 29.11.2011. Формат 60×84/16.  
Бумага офсетная. Печать офсетная. Гарнитура «Таймс».  
Усл. п.л. 9,5. Тираж 200 экз. Заказ № 514

Издательский центр  
РГУ нефти и газа им. И.М. Губкина  
Ленинский просп., 65  
Тел./Факс: (499) 233-95-44