

Министерство науки и высшего образования Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

В.А. Шелутко

МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА ГЕОЭКОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Санкт-Петербург
РГГМУ
2020

УДК 519.24:556.5
ББК 26.22

Рецензенты: д-р техн. наук А.Ю. Виноградов (генеральный директор НПО Гидротехпроект); д-р географ. наук, профессор В.В. Дмитриев (СПб ГУ, профессор); д-р физ.-мат. наук С.А. Кондратьев (Институт озераедения РАН, зам. директора).

Шелутко В.А.

Ш44 Методы обработки и анализа геоэкологической информации: учебник. – СПб.: РГГМУ, 2020. – 296 с.

Рассмотрены методы обработки и анализа геоэкологической информации. Большое внимание уделяется вопросам обоснования и граничных условий применения наиболее распространенных в геоэкологии методов обработки данных.

УДК 519.24:556.5
ББК 26.22

ISBN 978-5-86813-520-0

© Шелутко В.А., 2020
© Российский государственный гидрометеорологический университет (РГГМУ), 2020

Введение

Вместе с экономическим развитием, развитием городов и других населенных пунктов растет круг и сложность задач, связанных с решением геоэкологических проблем. При этом решение почти каждой задачи доводится до численного результата, который, чаще всего, и служит основанием для окончательного ответа на поставленные в задачах вопросы. Поэтому естественно, что объем вычислений в геоэкологии непрерывно возрастает, расширяется также круг используемых математических методов и приемов.

В том или ином объеме методы обработки и анализа информации излагаются во многих учебниках, статьях, руководствах и монографиях. Однако следует отметить, что сведения, интересующие геоэкологов, часто излагаются на уровне, недоступном для большинства специалистов, не имеющих специальной математической подготовки. Другие сведения, например, связанные с множественной корреляцией, анализом временных рядов и т.д., рассматриваются в общенаучной литературе в несколько другом аспекте, чем это нужно для геоэкологических исследований. Кроме того, как и в любой другой науке, достигшей определенной степени зрелости, в геоэкологии не только используются существующие математические методы, но в той или иной степени эти методы трансформируются и развиваются с наложением определенного отпечатка, вызванного особенностями изучаемых объектов. Наконец, ряд вопросов, таких как анализ колебаний рассматриваемых процессов по коротким выборкам, вообще имеет только принципиально общее решение. Применение его для конкретного анализа в геоэкологии требует дополнительных исследований и проработок, специфичных для наук гидрометеорологического цикла.

Все эти обстоятельства, а также необходимость систематически излагать особенности обработки массовых материалов наблюдений и математического моделирования процессов естественного и антропогенного воздействия на окружающую среду, привели к постепенному формированию самостоятельного курса лекций «Методы обработки и анализа геоэкологической информации».

Целью этого курса является подготовка специалистов-экологов к решению широкого круга научных и инженерных задач, связанных с обработкой и анализом геоэкологической информации.

Основные задачи курса:

- изучение теоретических основ современных методов обработки и анализа геоэкологической информации в объеме, достаточном для профессиональной деятельности;
- приобретение навыков практического применения указанных методов на базе широкого использования ЭВМ.

Учитывая особенности геоэкологической информации, в данном учебнике не только излагается техника обработки данных, но и разъясняется, зачем, почему, на каком основании необходимо использовать тот или иной статистический метод. Автор убежден, что без ответа на эти вопросы использование математических, особенно статистических методов, имеет формальный характер и часто приводит к грубым ошибкам.

В основу данного курса положены разработанные ранее автором учебные курсы и изданные по ним учебные пособия: «Техника статистических вычислений в гидрологии» (1977), «Численные методы в гидрологии» (1983), учебно-методическое пособие «Методы обработки и анализа гидрологической информации» (2007), а также учебник «Численные методы в гидрологии» (1991).

Как следует из названия перечисленных изданий, автор ранее занимался исследованиями в области гидрологии и, по-видимому, как бы автор не старался, читатель может заметить некоторый уклон в эту сторону. К счастью, исследования рек и водоемов составляют немалую часть геоэкологических исследований, и, кроме того, основные черты применяемых математических методов по разным направлениям геоэкологии и гидрологии зачастую хорошо совпадают.

При разработке данного и перечисленных выше курсов автор использовал работы ученых-гидрометеорологов, заложивших основы применения статистических методов для анализа состояния окружающей среды. Это, прежде всего, Е.Е. Слуцкий [40], впервые выдвинувший теорию, позволяющую дать достаточно обоснованное объяснение цикличности процессов природы; Д.Л. Соколовский [41], С.Н. Крицкий и М.Ф. Менкель [23], предложившие еще в начале тридцатых годов прошлого столетия использовать теоретические законы распределения для экстраполяции и интерполяции значений исходных рядов наблюдений. Большой вклад в развитие методов обработки и анализа исходной информации в России внесли труды П.П. Ефимовича [16], О.А. Дроздова [14], Н.А. Картвелишвили [19], Ю.М. Алехина [4], Г.А. Алексеева [1, 2], В.А. Рожкова [36], А.М. Рождественского [35], В.А. Румянцева [37],

Г.П. Калинина [18], Л.С. Кучмента [25], Д.Я. Ратковича [33], Г.Г. Сви-
нидзе [38] и др. Значительное влияние на развитие методов обработ-
ки данных наблюдений и анализа результатов оказали также пере-
водные монографии Р.К. Линслея, М.А. Колера, Д.Л. Паулюса [27],
Д. Бокса, Г. Дженкинса [7], Ч. Кайсла [17] и др.

Знания, полученные на основе изучения курса, могут быть ис-
пользованы для количественной оценки состояния окружающей
среды, количественной оценки влияния хозяйственной деятельно-
сти на состояние окружающей среды и прогноза состояния окружа-
ющей среды при том или ином виде хозяйственной деятельности.

При разработке курса лекций и настоящего учебника учитыва-
лось, что студенты знакомы с фундаментальными курсами геоэко-
логии и математики. Математика, включающая теорию вероятно-
стей, изучалась студентами геоэкологами ранее в объеме, предусмо-
тренном для высших технических учебных заведений.

Далее представлены основные разделы данного учебника и
краткое их содержание.

– ***Геоэкологическая информация и ряды наблюдений.***

Здесь даются общие понятия о геоэкологической информации,
геоэкологических процессах и рядах наблюдений, рассматриваются
источники геоэкологической информации, ее особенности и основ-
ные направления применения и изучения.

– ***Математические модели геоэкологических процессов.***

Здесь даются общие понятия о математических моделях, струк-
турах и параметров моделей, идентификации математических моде-
лей. Методы оптимизации структуры и параметров математических
моделей. Типы математических моделей в геоэкологии. Общее обо-
снование применений статистических методов в геоэкологии

– ***Теория вероятностей и математическая статистика
в экологических исследованиях.***

Вообще говоря, этот раздел лекций частично посвящается по-
вторению материалу, пройденного ранее по курсу «Теория вероят-
ностей». Многолетний опыт показал необходимость такого повторе-
ния как в смысле некоторых уточнений, связанных с особенностью
геоэкологической информации, так и в смысле освоения принятых
в геоэкологии и гидрометеорологии обозначений.

Здесь приводятся общие понятия о случайной величине и их
характеристиках, определениях и обозначениях; общая характери-
стика законов распределения. Подробно рассматриваются числовые
характеристики случайных величин и их свойства. Даются понятия

о системе случайных величин и о нормированной случайной величине.

– *Теоретические законы распределения в геоэкологических исследованиях (экстраполяция и интерполяция данных наблюдений).*

В разделе детально рассматриваются: назначение теоретических законов распределения; области, особенности и алгоритмы применения следующих законов распределения: нормальный закон распределения, закон распределения Пирсона III типа, трехпараметрическое гамма-распределение (распределение Крицкого–Менкеля), логарифмически нормальное распределение, распределение Гумбела, распределение Джонсона.

– *Методы оценки числовых характеристик в геоэкологических исследованиях.*

Формулируются требования к методам оценки, и, исходя из этих требований, дается их сравнительный анализ. Детально рассматриваются три метода: метод моментов, метод наибольшего правдоподобия, метод квантилей (графоаналитический метод Алексеева).

– *Оценка однородности рядов наблюдений и стационарности геоэкологических процессов на основе статистических гипотез.*

В разделе детально рассматриваются понятия и свойства гипотез и критерия значимости. Излагаются методы проверки гипотез о законе распределения (критерии согласия Колмогорова, Пирсона, Смирнова–Колмогорова) и о параметрах распределения (критерии Стьюдента–Госсета и Фишера), а также непараметрические методы проверки гипотез (критерии Уилкоксона, Манна–Уитни и критерий знаков).

– *Статистические зависимости и линейная корреляция в геоэкологии.*

В разделе рассматриваются общие задачи изучения взаимосвязей природных процессов, регрессионная математическая модель взаимосвязи двух переменных величин (линейное уравнение регрессии двух переменных величин, коэффициент корреляции, оценка коэффициента корреляции и уравнения регрессии по выборочным данным, граничные условия математической модели парной корреляции); регрессионная математическая модель взаимосвязи ряда переменных величин (уравнение множественной линейной корреляции, сводные параметры множественной линейной корреляции, частный коэффициент корреляции, оценка параметров уравнения множественной корреляции по выборочным данным, граничные условия регрессионной математической модели взаимосвязи

ряда процессов, алгоритм и программа расчетов на основе регрессионной математической модели, преобразование исходной информации в целях нормализации и линеаризации взаимосвязей, графические методы построения уравнения множественной корреляции).

– *Анализ колебаний геоэкологических процессов во времени и пространстве.*

Даются понятия о случайных функциях и случайных процессах, их характеристиках (определение, законы распределения, числовые характеристики, классификация математических моделей случайных функций).

Оценка общего характера колебаний геоэкологических процессов. Анализ стационарности, однородности и наличия тренда (оценка тренда по коэффициенту корреляции и по критерию Спирмена, определение даты нарушения однородности). Проверка выборочных рядов на случайность (по критерию длин и числа серий, по числу повышений и понижений ряда, по числу экстремумов).

Корреляционные функции геоэкологических процессов (общие сведения, корреляционные функции стационарных гидрологических процессов, корреляционные функции различных временных рядов, оценки корреляционных функций по выборочным данным).

Спектральный анализ геоэкологических процессов (общие сведения, численные методы расчета спектральной плотности, выборочная оценка, спектральная плотность различных временных рядов). Математические модели линейных стационарных процессов.

– *Численные эксперименты в геоэкологии.*

Методы экспериментальных исследований в геоэкологии. Общее представление о методе Монте-Карло. Происхождение метода, схема и некоторые особенности применения метода Монте-Карло.

Разыгрывание дискретной случайной величины. Воспроизведение последовательности значений случайной величины, равномерно распределенной в интервале $[0, 1]$, моделирование гидрологических рядов по различным математическим схемам. Понятие о моделировании ряда. Моделирование ряда по схеме случайной величины, по схеме простой цепи Маркова, графоаналитический способ моделирования.

Предваряя изложение курса, необходимо сделать несколько общих замечаний.

Многие численные методы, применяемые в настоящее время в геоэкологических исследованиях, могут быть реализованы на ЭВМ. При этом программы счета часто используются по схеме

«черного ящика». Есть вход (исходные данные), есть выход (результат расчета). Как происходит переработка информации для получения результата, казалось бы, можно и не знать. Между тем многолетний опыт показал, что незнание схемы расчетов и ограничений использования того или иного метода может привести к крупным просчетам и к парадоксальным результатам. С учетом этого обстоятельства возникла необходимость предупредить пользователей о возможных последствиях формального использования того или иного метода. Для этого в учебнике, там, где это оказалось возможным, приводится обзор граничных условий применения рассматриваемых методов и последствий их несоблюдения.

Применяемые методы расчетов и анализа процессов разделяются на два класса: детерминистические (динамические) и вероятностные. Первые основаны на применении классических методов вычислительной математики, вторые – на применении методов теории вероятностей и математической статистики. В силу особенностей рассматриваемых процессов (см. ниже) явное предпочтение при обработке информации в геоэкологии и гидрометеорологии отдается методам теории вероятностей и математической статистики. Вместе с тем в ряде случаев будут использоваться и методы классической вычислительной математики.

В заключение следует отметить, что результаты, полученные на основе применения методов обработки и анализа геоэкологической информации, зависят не только от правильности выбора самих методов, но и от репрезентативности и точности исходных данных. По замечанию английского натуралиста Гексли:

«Математику можно сравнить с мельницей превосходного устройства, которая перемалывает что угодно до любой тонкости. Тем не менее, то, что вы получите, зависит от того, что вы засыпаете. И как самая великолепная мельница в мире не доставит вам пшеничную крупчатку из лебеды, так и страницы формул не доставят вам определенного результата из сомнительных данных».

Очевидно, что это относится и к применению математических методов в геоэкологии. Именно поэтому автор является сторонником анализа, начиная с анализа первичных данных наблюдений. Многолетний опыт работы показал необходимость такого анализа. Так, даже в гидрологических ежегодниках, как правило, прошедших многоступенчатую проверку исходных данных и иногда даже уже использованных в проектах, все равно находились ошибки. Тем более это относится к геоэкологическим данным, контроль за

качеством которых гораздо менее эффективен, чем за гидрологическими данными. Следует также иметь в виду, что публикующиеся в настоящее время в открытой печати материалы часто получены на основе тех или иных методов и методик, которые со временем могут изменяться. Первичные данные наблюдений гораздо более устойчивы в этом отношении.

В заключение автор должен выразить глубокую благодарность своим учителям и наставникам, оказавшим большое влияние на формирование курса лекций и учебников автора. Особенно большое влияние на автора оказали идеи Д.Л. Соколовского, Ю.М. Алехина, Г.А. Алексеева, Г.Г. Сванидзе. Автор считает себя учеником и, в какой-то степени, наследником идей этих больших ученых.

Особую благодарность должен выразить недавно ушедшим из жизни профессорам Ю.Б. Виноградову и А.М. Рождественскому, дискуссии и споры с которыми помогли формированию взглядов автора, отраженных в данном учебнике.

Наконец, созданию и усовершенствованию этого курса во многом способствовало многолетнее сотрудничество автора с недавно ушедшей из жизни доцентом В.Г. Гутниченко и инженером А.В. Смирновой. Их замечания и предложения также отражены в этом учебнике.

Глава 1. Геоэкологическая информация и ряды наблюдений

1.1. Общие понятия

В основе научных исследований и их практических приложений, как правило, лежит информация.

Информация – сведения об исследуемом объекте, зафиксированные каким-либо образом.

Геоэкологическая информация (ГЭИ) – информация о геоэкологических объектах или объектах, изучаемых в геоэкологии.

Информация может фиксироваться в памяти человека или на каких-либо технических носителях: бумага, магнитные диски, лазерные диски, USB-flash-drive (флешки) и т.д.

Информация сама может служить объектом хранения, переработки, передачи и изучения. Информацию можно создавать, передавать, запоминать, искать, принимать, копировать, обрабатывать, разрушать. Информация может быть в самых разнообразных формах: в форме световых, звуковых или радиоволн, электрического тока или напряжения, магнитных полей, знаков на бумажном носителе. В принципе, информацию может переносить любая материальная структура или поток энергии.

Масштабы использования информации являются одним из основных признаков, отличающих мыслящие особи от всех остальных существ (В. Иллингворт, 1989).

1.2. Источники геоэкологической информации

Геоэкологическую информацию формируют многочисленные и разнообразные организации. Это законодательные и исполнительные органы власти РФ, аналогичные органы в субъектах федерации, которые издают различные нормативные акты, принимают планы и программы; службы соответствующих министерств и ведомств,

которые ведут те или иные наблюдения за состоянием окружающей среды, здоровьем людей, а также за факторами воздействия на окружающую среду. Это многочисленные научно-исследовательские организации – институты, станции и пр. Кроме того, сбором и анализом экологической информации занимаются проектировщики, поскольку, например, без знания характеристик окружающей среды невозможно спроектировать прочное и полезное сооружение. Эта же информация оказывается и в распоряжении заказчиков проектной документации – государственных организаций или коммерческих компаний.

Также сбором экологической информации занимаются международные организации, координирующие деятельность по охране окружающей среды, и общественные экологические организации.

Одним из важнейших понятий, связанных с экологической информацией, ее получением, является экологический мониторинг. В функции различных российских ведомств, определенные нормативными документами, входит целый ряд видов мониторинга: от мониторинга состояния и загрязнения окружающей среды до социально-гигиенического мониторинга.

Такая информация может содержать данные:

- о состоянии воды, воздуха, фауны, флоры, земли, почвы, недр, природных ландшафтов и комплексов;
- об экологической угрозе или риске для здоровья и жизни людей;
- о химических, физических и биологических воздействиях на состояние окружающей среды и их источниках;
- о хозяйственной деятельности, отрицательно влияющей или могущей повлиять на природные объекты и человека;
- о мерах по охране окружающей среды, в том числе правовых, административных и иных мерах;
- о деятельности государственных органов, юридических лиц и граждан-предпринимателей в сфере распоряжения природными ресурсами, природопользования, охраны окружающей среды, обеспечения соблюдения и защиты экологических прав и законных интересов физических и юридических лиц, если необходимость осуществления такой деятельности установлена законодательством России.

1.3. Методы изучения геоэкологических процессов

Геоэкологические объекты могут оставаться неизменными или меняться во времени и/или в пространстве. В свою очередь смена состояний экологических объектов может быть естественной или вызванной человеческой деятельностью. Для исследования изменений состояний объекта во времени вводится понятие «процесс».

***Процесс** – последовательная смена состояний или стадий развития объекта во времени или пространстве.*

Геоэкологический процесс разворачивается во времени в виде определенной последовательности событий. Представления об этом процессе получаются, как правило, по данным наблюдений или экспериментов.

До недавнего времени в естественных науках использовались почти исключительно два метода экспериментальных исследований: метод пассивного и метод активного эксперимента. В последние годы широкое распространение получил метод численных экспериментов.

Метод пассивного эксперимента. Метод заключается в наблюдениях за природными явлениями в естественных условиях на обширной сети станций и постов. Результаты наблюдений хранятся в различных фондах и частично публикуются в виде ежегодников и в различного рода справочниках. Значение таких данных трудно переоценить, так как они являются фундаментом, на котором развиваются геоэкологические и все другие исследования окружающей среды. Вместе с тем набор материалов, которые могут быть получены на основе пассивных экспериментов, имеет весьма существенные ограничения.

Во-первых, эти материалы являются выборочными во времени. Действительно, стационарные наблюдения охватывают период времени определенной продолжительности, не превышающий, как правило, 100–150 лет. В подавляющем числе случаев продолжительность этого периода меньше 40–50 лет. Кроме того, наблюдения в основном приурочены к нашему времени, т. е. охватывают только последний период развития рассматриваемых процессов. Вполне вероятно, что в него не попали годы экстремальной водности или годы с редким сочетанием тех или иных геофизических факторов. А именно эти данные часто интересуют потребителей геоэкологической информации.

Во-вторых, материалы наблюдений являются выборочными в пространстве. Как известно, преобладающее большинство пунктов наблюдений, особенно продолжительных во времени, приурочено к обжитым местам. В то же время в связи с развитием народного хозяйства происходит все большее смещение промышленности в северные и восточные, ранее малообжитые, районы. А именно в этих районах данные наблюдений отсутствуют или крайне немногочисленны. Кроме того, формирование климатических условий на отдельных территориях происходит, как правило, в верхней и средней части речных бассейнов, а наблюдения в основном приурочены к нижней, более обжитой части этих бассейнов.

В-третьих, многие природные явления, активно участвующие в формировании геоэкологических процессов, настолько завуалированы, что наблюдать их в природе практически невозможно.

Метод активного эксперимента. Метод заключается в постановке полевых или лабораторных физических экспериментов.

Так как формирование природных процессов имеет сложный, многоступенчатый и многофакторный характер, то такие эксперименты во многих случаях основаны на более или менее грубом приближении физических моделей к природным объектам. Вследствие этого экспериментальные результаты, полученные в лаборатории и в полевых условиях, не всегда могут быть перенесены в естественные условия.

Метод численных экспериментов. В настоящее время в связи с развитием вычислительной техники и повышением общего уровня исследований в практику расчетов и прогнозов широко внедряется третий метод экспериментальных исследований – метод численных экспериментов, осуществляемый на основе математического моделирования рассматриваемых процессов.

Численным экспериментом называется численное решение уравнений математической модели некоторой системы при тех или иных заданных условиях или значениях параметров.

Особенно важную роль численные эксперименты играют в следующих областях наук о Земле.

1. Исследование состояний окружающей среды с помощью математических моделей для тех условий, которые не могут быть изучены путем непосредственных наблюдений или путем постановки лабораторных или полевых экспериментов.

Например, одной из важнейших задач при водохозяйственном проектировании является определение возможной продолжительности маловодных периодов. Ряды непосредственных наблюдений,

всегда ограниченные по продолжительности, или результаты лабораторных и полевых экспериментов в этом случае помочь не могут. В то же время метод численных экспериментов, осуществляемый на основе статистической модели колебаний водности, – метод Монте-Карло – позволяет по статистическим характеристикам исследуемых рядов получить выборки практически неограниченной продолжительности, включающие в себя почти все возможные циклы колебаний водности.

2. Конструирование наиболее простых, физически обоснованных и достаточно точных моделей процессов в окружающей среде.

Действительно, численные эксперименты позволяют задавать любые сочетания гидрологических факторов в возможном диапазоне их изменения и предполагаемом соотношении и путем пересчета устанавливать рациональную структуру и параметры предполагаемой математической модели.

3. Планирование и постановка экспериментов в полевых и лабораторных условиях на основе разработанных путем численных экспериментов математических моделей.

В настоящее время во многих отраслях науки и техники все чаще ставится требование, чтобы перед тем, как перейти к дорогостоящим лабораторным или полевым экспериментам, возможные ситуации, возникающие в ходе опытов, были проиграны на математических моделях путем численных экспериментов.

1.4. Геоэкологические процессы и ряды наблюдений

Из данных, полученных в результате экспериментов наблюдений, объединенных по какому либо признаку, конструируются ряды наблюдений.

Ряд наблюдений – последовательность данных наблюдений или каких-либо других сведений, характеризующих данный объект во времени и/или в пространстве.

Ряды наблюдений могут быть первичными и вторичными.

Первичный ряд – ряд данных непосредственных наблюдений за теми или иными процессами.

Вторичный ряд – ряд данных, полученный на основе обработки первичного ряда данных наблюдений.

Деление рядов на первичные и вторичные во многом условно, так как в зависимости от имеющихся приборов и принятой методики наблюдений или измерений те же самые данные могут составлять и первичный и вторичный ряд. Так влажность воздуха может измеряться непосредственно, например, с помощью гигрографов. В этом случае ряд данных является первичным. Те же самые данные, полученные расчетом по данным наблюдений с помощью психрометров (сухой и смоченный термометр) являются вторичными.

Ряды наблюдений могут быть *дискретными* или *непрерывными*. Например, запись уровней воды самописцем дает непрерывный ряд наблюдений. Аналогичные ряды наблюдений могут быть получены при непрерывной записи и по другим процессам. Однако чаще значения уровней наблюдаются на простейших уровнемерных устройствах в конкретные сроки, например, ежедневно в 8 и в 20 ч. Временной ряд, составленный из таких наблюдений, называется дискретным.

Для практических расчетов непрерывные ряды обычно дискретизируются (квантуются). Это делается для удобства обработки и анализа. При этом используются два метода дискретизации: по равномерным интервалам или по характерным точкам.

Наблюдения могут производиться через равные интервалы времени. Такие дискретные ряды называются *эквилистенными*. Наблюдения могут производиться и не через равные интервалы. Такие дискретные ряды называются *неэквилистенными*.

Ряды наблюдений могут быть временными, пространственными и пространственно-временными.

Ряды наблюдений обычно обозначают заглавной латинской буквой X , Y , или Z , и т.д. В этих обозначениях X представляет всю последовательность значений ряда X , Y — всю последовательность значений ряда Y . Отдельные значения каждого ряда обозначаются соответствующими строчными буквами с опущенным индексом. Например первое значение ряда $X - x_1$, второе значение — x_2 , и т.д.; любое i -ое значение — x_i , где i может изменяться от 1 до n , n — число членов ряда. В соответствии с этими обозначениями значения ряда X можно представить в следующем виде:

$$x_1, x_2, \dots, x_n,$$

а отдельные значения ряда:

$$x_i, (i = 1, 2, \dots, n).$$

Данные в рядах наблюдений могут представляться в различном виде, чаще всего в цифровом и реже в символьном обозначении. Например, концентрация загрязняющих веществ в реке может

задаваться в количестве миллиграмм на литр воды, а влажность почвы – в символьном обозначении.

Следует отметить, что в геоэкологии используются ряды наблюдений над разнообразными объектами окружающей среды. Эти ряды имеют различную продолжительность периода и частоту наблюдений. Так при анализе уровней загрязнения окружающей среды следует иметь в виду, что по большинству городов и промышленных зон материалы наблюдений, как правило, освещают только период установившейся урбанизации. Так, в пределах г. Санкт-Петербурга регулярные наблюдения за химическим составом речного стока начались в основном лишь в период с 1959 по 1969 г.

Кроме того, материалы наблюдений за факторами формирования окружающей среды во многих случаях являются неоднородными во времени. Так, если в первые годы наблюдений за качественными характеристиками окружающей среды отбор проб и определение концентраций загрязняющих веществ или их характеристик проводились 4 раза в год, то затем 7 раз, а с 1989 г. – 12 раз в год. Естественно, что эти обстоятельства в значительной степени затрудняют ретроспективный анализ возможных средних годовых и экстремальных годовых значений, так и анализ тенденций изменяющегося качества воды во времени.

Следует также отметить, что определение химических и других веществ в пробах воды обычно определяется с достаточной для анализа точностью. Однако сами по себе пробы воды в некоторых случаях являются нерепрезентативными, то есть не отражают действительное содержание химических веществ в реке или водоеме.

1.5. Особенности геоэкологической информации

Предметом изучения большинства наук о земле являются естественные процессы, происходящие в природе. В геоэкологии особое внимание уделяется процессам, которые изначально формировались и поддерживались внешними естественными условиями и лишь в последние десятилетия или годы стали подвергаться антропогенному воздействию. В связи с этим геоэкологические наблюдения над многими процессами начались сравнительно недавно и часто носили отрывочный или нерегулярный характер.

В связи с этим информация, используемая в геоэкологии, зачастую во многом отличается от информации, используемой, например, в гидрологии и метеорологии.

В трудах РГГМУ неоднократно отмечалось, что первичная обработка данных гидрохимических и ряда других наблюдений, освещающих состояние природной среды, ведётся не совсем корректно, так как не учитывает особенности этих рядов наблюдений, что приводило и приводит к погрешностям в результатах расчетов и, конечно же, в какой-то степени сказывается на выводах на основе этих данных [49]. В частности нам представляется, что замечание о недостаточной корректности обработки данных в особой степени относится к расчетам среднегодовых концентраций содержащихся в речном стоке веществ.

Так методика расчёта среднегодовых концентраций содержащихся в воде веществ, по гидрохимическим наблюдениям, принятая в настоящее время, основана на следующих теоретических предположениях:

- 1) временные ряды измеренных значений концентраций содержащихся в воде веществ в каждом пункте наблюдений описываются моделью в виде ряда значений случайной величины;

- 2) данные ряды наблюдений являются стационарными, регулярными и однородными.

В последние годы выявилось, что при анализе среднегодовых концентраций, рассчитанных на этой основе, могут быть получены противоречивые результаты, не имеющие какого-либо физического объяснения. В связи с этим возник вопрос о репрезентативности и надёжности расчётов среднегодовых концентраций содержащихся в воде веществ существующими методами. В выполненных исследованиях было показано, что гидрохимические наблюдения, имеют ряд особенностей, которые не укладываются в рамки рассмотренных выше теоретических положений [49]. Во-первых, исходные ряды неоднородны по числу измерений в год; во-вторых, концентрации рассматриваемых в анализе веществ зависят от расходов воды; в третьих, исходные ряды являются неэквидистентными; в четвертых, ряды наблюдений являются неоднородными по генезису формирования.

Глава 2. Математические модели геоэкологических процессов

2.1. Общие понятия о математических моделях

Понятие «математическая модель» является в настоящее время не переменным атрибутом большинства исследований, в том числе и чисто теоретических. В последние годы расширилось и само понятие моделей. Когда-то раньше модель – это непременно какая-то материально реализованная система, имеющая черты внешнего сходства с рассматриваемым объектом. В настоящее время:

***Модель** – каким-либо образом реализованная система, которая способна замещать исследуемый объект так, что ее изучение даст новую информацию об этом объекте.*

***Моделирование** – процесс построения (или выбора) и изучения моделей в целях получения новых знаний об исследуемом объекте.*

В общем случае модель может быть нематериальной и не иметь никакого внешнего сходства с исследуемым объектом. То есть исследуемый объект и его модели могут быть любой природы: в том числе материальными (предметными) и идеальными (мысленными).

***Предметная модель** – материальный объект, отражающий свойства исследуемого объекта и способный его замещать в процессе эксперимента.*

***Мысленная модель** – мысленный образ, способный замещать исследуемый объект так, что его изучение дает новую информацию об этом объекте.*

Предметное моделирование применяется тогда, когда необходимо выяснить не общие физические закономерности, а детально изучить вполне конкретный процесс, происходящий на исследуемом объекте с определенными геометрическими и физическими свойствами при заданных режимных условиях. Проводить же эксперименты с самим этим процессом либо совсем не удастся, либо они крайне затруднены.

Среди предметных моделей обычно выделяют пространственно-временные (геометрические), физические и предметно-математические модели.

Физические и пространственно-временные модели сходны с оригиналом по физической природе, геометрической форме или характеру изменения во времени. Примеры таких моделей: уменьшенные копии ГЭС, уменьшенные копии участка реки, копии прибора.

Предметно-математические модели конструируются из элементов иной, чем исследуемый объект, физической природы, но описываются теми же математическими зависимостями. Они имеют ряд преимуществ по сравнению с физическими и пространственно-временными моделями:

- более универсальны;
- переход от задачи к задаче не требует каждый раз построения новой модели;
- изменение параметров моделируемой системы, как правило, не вызывает больших переделок модели;
- сами модели относительно просты и дешевы.

Предметно-математические модели могут строиться на основе прямой аналогии между величинами, присущими физически различным явлениям.

Так, в исследованиях речного стока в 1960–70-х годах был распространен расчет гидрографов стока на электронных моделирующих установках, основанный на использовании аналогии между генетической формулой стока и формулой изменения силы тока i на выходе электрической системы из RC -элементов в зависимости от изменения входного напряжения U [23, 46].

Действительно, первая из них имеет вид:

$$Q(t) = \int_{\tau=1}^{\tau=n} h_{t-\tau} f_{\tau} d\tau, \quad (2.1)$$

где Q – расход воды в замыкающем створе водосбора, h – слой дождевых осадков минус потери, f – распределение единичных площадей стекания, n – число единичных площадей, t – время, τ – время добегания.

Вторая:

$$i(t) = \int_{\tau=1}^{\tau=n} U_{t-\tau} P_n(\tau) d\tau, \quad (2.2)$$

где P_n – весовая функция системы, представляющая собой реакцию устройства на единичный импульс, поданный на его вход.

Решающий блок таких систем – это электрическая цепь из n последовательно соединенных RC -элементов, состоящих из активного сопротивления R и емкости C .

Мысленные модели конструируются с помощью языка знаков (математические формулы и другие обозначения), рисунков, схем и прочих материальных средств. Однако все преобразования в них, все переходы в другие состояния осуществляются мысленно, т. е. в сознании человека.

Из класса мысленных моделей для исследователей особый интерес представляют математические модели.

Математическая модель – совокупность математических и логических соотношений, с помощью которых устанавливаются количественные и качественные характеристики изучаемого явления и его связей с окружающей средой.

Широкое распространение математических моделей в науках о Земле объясняется следующими обстоятельствами.

1. Для решения задач, связанных с природными явлениями, необходимо использовать те или иные математические методы. Выбор этих методов нередко сам по себе представляет достаточно сложную задачу, решение которой далеко не всегда является бесспорным. Между тем от того, какой математический метод будет использован, в большой степени зависит конечный результат исследований, особенно в сложных условиях геоэкологического анализа. Эта зависимость с течением времени все больше возрастает.

Во многих отношениях задача выбора математического метода упрощается тем, что современная математика представляет собой колоссальный арсенал математических структур, заготовленных впрок. Каждую из этих структур можно использовать при выполнении определенных граничных условий. Если исследуемый процесс или явление соответствует этим условиям, то математическая структура может выступать в качестве полномочного представителя процесса или, иначе говоря, служить его математической моделью.

2. Если рассматриваемый процесс адекватно описывается какой-то математической моделью, то это означает, что для его описания (характеристики) можно использовать весь круг понятий, связанных с данной математической моделью. Этот круг понятий представляет собой дополнительную информацию, которая часто является весьма весомым вкладом в общие знания о рассматриваемом явлении или процессе. Например, решение о том, что для описания

данного ряда наблюдений может использоваться математическая модель в виде последовательности не связанных между собой значений случайной величины, часто означает, что для полной характеристики этого ряда достаточно знать вид одномерного закона распределения и три числовые характеристики: математическое ожидание, среднее квадратическое отклонение и коэффициент асимметрии. Иногда используется четвертая характеристика – эксцесс.

3. Во многих случаях математическая модель может замещать исследуемый объект, служить его полномочным представителем во всякого рода испытаниях. Так, часто невыполнимые на практике опыты с объектами природы или эксперименты с ними могут быть заменены численными экспериментами с математическими моделями.

Таким образом, использование математических моделей облегчает выбор и использование математического аппарата, предоставляет в распоряжение исследователя дополнительную информацию о рассматриваемом процессе или явлении, а также позволяет произвести эксперименты в тех случаях, когда на объектах природы они невозможны или затруднены в силу тех или иных обстоятельств.

В основе математического моделирования лежат явления изоморфизма и гомоморфизма [26].

Изоморфными называются такие две системы A и B с заданными в них отношениями, для которых выполняются следующие условия:

- каждому объекту a из системы A однозначно соответствует объект b из системы B и наоборот;
- каждому отношению в системе A однозначно соответствует отношение в системе B и наоборот;
- если некоторые объекты системы A ($a_1, a_2, \dots$) связаны отношением R , то соответствующие объекты системы B ($b_1, b_2, \dots$) должны быть связаны отношением R и наоборот.

Изоморфное отображение какого-либо объекта является адекватным или тождественным.

Гомоморфными называются такие две системы A и B , между объектами которых нет взаимнооднозначного соответствия, т. е. одна система является лишь упрощенной копией другой.

Гомоморфизм указывает на единство, связь, взаимодействие и взаимозаменяемость в определенных пределах различных явлений,

на сходство их формы и отдельных закономерностей. Поэтому в строго определенных границах и условиях (граничные условия) можно заменить изучение одного явления изучением другого, подобного ему по структуре.

При исследовании природных явлений, вследствие сложности рассматриваемых процессов, математические модели являются чаще всего гомоморфным образом объекта.

2.2. Структура и параметры модели

В математических моделях концентрированно выражаются: во-первых, общие теоретические (априорные) представления о классе, к которому относится рассматриваемое явление: во-вторых, устойчивые количественные соотношения, свойственные только данному конкретному явлению. Носителем априорной информации, вытекающей из принадлежности, допустим, водосбора или процесса формирования стока к некоторому классу, обычно выступает структура модели (общий вид математических и логических соотношений). Характеристики, свойственные каждому конкретному явлению, содержатся в параметрах модели.

Таким образом, информация, содержащаяся в математической модели, состоит из априорной информации, заключенной в структуре модели и информации, содержащейся в параметрах модели, которая определяется на основе наблюдений за конкретным объектом или явлением.

В связи с этим для любой математической модели характерно следующее противоречие. С одной стороны, чем более общей выбирается структура модели, тем меньше содержится в ней информации, относящейся к отдельному явлению; с другой стороны, при выборе структуры модели, наилучшим образом описывающей лишь наблюдаемые на конкретном водосборе явления, теряется априорная информация, полученная на основе анализа наблюдений за другими явлениями. Это противоречие разрешается в разных условиях по-разному и нередко зависит от направления исследований и возможности получения той или иной информации.

При выборе структуры модели должны учитываться следующие обстоятельства: объем имеющейся информации и точность определения данных, составляющих ее.

Любая математическая модель должна быть достаточно простой, обозримой, чтобы ее можно было конкретно записать, добыть все необходимые для ее записи данные и, если эта модель является расчетной, поместить эти данные в память ЭВМ.

2.3. Идентификация математических моделей

Математическое моделирование геоэкологических процессов связано, как правило, с задачами идентификации. Дело в том, что традиционные методы моделирования, основанные на строгом применении общих физических законов к описанию определенных явлений, предполагает наличие измеренных параметров системы. Однако параметры подавляющего большинства исследуемых в природе объектов, например, речного бассейна, установить полностью путем измерений практически невозможно. Поэтому при математическом моделировании такие объекты обычно рассматриваются как системы с не измеряемыми параметрами. Для определения системы в этом случае используются наблюдения за факторами формирования исследуемых процессов (вход) и режимом исследуемого процесса на выходе (рис. 2.1).

Такой подход позволяет заменить отсутствующие или ненадежные реальные физические характеристики параметрами, которые находятся по достаточно точным измеренным величинам. Эти параметры могут иметь различную физическую интерпретацию. В упрощенных моделях они обычно отражают влияние сразу нескольких факторов, и их следует рассматривать как эмпирические коэффициенты, позволяющие установить соответствие между доступными для измерения величинами.

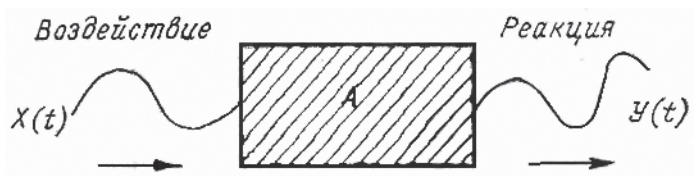


Рис. 2.1. Схема динамической модели

Таким образом, недоступность непосредственных измерений ставит перед исследователями задачу косвенных определений по наблюдениям на входе и выходе исследуемой системы, что является типичной задачей идентификации.

Идентификация в широком смысле – построение оптимальной математической модели динамической системы на основании наблюдений на ее входе и выходе.

Эта задача включает в себя установление структуры математической модели, т. е. выбор общего вида уравнений и связей, и параметров этих уравнений и связей (параметризация).

Идентификация по структуре математической модели обычно основывается на совпадении граничных условий, определяющих исследуемые стороны данного явления и математическую модель. Идентификация параметров или параметризация (задача идентификации в узком смысле) проводится, как правило, путем решения обратной задачи.

Задача идентификации структуры и параметров математической модели обычно ставится в следующем виде [23]:

Задан ряд наблюдений X за воздействием на систему и ряд Y на выходе системы (рис. 2.1). Необходимо найти оператор, наилучшим образом согласующий X и Y . Структура и параметры искомого оператора, как правило, неизвестны. Вместе с тем почти всегда существует информация о возможных классах структуры математической модели и возможных пределах изменений ее параметров. Источник информации – теоретические представления о механизме физических явлений данного класса. Эту информацию можно получить одним из следующих способов: просмотр литературы, составление библиографических справок, анализ источников информации и данных наблюдений, анализ экспериментальных данных, консультации со специалистами и экспертами и т.д. Также одним из возможных источников информации могут быть априорные предположения, включенные в процесс моделирования и основанные, как и предыдущие источники информации, на имеющемся уже опыте.

В тех случаях, когда имеющейся информации недостаточно для построения модели, выдвигаются рабочие гипотезы, принимаются различные допущения, которые в сумме представляют собой априорное математическое описание процессов исследуемой системы. При получении новой информации в процессе идентификации или работы модели ее структура и параметры могут быть пересмотрены.

Математическая модель становится адекватной, т. е. вполне соответствующей, и приобретает объективное содержание только при выполнении следующих условий.

- Во-первых, представления, составляющие априорное содержание модели, должны быть связаны определенными соотношениями с объективными характеристиками изучаемого объекта.

- Во-вторых, должна быть разработана и использована четкая процедура проверки справедливости системы умозаключений, следующих из данной математической модели. Этим гарантируется получение объективных результатов, не зависящих от произвольных желаний исследователя.

- В-третьих, должна быть обеспечена определенная степень согласованности между результатами, полученными при экспериментировании с объектами, и соответствующими значениями, вычисленными или «снятыми» с модели.

- В-четвертых, должно быть определенное соответствие между предсказаниями на основе математической модели и реальным поведением объекта.

Следует отметить, что определение параметров модели путем непосредственного решения уравнений, связывающих наблюдения на ее входе и выходе, может привести к большим погрешностям, так как небольшие изменения входных и выходных данных могут значительно изменить определяемые таким образом параметры. Поэтому в некоторых случаях при определении параметров математической модели привлекают теорию некорректных задач, позволяющую с помощью дополнительной информации получить более устойчивые решения с точностью, соответствующей точности задания исходных данных и адекватности математической модели реальному процессу [23].

Нередко при определении параметров рассматриваемой системы отсутствует гарантия, что будут получены фактические значения параметров. Это объясняется тем, что в значениях исходных данных могут быть большие погрешности, вызванные, с одной стороны, погрешностями наблюдений, с другой – выборочным характером исходных данных. Роль отдельных грубых погрешностей можно в значительной степени снизить, объединив все имеющиеся наблюдения и дополнительную информацию о параметрах. Однако при объединении наблюдений нередко число уравнений превышает количество искомых параметров, причем используются неточные, а иногда противоречивые соотношения. В этом случае постановку задачи меняют, ищут не точные значения параметров, а их оптимальные значения в соответствии с заранее выбранными условиями.

2.4. Методы оптимизации

В более общей постановке задачу определения оптимальных параметров можно рассматривать как задачу отыскания таких значений параметров, при которых некоторая заданная функция, зависящая от сходимости рассчитанных значений и данных наблюдений, достигает минимума. Такая функция называется критерием оптимизации или критерием качества.

Критерий качества – некоторая заданная функция, зависящая от сходимости рассчитанных значений и данных наблюдений.

В отличие от методов определения параметров, основанных на непосредственном решении уравнений относительно этих параметров и дающих вполне определенный набор решений для рассматриваемых входных и выходных данных, методы оптимизации могут давать различные значения параметров в зависимости от принятого критерия качества.

Большим достоинством метода оптимизации является возможность охватить практически неограниченный объем наблюдений и дополнительную априорную информацию, ограничивающую область, в которой должны искать решение. Это позволяет при достаточном объеме информации находить значения параметров, близких к фактическим.

Вследствие универсальности метода идентификации на основе оптимизации требуется, чтобы поставленная задача была достаточно обоснована физически, а также, чтобы была уверенность в существовании семейства сравнительно мало отличающихся решений. Без этого применение методов оптимизации создаст лишь иллюзию решения задачи идентификации параметров, а сами параметры, при которых получают хорошую сходимость поверочных расчетов, при испытаниях на независимом материале дают результаты, совершенно отличающиеся от фактических.

Важную роль при оптимизации играет выбор критерия качества, от которого, как отмечалось, во многом зависят конечные результаты расчетов.

Строгое физическое обоснование имеют критерии, характеризующие отклонения фактических значений от расчетных:

$$K(a) = \int_0^{t_{\max}} [x(t) - x_p(t)] dt, \quad (2.3)$$

где $x(t)$ и $x_p(t)$ – соответственно фактические и рассчитанные значения X .

Однако результаты оптимизации на основе этого критерия могут иметь неопределенный характер, если распределение разности $x(t) - x_p(t)$ является симметричным.

В последнее время в различных расчетах за критерий качества часто принимается абсолютная погрешность расчетов:

$$K(a) = \int_0^{t_{\max}} |x(t) - x_p(t)| dt. \quad (2.4)$$

Этот критерий дает более определенные результаты. Однако использовать его в расчетных формулах в некоторых случаях достаточно сложно.

Наиболее часто критерий качества представляется в виде среднего квадрата разностей фактических и рассчитанных значений:

$$K(a) = \int_0^{t_{\max}} [x(t) - x_p(t)]^2 dt. \quad (2.5)$$

Именно на этом критерии основывается принцип наименьших квадратов, широко используемый (см. ниже) в различных выводах.

Для придания большего веса высоким значениям исходной величины иногда используется критерий вида:

$$K(a) = \int_0^{t_{\max}} [x(t) - x_p(t)]^2 x(t) dt. \quad (2.6)$$

В некоторых работах используются нормированные критерии качества, например:

$$K(a) = \frac{(D - D_p)}{D}, \quad (2.7)$$

где D и D_p – соответственно дисперсия исходного и рассчитанного ряда.

В заключение следует отметить, что проблема выбора критерия качества для оптимизации параметров моделей стока тесно связана с проблемой объективной оценки точности расчета исследуемой величины. Например, если оценку точности расчета гидрографа паводка произвести по точности расчета отдельных ординат, то небольшие сдвиги фаз могут дать большие погрешности, и рассматриваемая методика окажется непригодной, несмотря на то, что, в общем, она дает верное представление о форме гидрографа.

На первых этапах развития геоэкологических исследований, когда структуры математических моделей были очень просты и включали, как правило, не более 2–5 параметров, оптимизационные расчеты заключались в последовательном переборе всех возможных значений параметров и выборе тех из них, которые дают минимальное значение критерия качества. Этот простой и надежный метод поиска экстремума критерия качества называется сканированием.

Сканирование – *последовательный перебор всех возможных значений параметров и выбор тех из них, которые дают минимальное значение критерия качества.*

Сканирование относится к методам пассивного поиска, т. е. к тем методам, при которых направление последующего поиска не зависит от результатов предшествующих шагов поиска. Преимущество сканирования состоит в том, что оно позволяет находить истинный минимум критерия качества при любой зависимости (топографии) критерия качества от значений параметров.

Большое распространение в технических приложениях получил метод покоординатного спуска (метод Гаусса–Зейделя). Этот метод состоит в минимизации критерия качества для каждого из параметров при постоянном значении остальных параметров, причем последовательность оптимизации каждого параметра устанавливается в зависимости от его влияния на функцию качества. Значение параметра, дающее минимум критерия качества, используется затем для оптимизации следующего параметра и так до тех пор, пока наблюдается заметное уменьшение критерия качества.

Метод покоординатного спуска дает хорошие результаты при сравнительно простой топографии критерия качества. Например [23], на рис. 2.2 представлены разные топографии критерия качества при двух параметрах a_1 и a_2 . В соответствии с алгоритмом названного метода принимаем начальное значение a_2 на левом графике рисунка равным нулю, и изменяем a_1 , вдоль оси абсцисс от начального значения до тех пор, пока будет наблюдаться уменьшение критерия качества (точка b); затем изменяем a_2 вдоль оси ординат от точки b до тех пор, пока наблюдается уменьшение критерия качества (точка c), затем снова изменяем a_1 и т.д. В случае, представленном на левом графике рис. 2.2.а, критерий качества быстро достигает минимума.

Гораздо сложнее обстоит дело: когда критерий качества имеет более чем один минимум, или если его топография имеет

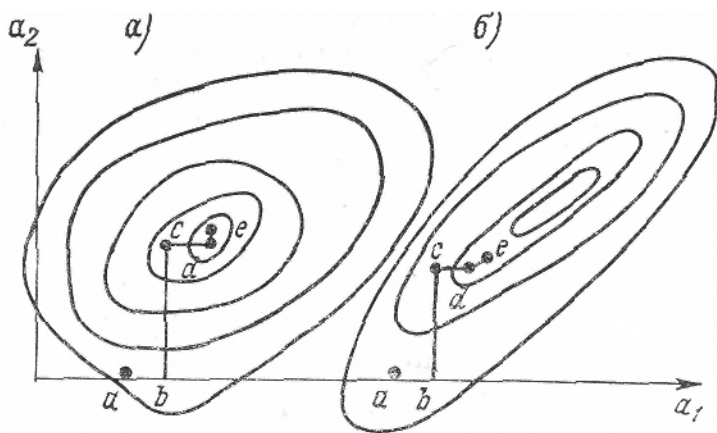


Рис. 2.2. Топография критерия качества: а) обычный тип, б) овражный тип

направление, вдоль которого критерий качества меняется намного медленнее, чем в других направлениях (см. рис. 2.2.б – овражный тип топографии). В этом случае эффективность метода покоординатного спуска резко снижается. На рис. 2.2.б видно, что продвижение к минимуму здесь прекратится задолго до достижения самого минимума.

Для ускорения движения к минимуму используется анализ топографии критерия качества в окрестностях какой-либо точки. Этот анализ заключается обычно в расчете тех или иных градиентов.

2.5. Типы математических моделей в геоэкологии

На выбор математической модели влияют цель и задачи исследования, прошлый опыт, ранее установленные научные истины, полученные знания и т.п. Несмотря на их большое разнообразие, они могут быть классифицированы по виду исследуемых закономерностей, по сложности описываемых процессов, способу задания параметров и т.д.

По виду исследуемых закономерностей математические модели разделяются на динамические (детерминистические), статистические и динамико-статистические.

Динамические математические модели основываются на аналитическом представлении закономерностей исследуемого процесса, при котором для данного множества входных значений может быть получен на выходе только один единственный результат, задаваемый в виде некоторой пространственной или временной функции. При анализе динамических моделей используют главным образом методы математического анализа. Например, одной из первых математических моделей водных экосистем, в основе которых лежал энергетический принцип, была модель Г.Г. Винберга и С.А. Анисимова [51]:

$$\frac{db_i}{dt_j} = c_i - u_i - q_i - m_i - \sum \lambda_{ij} c_j, \quad (2.8)$$

где t – время в сутках, b_i – биомасса, c_i – первичная продукция, u_i – неусвоенная пища, q_i – траты на обмен, m_i – суточная величина отмирания, λ_{ij} – доля i -го компонента в питании j -го; i, j – группы водорослей, зоопланктеров, рыб и бактерий.

В статистических математических моделях функциональные соотношения зависят от случайных параметров. Выходной сигнал такой модели для данного набора входных значений может быть предсказан только в вероятностном смысле. Для анализа статистических математических моделей используется в основном аппарат теории случайной величины. Ярким примером статистической модели в геоэкологии могут служить, например, модели процессов максимального и среднего годового стока в виде случайной величины.

Динамико-статистические модели являются промежуточными между первыми двумя моделями.

В зависимости от сложности описываемых процессов математические модели могут быть простыми и составными. Примерами простых математических моделей могут быть различного рода эмпирические формулы, определяющие, например, максимальный сток в зависимости от площади водосбора и т.д. Примерами составных математических моделей – модели формирования стока, разрабатываемые в последнее время в России и за рубежом.

Математические модели разделяются также по степени учета пространственной изменчивости исследуемых характеристик. Так, если возможно пренебречь пространственной изменчивостью, то используются математические модели с сосредоточенными параметрами. В таких случаях считается, что параметры сосредоточены

в точке, относящейся, например, к центру тяжести речного бассейна. Математические модели, в которых характеристики водосбора задаются изменяющимися в пространстве, называются моделями с распределенными параметрами.

Наконец, математические модели разделяются также по объекту исследования: модели водосбора, модели временного ряда, модели формирования стока и т.д.

2.6. Общее обоснование применений статистических методов в геоэкологии

Процессы, происходящие в окружающей среде, как правило, являются интегральными характеристиками или результатом взаимодействия многих геофизических процессов на сравнительно больших территориях (прямая и рассеянная радиация, осадки, температура воздуха и подстилающей поверхности, давление и влажность воздуха, скорость и направление ветра и т.д.), физико-географических условий бассейна (ландшафт, почвы, геологическое строение, растительность и т.д.), хозяйственной деятельности человека и т.д.

Исходя из этого, можно записать, что

$$A_t = f(B_{1,t-\Delta_1}, B_{2,t-\Delta_2}, \dots, B_{m,t-\Delta_m}), \quad (2.9)$$

где A – рассматриваемое явление, m – число его причин, $B_j (j = 1, 2, \dots, m)$ – причины или комплексы причин, Δ_j – запаздывание воздействия причин.

Если бы все факторы, характер их влияния на процесс A и взаимодействия между собой были известны и однозначны, то для решения уравнения (2.9) могли бы быть использованы методы классической математики, в частности, аппарат дифференциальных уравнений, и на их основе получены точные представления о процессе A в настоящем, будущем и прошлом.

Однако практически ни в одной задаче прикладных и теоретических исследований в геоэкологии и других науках о Земле получить однозначное решение уравнения (2.9) не удалось. Попытки повышения точности решения за счет учета новых факторов, установления новых причинно-следственных связей и т.д. все чаще оказываются безрезультатными. Собственно именно это обстоятельство привело

к широкому распространению методов теории вероятностей и математической статистики в науках о Земле.

В настоящее время практически все исследователи, работающие в этой области, признают необходимость применения статистических методов. Однако о характере их использования существуют различные точки зрения. В соответствии с одной из них статистические методы в науках о Земле есть временное средство, вызванное многофакторностью гидрометеорологических и других процессов и недостаточной изученностью многих из них. В соответствии с другой точкой зрения статистические закономерности объективно отражают природу рассматриваемых процессов и, следовательно, аппарат математической статистики является наиболее подходящим для их описания.

Рассмотрим более детально эти точки зрения. Если проанализировать, например, факторы, определяющие состояние окружающей среды, то, во-первых, окажется, что этих факторов очень много, и конкретное их число определить практически невозможно; во-вторых, более или менее точный учет большинства из этих факторов является пока, а возможно и в будущем, неразрешимой проблемой; при этом точность учета многих факторов (осадки, средняя влажность, почвы в бассейне и т.д.) нередко уступает точности непосредственных измерений самого исследуемого процесса; и, в-третьих, о законах взаимодействия этих факторов и характере их изменения во времени часто известно также мало или даже меньше, чем об изучаемой интегральной характеристике.

Исходя из этого анализа, формулу (2.9) следует переписать в виде:

$$A_t = f(B_{1,t-\Delta_1}, B_{2,t-\Delta_2}, \dots, B_{k,t-\Delta_k}) + \\ + f(B_{k+1,t-\Delta_{k+1}}, B_{k+2,t-\Delta_{k+2}}, \dots, B_{m,t-\Delta_{k+m}}), \quad (2.10)$$

где k – число известных, определенных факторов; $1, 2, \dots, k$ – номера известных факторов; $k+1, k+2, \dots, m$ – номера неопределенных факторов, которые не могут быть пока учтены при описании процесса A .

Именно такое представление о формировании процесса A послужило для многих исследователей обоснованием использования статистических методов в науках о Земле.

Вместе с тем приведенное обоснование вызвало несколько односторонний взгляд на статистические методы как на необходимое,

но временное средство. Действительно, необходимость использования статистических методов здесь объясняется только недостаточной на данном этапе изученностью процессов формирования стока. Исходя из этой точки зрения с повышением степени изученности область применения статистических методов будет постепенно сокращаться и, в конце концов, сойдет на нет.

Однако приведенные соображения совершенно не учитывают ряд обстоятельств, в частности появления элементов случайностей, сопровождающих весь процесс формирования природных процессов.

Если исходить из понятий классической физики, рассматривающей окружающий мир как совокупность отдельных частей, взаимодействующих по однозначным динамическим законам, то появление элементов случайностей невозможно. В мире классической физики представляется, что все связи имеют однозначный характер, и развитие процессов происходит по строгим динамическим законам. Отсюда для полного определения того или иного процесса достаточно установить только эти связи и законы. Однако в современной физике произошли коренные изменения взглядов на окружающий мир. Так, в частности, возникло представление о том, что «... части не являются отдельными независимыми образованиями и ни один объект не может быть индивидуализирован». Отсюда:

«Представление о том, что материю возможно представить в виде совокупности отдельных независимых частей ... заменяется представлением о том, что Вселенная является неделимым целым, отдельные части которого имеют смысл абстракций или приближений, справедливым лишь в классическом пределе» [29].

Интересно, что этот принцип был сформулирован еще Аристотелем в труде «Метафизика», в котором он отмечал, что «Целое является чем-то большим, нежели сумма его частей».

Сейчас, пожалуй, повсеместное признание получили свойства эмерджентности сложных систем

***Эмерджентность** – наличие у данной системы некоторых особых свойств, не присущих ее элементам: блокам и подсистемам, и даже не присущих их сумме.*

Таким образом, свойства больших систем, с которыми мы, несомненно, и в географии и в геоэкологии имеем дело, невозможно свести к сумме свойств их компонентов.

Признание окружающего мира «неделимым целым» и понимание конкретных объектов и процессов природы, как некоторой более

или менее оправданной абстракции, объясняет появление элементов случайностей, а, следовательно, и принципиально вероятностный характер поведения объектов, их вероятностную природу. [4]

Необходимо также отметить, что формирование природных процессов является, как правило, многоступенчатым, т. е. непосредственные причины любого явления в свою очередь имеют свои причины, вплоть до гипотетических абсолютных первопричин (профакторов), находящихся на формально бесконечно удаленной степени формирования. Даже если принять, что каждая причина данного процесса в свою очередь имеет не больше двух причин, то и в этом случае имеем геометрическую прогрессию числа причин от ступени к ступени вплоть до очень большого числа начальных профакторов (рис. 2.3). При этом в процессе взаимодействия причин возникают недетерминированные перекрещивания причинно-следственных цепей, приводящие к возникновению неконтролируемых элементов случайностей ЭС [4].

Возникая на всех ступенях формирования и передаваясь в цепном процессе воздействий со ступени на ступень, ЭС входят, в конце концов, в массовом количестве в рассматриваемое явление. Таким образом, любая элементарная связь природных процессов уже неоднозначна, она включает в себя элементы случайностей, потому что сама по себе не наблюдается в природе, а является лишь, абстракцией, используемой нами в тех или иных целях. Исходя из этого связь сложного процесса, являющегося по своему происхождению

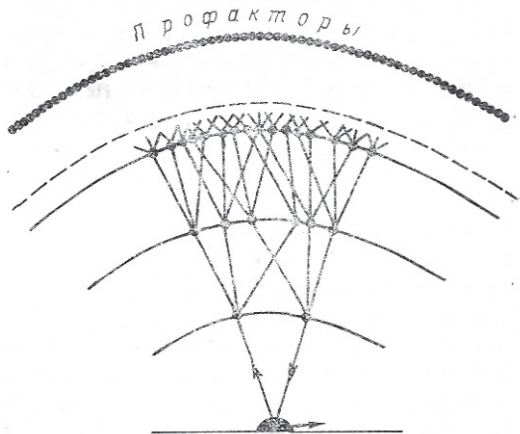


Рис. 2.3. Схема формирования природных явлений

многофакторным и многоступенчатым, с причинами может быть представлена в виде:

$$A_t = \sum_{j=1}^m f_j(B_{jt}) + \sum_{\zeta=1}^n \sum_{i=1}^{k_{\zeta}} \Psi_{\zeta i}, \quad (2.11)$$

где в данном случае n – число промежуточных ступеней формирования между явлением и рассматриваемыми причинами; m – число причин; k_{ζ} – число элементов случайностей на каждой ζ -й ступени формирования; $\Psi_{\zeta i}$ – элементы случайности, появляющиеся на ζ -й ступени при взаимодействии причин.

При этом, чем больше этих причин и большее число промежуточных ступеней имеет процесс формирования, тем больше элементов случайностей включается в данный процесс. Отсюда, например, при $n \rightarrow \infty$ и соответственно при $m \rightarrow \infty$ получим, что

$$A_t = \sum_{\zeta=0}^{\infty} \sum_{i=1}^{k_{\zeta}} \Psi_{\zeta i}, \quad (2.12)$$

т. е. что процесс A почти полностью определяется массой элементов случайностей, влияние же причин, лежащих в основе формирования этого процесса, почти или совсем не проявляется.

Таким образом, вследствие вероятностной природы колебаний стока и вероятностного характера воздействий окружающей среды, причинно-следственные связи колебаний факторов окружающей среды также должны быть вероятностными.

Из приведенных соображений следуют *два важных вывода*.

Во-первых, несмотря на то, что любой процесс формируется вполне определенными причинами и посредством вполне определенных причинно-следственных связей, в некоторых случаях, когда рассматриваются отдаленные причины, связь его с этими причинами из-за многоступенчатости и многофакторности формирования геоэкологических процессов, а также проявления элементов случайностей может стать исчезающе малой, и в конечном итоге определение влияния конкретной причины на развитие процесса может стать невозможным.

Во-вторых, даже если бы мы знали все или, более правильно, все основные причины формирования природных макропроцессов, имеющейся полной информации всё равно было бы недостаточно для полного динамического описания макропроцессов.

Общая неопределенность исследуемого процесса по отношению к внешним и внутренним причинам (в дальнейшем, для

краткости, случайность) может быть выражена через критерий случайности δ :

$$\delta = \frac{D_{\alpha}}{D}, \quad (2.13)$$

где D и D_{α} – соответственно дисперсии ряда значений A и его случайной составляющей, вызванной воздействием элементов случайности.

При $\delta = 0$ влияние внешних и внутренних причин полностью определяет процесс A . Вследствие вероятностной природы и вероятностного характера причинно-следственных связей процессов природы такая ситуация принципиально невозможна. При $\delta = 1$ влияние внешних и внутренних причин не проявляется, развитие процесса полностью определяется элементами случайности. Прогнозирование такого процесса на основе учета этих причин принципиально невозможно. Очевидно, что весь континуум природных процессов по случайности находится между этими крайними значениями δ . Здесь интересный парадокс: с одной стороны, если убрать причины формирования процесса, то он не появится, с другой стороны эти причины могут не влиять на изменения этого процесса

Следует отметить, что опытное значение критерия случайности процесса δ , определяемое по отношению к внутренним и внешним причинам, всегда больше действительного значения. Это объясняется неполнотой и несовершенством научных знаний, не учетом всех действующих факторов и т.д. Очевидно, что с развитием научных исследований значение δ , полученное в результате опыта, должно стремиться к действительному значению.

Таким образом, из проведенного анализа следует, что природа геоэкологических процессов, характер причинно-следственных связей и законов, управляющих ими, являются вероятностными. Именно поэтому обычные для классической теории анализа методы математических исследований, основанные на предположении о динамическом поведении объектов, оказались бессильными для анализа многих геоэкологических процессов. Так, например, не привело и не могло привести к серьезным результатам использование аппарата дифференциальных уравнений, связанного с классической физикой. Действительно, если ни характер изменений во времени, ни закон взаимодействия отдельных факторов не изучены и вряд ли могут быть изучены вообще в классическом измерении на основе однозначных причинно-следственных связей, то применение

аппарата дифференциальных уравнений, по выражению Б.В. Гнеденко [11], носит элементы грубого произвола.

Очень часто в различного рода исследованиях противопоставляются статистические и физические (генетические) методы. При этом указывается, что во всех науках естествознания общепринятыми являются физические методы и только в геоэкологии отдается предпочтение статистическим методам исследований. По этому поводу следует отметить, что противопоставление статистических и физических методов исследования было в какой-то степени оправдано до середины 30-х годов прошлого века. Однако с тех пор произошел коренной пересмотр отношения к использованию статистического аппарата во многих науках естествознания, в том числе в физике и генетике, связанный с пересмотром соотношений статистических и динамических закономерностей в природе. Если раньше назвать метод исследований физическим означало, что в основе его лежит использование динамической модели в виде той или иной функциональной зависимости, то сейчас это было бы неверно. Например, крупнейший физик нашего времени М. Борн в книге «Физика в жизни моего поколения» отмечает: «Статистические методы в физике по мере развития науки распространялись все больше и больше, и сегодня можно сказать, что современная физика полностью опирается на статистическую основу. Это является событием в истории человеческого мышления, значение которого выходит за рамки самой науки».

Глава 3. Теория вероятностей и математическая статистика в экологических исследованиях

3.1. Случайные величины и их характеристики

Одним из основных понятий теории вероятностей является понятие случайной величины. С этим понятием приходится иметь дело в самых разнообразных областях науки и техники. Во многом на этом понятии основываются и методы обработки и анализа экологической информации.

3.1.1. Определения и обозначения

Существует множество определений этого понятия. В технических приложениях теории вероятностей обычно используются следующие определения [34].

Случайная величина – величина, которая в результате испытания может принимать различные значения в зависимости от случайного исхода испытания.

Случайная величина – величина, которая в результате опыта может принять одно из возможных значений, неизвестно заранее, какое именно.

Следует обратить внимание на то, что это чисто математическое понятие. В определениях случайной величины ничего не говорится о происхождении отдельных ее значений. Случайны они в житейском смысле или достаточно точно определены внешними причинно-следственными связями в данном случае безразлично. Может быть и то и другое. Важно лишь то, что, зная n предшествующих значений случайной величины, мы не можем сказать, какое из возможных значений примет ее $n + i$ значение, где i может быть равно 1, 2 и т.д.

Случайные величины могут быть одномерными и многомерными.

Одномерная случайная величина – величина, которая в результате испытания может приобретать только одно возможное значение.

Многомерная случайная величина – величина, которая в результате испытания может принимать два и более возможных значения.

Одномерная случайная величина, а следовательно и процесс, который представляют в качестве последовательности значений одномерной случайной величины, обычно обозначается прописной латинской буквой X , Y и т.д. Последовательность значений случайной величины, принимаемая ею в результате опыта, обозначают строчными буквами с индексом $x_1, x_2, \dots, x_n$, где $1, 2, \dots, n$ – номер испытания.

Например, процесс изменения годовых сумм осадков X в каком-то конкретном пункте за имеющийся период наблюдений в n лет может быть представлен последовательностью значений $x_1, x_2, \dots, x_n$, или в другой записи $x_i (i = 1, 2, \dots, n)$.

Примером многомерной случайной величины может служить распределение дождевых осадков в каком-то районе. Пусть в этом районе имеется m пунктов наблюдений с периодом наблюдений n лет.

Тогда ряд наблюдений в первом пункте: $x_{11}, x_{12}, \dots, x_{1n}$ или $x_{1i} (i = 1, 2, \dots, n)$; соответственно в j -м пункте: $x_{j1}, x_{j2}, \dots, x_{jn}$ или $x_{ji} (i = 1, 2, \dots, n)$ и в m -м пункте $x_{m1}, x_{m2}, \dots, x_{mn}$ или $x_{mi} (i = 1, 2, \dots, n)$.

Например, в общем виде осадки в каком-то районе могут быть представлены в виде таблицы:

$$\begin{array}{cccccc} x_{11} & x_{12} & \dots & x_{1i} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2i} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{j1} & x_{j2} & \dots & x_{ji} & \dots & x_{jn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mi} & \dots & x_{mn} \end{array}, \quad (3.1)$$

которая называется матрицей наблюдений.

В краткой записи данные этой таблицы обозначаются:

$$x_{ji} (j = 1, 2, \dots, m; i = 1, 2, \dots, n). \quad (3.2)$$

3.1.2. Дискретные и непрерывные случайные величины

Различают два типа случайных величин: непрерывные и дискретные.

Непрерывная случайная величина может принимать любые численные значения в некотором диапазоне, который, в частности, может быть и бесконечным.

Примером непрерывной случайной величины служат многие характеристики стока; средний годовой, максимальный, минимальный и т.д. Значения этих характеристик могут быть по сути дела любыми в заданном интервале.

Дискретная случайная величина может принимать только конечное или счетное множество значений.

Примером дискретной случайной величины служит современная шкала облачности. Как известно, облачность измеряется в баллах от 0 до 10. В соответствии с этим и ряд наблюдений значений облачности может состоять только из следующих чисел: 0, 1, 2, ..., 10, каждое из которых может повторяться множество раз.

На практике непрерывные случайные величины, учитывая точность измерений и необходимые округления, часто заменяются дискретными. Например, значения стока регистрируются с точностью до трех значащих цифр. Следовательно, период дискретности значения стока, например, при изменении его от 10,0 до 99,9 равен 0,1.

Рассмотрим одномерную дискретную случайную величину X , которая имеет k возможных значений $\zeta_1, \zeta_2, \dots, \zeta_k$. События, в результате которых X примет каждый раз одно и только одно из возможных значений $\zeta_i (i = 1, 2, \dots, k)$, являются несовместными (никакие два из них не могут появиться вместе) и образуют полную группу событий (в результате опыта непременно должно появиться хотя бы одно из них). Каждое из этих событий возможно, но не достоверно, т. е. X может принять каждое из них с какой-то вероятностью p . Обозначим вероятность события $X = \zeta_i$, т. е. $p(X = \zeta_i)$, через p_i . Так как множество всех несовместных событий образуют полную группу, то

$$\sum_{i=1}^k p_i = 1, \quad (3.3)$$

т. е. сумма вероятностей всех возможных значений случайной величины равна 1.

Так, например, в каждый данный момент времени обязательно наблюдается одно из значений облачности от 0 до 10. Представляет практический интерес вероятность каждого из этих значений облачности, т. е. каким образом вероятность, равная 1, распределена между отдельными возможными значениями облачности. Конечно, было бы лучше указать точное значение облачности, допустим на завтра, однако прогнозы с такой точностью чаще всего невозможны. Поэтому на практике приходится обходиться именно такой вероятностной оценкой случайной величины.

Считается, что случайная величина полностью описана с вероятностной точки зрения, если задана вероятность каждого из ее возможных значений. Этим устанавливается так называемый закон распределения случайной величины.

3.1.3. Общая характеристика законов распределения

***Закон распределения случайной величины** – всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.*

Законы распределения можно задавать в табличной, графической и аналитической форме. В последнее время широко распространены алгоритмические методы задания.

Простейшей формой задания закона распределения дискретной случайной величины является таблица, называемая **рядом распределения**, в которой перечислены возможные значения случайной величины ζ и соответствующие им вероятности p (табл. 3.1).

Таблица 3.1. Ряд распределения случайной величины X

x_i	ζ_1	ζ_2	ζ_3	...	ζ_k
p_i	p_1	p_2	p_3	...	p_k

Более наглядной, но менее компактной формой представления закона распределения является графическое изображение, при этом по оси абсцисс откладываются возможные значения случайной величины, а по оси ординат – вероятности этих значений. Для наглядности полученные точки соединяются отрезками прямых линий. Полученная фигура называется **многоугольником** или **полигоном распределения** (рис. 3.1).

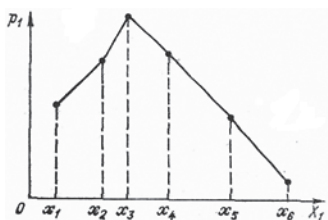


Рис. 3.1. Многоугольник распределения

Многоугольники распределения могут иметь самую различную форму, однако все они обладают одним общим свойством: сумма ординат многоугольника распределения, представляющая собой сумму вероятностей всех возможных значений случайной величины, всегда равна единице.

Ряд и многоугольник распределения являются исчерпывающими характеристиками дискретной случайной величин. Для непрерывной случайной величины использование этих характеристик невозможно, так как непрерывная величина имеет бесчисленное множество значений, сплошь заполняющих некоторый промежуток (см. выше). Составить таблицу, в которой были бы перечислены все возможные значения такой величины, невозможно. Кроме того, каждое отдельное значение непрерывной случайной величины обычно не обладает никакой отличной от нуля вероятностью.

Для количественного описания закона распределения непрерывной случайной величины используется вероятность события $X < x$, где x — некоторое значение переменной X . Эта вероятность зависит от значения x , то есть является некоторой функцией x , отсюда и название новой характеристики — функция распределения.

Функция распределения $F(x)$ случайной величины X — вероятность выполнения неравенства $X < x$ при различных значениях x , то есть:

$$F(x) = p(X < x). \quad (3.4)$$

Часто функцию распределения называют **интегральной функцией распределения или интегральным законом распределения**.

Функция распределения — общая форма описания случайной величины, полностью характеризующая ее с вероятностной точки зрения. Она может быть использована как для дискретных, так и для непрерывных случайных величин.

В экологических расчетах вместо функции распределения принято рассматривать функцию:

$$P(x) = p(X \geq x) = 1 - F(x). \quad (3.5)$$

Для отличия от функции распределения, обычно используемой в математической литературе, назовем эту функцию — **функцией**

обеспеченности, а ее графическое изображение – **кривой обеспеченности**.

Далее следует отметить некоторые общие свойства функции распределения.

1. Функция распределения $F(x)$ есть неубывающая функция своего аргумента, т. е. при $x_2 \geq x_1$:

$$F(x_2) \geq F(x_1). \quad (3.6)$$

2. На минус бесконечности ($x = -\infty$) функция распределения равна нулю, т. е.

$$F(-\infty) = 0. \quad (3.7)$$

3. На плюс бесконечности ($x = +\infty$) функция распределения равна единице, т. е.

$$F(+\infty) = 1. \quad (3.8)$$

Доказательства перечисленных свойств приводятся в курсе теории вероятностей, например [11]. Упрощенное доказательство этих свойств производится с помощью геометрической интерпретации случайной величины в виде случайной точки X на оси OX . Точка X в результате опыта может занять то или иное положение. Во всех случаях функция распределения $F(x)$ есть вероятность того, что случайная точка X будет находиться левее заданного x .

Будем увеличивать X , т. е. перемещать точку x вправо по оси OX . Очевидно, что с увеличением области значений X по числовой оси OX , вероятность попадания X в эту область не может уменьшаться, она или возрастает, или остается неизменной.

Будем дальше смещать точку x вправо по оси OX . Вместе с этим будет возрастать область возможных значений X , и в пределе при $x = \infty$ все возможные значения X окажутся левее x , т. е. события $X < x$ станут достоверными. Отсюда вероятность значений $X < \infty$, т. е. функция распределения при $x = +\infty$ $F(X < \infty)$ равна единице.

Аналогично, перемещая точку x влево, получаем, что $F(X = \infty) = 0$.

Исходя из этих свойств можно получить общее представление о графике функции распределения. Очевидно, он должен представлять собой график неубывающей функции, значения которой начинаются от 0 и доходят до 1 (рис. 3.2).

Можно показать, что для функции обеспеченности, установленные выше соотношения будут иметь обратный вид.

1. Функция обеспеченности $P(x)$ есть убывающая (точнее, невозрастающая) функция своего аргумента, т. е. при $x_2 > x_1$ $P(x_2) \leq P(x_1)$.

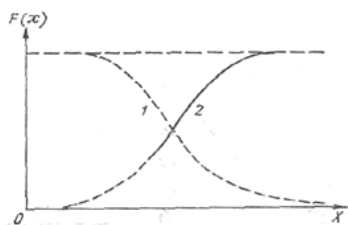


Рис. 3.2. График функции обеспеченности (1) и распределения (2)

2. При $x = \infty$ функция обеспеченности равна единице, т. е., $P(\infty) = 1$.

3. При $x = -\infty$ функция обеспеченности равна нулю, т. е., $P(-\infty) = 0$.

Исходя из этого, очевидно, что кривая обеспеченности симметрична кривой функции распределения и пересекается с ней при $F(x) = P(x) = 0,5$ (рис. 3.2).

По известным функциям распределения или обеспеченности может быть установлена вероятность появления случайной величины в заданном интервале, полузамкнутом слева. Как следует из курса теории вероятностей, она равна разности значений функции распределения в концах интервала, т. е.

$$p(\alpha \leq x < \beta) = F(\beta) - F(\alpha). \quad (3.9)$$

Тогда из выражения (3.5) также следует, что

$$p(\alpha \leq x < \beta) = P(\alpha) - F(\beta). \quad (3.10)$$

Часто событие, заключающееся в появлении случайной величины в интервале (α, β) , называют попаданием случайной величины на заданный участок.

Функция распределения случайной величины является ее исчерпывающей характеристикой, но по ней трудно судить о характере распределения случайной величины в небольшой окрестности той или другой точки числовой оси. Для дискретной случайной величины с этой целью можно воспользоваться многоугольником или полигоном распределения. Для непрерывной случайной величины наглядное представление дает функция, называемая плотностью распределения вероятности или дифференциальным законом распределения.

Плотность распределения — производная функции распределения вероятностей

$$f(x) = F'(x) = \frac{dF(x)}{dx} \quad (3.11)$$

в тех точках, где она существует.

Кривая, изображающая плотность распределения случайной величины, называется **кривой распределения**.

Понятие плотности вероятности можно обобщить и на дискретные случайные величины, если воспользоваться дельта-функцией Дирака $\delta(x)$:

$$\delta(x) = \begin{cases} \infty, & x = 0, \\ 0, & x \neq 0. \end{cases} \quad (3.12)$$

Так, для дискретной случайной величины, закон распределения которой задан рядом распределения x_j , ($j = 1, N$), плотность вероятности определяется по формуле:

$$f(x) = \sum_{j=1}^N P_j \delta(x - x_j). \quad (3.13)$$

Таким образом, плотность дискретной случайной величины равна нулю всюду, кроме точек $x_1, x_2, \dots, x_N$, в которых она «делает всплески в бесконечность», причем «площади всплесков» в точках $x_1, x_2, \dots, x_N$ равны соответственно $p_1, p_2, \dots, p_N$.

Рассмотрим непрерывную случайную величину X с плотностью вероятности $f(x)$. Вероятность попадания случайной величины X на элементарный участок dx , примыкающий к точке X (рис. 3.3), с точностью до бесконечно малых высшего порядка равна $f(x)dx$. Величина $f(x)dx$ называется элементом вероятности. Геометрически это площадь элементарного прямоугольника, опирающегося на отрезок dx (рис. 3.3).

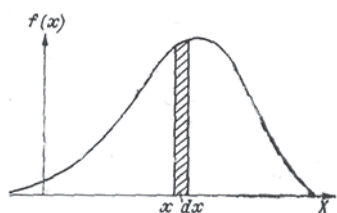


Рис. 3.3. Кривая распределения и элемент вероятности

Вероятность попадания значений x на отрезок от α до β , очевидно, будет равна сумме элементов вероятности на всем этом участке, т. е.

$$p(\alpha \leq x < \beta) = \int_{\alpha}^{\beta} f(x)dx. \quad (3.14)$$

Геометрически вероятность попадания значений X на участок (α, β) равна площади под кривой распределения, опирающейся на этот участок (см. рис. 3.3).

Теперь выразим функцию распределения через плотность вероятности. По определению $F(x) = p(X < x) = p(-\infty \leq X < x)$. Отсюда с учетом формулы (3.14):

$$F(x) = \int_{-\infty}^x f(x) dx. \quad (3.15)$$

Геометрически функция распределения, очевидно, есть площадь кривой распределения, лежащая левее точки X .

Рассмотрим основные свойства плотности распределения.

1. Плотность распределения есть неотрицательная функция, т. е.

$$f(x) \geq 0. \quad (3.16)$$

Это свойство является следствием того, что функция распределения есть неубывающая функция. Поэтому числитель выражения плотности вероятности (3.11) больше или равен нулю, а знаменатель ($\Delta x_i = x_i - x_{i-1}$) тоже положителен.

2. Интеграл в бесконечных пределах от плотности распределения равен единице:

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (3.17)$$

Это следует из формулы (3.15) и из того, что $F(x = \infty) = 1$. Геометрически (см. рис. 3.3) основные свойства плотности распределения означают, что вся кривая распределения лежит не ниже оси абсцисс, и полная площадь, ограниченная кривой распределения и осью абсцисс, равна единице.

3.1.4. Числовые характеристики случайных величин

Полное представление о случайной величине дает ряд распределений или функция распределения. Однако ряд и функция распределения слишком громоздки и неудобны для практического использования, особенно при сравнительном анализе различных рядов наблюдений. Между тем, во многих случаях бывает вполне достаточно указать только отдельные параметры, характеризующие какие-либо существенные в данном случае черты распределения исследуемой случайной величины. Например, среднее значение,

относительно которого располагаются все возможные значения случайной величины; характеристики рассеивания этих значений относительно среднего и т.д.

Следует также отметить, что при описании конкретных рядов наблюдений законы распределения (см. гл. 4) часто задают в аналитической форме, для чего необходимы какие-то параметры, иначе числовые характеристики, отражающие те или иные стороны этого закона.

Числовые характеристики или параметры распределения случайной величины – характеристики, назначение которых – выразить в сжатой форме наиболее существенные особенности распределения.

В инженерных расчетах числовые характеристики и операции с ними имеют чрезвычайно большое значение. Очень часто удается решить ту или иную задачу до конца, оставляя в стороне законы распределения и оперируя одними числовыми характеристиками.

В теории вероятностей и математической статистике используются характеристики, имеющие различное назначение и различные области применения. Большинство из них основано на понятии моментов распределения.

Моменты статистического распределения

Понятие моментов в статистических расчетах во многом аналогично понятию моментов в механике, применяемому для описания распределения масс относительно какой-либо точки отсчета.

Моменты статистических распределений могут определяться относительно любой точки отсчета. Однако чаще всего используются моменты двух видов: начальные, определяемые относительно нуля исследуемой величины, и центральные, определяемые относительно среднего значения этой величины.

Начальный момент S -го порядка дискретной случайной величины X – сумма вида

$$\alpha_s[X] = \sum_{j=1}^k \zeta_j^s p_j, \quad (3.18)$$

где ζ_j – возможные значения X ; p_j – вероятность появления значения X , равного ζ_j ; k – число возможных значений X .

Начальный момент s -го порядка непрерывной случайной величины X – интеграл вида:

$$\alpha_s[X] = \sum_{-\infty}^{\infty} x^s f(x) dx, \quad (3.19)$$

где $f(x)$ – плотность распределения X .

Первый начальный момент, представляющий собой сумму произведений всех возможных значений случайной величины на их вероятности, называется **математическим ожиданием (m_x)**. Для дискретных случайных величин:

$$m_x = M[X] = \alpha_1[X] = \sum_{j=1}^k \zeta_j p_j, \quad (3.20)$$

для непрерывных:

$$m_x = \sum_{-\infty}^{\infty} x f(x) dx. \quad (3.21)$$

С учетом выражений (3.18) и (3.20) можно записать, что

$$\alpha_s[X] = M[X^s]. \quad (3.22)$$

Отсюда другое определение начального момента:

Начальный момент S -го порядка случайной величины X – математическое ожидание (м. о.) S -й степени этой величины.

Для определения центральных моментов вводится понятие центрированной случайной величины:

$$\overset{\circ}{x}_i = \delta_i = x_i - m_x, \quad (3.23)$$

представляющей из себя отклонение от математического ожидания.

При анализе графиков центрирование аналогично переносу начала координат в среднюю, «центральную», точку, абсцисса которой равна m_x . В гидрометеорологических расчетах центрированные величины часто называют отклонениями от среднего или от нормы.

Моменты центрированной случайной величины называются центральными моментами. Они аналогичны моментам относительно центра тяжести в механике.

Центральный момент порядка s случайной величины X – математическое ожидание (м. о.) S -й степени соответствующей центрированной случайной величины:

$$\mu_s[X] = M\left[\overset{\circ}{X}^s\right] = M\left[(X - m_x)^s\right] \quad (3.24)$$

или, согласно выражению (3.20), для дискретной случайной величины:

$$\mu_s[X] = \sum_{j=1}^k (\zeta_j - m_x)^s p_j, \quad (3.25)$$

и с учетом формулы (3.21) для непрерывной случайной величины:

$$\mu_s[X] = \int_{-\infty}^{\infty} (x - m_x)^s f(x) dx = \int_{-\infty}^{\infty} \overset{\circ}{x}^s f(x) dx. \quad (3.26)$$

Второй центральный момент называется **дисперсией**.

Легко показать, что первый центральный момент равен нулю. Действительно:

$$\mu_s[X] = \sum_{j=1}^k (\zeta_j - m_x)^1 p_j = \sum_{j=1}^k \zeta_j p_j - \sum_{j=1}^k m_x p_j = 0.$$

Точно также можно показать, что нулевой центральный момент равен единице.

В дальнейшем нам понадобятся соотношения, связывающие второй и третий центральный и начальный момент:

$$\begin{aligned} \mu_2 &= \sum_{i=1}^N (x_i - m_x)^2 p_i = \sum_{i=1}^N x_i^2 p_i - 2m_x \sum_{i=1}^N x_i p_i + m_x^2 \sum_{i=1}^N p_i = \\ &= \alpha_2 - 2m_x^2 + m_x^2 = \alpha_2 - m_x^2. \end{aligned}$$

Аналогично для третьего центрального момента:

$$\begin{aligned} \mu_3 &= \sum_{i=1}^N (x_i - m_x)^3 p_i = \sum_{i=1}^N x_i^3 p_i - 3m_x \sum_{i=1}^N x_i^2 p_i + 3m_x^2 \sum_{i=1}^N x_i p_i - m_x^3 \sum_{i=1}^N p_i = \\ &= \alpha_3 - 3\alpha_2 m_x + 2m_x^2. \end{aligned}$$

Таким образом, для начальных и центральных моментов случайной величины характерны следующие соотношения:

$$\mu_0 = 1,$$

$$\mu_1 = 0,$$

$$\mu_2 = \alpha_2 - m_x^2,$$

$$\mu_3 = \alpha_3 - 3\alpha_2 m_x + 2m_x^2.$$

Числовые характеристики распределения

Числовые характеристики, используемые в практике расчетов, разделяют на характеристики положения, показывающие расположение определенных характеристик распределения на числовой оси X , и характеристики рассеивания.

Из характеристик положения важнейшую роль во всех расчетах играет математическое ожидание m_x . (Еще одно возможное обозначений $M[X]$). Кроме того, часто используются еще две характеристики положения: мода M и медиана Me .

Мода дискретной случайной величины – значение, имеющее наибольшую вероятность.

Для непрерывной случайной величины это определение не подходит, так как вероятности каждого отдельного значения случайной величины равны нулю (см. определение непрерывной случайной величины).

Мода непрерывной случайной величины – значение, имеющее наибольшую плотность вероятности.

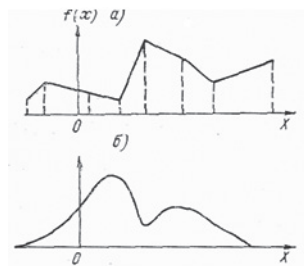


Рис. 3.4. Полимодальное распределение дискретной (а) и непрерывной (б) случайной величины.

Наиболее удобно определять моду по многоугольнику или соответственно кривой распределения.

Если многоугольник или кривая распределения имеют более одного максимума (рис. 3.4.а), то распределение называется полимодальным. Иногда встречаются распределения, имеющие посередине не максимум, а минимум (рис. 3.4.б). Такое распределение называется амодальным.

Другая часто применяемая характеристика положения – медиана Me .

Медиана случайной величины X – такое ее значение, для которого $p(X < Me) = p(X > Me)$, т. е. вероятность значений X , больших и меньших Me , одинакова и, следовательно, равна 0,5.

Геометрически медиана – это абсцисса точки, в которой площадь, ограниченная кривой плотности вероятности, делится пополам. Обычно значение Me определяют по функции распределения или обеспеченности.

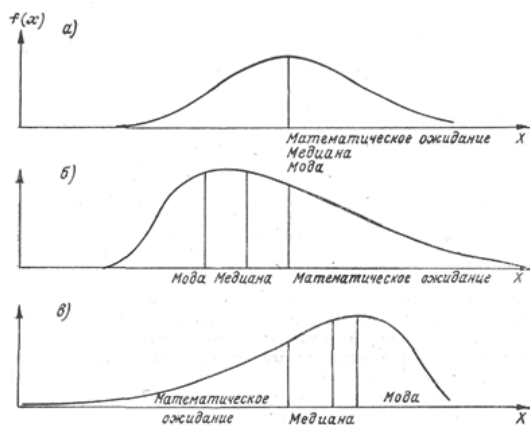


Рис. 3.5. Соотношение между математическим ожиданием, медианой и модой для симметричного распределения (а), распределение с положительной (б) и отрицательной (в) асимметрией

В общем случае математическое ожидание, мода и медиана не совпадают (рис. 3.5), однако в случае симметричного одномодального распределения эти характеристики могут совпадать. В частности, они совпадают для нормального закона распределения.

Другим классом числовых характеристик являются характеристики рассеивания: дисперсия, среднее квадратическое отклонение, коэффициент вариации, коэффициент асимметрии и реже эксцесс.

Дисперсия (см. выше) представляет собой второй центральный момент:

$$D_x = \mu_2[X] = M[\overset{\circ}{X}^2]. \quad (3.28)$$

Для непосредственного вычисления дисперсии служат следующие формулы.

Для непрерывных случайных величин:

$$D_x = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx; \quad (3.29)$$

для дискретных случайных величин:

$$D_x = \sum_{j=1}^k (\zeta_j - m_x)^2 p_j. \quad (3.30)$$

Отсюда определение:

Дисперсия случайной величины X – математическое ожидание квадрата соответствующей центрированной величины.

Для характеристики рассеивания часто удобнее пользоваться величиной, размерность которой совпадает с размерностью случайной величины. Это дает возможность сопоставить рассеивание со значениями самих величин X . В качестве такой величины может служить среднее квадратическое отклонение или стандарт.

Среднее квадратическое отклонение или стандарт – корень квадратный из дисперсии:

$$\sigma_x = \sqrt{D_x} = \sqrt{\mu_2[X]}. \quad (3.31)$$

Для описания заведомо положительных величин большое пространство получила безразмерная характеристика рассеивания – **коэффициент вариации**:

$$C_v = \frac{\sigma_x}{m_x}. \quad (3.32)$$

Коэффициент вариации не зависит от среднего значения и может быть использован для сопоставления изменчивости различных процессов.

Для характеристики симметричности рассеивания значений случайной величины относительно математического ожидания применяется безразмерная величина – **коэффициент асимметрии**:

$$C_s = \frac{\mu_3}{\sigma^3}. \quad (3.33)$$

В соответствии с этой формулой для симметричных распределений, когда каждому $+x$ соответствует $-x$, кубы x при суммировании уничтожаются и $C_s = 0$. Асимметрия может быть положительной и отрицательной. Знак асимметрии зависит от соотношения числовых характеристик положения: моды и математического ожидания. Если $m_x > M$, то $C_s > 0$ – положительная асимметрия. Если $m_x < M$ – отрицательная асимметрия (см. рис. 3.5).

Примером распределений с положительной асимметрией могут служить ряды концентраций химических элементов в речной воде. В данном случае причиной положительной асимметрии является то, что значения концентраций всегда больше и равны нулю и могут в несколько раз превосходить математическое ожидание. Таким

образом, максимум отрицательных отклонений от математического ожидания по абсолютной величине всегда меньше математического ожидания, в то время как положительные отклонения от математического ожидания нередко в несколько раз больше него.

Примером распределений с отрицательной асимметрией могут служить многие ряды максимальных уровней, особенно на участках рек с поймой. Причиной отрицательной асимметрии в данном случае является резкое возрастание ширины живого сечения при выходе воды на пойму.

Для характеристики крутости (островершинности или плоско-вершинности) распределения используется эксцесс, основанный на учете 4-го центрального момента:

$$E_x = \frac{\mu_4}{\sigma^4 - 3}. \quad (3.34)$$

Для нормального закона распределения $E_x = 0$.

В практических расчетах в формулах математического ожидания (3.20), дисперсии (3.30), среднего квадратического отклонения (3.31), коэффициента вариации (3.32), коэффициента асимметрии (3.33) вместо p_j – вероятности появления значения X , равного ζ_j в N испытаниях – подставляется ее значение по формуле:

$$P_j = \frac{m_j}{N}, \quad (3.35)$$

где m_j число появлений значений X , равных ζ_j в N испытаниях.

Легко показать, что в этом случае перечисленные формулы принимают следующий вид:

- математическое ожидание:

$$m_x = \sum_{i=1}^N \frac{x_i}{N}; \quad (3.36)$$

- дисперсия:

$$D_x = \sum_{i=1}^N \frac{(x_i - m_x)^2}{N}; \quad (3.37)$$

- среднее квадратическое отклонение:

$$\sigma_x = \sqrt{D_x} = \sqrt{\sum_{i=1}^N \frac{(x_i - m_x)^2}{N}}; \quad (3.38)$$

- коэффициент вариации:

$$C_v = \frac{\sigma_x}{m_x} = \sqrt{\sum_{i=1}^N \frac{(x_i - m_x)^2}{m_x^2}}; \quad (3.39)$$

- коэффициент асимметрии:

$$C_s = \frac{\sum_{i=1}^N \frac{(x_i - m_x)^3}{\sigma_x^3}}{N}. \quad (3.40)$$

3.1.5. Свойства числовых характеристик случайных величин

Рассмотрим основные свойства числовых характеристик, используемые в расчетах.

Свойства математического ожидания

1. Математическое ожидание постоянной величины равно самой этой величине.

Действительно, постоянная величина c может рассматриваться как частный вид случайной величины при одном возможном значении с вероятностью $p = 1$. Но тогда по формуле математического ожидания (3.20):

$$M[c] = c * 1 = c. \quad (3.41)$$

2. Постоянный множитель можно вынести за знак математического ожидания, т. е.

$$M[cX] = cM[X]. \quad (3.42)$$

Приведем доказательство этого положения для дискретных случайных величин, используя формулу математического ожидания (3.20):

$$M[cX] = \sum_{i=1}^N cX_i p_i = c \sum_{i=1}^N X_i p_i = cM[X].$$

3. Математическое ожидание суммы ряда случайных независимых величин равно сумме их математических ожиданий, т. е.

$$M\left[\sum_i X_i\right] = \sum_{i=1}^N M[X_i]. \quad (3.43)$$

Докажем это свойство для двух случайных величин X и Y .

$$\begin{aligned} M[X+Y] &= \sum_j \sum_i (\zeta_j + \xi_i) p_{ij} = \sum_j \sum_i \zeta_j p_{ij} + \sum_j \sum_i \xi_i p_{ij} = \\ &= \sum_j \zeta_j \sum_i p_{ji} + \sum_i \xi_i \sum_j p_{ij}, \end{aligned}$$

где ζ_j и ξ_i – возможные значения случайных величин X и Y .

Отсюда, так как p_{ji} – вероятность совместного появления величин ζ_j и ξ_i , получаем:

$$\sum_i p_{ji} = p_j \sum_j p_{ji} = p_i$$

и

$$M[X+Y] = \sum_j \zeta_j p_j + \sum_i \xi_i p_i = M[X] + M[Y].$$

Аналогично методом полной индукции можно доказать соотношение (3.43) для произвольного числа слагаемых.

4. Математическое ожидание линейной функции нескольких случайных величин равно той же линейной функции от математических ожиданий аргументов, т. е.

$$M\left[\sum_i a_i X_i + b\right] = \sum_i a_i M[X_i] + b. \quad (3.44)$$

Докажем это предположение на основе первых трех свойств математического ожидания:

$$M\left[\sum_i a_i X_i + b\right] = M\left[\sum_i a_i X_i\right] + M[b] = \sum_i a_i [X_i] + b.$$

5. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий, т. е.

$$M\left[\prod_i X_i\right] = \prod_i M[X_i]. \quad (3.45)$$

Докажем это свойство для двух дискретных случайных величин X и Y :

$$M[XY] = \sum_i \sum_j \xi_i \zeta_j p_{ij},$$

где p_{ij} – вероятность совместного появления ξ_i и ζ_j , равная для независимых случайных величин произведению $p_i p_j$. Исходя из этого:

$$M[XY] = \sum_i \zeta_i p_i \sum_j \xi_j p_j = M[X]M[Y].$$

Отсюда методом полной индукции можно получить формулу произведения математических ожиданий и для произвольного числа независимых случайных величин.

Свойства дисперсии случайной величины

1. Дисперсия постоянной величины равна нулю, т. е.

$$D[c] = 0. \quad (3.46)$$

Доказательство: так как математическое ожидание c (см. первое свойство математического ожидания) равно c , то

$$D[c] = M[(c - m_c)^2] = M[(c - c)^2] = 0.$$

2. Постоянную величину можно выносить за знак дисперсии, возведя ее в квадрат, т. е.

$$D[cX] = c^2 D[X]. \quad (3.47)$$

Доказательство:

$$D[cX] = M[(cX - M[cX])^2] = M[(cX - cm_x)^2] = c^2 D[X].$$

Отсюда же следует, что

$$\sigma[cX] = c\sigma[X], \quad (3.48)$$

т. е. постоянную величину можно выносить за знак среднего квадратического отклонения,

3. Дисперсия суммы случайных независимых величин равна сумме дисперсий слагаемых:

$$D\left[\sum_i X_i\right] = \sum_i D[X_i]. \quad (3.49)$$

Докажем это положение для двух случайных величин X и Y :

$$\begin{aligned} D[X + Y] &= M\left\{\left[(X + Y) - M(X + Y)\right]^2\right\} = \\ &= M\left\{\left[(X - m_x) + (Y - m_y)\right]^2\right\} = M\left[(X - m_x)^2\right] + \\ &\quad + 2M\left[(X - m_x)(Y - m_y)\right] + M\left[(Y - m_y)^2\right]. \end{aligned}$$

Так как X и Y независимы, то

$$M\left[(X - m_x)(Y - m_y)\right] = M[X - m_x] + M[Y - m_y] = 0.$$

и

$$D[X + Y] = M\left[(X - m_x)^2\right] + M\left[(Y - m_y)^2\right] = D[X] + D[Y]. \quad (3.50)$$

4. Дисперсия линейной функции нескольких независимых случайных величин выражается формулой:

$$D\left[\sum_i a_i X_i + b\right] = \sum_i a_i^2 D[X_i]. \quad (3.51)$$

Для доказательства используем формулу дисперсии суммы случайных независимых величин:

$$D\left[\sum_i a_i X_i + b\right] = \sum_i D[a_i X_i] = \sum_i a_i^2 D[X_i].$$

3.1.6. Нормированная (стандартизованная) случайная величина

Во многих практических задачах, например, при сопоставлении асимметричности рядов стока и их законов распределения, формализованной записи того или иного ряда распределения требуется освободиться от влияния среднего значения m_x и среднего квадратического отклонения σ_x . Это достигается нормированием значений случайной величины.

Нормированной случайной величиной называется переменная вида:

$$t_i = \frac{x_i - m_x}{\sigma_x} = \frac{(x_i - m_x)}{\sigma_x}. \quad (3.52)$$

Если X положительно ($X \geq 0$) то, разделив числитель и знаменатель на m_x , получим:

$$t_i = \frac{(k_i - 1)}{C_v}, \quad (3.53)$$

где $k_i = x_i/m_x$ – модульный коэффициент.

Последовательность нормированных случайных величин обладает рядом замечательных свойств.

1. Математическое ожидание нормированной случайной величины равно 0, т. е. $m[t] = 0$. Действительно:

$$\begin{aligned} m[t] &= \sum_{j=1}^k t_j p_j = \left(\frac{1}{\sigma_x} \right) \sum_{j=1}^k (\zeta_j - m_x) p_j = \\ &= \left(\frac{1}{\sigma_x} \right) \left(\sum_{j=1}^k \zeta_j p_j - \sum_{j=1}^k m_x p_j \right) = 0. \end{aligned} \quad (3.54)$$

2. Дисперсия нормированной случайной величины равна единице, т. е. $D_t = 1$. Действительно:

$$D_t = \sum_{j=1}^k t_j^2 p_j = \left(\frac{1}{\sigma_x^2} \right) \sum_{j=1}^k (\zeta_j - m_x)^2 p_j = \frac{\sigma_x^2}{\sigma_x^2} = 1. \quad (3.55)$$

3. Если X – нормально распределенная случайная величина с параметрами m_x и σ^2 , то нормированная случайная величина T , представленная значениями t_i ($i = 1, 2, \dots, N$), также распределена нормально и имеет параметры 0 и 1.

3.2. Системы случайных величин и их характеристики

3.2.1. Общая характеристика законов распределения системы случайных величин

Экологические процессы определяются совокупным воздействием множества различных факторов (см. гл. 2). При их математическом анализе совокупность различных факторов часто рассматривается как система случайных величин.

Система m случайных величин $(X_1, X_2, \dots, X_m)$ геометрически интерпретируется как координаты случайной точки в m -мерном пространстве. Иногда в этих целях используют образ случайного вектора, составляющие которого по осям представляют собой случайные величины $X_1, X_2, \dots, X_m$.

Функция распределения m -мерного случайного вектора или, иначе, совместная функция распределения определяется как вероятность совместного выполнения m неравенств:

$$F(x_1, x_2, \dots, x_m) = P(X_1 < x_1, X_2 < x_2, \dots, X_m < x_m). \quad (3.56)$$

Свойства совместной функции распределения аналогичны свойствам функции распределения одной случайной величины (см. разд. 3.1.3). Кроме того, так как события $X_j < -\infty$ ($j = 1, 2, \dots, m$) невозможны, то при стремлении хотя бы одного аргумента к $-\infty$ совместная функция распределения стремится к нулю. Важно также отметить, что при стремлении какого-то j -го аргумента к ∞ , событие становится достоверным и совместная функция распределения уже не зависит от j -го аргумента. Отсюда для получения функции распределения одной или нескольких из имеющихся случайных величин нужно все остальные аргументы положить равными бесконечности.

Плотность распределения системы случайных величин или совместная плотность распределения – смешанная частная производная, если она существует, от совместной функции распределения, взятая 1 раз по каждому аргументу:

$$f(x_1, x_2, \dots, x_m) = \frac{\partial^m F(x_1, x_2, \dots, x_m)}{\partial x_1 \partial x_2 \dots \partial x_m}. \quad (3.57)$$

Свойства совместной плотности распределения во многом аналогичны свойствам плотности распределения одной случайной величины (см. разд. 3.1.3). Зная плотность распределения, можно определить совместную функцию распределения:

$$f(x_1, x_2, \dots, x_m) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_m} f(x_1, x_2, \dots, x_m) dx_1 dx_2 \dots dx_m. \quad (3.58)$$

Зная совместную плотность распределения, можно определить вероятность попадания случайной точки $(X_1, X_2, \dots, X_m)$ в пределы m -мерной области D по формуле:

$$p(D) = \int_{(D)} \dots \int f(x_1, x_2, \dots, x_m) dx_1 dx_2 \dots dx_m. \quad (3.59)$$

3.2.2. Зависимость между случайными величинами

Для исчерпывающей характеристики системы недостаточно знать распределение каждой из величин, входящих в эту систему. Нужно еще знать зависимость между этими величинами. В наиболее общем виде эта зависимость может быть охарактеризована с помощью так называемых условных законов распределения.

Условный закон распределения X_k – закон распределения, определенный при условии, что другие случайные величины X_j ($j \neq k$) приняли заданные значения.

Условный закон распределения можно задавать как функцией $F(x_k/x_j, j \neq k)$, так и плотностью распределения $f(x_k/x_j, j \neq k)$. Случайные величины, входящие в систему могут быть зависимыми и независимыми.

Независимые случайные величины – закон распределения каждой из них или любой их подсистемы не зависит от того, какие значения приняли остальные величины.

Функция распределения системы независимых случайных величин равна произведению их функций распределения, т. е.

$$f(x_1, x_2, \dots, x_m) = F_1(x_1)F_2(x_2)\dots F_m(x_m). \quad (3.60)$$

Аналогично плотность распределения системы независимых случайных величин определяется по формуле:

$$f(x_1, x_2, \dots, x_m) = f_1(x_1)f_2(x_2)\dots f_m(x_m). \quad (3.61)$$

Условия, выраженные формулами (3.60) и (3.61), являются необходимыми и достаточными условиями независимости случайных величин, входящих в систему.

При наличии линейной связи для характеристики связи двух случайных величин X и Y используется второй смешанный центральный момент, называемый корреляционным моментом или моментом связи случайных величин:

$$\mu_{11} = m[\overset{\circ}{X}, \overset{\circ}{Y}] = K_{xy}, \quad (3.62)$$

где $\overset{\circ}{X}$ и $\overset{\circ}{Y}$ – центрированные случайные величины [см. формулу (3.23)]. Для дискретных случайных величин:

$$K_{xy} = \sum_i \sum_j (\zeta_{xi} - m_x)(\zeta_{yj} - m_y) p_{ij}, \quad (3.63)$$

где $p_{ij} = p(X = x_i, Y = y_j)$ – вероятность совместного появления событий x_i и y_j .

Если X и Y независимые случайные величины, то $p_{ij} = p_i p_j$ и K_{xy} , как следует из формулы (3.63), равно нулю.

Для непрерывных случайных величин:

$$K_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x, y) dx dy. \quad (3.64)$$

Если X и Y независимы, то $f(x, y) = f(x)f(y)$ и, как видно из формулы (3.64), $K_{xy} = 0$.

Из формул (3.63) и (3.64) следует, что корреляционный момент характеризует не только линейную зависимость X и Y , но и их рассеивание относительно математического ожидания. Так, если одна из величин X или Y очень мало отклоняется от своего математического ожидания, то какой бы тесной ни была связь X и Y , корреляционный момент K_{xy} будет мал. Поэтому для более объективной характеристики связи используется безразмерная характеристика:

$$r_{xy} = \frac{K_{xy}}{(\sigma_x \sigma_y)}, \quad (3.65)$$

называемая коэффициентом корреляции.

Для независимых случайных величин $r_{xy} = 0$ (так как $K_{xy} = 0$). Случайные величины, для которых $r_{xy} = 0$, называются некоррелированными. Из независимости случайных величин следует их некоррелированность. Из некоррелированности случайных величин их независимость не следует. В случае, если $r_{xy} > 0$, говорят о положительной корреляции, если $r_{xy} < 0$ – об отрицательной. Положительная корреляция между случайными величинами означает, что при возрастании одной из них другая имеет тенденцию к возрастанию; отрицательная корреляция означает, что при возрастании одной из случайных величин другая имеет тенденцию к убыванию.

3.3. Математическая статистика в экологических исследованиях

3.3.1. Основные задачи математической статистики

Как отмечалось выше описание рассматриваемых процессов в геоэкологии, выводы и прогнозы об их возможных состояниях делаются на основе рядов наблюдений. В то же время ряды наблюдений, как правило, ограничены во времени и освещают лишь какой-то ограниченный период рассматриваемых процессов. Нам же необходимо на основе этих рядов, то есть, по сути, на основе имеющихся наблюдений, сделать выводы обо всем процессе.

***Математическая статистика** – наука, занимающаяся разработкой методов получения научно обоснованных выводов о массовых явлениях и процессах на основе данных наблюдений или экспериментов.*

Эти выводы относятся не только к имеющемуся ряду наблюдений или серии испытаний, а представляют утверждения об общих вероятностных характеристиках данного процесса, т. е. об его законе распределения и числовых характеристиках. Именно такое использование исходных данных является отличительной чертой математической статистики.

При переходе от данных наблюдений или серии опытов, как правило, ограниченных по объему, к характеристике всего процесса возникает ряд трудных задач. Наиболее типичными из них являются следующие.

1. Оценка неизвестной функции распределения рассматриваемого процесса на основе данных наблюдений.

Задача ставится так: за ограниченный период наблюдений в n лет получен ряд значений случайной величины X ($x_1, x_2, \dots, x_n$). Требуется по этому ряду наблюдений (выборке) определить неизвестную функцию распределения $F(x)$ случайной величины X .

2. Оценка неизвестных параметров распределения.

Известно, что случайная величина X имеет функцию распределения $F(x)$, зависящую от k неизвестных параметров. Требуется на основании опытных данных, т. е. по ряду ограниченной продолжительности, определить значения этих параметров.

3. Статистическая проверка гипотез.

– Предполагается, что функция распределения ряда X есть F_x (наша гипотеза). Спрашивается: совместимы ли наблюдаемые значения $X(x_1, x_2, \dots, x_n)$ с этим предположением.

– Предполагается, что значения параметров распределения генеральной совокупности имеют некоторые определенные значения (наша гипотеза). Не опровергают ли опытные данные эту гипотезу?

В основе применения математической статистики лежит ряд исходных понятий, без предварительного изучения которых невозможно использование современных методов обработки результатов наблюдений или опытных данных. Остановимся на некоторых из них.

В исследованиях окружающей среды анализ исходных данных производится таким образом, что в качестве репера, к которому привязываются результаты исследований, обычно выступает генеральная совокупность.

3.3.2. Генеральная совокупность и выборка

Генеральная совокупность – множество, включающее в себя все однородные объекты (например, все деревья в лесу, все данные об уровнях за период существования водного объекта и т.д.), которым присущи или не присущи определенные количественные или качественные признаки.

Наряду с генеральной совокупностью объектов можно рассматривать генеральную совокупность их признаков. В соответствии с числом объектов генеральные совокупности могут быть конечными и бесконечными.

Основное отличие генеральной совокупности заключается в том, что все числовые характеристики и закон распределения случайной величины X могут быть установлены по этой совокупности значений с точностью, равной точности определения исходных данных.

В исследованиях, связанных с анализом состояния окружающей среды, в качестве генеральной совокупности обычно принимается ряд наблюдений бесконечной продолжительности. Однако продолжительность рядов наблюдений за природными явлениями, как уже отмечалось, ограничена. Так, период непрерывных инструментальных наблюдений за стоком и другими экологическими процессами лишь в исключительных случаях достигает 100–150 лет. По всей территории СССР только около 80 пунктов имеют продолжительность наблюдений за стоком 60 лет и более. Все другие ряды гораздо короче. Ряды регулярных наблюдений за концентрациями загрязняющих веществ в реках в лучшем случае достигают 50–60 лет. Таким образом, если обозначить имеющуюся длительность экологических рядов через n , то во всех случаях $n \ll \infty$.

Выборочная совокупность (выборка) – ряд наблюдений ограниченной продолжительности, взятый из генеральной совокупности.

Таким образом, ряды наблюдений, которыми располагают геоэкологи, в подавляющем числе случаев, являются выборочными, так как освещают лишь какой-то короткий период развития рассматриваемых процессов. Причем характерной особенностью выборки в данном случае, является то, что их берут не произвольно, как обычно принято в статистике, а лишь за последний период развития данного процесса.

Располагая ограниченным по времени материалом наблюдений, мы хотим распространять полученные по ним выводы на весь процесс в целом, а не только на ту часть, которая попала в орбиту нашего внимания. Такая экстраполяция выводов может быть, оправдана только в том случае, если ограниченный материал наблюдений является представительным или репрезентативным, для всего процесса.

Представительная выборка – выборка, достаточно хорошо представляющая пропорции генеральной совокупности.

Все современные методы расчетов колебаний стока связаны с гипотезой, полагающей, что ряды наблюдений за стоком при отсутствии антропогенных изменений последнего являются

представительными выборками генеральной совокупности, освещающей весь процесс стока.

3.3.3. Статистические характеристики

Имеющиеся ряды наблюдений необходимо представить в наиболее наглядном виде и каким-то образом охарактеризовать. В статистике разработан ряд достаточно простых и наглядных характеристик. При использовании их необходимо иметь в виду, что, так как исходные ряды наблюдений являются выборочными, их характеристики также должны быть выборочными или статистическими. Естественно, что эти характеристики в той или иной мере могут отличаться от действительных характеристик, полученных по рядам бесконечной продолжительности (см. разд. 3.1).

Статистический ряд

Статистический ряд – первичная форма записи статистического материала x_i ($i = 1, 2, \dots, n$) в виде таблицы из двух-трех строк или колонок.

В первой строке или колонке таблицы указывают номер члена ряда, во второй – календарный срок или какую-то другую характеристику наблюдений (можно опустить), в третьей – в календарной последовательности располагают значения исходного ряда наблюдений. В табл. 3.2 в качестве примера представлен статистический ряд измеренных значений БПК₅ в мг/л по р. Великой (г. Псков).

Таблица 3.2. Статистический ряд измеренных значений БПК₅ в мг/л, р. Великая, г. Псков, верхний створ, за 1995 и 1996 г.

Номер п/п	Дата	БПК ₅	Номер п/п	Дата	БПК ₅	Номер п/п	Дата	БПК ₅
1	24.01.95	2,1	8	4.08.95	3,6	15	3.05.96	1,9
2	22.02.95	4,0	9	28.09.95	0,8	16	10.06.96	3,7
3	17.03.95	3,4	10	30.10.95	1,0	17	12.07.96	2,1
4	26.04.95	0,9	11	22.11.95	1,9	18	21.08.96	2,4
5	25.05.95	2,2	12	13.02.96	1,2	19	23.09.96	1,1
6	22.06.95	1,3	13	29.03.96	2,5	20	11.10.96	1,0
7	27.07.95	3,0	14	24.04.96	0,9	21	13.11.96	2,2

Для более наглядного представления материала по данным статистического ряда строится график последовательности значений

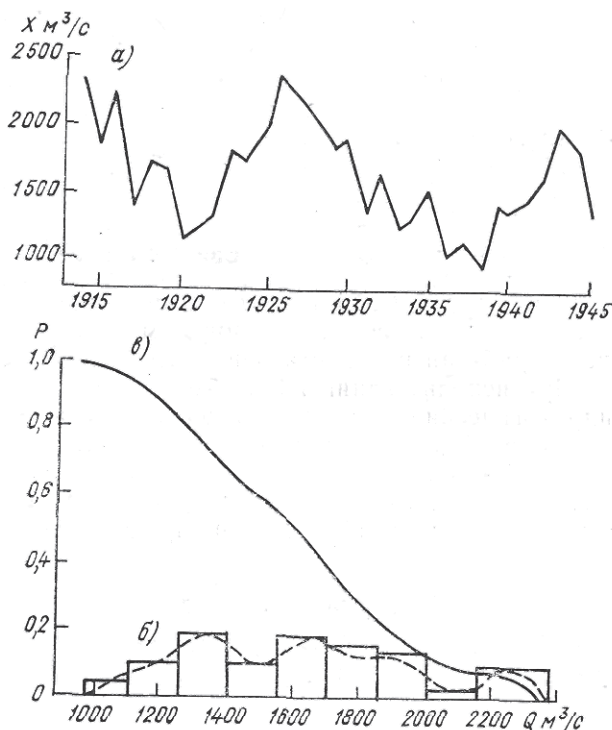


Рис. 3.6. Гидрограф (а), гистограмма (б) и эмпирическая кривая обеспеченности (в) годового стока р. Кама (г. Пермь)

рассматриваемой величины во времени или пространстве. Таким графиком является, например, гидрограф стока (рис. 3.6), показывающий изменение стока во времени.

По этому графику можно составить определенное представление о рассматриваемом процессе: об амплитуде его колебаний, тенденции изменений среднего значения и размаха колебаний, регулярности наблюдений и т.д.

Статистическая совокупность. Гистограмма (статистический многоугольник распределения)

Одной из основных задач статистической обработки материалов наблюдений является определение вероятности того или иного значения или диапазона значений случайной величины. Приближенное

представление об этой вероятности можно получить с помощью специальной таблицы. Для построения ее необходимо подсчитать количество наблюдений m_j , попадающих в тот или иной заранее заданный j -й интервал (группу) возможных значений данной случайной величины, и рассчитать частоту $p_j = m_j/n$ попаданий наблюдаемых величин в каждый j -й интервал, где n – период наблюдений.

Статистическая совокупность – совокупность групп, на которые разбивают значения исследуемой случайной величины, и частот попадания данных наблюдений в каждую группу (табл. 3.3).

Графическое изображение статистической совокупности называется **гистограммой** или **статистическим многоугольником**.

Таблица 3.3. Статистическая совокупность измеренных значений БПК₅ в мг/л, р. Великая, г. Псков, верхний створ, за период наблюдений с 1995 по 1998 г.

Группа	0,6–1,1	1,1–1,6	1,6–2,1	2,1–2,6	2,6–3,1	3,1–3,6	3,6–4,1
m	7	13	6	5	4	2	3
$\hat{p}$	0,18	0,33	0,15	0,13	0,10	0,05	0,08

Для построения гистограммы по оси абсцисс последовательно откладывают границы групп или интервалов исследуемой величины. На каждом из этих интервалов, как на основании, строят прямоугольник высотой p . Последовательность полученных прямоугольников, сумма высот которых равна 1, и представляет собой гистограмму (см. рис. 3.9.б).

Гистограмма позволяет сделать ряд достаточно определенных предположений об имеющемся ряде наблюдений: об амплитуде колебаний относительно средних значений, о наиболее вероятных значениях, симметричности колебаний относительно среднего значения, характере изменений частоты с изменением длин интервалов значений и т. д.

Согласно теореме Бернулли, если неограниченно увеличивать число опытов, то частота события $x_i \leq X \leq x_{i+1}$ $\hat{p}(X)$ сходится по вероятности к вероятности этого события $p(X)$, т. е.

$$\lim_{n \rightarrow \infty} p \left\{ \left| \hat{p}(X) - p(X) \right| < \varepsilon \right\} = 1. \quad (3.66)$$

Отсюда, соединяя точки вершин прямоугольников, соответствующие средним значениям интервалов, получим приближенный

график плотности вероятности тем более точный, чем ближе n к N и чем больше количество интервалов взято в расчет.

Следует отметить, что результаты анализа ряда наблюдений по гистограммам в некоторой степени зависят от числа групп или интервалов. Чрезмерное увеличение числа групп при ограниченном материале наблюдений приведет к истощению информации о каждой группе в отдельности и большим погрешностям в определении частоты ее значений. Уменьшение числа групп приведет к излишней схематизации анализа.

На основании имеющегося опыта обычно рекомендуется выбирать число групп таким, чтобы среднее число точек наблюдений в группе было не меньше семи-восьми. Иногда для определения числа групп или интервалов используется формула:

$$K = 5 \lg n, \quad (3.67)$$

где n – число членов ряда.

Эмпирическая (статистическая) функция распределения и функция обеспеченности

Эмпирической (статистической) функцией распределения $F(X)$ случайной величины X называется закон изменения частоты события $X < x$ в данном статистическом материале (выборке):

$$\hat{F}(X) = \hat{p}(X < x), \quad (3.68)$$

где $\hat{p} = \frac{m}{n}$ – частота события $X < x$; m – число событий $X < x$ в данной выборке.

В практике часто (см. разд. 3.1.3) используется обратное соотношение X и x . Преобразованная таким образом функция называется эмпирической функцией обеспеченности:

$$\hat{P}(x) = \hat{p}(X \geq x), \quad (3.69)$$

где $\hat{p}$ в данном случае частота событий $X \geq x$.

Графическое изображение эмпирической функции обеспеченности называется **эмпирической кривой обеспеченности** (рис. 3.6.б).

При $n \rightarrow \infty$ эмпирическая обеспеченность $\hat{P}$ сходится по вероятности к действительному значению обеспеченности P [см.

формулу (3.66)], где P – вероятность события $X < x$ или $X \geq x$ и, следовательно, $\hat{F}(x)$ сходится к $F(x)$ и $\hat{P}(x)$ к $P(x)$.

Из формулы (3.71) следует, что для расчета эмпирической функции обеспеченности необходимо для каждого данного значения случайной величины $\zeta_j (j = 1, 2, \dots, k)$ сосчитать число опытов m_j , в которых случайная величина X приняла значение большее или равное ζ_j , и разделить на общее число опытов n .

В табл. 3.4 представлен расчет эмпирической функции обеспеченности. В верхней строке таблицы указана нижняя граница этой группы, затем число случаев m , когда значения X больше или равны нижней границе интервала. В нижней строке по формуле (3.71) дана частота события $X \geq x$. По данным этой таблицы построен график эмпирической функции обеспеченности (см. рис. 3.6.в).

Таблица 3.4. Расчет эмпирической функции обеспеченности измеренных значений БПК₅ в мг/л, р. Великая, г. Псков, верхний створ, за период наблюдений с 1995 по 1998 г.

Нижняя граница группы	0,6	1,1	1,6	2,1	2,6	3,1	3,6	4,1
m	40	33	20	14	9	5	3	0
$\hat{p}(X \geq x)$	1	0,82	0,50	0,35	0,22	0,12	0,08	0

Однако построения на основе статистической совокупности дают удовлетворительные результаты только в том случае, когда имеющийся объем выборки достаточен для детального освещения частоты каждого интервала значений. Как отмечалось выше, для этого требуется, чтобы в каждом интервале было не менее семи-восьми точек наблюдений. Продолжительность рядов обычно недостаточна для таких построений. Кроме того, в большинстве случаев потребителей интересует обеспеченность экстремальных значений экологических характеристик, а при построении указанным способом эти данные получить невозможно. Поэтому в практике для расчета эмпирической кривой обеспеченности применяется другой прием, заключающийся в ранжировании исходного ряда наблюдений путем расположения его значений в убывающем порядке. При этом принимается, что каждое значение ряда представляет собой определенную группу, частота которой p_i равна $1/n$. Тогда эмпирическая обеспеченность может быть определена так же, как и в предыдущем случае, путем последовательного суммирования от наибольшего (первого) члена к интересующему нас m -ому значению ранжированного ряда (см. табл. 4.3), т. е.

$$\hat{P}(x) = \sum_{i=1}^m \hat{p}_i = \sum_{i=1}^m \frac{1}{n} = \frac{m}{n}. \quad (3.70)$$

В экологии широко распространилось мнение, что формула (3.70) верна только в том случае, если все возможные значения представлены в данной выборке. Иначе, при оценке по эмпирическим данным в краевых частях распределения возможны большие погрешности. Так, по формуле (3.70) эмпирическая обеспеченность минимального значения ряда $P_{m=n} = 1$. Отсюда вероятность значений, меньших минимального за период наблюдений, равна 0. Таким образом, из формулы (3.70) следует, что получить значение X , меньшее, чем наблюдаемое, невозможно¹. Однако опыт показывает, что какие бы значения, например стока, не отмечались в прошлом, в будущем возможны значения меньше наблюдаемых. В связи с этим для расчетов эмпирической обеспеченности был предложен ряд приближенных формул, более полно, по мнению авторов, учитывающих особенности экологической информации:

- формула Хазена:

$$\hat{P} = \left[\frac{(m-0,5)}{n} \right] 100\%; \quad (3.71)$$

- формула Крицкого–Менкеля:

$$\hat{P} = \frac{m}{(n+1)} 100\%; \quad (3.72)$$

- формула Чегодаева:

$$\hat{P} = \left[\frac{(m-0,3)}{(n+0,4)} \right] 100\%. \quad (3.73)$$

Формула Хазена основывается на предположении о том, что обеспеченность каждого интервала значений переменной величины можно отнести к середине этого интервала. Формула Крицкого–Менкеля основывается на представлении, что наиболее надежное значение эмпирической обеспеченности отвечает среднему значению эмпирических обеспеченностей заданной величины по множеству выборок. Если исходить из того, что рассматриваемая выборка

¹ Сам по себе этот вывод недостаточно обоснован, так как эмпирическая обеспеченность только по вероятности сходится к обеспеченности [см. формулу (3.68)].

занимает медианное положение среди других выборок, то получаем формулу Чегодаева.

Перечисленные формулы в средних частях кривых обеспеченностей дают практически одинаковые результаты, существенные расхождения отмечаются в верхней и нижней части этих кривых.

При этом формула (3.73) дает по сравнению с другими формулами в верхней части кривой обеспеченности при той же обеспеченности завышенные значения X . Формула (3.74) дает заниженные значения X при той же обеспеченности. Наконец медианная формула (3.75) дает значения X в верхней части близкие к значениям, вычисленным по формуле Хазена.

Дискуссия о применении различных формул эмпирической обеспеченности продолжалась довольно длительное время (см. работы С.Н. Крицкого и М.Ф. Менкеля [22, 23], Г.А. Алексеева [1, 2], В.Ф. Крюкова [24], Е.Г. Блохинова [6], А.В. Рождественского [35] и других авторов).

В этой связи можно отметить, что при оценке эмпирической обеспеченности возможны две ситуации.

1. Имеется только ряд значений X ($x_1, x_2, \dots, x_n$). Какая-либо другая информация отсутствует.

Тогда, по-видимому, расчеты эмпирической обеспеченности необходимо производить по формуле (3.72). Действительные значения $P(x)$ определяются с помощью доверительных интервалов на основе теорем Гливленко и Колмогорова (см. гл. 6). Какие-либо другие паллиативные решения могут только увеличить погрешность расчетов.

2. Имеется ряд значений X . Также известно, что, по крайней мере, одно из возможных значений X может быть меньше имеющихся.

В этом случае вариационный ряд представляется в виде:

$$x_1 \geq x_2 \geq x_3 \geq \dots \geq x_n \geq \overset{\bullet}{x_{n+1}} + 1,$$

где $x_1, x_2, \dots, x_n$ — имеющиеся значения ряда X ; x_{n+1} — член ряда X , относительно которого известно только что $x_{n+1} \leq x_n$, и для расчета эмпирической обеспеченности может быть использована формула (3.73).

Однако принимать формулу (3.73) в качестве основной было бы неверно, в частности, в некоторых случаях наблюдаемое значение x_n может быть равно 0. Тогда формула (3.74) теряет всякий смысл.

При использовании эмпирических обеспеченностей необходимо учитывать их выборочный характер, особенно при небольших n . Так, например, эмпирическая обеспеченность наиболее высокого максимума максимального стока р. Зеи у г. Зея за маловодный десятилетний период 1901–1910 гг., равный $5440 \text{ м}^3/\text{с}$, составляет по формуле (3.75) 6,7 %. За период 1901–1920 гг. обеспеченность этого максимума составляет 30,9 %, а за период 1901–1958 гг. – уже 41,0 %.

Таким образом, для анализа эмпирического материала используются три вида табличных и графических построений: статистический ряд, гистограмма, эмпирическая функция распределения. Эти построения дают достаточно наглядное представление о характере имеющегося материала и позволяют выявить некоторые важные особенности исследуемой величины.

Глава 4.

Теоретические законы распределения в геоэкологических исследованиях (экстраполяция и интерполяция рядов наблюдений)

4.1. Назначение теоретических законов распределения

При проектировании, строительстве и эксплуатации хозяйственных систем, а также при решении ряда научных задач необходимо произвести оценку различных характеристик состояния окружающей среды и ее возможных естественных и антропогенных изменений. При этом особенно важное значение имеет оценка экстремальных значений. Для потребителей было бы желательно, чтобы такая информация представляла собой абсолютно точное значение, например, величины и времени прохождения экстремальных значений стока на весь период работы каждой конкретной водохозяйственной установки. Однако, как отмечалось выше (см. гл. 2), такие однозначные прогнозы в принципе невозможны вследствие вероятностной природы экологических процессов. В то же время имеющиеся наблюдаемые максимальные и минимальные значения в этом смысле часто не удовлетворяют потребителя информации. Дело в том, что с увеличением периода наблюдений могут быть получены новые максимальные значения рассматриваемых характеристик, в той или иной степени, превышающие прежние, и новые минимальные значения, меньшие, чем ранее наблюдаемые. Таким образом, и максимальные и минимальные значения есть соответственно неубывающие или невозрастающие функции продолжительности периода наблюдений. Нам надо, зная прежние наблюдаемые, в том числе экстремальные, значения рассматриваемого процесса, определить возможные будущие экстремальные значения.

Допустим, в распоряжении исследователей имеется 40-летний ряд наблюдений. Оценка вероятности равенства или превышения

(обеспеченности) максимального значения этого ряда составляет 2,44 %. (см. формулу 3.72). Согласно этой оценке будущие максимальные значения могут быть равными или превышать наблюдаемые значения примерно раз в 40 лет. В то же время допустим, для проектирования плотины водохранилища надо знать расход воды обеспеченностью 0,01 %, т. е. повторяемостью раз в 10 000 лет. Таким образом, перед исследователями встает задача: имея ряд ограниченной продолжительности (выборку), определить максимальные значения ряда не ограниченной продолжительности (генеральной совокупности), то есть проэкстраполировать значения наблюдаемого ряда на более продолжительный период наблюдений.

Задачи, которые встают перед экологами, при оценке экстремальных значений характеристик окружающей среды, весьма разнообразны. Так, для обеспечения потребителей необходимым минимумом речной воды нужно знать, как часто расход воды будет равен или превысит заданный минимум. Если, например, минимальный сток равен или превышает необходимый минимум $Q_{\min}$ только в 80 % случаев (т. е. обеспеченность $Q_{\min}$ составляет 80 %), то, по-видимому, при проектировании следует предусмотреть другие источники водоснабжения, так как, примерно, 1 раз в пять лет речной воды будет недостаточно. Если обеспеченность $Q_{\min}$ составляет 99 %, то можно считать, что, за редким исключением, других источников водоснабжения в данном случае не потребуются.

Другой пример. Для пропусков больших расходов воды при строительстве и эксплуатации плотин необходимо рассчитать водосбросные отверстия. Теоретически можно предусмотреть такие большие размеры этих отверстий, что они будут всегда, в любой по водности год, достаточны для сброса воды. Но, так как максимально возможные расходы воды и гидрограф стока неизвестны, то в этом случае размер отверстий должен быть очень большим. Однако такое решение задачи практически невозможно из-за большой стоимости. Поэтому для ее решения используют другой подход, который заключается в том, что в качестве расчетного значения расхода при определении размера отверстий принимается расход Q_p обеспеченностью P 0,01; 0,1; 1,0; 3,0 % и т.д. Причем обеспеченность, т. е. вероятность равенства или превышения, назначается с учетом класса сооружений и тех последствий, которые могут быть при превышении фактического расхода воды над рассчитанным [19, 38].

Можно привести еще множество примеров, когда потребителю информации, в том числе и самим экологами, необходимо знать

режимные характеристики тех или иных объектов той или иной обеспеченности.

Однако определение таких характеристик имеет вид сложной, далеко не тривиальной задачи. Дело в том, что, как отмечалось ранее, имеющиеся ряды наблюдений обычно непродолжительны и определение, например, значений заданной обеспеченности непосредственно по эмпирическим данным не дает удовлетворительных результатов. Действительно, по любому ряду $X(x_1, x_2, \dots, x_n)$ могут быть рассчитаны только эмпирические обеспеченности (см. разд. 3.3). Как отмечалось выше, лишь при неограниченном увеличении числа членов в выборке, или в данном случае периода наблюдений, эмпирическая обеспеченность $\hat{p}$, согласно теореме Бернулли [11], сходится по вероятности к действительной обеспеченности (вероятности равенства или превышения) p .

Таким образом, если для наблюдаемых значений ряда $x_1, x_2, \dots, x_n$ можно с какой-то степенью точности указать их обеспеченность, то есть вероятность их превышения в будущий период, то определить обеспеченность возможных значений X , которые в имеющийся период не наблюдались, без дополнительной информации вообще невозможно. Так, продолжительность имеющихся рядов годового стока позволяет вычислить значения расхода воды, эмпирическая обеспеченность которых составляет, за редким исключением, не менее 1–10 %. Между тем для проектирования водохозяйственных сооружений необходимо знать, в зависимости от класса сооружения, расходы обеспеченностью 0,01, 0,1, 1, 99 % и др. Таким образом, возникает задача экстраполяции кривых обеспеченностей в верхней и нижней части.

В настоящее время эти задачи решаются на основе анализа законов распределения случайной величины.

Из множества известных законов распределения в гидрометеорологии наибольшее распространение получили: дискретные законы распределения: биномиальный и Пуассона; непрерывные законы распределения: нормальный, Пирсона III типа, Крицкого–Менкеля, логарифмически нормальный, Гумбела, Гудрича, Джонсона.

4.2. Дифференциальное уравнение кривых плотности вероятности Пирсона

В 1930 г. Д.Л. Соколовским [41] было предложено для расчета экстремальных значений стока использовать семейство кривых трехпараметрического гамма-распределения, значения которых изменяются от 0 до ∞ .

На рис. 4.1 представлен общий вид графика плотности распределения вероятностей случайной величины X , относительно которой известно, что она подчиняется одномодальному закону распределения и ее коэффициент асимметрии $C_s \geq 0$. По оси абсцисс здесь отложены значения X в модульных коэффициентах $k_i = x_i/m_x$. По оси ординат – плотность распределения $y = f(k)$. Расстояние по оси абсцисс от математического ожидания до моды – радиус асимметрии – d ; расстояние от моды до минимального значения – a . Отметим основные свойства представленного графика плотности распределения вероятностей.

1. Значения случайной величины X заключаются в определенных пределах. Плотность распределения, соответствующая значениям меньше наименьшего и больше наибольшего, равна нулю.

2. Так как в данном случае распределение одномодальное, то в промежутке между наибольшим и наименьшим значениями X плотность вероятности сначала возрастает, начиная от нуля до некоторого максимального значения, а затем снова убывает до нуля.

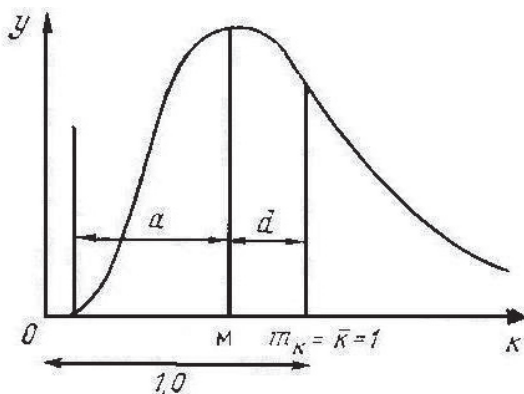


Рис. 4.1. График плотности распределения случайной величины ($x \geq 0$) в модульных коэффициентах ($k_i = x_i/m_x$)

3. Полная площадь, ограниченная кривой распределения и осью абсцисс (см. разд. 3.1.3), равна 1.

4. Математическое ожидание случайной величины K равно единице, то есть $m_k=1$.

Таким образом, график плотности распределения вероятностей $y = f(k)$ должен удовлетворять следующим условиям: во-первых, в начале и в конце графика по оси абсцисс $y = 0$; во-вторых, где-то между началом и концом графика значение y должно достигать максимума. Отсюда скорость изменения функции $y = f(k)$, или выражающая эту скорость первая производная dy/dx , должна равняться нулю в трех точках: в начале и в конце распределения, где $y = 0$, и в моде, где плотность распределения достигает максимума.

Для дальнейшего описания графика плотности вероятностей перенесем начало отсчета по оси k в центр распределения. Очевидно, это аналогично тому, что из каждого отсчета k_i , вычитается m_k , равное (см. выше) 1. Обозначим отсчеты по оси абсцисс относительно m_k через z , т. е. $z_i = k_i - 1$.

На основе вышеизложенных соображений и обозначений Пирсон представил дифференциальное уравнение кривых плотности вероятности централизованных случайных величин в модульных коэффициентах в следующем виде:

$$\frac{dy}{dz} = \frac{y(z+d)}{F(z)}, \quad (4.1)$$

где $F(z)$ – функция переменной величины z .

Это уравнение действительно соответствует изложенным выше условиям, так как на концах распределения $y = 0$ и, следовательно, $dy/dz = 0$, а при $z = -d$, т. е. в точке максимума, dy/dz также равно нулю.

Для расчета значений y по z необходимо определить $F(z)$. Для этого будем рассматривать такие функции $F(z)$, которые раскладываются в строку Маклорена:

$$F(z) = b_0 + b_1 z + b_2 z^2 + \dots + b_k z^k. \quad (4.2)$$

В данном случае вполне достаточно взять три первых члена разложения. Тогда

$$\frac{dy}{dz} = \frac{y(z+d)}{(b_0 + b_1 z + b_2 z^2)}. \quad (4.3)$$

Значения z и d в формуле (4.3) известны. Необходимо найти значения b_0, b_1, b_2 . Для этого приведем обе части уравнения (4.3) к общему знаменателю и умножим левую и правую части на z^n . Тогда

$$\int z^n (b_0 + b_1 z + b_2 z^2) dy = \int z^n y (z + d) dz. \quad (4.4)$$

Введем следующие обозначения:

$$u = z^n (b_0 + b_1 z + b_2 z^2), \quad v = y, \quad dv = dy$$

и проинтегрируем левую часть уравнения (4.4) по частям. Получаем:

$$\frac{Z^n (b_0 + b_1 z + b_2 z^2) y}{z_{\min}^{\max}} - \int y d \left[z^n (b_0 + b_1 z + b_2 z^2) \right]. \quad (4.5)$$

Так как на концах распределения $y = 0$ (см. выше), то первое слагаемое данного выражения равно нулю. Продифференцировав произведение в квадратных скобках, получим выражение для второго слагаемого:

$$-nb_0 \int y z^{n-1} dz - (n+1)b_1 \int y z^n dz - (n+2)b_2 \int y z^{n+1} dz. \quad (4.6)$$

Выразим полученные слагаемые через центральные моменты.

Как следует из формулы (3.24), центральные моменты различного порядка могут быть представлены в виде:

$$\mu_n = \int z^n y dz, \quad (4.7)$$

где z (см. выше) – центрированная случайная величина, y – плотность распределения.

Подставляя эти значения в формулу (4.6), получаем решение интеграла в левой части уравнения (4.4):

$$-nb_0 \mu_{n-1} - (n+1)b_1 \mu_n - (n+2)b_2 \mu_{n+1}. \quad (4.8)$$

Проинтегрируем теперь правую часть уравнения (4.4), подставив значения моментов по формуле (4.7):

$$\int y z^n (z + d) dz = \int y z^{n+1} dz + \int y z^n dz = \mu_{n+1} + d \mu_n. \quad (4.9)$$

Объединим преобразованные правую и левую части уравнения (4.4):

$$-nb_0 \mu_{n-1} - (n+1)b_1 \mu_n - (n+2)b_2 \mu_{n+1} = \mu_{n+1} + d \mu_n. \quad (4.10)$$

и рассмотрим полученное уравнение при значениях n , равных 0, 1, 2.

Так, из уравнения 4.10 получаем:

$$\text{при } n = 0 \quad -b_1\mu_0 - 2b_2\mu_1 = \mu_1 + d\mu_0,$$

$$\text{при } n = 1 \quad -b_0\mu_0 - 2b_1\mu_1 - 3b_2\mu_2 = \mu_2 + d\mu_1,$$

$$\text{при } n = 2 \quad -2b_0\mu_1 - 3b_1\mu_2 - 4b_2\mu_3 = \mu_3 + d\mu_2.$$

Учитывая, что $\mu_0 = 1$, $\mu_1 = 0$ (см. раздел 3.1.4), получаем для определения параметров b_0 , b_1 , b_2 систему из трех уравнений:

$$\begin{aligned} -b_1 &= d, \\ -b_0 - 3b_2\mu_2 &= \mu_2, \\ -3b_1\mu_2 - 4b_2\mu_3 &= \mu_3 + d\mu_2. \end{aligned} \tag{4.11}$$

Очевидно, что учет других значений b практически излишен, так как даже для расчета b_2 , как следует из полученной системы, уже требуется знать 4 момента: μ_0 , μ_1 , μ_2 , μ_3 . В то же время точность расчета уравнений в значительной степени понижается с увеличением порядка μ .

В зависимости от значений коэффициентов b_0 , b_1 , b_2 можно получить различные типы кривых обеспеченностей Пирсона. Из них очень широкое распространение получил III тип кривых Пирсона. Поэтому детальный анализ его с выводом основных расчетных формул дается ниже. Кроме того, из дифференциального уравнения кривых Пирсона можно вывести уравнение нормального закона распределения, имеющего в экологии самостоятельное значение.

4.3. Нормальный закон распределения

4.3.1. Уравнение кривой плотности вероятности нормального закона распределения

Если принять $F(z) = b_0$, то дифференциальное уравнение Пирсона (4.3) можно представить в виде:

$$\frac{dy}{dz} = \frac{y(z+d)}{b_0}.$$

Так как в этом случае принимается, что $b_1 = b_2 = \dots = 0$, то из системы (4.11) следует, что радиус асимметрии $d = 0$, а коэффициент $b_0 = -\mu_2$. Отсюда:

$$\frac{dy}{dz} = \frac{-yz}{\mu_2}. \quad (4.12)$$

Проинтегрировав это выражение, получим:

$$\ln y = \frac{-z^2}{(2\mu_2) + C}. \quad (4.13)$$

В точке $z = 0$ $\ln y = c$. Обозначим это значение y через y_0 . Очевидно, что y_0 будет максимальным значением плотности вероятности, так как при всех других z $\ln y < \ln y_0$. Учитывая это, потенцируем выражение (4.13):

$$Y = y_0 \exp\left(\frac{-z^2}{(2\mu_2)}\right). \quad (4.14)$$

Определим теперь максимальное значение y_0 , исходя из ранее установленного соотношения (3.17). Подставив в формулу (3.17) значение y по формуле (4.14), получаем:

$$y_0 \int_{-\infty}^{\infty} e^{\frac{-z^2}{2\mu_2}} dz = 1. \quad (4.15)$$

Обозначим:

$$t = \frac{z}{(2\mu_2)^{0,5}}$$

и, учитывая, что $\mu_2 = D_z = \sigma_z^2$,

$$t = \frac{z}{(\sigma_z \sqrt{2})} \text{ и } dt = \frac{\sigma_z}{\sqrt{2} dt}. \quad (4.16)$$

Подставив значения t и dz в формулу (4.15) после несложных преобразований, используя интеграл Эйлера-Пуассона, получим:

$$y_0 = \frac{1}{(\sigma_z \sqrt{2\pi})}. \quad (4.17)$$

Наконец, подставив значение y_0 в формулу (4.14), получим общий вид уравнения нормальной кривой распределения:

$$y = \frac{1}{(\sigma_z \sqrt{2\pi}) \exp \left[\frac{-z^2}{(2\sigma_z)} \right]}. \quad (4.18)$$

Так как z – значение исходного ряда X в отклонениях модульных коэффициентов от единицы (см. разд. 4.2), то $\sigma_z = C_v$ и

$$y = \frac{1}{(C_v \sqrt{2\pi}) \exp \left[\frac{-(k-1)^2}{(2C_v^2)} \right]}. \quad (4.19)$$

В соответствии с этим для исходного ряда X :

$$y = \frac{1}{(\sigma_x \sqrt{2\pi}) \exp \left[\frac{-(x-m_x)^2}{(2\sigma_x^2)} \right]}. \quad (4.20)$$

Как следует из полученного выражения, плотность распределения нормального закона, а следовательно и функция распределения y зависит от двух параметров: m_x , σ_x . Значение $c_s = 0$.

Если вместо значений X взять значения нормированной случайной величины (см. разд. 3.1.4), то, учитывая, что $\sigma_t = 1$ (см. разд. 3.1.6), получаем:

$$y = \frac{1}{\sqrt{2\pi} \exp \left(\frac{-t^2}{2} \right)}. \quad (4.21)$$

Здесь плотность вероятности y зависит уже только от одного аргумента t , эта зависимость может быть представлена в компактной табличной форме. По этой таблице производится расчет кривой плотности вероятности для любого исходного ряда X , подчиняющегося нормальному закону распределения. Переход от нормированных значений ординат к исходным выполняется в соответствии с выражением (3.53).

Впервые нормальный закон распределения был разработан для анализа погрешностей измерений. На этой основе он и получил наибольшее распространение во многих отраслях науки и техники, в том числе в экологии, где он широко используется для оценки точности расчетов, определения доверительных интервалов и т.д. Следует также отметить, что методы оценки параметров нормального

закона распределения хорошо разработаны и доступны. Для его применения достаточно знать два параметра: m_x и σ_x . Поэтому в практических исследованиях иногда допускается, что при описании некоторых рядов, имеющих асимметричное распределение, без особого ущерба может использоваться нормальное распределение. Однако погрешность вследствие ошибочно принятого допущения о нормальности распределения будет различной в каждом конкретном случае. Многие статистические методы, разработанные при этом допущении, остаются справедливыми в случае умеренных отклонений от него, например, дисперсионный анализ. В других методах, особенно при корреляционном анализе, это допущение может привести к серьезным погрешностям.

4.3.2. Функции распределения и обеспеченности

Подставляя в формулу функции распределения (3.15) значение y по формуле (4.21), получаем для нормированного ряда значений:

$$F(t) = \frac{1}{(2\sqrt{\pi}) \int_{-\infty}^t \exp\left(\frac{-t^2}{2}\right) dt}. \quad (4.22)$$

Это выражение показывает вероятность того, что случайно выбранное значение из нормальной совокупности нормированного ряда меньше значения t , рассчитанного по заданному x .

Расчет вероятности непревышения заданной величины, очевидно, сводится к решению интеграла в правой части формулы (4.22). Этот интеграл не выражается через элементарные функции. Обычно для его вычисления пользуются таблицами специальной функции интеграла вероятностей или функции Лапласа:

$$\Phi(Z) = \frac{2}{(\sqrt{\pi}) \int_{-\infty}^Z \exp\left(\frac{-z^2}{2}\right) dz}. \quad (4.23)$$

Функция $F(x)$ связана с $\Phi(Z)$ (см. [64]) следующим соотношением:

$$F(x) = \frac{[\Phi(z) + 1]}{2}, \quad (4.24)$$

где $z = \frac{t}{\sqrt{2}}$.

Эта формула может быть использована для непосредственного расчета вероятности непревышения [64] и функции распределения.

Изображение кривой обеспеченности в поле декартовых координат имеет довольно сложные очертания и не дает четкого представления об изменении расходов в области малой и высокой обеспеченности. Поэтому обычно для анализа законов распределения применяют так называемые клетчатки вероятности. Чаще всего используются клетчатки вероятностей для кривых обеспеченности с умеренной асимметричностью. Построение их производится аналитическим или графическим способом [46] таким образом, чтобы кривая обеспеченности нормального закона распределения была прямолинейной.

При использовании формулы расчета функции распределения (4.23) можно рассчитать вероятность попадания случайной величины, подчиняющейся нормальному закону распределения, на заданный участок значений X от α до β .

Действительно,

$$p(\alpha \leq x < \beta) = F(\beta) - F(\alpha). \quad (4.25)$$

Подставляя в это выражение значения $F(\beta)$ и $F(\alpha)$ по формуле (4.21), получаем:

$$p(\alpha \leq X < \beta) = \frac{\left[\Phi\left(\frac{t_\beta}{\sqrt{2}}\right) - \Phi\left(\frac{t_\alpha}{\sqrt{2}}\right) \right]}{2}. \quad (4.26)$$

Рассчитаем, например, по этой формуле вероятность попадания величины X в интервалы $\pm\sigma$, $\pm 2\sigma$, $\pm 3\sigma$:

$$\begin{aligned} p(-\sigma \leq x < \sigma) &= \frac{\left[\Phi\left(\frac{1}{\sqrt{2}}\right) - \Phi\left(\frac{-1}{\sqrt{2}}\right) \right]}{2} = 0,678; \\ p(-2\sigma \leq X < 2\sigma) &= \frac{\left[\Phi\left(\frac{2}{\sqrt{2}}\right) - \Phi\left(\frac{-2}{\sqrt{2}}\right) \right]}{2} = 0,956; \\ p(-3\sigma \leq x < 3\sigma) &= \frac{\left[\Phi\left(\frac{3}{\sqrt{2}}\right) - \Phi\left(\frac{-3}{\sqrt{2}}\right) \right]}{2} = 0,997. \end{aligned} \quad (4.27)$$

Таким образом, с вероятностью 67,8 % возможное значение X лежит в пределах $\pm \sigma$, с вероятностью 95,6 % в пределах $\pm 2\sigma$ и с вероятностью 99,7 % в пределах $\pm 3\sigma$.

4.4. Распределение Пирсона III типа

4.4.1. Уравнение кривой распределения

Вернемся к анализу дифференциального уравнения Пирсона (4.3) для кривых плотности распределения вероятности центрированных случайных величин, представленных в модульных коэффициентах. Для определения коэффициентов b_i в знаменателе правой части уравнения (4.3) была выведена система уравнений (4.11).

Пусть теперь случайная величина X является заведомо положительной, т. е. $x_{\min} \geq 0$, и положительно асимметричной. Заметим сразу, что положительная асимметричность распределения является характерным свойством многих экологических рядов наблюдений. Так, например, значения концентрации растворенных веществ, стока воды и осадков, как правило, больше или равны нулю, и, следовательно, имеют положительную асимметрию.

Как показано ранее (см. рис. 4.1),

$$a + d = m_k - k_{\min} = 1 - k_{\min}. \quad (4.28)$$

Так как по условию $k_{\min} \geq 0$, то естественно, что $a + d \leq 1$. Предположим, что $b_2 = b_3 = \dots = 0$, тогда из системы (4.11) следует:

$$\begin{aligned} -b_1 &= d, \\ -b_0 &= \mu_2, \\ -3b_1\mu_2 &= \mu_3 + d\mu_2. \end{aligned} \quad (4.29)$$

Исходя из этих соотношений можно получить выражение радиуса асимметрии d через второй и третий моменты распределения. Действительно, заменив в третьем уравнении системы (4.27) b_1 на $-d$, получим:

$$d = \frac{\mu_3}{(2\mu_2)}. \quad (4.30)$$

Теперь предположим, что в уравнении (4.3) $b_2 = b_3 = \dots = 0$, тогда, предварительно разделив переменные, получим:

$$\int \frac{dy}{y} = \int \frac{(z+d)}{(b_0 + b_1 z)} dz.$$

Преобразуем подынтегральное выражение в правой части полученного равенства, умножая числитель и знаменатель на b_1 и прибавляя и отнимая в числителе b_0 . Тогда:

$$\int \frac{dy}{y} = \int \left\{ \frac{(b_1 z + b_0 - b_0 + b_1 d)}{[b_1 (b_0 + b_1 z)]} \right\} dz = \int \frac{dz}{b_1 + \left(\frac{d-b_0}{b_1} \right) \int \frac{dz}{(b_0 + b_1 z)}}$$

и, решая интегралы в левой и правой частях, уравнения, получаем:

$$\ln y = \frac{z}{b_1 + \left(\frac{1}{b_1} \right) \left(\frac{d-b_0}{b_1} \right) \ln(b_0 + b_1 z) + c}. \quad (4.31)$$

Отсюда:

$$y = c \exp \left(\frac{z}{b_1} \right) (b_0 + b_1 z)^{\left(\frac{d-b_0}{b_1} \right)}. \quad (4.32)$$

Заметим теперь, что в левой крайней точке распределения, т. е. при минимальном значении z , $y = 0$. Тогда из уравнения (4.32) при $z = z_{\min}$ получим $b_0 + b_1 z_{\min} = 0$, т. е. при $b_1 \neq 0$ для асимметричного распределения $z_{\min} = -b_0/b_1$. Так как расстояние от $z_{\min}$ до центра распределения m_k равно $(a+d)$ (см. рис. 4.1), то можно записать:

$$a+d = \frac{b_0}{b_1}. \quad (4.33)$$

И подставляя значение b_0/b_1 в показатель степени уравнения (4.32) и определяя b_1 по системе (4.29), получаем:

$$\frac{\left(\frac{d-b_0}{b_1} \right)}{b_1} = \frac{a}{d}. \quad (4.34)$$

Перенесем начало отсчетов по оси абсцисс в точку, соответствующую моде распределения. Тогда отсчеты относительно моды: $z' = z + d$.

Преобразуем множитель уравнения (4.32), стоящий в скобках, заменив значения z на z' и подставив значение a по формуле (4.33):

$$b_0 + b_1 z = b_1 \left(\frac{b_0}{b_1 + z} \right) = -ad \left(\frac{1 + z'}{a} \right). \quad (4.35)$$

Подставим выражения (4.34) и (4.35) в исходное уравнение (4.32):

$$y = c \exp \left[\frac{(z' - d)}{d} \right] \left[ad \left(\frac{1 + z'}{a} \right) \right]^{\frac{a}{d}}. \quad (4.36)$$

Дальше для удобства будем обозначать z' через z , имея в виду, что z теперь – отсчет по оси абсцисс от моды. Преобразуем полученное выражение, сгруппировав отдельно все параметры, не зависящие от z , и обозначив их через C . Тогда

$$y = C \exp \left(\frac{-z}{d} \right) \left(\frac{1 + z}{a} \right)^{\frac{a}{d}}. \quad (4.37)$$

При $z = 0$, т. е. в моде распределения значение y (плотность вероятности) будет максимальным. Обозначим это значение через y_0 . Из уравнения (4.35) получаем для $z = 0$, $y_0 = c$. Тогда окончательный вид уравнения кривой Пирсона III типа при отсчете z относительно моды распределения:

$$y = y_0 \exp \left(\frac{-z}{d} \right) \left(\frac{1 + z}{a} \right)^{\frac{a}{d}}, \quad (4.38)$$

где y_0 – максимальная ордината распределения.

Выразим параметры a и d через C_v и C_s . При этом, согласно формулам (3.32), (3.33):

$$\mu_2 = \sigma^2 = C_v^2 m_k^2 = C v^2;$$

$$\mu_3 = C s C v^3.$$

Подставим эти выражения моментов в формулу (4.28). Тогда радиус асимметрии:

$$d = \frac{C_s C_v^3}{(2C_v^2)} = \frac{C_s C_v}{2}. \quad (4.39)$$

После подстановки в формулу (4.33) значений b_0 и b_1 из системы (4.27) получим:

$$a = \frac{\mu_2}{d} = \frac{2C_v}{\frac{C_s - C_s C_v}{2}}. \quad (4.40)$$

Из формул (4.33) и (4.40) следует, что расстояние от центра распределения до его левого конца:

$$a + d = \frac{2C_v}{C_s}. \quad (4.41)$$

В формуле (4.41) $a + d = 1 - k_{\min}$ (см. формулу 4.26). Отсюда получаем выражение, связывающее коэффициент асимметрии с величиной $k_{\min}$:

$$C_s = \frac{2C_v}{(1 - k_{\min})}, \quad (4.42)$$

Из формулы (4.42) следует, что $C_s \geq 2C_v$, а если $k_{\min} = 0$, то $C_s = 2C_v$.

Таким образом, значения C_s для кривой Пирсона III типа всегда больше или равны $2C_v$.

4.4.2. Запись уравнения кривой распределения через гамма-функцию

Вернемся к уравнению кривой Пирсона III типа относительно центра тяжести (4.30) и перенесем начало отсчета по оси абсцисс в точку $z_{\min}$ где плотность вероятности $y = 0$. Тогда расстояние по оси абсцисс в новой системе координат будет следующим:

$$z' = z + a + d, \quad (4.43)$$

где z – отсчет от центра распределения $m_k = 1$.

С учетом формул (4.34) и (4.43) получим для значения суммы в скобках формулы (4.35) следующие выражения:

$$b_0 + b_1 z = b_1 (z' - a - d + a + d) = -dz' \quad (4.44)$$

$$\frac{d-b_0}{b_1} = d - a - d = -a. \quad (4.45)$$

Подставляя эти значения в формулу (4.34), получаем:

$$y = c \exp \left[\frac{-(z' - a - d)}{d} \right] (dz)^{\frac{-a}{d}}. \quad (4.46)$$

Сгруппируем параметры, не зависящие от z , и обозначим их через сводный параметр c , а z' — через z , тогда:

$$y = c \exp \left(\frac{-z}{d} \right) z^{\frac{a}{d}}. \quad (4.47)$$

Выразим теперь значения d и a через C_s и C_v [см. формулы (4.39) и (4.40)]:

$$y = c \exp \left(\frac{-2z}{\frac{C_s}{C_v}} \right) z^{\frac{4}{C_s^2 - 1}} \quad (4.48)$$

и обозначим:

$$\beta = \frac{2}{\frac{C_s}{C_v}}, \quad \alpha = \frac{4}{C_s^2}. \quad (4.49)$$

Тогда:

$$y = c \exp(-\beta z) z^{\alpha-1}. \quad (4.50)$$

Можно показать, что

$$c = \frac{\beta^\alpha}{\Gamma(\alpha)}, \quad (4.51)$$

где $\Gamma(\alpha)$ — гамма-функция [16].

После подстановки значения c в уравнение (4.50) получаем уравнение биномиальной кривой распределения или кривой Пирсона III типа для условия $C_s = 2C_v$ в отсчетах z от нуля графика ($K_{\min} = 0$). С достаточной точностью, так как возможное $K_{\min}$ для многих рядов стока достаточно близко к нулю, это уравнение можно использовать и для случая, когда $C_s > 2C_v$:

$$y = \frac{\beta^\alpha}{\Gamma(\alpha) \exp(-\beta z) z^{\alpha-1}}. \quad (4.52)$$

Запись кривой Пирсона III типа через гамма-функцию нередко называют трехпараметрическим гамма-распределением. При использовании его следует учитывать, что X задается в модульных коэффициентах.

Если принять, что $C_s = 2C_v$, то из формул (4.37), (4.38) получаем:

$$\alpha = \beta = \frac{1}{C_v^2}, \quad (4.53)$$

и тогда:

$$y = \frac{\alpha^\alpha}{\Gamma(\alpha) \exp(-\alpha z) z^{\alpha-1}}. \quad (4.54)$$

Полученное для частного случая $C_s = 2C_v$ распределение называется гамма-распределением. В этом случае $z_i = k_i$.

4.4.3. Свойства кривой распределения Пирсона III типа

Кривая распределения Пирсона III типа предназначена для описания существенно положительных величин ($0 \leq x < \infty$). Так как многие ряды наблюдений в экологии, в частности ряды концентраций загрязняющих веществ, ряды речного стока и др., всегда положительны, то это распределение, впервые в СССР примененное в экологии Д.Л. Соколовским, получило весьма широкое распространение.

Коэффициент асимметрии распределения Пирсона III типа связан с коэффициентом вариации соотношением (4.42). Из этого соотношения при $k_{\min} = 0$ получаем $C_s = 2C_v$, при $k_{\min} > 0$ $C_s > 2C_v$ и во всех случаях $C_s \geq 2C_v$.

При $C_v < 1$ дифференциальная кривая распределения имеет куполообразную форму. При $C_s = 2C_v$, как следует из формулы (4.39), мода кривой находится в точке $k_M = 1 - C_v^2$.

При увеличении C_v мода смещается влево, и при $C_v = 1$ наибольшая плотность вероятности соответствует случаю $k_M = \alpha$ (рис. 4.2).

Как следует из перечисленных свойств, основным ограничением, которое накладывается на использование кривой распределения

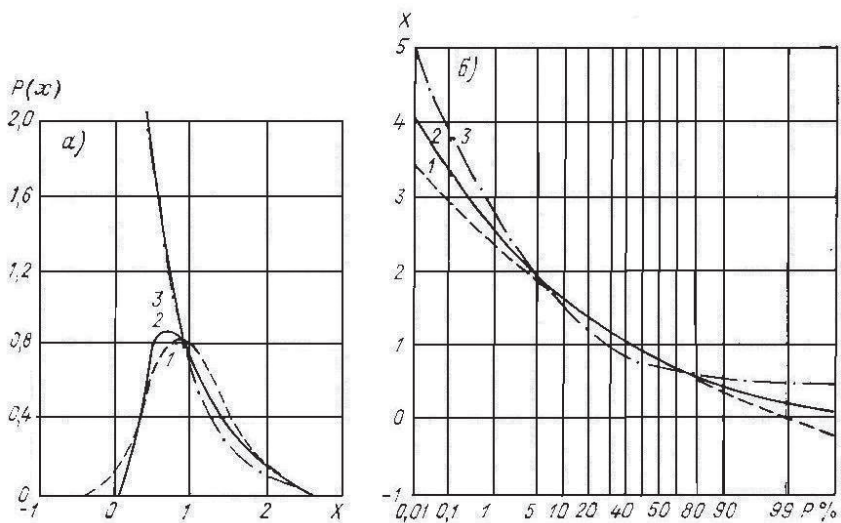


Рис. 4.2. Влияние параметров C_v и C_s на форму кривой распределения (а) и кривой обеспеченности (б) Пирсона III типа при $C_v = 0,50$: 1) $C_s = C_v$, 2) $2C_v$, 3) $C_s = 4C_v$

Пирсона III типа, является требование $C_s \geq 2C_v^1$. Если это требование не соблюдается, то при расчете могут быть получены отрицательные значения расхода воды, что противоречит природе явлений речного стока.

4.4.4. Интегральный закон распределения

В практических расчетах интегральные законы распределения обычно задают (см. разд. 3.1) в виде функции (3.5), называемой здесь функцией обеспеченности. Ее уравнение [см. формулы (3.13), (4.36)] в данном случае имеет вид:

$$P(z) = y_0 \int_z^{\infty} \exp\left(\frac{-z}{d}\right) \left(\frac{1+z}{a}\right)^{\frac{a}{d}} dz. \quad (4.55)$$

В соответствии с этой формулой составлена таблица ординат кривой обеспеченности значений X в нормированном виде $t_p = f(P, C_s)$ для значений C_s от 0 до 3,0.

В экологической литературе таблица зависимости $t_p = f(P, C_s)$ приводится под различными названиями: «Отклонения ординат

биномиальной кривой обеспеченности от середины при $C_v = 1$ », «Нормированные отклонения от среднего значения ординат распределения Пирсона III типа $\frac{(x_p - x)}{\sigma} = \frac{(k_p - 1)}{C_v} = \Phi(P, C_s)$ ¹ биномиальная кривая распределения)» и др. Нам представляется, что более удачным является название «Нормированные ординаты кривой обеспеченности Пирсона III типа $t_p = f(P \%, C_s)$ ». Переход от нормированных значений t_p к значениям x_p производится по формуле:

$$x_p = t_p \sigma_x + m_x = (t_p C_v + 1) m_x. \quad (4.56)$$

Для значений $C_s = 2C_v$ и $C_s = 4C_v$ имеются таблицы ординат кривой обеспеченности в модульных коэффициентах $k_p = f(P \%, C_v)$ (см., например, [35, 39, 47]). В некоторых случаях при машинном счете представляет интерес непосредственный расчет ординат кривой обеспеченности по формуле (4.55). Алгоритм и пример такого расчета даны в работе Г.А. Алексеева [1].

4.5. Трехпараметрическое гамма-распределение (распределение Крицкого–Менкеля)

4.5.1. Уравнение кривой распределения

Семейство кривых трехпараметрического гамма-распределения, предназначенных для расчета статистических рядов, значения которых изменяются от 0 до ∞ , было предложено в 1946 г. С.Н. Крицким и М.Ф. Менкелем [22]. Оно основано на модификации уравнения кривой Пирсона III типа при $C_s = 2C_v$. Подробное изложение принципов этого распределения приводится в монографиях [6, 22].

Уравнение кривой Пирсона III типа при $C_s = 2C_v$ является двухпараметрическим, так как зависит только от двух аргументов

¹ Это название, принятое в «Пособии по определению расчетных гидрологических характеристик», 1984 г., является неудачным, так как из названия следует, что в таблице представлен дифференциальный закон распределения, что не соответствует действительности.

m_x и C_v . Положительным качеством гамма-функции является то, что значение X изменяется, как и требуется по условию задачи, от 0 до ∞ , т. е. не имеет отрицательных значений, что в большей степени соответствует природе речного стока. Однако соотношение $C_s = 2C_v$ является только одним из частных случаев возможных распределений природных процессов. Поэтому для того, чтобы сделать уравнение (4.53) более общим, С.Н. Крицкий и М.Ф. Менкель прибегли к трансформации исходного признака распределения z . С этой целью в уравнение (4.52) вместо параметра z был введен параметр u , т. е.

$$y_0 = \frac{\alpha^\alpha}{\Gamma(\alpha) \exp(-\alpha u) u^{\alpha-1}}, \quad (4.57)$$

где u связано с z соотношением

$$u = \left(\frac{z}{a} \right)^{\frac{1}{b}}, \quad (4.58)$$

а параметры a и b являются функциями C_v и C_s и подлежат определению на основании данных наблюдений.

Пусть плотность кривой распределения u задана выражением (4.57). Найдем плотность вероятностей распределения z .

Как известно, плотность распределения монотонно возрастающей функции одного случайного аргумента записывается в виде:

$$f(z) = \frac{y du}{dz}. \quad (4.59)$$

Так как функция $z = f(u)$ является монотонно возрастающей, то уравнение кривой распределения z после подстановки значения du/dz по формуле (4.58):

$$\frac{du}{dz} = b^{-1} a^{\frac{-1}{b}} z^{\frac{1}{b-1}}, \quad (4.60)$$

может быть представлено в виде:

$$f(z) = \frac{\alpha^\alpha}{\Gamma(\alpha) \exp(-\alpha u) u^{\alpha-1} b^{-1} a^{\frac{-1}{b}} z^{\frac{1}{b-1}}}. \quad (4.61)$$

Если выразить значения u через z и учесть, что при $C_s = 2C_v$ $z_i = K_p$, то

$$f(z) = \frac{\alpha^\alpha}{\Gamma(\alpha) a^{\frac{-a}{b}} b^{-1} \exp \left[-\alpha \left(\frac{k}{a} \right)^{\frac{1}{b}} \right] k^{ub-1}}. \quad (4.62)$$

Координаты кривой обеспеченности трехпараметрического закона распределения задаются в виде ряда таблиц (см., например, [39, 48]), составленных Д.В. Коренистовым, Е.Г. Блохиновым и Н.В. Никольской для определенного соотношения $j = C_s/C_v$ ($j = 0,5; 1,0; 1,5; 2,0; 2,5; 3,0; 3,5; 4,0; 5,0; 6,0$). В каждой таблице приводятся для данного значения j координаты кривой обеспеченности в модульных коэффициентах: $k_p = f(P, C_v)$, для C_v , изменяющегося от 0,1 до 2,0 с интервалом $\Delta C_v = 0,1$.

4.5.2. Свойства распределения

Рассматриваемое распределение определяют три параметра: математическое ожидание m_x , коэффициент вариации C_v и коэффициент асимметрии C_s , причем C_s может принимать как положительные, так и отрицательные значения. При всех сочетаниях C_s ($C_s > 0$) и C_v изменение исследуемой величины ограничено интервалом $0 \leq x < \infty$. Вид кривых распределения зависит от соотношения между параметрами a и b . При $a > b$ для $b > 0$ и произвольном (положительном) значении a для $b < 0$ мода распределения находится в точке:

$$M = \left[\frac{\Gamma(a)}{\Gamma(a+b)} \right] (a-b)^b, \quad (4.63)$$

а малым значениям X соответствуют малые значения плотности вероятности. При $a \leq b$ и $b > 0$ мода распределения совпадает с началом кривой, т. е. $M = 0$.

При $C_s = 2C_v$ и $a = b = 1$ кривые обеспеченности Крицкого–Менкеля совпадают с аналогичными кривыми гамма-распределения.

Основное преимущество кривых обеспеченности Крицкого–Менкеля заключается в том, что при любых соотношениях между C_s и C_v ($C_s > 0$) нижний предел варьирующего признака равен нулю. Важным обстоятельством является также то, что, в отличие от кривых Пирсона III типа, в данном случае значение C_s не связано с началом кривых.

На рис. 4.3 представлены кривые обеспеченности Крицкого–Менкеля при различных значениях C_s и C_v .

Примеры расчета кривых обеспеченности Крицкого–Менкеля приведены в работах [9, 30].

4.6. Логарифмически нормальное распределение

Для вероятностного описания и расчета метеорологических и экологических величин, которым свойственна повышенная асимметрия (максимальный сток, осадки), нередко применяется логарифмически нормальное распределение. Подробное изложение теоретических основ и приемов практического использования этого закона дается в работах Г.А. Алексеева [2] и Е.Г. Блохинова [6].

Используется две разновидности логарифмически нормального распределения. Первая из них заключается в логарифмическом преобразовании всех членов ряда наблюдений X . При этом, если значения исходного ряда $x_1, x_2, x_3, \dots, x_n$ имели область изменения от 0 до ∞ ($0 \leq x < \infty$) (т. е. ряд заведомо асимметричен, так как изменения ниже среднего значения ограничены 0 ($x \geq 0$), а выше – неограничены), то преобразованные значения $u = \ln x$ будут уже находиться в пределах от $-\infty$ до ∞ ($-\infty < u < \infty$), т. е. область существования будет приближаться к нормальному закону распределения.

Такой подход весьма прост в техническом отношении и обладает важным принципиальным достоинством: он позволяет использовать весь арсенал достаточно хорошо разработанных методических средств нормального закона, но в то же время подобный подход неудобен тем, что при проведении анализа приходится пользоваться логарифмическим языком. В единицах логарифмов выражаются рассматриваемые величины и соответственно их параметры, в логарифмической шкале строится функция обеспеченности и производятся необходимые сопоставления и лишь в конце расчетов итоговые характеристики выражаются в натуральных единицах.

Для построения логарифмически нормальной кривой обеспеченности при логарифмическом преобразовании всех членов ряда X необходимо выполнить следующие преобразования:

- 1) прологарифмировать исходный ряд $u_i = \ln x_i$;

2) рассчитать числовые характеристики ряда логарифмов исходных величин:

- математическое ожидание:

$$m_u = \frac{\left(\sum_{i=1}^N u_i \right)}{N} = \frac{\left(\sum_{i=1}^N \ln x_i \right)}{N}; \quad (4.64)$$

- среднее квадратичное отклонение:

$$\sigma_u = \left[\left(\sum_{i=1}^N \frac{(u_i - m_u)^2}{N} \right) \right]^{0,5}; \quad (4.65)$$

3) рассчитать ординаты нормальной кривой обеспеченности по формуле (4.20), где

$$y = \frac{1}{\left(\sigma_u \sqrt{2\pi} \right) \exp \left[\frac{-(u - m_u)^2}{(2\sigma_u^2)} \right]}. \quad (4.66)$$

Обычно для этого используется таблица значений нормальной кривой обеспеченности $t_p = f(P \%)$ ([64в], прилож. 2, $C_s = 0$). По значениям t_p рассчитываются значения u_p заданной обеспеченности:

$$u_p = t_p \sigma_u + m_u; \quad (4.67)$$

и затем определяются действительные значения X :

$$x_p = \exp(u_p). \quad (4.68)$$

В ряде случаев нижний предел значений рассматриваемого ряда X существенно больше нуля. Тогда нижняя граница полученного в результате логарифмирования ряда U будет положительной. Для учета этого обстоятельства предлагается следующее преобразование:

$$u = \ln(x - x_0), \quad (4.69)$$

где x_0 – нижний предел случайной величины X .

Если, как часто бывает в практике, x_0 неизвестно, то при его определении исходят из положения о том, что при соответствующем подборе x_0 распределение U должно стать симметричным, т. е. значения C_s и μ_3 должны быть равны нулю:

$$\mu_3 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[\ln(x_i - x_0) - \left(\frac{1}{N} \right) \sum_{i=1}^N \ln(x_i - x_0) \right]^3 = 0. \quad (4.70)$$

Решение уравнения (4.70) обычно производится подбором. Для этого последовательно назначают несколько значений x_0 и для каждого рассчитывают значение μ_3 . Затем строят график зависимости $\mu_3 = f(x_0)$ и по графику определяют то значение x_0 , для которого μ_3 равно или близко к нулю.

Возможны такие ситуации, когда полученное в результате расчетов, x_0 не имеет физического смысла. В этом случае следует или отказаться от использования данного закона распределения или же при достаточном физическом обосновании назначить приближенное значение $x_0 \leq x_{\min}$, где $x_{\min}$ – наблюдаемое минимальное значение X .

Другая разновидность логарифмической модификации исходного ряда X заключается в преобразовании плотности вероятности нормального распределения.

Пусть случайная величина X подчиняется нормальному закону распределения [см. формулу (4.20)]. Введем замену $u = \ln x$. Тогда плотность вероятностей случайной величины X примет следующий вид:

$$y(x) = \left(\sigma_u x \sqrt{2\pi} \right)^{-1} \exp \left[\frac{-(\ln x - m_u)^2}{(2\sigma_u^2)} \right], \quad (4.71)$$

где σ_u и m_u уже не имеют того смысла как в исходном уравнении нормального распределения, а связаны с параметрами ряда X следующими соотношениями:

$$m_x = \exp \left(\frac{m_u + \sigma_u^2}{2} \right); \quad (4.72)$$

$$Cv_x = \left[\exp(\sigma_u^2 - 1) \right]^{\frac{1}{2}}. \quad (4.73)$$

Полученное распределение имеет свойства, существенно отличающиеся от исходного нормального распределения. Так, изменение случайной величины X ограничено положительной областью. Коэффициент асимметрии связан с коэффициентом изменчивости C_v соотношением:

$$Cs = 3Cv + Cv^3. \quad (4.74)$$

Таким образом, логарифмически нормальное распределение более асимметрично, чем распределение Пирсона III типа при $C_s = 2C_v$.

Вид дифференциальных и интегральных кривых логарифмически нормального закона распределения представлен на рис. 4.3.

Очевидно, что кривая распределения одномодальна, причем мода распределения находится в точке:

$$k_m = \frac{1}{(1 + Cv^2)^{\frac{3}{2}}}. \quad (4.75)$$

При увеличении C_v мода приближается к начальной точке $x = 0$ и C_s возрастает. Логарифмически нормальным распределением

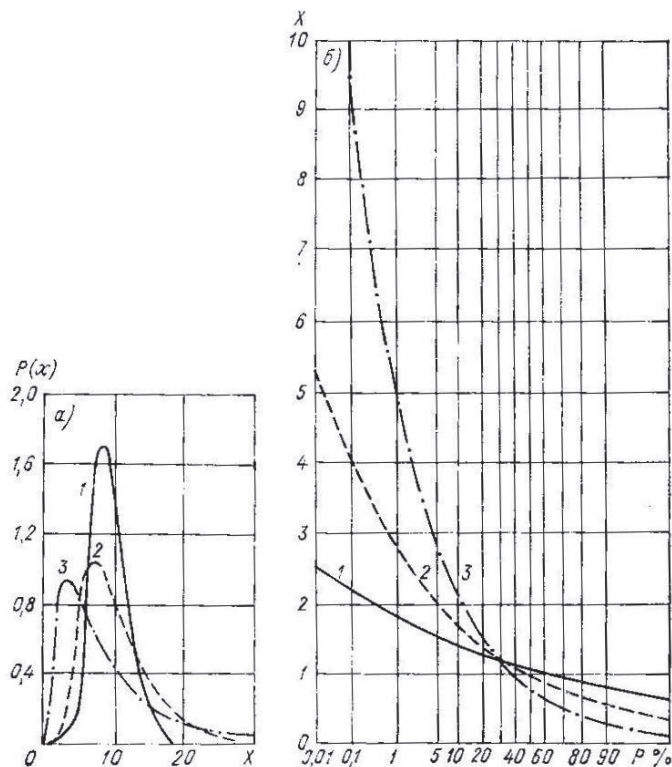


Рис. 4.3. Влияние C_v на форму кривой распределения (а) и кривой обеспеченности (б) логарифмически нормального закона распределения:
1) $C_v = 0,25$, 2) $C_v = 0,50$, 3) $C_v = 1,0$

хорошо аппроксимируются эмпирические распределения случайных величин, представляющих собой произведения большого числа независимых или слабо зависимых неотрицательных величин, из которых ни одна не является преобладающей.

4.7. Распределение Гумбела

В практике расчетов экстремальных значений гидрометеорологических характеристик различной вероятности превышения широкое распространение за рубежом (главным образом в США и Франции) получили кривые обеспеченности Гумбела, основанные на законах распределения крайних членов выборки [15, 43].

Пусть, например, имеется ряд наблюдений за стоком за n лет, представленный в виде последовательности ежедневных расходов. Тогда ежедневные расходы $x_{1i}, x_{2i}, \dots, x_{ni}$ в каждом i -м году можно принять за последовательность значений случайной величины X_i . Как и каждая случайная величина, величина X_i может быть описана функцией распределения:

$$F_i(x) = p(X_i < x). \quad (4.76)$$

Предполагается, что функция распределения стока в каждом году одинакова и описывается показательным законом:

$$F(x) = \begin{cases} 1 - \exp\left(\frac{-x}{v}\right), & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (4.77)$$

Можно показать, что в приведенной формуле v равно математическому ожиданию.

Рассмотрим теперь событие, состоящее в том, что какое-то значение X , равное u_n , не будет превышено в течение n лет, т. е. $U_n \geq \max(X_1, X_2, \dots, X_n)$. Тогда очевидно, что $U_n < x$ одновременно означает, что $X_1 < x, X_2 < x, \dots, X_n < x$, т. е. данное значение x больше, чем экстремальное значение за весь период наблюдений. Отсюда, по правилу умножения вероятностей независимых событий, вероятность непревышения $p(U_n < x)$ или иначе функция распределения может быть определена по формуле:

$$F_n(x) = p(U_n < x) = p(X_1 < x, X_2 < x, \dots, X_n < x) = F^n(x) \quad (4.78)$$

или с учетом формулы (4.77):

$$p(U_n < x) = \left[1 - \exp\left(\frac{-x}{v}\right) \right]^n. \quad (4.79)$$

При любом постоянном $x > 0$ и достаточно большом n ($n \rightarrow \infty$) $p(U_n < x) \rightarrow 0$, $p(U_n \geq x) \rightarrow 1$. Это означает, что при достаточно большом n , U_n , наверное, превзойдет любую наперед заданную границу. Исходя из этого, можно доказать, что среднее значение U_n должно расти с ростом n , будучи равным [11]:

$$mU_n = v_n \approx v \ln n. \quad (4.80)$$

Представим значение X в отклонениях от mU_n , т. е.

$$x = v_n + z \approx v \ln n + z, \quad (4.81)$$

где z – отклонение от m_{U_n} любого знака. Тогда

$$p(U_n < x) = p(U_n < v \ln n + z) = \left[1 - \exp\left(\frac{-\ln n - z}{v}\right) \right]^n \quad (4.82)$$

или, так как $\exp(-\ln n) = \frac{1}{n}$, то

$$p(u_n < x) = \left[\frac{1 - \exp\left(\frac{-z}{v}\right)}{n} \right]^n. \quad (4.83)$$

Как известно, при $n \rightarrow \infty$ $\left(1 - \frac{u}{n}\right)^n = \exp(-u)$. Отсюда, при том

же условии:

$$p(U_n < x) = \left[-\exp\left(\frac{-z}{v}\right) \right]. \quad (4.84)$$

Полученный закон распределения максимальных значений $X - U_n$ называется двойным показательным законом распределения. Общий вид этого закона:

$$F_n(x) = \exp[-\exp(-y)], \quad (4.85)$$

где $y = z/v$ представляет собой нормированное отклонение от моды.

По формуле (4.84) рассчитаны значения $F_n(x)$ для различных значений y :

$$y = \ln \ln [F_n(x)]. \quad (4.86)$$

В свою очередь по значениям $F_n(x)$ при различных значениях рассчитаны координаты функции обеспеченности $P_n(x) = 1 - F_n(x)$.

Таблица 4.1. Значения функции $P_n(x)$ при различных значениях y

y	P_n	y	P_n	y	P_n	y	P_n	y	P_n
9,10	0,01	4,00	1,80	1,50	20,0	0,09	60	-1,00	95
6,90	0,12	3,49	3,00	1,25	25,0	0,00	63	-1,10	97
6,00	0,24	3,00	4,90	1,03	30,0	-0,19	70	-1,53	99
5,30	0,50	2,97	5,00	1,00	31,0	-0,27	75	-1,93	99,9
5,00	0,67	2,25	10,0	0,67	40,0	-0,48	80	-2,00	99,94
4,60	1,00	2,00	13,0	0,37	50,0	-0,83	90		

Связь значений y и x определяется по уравнению прямой линии:

$$y = \alpha(x - q). \quad (4.87)$$

В этом уравнении:

$$\alpha = \frac{\sigma_{y_n}}{\sigma_x}, \quad (4.88)$$

где σ_{y_n} – среднее квадратическое отклонение значений y при данном объеме выборки n , σ_x – среднее квадратическое отклонение исходного ряда X .

Значение q при известном значении α может быть рассчитано по формуле:

$$q = \frac{\overline{x} - \overline{y_n}}{\alpha}, \quad (4.89)$$

где $\overline{y_n}$ и $\overline{x}$ – среднее значение ряда Y и X .

Для получения значений $\overline{y_n}$ и σ_{y_n} рассчитываются координаты эмпирической кривой обеспеченности по формуле (3.89), и по формуле (4.85) или по таблице $y = f(P)$ определяется последовательность значений $y_1, y_2, \dots, y_n$. По полученной последовательности рассчитываются среднее $\overline{y_n}$ и среднее квадратическое отклонение σ_{y_n} . Очевидно, что $\overline{y_n} = f(n)$ и $\sigma_{y_n} = f_n$. Имеются специальные таблицы этих зависимостей (см., например, работу [35, 45], табл. 12).

Для практического применения метода Гумбела используется процедура, приводящая к построению графика зависимости исследуемых максимальных величин (например, максимальных расходов) от нормированных отклонений от моды y . При этом совокупность точек, отвечающих на графике имеющимся наблюдениям, аппроксимируется в поле декартовых координат прямой линией. По этой линии и производится экстраполяция значений максимального стока до значений, отвечающих заданной вероятности. Уравнение теоретической прямой, согласно вышеизложенному, принимается в виде:

$$x = \frac{q + y}{\alpha}, \quad (4.90)$$

а оценка эмпирической обеспеченности производится по формуле (3.70).

В практике расчетов кривые обеспеченности Гумбела строятся в поле клетчатки вероятности для кривых с умеренной асимметричностью. При этом расчет производится по следующей схеме.

1) По соответствующим значениям y и P параллельно оси абсцисс клетчатки вероятности строится шкала значений y (рис. 4.4).

2) Члены исходного ряда X располагаются в убывающем порядке, и по формуле (3.89) рассчитываются их эмпирические обеспеченности.

Рассчитанные эмпирические точки (x_i, P_i) наносятся в поле клетчатки вероятности. Очевидно, что в отличие от эмпирической кривой обеспеченности в декартовых координатах (по оси абсцисс $y = f(P)$, по оси ординат – x), здесь, при $C_s \neq 0$, кривая обеспеченности будет иметь вид кривой линии.

3) По формулам (3.20) и (3.31) рассчитывается среднее арифметическое значение $\bar{X}$ и среднее квадратическое отклонение σ_x .

4) По таблице $\bar{y}_n, \sigma_y = f(n)$ [64] по числу членов исходного ряда определяется среднее значение $\bar{y}_n$ и среднее квадратичное отклонение σ_{y_n} .

5) По формулам (4.87) и (4.88) определяются коэффициенты уравнения α и q .

6) Значения α и q подставляются в уравнение (4.89), и по различным значениям y рассчитываются значения x .

7) Значения $x = f(y)$ наносятся на клетчатку вероятностей, причем в качестве абсциссы принимается ранее построенная ось

(см. рис. 4.4). Вид дифференциальных и интегральных кривых распределения Гумбела представлен на рис. 4.5. Очевидно, что кривая распределения одномодальна, мода распределения находится в точке $x = q$. Значение y в этой точке равно нулю, а плотность распределения, как следует из формулы (4.84), равна 0,37. При достаточно большом n ($n \rightarrow \infty$) значения $\bar{y}_n$ и σ_y стремятся к пределу соответственно 0,86 и 1,28.

Параметры более высокого порядка для рассматриваемого распределения постоянны: коэффициент асимметрии $C_s = 1,14$, эксцесс $E = 2,40$. Таким образом, распределение Гумбела является двухпараметрическим.

При использовании кривых обеспеченности Гумбела для расчетов экстремальных значений стока и других гидрометеорологических характеристик необходимо иметь в виду, что при выводе формулы функции распределения были приняты следующие допущения.

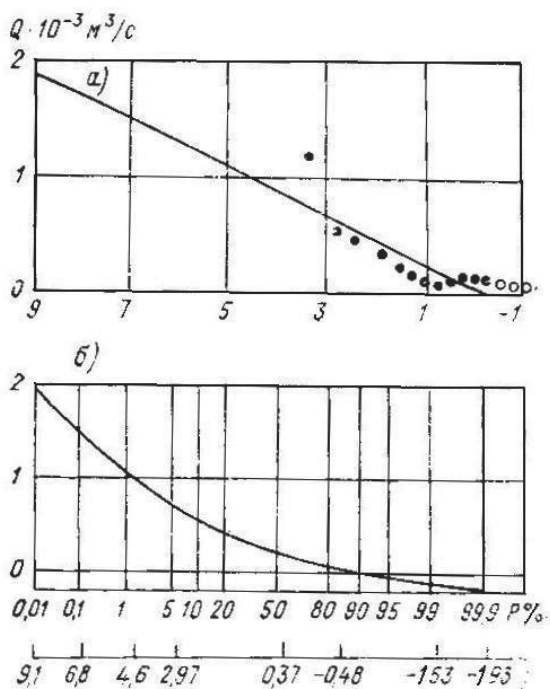


Рис. 4.4. Кривая обеспеченности Гумбела в поле декартовых координат (а) и клетчатки вероятности (б)

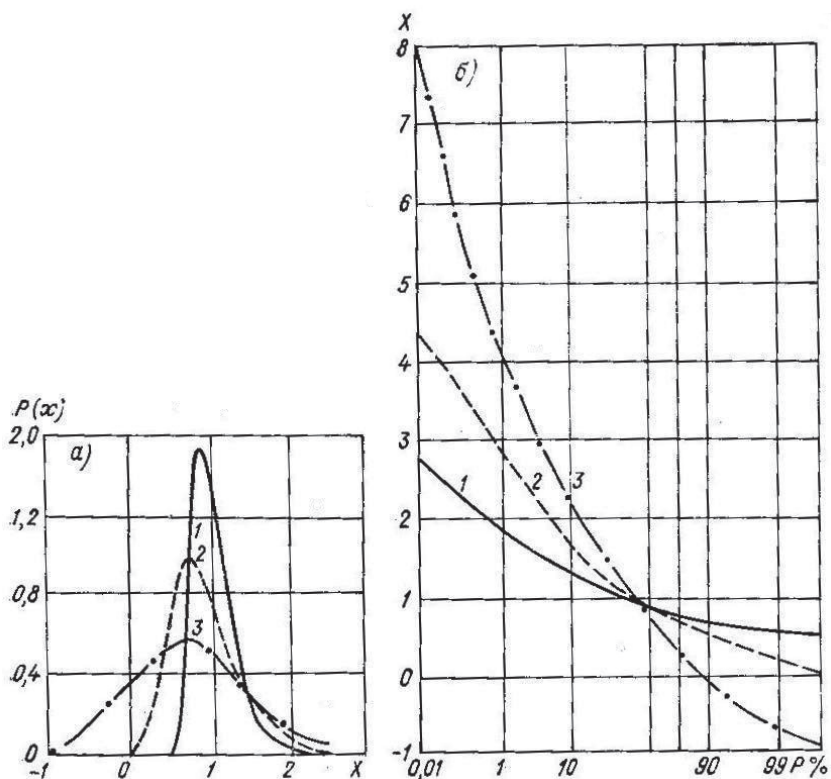


Рис. 4.5. Влияние C_v на форму кривой распределения (а) и кривой обеспеченности (б) Гумбела: 1) $C_v = 0,25$, 2) $C_v = 0,50$, 3) $C_v = 1,0$

1. Отдельные значения каждой случайной величины X_i ($i = 1, 2, \dots, n$) взаимно независимы. При расчетах стока это равносильно утверждению о том, что суточные расходы воды в каждом году взаимно независимы.

2. Функции распределения суточных значений стока в каждом году одинаковы и имеют экспоненциальный характер.

3. Число элементов выборки значений каждой случайной величины X_i для применения асимптотической теории достаточно велико.

Указанные допущения, принимаемые при применении функции распределения Гумбела для рядов значений гидрометеорологических процессов, вряд ли можно считать с физической точки зрения достаточно обоснованными. Например, при расчете максимального стока утверждение о независимости ежедневных расходов внутри

года явно не соответствует фактическим данным, подтверждающим наличие достаточно высокой корреляции. Одним из следствий наличия высокой внутрирядной корреляции между суточными расходами воды является уменьшение объема независимой информации, которая может быть гораздо меньше, чем число дней в году, а следовательно и недостаточна для утверждения об асимптотическом приближении. Приближенным является также утверждение об одинаковом распределении расходов в каждом году.

Наконец из теоретического анализа, выполненного Г.А. Алексеевым [4], следует, что предельное значение третьего центрального момента μ_3 для кривой распределения наибольшего члена, выборки

при $n \rightarrow \infty$ равно 2,40, отсюда $C_{sy} = \frac{\mu_3}{\sigma_y^3 - 1,14}$.

Более детально отмеченные свойства кривой Гумбела на ряде примеров рассмотрены в работе [43].

4.8. Распределение Джонсона

В трудах Г.Г. Сванидзе и Г.Л. Григолия рекомендуется использовать для анализа колебаний речного стока кривые распределения Джонсона [38].

Джонсон предложил находить эмпирическое распределение путем преобразования нормированной нормально распределенной случайной величины. Как уже отмечалось выше, при анализе логарифмически нормального закона распределения преимущество такого преобразования состоит в том, что оценки обеспеченности в этом случае можно получить на основе нормального закона распределения.

В общем случае преобразование Джонсона имеет вид:

$$\delta > 0, \quad -\infty < v < \infty,$$

$$z = v + \delta \tau(x, \xi, \lambda),$$

$$\lambda > 0, \quad -\infty < \xi < \infty,$$

где X – случайная величина, для которой подбирается кривая Джонсона; τ – некоторая произвольная функция; v, δ, ξ, λ – параметры распределения; z – нормированная случайная величина, подчиняющаяся нормальному закону распределения.

Исходя из вида функции $\tau(x, \xi, \lambda)$, Джонсоном предложены три различные формы или семейство функций:

$$\tau_1(x, \xi, \lambda) = \ln \left[\frac{(x - \xi)}{\lambda} \right], \quad x \geq \xi, \quad (4.91)$$

$$\tau_2(y, \xi, \lambda) = \ln \left[\frac{(x - \xi)}{(\lambda + \xi - x)} \right], \quad \xi \leq x < \xi + \lambda, \quad (4.92)$$

$$\tau_3(x, \xi, \lambda) = \ln \left[\frac{(x - \xi)}{\lambda} \right], \quad -\infty < x < \infty. \quad (4.93)$$

Распределение, основанное на преобразовании τ_1 , представляет собой логарифмически нормальное распределение с тремя параметрами, называемое семейством распределения S_L Джонсона; распределение, основанное на преобразовании τ_2 , представляется в виде семейства распределения S_B Джонсона с четырьмя параметрами распределения; а τ_3 – семейство распределений S_v Джонсона также с четырьмя параметрами распределения. Как следует из приведенных формул, распределения S_L , S_B , S_v применяются соответственно для случайных величин, ограниченных с одной стороны (I тип), случайных величин, ограниченных сверху и снизу (II тип) и, наконец, для неограниченных случайных величин.

При исследовании экологических процессов наибольший интерес представляет ограниченное с двух сторон распределение S_B (II тип).

Плотность распределения S_B Джонсона можно записать в следующем виде:

$$f(x) = \frac{(b - a)}{\left[(\sigma_\tau \sqrt{2\pi})(x - a)(b - x) \right] \exp \left\{ -\frac{1}{2\sigma_\tau^2} \left[\ln \left[\frac{(x - a)}{(b - x)} \right] - m_\tau \right]^2 \right\}}, \quad (4.94)$$

где $b = \lambda + \xi$ и $a = \xi$, – верхняя и нижняя граница случайной величины X ; m_τ и σ_τ – математическое ожидание и среднее квадратическое отклонение T .

Отдельные значения T с учетом введенных обозначений нижней и верхней границы [см. формулу (4.91)] рассчитываются по формуле:

$$\tau_i = \ln \left[\frac{(x_i - a)}{(b - x_i)} \right], \quad (4.95)$$

где i – порядковый номер члена ряда X и T ($i = 1, 2, \dots, N$) (индекс, обозначающий номер семейства функций, в дальнейшем опускается, так как рассматривается только семейство S_B).

Числовые характеристики T (m_τ и σ_τ) вычисляются по формулам:

$$m_\tau = \left(\frac{1}{N} \right) \sum_{i=1}^N \tau_i = \left(\frac{1}{N} \right) \sum_{i=1}^N \ln \left[\frac{(x_i - a)}{(b - x_i)} \right], \quad (4.96)$$

$$\sigma_\tau = \sqrt{\sum_{i=1}^N \frac{(\tau_i - m_\tau)^2}{(N-1)}} = \sqrt{\sum_{i=1}^N \frac{\left\{ \ln \left[\frac{(x_i - a)}{(b - x_i)} \right] - m_\tau \right\}^2}{N}}. \quad (4.97)$$

Расчеты ординат кривой Джонсона обычно проводятся по нормированным ординатам (3.35):

$$t_i = \frac{(\tau_i - m_\tau)}{\sigma_\tau} = \frac{\left\{ \ln \left[\frac{(x_i - a)}{(b - x_i)} \right] - m_\tau \right\}}{\sigma_\tau}. \quad (4.98)$$

t_i в этом случае, так же, как и τ_i , подчиняется нормальному закону распределения и, кроме того, $m_t = 0$, $\sigma_t = 1$.

При построении кривой обеспеченности Джонсона возможны следующие ситуации:

- параметры a и b известны,
- известен параметр a или b ,
- ни одно из крайних значений неизвестно.

В первом случае по ряду X рассчитывается последовательность значений τ [см. формулу (4.96)] и числовые характеристики этой последовательности m_τ и σ_τ [см. формулы (4.96) и (4.97)]. Затем из таблицы нормированных ординат нормального закона распределения для заданных значений обеспеченности берутся значения нормированных ординат t_p . По известным значениям t_p определяются [см. формулу (4.98)] значения:

$$\tau_p = t_p \sigma_\tau + m_\tau, \quad (4.99)$$

и на основе формулы (4.95) значения ряда X обеспеченностью $P\%$:

$$x_p = \frac{[b \exp(\tau_p + a)]}{[\exp(\tau_p + 1)]}. \quad (4.100)$$

Во втором случае, когда известно одно из крайних значений, в качестве которого чаще всего выступает параметр a – нижняя граница значения X , на первом этапе подбором определяется параметр b . В качестве критерия подбора (критерий качества) обычно принимается какой-либо количественный показатель согласия (совпадения) кривой обеспеченности Джонсона с эмпирическими точками. В качестве такого критерия обычно принимается критерий согласия Пирсона (χ^2). Нам представляется более целесообразным использование критерия $n\omega^2$. (Описание критериев согласия дано в гл. 6.)

Таким образом, процедура расчетов заключается в следующей последовательности действий: при известном значении a (или b) задаем какое-либо значение $b = b_1$. В качестве b_1 может выступать, например, максимальное наблюдаемое значение $X - x_{\max}$. Для значения a и b_1 по изложенной для первого случая схеме рассчитываются значения координат теоретической кривой обеспеченности и определяется критерий χ_1^2 или $n\omega_1^2$. Затем задаем новое значение $b = b_2$, $b_2 > b_1$, снова рассчитываем координаты теоретической кривой и находим χ_2^2 и $n\omega_2^2$ и т.д. По данным расчетов строится график зависимости $\chi_2^2 = f(b)$ или $n\omega_2^2 = f(b)$ или просто производится сравнение значений этих критериев. В качестве расчетного выбирается то значение b , при котором значения χ^2 или $n\omega^2$ являются наименьшими, т. е. отклонение эмпирических точек от теоретической кривой является наименьшим.

В качестве примера на рис. 4.6 представлен график зависимости $n\omega^2 = f(b)$ для минимального стока р. Иртыш, для которого можно принять $a = 0$.

В третьем случае, когда неизвестны оба предела колебаний X , производится подбор параметров a и b . В этом случае назначается какое-то малое предельно возможное a , или в качестве a принимается $x_{\min}$. Затем, так же как во втором случае, подбирается параметр b . Затем назначается новое значение a и снова выполняется подбор b , и т.д. На основе расчетов строится график $\chi^2 = f(a, b)$ или $n\omega^2 = f(a, b)$ и выбираются такие значения a и b , при которых названные критерии являются минимальными (рис. 4.6).

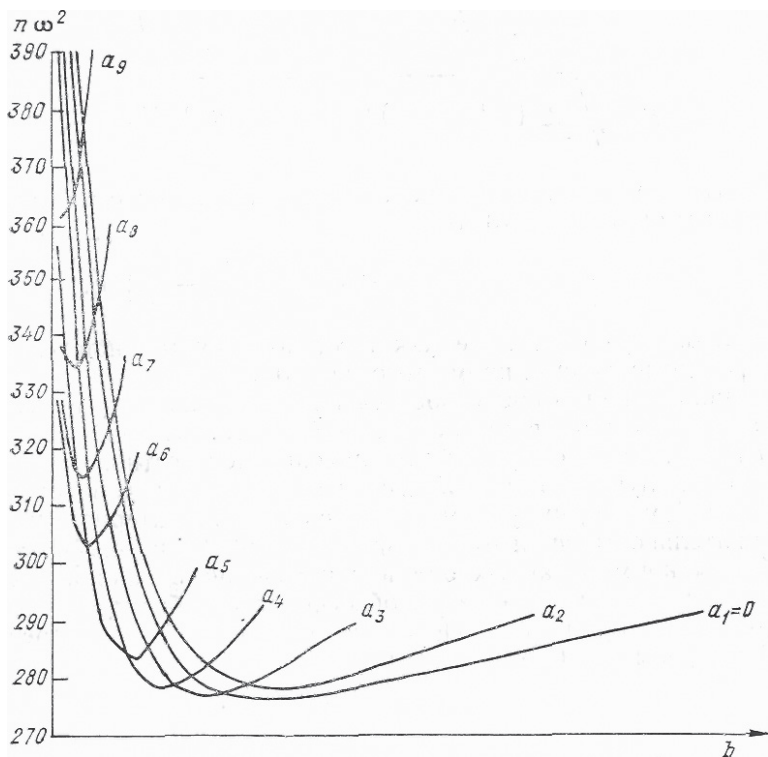


Рис. 4.6. Изменение критерия согласия $\pi\omega^2$ в зависимости от нижнего (а) и верхнего (б) предела распределения S_b Джонсона

4.9. Области и особенности применения различных законов распределения

В настоящее время, как известно, в строительном проектировании рекомендуется использовать кривые обеспеченности Крицкого–Менкеля и Пирсона III типа [31]. Другие законы распределения допускается использовать только при надлежащем обосновании. Такие рекомендации, записанные в нормативном документе, привели к тому, что в расчетах строительных организаций почти исключительно применяются кривые обеспеченности Крицкого–Менкеля, что естественно, так как при их использовании никаких обоснований, т. е. дополнительных работ, не нужно.

В научных исследованиях по-прежнему довольно часто используются кривые обеспеченности Пирсона III типа. Это вызвано тем, что уравнения закона распределения Пирсона III типа сравнительно легко поддаются различным преобразованиям и, таким образом, их проще использовать в исследованиях.

За рубежом в практических расчетах и научных исследованиях используются в основном кривые обеспеченности Пирсона III типа, логарифмически нормальные и кривые обеспеченности Гумбела. Кривые обеспеченности Крицкого–Менкеля там признания не получили.

Сопоставление результатов расчетов по перечисленным законам распределения проводилось неоднократно разными авторами.

При рассмотрении отдельных законов распределения отмечались их достоинства и недостатки. В частности, при анализе кривой обеспеченности Пирсона III типа отмечалось, что основным ее недостатком является требование $C_s \geq 2C_v$. При $C_s < 2C_v$ в некоторых случаях могут быть получены отрицательные значения стока большой обеспеченности, что, естественно, противоречит природе явлений речного стока. Однако до сих пор причины появления низких значений C_s ($C_s < 2C_v$) исследованы недостаточно, так остается неясным, не является ли $C_s < 2C_v$ следствием выборочности исходных данных или следствием неоднородности исходной информации (уменьшение максимальных расходов за счет неполного учета стока, естественной зарегулированности стока и т.д.).

Следует также обратить внимание на противоречие, возникающее при использовании соотношений $C_s < 2C_v$ и $0 \leq X < \infty$. Действительно, при $C_s = 0$ значения X могут изменяться от $-\infty$ до ∞ , при $C_s = 2C_v$ $0 \leq X < \infty$. Следовательно, при $0 < C_s < 2C_v$ нижний предел значений X в соответствии с рассмотренными выше законами распределения должен изменяться от 0 до $-\infty$, и с уменьшением C_s до нуля стремится к $-\infty$. Таким образом, какой бы трехпараметрический закон мы не использовали, очевидно, что с уменьшением C_s ($C_s \rightarrow 0$) нижний предел значений X должен становиться отрицательным. В кривой обеспеченности Крицкого–Менкеля такая переходная зона отсутствует, получается ничем не оправданный разрыв. Возможно, это противоречие носит формальный характер, однако разъяснение его, по-видимому, необходимо.

В заключение необходимо отметить, что все рассмотренные выше законы распределения, используемые в практике разных стран, основаны на предположении, что анализируемые в расчетах

ряды наблюдений являются чисто случайными, то есть внутрирядные связи в них отсутствуют. Однако это предположение наблюдается далеко не всегда. В этих случаях могут появляться отрицательные значения. Пока этот вопрос исследован не достаточно.

Наконец, нельзя не отметить, что в разных странах для оценки экстремальных значений используются разные законы распределения. Расхождения экстремальных значений, рассчитанные по разным законам распределения довольно значительны (см. например, [39]).

Глава 5. Оценка параметров законов распределения в геоэкологических исследованиях

5.1. Постановка задачи

Имеется случайная величина X , закон распределения которой содержит неизвестный параметр a . Требуется на основании опытных данных найти подходящую оценку этого параметра.

Чтобы понять суть этой задачи, представим себе, что в нашем распоряжении имеется ряд наблюдений бесконечной продолжительности $x_1, x_2, \dots, x_N$ ($N \rightarrow \infty$). Этот ряд наблюдений может рассматриваться как генеральная совокупность значений исследуемого процесса X . Разобьем генеральную совокупность на l частных выборок ($l \rightarrow \infty$) объемом в n ($n \ll N$) членов каждая. Пронумеруем значения X в соответствии с порядковым номером выборки j и порядковым номером значений X в данной выборке – i .

По материалам наблюдений, представленным в первой выборке: $x_{11}, x_{12}, \dots, x_{1n}$, рассчитано значение параметра a . Это значение называется статистической или выборочной (вычислено по выборке) оценкой параметра a . Обозначим эту оценку через $\hat{a}_1$. Естественно, что значение $\hat{a}_1$ является функцией значений входящих в первую выборку, т. е. $\hat{a}_1 = f(x_{11}, x_{12}, \dots, x_{1n})$. Аналогично $\hat{a}_j$ является функцией значений, входящих в J -ую выборку, т. е. $\hat{a}_j = f(x_{j1}, x_{j2}, \dots, x_{jn})$ и т.д.

Значения случайной величины, очевидно, могут быть разными в разных выборках, и появление именно их в данной выборке совершенно случайно. Отсюда следует, что последовательность значений $\hat{a}_j$, являющееся функцией входящих в выборку значений x_{ji} , также является значением случайной величины, имеющей, как и каждая случайная величина, свой закон распределения F_a . Этот закон распределения, разный для различных параметров распределения, при прочих равных условиях, зависит также от закона распределения случайной величины X и от объема выборки.

Таким образом, по каждой выборке одинаковой продолжительности из одной и той же генеральной совокупности может быть получена своя оценка $\hat{a}$ искомого действительного значения параметра a . Совокупность этих оценок может быть представлена в виде кривой распределения оценок a .

На рис. 5.1 представлены кривые распределения оценок a при различной продолжительности выборок.

Очевидно, что с увеличением объема выборок n кривая распределения $\hat{a}$ будет сужаться по оси абсцисс, т. е. рассеивание выборочных оценок относительно действительного значения будет в соответствии с законом больших чисел уменьшаться, и при $n \rightarrow \infty$, с вероятностью $p \rightarrow 1$ $\hat{a}$ должно быть равно a , т. е. кривая распределения должна превратиться в линию, проходящую через ось абсцисс в точке $\hat{a} = a$, равную по оси ординат 1.

Из приведенных графиков распределения оценок a при различных n видно, что при малых значениях n выборочные оценки a могут весьма существенно отличаться от действительных. Очевидно, что, чем меньше, при прочих равных условиях, отличие выборочных оценок от действительных и чем меньше вероятность больших отличий, тем лучше оценка.

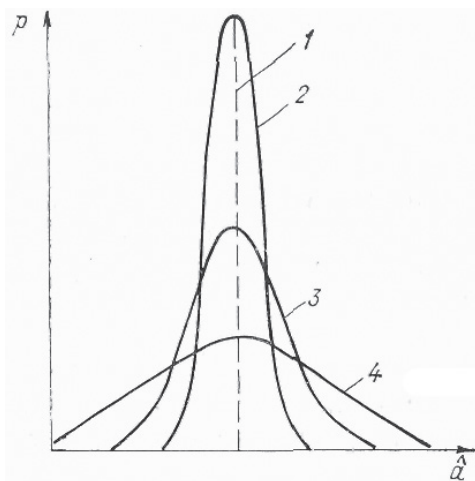


Рис. 5.1. Кривые распределения оценок статистического параметра:

1) $n = N$; 2) $n_2 < n_1$; 3) $n_3 < n_2$; 4) $n_4 < n_3$

Точно так же ряды данных наблюдений, используемые в науках о Земле, являются выборками из генеральной совокупности и рассчитанные по ним значения числовых характеристик (оценки), как правило, отличаются от действительных.

Принятые оценки параметров распределения, по-видимому, должны наилучшим образом отражать их действительные значения. При этом, исходя из приведенных выше рассуждений, каждая оценка параметра является функцией выборки, и ее следует считать наблюдаемым значением некоторой случайной величины. Следовательно, о качестве оценки параметра распределения с помощью данной статистики необходимо судить не по индивидуальным значениям этой статистики, а по ее выборочному распределению.

Исходя из этого, рассмотрим основные требования к выборочным оценкам.

Для того, чтобы какая-то статистическая оценка $\hat{a}(x_1, x_2, \dots, x_n)$ имела наибольшую практическую ценность, она должна удовлетворять следующим требованиям.

1. Быть несмещенной.

Оценка $\hat{a}$ называется **несмещенной**, если математическое ожидание $M[\hat{a}]$ равно оцениваемому параметру, т. е.

$$M[\hat{a}] = a. \quad (5.1)$$

Это означает, что центр распределения оценок числовой характеристики по множеству выборок (см. рис. 5.1) должен совпадать с действительным значением этой характеристики.

Естественно, что в качестве приближения неизвестного параметра лучше брать несмещенные оценки, чтобы не делать систематических погрешностей в сторону завышения или занижения.

Бывают случаи, когда возникает необходимость использовать смещенные оценки. Тогда, если $M[\hat{a}] > a$ оценка называется положительно смещенной, а если $M[\hat{a}] < a$ – отрицательно смещенной.

2. Оценка должна быть состоятельной.

Оценка $\hat{a}$ называется **состоятельной**, если она сходится по вероятности к оцениваемому параметру при неограниченном возрастании числа опытов n , т. е.

$$\lim_{n \rightarrow \infty} p \left[\left| \hat{a} - a \right| < \xi \right] = 1, \quad (5.2)$$

где ξ – сколь угодно малое положительное число.

Для выполнения этого требования достаточно, чтобы дисперсия оценки при увеличении n стремилась к нулю, т. е.

$$\lim_{n \rightarrow \infty} D \left[\hat{a} \right] = 0, \quad (5.3)$$

и, кроме того, чтобы оценка была несмещенной.

Требование состоятельности практически означает, что с увеличением объема выборок рассеивание оценок относительно математического ожидания (см. рис. 5.1), совпадающего с действительным значением числовой характеристики, должно уменьшаться и при достаточно большом n отклонение $\hat{a}$ от a с $p \rightarrow 1$ должно быть меньше любого наперед заданного значения.

Желательно, чтобы этому требованию удовлетворяла всякая оценка, пригодная для практического использования.

3. Оценка числовой характеристики распределения должна быть эффективной.

Оценки, обладающие свойством несмещенности и состоятельности, при ограниченном числе опытов могут иметь разные дисперсии. Очевидно, что, чем меньше дисперсия оценки, т. е. чем меньше разброс относительно действительного значения (см. рис. 5.1), тем меньше вероятность грубой погрешности при определении приближенного значения параметра.

Оценка, обладающая свойством

$$D \left[\hat{a} \right] = D_{\min}, \quad (5.4)$$

называется *эффективной*.

В практических приложениях теории вероятностей в для описания случайной величины X обычно применяются следующие числовые характеристики: математическое ожидание m_x , мода M , медиана Me , дисперсия D_x , среднее квадратическое отклонение σ_x , коэффициент вариации C_v , коэффициент асимметрии C_{sx} и реже эксцесс E_x (см. гл. 3).

Эти характеристики m_x , D_x , σ_x , C_v , C_{sx} разработаны на основе понятий начальных и центральных моментов распределения (см. гл. 3). Поэтому естественно, что одним из методов статистических оценок этих параметров является метод моментов. Однако этот

исторически более ранний, часто называемый классическим метод не полностью соответствует представленным выше требованиям к наилучшей оценке.

С середины 1950-х годов в гидрологии получил широкое распространение графоаналитический метод (метод квантилей), предложенный Г.А. Алексеевым [2].

В последнее время в практику расчетов все больше входит метод наибольшего правдоподобия, разработанный Фишером.

5.2. Метод моментов

Наибольшее распространение для вычисления оценок параметров распределения получил метод моментов. До недавнего времени другие методы оценок вообще не использовались в расчетах. Формулы расчета числовых характеристик методом моментов (3.36–3.40) представлены в гл. 3.

Удовлетворяют ли оценки на основе этих формул приведенным выше требованиям? Оказывается не всегда. Так, оценки по формулам (3.37–3.40) являются смещенными и в ряде случаев неэффективными. Рассмотрим этот вопрос подробнее.

5.2.1. Оценка математического ожидания

Проверка состоятельности. Согласно закону больших чисел (теорема Чебышева) [11], при неограниченном увеличении числа независимых испытаний среднее арифметическое наблюдаемых значений случайной величины, имеющей конечную дисперсию, сходится по вероятности к ее математическому ожиданию, т. е.

$$\lim_{n \rightarrow \infty} P \left[\left| \left(\frac{1}{n} \right) \sum_{i=1}^n x_i - m_x \right| < \xi \right] = 1, \quad (5.5)$$

где ξ – сколь угодно малое положительное число.

Нетрудно заметить, что вместе с выполнением равенства (5.5) выполняется и условие состоятельности (5.2).

Проверка несмещенности. Пусть имеется l периодов наблюдений ($l \rightarrow \infty$) продолжительностью n лет. Найдем математическое ожидание оценок математического ожидания методом моментов:

$$M[\hat{m}_x] = \left(\frac{1}{l}\right) \sum_{j=1}^l \hat{m}_{xj} = \left(\frac{1}{l}\right) \sum_{j=1}^l \left(\frac{1}{n}\right) \sum_{i=1}^n x_{ji} = \left(\frac{1}{n}\right) \sum_{i=1}^n \left(\frac{1}{l}\right) \sum_{j=1}^l x_{ji} = m_x.$$

Таким образом, $M[\hat{m}_x] = m_x$. Отсюда, согласно выражению (5.1), оценка математического ожидания методом моментов является несмещенной.

5.2.2. Оценка дисперсии

Проверка состоятельности. Раскроем квадрат в числителе формулы (3.37). Тогда

$$\hat{D}_x = \left(\frac{1}{n}\right) \sum_{i=1}^n x_i^2 - \left(\frac{2}{n}\right) \bar{x} \sum_{i=1}^n x_i + \left(\frac{1}{n}\right) \sum_{i=1}^n \bar{x}^2 = \left(\frac{1}{n}\right) \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Сумма $\left(\frac{1}{n}\right) \sum_{i=1}^n x_i^2$ представляет собой среднее арифметическое n наблюденных значений случайной величины X^2 , или, иначе, оценку математического ожидания X^2 методом моментов $\hat{m}_{x^2}$. Как показано выше, эта оценка сходится по вероятности к m_x^2 . В свою очередь $\hat{m}_x$ сходится по вероятности к m_x , и тогда вся правая часть полученного равенства сходится по вероятности к

$$M[X^2] - M^2[X] = D_x,$$

где $M[X^2]$ – начальный момент второго порядка [см. формулу (3.22)].

Следовательно, эмпирическая оценка дисперсии D_x методом моментов является состоятельной.

Проверка несмещенности. Преобразуем формулу оценки дисперсии (3.37), отняв и прибавив в скобках математическое ожидание:

$$\begin{aligned} \hat{D}_x &= \left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - m_x + m_x - \bar{x})^2 = \left(\frac{1}{n}\right) \sum_{i=1}^n [(x_i - m_x) - (\bar{x} - m_x)]^2 = \\ &= \left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - m_x)^2 - \left(\frac{2}{n}\right) \sum_{i=1}^n (x_i - m_x)(\bar{x} - m_x) + \left(\frac{1}{n}\right) \sum_{i=1}^n (\bar{x} - m_x)^2 \end{aligned}$$

и найдем математическое ожидание $\hat{D}_x$:

$$\begin{aligned}
M[\widehat{D}x] &= M\left[\left(\frac{1}{n}\right)\sum_{i=1}^n (x_i - m_x)^2\right] - \left(\frac{2}{n}\right)M\left[(x - m_x)\sum_{i=1}^n (x_i - m_x)\right] + \\
&+ \left(\frac{1}{n}\right)M\left[\sum_{i=1}^n (\bar{x} - m_x)^2\right] = \left(\frac{1}{n}\right)\sum_{i=1}^n M[(x_i - m_x)^2] - \\
&- \left(\frac{2}{n}\right)M[n(x - m_x)^2] + \left(\frac{1}{n}\right)\left[\sum_{i=1}^n [(\bar{x} - m_x)^2]\right].
\end{aligned}$$

В полученном выражении:

$$M(x_i - m_x)^2 = D_x, \quad M(\bar{x} - m_x)^2 = D_{\bar{x}}.$$

В то же время (см. гл. 3) $D_{\bar{x}} = \frac{D_x}{n}$. Отсюда:

$$M[\widehat{D}x] = D_x - D_{\bar{x}} = \frac{D_x - D_x}{n} = \left[\frac{(n-1)}{n}\right]D_x,$$

т. е.

$$D_x = \left[\frac{n}{(n-1)}\right]M[\widehat{D}_x]. \quad (5.6)$$

Таким образом, оценка дисперсии имеет отрицательную смещенность, равную в среднем D_x/n . Для ее учета следует правую часть формулы (3.37) умножить на $n/(n-1)$. В результате получаем:

$$\widehat{D}x = \left[\frac{1}{(n-1)}\right]\sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.7)$$

В соответствии с этим для оценки среднего квадратического отклонения должна использоваться формула:

$$\hat{\sigma}_x = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{(n-1)}}, \quad (5.8)$$

а для оценки коэффициента вариации:

$$\hat{C}_v = \sqrt{\sum_{i=1}^n \frac{(k_i - 1)^2}{(n-1)}}. \quad (5.9)$$

Для определения коэффициента асимметрии на основании аналогичных выводов получена формула:

$$\hat{C}_s = \frac{\left[n \sum_{i=1}^n (k_i - 1)^3 \right]}{\left[(n-1)(n-2) \hat{C}_v^3 \right]}. \quad (5.10)$$

Следует отметить, что при $n \geq 30$ во многих расчетах с достаточной точностью можно принимать в знаменателе формул (5.7)–(5.10) не $n - 1$, а n .

Практически на всех языках программирования имеются пакеты программ для оценки числовых характеристик методом моментов на ЭВМ.

5.2.3. Особенности применения метода моментов

1. Моментные оценки параметров m_x , D_x , σ_x , C_s , как это следует из формул (5.7–5.10), не зависят от закона распределения (в других методах оценки параметров могут зависеть от закона распределения). Это свойство моментных оценок в значительной степени упрощает их практическое использование и, во многом, именно поэтому эти оценки первыми вошли в практику статистических расчетов и получили такое большое распространение.

2. Эмпирическое математическое ожидание или среднее значение является несмещенной и состоятельной оценкой математического ожидания.

3. Рассмотренные оценки дисперсии, коэффициента вариации и асимметрии смещены. Поправочные множители, которые вводятся в выражения (5.7–5.10), ликвидируют погрешность в значениях выборочных моментов μ_2 и μ_3 , однако при этом соответствующее исправление оценок C_v и C_s достигается не всегда.

Как показано Е.Г. Блохиновым [6], смещенность (отрицательная) оценки коэффициента вариации C_v , определенная по формуле (5.9), невелика и становится практически ощутимой лишь при большой асимметрии.

Смещенность (отрицательная) оценки коэффициента асимметрии C_s по формуле (5.10) и соответственно отношения C_s/C_v гораздо больше и составляет десятки процентов. При большой асимметрии

($C_s > 3 - 4C_v$) и больших C_v смещенность так велика (при этом уже какие-либо поправки не могут быть введены принципиально), что метод моментов теряет право на применение.

4. Эффективность моментных оценок часто невысока.

Все это позволяет сделать вывод, что применение метода моментов в экологических расчетах должно быть ограничено, и в некоторых случаях он должен быть заменен методами, дающими оценки более высокой эффективности.

5.2.4. Оценка случайных погрешностей выборочных числовых характеристик

Погрешность средних значений в общем виде определяется по формуле:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \quad (5.11)$$

При нормальном законе распределения распределение $\sigma_{\bar{x}}$ также подчиняется нормальному закону распределения. При асимметричном законе распределения распределение $\sigma_{\bar{x}}$ имеет асимметричный характер. В этом случае для оценки погрешностей той или иной обеспеченности может быть использован закон распределения Стьюдента (см. гл. 6).

В общем случае средняя квадратическая погрешность оценки дисперсии может быть аналогично выражению (5.11) представлена в виде:

$$\sigma_D = \left(\frac{\sigma_{\delta^2}}{\sqrt{n}} \right), \quad (5.12)$$

где σ_{δ^2} – среднее квадратическое отклонение δ^2 :

$$\delta^2 = \left[\frac{1}{(n-1)} \right] \sum_{i=1}^n \left(\delta_i^2 - \bar{\delta}^2 \right)^2 = \mu_4 - \sigma_x^4, \quad (5.13)$$

$\bar{\delta}$ – центрированное значение случайной величины (см. формулу 3.25).

Для распределения Пирсона III типа в соответствии с формулами (5.12) и (5.13) получаем:

$$\sigma_D = \left(\frac{D_x}{\sqrt{n_\sigma}} \right) \sqrt{2 + \left(\frac{3}{2} \right) C_s^2}, \quad (5.14)$$

откуда при $C_s = 0$, т. е. для нормального распределения:

$$\sigma_D = \frac{D\sqrt{2}}{\sqrt{n_\sigma}}, \quad (5.13)$$

а при $C_s = 2C_v$, т. е. для гамма-распределения:

$$\sigma_D = \frac{D\sqrt{2 + 6C_v^2}}{\sqrt{n}}. \quad (5.14)$$

Средняя квадратическая погрешность среднего квадратического отклонения вычисляется по формуле:

$$\sigma_{\sigma_x} = \sigma_x \sqrt{1 + \sqrt{\frac{\left(2 + \left(\frac{3}{2} \right) C_s^2 \right)}{n-1}}}. \quad (5.15)$$

Средняя квадратическая погрешность определения коэффициента вариации в общем случае для распределения Пирсона III типа рассчитывается по формуле [1, 27]:

$$\sigma_{C_v} = C_v \sqrt{\frac{(1 + 2C_v^2)}{(2n) + \left(\frac{C_s}{n} \right) \left[\left(\frac{3}{8} \right) C_s - C_v \right]}}. \quad (5.16)$$

Отсюда при отсутствии внутрирядной связи и нормальном законе распределения:

$$\sigma_{C_v} = C_v \sqrt{\frac{(1 + 2C_v^2)}{(2n)}}, \quad (5.17)$$

и при гамма-распределении, т. е. при $C_s = 2C_v$:

$$\sigma_{C_v} = C_v \sqrt{\frac{(1 + C_v^2)}{(2n)}}. \quad (5.18)$$

Сопоставление погрешностей C_v распределения Крицкого–Менкеля и Пирсона III типа, показало, что для их оценки могут использоваться ранее представленные формулы.

5.3. Метод наибольшего правдоподобия

5.3.1. Общие положения

Метод разработан датским математиком Р. Фишером и впервые в отечественной экологии применен С.Н. Крицким и М.Ф. Менкелем [22] для определения параметров функций обеспеченности стока по исходным рядам с большим коэффициентом вариации. В дальнейшем большой вклад в развитие этого метода внес Е.Г. Блохин [6].

Метод основан на следующих положениях: пусть имеем выборку значений X объемом в n членов. Плотностью распределения вероятностей X является функция $f(x, a)$, зависящая, допустим, от одного постоянного параметра a . Если задан закон распределения и верно определен параметр a , то вероятность получить значения данного ряда X по параметру a является максимальной.

Функция правдоподобия – совместная плотность вероятности, или вероятность получить именно данную выборку при данном значении параметра a в результате n опытов:

$$\begin{aligned} L(x_1, x_2, \dots, x_n; a) &= \\ &= f(x_1, a) f(x_2, a) \dots f(x_n, a) = \prod_{i=1}^n f(x_i, a), \end{aligned} \quad (5.19)$$

где $\prod$ – знак произведения.

Если случайная величина X – дискретная с возможными значениями $\zeta_1, \zeta_2, \dots, \zeta_r$ а $m_1, m_2, \dots, m_r$ – соответственно число опытных значений X , равных $\zeta_1, \zeta_2, \dots, \zeta_r$, то функция правдоподобия определяется соотношением:

$$L(x_1, x_2, \dots, x_n; a) = p_1^{m_1}(a) p_2^{m_2}(a) \dots p_r^{m_r}(a) = \prod_{i=1}^n p_i^{m_i}(a), \quad (5.20)$$

где $p_i(a) = p(X = \zeta_i)$.

В этом случае функция правдоподобия L есть не что иное, как вероятность совместного появления значений $x_1, x_2, \dots, x_n$ при данном значении a .

Сущность метода максимального правдоподобия состоит в том, что в качестве оценки неизвестного параметра a выбирается такое значение, которое максимизирует функцию правдоподобия. Это

значение является функцией $x_1, x_2, \dots, x_n$ и называется **оценкой наибольшего правдоподобия**.

Отсюда, согласно известным правилам дифференциального исчисления, для нахождения оценки наибольшего правдоподобия необходимо решить уравнение

$$\frac{\partial L}{\partial a} = 0 \quad (5.21)$$

и отобрать то решение $a = a(x_1, x_2, \dots, x_n)$, которое обращает L в максимум.

Обычно в целях упрощения функцию правдоподобия заменяют ее логарифмом и рассматривают уравнение:

$$\left(\frac{1}{L}\right)\left(\frac{\partial L}{\partial a}\right) = \frac{\partial \ln L}{\partial a} = 0. \quad (5.22)$$

При двух неизвестных параметрах a_1 и a_2 их оценки определяются по двум совместно решаемым уравнениям:

$$\frac{\partial \ln L}{\partial a_1} = 0, \quad \frac{\partial \ln L}{\partial a_2} = 0. \quad (5.23)$$

При m параметрах a_j ($j = 1, 2, \dots, m$) их оценки получают из решений системы m уравнений:

$$\frac{\partial \ln L}{\partial a_j} = 0, \quad (j = 1, 2, \dots, m). \quad (5.24)$$

Иногда на практике не удается получить решение в аналитическом виде и тогда для нахождения максимума функции правдоподобия прибегают к численным расчетам.

5.3.2. Оценка параметров нормального закона распределения

Пусть имеем выборку $x_1, x_2, \dots, x_n$ значений ряда X , подчиняющегося нормальному закону распределения, т. е.

$$\frac{\partial \ln L}{\partial a_j} = 0, \quad (j = 1, 2, \dots, m). \quad (5.25)$$

Функция правдоподобия $L(x_1, x_2, \dots, x_n; \bar{x}, D_x)$, равная вероятности получить значения выборки $x_1, x_2, \dots, x_n$ при данных

значениях $\bar{x}$ и D_x в n испытаниях, согласно формуле (5.20), будет равна произведению вероятностей отдельных x_i ($i = 1, 2, \dots, n$), т. е.

$$L = \prod_{i=1}^n p_i. \quad (5.26)$$

Подставляя сюда значения p_i по формуле (4.18), получаем:

$$L = \left[\frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \right] \exp \left[\left(\frac{-1}{(2\sigma^2) \sum_{i=1}^n (x_i - \bar{x})^2} \right) \right]. \quad (5.27)$$

Отсюда:

$$\ln L = -\left(\frac{n}{2}\right) \ln 2\pi - \left(\frac{n}{2}\right) \ln \sigma^2 - \left(\frac{1}{2\sigma^2}\right) \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.28)$$

Теперь, в соответствии с формулой (5.25), найдем формулу оценки математического ожидания и дисперсии:

$$\frac{\partial \ln L}{\partial \bar{x}} = \left(\frac{1}{\sigma^2}\right) \sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Отсюда:

$$\bar{x} = \hat{m}_x = \frac{\sum_{i=1}^n x_i}{n}. \quad (5.29)$$

Определим теперь формулу для оценки дисперсии:

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{(2\sigma)^2 + \left(\frac{1}{(2\sigma)^4}\right) \sum_{i=1}^n (x_i - \bar{x})^2} = 0.$$

Отсюда:

$$\hat{\sigma}^2 = \hat{D} = \left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.30)$$

Таким образом, оценки параметров нормального закона распределения методом наибольшего правдоподобия совпадают с моментными оценками.

5.3.3. Оценка параметров распределения Пирсона III типа

Имеем выборку значений ряда X , подчиняющегося закону гамма-распределения. Вероятность i -го значения X оценивается по формуле (4.51). Подставляя эту формулу в формулу (5.20), получаем формулу функции правдоподобия:

$$L(x_1, x_2, \dots, x_n; a) = \left(\frac{\beta^{na}}{\Gamma^n(a)} \right) \exp \left(-\beta \sum_i z_i \right) \prod z_i^{\alpha-1}.$$

Откуда, с учетом того, что $\sum z_i = \sum k_i = n$ (см. гл. 4), следует:

$$L = \left(\frac{\beta^{na}}{\Gamma^n(a)} \right) \exp(-\beta n) \prod_i z_i^{\alpha-1}. \quad (5.31)$$

Прологарифмируем это выражение:

$$\ln L = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha - 1) \sum_i \ln z_i - \beta n. \quad (5.32)$$

Отсюда можно найти значения оценок параметров m_x и D_x для различных соотношений C_s и C_v . Рассмотрим, например, случай, когда $C_s = 2C_v$. В этом случае (см. гл. 4):

$$\ln L = n\alpha \ln \alpha - n \ln \Gamma(\alpha) - \alpha n + (\alpha - 1) \sum_i \ln z_i. \quad (5.33)$$

Найдем оценку коэффициента вариации:

$$\frac{\ln L}{\partial \alpha} = \frac{n(\ln \alpha + 1) - n \partial \ln \Gamma(\alpha)}{\partial \alpha - \sum_i \ln z_i - n} = 0. \quad (5.34)$$

Отсюда:

$$\frac{\partial \ln \Gamma(\alpha)}{\partial \alpha - \ln \alpha} = \left(\frac{1}{n} \right) \sum_{i=1}^n \ln k_i. \quad (5.35)$$

Введем обозначение:

$$\lambda = \left(\frac{1}{n} \right) \sum_i \ln k_i. \quad (5.36)$$

Тогда:

$$\frac{\partial \ln \Gamma(\alpha)}{\partial \alpha - \ln \alpha} = \lambda, \quad (5.37)$$

т. е. α , а следовательно, и C_v могут быть установлены по сумме логарифмов исходного ряда. Можно показать, что и при других соотношениях C_s и C_v $C_v = f(\lambda)$. Зависимость между λ и C_v обычно представляется в виде графика или таблицы $C_v = f(\lambda)$ (см., например, работу [39], прилож. 1) в десятичных логарифмах, позволяющих с помощью простых вычислений находить оценку C_v при практических расчетах.

В явном виде коэффициент изменчивости C_v выражается через λ следующим соотношением, указанным К.А. Семендяевым [11]:

$$\hat{C}_v = \sqrt{\frac{[-(5-2\lambda)\lambda]}{(3-2\lambda)}} \approx \sqrt{\frac{-6\lambda}{(3-2\lambda)}}, \quad (5.38)$$

где λ задают в натуральных логарифмах.

Оценка C_v и C_s методом приближенного наибольшего правдоподобия для распределения Крицкого–Менкеля производится по специальным номограммам, в качестве входа в которые используются значения λ_2 и λ_3 , где λ_2 определяется по формуле (5.54), а λ_3 – по формуле:

$$\lambda_3 = \left[\frac{1}{(n-1)} \right] \sum_{i=1}^n K_i \ln K_i. \quad (5.39)$$

Номограммы даны в работе [39], прилож. 1. Схема построения и соответствующие выводы представлены в монографии Е.Г. Блохинова [6].

5.3.4. Особенности применения метода наибольшего правдоподобия

По сравнению с другими методами оценки числовых характеристик метод наибольшего правдоподобия обладает рядом важных достоинств, главными из которых являются следующие:

- 1) если для параметра a существует эффективная оценка $\hat{a}$, то она получается как единственное в этом случае решение уравнения правдоподобия;
- 2) статистики, найденные методом правдоподобия, основаны на использовании всей информации о данных параметрах, содержащейся в выборке;
- 3) метод всегда дает состоятельные оценки.

К недостаткам метода наибольшего правдоподобия, в значительной мере сдерживающих его применение в повседневной практике, относятся следующие.

1) Формулы оценки статистических параметров методом наибольшего правдоподобия зависят от закона распределения исследуемой величины. Так как закон распределения во многих случаях устанавливается по самим статистическим характеристикам, то это вносит некоторый элемент неопределенности в расчетную схему метода.

2) Расчеты методом наибольшего правдоподобия нередко весьма трудоемки и требуют наличия специальной разработки расчетного метода для каждого нового закона распределения.

5.4. Метод квантилей (графоаналитический метод Алексеева)

Большое распространение для определения статистических параметров распределения с 1960 г. получил метод квантилей, предложенный Г.А. Алексеевым [2]. В то время вследствие своей простоты и меньшей трудоемкости, а также при почти полном отсутствии вычислительной техники этот метод позволял в три-четыре раза сократить время на расчет числовых характеристик по рядам продолжительностью в 60 лет.

Метод основан на определении параметров кривых распределения непосредственно по сглаженной эмпирической кривой обеспеченности. При этом выборочные параметры $\hat{m}_x$, $\hat{C}_v$, $\hat{C}_s$ находятся как соответствующие функции от выборочных квантилей x_p , отвечающих заданному уровню вероятности привышения.

Порядок расчета при применении метода квантилей Г.А. Алексеева к кривой обеспеченности Пирсона III типа сводится к следующему.

1. В поле клетчатки вероятности строится эмпирическая кривая обеспеченности ряда X . В поле эмпирических точек проводится плавная осредняющая линия.

2. По осредняющей линии определяются квантили обеспеченностью 5, 50, 95 %: $\hat{X}_5$, $\hat{X}_{95}$, $\hat{X}_{50}$, и вычисляется коэффициент скошенности:

$$S = \frac{(\widehat{X}_5 + \widehat{X}_{95} - 2\widehat{X}_{50})}{(\widehat{X}_5 - \widehat{X}_{95})}, \quad (5.40)$$

3. По таблице $C_s = f(S)$, разработанной для биномиального закона распределения, определяется коэффициент асимметрии (между S и C_s – функциональная связь, приведенная на рис. 5.3).

По формуле

$$\hat{\sigma} = \frac{(\widehat{X}_5 - \widehat{X}_{95})}{(\hat{t}_5 - \hat{t}_{95})}, \quad (5.41)$$

где t_5 и t_{95} – нормированные отклонения от среднего, определяемые по C_s и данным вероятностям 5 и 95 % (см., например, работу [47], прилож. 5) определяется среднее квадратическое отклонение.

По формуле

$$\bar{x} = \widehat{X}_{50} - \hat{\sigma}t_{50} \quad (5.42)$$

находится оценка математического ожидания.

Следует отметить, что метод очень прост в использовании, и во многих случаях на его основе можно получить результаты расчетов, не уступающие по точности другим методам.

Вместе с тем метод имеет и следующие серьезные недостатки.

1) Результаты расчетов этим методом во многом зависят от проведения сглаживающей эмпирической кривой обеспеченности. При коротком ряде распределения и значительном разбросе точек в поле графика клетчатки вероятностей разные авторы могут провести сглаживающую линию по-разному.

2) Большое влияние на результаты расчетов оказывают отдельные точки, расположенные в краевых частях графика.

3) При расчетах по коротким рядам наблюдений часто используются карты соотношений C_s/C_v . В данном случае это соотношение существенно меняется для различных створов.

В настоящее время этот метод используется редко и главным образом в научных исследованиях.

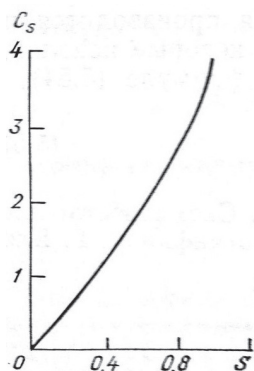


Рис. 5.3 Связь $C_s = f(S)$ для закона распределения Пирсона III типа

Глава 6. Статистическая проверка гипотез и оценка однородности геоэкологической информации

6.1. Постановка задачи

6.1.1. Общие соображения

По генеральной совокупности тем или иным способом можно установить закон распределения исследуемой случайной величины и ее числовые характеристики.

Совершенно иначе (см. гл. 5) обстоит дело, когда в распоряжении исследователя имеются ряды данных ограниченной продолжительности. Определенные по этим выборкам числовые характеристики закона распределения сами по себе являются случайными величинами. Как и любая другая случайная величина, эти выборочные, или как часто называют эмпирические, характеристики в свою очередь могут быть описаны тем или иным законом распределения.

Таким образом, по имеющимся рядам исходных данных получаются лишь приближенные представления о законе распределения. При этом, если бы в нашем распоряжении был новый ряд данных о том же процессе, не совпадающий с первым, допустим, по времени, то мы бы получили новые значения характеристик исследуемого процесса.

В условиях, когда объем выборок ограничен, и данные об исследуемом массовом явлении по отдельным выборкам обнаруживают значительное рассеяние, объективное суждение о процессе или преимуществах того или иного метода его описания можно вынести лишь на основе дополнительного статистического анализа.

Например, за разные периоды времени в одном и том же пункте наблюдений получены разные средние значения рассматриваемого процесса. Встает вопрос, является ли разность средних значений следствием, допустим, влияния изменений хозяйственной деятельности или эта разность есть следствие естественных колебаний выборочных оценок.

Отсутствие аппарата для объективного суждения по названому и многим другим вопросам приводило к разным выводам на основе того же самого исходного материала и в конечном итоге к дискуссиям, например, о влиянии антропогенных факторов на состояние окружающей среды.

В настоящее время в математической статистике разработаны приемы анализа, часто позволяющие прийти к определенным выводам в этих сложных случаях. Это достигается за счет более полного использования имеющейся информации, в частности использования гипотез того или иного вида, которые должны быть доказаны или опровергнуты в ходе статистической проверки.

Как известно, при любом научном исследовании обычно заранее имеется какое-то определенное, верное или неверное, представление об изучаемом процессе или явлении и его числовых характеристиках. Эти представления создаются на основании имеющихся публикаций, интуиции и опыта, теоретических соображений и др. Наличие таких представлений (гипотез) позволяет в значительной степени сузить направление исследования и от более широкой задачи перейти к частной проверке конкретной гипотезы.

Гипотеза – определенное представление или предположение относительно объективных свойств изучаемого явления.

В необходимых случаях можно перейти к проверке ряда гипотез и последовательным исключением их прийти, в конце концов, к однозначному решению.

Например, ниже приводятся некоторые практические ситуации в геоэкологических исследованиях, при которых возникает необходимость использовать гипотезы.

1. **Гипотеза о законе распределения.** По имеющемуся ряду значений случайной величины $X(x_1, x_2, \dots, x_n)$ необходимо определить функцию распределения $F(x)$.

Решение задачи в такой общей постановке вызывает значительные трудности и в большинстве случаев, не является необходимым. Обычно, исходя из некоторых дополнительных соображений, например, основанных на общей изученности рассматриваемых процессов, заранее делается предположение (гипотеза) о виде искомой функции распределения. Таким простейшим и в то же время наиболее весомым является предположение, что искомая функция распределения есть вполне определенная функция, допустим, $F_c(x)$.

Конкретные методы решения задач подобного типа представлены в разд. 6.1.

2. **Гипотеза однородности в пространстве.** В каком-то районе имеются данные наблюдений, допустим, за осадками по l станциям. Необходимо выяснить, подчиняются ли все ряды наблюдений одному и тому же закону распределения, то есть являются ли они однородными. Обычно вопрос в этом случае ставится так: существенны или несущественны расхождения между данными наблюдений различных станций.

Для выяснения этого вопроса по исходным рядам наблюдений строятся совмещенные эмпирические кривые обеспеченности. Эти кривые несколько отклоняются друг от друга. Могут ли названные отклонения служить доказательством неоднородности распределения дождевых осадков в рассматриваемом районе?

При решении этой задачи можно исходить из предположения, что осадки по всей территории района описываются одним и тем же конкретным законом распределения. Тогда решение поставленной задачи сводится к проверке этого или, в случае необходимости, такого же последующего предположения.

Методы решения задач подобного типа представлены в разд. 6.2.

3. **Гипотеза однородности во времени.** Данные наблюдений за речным стоком представлены в виде отдельных групп, полученных в разное время (перерывы в наблюдениях) или в разных створах (перенос створа). Требуется выяснить, различаются ли эти группы наблюдений между собой более, чем можно было бы ожидать в связи с чисто случайными колебаниями, имеющими место в выборочных совокупностях.

Обычно вопрос в этом случае ставится так: существенны или несущественны расхождения между данными наблюдений в различные периоды времени, или иначе: является ли ряд наблюдений однородным во времени.

Для ответа на этот вопрос выдвигается предположение о том, что ряд наблюдений однороден во времени. Теперь достаточно проверить правильность этого предположения.

Методы решения задач подобного типа представлены в разд. 6.5–6.7.

4. **Гипотеза независимости.** Имеются выборки значений случайной величины X и Y . Есть основания предполагать, что значения X и Y независимы, т. е. что $F(x, y) = F(x) F(y)$, где $F(x)$ и $F(y)$ – одномерные функции распределения.

Подробнее содержание этой гипотезы и методы ее проверки изложены в гл. 7.

5. **Гипотеза случайности.** По имеющейся выборке значений X рассчитаны коэффициенты корреляции внутрирядной связи $r(X_p, X_j)$.

Есть основания считать, что действительные значения коэффициентов корреляции $r(X_p, X_j)$ равны нулю, т. е. внутрирядные связи отсутствуют. Необходимо проверить эту гипотезу.

Более подробно содержание гипотезы и методы ее проверки изложены в этой главе.

Во всех приведенных примерах на первый план выдвигается какое-то предположение или гипотеза и производится ее проверка. От того, насколько верно сформулирована гипотеза, во многом зависят результаты анализа. Поэтому представляется важным детально рассмотреть понятие гипотезы в статистике и методы их использования в геоэкологических исследованиях.

6.1.2. Основные обозначения и определения

Нулевая гипотеза – гипотеза, имеющая в данном случае особенно важное значение, называется нулевой или основной.

Гипотеза обычно обозначается через H , нулевая гипотеза – через H_0 .

Если закон распределения задан отдельными параметрами и гипотеза строится именно по этим параметрам (вид закона распределения рассматриваемых случайных величин предполагается известным), то говорят о **параметрических гипотезах**. Например, предположение о значении какого-то параметра a закона Пирсона III типа является параметрической гипотезой. Символически ее можно записать так:

$$H_0 : a = c, \quad (6.1)$$

где c – предполагаемое действительное значение параметра a .

Статистические гипотезы, в основе которых лежит предположение о конкретном виде закона распределения, называются **непараметрическими**. Символически эти гипотезы записываются в следующем виде:

$$H_0 : F(x) = F_c(x). \quad (6.2)$$

Здесь $F(x)$ – искомый вид закона распределения, $F_c(x)$ – предполагаемый вид закона.

Непараметрические гипотезы являются более общими, чем параметрические. Более общими, но менее эффективными, являются и методы их проверки.

Проверка как параметрических, так и непараметрических гипотез чаще всего основывается на использовании хорошо изученных законов распределения некоторых статистик.

Например, нормированная разность средних значений рядов Y и X , соответственно $\bar{y}$ и $\bar{x}$, определяется по формуле:

$$\hat{t} = \frac{(\bar{y} - \bar{x})}{\sigma_{(\bar{y} - \bar{x})}}, \quad (6.3)$$

где $\sigma_{(\bar{y} - \bar{x})}$ – среднее квадратическое отклонение разности $\bar{y} - \bar{x}$. Эта статистика, при условии, что математические ожидания сопоставляемых рядов равны, подчиняется закону распределения Стьюдента–Госсета (см. разд. 6.3). Тогда, зная нормированную разность $\hat{t}$ по сопоставляемым рядам Y и X , можно по закону распределения Стьюдента–Госсета определить вероятность этого значения в предположении, что гипотеза верна. Если эта вероятность окажется очень мала, то, по-видимому, гипотеза о равенстве действительных средних опровергается.

Таким образом, в данном случае применение нулевой гипотезы $H_0 : m_Y = m_X$ позволяет воспользоваться известным распределением выборочных оценок статистики $\hat{t}$ и по этому распределению оценить вероятность $P(m_Y = m_X)$.

Заметим, что в данном случае закон распределения выборочных значений известной статистики $\hat{t}$ является дополнительной информацией, которую получают за счет использования нулевой гипотезы.

Правило, согласно которому проверяемая гипотеза H_0 принимается или опровергается, называется **статистическим критерием** или просто **критерием проверки гипотез**.

В заключение необходимо еще раз подчеркнуть причину широкого использования гипотез и в быту и в науке. Основная причина – гипотеза, верная, а часто и неверная, дает дополнительную информацию. В приведенном выше примере дополнительной информацией является распределение выборочных значений статистики $\hat{t}$.

6.1.3. Уровень значимости

При проверке статистических гипотез часто встает вопрос о том, мала или нет вероятность рассматриваемого события в предположении, что гипотеза верна. Заслуживающие доверия ответы на этот вопрос даются при помощи критериев значимости, основанных на законах распределения различных статистик.

Критерии значимости применяются для определения, будут ли некоторые статистики, характеризующие расхождение значений случайных величин, отличаться друг от друга или от других статистик более, чем можно было бы ожидать в связи со случайными колебаниями выборочных оценок. Если сравниваемые величины отличаются друг от друга более, чем это можно разумно приписать их случайной вариации, то разность между ними называется *значимой* или *существенной*. В противном случае разность называется *несущественной* или *случайной*.

Уровень значимости – граница между вероятностью существенного и несущественного различия.

Если вероятность полученной статистики равна или меньше уровня значимости, то гипотеза опровергается.

Можно дать другое определение:

Уровень значимости – вероятность, события, которой решено пренебречь в данной области исследований.

Уровень значимости обозначается, например, через α . Типичный смысл уровней значимости заключается в следующем. Пусть уровень значимости α равен 5%. Тогда, в предположении, что нулевая гипотеза верна, появление значений статистики, больших, чем при уровне значимости α , можно ожидать не более чем 5 раз на 100 испытаний. Появление такого события в одном испытании считается практически невозможным.

Значения статистики, при которых гипотеза опровергается, т. е. вероятность которых меньше заданного уровня значимости α , образуют *критическую область* проверяемой гипотезы.

Значения статистики, при которых гипотеза не опровергается, образуют *доверительную область*.

Границы между критической и доверительной областью называются *доверительными границами*.

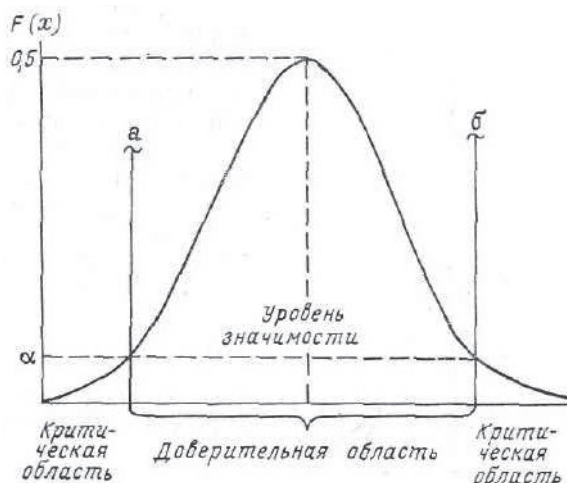


Рис. 6.1. Двухсторонний уровень значимости и доверительная область симметричного распределения: α – уровень значимости, a – область положительных отклонений от среднего, b – область отрицательных отклонений

Задача проверки гипотезы сводится к построению критической области рассматриваемой статистики для данного уровня значимости (рис. 6.1). Если статистика попадает в критическую область, то этот факт указывает на несоответствие гипотезы наблюдаемым данным, и гипотеза опровергается.

При помощи критерия значимости нулевая гипотеза может быть опровергнута. Но даже если гипотеза не опровергнута, это еще не значит, что она верна.

Действительно, пусть, например, уровень значимости $\alpha = 5\%$, а вероятность полученной статистики равна 6% . Опровергнуть гипотезу H_0 вы не можете, так как полученное значение не выходит в критическую область. Но считать ее верной тоже нельзя, так как вероятность получить такое значение статистики, в предположении, что гипотеза верна, составляет всего 6% , то есть в данном случае событие является маловероятным.

На это необходимо обратить особое внимание. Только в случае достаточно высокой вероятности полученного результата при данной нулевой гипотезе можно утверждать, что рассматриваемая гипотеза, по-видимому, верна.

Выбор уровня значимости или вероятности события, которой решено пренебречь в данной области исследования, производится

в некоторой степени произвольно. Он устанавливается на основании опыта, как уровень, дающий определенную практическую уверенность, что ошибочные заключения будут сделаны только в очень редких случаях. Например, в расчетах и прогнозах широко используются 10, 5, реже используется 1%-й уровень значимости ($\alpha = 10, 5, 1 \%$). В других областях научных исследований уровень значимости может быть другим.

Следует отметить, что при увеличении уровня значимости (рис. 6.1) увеличивается критическая область, а следовательно, возрастает и вероятность попадания исследуемой статистики в критическую область. Вместе с тем, очевидно, возрастает и вероятность ошибочного отбрасывания гипотезы. Действительно, при уровне значимости α , равном, допустим, 20 %, будут опровергаться все гипотезы, вероятность статистик которых составляет 20 % и менее. Однако значение этих статистик и в этом случае, если гипотеза верна, возможно в среднем в одном из пяти случаев. Поэтому к отбрасыванию гипотез при больших уровнях значимости надо подходить очень осторожно.

С уменьшением уровня значимости возрастает число испытаний, необходимых для эффективного применения критерия значимости, так как при малом объеме выборок n возможна такая ситуация, когда гипотеза вообще не может быть опровергнута.

В качестве основного критерия при выборе того или иного уровня значимости выступает только возможность его использования на практике. С одной стороны, этот уровень должен быть достаточно велик для отбрасывания ложных гипотез, а с другой – он должен быть достаточно мал, чтобы приводить к отбрасыванию лишь немногих верных гипотез. Поэтому в каждой области исследований обычно вводится несколько уровней значимости и в каждом конкретном случае надо выбирать тот из уровней, который наиболее отвечает задаче данного исследования.

При этом, при всех прочих условиях, необходимо учитывать следующие важные обстоятельства.

1. Уровень значимости уменьшается с повышением важности расчета. Так, если проверяются, например, гипотезы о надежности или возможности отказа какой-то важной конструкции, то уровень значимости может быть принят равным даже 0,01 %.

2. Уровень значимости должен быть в какой-то степени согласован с точностью исследуемых исходных данных и возможностью увеличения объема выборки. Так как в геоэкологии погрешность

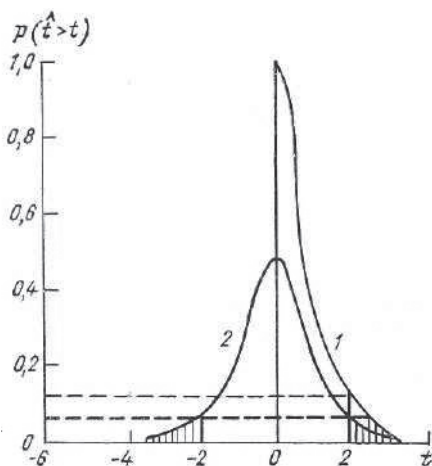


Рис. 6.2. Односторонний (1) и двухсторонний (2) критерий значимости

измерений часто превышает или равна 5 %, то использование уровня значимости меньше 5 % не всегда достаточно обосновано.

При проверке гипотез необходимо различать **двухсторонний (2а)** и **односторонний (а)** уровень значимости. Так, если при сравнении двух случайных величин требуется оценить их расхождение, т. е. одинаковый интерес представляют как положительная, так и отрицательная разность между изучаемыми величинами, то, допустим, при 5%-м уровне значимости, берется по 2,5 % на каждом конце (рис. 6.2). Наиболее часто эта ситуация встречается при проверке средних значений, когда одинаково важны отклонения как в положительную, так и в отрицательную сторону.

Односторонний критерий значимости используется в тех случаях, когда надо убедиться, что одно значение строго больше (меньше) другого.

При проверке гипотез возможны погрешности двух родов. Погрешность первого рода состоит в том, что нулевая гипотеза отвергается в то время, когда она в действительности верна. Чем меньше уровень значимости, тем меньше вероятность отвергнуть верную гипотезу.

Погрешность второго рода состоит в том, что нулевая гипотеза принимается, а на самом деле она неверна. Критерий значимости указывает на величину риска сделать погрешность первого (второго) рода.

6.2. Проверка гипотез о законе распределения

6.2.1. Постановка задачи

В гидрометеорологических исследованиях используется целая группа законов распределения и созданных на их основе кривых обеспеченностей. Из них наибольшее распространение получили кривые обеспеченностей нормального закона распределения; распределения Пирсона III типа, Крицкого–Менкеля, Гумбела, Джонсона; логарифмически нормального закона и некоторых других. Описание этих законов распределения и кривых обеспеченностей дано в гл. 4.

По имеющемуся ряду наблюдений необходимо выбрать один из перечисленных законов распределения. Однако для объективного суждения о лучшем или худшем соответствии того или иного закона распределения по имеющимся данным наблюдений, особенно при ограниченном объеме выборок, необходимо иметь какой-то количественный показатель, который измерял бы степень случайного расхождения (или согласия) между наблюдаемым рядом и предполагаемым распределением случайной величины в целом. Такой показатель также, очевидно, должен являться случайной величиной.

Подобные показатели, называемые *критериями согласия*, представляют собой определенную меру расхождения между теоретическим (или гипотетическим) и эмпирическим распределением и основаны на непараметрических гипотезах вида (6.2). Если расхождение в рассматриваемом случае превосходит установленный предел, то гипотеза опровергается.

Имеется несколько видов критериев согласия, используемых в различных ситуациях.

6.2.2. Критерий согласия Колмогорова

В качестве показателя степени согласия эмпирической $\hat{P}(x)$ и теоретической $P(x)$ кривых обеспеченностей принята величина:

$$\hat{\lambda} = \hat{D}\sqrt{n}, \quad (6.4)$$

где $\hat{D}$ – наибольшая разность значений эмпирической $\hat{P}(x)$ и теоретической кривой обеспеченности:

$$\widehat{D} = \max |\widehat{P}(x) - P(x)|; \quad (6.5)$$

($\widehat{P}(x)$ и $P(x)$ – в долях единицы). Очевидно, что $\widehat{D}$, являющееся функцией случайной величины $\widehat{P}(x)$, также представляет собой значение случайной величины. Предельное распределение этой величины вне зависимости от гипотетического распределения, для условия, когда нулевая гипотеза верна, было установлено А.Н. Колмогоровым в следующем виде:

$$\lim P\{\widehat{D}\sqrt{n} \leq \lambda\} = \sum_{k=-\infty}^{\infty} (-1)^k (-2k^2\lambda^2) = k(\lambda). \quad (6.6)$$

Отсюда при достаточно больших n вероятность получить значение $\widehat{\lambda} = \widehat{D}\sqrt{n}$, большее, чем λ , равна $1 - k(\lambda)$, т. е.

$$P\{\widehat{\lambda} \leq \lambda\} = 1 - k(\lambda). \quad (6.7)$$

Значения вероятностей превышения $\widehat{\lambda} > \lambda - P = 1 - k(\lambda)$ для случая, если гипотеза (6.3) верна, табулированы. Задавая уровень значимости $\alpha = P = 1 - k(\lambda)$, можно по значению $1 - k(\lambda)$ определять критические значения λ_α . Если рассчитанное значение $\widehat{\lambda}$ больше критического при данном уровне значимости, то гипотеза H_0 опровергается, то есть расхождение эмпирического и гипотетического распределения считается существенным.

Если обеспеченность значения $\widehat{\lambda}$ по кривой обеспеченностей $\lambda_\alpha = f(1 - k(\lambda))$ меньше, чем 1 % и 0,1 %, то расхождение считается весьма существенным. Если обеспеченность $\widehat{\lambda}$ больше 1 % и меньше 5 %, то расхождение считается существенным. В противном случае, т. е. $\widehat{\lambda} \leq \lambda_\alpha$, гипотеза не опровергается, но это еще не значит, что она доказана.

На основании формул (6.5) и (6.6) может быть получена формула для доверительных границ $\widehat{\lambda}$ относительно гипотетического закона распределения в точке $\widehat{D}$:

$$P\left\{\frac{P(x) - \lambda}{\frac{\sqrt{n} \leq \widehat{P}(x) \leq P(x) + \lambda}{\sqrt{n}}}\right\} = k(\lambda). \quad (6.8)$$

Если рассчитанное значение $\hat{P}(x)$ попадает в доверительный интервал, то гипотеза (6.2) не опровергается, если это значение попадает в критическую область, то гипотеза опровергается.

Практическое применение критерия согласия Колмогорова для оценки правильности гипотезы H_0 состоит в следующем:

1) для отдельных значений ряда $x_i (i = 1, 2, \dots, n)$, расположенных в убывающем порядке, определяются (в долях единицы) эмпирические обеспеченности;

2) для этих же значений определяются обеспеченности по предполагаемой теоретической кривой обеспеченности;

3) находится наибольшее значение разности $\hat{D}$ [форм. (6.4)] и $\hat{\lambda} = \hat{D}\sqrt{n}$;

4) по кривой обеспеченностей или таблице значений λ_α определяется обеспеченность $P_{\hat{\lambda}}$ рассчитанного значения $\hat{\lambda}$. Если $P_{\hat{\lambda}} < \alpha$, то гипотеза H_0 опровергается, в противном случае не опровергается.

Критерий согласия Колмогорова, как следует из формулы (6.5), предназначен для случая, когда n достаточно велико. Однако исследования, выполненные в последнее время, показали, что распределение статистики $\hat{\lambda}$ уже при $n \geq 20$ практически не зависит от n .

При небольших значениях n необходимо учитывать следующие обстоятельства:

1) критерий согласия $\hat{\lambda}$, определенный фактически по одной точке, в некоторых случаях может не отражать общую согласованность эмпирического и теоретического распределения;

2) наибольшее значение $\hat{D}$, как правило, отмечается в средней части кривой обеспеченностей. Однако именно в этой части почти все кривые обеспеченностей, используемые в практике, достаточно хорошо совпадают. Поэтому суждение о совпадении теоретических и эмпирических кривых обеспеченностей в краевых частях (именно это и нужно для практических расчетов) на основании критерия Колмогорова следует выносить очень осторожно.

6.2.3. Критерий согласия Пирсона $P[\chi^2]$

Для оценки согласия эмпирических данных $(x_1, x_2, \dots, x_n)$ и предполагаемого закона распределения используется расхождение между эмпирической $\hat{P}(x)$ и теоретической $P(x)$ вероятностью отдельных значений или интервалов значений случайной величины.

Если число значений дискретной случайной величины X ограничено $(\xi_1, \xi_2, \dots, \xi_k)$ и каждое $\xi_j (j = 1, 2, \dots, k)$ встречается в выборке (имеющегося ряда наблюдений) не менее 5 раз, то расчеты критерия согласия производятся по эмпирической и теоретической вероятности каждого значения ξ_j .

Если X является непрерывной случайной величиной или возможное число значений X велико так, что каждое или некоторые значения ξ встречаются менее 5 раз, то необходимо использовать метод группировки данных и перейти к рассмотрению дискретной системы, в которой рассматриваются вероятности отдельных интервалов значений X .

Пусть имеется выборка $x_1, x_2, \dots, x_n$. Значения выборки оформлены в виде статистического ряда распределения (см. разд. 3.3), в котором указаны эмпирические вероятности $\hat{P}$ (относительная частота) отдельных значений или интервалов значений X . Здесь же приводятся вероятности P_j значений ξ_j , определенные по предполагаемому теоретическому закону распределения.

В качестве меры расхождения между теоретической и эмпирической вероятностью принимается сумма квадратов отклонений

$$\sum_{j=1}^k [\hat{P}_j - P_j]^2, \text{ взятых с некоторыми весами } C_j:$$

$$\sum_{j=1}^k C_j [\hat{P}_j - P_j]^2.$$

Веса C_j вводятся потому, что в общем случае относящиеся к различным разрядам отклонения нельзя считать равноправными по значимости. Действительно, одно и то же ΔP_j может быть малозначительным, если сама вероятность $P_j(\xi)$ велика, и очень заметным, если она мала. Поэтому естественно, что C_j берутся обратно пропорционально значениям $P_j(\xi)$. Пирсон показал, что закон распределения:

$$\frac{n \sum_{j=1}^k [\hat{P}_j - P_j]^2}{P_j}, \quad (6.9)$$

если нулевая гипотеза верна, практически не зависит от закона распределения X и от числа опытов n , а зависит только от числа k . При больших значениях n этот закон приближается к так называемому распределению χ^2 :

$$\hat{\chi}^2 = \frac{n \sum_{j=1}^k [\hat{P}_j - P_j]^2}{P_j}. \quad (6.10)$$

Распределением χ^2 с ν степенями свободы называется распределение суммы квадратов ν независимых нормированных случайных величин, каждая из которых подчинена нормальному закону распределения с математическим ожиданием, равным нулю, и дисперсией, равной единице. Это распределение характеризуется плотностью вероятностей:

$$f(\chi^2) = \left\{ \frac{1}{\left[2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right) \right]} \right\} \chi^{2\left(\frac{\nu}{2}-1\right)} \exp\left(\frac{-\chi^2}{2}\right), \quad (6.11)$$

где ν – число степеней свободы:

$$\nu = k - 2; \quad (6.12)$$

$\Gamma(\nu/2)$ – значение гамма-функции (см. гл. 4).

На основе формул (6.11) и (6.12) составлены таблицы значений χ^2 для различных значений обеспеченности и числа степеней свободы (см., например [47], прилож.). По этой таблице можно определить вероятность события $\chi^2 \geq \hat{\chi}^2$, где χ^2 – значение, рассчитанное для данной нулевой гипотезы. Если вероятность окажется меньше заданного уровня значимости, то гипотеза опровергается. Аналогично можно сравнивать рассчитанное значение χ^2 с табличным при данных ν и α . Если $\hat{\chi}^2 > \chi^2(\alpha, \nu)$, то гипотеза опровергается.

Для обоснованного применения критерия согласия χ^2 необходимо иметь ввиду, что при выводе закона распределения χ^2

предполагается, что биномиальное распределение частоты превышения $\hat{p}_j(\zeta)$ может быть сведено к нормальному. Однако соответствующий предельный переход осуществляется достаточно быстро, если ни одна из вероятностей $\hat{p}_i(x)$ или $P_i(x)$ не очень мала. Поэтому при практическом применении критерия согласия $P(\chi^2)$ необходимо частоты крайних разрядов, представляющих обычно малые числа, объединять между собой.

Считается, что использование распределения χ^2 в качестве критерия согласия возможно только в тех случаях, когда длина выборки $n \geq 50$, а частота отдельных значений или интервалов значений X (в том числе полученная за счет объединения крайних интервалов) не меньше 5.

Недостатком метода является то, что группировка данных по интервалам в случае непрерывной случайной величины или большого числа возможных ее значений приводит к некоторой потере информации. Кроме того, элементы неопределенности и возможной неоднозначности решений вносятся при назначении числа интервалов и длины самих интервалов.

К достоинствам критерия χ^2 следует отнести универсальность – независимость от закона распределения и возможность использования для данных нечислового характера.

6.2.4. Критерий согласия $n\omega^2$

Критерий согласия $n\omega^2$ в отличие от критерия χ^2 основывается на оценке расхождения обеспеченностей непосредственно наблюдаемых (несгруппированных) значений рассматриваемой величины X .

Расчет критерия $n\omega^2$ производится по формуле:

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n (\hat{P}_i - P_i)^2, \quad (6.15)$$

где $\hat{P}_i$ и p_i – значения обеспеченности по эмпирической и предполагаемой теоретической кривым обеспеченностей для каждого данного $x_i (i = 1, 2, \dots, n)$. При этом $\hat{P}_i$ определяется по формуле:

$$\hat{P}_i = \frac{(i - 0,5)}{n}, \quad (6.16)$$

а математическое ожидание и дисперсия $n\omega^2$ по формулам:

$$M[n\omega^2] = \frac{1}{6}, \quad (6.17)$$

$$D(n\omega^2) = \frac{(4n-3)}{(180n^2)}. \quad (6.18)$$

При $n \geq 40$ распределение произведения $n\omega^2$ близко к некоторому предельному распределению, для которого определены верхние доверительные границы критерия $n\omega^2$ (табл. 6.1).

Таблица 6.1. Значения верхних доверительных границ критерия $n\omega^2$ в зависимости от уровня значимости α , %

α	20	10	5	3	2	1	0,1
$n\omega^2$	0,2412	0,3473	0,4614	0,5489	0,6198	0,7435	1,168

Оценка гипотезы о законе распределения на основе критерия $n\omega^2$ производится по следующей схеме:

- 1) ранжирование исходного ряда в убывающем порядке;
- 2) расчет эмпирических обеспеченностей членов ранжированного ряда по формуле (6.16);
- 3) определение обеспеченностей значений ранжированного ряда по предполагаемому закону распределения;
- 4) расчет критерий $n\omega^2$ по формуле (6.15);
- 5) сопоставление рассчитанных значений $n\omega^2$ с критическим $n\omega_{\alpha}^2$ при данном уровне значимости α . Если $\hat{n\omega}^2 > \hat{n\omega}_{\alpha}^2$, то гипотеза (6.2) о согласии кривых обеспеченности опровергается.

Критерий $n\omega^2$ обладает рядом преимуществ по сравнению с критерием χ^2 . Во-первых, при его применении полнее используется информация, заключающаяся в данных выборки, так как его расчет основывается на анализе непосредственно наблюдаемых величин. Во-вторых, его распределение значительно быстрее сходится к предельному закону, особенно при больших значениях $n\omega^2$.

6.3. Проверка гипотез о параметрах распределения

6.3.1. Постановка задачи

Пусть известно, что, начиная с момента t_0 , в бассейне реки началась или усилилась хозяйственная деятельность. Материалы наблюдений показывают, что статистические характеристики, полученные по наблюдениям в последние годы, несколько отличаются от прежних. Требуется выяснить, могут ли обнаруженные различия быть приписаны случайным колебаниям выборочных характеристик, или же нужно сделать вывод, что эти различия вызваны какими-то изменениями условий формирования рассматриваемого процесса.

Вопрос осложняется тем, что действительные значения параметров распределения исследуемой случайной величины, как правило, неизвестны. В связи с этим возникают три задачи:

- 1) получить наиболее точное представление о параметрах распределения случайной величины по эмпирическим данным;
- 2) определить соотношение параметров распределения различных процессов или процессов в различные периоды времени по эмпирическим данным;
- 3) оценить вероятность обнаруженных различий, если они являются следствием чисто случайных колебаний.

Чаще всего при наличии выборок достаточно большого объема для оценки параметров распределения применяются так называемые точечные оценки (оценки, рассчитанные по рядам наблюдений). Однако в большинстве случаев точечная оценка параметров еще мало определена, так как эти оценки представляют собой частные значения некоторой случайной величины. Поэтому решение первой задачи о наиболее точном представлении параметров распределения производится на основе построения доверительных интервалов. Решение второй задачи выполняется с помощью различных статистик, определяющих согласованность или рассогласованность сравниваемых по эмпирическим данным действительных значений исследуемых параметров.

***Доверительный интервал** – интервал, в котором с заданной вероятностью $1 - \alpha$ находится действительное значение искомого параметра.*

или

Доверительный интервал – интервал $(a - \delta, a + \delta)$ со случайными концами с вероятностью, равной $1 - \alpha$, покрывающий неизвестный параметр.

Для того, чтобы найти доверительный интервал какой-либо статистической характеристики, как следует из вышесказанного, необходимо, прежде всего, знать возможное распределение этой статистики. Именно поэтому главная задача статистического исчисления при малом числе наблюдений (ограниченный объем выборки) состоит в установлении точных распределений разных статистик. При таком толковании статистических оценок изменяются как сами вопросы, возникающие при исследовании точности расчетов, так и способы их решения.

6.3.2. Проверка гипотез о математических ожиданиях

Постановка задачи

При оценке гипотез о средних значениях используется статистика t , являющаяся нормированным отклонением выборочного среднего значения от математического ожидания (см. разд. 3.1.6):

$$\hat{t} = \frac{(\bar{x} - m_x)}{\sigma_x} = \frac{(\bar{x} - m_x)}{\left(\frac{\sigma_x}{\sqrt{n}}\right)}. \quad (6.20)$$

Математическое ожидание статистики t и ее среднее квадратическое отклонение равны соответственно 0 и 1 (см. разд. 3.1.3). Распределение статистики t подчиняется при некоторых условиях закону t -распределения Стьюдента:

$$S_v(t) = \left\{ \frac{\Gamma(v+1)}{2 \left[\pi v \Gamma\left(\frac{v}{2}\right) \right]} \right\} \int_{-\infty}^t \left(\frac{1+\tau^2}{v} \right)^{\frac{-(v+1)}{2}} d\tau, \quad (6.21)$$

где Γ – гамма; v – число степеней свободы:

$$v = n - 1. \quad (6.22)$$

Одним из этих условий является принадлежность сопоставляемых оценок математических ожиданий к одной генеральной совокупности.

Формула (6.21) выражает вероятность случайных значений τ , меньших, чем заданные значения t , т. е.

$$S_v(t) = p(\tau < t). \quad (6.23)$$

Указанное соотношение может быть использовано для определения значимости или существенности расхождения между выборочным и предполагаемым генеральным значением параметра распределения a , или между действительными значениями параметра a по разным выборкам. Так, если вероятность расхождения статистики t двух выборок с предполагаемым значением составляет менее 1–5 %, т. е. очень мало, то, по-видимому, расхождение существенно или значимо.

Обычно вопрос о значимости решается с помощью таблицы значений t , соответствующих данному уровню значимости α при данном числе степеней свободы v (см., например, работу [47], прилож. 7). В заголовке таблицы в первой строке указаны уровни значимости при использовании одностороннего критерия значимости (см. разд. 6.3), когда необходимо знать вероятность того, что t превосходит некоторое значение только в положительном (или только в отрицательном) направлении, а во второй строке указаны уровни значимости при использовании двухстороннего критерия значимости, когда определяется вероятность того, что t будет по абсолютной величине больше некоторого значения.

При малых значениях степеней свободы t -распределение Стьюдента заметно отличается от нормального закона распределения (рис. 6.3).

Основное различие заключается в том, что для формы распределения Стьюдента характерна большая острровершинность и более

длинные хвосты на концах (рис. 6.3). По мере возрастания v распределение стремится к нормальному и уже при $v = 30$ t -распределение практически не отличается от него.

t -распределение Стьюдента играет важную роль в геоэкологических исследованиях. Наибольшее распространение оно получило для

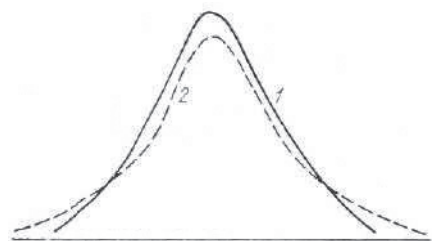


Рис. 6.3. Кривая нормального (1) и t -распределения (2)

оценки доверительных границ математического ожидания, оценки значимости среднего, оценки расхождения средних значений по двум и более рядам значений исследуемых процессов.

Определение доверительных границ математического ожидания

Пусть имеется выборка значений случайной величины X объемом в n членов. Среднее значение выборки: $\bar{x}$, несмещенная оценка дисперсии: $\hat{D}_x$. На основании этих данных требуется найти границы, внутри которых с определенной степенью надежности находится среднее значение генеральной совокупности m_x .

Выбрав двухсторонний 10%-й уровень значимости ($2\alpha = 10\%$), при данном числе степеней свободы ν находим $t_{0,05}$ такое, что $p(|\hat{t}| \geq t_{0,05}) = 0,05$, т. е. вероятность того, что $|\hat{t}|$ будет больше или равно табличному $t_{0,05}$ составляет 0,05. Тогда вероятность противоположного неравенства будет равна $p(|\hat{t}| < t_{0,05}) = 0,95$, т. е. вероятность того, что t находится в пределах $\pm t_{0,05}$ составляет 90 %. Подставляя в это равенство значение t , находим:

$$-t_{0,05} < \frac{\bar{x} - m_x}{\frac{\sigma}{\sqrt{n}}} \leq t_{0,05}. \quad (6.24)$$

Отсюда:

$$\frac{\bar{x} - t_{0,05}\sigma}{\sqrt{n}} \leq m_x < \frac{\bar{x} + t_{0,05}\sigma}{\sqrt{n}}. \quad (6.25)$$

Таким образом, с вероятностью 0,90 можно утверждать, что среднее значение совокупности лежит между значениями $\bar{x}_н$ и $\bar{x}_в$, равными соответственно:

$$\bar{x}_н = \frac{\bar{x} - t_{0,05}\sigma}{\sqrt{n}}, \quad \bar{x}_в = \frac{\bar{x} + t_{0,05}\sigma}{\sqrt{n}}. \quad (6.26)$$

Эти значения называются доверительными границами среднего значения общей совокупности при двухстороннем 5%-м уровне значимости, а промежуток с этими крайними

значениями накроет среднее значение общей совокупности, называется доверительной вероятностью.

Оценка значимости среднего значения

Под значимостью среднего значения понимается существенность отличия среднего значения X по данной выборке $\bar{x}$ от предполагаемого, значения математического ожидания $m_x = m_0$. Отличие, того или иного знака считается существенным (см. разд. 6.3), если вероятность его меньше заданного уровня значимости.

Схема оценки значимости заключается в следующем.

1. Задается нулевая гипотеза, что математическое ожидание m_x генеральной совокупности случайной величины X равно c , т. е. $H_0: m_x = c$.

2. По формуле (6.20) рассчитывается нормированное отклонение $\hat{t}$ от c :

$$\hat{t} = \frac{(\bar{x} - c)}{\left(\frac{\sigma_x}{\sqrt{n_1}} \right)}.$$

3. Если нулевая гипотеза H_0 верна, то распределение t должно подчиняться закону t -распределения Стьюдента. Находим по этому распределению при данном числе степеней свободы:

$$v = n - 1 \quad (6.27)$$

вероятность полученного $\hat{t}$. Если эта вероятность окажется меньше заданного уровня значимости α , то гипотеза H_0 опровергается, если вероятность окажется больше α , то гипотеза не опровергается.

На практике при оценке значимости часто поступают несколько иначе. Вместо определения вероятности $p(\hat{t} \geq t_\alpha)$ определяется критическое при проверке H_0 значение статистики t : $t = t_{2\alpha}$ и сравнивается с $\hat{t}$. Если $\hat{t} > t_{2\alpha}$, то гипотеза опровергается.

Оценка расхождения между средними значениями

Очень часто в исследованиях встает вопрос о сравнении математических ожиданий двух или более случайных величин по их выборкам.

Пусть, например, имеются две независимых ряда объемом n_1 и n_2 , взятые соответственно из генеральных совокупностей X и Y . Нулевая гипотеза в данном случае состоит в предположении, что математические ожидания рядов X и Y равны, т. е. $H_0: m_x = m_y$. В качестве критерия проверки H_0 естественно взять нормированную разность выборочных значений:

$$\hat{t} = \frac{(\bar{y} - \bar{x})}{\sigma_{(\bar{y} - \bar{x})}}, \quad (6.27)$$

где $\bar{y}$ и $\bar{x}$ – средние значения выборок Y и X ; $\sigma_{(\bar{y} - \bar{x})}$ – среднее квадратическое отклонение разности средних значений.

Если известны или предполагаются известными действительные значения средних квадратических отклонений рядов Y и X , то среднее квадратическое отклонение разности выборочных средних $\bar{y}$ и $\bar{x}$, при условии независимости рядов X и Y , может быть определено по формуле:

$$\sigma_{(\bar{y} - \bar{x})}^2 = \sigma_y^2 + \sigma_x^2 = \frac{\sigma_y^2}{n_1} + \frac{\sigma_x^2}{n_2} = \frac{(n_2 \sigma_y^2 + n_1 \sigma_x^2)}{(n_1 n_2)}. \quad (6.28)$$

В случае, если действительные значения σ_y и σ_x неизвестны, проверка H_0 производится, как правило, на основе предположения, что $\sigma_y = \sigma_x = \sigma$. В этом случае из формулы (6.28) следует:

$$\sigma_{(\bar{y} - \bar{x})}^2 = \frac{(n_1 + n_2) \sigma^2}{(n_1 n_2)}, \quad (6.29)$$

где σ определяется путем взвешивания по числу лет наблюдений:

$$\sigma^2 = \frac{(\sigma_y^2 n_1 + \sigma_x^2 n_2)}{(n_1 + n_2)}. \quad (6.30)$$

Из формул (6.27)–(6.30) получаем: для случая, когда σ_y и σ_x известны:

$$\hat{t} = \frac{(\bar{y} - \bar{x}) \sqrt{n_1 n_2}}{(n_2 \sigma_y^2 + n_1 \sigma_x^2)}, \quad (6.31)$$

для случая, когда σ_y и σ_x не известны:

$$\hat{t} = \frac{(\bar{y} - \bar{x})\sigma^2\sqrt{n_1 n_2}}{(n_2 + n_1)}, \quad (6.32)$$

где σ определяется по формуле (6.30).

Нормированная разность выборочных средних следует, закону t -распределения Стьюдента при числе степеней свободы:

$$v = n_1 + n_2 - 2. \quad (6.33)$$

Отсюда по таблице распределения t -статистики, так же как и в предыдущем случае, может быть определено критическое для данной гипотезы значение t_a при данном уровне значимости a .

Если значение $\hat{t}$ окажется больше t_a при данном уровне значимости a или $p(\hat{t} > t_a)$ окажется меньше a , то гипотеза $H_0 : m_y = m_x$ опровергается.

6.3.3. Оценка дисперсии

Постановка задачи

Оценка равенства действительных значений дисперсий по выборкам или отдельным частям одной выборки производится при помощи статистики F , называемой дисперсионным отношением:

$$\hat{F} = \frac{\hat{D}_1}{\hat{D}_2}, \quad (6.34)$$

где $\hat{D}_1$ и $\hat{D}_2$ – выборочные значения дисперсий. При этом в числителе берется большее значение.

Функция обеспеченности статистики F , выражающая вероятность того, что значение $\hat{F}$ будет больше или равно некоторому значению F :

$$P(\hat{F}) = p(\hat{F} \geq F) = \alpha, \quad (6.35)$$

где a – уровень значимости, установлена Фишером.

Функция плотности вероятности этого распределения (F -распределение) описывается формулой:

$$f(F) = \frac{v_1^{\frac{v_1}{2}} v_2^{\frac{v_2}{2}} \Gamma\left[\frac{(v_1 + v_2)}{2}\right]}{\left[\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)\right]}, \quad (6.36)$$

где v_1 и v_2 – число степеней свободы первой и второй выборки:

$$v_1 = n_1 - 1, \quad v_2 = n_2 - 1. \quad (6.37)$$

Как следует из формулы (6.36), F -распределение не зависит от дисперсии исходных рядов, а зависит только от числа степеней свободы. Это обстоятельство является очень важным, так как именно характеристику дисперсий и требуется установить в результате тех или иных исследований.

Для F -распределения составлены таблицы значений в зависимости от числа степеней свободы и уровней значимости. Имеются таблицы значений F , которые могут быть превзойдены с вероятностью $\alpha = 0,05; 0,025; 0,01; 0,005$, где число v_1 определяет столбец, а v_2 строку (см., например, работу [47], прилож. 8). В таблицах указаны значения для одностороннего критерия значимости, т. е. для оценки, например, $D_1 > D_2$. Если требуется оценить равенство дисперсий по двум выборкам, т. е. проверить $D_1 = D_2$, то уровень значимости следует удвоить.

Для оценки доверительного интервала и значимости дисперсии используется отношение:

$$\chi^2 = \frac{(n-1)\hat{D}}{D}, \quad (6.38)$$

которое удовлетворяет χ^2 -распределению с $\nu = n - 1$ степенями свободы, если гипотеза $H_0: \hat{D} = D$ верна [60].

Определение доверительных границ дисперсии

Пусть имеется выборка значений случайной величины X объемом в n членов. Среднее значение выборки $\bar{x}$, несмещенная оценка дисперсии $\hat{D}_c$. На основании этих данных требуется найти границы, внутри которых с определенной вероятностью находится действительное значение дисперсии D . Так как значение $\hat{D}$ может быть как больше, так и меньше D , то в данном случае при построении

доверительного интервала необходимо использовать двухсторонний критерий значимости при уровне значимости 2α .

Выбирая уровень значимости 2α , при данном числе степеней свободы ν находим такое значение χ_α^2 , при котором:

$$p\left(\hat{\chi}^2 \geq \chi_\alpha^2\right) = \alpha, \quad (6.39)$$

аналогично:

$$p\left(\hat{\chi}_\alpha^2 < \chi_{1-\alpha}^2\right) = 1 - \alpha. \quad (6.40)$$

Отсюда получаем, что вероятность, что значение $\hat{\chi}^2$ находится в пределах от $\hat{\chi}_\alpha^2$ до $\hat{\chi}_{1-\alpha}^2$:

$$p\left(\hat{\chi}_\alpha^2 \leq \hat{\chi}^2 < \hat{\chi}_{1-\alpha}^2\right) = 1 - 2\alpha. \quad (6.41)$$

Подставляя в это равенство значение $\hat{\chi}^2$, находим:

$$\hat{\chi}_\alpha^2 \leq \frac{(n-1)\hat{D}}{D} = \chi_{1-\alpha}^2,$$

и отсюда:

$$\left. \frac{(n-1)\hat{D}}{\chi_{1-\alpha}^2} \right\} < D < \frac{(n-1)\hat{D}}{\chi_\alpha^2}. \quad (6.42)$$

Оценка равенства дисперсий

Оценки гипотезы $H_0 : D_1 = D_2$ производятся по критерию Фишера (6.34). Если значение $\hat{F} > F_\alpha$, определенного по таблице при данном уровне значимости и числах степеней свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$, то расхождение между $\hat{D}_1$ и $\hat{D}_2$ можно считать существенным и, следовательно, гипотезу о том, что выборки взяты из нормальных общих совокупностей с одинаковой дисперсией опровергнутой. Если же $\hat{F} < F_\alpha$, то расхождение не существенно, и гипотеза о равенстве дисперсий не опровергается.

В некоторых случаях для оценки равенства дисперсий используется критерий Романовского. Расчет производится в следующей последовательности.

Сначала рассчитывается вспомогательная величина Θ :

$$\Theta = \left[\frac{(v_2 - 2)}{v_2} \right] \hat{F}, \quad (6.43)$$

где $\hat{F}$ – тисперсионное отношение, v_2 – число степеней свободы больше 4:

$$M[\Theta] = 1, \quad (6.44)$$

а среднее квадратическое отклонение:

$$\sigma(\Theta) = \sqrt{\frac{2(v_1 + v_2 - 2)}{[v_2 - 4]}}. \quad (6.45)$$

С вероятностью $p \geq 0,99$ можно ожидать, что абсолютная величина отклонения Θ от 1 по модулю не превышает 3σ . Отсюда следует, что, если выборки относятся к одной генеральной совокупности, то

$$R = \frac{|\Theta - 1|}{\sigma(\Theta)} < 3, \quad (6.46)$$

т. е. расхождение выборочных оценок дисперсий несущественно. Если $R \geq 3$, то расхождение существенно.

6.4. Непараметрические методы проверки гипотез

6.4.1. Постановка задачи

Рассмотренные выше статистические критерии, за исключением критериев оценки согласия эмпирического и теоретического распределения, связаны с проверкой параметрических гипотез. При проверке параметрических гипотез предлагается известный вид функции распределения генеральной совокупности (как правило, принимается нормальный закон распределения) и отдельные параметры.

В противоположность этому (см. разд. 6.2) при проверке непараметрических гипотез не вводится никаких допущений о конкретном

виде распределения. Таким образом, непараметрические гипотезы являются более общими, чем параметрические.

Большинство непараметрических критериев требует меньшего объема вычислений, чем параметрические. Их особенно удобно применять для быстрой оценки нулевой гипотезы.

Непараметрические критерии подразделяются на порядковые (ранговые) критерии и критерии, основанные на разности значений двух эмпирических распределений. К последним относятся рассмотренные выше критерии согласия.

6.4.2. Критерий Уилкоксона

В 1945 г. Уилкоксон предложил непараметрический метод проверки гипотез об однородности выборок, позволяющий осуществлять проверку этого предположения на основе рангового критерия.

Ранговый критерий – критерий основанный на анализе номера места членов рассматриваемых рядов в объединенном ряду, упорядоченным каким-либо образом.

Пусть даны две независимые выборки объема n_1 и $n_2(x_1, x_2, \dots, x_{n_1}$ и $y_1, y_2, \dots, y_{n_2})$. Требуется определить принадлежность этих выборок к одной генеральной совокупности, т. е. проверить гипотезу $H_0: F(x) = F(y)$. Так как критерий Уилкоксона принадлежит к категории ранговых критериев, то для его применения $n_1 + n_2$ данных обеих выборок совместно упорядочиваются. Например, наблюдения, полученные в выборках Y и X , располагаются в общую последовательность в порядке возрастания их значений:

$$y_1 x_1 x_2 y_2 y_3 y_4 x_3 y_5 y_6 x_4 \dots \quad (6.47)$$

Статистика ω критерия Уилкоксона задается выражением:

$$\omega = S(r_1) + S(r_2) + \dots + S(r_n), \quad (6.48)$$

где $r_1, r_2, \dots, r_n$ – ранги (порядковые номера), соответствующие значениям X ; S – некоторая функция.

Выбор функции $S(r)$ осуществляется так, чтобы чувствительность критерия в данном конкретном случае была бы, по возможности, наибольшей. Если предполагается что математические ожидания X и Y не равны между собой, то целесообразно положить $S(r) = r$ и вычислять ω по формуле:

$$\omega = r_1 + r_2 + \dots + r_m. \quad (6.49)$$

В выражении (6.48) ранги могут быть записаны следующим образом: 12345678910, где для значений выборки X значения рангов будут 2, 3, 7, 10, ...

Математическое ожидание и дисперсия статистики ω определяются по формулам:

$$M[\omega] = \frac{n_1(n_1 + n_2 + 1)}{2}; \quad (6.50)$$

$$D[\omega] = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}. \quad (6.51)$$

Критические значения двухстороннего критерия Уилкоксона ω с уровнем значимости 2α определяются нижним и верхним пределом. Значения нижнего предела $\omega_n(\alpha, n_1, n_2)$ при объемах выборок n_1 и n_2 , не превышающих 25, приводятся в соответствующих таблицах (см., например, работу [3], прилож. 9).

Поскольку распределение статистики ω симметрично относительно математического ожидания, то верхнее критическое значение ω_v связано с ω_n следующим соотношением:

$$\omega_v(\alpha; x_1, n_1) = 2M[\omega] + \omega_n(\alpha; n_1, n_2). \quad (6.52)$$

Если n_1 или n_2 больше 25, то для вычисления критических значений ω_n и ω_v используется нормальное распределение с параметрами, заданными формулами (6.51), (6.52).

6.4.3. Критерий Манна–Уитни

Манн и Уитни распространили указанный критерий Уилкоксона на сравнение двух независимых выборок. Этот критерий является наиболее распространенным и популярным ранговым критерием.

Значением критерия служит общее число инверсий U . Об инверсии мы говорим, когда в упорядоченном ряде $n_1 + n_2$ членов в порядке возрастания для произвольной пары (y_i, x_k) выполняется соотношение $x_k > y_i$, т. е. когда y_i стоит перед x_k . Иначе говоря, если какому-либо значению x в упорядоченном ряде предшествует y , то эта пара дает инверсию. Так, в последовательности (6.47) x_1 и x_2 дают по одной инверсии с y_1 , x_3 дает четыре инверсии с y_1, y_2, y_3, y_4 и x_4 дает шесть инверсий с $y_1, y_2, y_3, y_4, y_5, y_6$. Всего инверсий в перечисленных значениях упорядоченного ряда (6.48) будет:

$$U = 1 + 1 + 4 + 6 = 12.$$

Теоретически доказано, что при объемах выборок $n_1 > 10$ и $n_2 > 10$ число инверсий U распределено приблизительно нормально с математическим ожиданием:

$$M[U] = \frac{n_1 n_2}{2} \quad (6.53)$$

и дисперсией

$$D_u = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}. \quad (6.54)$$

Критическая область значений U для нулевой гипотезы об отсутствии существенных различий в сопоставляемых выборках с учетом нормального закона распределения U может быть представлена в виде:

$$U \leq M[U] - t_p \sigma_u; \quad (6.55)$$

$$U \geq M[U] + t_p \sigma_u. \quad (6.56)$$

6.4.4. Критерий знаков

Основан на учете различия между соответствующими значениями ряда X и Y . Нулевая гипотеза состоит в предположении, что рассматриваемые X и Y подчиняются одному и тому же закону распределения. Отсюда должно следовать, что разности $\Delta_i = x_i - y_i$ распределены симметрично около нуля, т. е. отклонение положительного и отрицательного знаков равновероятны. При этом, естественно, абсолютные величины разностей не имеют значения, а разности, равные нулю, исключаются из рассмотрения.

Таким образом, сравнивая ряды X и Y , можно определить знак разности и подсчитать число случаев со знаком «плюс» и со знаком «минус». Затем наименьшее число случаев сравнивается с критическим числом [47].

Глава 7.

Статистические зависимости и линейная корреляция в геоэкологии

7.1. Статистические зависимости

7.1.1. Задачи изучения взаимосвязей природных процессов

В геоэкологических исследованиях большое внимание уделяется анализу связей между различными природными процессами, определяющими общий фон режима исследуемых объектов.

Для изучения и возможного прогнозирования этих процессов необходимо установить их причины, движущие силы и тенденцию развития. Известно, что большинство из них формируется цепью причинно-следственных явлений, изменяющихся во времени и пространстве. Именно при этом возникает необходимость в исследовании зависимостей, связывающих рассматриваемые процессы между собой и с другими процессами и явлениями. Например, при расчетах экологического риска затопления территорий в период весеннего половодья требуется исследовать связи колебаний водности реки с колебаниями слоя дождевых и снеговых осадков, температуры воздуха, испарения с поверхности суши и водоемов, запасов воды в снежном покрове и т.д.

Задачи изучения причинно-следственных связей природных процессов можно условно разделить на два класса.

1) Выяснение причин или факторов развития процессов во времени и пространстве. Решение задач этого класса помимо научного значения, заключающегося в раскрытии процесса формирования исследуемых явлений, имеет и большое практическое значение, связанное с возможностью определения режима неизученных или слабо изученных процессов по косвенным данным.

2) Определение конкретных значений или тенденции развития данного процесса в будущем. Или, иначе, решение задач предсказания и прогноза будущих значений исследуемых явлений и восстановление пропущенных данных.

При решении этих задач устанавливается наличие и характеристики связей рассматриваемого процесса Y с другими процессами X_j , по которым можно получить недостающую информацию. Для этого обычно используются ряды наблюдений Y и X_j ($j = 1, 2, \dots, m$) за совместный период наблюдений. В практике ряд Y обычно называется зависимой переменной или функцией, ряды X – независимыми переменными или аргументами. Нередко также, особенно в прогнозах, ряд Y называется предиктантом, а ряды X – предикторами.

Следует отметить, что задачи исследования зависимостей природных процессов, вообще говоря, не являются чисто математическими, так как математический анализ в данном случае обязательно должен сопровождаться физическим анализом, отсутствие которого может привести к серьезным просчетам и, нередко, к парадоксальным результатам.

7.1.2. Функциональные и стохастические связи

Связь исследуемых процессов оценивается по соответствию изменений их значений во времени или пространстве или и во времени и в пространстве. Эта связь может быть линейной или нелинейной, более или менее тесной. Линейные связи более просты и для описания их требуется меньший объем информации. Однако многие попарные корреляционные связи являются нелинейными. Поэтому ниже, наряду с линейными, рассматриваются некоторые способы определения нелинейных связей, основанные на их линеаризации путем того или иного функционального преобразования.

По тесноте связи или степени определенности одного из сопоставляемых процессов относительно другого связи природных процессов могут быть разделены на две группы: функциональные и стохастические. Во многих случаях связь между сопоставляемыми рядами может отсутствовать вообще.

Функционально зависимые величины

Зависимость между рассматриваемыми процессами бывает настолько тесной, что, зная значения одного из процессов, можно указать точные значения другого.

Случайная величина Y называется **функционально зависимой от случайной величины X** если каждому значению X соответствует только одно возможное значение Y .

Функциональные связи встречаются в различного рода технических приложениях, в физике и т.д. Например, широко известен закон, связывающий давление P , температуру T и объем газа V :

$$P = \frac{RT}{V},$$

где R – постоянный коэффициент. В соответствии с этой функциональной зависимостью принимается, что заданной паре значений T и V соответствует только одно возможное значение P .

Широко известны функциональные связи различных величин из области гидравлики. Так функциональной обычно принимается зависимость гидростатического давления от высоты столба жидкости:

$$P = P_0 + \gamma h,$$

где P – гидростатическое давление; P_0 – давление на единицу площади свободной поверхности жидкости; γ – масса единицы объема жидкости; h – высота столба жидкости. В соответствии с этой зависимостью принимается, что заданному значению параметра h при фиксированных P_0 и γ соответствует только одно значение P . Действительно, если нанести соответствующие пары значений сопоставляемых элементов в виде точек (h, P) в поле графика в декартовых координатах (рис. 7.1), то полученная группа точек лежит строго на одной линии и каждому значению h соответствует одно вполне определенное значение P .

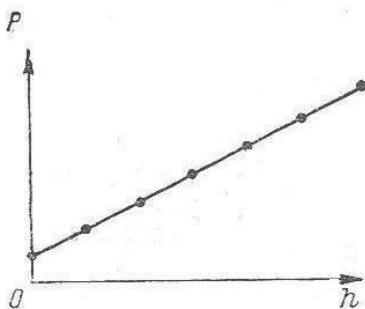


Рис. 7.1. График связи функционально зависимых величин

Независимые величины

Исследуемые процессы часто бывают независимыми, т. е. изменение значений одного процесса не зависит от изменений другого. Пусть, например, исследуемые процессы описываются случайными величинами X и Y , которые заданы последовательностями своих значений: $x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_n$.

Случайная величина Y называется **независимой от случайной величины X** , если закон распределения Y не зависит от того, какое значение приняла X , т. е. при любом x :

$$f\left(\frac{y}{x}\right) = f(y), \quad (7.1)$$

где $f(y/x)$ – условная плотность распределения, $f(y)$ – плотность распределения Y .

В более общем виде условие независимости выражается через функции распределения:

$$f(y, x) = F_1(y)F_2(x), \quad (7.2)$$

т. е. функция распределения системы независимых случайных величин равна произведению функций распределения отдельных величин, входящих в систему.

Условия (7.1) и (7.2) могут рассматриваться как необходимые и достаточные условия независимости.

График связи независимых величин (рис. 7.2) представляет собой поле точек, в котором каждому значению X принадлежит с соответствующей плотностью весь диапазон возможных значений Y и наоборот.

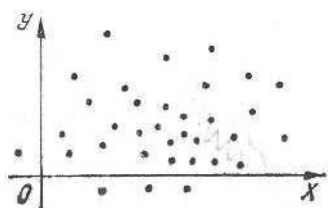


Рис. 7.2. График связи независимых величин

Вероятностная (стохастическая) зависимость

Чаще всего при исследовании характеристик окружающей среды встречается так называемая вероятностная или стохастическая зависимость.

Случайные величины Y и X называются **стохастически связанными**, если каждому значению X соответствует определенный диапазон значений Y и, наоборот, каждому значению Y соответствует определенный диапазон значений X .

Таким образом, если при функциональной связи заданному значению аргумента принадлежит только одно возможное значение функции, при отсутствии связи заданному значению аргумента соответствует весь диапазон возможных значений функции, то при стохастической связи заданному значению аргумента соответствует определенный диапазон значений функции.

Вероятностная зависимость может быть более или менее тесная. По мере увеличения тесноты вероятностной зависимости она все более приближается к функциональной, и диапазон возможных значений Y при заданном значении X уменьшается. По мере уменьшения тесноты сопоставляемые величины становятся независимыми.

Таким образом, функциональная зависимость – крайний предельный случай наиболее тесной вероятностной зависимости. Другой крайний случай – полная независимость. Между ними все градации вероятностной зависимости от самой сильной до самой слабой.

Если Y связан с X вероятностной зависимостью, то, зная значение X , нельзя указать точное значение Y , а можно указать только закон распределения Y , зависящий от того, какое значение приняло X .

Для стохастически связанных случайных величин X и Y можно записать:

$$f(x, y) = f_1(x) f_2\left(\frac{y}{x}\right), \quad (7.3)$$

т. е. плотность распределения системы двух случайных величин равна плотности распределения одной из величин, входящих в систему, умноженную на условную плотность другой величины, вычисленной в предположении, что первая величина приняла заданное значение.

На рис. 7.3 представлен график связи двух зависимых величин. Точки каждой пары значений этих величин располагаются в поле графика относительно какой-то линии. При этом каждому значению X соответствует своя полоса значений Y . Вместе с изменением X меняется среднее из возможных значений Y , при данном X меняются и экстремальные значения Y .

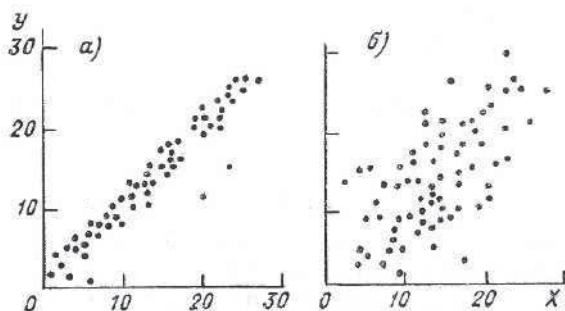


Рис. 7.3. График связи стохастически зависимых величин:
а – сильная связь, б – слабая связь

7.1.3. Различные виды связи природных процессов

Стохастическая связь между двумя процессами появляется обычно тогда, когда наряду с другими имеются общие факторы, влияющие как на первый, так и на второй процесс.

Так, например, если X зависит от случайных величин Z и U , т. е. $X = f(z_1, z_2, \dots, z_m; u_1, u_2, \dots, u_k)$, а Y зависит от той же случайной величины Z и некоторой другой величины V , то есть $Y = f(z_1, z_2, \dots, z_m; v_1, v_2, \dots, v_k)$, то, очевидно, за счет общей составляющей Z между X и Y появляется взаимосвязь, причем тем более тесная, чем больше влияние Z на X и Y .

Например, пусть имеется n речных бассейнов: $A_1; A_2, \dots, A_n$ (рис. 7.4). Каждый из бассейнов схематизируется в виде динамической системы. На вход динамической системы подаются сигналы, характеризующие воздействие различных факторов формирования стока – $\alpha_{\text{общ}}(t)$ и $\alpha_i(t)$. Здесь $\alpha_{\text{общ}}(t)$ – общие для группы бассейнов климатические и зональные бассейновые факторы, меняющиеся во времени, $\alpha_i(t)$ – местные бассейновые факторы. Бассейн преобразует эти сигналы, и на выходе получаем сигнал, характеризующий сток $Q(t)$.

Для представленной системы случайных величин все взаимосвязи $Q_i(t) = f[\alpha_{\text{общ}}(t)]$, $Q_i(t) = f[\alpha_i(t)]$, $Q_i(t) = f[Q_j(t)]$, где $i \neq j$, носят стохастический характер. При этом связи между $Q_i(t)$ и $\alpha_{\text{общ}}(t)$, а также между $Q_i(t)$ и $\alpha_i(t)$ объясняются непосредственным воздействием $\alpha_{\text{общ}}(t)$ и $\alpha_i(t)$ на $Q_i(t)$. Таким образом, в данном случае изменения $\alpha_{\text{общ}}(t)$ и $\alpha_i(t)$ выступают как причины изменения $Q_i(t)$. Поэтому эти связи называются **причинно-следственными**.

Связь между стоком воды различных рек $Q_i(t)$ и $Q_j(t)$ по этой схеме объясняется воздействием сигнала $\alpha_{\text{общ}}(t)$, являющегося общей

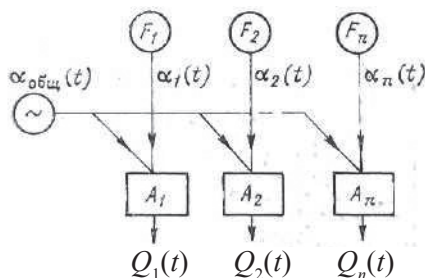


Рис. 7.4. Схема взаимосвязи стока системы речных бассейнов

причиной для всей рассматриваемой группы бассейнов. Связи разных следствий общих причин получили название **связи сопряженности**. Эти связи, очевидно, будут тем более тесными, чем больше влияние $\alpha_{\text{общ}}(t)$ и меньше влияние индивидуальных бассейновых факторов $\alpha_i(t)$.

7.2. Регрессионная математическая модель взаимосвязи двух переменных величин (парная корреляция)

Для исследования взаимосвязи двух природных процессов по имеющимся рядам наблюдений чаще всего используется математическая модель в виде уравнения регрессии. При этом решаются две задачи.

1) Выявление на основе большого количества наблюдений характера изменений Y в зависимости от изменения X . Эта задача кратко формулируется как задача определения формы (линейная или нелинейная) и нахождения параметров уравнения связи двух переменных величин.

2) Определение степени взаимосвязи исследуемых явлений (если это связи сопряженности) или степени влияния главных учитываемых факторов на исследуемое явление (если эти связи носят причинно-следственный характер).

7.2.1. Линейное уравнение регрессии двух переменных величин

Решение первой задачи чаще всего основывается на изучении линейной зависимости двух переменных величин. Общий вид линейной функциональной зависимости в этом случае:

$$Y = aX + b, \quad (7.4)$$

где a и b – параметры уравнения связи, легко и точно устанавливаемые по двум точкам линии связи исходных рядов наблюдений. В случае стохастической связи, когда точки связи имеют больший или меньший разброс относительно той или иной линии связи, по каждой паре точек в поле графика получаются различные значения

параметров a и b . Поэтому для определения прямой линии в поле точек (линии связи) в данном случае используется метод оптимизации, основанный на принципе наименьших квадратов (см. разд. 2.4).

Полученное в результате оптимизации уравнение, выражающее связь между значениями аргумента и функции, называется **уравнением регрессии**:

$$Y(\bar{X}) = aX + b. \quad (7.5)$$

В отличие от уравнения (7.4) в этом уравнении $Y(X)$ – *возможное значение* Y определенное на основе принципа наименьших квадратов. Следует отметить, что при выборе другого критерия качества значения параметров могли бы быть другими.

Действительное значение Y будет равно значению, полученному по уравнению (7.5), плюс возможное отклонение от линии связи – Δ , (по сути – **погрешность расчета по уравнению регрессии**), т. е.

$$Y(\bar{X}) = Y(X) + \Delta = aX + b + \Delta. \quad (7.6)$$

Для расчетов Y по уравнению (7.5) необходимо определить параметры a и b .

Пусть имеются последовательности значений X и Y , соответственно $x_i (i = 1, 2, \dots, N)$ и $y_i (i = 1, 2, \dots, N)$. Представим себе, что параметры a и b известны. Тогда, подставляя в уравнение регрессии значения X , можем получить для каждого значения X рассчитанное значение $Y - \bar{y}(x_i)$. Это значение, естественно, чаще всего не совпадает с фактическим значением, так как связь $Y = f(x)$ не является функциональной.

Как следует из формулы (7.6), разность фактических и рассчитанных значений Δ определяется по формуле:

$$\Delta_i = y_i - \bar{y}(x_i) = y_i - ax_i - b. \quad (7.7)$$

Дальше, исходя из принципа наименьших квадратов, находим такие значения параметров a и b , при которых среднее значение квадратов разностей рассчитанных и фактических значений S_0 является **минимальным**, т. е.

$$S_0^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \Delta_i^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N [y_i - ax_i - b]^2 = \min. \quad (7.8)$$

Для отыскания минимума функций S_0^2 от параметров a и b определяются частные производные по каждому из параметров и приравняются нулю. При этом предполагается, что функции $S_0^2 = f(a)$ и $S_0^2 = f(b)$ имеют по одному экстремуму, являющимися минимумом значений S_0^2 .

Исходя из этого, для определения параметра b получаем следующее выражение:

$$\frac{\partial S_0^2}{\partial b} = \frac{1}{N} \frac{\partial}{\partial b} \left[\sum_{i=1}^N (y_i - ax_i - b)^2 \right] = 0. \quad (7.9)$$

Отсюда:

$$\left(\frac{1}{N} \right) \sum_{i=1}^N y_i - \left(\frac{a}{N} \right) \sum_{i=1}^N x_i - \left(\frac{1}{N} \right) \sum_{i=1}^N b = 0.$$

и

$$b = \bar{y} - a\bar{x}. \quad (7.10)$$

Аналогично, обозначив отклонения значений Y и X от математического ожидания через δy и δx и, подставив найденные значения b , продифференцируем S_0^2 по параметру a :

$$\begin{aligned} \frac{\partial S_0^2}{\partial a} &= \frac{1}{N} \frac{\partial}{\partial a} \left[\sum_{i=1}^N (y_i - ax_i - \bar{y} - a\bar{x})^2 \right] = \\ &= \frac{1}{N} \frac{\partial}{\partial a} \left[\sum_{i=1}^N (\delta y_i - a\delta x_i)^2 \right] = 0. \end{aligned} \quad (7.11)$$

Отсюда:

$$\left(\frac{1}{N} \right) \sum_{i=1}^N (\delta y_i - a\delta x_i)(-\delta x_i) = 0$$

и

$$a = \frac{\sum_{i=1}^N \delta x_i \delta y_i}{\sum_{i=1}^N \delta x_i^2}. \quad (7.12)$$

Подставим найденные значения a и b (формулы 7.10 и 7.12) в уравнение регрессии (7.5), тогда

$$y_i - \bar{y} = \left(\frac{\sum_{i=1}^N \delta x_i \delta y_i}{\sum_{i=1}^N \delta x_i^2} \right) \delta x_i. \quad (7.13)$$

В рассмотренном выводе ряд X играл роль аргумента, ряд Y – роль функции. Однако возможны и такие случаи, когда X выступает в качестве функции, а Y – аргумента, т. е.

$$\bar{x}(y) = a'y + b'. \quad (7.14)$$

Повторяя предыдущий вывод, для этого уравнения получаем:

$$b' = \bar{x} - a'\bar{y}, \quad (7.15)$$

$$a' = \frac{\sum_{i=1}^N \delta x_i \delta y_i}{\sum_{i=1}^N \delta y_i^2} \quad (7.16)$$

и уравнение регрессии $x = f(y)$ имеет вид:

$$x_i - \bar{x} = \left(\frac{\sum_{i=1}^N \delta x_i \delta y_i}{\sum_{i=1}^N \delta y_i^2} \right) \delta y_i. \quad (7.17)$$

Таким образом, для определения формы связи процессов Y и X нами получены два уравнения: (7.13), описывающее линию регрессии Y по X , и (7.17), описывающее линию регрессии X по Y .

7.2.2. Коэффициент корреляции

Второй задачей определения взаимосвязи двух величин является определение тесноты связи.

Выражения (7.13) и (7.17) описывают две существенно отличающиеся друг от друга прямые регрессии, пересекающиеся в точке $(\bar{x}, \bar{y})$ (рис. 7.5) под углом φ . Чем более тесная связь между Y и X и чем, соответственно, меньше разброс точек в поле графика связи, тем меньше угол между линиями регрессии. Для функциональной

связи угол $\varphi = 0$, так как линии регрессии Y по X и X по Y совпадают; в случае отсутствия связи угол $\varphi = 90^\circ$, то есть линии регрессии проходят перпендикулярно друг другу и пересекаются в точке $(\bar{x}, \bar{y})$.

Направление линий регрессии в поле графика (см. рис. 7.5) определяется коэффициентами регрессии a и a' . Первый из них представляет собой тангенс угла наклона линии регрессии $y = f(x)$ к оси x ; второй – тангенс угла наклона линии регрессии $x = f(y)$ к оси y . Обозначим эти углы через α и β . Тогда коэффициенты регрессии:

$$a = \operatorname{tg} \alpha, a' = \operatorname{tg} \beta. \quad (7.18)$$

Сумма углов α , β , φ (см. рис. 7.5) составляет 90° .

В случае, если связь между Y и X функциональная, то $\angle \varphi = 0^\circ$ и $\angle \alpha + \angle \beta = 90^\circ$. Отсюда $\angle \alpha = 90^\circ - \angle \beta$, $\operatorname{tg} \alpha = \operatorname{tg}(90^\circ - \beta) = \operatorname{ctg} \beta$ и $\operatorname{tg} \alpha \operatorname{tg} \beta = 1$. Если связь между Y и X отсутствует, то $\angle \varphi = 90^\circ$, $\angle \alpha = \angle \beta = 0$ и $\operatorname{tg} \alpha \operatorname{tg} \beta = 0$.

С увеличением тесноты связи угол φ уменьшается от 90 до 0° ; вместе с тем увеличиваются углы α и β , а следовательно, тангенсы этих углов и их произведение.

Таким образом, произведение тангенсов углов α и β может служить мерой тесноты связи X и Y . Обычно в качестве критерия степени близости корреляционной связи к линейной функциональной зависимости используется корень квадратный из произведения $\operatorname{tg} \alpha \operatorname{tg} \beta$:

$$r = \sqrt{\operatorname{tg} \alpha \operatorname{tg} \beta}. \quad (7.19)$$

Это произведение называется **коэффициентом корреляции двух переменных величин** и обозначается через r или R . Чаще всего здесь будет использоваться первое обозначение.

Коэффициент корреляции принимает положительное значение, если Y возрастает с увеличением X (прямая связь), и отрицательное, если Y уменьшается с увеличением X (обратная связь).

Подставляя значения a и a' , вычисленные по формулам (7.12) (7.16), в формулу (7.19), получаем:

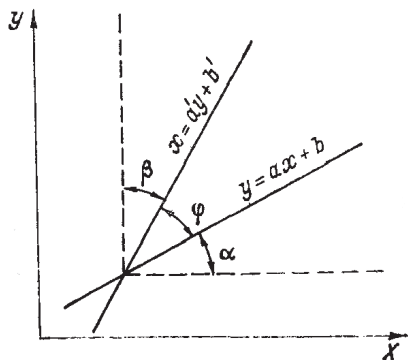


Рис. 7.5. Линии регрессии в поле графика (y, x)

$$r = \frac{\sum_{i=1}^N \delta x_i \delta y_i}{\sqrt{\sum_{i=1}^N \delta x_i^2 \sum_{i=1}^N \delta y_i^2}}. \quad (7.20)$$

Чаще используется другой вид записи формулы коэффициента корреляции – через σ_x и σ_y :

$$r = \frac{\sum_{i=1}^N \delta x_i \delta y_i}{(N \sigma_x \sigma_y)}. \quad (7.21)$$

Преобразуем уравнения регрессии Y по X и X по Y , выразив коэффициенты a и a' в формулах (7.13) и (7.17) через коэффициенты корреляции:

$$\begin{aligned} y_i - \bar{y} &= \left(r \frac{\sigma_x}{\sigma_y} \right) (x_i - \bar{x}) \\ x_i - \bar{x} &= \left(r \frac{\sigma_y}{\sigma_x} \right) (y_i - \bar{y}) \end{aligned} \quad (7.22)$$

или

$$\begin{aligned} \delta y_i &= \left(\frac{r \sigma_y}{\sigma_x} \right) \delta x_i, \\ \delta x_i &= \left(\frac{r \sigma_x}{\sigma_y} \right) \delta y_i, \end{aligned}$$

где σ_y и σ_x – средние квадратические отклонения Y и X ; $\bar{x}$ и $\bar{y}$ – средние значения этих рядов, r – коэффициент корреляции.

7.2.3. Оценка коэффициента корреляции и уравнения регрессии. Преобразование Фишера

Параметры уравнения регрессии в практических расчетах определяются по выборкам. Поэтому естественно, что они носят выборочный характер и могут использоваться лишь как более или

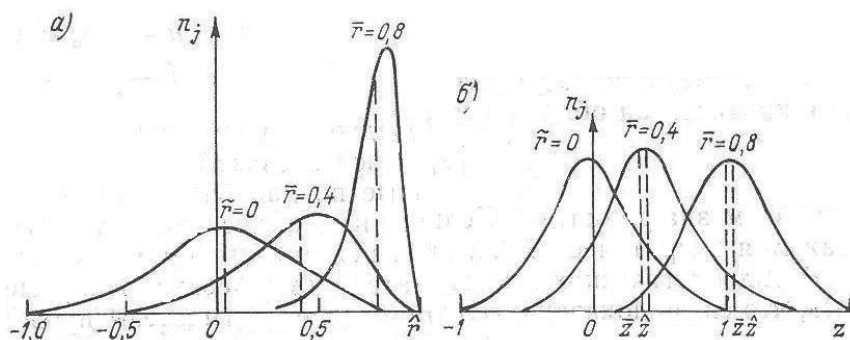


Рис. 7.6. Распределения: а) выборочного коэффициента корреляции r ; б) преобразования Фишера z при $n = 12$ и различных значениях r

менее достоверные оценки действительных значений. Обозначим выборочные оценки параметров a и b и коэффициента корреляции r через $\hat{\alpha}$, $\hat{b}$, $\hat{r}$ соответственно.

Известно [5, 11], что эмпирический (выборочный) коэффициент корреляции $\hat{r}$ представляет собой состоятельную оценку. Однако точность этой оценки или близость ее к действительному значению может быть установлена только при достаточно большом объеме выборки и невысоком коэффициенте корреляции, когда распределение его оценок может приниматься нормальным. Если же число испытаний невелико ($n < 30$), а коэффициент корреляции $\hat{r} \geq 0,4$, то распределение выборочных значений коэффициента корреляции существенно отличается от нормального.

Например, на рис. 7.6 представлено распределение оценок коэффициента корреляции r для значений $r = 0; 0,4; 0,8$ при $n = 12$.

Оценка коэффициентов корреляции и регрессии при $r < 0,4$ и $n > 30$

В этом случае закон распределения выборочных оценок коэффициента корреляции принимается нормальным. Поэтому средние квадратические погрешности оценки r , a и b при данном значении n и $\hat{r}$ могут определяться по формулам:

$$\sigma_r \approx \frac{(1 - \hat{r}^2)}{\sqrt{n - 1}}, \quad (7.23)$$

$$\sigma_a = \left(\frac{\sigma_y}{\sigma_x} \right) \sqrt{\frac{(1-r^2)}{(n-2)}}, \quad (7.24)$$

$$\sigma_b = \sigma_{\Delta y} \sqrt{\frac{1-r^2}{n-2}} \sqrt{1 + \left(\frac{\bar{x}}{\sigma_x} \right)^2}. \quad (7.25)$$

С учетом того, что распределение оценки коэффициента корреляции при указанных условиях подчиняется нормальному закону, то есть является симметричным, доверительные интервалы коэффициентов корреляции и регрессии при заданном уровне значимости можно определить по формулам:

$$\hat{r} - t_\alpha \sigma_r < r < \hat{r} + t_\alpha \sigma_r; \quad (7.26)$$

$$\hat{a} - t_\alpha \sigma_a < a < \hat{a} + t_\alpha \sigma_a; \quad (7.27)$$

$$\hat{b} - t_\alpha \sigma_b < b < \hat{b} + t_\alpha \sigma_b, \quad (7.28)$$

где t_α — статистика Стьюдента при двухстороннем уровне значимости α .

Нередко в исследованиях возникает вопрос о реальности связей вообще, т. е. вопрос о том, являются ли полученные по данным наблюдений коэффициенты корреляции значимыми и не объясняются ли они случайностями выборки.

Обычно этот вопрос решается с помощью нулевой гипотезы относительно рассматриваемой связи:

$$H_0 : r = 0.$$

То есть, исходим из предположения, что действительное значение коэффициента корреляции равно 0 и линейная связь отсутствует. Оценка гипотезы проводится по эмпирическому коэффициенту корреляции r . Для этого с учетом соотношения (7.26) строится критическая область вида:

$$|r| > t_\alpha \sigma_r, \quad (7.29)$$

с уровнем значимости α . Если полученное выборочное значение r окажется в критической области, то есть больше чем r_α , то гипотеза отбрасывается.

Во многих исследованиях, раньше, принималось, что связь доказана, если

$$\hat{r} \geq 4E_r, \quad (7.30)$$

где $E_r = 0,674\sigma_r$ – вероятная погрешность коэффициента корреляции. Однако эта оценка имеет излишне жесткий характер, так как уровень значимости такого значения r в предположении, что гипотеза верна, равен 1 %.

Оценка коэффициента корреляции при $r > 0,4$. Преобразование Фишера

Как отмечалось выше, при $r > 0,4$ для оценки действительного значения коэффициента корреляции ввиду асимметричного распределения выборочных оценок $\hat{r}$ нельзя использовать формулу средней квадратической погрешности коэффициента корреляции (7.23). Поэтому в этих случаях для оценки коэффициента корреляции применяется преобразование Фишера, при котором коэффициент корреляции выборочных совокупностей приравняется к гиперболическому тангенсу $\text{th}z$ некоторого значения z :

$$r = \text{th}z = \frac{\text{sh}z}{\text{ch}z}, \quad (7.31)$$

$$\text{где } \text{sh}(z) = \frac{(e^z - e^{-z})}{2}; \quad \text{ch}z = \frac{(e^z + e^{-z})}{2}.$$

Отсюда:

$$z = 0,5 \ln \left[\frac{(1+r)}{(1-r)} \right]. \quad (7.32)$$

Статистика Z имеет ряд замечательных свойств, а именно закон распределение Z почти не зависит от n и r и с возрастанием n быстро приближается к нормальному закону (рис. 7.6.б) со средним значением:

$$\bar{z} = \frac{\ln \left[\frac{(1+\hat{r})}{(1-\hat{r})} \right]}{2} + \frac{\hat{r}}{[2(n-1)]} \quad (7.33)$$

и средним квадратическим отклонением:

$$\sigma_z = \frac{1}{\sqrt{(n-3)}}. \quad (7.34)$$

К тому же, как следует из формулы (7.34), погрешность оценки z не зависит от $\hat{r}$.

При помощи статистики z можно выяснить, существенно ли отличается рассчитанное значение $\hat{r}$ от предполагаемого действительного значения, и найти его доверительные границы.

Для этого необходимо по r и $\hat{r}$ определить z и $\hat{z}$. Тогда ввиду нормального распределения статистики z легко решить, будет ли разность $r - \hat{r}$ существенной или несущественной.

При ручном счете переход от r к z и обратно достаточно сложен. Поэтому он совершается обычно по таблицам или по графикам связи $r = f(z)$ и $z = f(r)$ (рис. 7.7). При расчете на ЭВМ определение z по r или r по z производится непосредственно по формулам (7.31), (7.32), которые включаются в алгоритм программы.

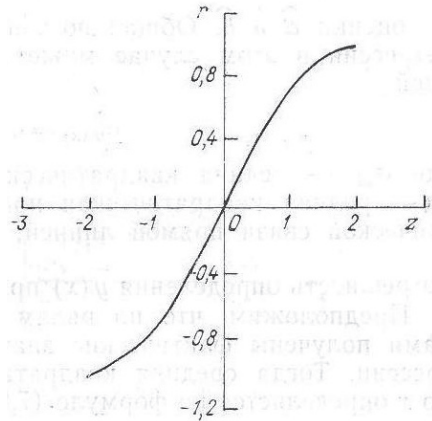


Рис. 7.7. Соотношение между коэффициентом корреляции r и преобразованием Фишера z

7.2.4. Анализ погрешностей расчетов по уравнению регрессии

Расчетные формулы

В результате расчетов по уравнениям регрессии (7.22) получают оценки значений Y . Фактические значения Y могут отличаться от этих оценок, причем тем больше, чем меньше теснота связи между Y и X и объем сопоставляемых выборок n . Таким образом, оценка Y по X всегда сопровождается погрешностями, во-первых, за счет аппроксимации стохастической связи прямой линией,

отображающей зависимость Y от X ; во-вторых, за счет выборочности исходных данных, вследствие чего в уравнениях регрессии (7.6) вместо действительных (генеральных) значений коэффициентов a и b используются их оценки $\hat{a}$ и $\hat{b}$. Общая погрешность расчетов Y по уравнению регрессии в этом случае может быть охарактеризована дисперсией:

$$\sigma_{y(x)}^2 = \sigma_{\Delta}^2 + \sigma_{\bar{y}(x)}^2, \quad (7.35)$$

где $\sigma_{y(x)}^2$ – дисперсия погрешности определения Y ; σ_{Δ}^2 – дисперсия погрешности аппроксимации стохастической связи прямой линией, характеризующая разброс точек связи относительно линии регрессии; $\sigma_{\bar{y}(x)}^2$ – средняя квадратическая погрешность определения $y(x)$ при данном x .

Предположим, что по рядам бесконечной продолжительности нами получены фактические значения параметров уравнения регрессии. Тогда средняя квадратическая погрешность расчетов y по x определяется по формуле (7.8). Подставим в эту формулу значение $\bar{y}_i(x)$, определенное по формуле (7.22):

$$\sigma_{\Delta}^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[y_i - \bar{y} - r \left(\frac{\sigma_y}{\sigma_x} \right) \delta x \right]^2. \quad (7.36)$$

Отсюда, после возведения в квадрат, получим:

$$\sigma_{\Delta}^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[\delta^2 y - 2r \left(\frac{\sigma_y}{\sigma_x} \right) \delta y_i \delta x_i + r^2 \left(\frac{\sigma_y^2}{\sigma_x^2} \right) \delta^2 x \right]$$

и, раскрывая квадратные скобки:

$$\sigma_{\Delta}^2 = \sigma_y^2 - \left(\frac{2}{N} \right) r \left(\frac{\sigma_y}{\sigma_x} \right) \sum_{i=1}^N \delta y_i \delta x + \left(\frac{1}{N} \right) r^2 \left(\frac{\sigma_y^2}{\sigma_x^2} \right) \sum_{i=1}^N \delta^2 x.$$

Дальше, учитывая формулы σ_y и σ_x , получаем:

$$\begin{aligned} \sigma_{\Delta}^2 &= \sigma_y^2 - \left(\frac{2}{N} \right) r \left(\frac{\sigma_y}{\sigma_x} \right) r N \sigma_y \sigma_x + \left(\frac{1}{N} \right) r^2 \left(\frac{\sigma_y^2}{\sigma_x^2} \right) \sigma_x^2 N = \\ &= \sigma_y^2 - 2r^2 \sigma_y^2 + r^2 \sigma_y^2 = \sigma_y^2 (1 - r^2). \end{aligned}$$

Таким образом, погрешность расчетов по уравнению регрессии за счет разброса точек относительно линии связи может быть определена по формуле:

$$\sigma_{\Delta} = \sigma_y \sqrt{1 - r^2}. \quad (7.37)$$

Рассмотрим теперь погрешность определения $\bar{y}(x)$ в формуле (7.35). Подставляя в формулу (7.6) значение b по формуле (7.10), получаем:

$$\bar{y}(x) = \bar{y} + \hat{a}(x - \bar{x}). \quad (7.38)$$

Здесь мы имеем линейную комбинацию $\bar{y}$ и $\hat{a}$. Поскольку последние не коррелированы, дисперсия $\bar{y}(x)$ равна сумме дисперсий $\bar{y}$ и $\hat{a}(x - \bar{x})$, т. е.

$$\begin{aligned} \sigma_{\bar{y}(x)}^2 &= \sigma_{\bar{y}}^2 + \sigma_{\hat{a}(x-\bar{x})}^2 = \frac{\sigma_{\Delta}^2}{n + \left(\frac{\sigma_{\Delta}^2}{(n\sigma_x^2)} \right) (x - \bar{x})^2} = \\ &= \left(\frac{\sigma_{\Delta}^2}{n} \right) \left[\frac{1 + (x - \bar{x})^2}{\sigma_x^2} \right] = \left(\frac{\sigma_{\Delta}^2}{n} \right) (1 + t^2), \end{aligned} \quad (7.39)$$

где t – нормированное значение x .

Учитывая (7.35) и (7.39), получаем:

$$\sigma_{\bar{y}(x)}^2 = \sigma_{\Delta}^2 \left(1 + \frac{1}{n} + \frac{t^2}{n} \right). \quad (7.40)$$

Теперь построим доверительные интервалы $y(x)$ и $\bar{y}(x)$. Очевидно, что если исходить из предположения о нормальном законе распределения Y и X , то доверительный интервал значений $\bar{y}(x)$ может быть рассчитан по формуле:

$$\bar{\bar{y}}(x) - t_{\alpha} \sigma_{\bar{y}(x)} \leq \bar{y}(x) < \bar{\bar{y}}(x) + t_{\alpha} \sigma_{\bar{y}(x)} \quad (7.41)$$

или

$$\frac{\bar{y}(x) - t_{\alpha} \sigma_{\Delta} (1 - t^2)^{\frac{1}{2}}}{\sqrt{n}} \leq \bar{y}(x) < \frac{\bar{y}(x) + t_{\alpha} \sigma_{\Delta} (1 - t^2)^{\frac{1}{2}}}{\sqrt{n}}. \quad (7.42)$$

Доверительный интервал значений $y(x)$ может быть аналогично выражениям (7.41) и (7.42) построен по формуле:

$$\bar{y} - t_{\alpha} \sigma_{\Delta} \left(1 + \frac{1}{n} + \frac{t^2}{n} \right)^{\frac{1}{2}} < y(x) < \bar{y} + t_{\alpha} \sigma_{\Delta} \left(1 + \frac{1}{n} + \frac{t^2}{n} \right)^{\frac{1}{2}}. \quad (7.43)$$

Заметим, что из формул (7.42) и (7.43) следует, что доверительный интервал $\bar{y}(x)$ и $y(x)$ растет вместе с увеличением отклонения x от $\bar{x}$, т. е. оценка этих величин с удалением x от $\bar{x}$ становится менее точной.

7.2.5. Проверка адекватности математической модели линейной корреляции на основе анализа погрешностей расчетов

В заключение исследования связи рассматриваемых процессов обычно проводится анализ адекватности подобранной математической модели линейной парной корреляции, то есть определяется насколько хорошо и в достаточной ли мере определяет эта модель характеристики связи между рядами Y и X , и может ли она использоваться на практике для определения Y по X .

Здесь следует отметить, что понятие адекватности нередко подменяется понятием значимости связи $Y = f(X)$. Однако значимость связи Y с X еще не означает, что ее использование может дать достаточно полное для данных целей исследования описание процесса Y .

Любая математическая модель, в том числе модель парной корреляции, становится адекватной и приобретает объективное содержание только при выполнении, как минимум, следующих условий [1]:

- 1) представления, составляющие априорное содержание модели, должны быть связаны конкретными соотношениями с объективными характеристиками изучаемого объекта;
- 2) должна быть разработана и использована четкая процедура проверки справедливости системы умозаключений, следующих из данной математической модели;

3) должна быть обеспечена определенная степень согласованности между результатами, полученными при экспериментировании с объектами, и соответствующими значениями, вычисленными или снятыми с модели;

4) должно быть определенное соответствие между предсказаниями на основе математической модели и реальным поведением объекта.

Если первые два условия формулируются и проверяются специалистами в той или иной области для решения конкретной задачи и могут иметь самые различные алгоритмы проверки, то остальные условия формулируются и проверяются на основе статистических методов. Однако и в этом случае условия адекватности во многом определяются конечной целью исследования. Поэтому в каждой области исследования могут быть выдвинуты свои критерии выполнения двух последних условий.

Например, в экологических прогнозах методика, основанная на той или иной математической модели, в том числе на математической модели линейной парной корреляции, во многих случаях считается эффективной (адекватной), если выполняются следующие условия:

$$\frac{S}{\sigma} \leq 0,70, \quad n \leq 15;$$

$$\frac{S}{\sigma} \leq 0,75, \quad 15 < n < 25;$$

$$\frac{S}{\sigma} \leq 0,80, \quad n \geq 25;$$

где S – средний квадрат отклонений значений, рассчитанных на основе уравнения регрессии (y_{pi}) от фактических (y_i), определяется или по формуле (5.28) или непосредственно по формуле:

$$S = \left[\frac{\sum (y_{pi} - y_i)^2}{(n-1)} \right]^{0,5},$$

σ – среднее квадратическое отклонение ряда предиктанта или во многих случаях – среднее квадратическое изменение предиктанта за период заблаговременности; n – период совместных наблюдений.

В экологических расчетах в соответствии с действующими нормативами математическая модель парной корреляции может использоваться при решении ряда задач, если $|r| \geq 0,7$.

В последнее время в целях более объективного анализа адекватности той или математической модели большое внимание уделяется проверке выполнения четвертого из названных выше условий. Эта проверка производится на независимых материалах, то есть материалах, не использованных при идентификации структуры и параметров модели.

Следует отметить, что для линейных связей коэффициент корреляции функционально связан с отношением дисперсии погрешностей расчетов (D_{Δ}) к дисперсии исходного ряда (D) или с отношением дисперсии ряда, рассчитанного по уравнению регрессии ($D_{y/x}$), к дисперсии исходного ряда:

$$r^2 = \frac{1 - D_{\Delta}}{D},$$

$$r^2 = \frac{D_{y/x}}{D}.$$

Если связь нелинейная, то эти равенства нарушаются, и несоответствие коэффициента корреляции и отношений указанных дисперсий может служить хорошим признаком нелинейности связей.

В практике экологических расчетов отношение D_{Δ}/D называется критерием случайности, а отношение $D_{y/x}/D$ – критерием детерминации:

$$\frac{D_{\Delta}}{D} = \frac{1 - D_{y/x}}{D}. \quad (5.42)$$

Следует отметить, что одним из методов проверки адекватности математической модели линейной регрессии является анализ остатков или погрешностей расчетов $\Delta_i = y_i - \bar{y}_i(x)$. В частности, для этого используются графики Δ_i в зависимости от x_i или $\bar{y}_i(x)$ (рис. 7.8). Если погрешности расчетов отдельных значений Y располагаются в поле графика в виде горизонтальной полосы, с центром на оси абсцисс, то модель можно рассматривать как адекватную (рис. 7.8.а).

Если полоса погрешностей в поле графика расширяется с возрастанием Y или X (см. рис. 7.8.б), то это указывает на гетероскедастичность связи, т. е. на отсутствие постоянства дисперсии σ_{Δ}^2 в зависимости от $\bar{x}$, что делает необходимым преобразование переменной Y .

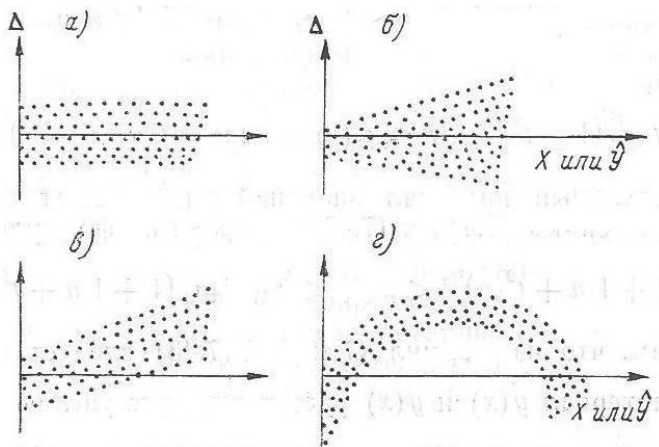


Рис. 7.8. Формы зависимости погрешности расчетов по уравнению регрессии от значений X или Y

График, показывающий линейный тренд (см. рис. 7.8.в), дает основание для введения в модель дополнительной переменной.

График, представленный на рис. 7.8.г, показывает, что в модель должен быть добавлен нелинейный или квадратический член.

7.2.6. Граничные условия математической модели парной корреляции

Многолетний опыт использования модели парной корреляции показал, что для корректного использования регрессионной математической модели необходимо выполнение ряда условий.

1. *Связь между сопоставляемыми рядами должна быть линейной.*

Условие следует непосредственно из того, что уравнения регрессии (7.22) получены исходя из уравнения прямой линии (7.4). Часто выполнение этого условия наглядно проверяется по графику зависимости $Y = f(X)$. Однако при малых коэффициентах корреляции или небольших объемах выборки этот прием оказывается неэффективным, так как картина связи носит размытый характер. В этом случае используется какой-либо другой вид контроля линейности связи. В частности, следует попытаться использовать различные приемы линеаризации связи. Если при этом коэффициент

корреляции значительно увеличивается, то можно считать, что связь имела нелинейный характер, и форма ее описывается формулой, использованной для линеаризации.

Например, зависимости вида

$$y = b \exp(ax), y = ax^b \quad (7.44)$$

могут быть преобразованы к линейному виду путем логарифмирования:

$$\ln y = \ln b + ax;$$

$$\ln y = \ln a + b \ln x.$$

В монографии [14] приводится достаточно полная таблица возможных преобразований, применяемых для линеаризации связи. Некоторые из них, например, логарифмирование, ведут также к частичной нормализации исходной информации.

При использовании названных преобразований необходимо обратить внимание на следующее обстоятельство [6, 40]: если переменные Y и X преобразованы, то параметры уравнения регрессии находятся не по минимуму среднего квадрата погрешности определения зависимой переменной Y , а по минимуму среднего квадрата погрешности определения преобразованной величины $U(Y)$. Например, в случае логарифмического преобразования минимизация ведется относительно среднего квадрата погрешности определения логарифмов исходного ряда, т. е.

$$S_{\Delta}^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[\lg y_i - \lg \bar{y}(x_i) \right]^2.$$

Поэтому независимо от того, какое преобразование используется, для линеаризации возникает необходимость введения поправок, чтобы избежать смещения.

Необходимо также иметь в виду, что в случае логарифмического преобразования необходимо в любых оценочных зависимостях увеличивать остаточную дисперсию в 1,15 раза [14].

2. Сопоставляемые переменные должны подчиняться нормальному закону распределения.

При выводе уравнений регрессии (7.22) использовался принцип наименьших квадратов. В математической литературе, например в работах [1, 14], доказывается, что корректное использование принципа наименьших квадратов предполагает нормальное

распределение сопоставляемых величин. Проиллюстрируем это требование.

1) На основе принципа наименьших квадратов параметры уравнения регрессии определяются с ориентиром на средний квадрат погрешностей расчета, т. е. погрешности расчета принимаются одинаковыми для всех значений Y или X , что верно только для нормального распределения. Пусть, например, Y имеет положительную асимметрию ($0 < Y < \infty$), а средняя квадратическая погрешность равна S_0^2 . Тогда действительные значения Y при данном X могут быть равны $y(x) - t_\alpha S_0^*$. Отсюда следует, что при $y(x) < t_\alpha S_0$ могут быть получены отрицательные значения Y , но это противоречит начальному условию, что Y заведомо положительно. Таким образом, ориентирование на средний квадрат погрешности, а следовательно, и использование принципа наименьших квадратов при асимметричном законе распределения некорректны и результаты расчетов по уравнению регрессии носят заведомо приближенный характер.

2) Положительная асимметрия предполагает, что количество малых значений, то есть значений меньших чем математическое ожидание, в любой выборке существенно больше, чем больших, которых может быть вообще очень мало. При этом несколько или даже одно большое значение оказывают значительное влияние на наклон линии регрессии и коэффициенты корреляции.

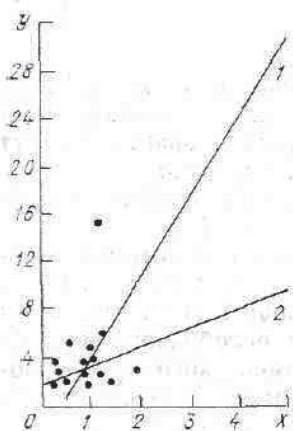


Рис. 7.9. График связи стока с осадками с учетом (1) и без учета (2) экстремальной величины

В этом случае применение рассматриваемой регрессионной модели может привести к неопределенным результатам. Так, на рис. 7.9 приводится пример взаимосвязи двух процессов: сток Y и осадки X [17]. Если включить все наблюдения, то сток хорошо коррелируется с осадками ($r = 0,797$, $\sigma_r = 0,08$, коэффициент веса $a = 6,04$). Если же исключить из расчетов максимальный паводок, то корреляционная связь резко нарушится ($r = 0,310$, $\sigma_r = 0,20$, коэффициент веса $a = 1,59$).

При этом в первом случае, когда учитываются экстремумы, значения стока выше среднего получаются в 2–3 раза чаще, чем во втором случае при исключении одного паводка. Таким образом,

учитывая асимметричность рядов наблюдений при построении уравнения регрессии, необходимо проводить тщательный анализ данных за расчетный период и репрезентативности его по отношению к максимальным значениям. В частности, в данном случае было бы необходимо провести изучение ливней по всему району.

Однако в практике расчетов регрессионная математическая модель часто используется и для рядов с асимметричным распределением. Естественно, что при этом в расчеты вносятся определенные погрешности, которые тем больше, чем больше асимметрия. Как показали расчеты, при $C_s \geq 2C_v$ и $C_v \geq 0,5$ погрешности могут быть настолько велики, что использование данной модели уже неэффективно. В этом случае рекомендуется перейти к одному из известных способов нормализации (получение по исходному ряду ряда с нормальным распределением) с последующим переходом от рассчитанных по уравнению регрессии величин, подчиняющихся нормальному закону распределения, к величинам, подчиняющихся закону распределения исходного ряда наблюдений.

3. Связь между рядами Y и X должна быть стационарной, т. е. коэффициент корреляции и коэффициенты регрессии не должны существенно меняться с изменением объема выборки или от выборки к выборке.

Проверка этого условия представляет известные трудности, поэтому чаще всего на практике не проводится. Между тем уже неоднократно встречались случаи, когда связи, установленные по имеющимся рядам наблюдений, с течением времени весьма значительно менялись, вплоть до изменения знака связи на обратный. Один из возможных путей проверки стационарности связи заключается в определении параметров уравнения регрессии по различным выборкам из имеющихся рядов наблюдений Y и X и проверки их равенства (см. разд. 7.2.3). Исходя из нулевой гипотезы, допустим:

$$H_0 : r_1 = r_2,$$

где r_1 – коэффициент корреляции по одним выборкам значений Y и X , а r_2 – по другим. При больших коэффициентах корреляции необходимо воспользоваться преобразованием Фишера.

4. Отклонения $\Delta_i = \bar{y}_i(x) - y_i$ должны быть взаимно независимы.

В практике исследований проверка этого условия, как правило, не производится. Между тем, если такая зависимость наблюдается, то характеристики оценок по уравнению регрессии существенно меняются [6].

5. Условная дисперсия значений Y при данном $X - \sigma_{\Delta}^2$ должна быть постоянной и не зависеть от X .

Это свойство называется *гомоскедастичностью*, а сами дисперсии – гомоскедастичными. Незамеченное отсутствие гомоскедастичности не приводит к смещенности оценок a и b , но в этом случае метод наименьших квадратов не гарантирует минимальности дисперсий оценок. В некоторых случаях, если известно, что σ_{Δ} пропорционально некоторой функции $G(x)$, можно воспользоваться методом взвешенных наименьших квадратов.

В практике расчетов по различным причинам граничные условия часто не выполняются или выполняются не полностью. К сожалению, нередко их выполнение вообще не проверяется. Следует помнить, что в этом случае применение регрессионной модели в исследовании данных процессов недостаточно оправдано и результаты расчетов носят приближенный характер; использование их при анализе взаимосвязей в некоторых случаях может привести к серьезным погрешностям.

7.3. Регрессионная математическая модель взаимосвязи ряда переменных величин (множественная корреляция)

7.3.1. Задачи исследования многофакторных связей

При исследовании различных природных явлений, значения которых во времени или пространстве представляются в виде последовательности значений случайных величин, часто встречаются ситуации, когда одна случайная величина, допустим Y , каким-либо образом определена по отношению к множеству случайных величин: $X_1, X_2, \dots, X_m$. Часто эта определенность используется для описания различных природных явлений на основе регрессионной математической модели по определенному множеству ряду переменных. Регрессионная модель часто позволяет провести полный статистический анализ исследуемого явления в его связи с внешней средой. Такой анализ является основой для физической интерпретации и

практического использования математической модели, например, в целях прогноза.

К настоящему времени опубликовано значительное число работ, в которых множественный регрессионный анализ применялся для идентификации математических моделей рассматриваемых процессов. Наряду с работами, содержащими положительный опыт применения регрессионного анализа, отмечались случаи, когда достичь желаемой точности математического описания не удавалось. Последнее обстоятельство послужило поводом для скептического отношения некоторых авторов к методу множественной корреляции как средству идентификации экологических процессов. По нашему мнению, основной причиной неудач является механическое применение схемы регрессионного анализа без учета специфических особенностей исследуемых объектов. Эти особенности весьма существенны и их нельзя игнорировать без риска получить отрицательный результат. Для учета их требуется разработка специальных мер, обеспечивающих работоспособность регрессионной математической модели при анализе процессов природы (см. ниже).

Задачи исследования многофакторных связей обычно решаются в два этапа. На первом этапе определяется степень влияния каждого фактора и их совокупности на исследуемое явление. На втором этапе по известным функциям влияния каждой переменной величины определяется неизвестное значение или значения исследуемого процесса (решение задачи интерполяции или экстраполяции).

Как и раньше обозначим значения исследуемого процесса (предиктанта) во времени или пространстве через Y . Значения факторов, влияющих на исследуемый процесс или связанных с исследуемым процессом обозначим через $X_1, X_2, \dots, X_m$. Эти факторы называются предикторами или независимыми переменными. Требуется установить форму и тесноту связи Y с $X_1, X_2, \dots, X_m$ и затем по известным значениям X_j найти неизвестное значение Y .

Например, требуется дать прогноз среднего притока в водохранилище на второй квартал. Известны значения 10 потенциальных предикторов: приток в течение предшествующего месяца X_1 , осадки на двух станциях в течение трех предшествующих месяцев: $X_2, X_3, X_4, X_5, X_6, X_7$, температура воздуха в течение предшествующего месяца по двум станциям X_8, X_9 , показатель циркуляции X_{10} , т. е. $Y = f(X_1, X_2, \dots, X_{10})$. Обычные методы построения графических связей в данном случае оказываются неприемлемыми. Необходимо искать какой-то аналитический метод расчета. Однако в этом случае

задача оказывается чрезвычайно сложной. В настоящее время наиболее корректно она решается лишь для случая линейных связей и при нормальном законе распределения сопоставляемых рядов.

При расчете многофакторных связей решаются следующие задачи.

1. Оценка тесноты связи между отдельными предикторами (факторами) или, иначе, независимыми переменными:

$$r_{hj} = \frac{\sum_{i=1}^N (x_{ji} - \bar{x}_j)(x_{hi} - \bar{x}_h)}{N\sigma_j\sigma_h} = \frac{\sum_{i=1}^N \delta x_{ji} \delta x_{hi}}{N\sigma_j\sigma_h}, \quad (7.45)$$

где r_{hi} – коэффициент линейной парной корреляции между h -им и j -им аргументами; N – число совместных значений ряда X_h и X_j (часто период совместных наблюдений); $\bar{x}_j$, $\bar{x}_h$, σ_j , σ_h – соответственно средние значения и средние квадратические отклонения h -го и j -го ряда.

2. Оценка тесноты связи каждого из предикторов с предиктантом:

$$r_{0h} = \frac{\sum_{i=1}^N (x_{hi} - \bar{x}_h)(y_i - \bar{y})}{N\sigma_h\sigma_y} = \frac{\sum_{i=1}^N \delta x_{hi} \delta y_i}{N\sigma_h\sigma_y}, \quad (7.46)$$

где r_{0h} – коэффициент линейной парной корреляции между последовательностью значений Y и аргументом X_h ; $\bar{y}$ и σ_y – соответственно среднее значение и среднее квадратическое отклонение ряда значений Y .

3. Оценка общей связи или определенности исследуемого явления с его предикторами или факторами по так называемому общему или сводному коэффициенту корреляции:

$$R_0 = \frac{\sum_{i=1}^N (y_i - \bar{y})(\bar{y}(x) - \bar{y})}{N\sigma_y\sigma_{\bar{y}(x)}} = \frac{\sum_{i=1}^N \delta y_i \delta \bar{y}(x)}{N\sigma_y\sigma_{\bar{y}(x)}}, \quad (7.47)$$

где $\bar{y}_i(x)$ – рассчитанные по данным X значения Y ; $\sigma_{\bar{y}(x)}$ – среднее квадратическое отклонение ряда рассчитанных значений.

При анализе парной линейной корреляции ранее было получено, что нормированные коэффициенты корреляции изменяются

в пределах от -1 до 1 , т. е. $-1 \leq r \leq 1$. Ниже будет показано, что сводный коэффициент корреляции меняется от 0 до 1 , т. е. $0 \leq R_0 \leq 1$.

4. Определение параметров уравнения множественной линейной корреляции:

$$\bar{y}(x_i) - \bar{y} = \sum_{j=1}^m a_j (x_{ji} - \bar{x}_j), \quad (7.48)$$

где a_j – коэффициент веса или влияния j -го аргумента на исследуемое явление, которое позволяет по заданным значениям предикторов X_j определить вероятное значение исследуемого явления Y .

7.3.2. Уравнение множественной линейной корреляции

Имеются выборки наблюдений, каждая объемом в N членов, над исследуемым явлением (ряд Y) и над m его предикторами (ряды $X_1, X_2, \dots, X_m$). Обозначим порядковое значение членов каждого из рядов через i , а номер ряда X через j ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$).

Идентификация структуры и параметров уравнения множественной линейной корреляции, как и раньше уравнения парной корреляции, будет производиться, исходя из принципа наименьших квадратов, основанном на том, что при нормальном законе распределения оценка коэффициентов уравнений a_j считается наилучшей, если средний квадрат разности между фактическими и расчетными значениями [см. формулу (7.8)] является наименьшим.

Подставим в формулу (7.8) значение, полученное по формуле (7.48). Тогда

$$S_0^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[y_i - \bar{y} - \sum_{j=1}^m a_j (x_{ji} - \bar{x}_j) \right]^2 \quad (7.49)$$

или

$$S_0^2 = \left(\frac{1}{N} \right) \sum_{i=1}^N \left[\delta y_i - \sum_{j=1}^m a_j \delta x_{ji} \right]^2, \quad (7.50)$$

где $\delta y_i = y_i - \bar{y}$, $\delta x_{ji} = x_{ji} - \bar{x}_j$.

Для определения значений a_j , при которых S_0^2 будет минимальным, достаточно взять частные производные S_0^2 по a_j и приравнять их к нулю.

Определим, например, коэффициент a_j для какого-то $j = h$:

$$\frac{\partial S_0^2}{\partial a_h} = \left(\frac{2}{N} \right) \sum_{i=1}^N \left(\delta y_i - \sum_{j=1}^m a_j \delta x_{ji} \right) (-\delta x_{hi}) = 0. \quad (7.51)$$

Отсюда

$$\sum_{i=1}^N \delta y_i \delta x_{hi} = \sum_{j=1}^m a_j \sum_{i=1}^N \delta x_{ji} \delta x_{hi}. \quad (7.52)$$

Из формул (7.45), (7.46) следует, что

$$\sum_{i=1}^N \delta y_i \delta x_{hi} = N r_{0h} \sigma_y \sigma_h, \quad (7.53)$$

$$\sum_{i=1}^N \delta x_{ji} \delta x_{hi} = N r_{hj} \sigma_j \sigma_h. \quad (7.54)$$

Подставив эти значения в формулу (7.52), получим:

$$r_{0h} \sigma_y = \sum_{j=1}^m a_j r_{hj} \sigma_j. \quad (7.55)$$

Значения h и j могут изменяться от 1 до m . Отсюда, расписав формулу (7.55) по всем значениям h и j , получим систему уравнений относительно коэффициентов веса $a_1, a_2, \dots, a_m$:

$$\begin{aligned} r_{01} \sigma_y &= a_1 r_{11} \sigma_1 + a_2 r_{12} \sigma_2 + \dots + a_m r_{1m} \sigma_m; \\ r_{02} \sigma_y &= a_1 r_{21} \sigma_1 + a_2 r_{22} \sigma_2 + \dots + a_m r_{2m} \sigma_m; \\ &\dots \dots \dots \\ r_{0m} \sigma_y &= a_1 r_{m1} \sigma_1 + a_2 r_{m2} \sigma_2 + \dots + a_m r_{mm} \sigma_m. \end{aligned} \quad (7.56)$$

Учитывая, что $r_{jj} = 1$, приводим эту систему к нормальному виду:

$$\begin{aligned} r_{01} &= \beta_1 + \beta_2 r_{12} + \dots + \beta_m r_{1m}; \\ r_{02} &= \beta_1 r_{21} + \beta_2 + \dots + \beta_m r_{2m}; \\ &\dots \dots \dots \\ r_{0m} &= \beta_1 r_{m1} + \beta_2 r_{m2} + \dots + \beta_m, \end{aligned} \quad (7.57)$$

$$\text{где } \beta_j = \left(\frac{\sigma_j}{\sigma_y} \right) a_j. \quad (7.58)$$

Поскольку имеется m уравнений и m неизвестных, то система имеет однозначное решение относительно коэффициентов a . Это решение может быть получено различными путями. В частности, могут быть использованы точные методы (метод Гаусса или Крамера) и приближенные: метод итерации, или, иначе, метод последовательных приближений, и как его развитие – метод Зейделя [47].

Расчет системы уравнений методом Крамера в конечном итоге сводится к расчету отношения определителей:

$$\beta_j = \frac{D_{0j}}{D_{00}}. \quad (7.59)$$

Отсюда с учетом выражения (7.58):

$$a_j = \left(\frac{\sigma_y}{\sigma_j} \right) \left(\frac{D_{0j}}{D_{00}} \right), \quad (7.60)$$

где D_{00} – главный определитель системы уравнений (7.57):

$$D_{00} = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{vmatrix}; \quad (7.61)$$

D_{0j} – определитель, получаемый заменой j -го столбца или j -й строки вектором свободных членов ($r_{01}, r_{02}, \dots, r_{0m}$):

$$D_{0j} = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{01} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{02} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & r_{0m} & \dots & 1 \end{vmatrix}. \quad (7.62)$$

Подставляя выражение коэффициентов a_j в уравнение регрессии (7.48), получаем:

$$\bar{\delta y}_i(x) = \sum_{j=1}^m \left(\frac{\sigma_0}{\sigma_j} \right) \left(\frac{D_{0j}}{D_{00}} \right) \delta x_{ji}. \quad (7.63)$$

Определенное таким образом уравнение дает, согласно нашему выводу, наилучшее по принципу наименьших квадратов линейное представление наблюдаемых значений Y в зависимости от X .

Для решения системы уравнений типа (7.57) на ЭВМ используются имеющиеся стандартные программы.

7.3.3. Сводные параметры множественной линейной корреляции

Дисперсия ряда Y может быть вычислена по формуле:

$$D_y = \left(\frac{1}{N} \right) \sum_{i=1}^N (y_i - \bar{y})^2 = \left[\frac{1}{(N-1)} \right] \sum_{i=1}^N [\bar{y}_i(x) + \Delta_i y - \bar{y}]^2, \quad (7.64)$$

где $\Delta_i y = y_i - \bar{y}_i(x)$ – погрешность расчетов по уравнению регрессии.

Отсюда, учитывая независимость $\delta y_i(x) = \bar{y}_i(x) - \bar{y}$ и Δy , получаем:

$$D_y = D_{y(x)} + D_{\Delta}, \quad (7.65)$$

т. е. дисперсия исходного ряда D_y равна сумме дисперсий ряда значений, рассчитанных по уравнению регрессии, и ряда погрешностей этих расчетов.

Для оценки точности вычислений по уравнению регрессии, или оценки определенности значений ряда y , по отношению к множеству рядов $X_j (j = 1, 2, \dots, m)$ наряду с коэффициентом корреляции R_0 широко используется отношение:

$$\delta = \frac{D_{\Delta}}{D_y}, \quad (7.66)$$

называемое дисперсионным отношением или критерием случайности. При $\delta = 1$, $D_{\Delta} = D_y$ и, следовательно [см. формулу (7.65)], $D_{y(x)} = 0$, т. е. значения Y не зависят от X_j . При $\delta = 0$, $D_{\Delta} = 0$ и $D_y = D_{y(x)}$, т. е. связь является функциональной. Таким образом, $0 \leq \delta \leq 1$.

Установим теперь взаимосвязь критерия случайности δ со сводным коэффициентом корреляции R_0 . Для этого подставим в формулу (7.66) значение D_{Δ} по формуле, аналогичной (7.37), получаем:

$$\delta = \frac{D_y (1 - R_0^2)}{D_y} = 1 - R_0^2.$$

Отсюда:

$$R_0^2 = 1 - \delta = \frac{1 - D_{\Delta}}{D_y}. \quad (7.67)$$

Следует отметить, что условие (7.67) выполняется только для линейных связей.

Как и для парной корреляции, особенно в последние годы, для характеристики эффективности расчетов по уравнению регрессии используется коэффициент детерминации:

$$F = 1 - \delta = \frac{D_{y(x)}}{D_y} = R_0^2. \quad (7.68)$$

Для расчетов R_0 по формулам (7.47) и (7.67) предварительно требуется определить a_j и $\bar{y}_j(x)$, что не всегда удобно. Поэтому обычно используется другая формула R_0 , параметры которой определяются непосредственно по матрицам значений r_{0j} и r_{hj} :

$$R_0 = \sqrt{\frac{1 - D_0}{D_{00}}}, \quad (7.69)$$

где D_{00} – определитель матрицы (7.61), D_0 – расширенный определитель системы (7.57), получаемый путем приписывания сверху и слева соответственно вектор строки и вектор столбца из коэффициентов корреляции $r_{0j} (j = 0, 1, 2, \dots, m)$ к определителю D_{00} :

$$D_0 = \begin{vmatrix} 1 & r_{01} & r_{02} & \dots & r_{0m} \\ r_{10} & 1 & r_{12} & \dots & r_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m0} & r_{m1} & r_{m2} & \dots & 1 \end{vmatrix}. \quad (7.70)$$

Можно показать также [2], что

$$R_0^2 = \left(\frac{1}{m} \right) (\alpha_1 r_{01} + \alpha_2 r_{02} + \dots + \alpha_m r_{0m}).$$

Эта формула замечательна тем, что она позволяет не только вычислять R_0 по известным a_j и r_{0j} , но и оценивать доли (проценты) вклада каждого из аргументов x_j в уравнение регрессии.

7.3.4. Частный коэффициент корреляции

При исследовании связи двух случайных величин Y и X коэффициент корреляции между ними может быть использован в качестве меры взаимозависимости. Однако при исследовании связей более чем двух случайных величин интерпретация коэффициента корреляции r в качестве меры физической взаимозависимости Y и X_j далеко не всегда достаточно обоснована. В этом случае, вследствие особенностей используемого математического аппарата, вполне возможно, что корреляция между Y и X_j наводится влиянием какой-то третьей переменной или группы переменных. Указанное обстоятельство приводит к необходимости исследования условных корреляций между двумя величинами при фиксированных значениях других величин. Коэффициент корреляции таких связей называется **частным коэффициентом корреляции**.

Если коэффициент корреляции Y и X_j уменьшается, когда фиксируется некоторая другая случайная величина или группа случайных величин, то это означает, что их взаимозависимость возникает частично, через воздействие этих величин; если же корреляция очень мала или вообще равна нулю, то, по-видимому, взаимозависимость Y и X_j целиком обусловлена этим воздействием. Наоборот, когда частная корреляция больше r_{0j} , то, вероятно, другие величины ослабляли связь, т. е. «маскировали» корреляцию [5].

Частный коэффициент корреляции ряда Y и X_1 обычно обозначается $r_{01.23\dots m}$, где первые две цифры – номера непосредственно исследуемых случайных величин, следующие после точки цифры означают номера фиксированных случайных величин. В этих обозначениях для расчета частного коэффициента корреляции используется формула:

$$r_{01.23\dots m} = \frac{-A_{01}}{(A_{00}A_{11})^{\frac{1}{2}}}, \quad (7.71)$$

где A_{jh} – алгебраическое дополнение для r_{jh} в расширенном определителе (7.69).

Например, если $Y = f(x_1, x_2)$, то частный коэффициент корреляции Y и X_1 будет определяться по формуле:

$$r_{01.2} = \frac{(r_{01} - r_{02}r_{12})}{\left[(1 - r_{12}^2)(1 - r_{02}^2)\right]^{\frac{1}{2}}}. \quad (7.72)$$

Следует отметить, что пока понятие частного коэффициента корреляции редко используется в исследованиях. Возможно, это вызвано тем, что оценка частного коэффициента корреляции менее точна, чем оценка любого парного коэффициента корреляции.

7.3.5. Оценка параметров уравнения множественной корреляции по данным наблюдений

Первичный анализ уравнений множественной корреляции производится на основе точечных оценок параметров этих уравнений. При этом для определения средних квадратических погрешностей эмпирического коэффициента множественной корреляции σ_{R_0} , средней квадратической погрешности эмпирических коэффициентов регрессии σ_{a_j} , средней квадратической погрешности расчетов $\bar{y}(x)$ за счет аппроксимации стохастических связей уравнением множественной линейной корреляции $\sigma_{\bar{y}(x)}$ обычно используются следующие формулы:

$$\sigma_{R_0} \cong \frac{\left(1 - \hat{R}_0^2\right)}{\sqrt{n - m - 1}}, \quad (7.73)$$

$$\sigma_{a_j} = \left(\frac{\sigma_0}{\sigma_j}\right) \sqrt{\left[\frac{\left(1 - \hat{R}_0^2\right)}{(n - m - 1)}\right] \left[\frac{\Delta_{0j}}{D_{00}}\right]}, \quad (7.74)$$

$$\sigma_{\bar{y}(x)} = \sigma_y \sqrt{1 - R_0^2}, \quad (7.75)$$

где Δ_{0j} – минор определителя, соответствующий его элементу r_{ij} , получается вычеркиванием j -й строки и i -го столбца в определителе (7.61); m – число предикторов.

Эти формулы применимы, строго говоря, в том случае, когда все m предикторов подчиняются порознь и в совокупности нормальному закону распределения, т. е. связаны нормальной корреляцией.

Как и ранее, при оценке парной корреляции, если сводный коэффициент корреляции $R_0 > 0,4$ или $n < 30$, то необходимо использовать

преобразование Фишера, с помощью которого решаются и все другие задачи оценки коэффициента корреляции (см. разд. 7.2.3).

Практически точечная оценка параметров уравнений регрессии всегда отличается от истинного значения. В некоторых случаях важно иметь надежное представление о возможных границах, т. е. границах доверительных интервалов истинного значения оцениваемого параметра, а также рассчитываемого значения Y .

Доверительный интервал сводного коэффициента корреляции R_0 оценивается так же, как для парного коэффициента корреляции r (см. разд. 7.2.3), по формуле:

$$\hat{R}_0 - t_\alpha \sigma_{R_0} \leq R_0 \leq \hat{R}_0 + t_\alpha \sigma_{R_0}, \quad (7.76)$$

когда $\hat{R}_0 \leq 0,4$, $n \geq 30$, либо с помощью преобразования Фишера. В последнем случае в начале рассчитывается доверительный интервал:

$$\hat{z} - t_\alpha \hat{\sigma}_z \leq z \leq \hat{z} + t_\alpha \hat{\sigma}_z, \quad (7.77)$$

где $\hat{z}$ – преобразование Фишера, вычисляемое по формуле (7.32); $\hat{\sigma}_z$ – средняя квадратическая погрешность z , определяемая по формуле (7.34); затем по нижним и верхним границам z находятся нижние и верхние границы сводного коэффициента корреляции: (см. разд. 7.2.3).

Доверительный интервал коэффициента веса a_j ($j = 1, 2, \dots, m$) определяется по формуле:

$$\hat{a}_j - t_\alpha \sigma_\Delta \sqrt{C_{jj}} \leq a_j \leq \hat{a}_j + t_\alpha \sigma_\Delta \sqrt{C_{jj}}, \quad (7.78)$$

где C_{jj} – соответствующий диагональный элемент обратной корреляционной матрицы:

$$C_{jj} = \frac{\tilde{A}_{jj}}{|A|}, \quad (7.79)$$

где $\tilde{A}_{jj}$ – алгебраическое дополнение диагонального элемента a_{jj} корреляционной матрицы; $|A|$ – определитель этой матрицы. В практических приложениях часто оказывается достаточным более простое приближенное выражение для определения доверительного интервала a :

$$\hat{a} - t_\alpha \hat{\sigma}_a \leq a \leq \hat{a} + t_\alpha \hat{\sigma}_a. \quad (7.80)$$

Доверительный интервал средних из возможных значений Y при заданных значениях $X_j (j = 1, 2, \dots, m) - \bar{y}(x)$ рассчитывается по выражению:

$$\bar{\hat{y}}_i(x) - t_a \sigma_{\Delta} \left(\frac{1}{n + d'A^{-1}d} \right) \leq \bar{y}_i(x) \leq \bar{\hat{y}}_i(x) + t_a \sigma_{\Delta} \left(\frac{1}{n + d'A^{-1}d} \right), \quad (7.81)$$

где d – вектор строки i -х значений рядов X_j в отклонениях от среднего:

$$d = (x_{1i} - \bar{x}_1, x_{2i} - \bar{x}_2, \dots, x_{mi} - \bar{x}_m); \quad (7.82)$$

A – ковариационная матрица, составленная из элементов a_{ik} :

$$a_{jk} = r_{jk} \sigma_j \sigma_k. \quad (7.83)$$

Доверительный интервал значений $\bar{y}_i(x)$ вычисляется по выражению:

$$\begin{aligned} \bar{\hat{y}}_i(x) - t_a \sigma_{\Delta} \left(1 + \frac{1}{n + d'A^{-1}d} \right) \leq y_i(x) \leq \bar{\hat{y}}_i(x) + \\ + t_a \sigma_{\Delta} \left(1 + \frac{1}{n + d'A^{-1}d} \right). \end{aligned} \quad (7.84)$$

В практических приложениях часто для определения доверительных интервалов оказывается достаточным использование упрощенного выражения:

$$\bar{\hat{y}}_i(x) - t_a \sigma_{\Delta} \leq y \leq \bar{\hat{y}}_i(x) + t_a \sigma_{\Delta}. \quad (7.85)$$

Здесь σ_{Δ} – погрешность расчетов по уравнению регрессии (7.75).

7.3.6. Определение оптимального числа аргументов. Исключение дублирующих и неэффективных аргументов

Главное затруднение при использовании методов множественной корреляции в экологических расчетах – это часто встречающаяся плохая обусловленность получаемых систем уравнений. Плохая обусловленность может быть вызвана двумя причинами.

1. Элементы корреляционной матрицы (7.61), образующие систему, находятся по эмпирическим данным, содержащим погрешности

определения. В результате привлечение дополнительных данных может привести не к уменьшению, а к увеличению общей погрешности элементов корреляционной матрицы.

Отсюда следует, что в расчетах по уравнению множественной корреляции необходимо использовать только те аргументы, которые дают объем информации, превышающий возможную погрешность расчетов. Эти аргументы называются эффективными. Оставшиеся неэффективные аргументы исключаются.

Вклад каждого аргумента в общую взаимосвязь может быть оценен по формуле [1, 5]:

$$\delta_j = \frac{r_{0j}a_j\sigma_j}{(R_0^2\sigma_0)} = \frac{r_{0j}a_j}{R_0^2}. \quad (7.86)$$

Если окажется, что δ_j незначительно, например, меньше удвоенной средней относительной квадратической погрешности квадрата сводного коэффициента корреляции:

$$\sigma_{R_0}^2 = 2R_0\sigma_{R_0} + \sigma_{R_0}^2, \quad (7.87)$$

то есть

$$\delta_j < \frac{2\sigma R_0^2}{R_0^2} \approx \frac{2\sigma R_0}{R_0}, \quad (7.88)$$

то аргумент X_j неэффективен и его целесообразно исключить из числа рассматриваемых.

2. Если какой-либо коэффициент корреляции между аргументами X_j и X_h , входящий в расчет определителя D_{00} (7.60), близок к единице, то аргументы X_j и X_h связаны функциональной или близкой к ней линейной зависимостью. Если это так, то очевидно, что коэффициенты корреляции связи j -го и i -го аргумента со всеми другими будут примерно равны, т. е. $r_{jk} \approx r_{hk}$ ($k = 1, 2, \dots, m$).

Отсюда следует, что в определителе D_{00} , построенном из парных коэффициентов корреляции, j и h – строка и столбцы – будут одинаковыми или почти одинаковыми. Но тогда, как следует из свойств определителей, определитель D_{00} системы линейных уравнений (7.57) будет близок или равен нулю.

В таких случаях, как следует из формулы Крамера (7.58), система (7.57) будет неопределенной, т. е. будет иметь неустойчивое решение относительно коэффициента a_j . Во избежание неустойчивости решения системы, а следовательно, и самого уравнения

множественной корреляции из каждой пары дублирующих друг друга аргументов необходимо исключить один из аргументов – либо X_h , либо X_j .

При этом в результате исключения дублирующих аргументов объем независимой информации, содержащейся в исходной матрице, уменьшается незначительно, но зато система уравнений (7.57) будет иметь устойчивое решение относительно коэффициентов регрессии.

В качестве критерия для определения дублирующих аргументов Г.А. Алексеев предлагает принять выполнение следующего неравенства [2]:

$$|r_{jh}| > 1 - 2\sigma_r, j \neq h \neq 0. \quad (7.90)$$

Следует отметить, что критерий, рассчитанный по этой формуле, зависит от периода наблюдений. Поэтому использовать его надо крайне осторожно. Опыт показывает, что лучше в качестве критерия взять коэффициент корреляции r_{jk} , равный 0,90 и проверить устойчивость результатов расчетов уже в ходе расчетов.

7.3.7. Граничные условия регрессионной математической модели взаимосвязи ряда процессов

Множественная корреляция является одним из немногих количественных методов, которые могут быть использованы для исследования взаимосвязей природных процессов, в том числе для оценки одновременного влияния нескольких факторов на данный процесс с целью его прогнозов и расчетов. Кроме того, этот метод позволяет определять относительное влияние каждого параметра и измерять полный эффект с помощью коэффициентов. Можно также оценить значимость связи между каждой зависимой и независимой переменной и получить «лучшее» расчетное уравнение.

Вместе с тем возможности множественной корреляции во многом не ясны. Одни ученые сообщают об успешных результатах применения множественной корреляции, другие – о неудовлетворительных. По-видимому, такая противоречивость вызывается тем, что использование регрессионной математической модели предъявляет к исходной информации ряд требований, соблюдение которых часто вообще не проверяется, а во многих случаях эти требования

не выполняются или выполняются не полностью. Коротко требования заключаются в следующем.

1. *Связи ряда – предиктанта Y с рядами-предикторами X_j ($j = 1, 2, \dots, m$) и между рядами-предикторами должны быть линейными.*

Так же, как и для математической модели парной корреляции (см. разд. 7.2), это требование следует из исходного уравнения, описывающего прямую линию (7.48). Однако, если для той математической модели проверка линейности связи часто могла проводиться визуально, то в данном случае такая проверка не представляется возможной, так как парные коэффициенты корреляции r_{0j} и r_{hj} редко превышают 0,5, и в то же время объем информации в исходных рядах ограничен. Поэтому используются косвенные методы проверки, основанные, во-первых, на физических соображениях о взаимосвязи тех или иных геофизических процессов, во-вторых, на учете эффективности того или иного предиктора или всей их совокупности по исходным и тем или иным методом линеоризованным рядам.

2. *Сопоставляемые переменные Y и X_j должны подчиняться нормальному закону распределения.*

Требование следует из того, что математическая модель множественной корреляции, так же как и модель парной корреляции, разработана на основе принципа наименьших квадратов (см. разд. 7.2). Большинство данных наблюдений, характеризующих природные явления, этому требованию не удовлетворяют. Обычно делается допущение: так как асимметричность законов распределения многих рядов невелика, то она сказывается при расчетах незначительно. Однако это допущение не всегда оправдано и может служить источником серьезных ошибок. Близость законов распределения выборок к нормальному является одним из главных показателей надежности любой математической модели, основанной на принципе наименьших квадратов. В связи с этим при ее применении требуется провести тщательный анализ асимметричности и при необходимости использовать один из методов нормализации, особенно объективный метод нормализации Г.А. Алексеева (см. ниже).

3. *Установленные связи должны быть стационарными (см. разд. 7.2.6).*

4. *Независимые переменные или предикторы должны измеряться без погрешностей.*

Этому требованию геоэкологические и другие данные, используемые в качестве предикторов, очевидно не удовлетворяют. И данные по стоку, и данные по осадкам и другим геофизическим

явлениям содержат определенные погрешности. Причем точность измерения или вычисления предиктанта нередко, особенно в экологии, может быть более высокой, чем точность предикторов. Чтобы избежать несоответствия исходных рядов, приходится рассматривать каждую совокупность независимых переменных как некоторые индексы, а не как точные значения.

5. Дисперсия зависимой переменной (предиктанта) не должна зависеть от независимых переменных, т. е. связь должна быть гомоскедастичной.

Это допущение также часто не оправдывается. Например, небольшие осадки ассоциируются с низкими значениями и небольшими вариациями стока, в то время как большие осадки вызывают большой сток и значительные его вариации.

6. Корреляция между рядами-предикторами должна отсутствовать или быть незначительной.

В действительности ряды-предикторы часто скоррелированы. Причем коэффициенты корреляции могут быть достаточно высокими. Наличие тесной связи приводит к неустойчивости коэффициентов регрессии, так как в этом случае корреляционная матрица (или система нормальных уравнений) становится плохо обусловленной, вырождающейся, ее детерминант стремится к нулю. Поэтому при обращении матрицы из-за косоугольности системы возникают погрешности, и обратная матрица в статистическом смысле определяется неточно. Сильная взаимосвязь предикторов приводит к затруднениям и при оценке результатов регрессии; во-первых, коэффициенты регрессии нельзя рассматривать как характеристику вклада соответствующих предикторов, независимую от остальных предикторов, с исключением одного или нескольких предикторов меняются остальные коэффициенты регрессии; во-вторых, форма прогностических уравнений может противоречить здравому смыслу.

Как указывалось выше, несколько улучшить решение можно путем исключения дублирующих предикторов – уменьшение многомерности статистической модели. С этой целью применяются какие-либо из известных методов просеивания исходных факторов, в том числе и изложенные выше методы исключения дублирующих и неэффективных аргументов.

7. Ряд-предиктант должен представлять собой выборку значений случайной величины, т. е. отдельные его значения должны быть некоррелированы между собой.

Это допущение в применении к геоэкологическим рядам часто не выполняется (см. гл. 8). Например, сток (годовой, месячный или ежедневный) часто коррелируется с предшествующим значением стока. В этом случае информация, содержащаяся во временных рядах, является в действительности меньшей, чем та, которая могла бы отвечать данному числу наблюдений.

8. *Объем выборки должен в несколько раз превосходить число независимых переменных.*

Помимо учета названных ограничений рассматриваемой математической модели следует также иметь в виду, что сам выбор модели и идентификация ее параметров содержат некоторые элементы неопределенности. Например, может случиться так, что будут использованы предикторы X_1, X_2, X_m и связь предиктанта Y с ними окажется настолько тесной, что включение в уравнение множественной корреляции других предикторов не имеет смысла. Но может случиться и так, что комбинация предикторов X_1, X_2, X_3 вообще не войдет в расчет, и эти предикторы никогда не будут использованы при определении Y .

В последнее время для того, чтобы найти именно те предикторы, которые как будто бы в наибольшей степени влияют на рассматриваемый процесс, на основе использования возможностей ЭВМ производится перебор всех возможных потенциальных предикторов и как расчетное принимается то уравнение и тот набор предикторов, которые дают наименьшую остаточную сумму квадратов. Этот метод имеет следующие недостатки.

1) Вследствие выборочности исходных данных в набор предикторов могут попасть ряды и явления, не имеющие ничего общего с рассматриваемым явлением. Независимо от этого сводный коэффициент корреляции будет возрастать с включением новых предикторов, но сама полученная взаимосвязь может быть фиктивной, сразу же нарушающейся с увеличением объема выборки. Поэтому первый этап набора потенциальных предикторов должен заключаться в физическом анализе и обосновании возможности взаимосвязи.

2) Большое число расчетных уравнений регрессии для определения коэффициентов веса делает определение статистической значимости влияния того или иного аргумента трудным, а результаты сомнительными.

3) Перебор большого количества предикторов может занять необоснованно много машинного времени.

7.3.8. Алгоритм расчетов на основе регрессионной математической модели

С изменением цели расчетов алгоритм и программа расчетов множественной линейной корреляции в некоторой степени меняются, однако, их общие черты во всех случаях остаются одинаковыми.

Алгоритм расчетов обычно представляется в следующем виде.

1. Комплектование массива исходных данных и ввод в память ЭВМ ряда-предиктанта Y и потенциальных предикторов $X_j (j = 1, 2, \dots, k)$.

В случае, если ряды Y и X_j имеют различную продолжительность, то в программе должен быть предусмотрен отбор значений рядов за совместный период. При этом совместный период может назначаться общим для всех рядов или только для данной группы рядов. Это определяется исключительно целями анализа.

2. Расчет числовых характеристик рядов Y и X_j : $\bar{y}$, $\bar{x}_j$, σ_y , $\sigma_j (j = 1, 2, \dots, k)$.

3. Расчет коэффициентов корреляции r_{0j} ряда Y с рядами $X_j (j = 1, 2, \dots, k)$ [см. формулу (7.46)].

4. Расчет коэффициентов корреляции r_{hj} потенциальных рядов-предикторов ($h = 1, 2, \dots, k; j = 1, 2, \dots, k$) [см. формулу (7.45)].

5. Расчет средних квадратических погрешностей коэффициентов корреляции r_{hj} и исключение дублирующих предикторов. При этом из каждой пары дублирующих предикторов, удовлетворяющих условию (7.90), исключается предиктор, имеющий меньшую скоррелированность с предиктантом.

6. Расчет сводного (множественного) коэффициента корреляции R_0 (7.47) и его средней квадратической погрешности σ_{R_0} (7.73) по оставшейся совокупности k' предикторов.

Если значение R_0 оказалось меньше заданного критического значения $R_{кр}$, то дальнейшие расчеты с таким набором потенциальных предикторов прекращаются. Значение $R_{кр}$ назначается исходя из целей экологического анализа. Например, при приведении параметров распределения к расчетному периоду $R_{кр}$ может приниматься равным 0,6–0,7, при восстановлении конкретных значений ряда Y $R_{кр}$ назначается 0,7–0,8 [42], и т.д. В некоторых случаях вводится дополнительное обязательное требование:

$$\frac{R}{\sigma_{R_0}} \geq 2, \quad (7.90)$$

при невыполнении которого дальнейшие расчеты с данным набором потенциальных предикторов также прекращаются.

7. Отбор оптимального числа m наиболее эффективных предикторов из набора потенциальных предикторов $X_j (j = 1, 2, \dots, k')$, оставшегося после исключения дублирующих предикторов. Отбор может производиться различными, методами. В частности, может быть использован метод сканирования (см. гл. 1.), когда рассматриваются все возможные сочетания предикторов при различном их числе и составе. Однако чаще всего используется более экономичный метод наибольшего градиента (см. гл. 1.), когда к отобранной группе из j предикторов ($j = 1, 2, \dots, k'$) присоединяется тот из оставшихся потенциальных предикторов, который дает наибольшее приращение (положительное или отрицательное) критерия качества. Обычно в качестве критерия качества принимается R_0 и σ_{Ra} .

Расчеты по отбору оптимального числа m предикторов при использовании метода наибольшего градиента производятся в следующем порядке.

- Из k' предикторов, оставшихся после исключения неэффективных и дублирующих, отбирается предиктор, имеющий наибольшее значение r_{0j} , равное в данном случае R_0 . Это первый эффективный предиктор. Обозначим его Z_1 .
- Из $k' - 1$ оставшихся предикторов отбирается предиктор Z_2 , дающий наряду с Z_1 наибольшее приращение R_0 .
- Из $k' - 2$ оставшихся предикторов отбирается предиктор Z_3 , дающий наряду с Z_1, Z_2 наибольшее приращение R_0 и уменьшение σ_{R0} , и т.д.

Процесс такой пошаговой регрессии продолжается до тех пор, пока не прекращается приращение R_0 и начинается увеличение σ_{R0} .

8. По отобранной группе из m наиболее эффективных предикторов производится расчет коэффициентов уравнения регрессии a_j и их средних квадратических погрешностей σ_{a_j} по формуле (7.75).

Если отношение $\frac{a_j}{\sigma_{a_j}} < k$, где k назначается исходя из целей анализа

и чаще всего принимается равным 2, то j -й предиктор исключается из расчета. Следует отметить, что обычно это не приводит к значительному уменьшению R_0 .

9. Расчет по полученным параметрам уравнения регрессии в соответствии с поставленной задачей экологического анализа.

10. Печать.

7.3.9. Пример расчета на основе регрессионной математической модели

Требуется на основании имеющихся рядов наблюдений продолжительностью в 49 лет идентифицировать структуру и параметры регрессионной математической модели для прогнозирования полезного притока в водохранилище с заблаговременностью в 1 месяц.

Пусть известно, что средний месячный приток в водохранилище Y может зависеть от десяти потенциальных предикторов: X_1 – приток в течение предшествующего месяца; X_2, X_3, X_4 – осадки по данным первой станции в течение трех предшествующих месяцев; X_5, X_6, X_7 – осадки по данным второй станции в течение трех предшествующих месяцев; X_8, X_9 – температура воздуха в два предшествующих месяца; X_{10} – показатель атмосферной циркуляции.

Требуется на основании имеющихся рядов наблюдений продолжительностью в 49 лет идентифицировать структуру и параметры регрессионной математической модели.

Расчеты производятся по следующей схеме.

1. Рассчитываются статистические характеристики всех рядов наблюдений. Результаты расчетов по рядам X представлены в табл. 7.1.

Таблица 7.1. Основные числовые характеристики по рядам предикторам

Номер ряда j	$\bar{x}_j$	σ_j	Номер ряда j	$\bar{x}_j$	σ_j
Y	3,49	0,45	X_6	3,81	0,60
X_1	3,64	0,47	X_7	3,85	0,56
X_2	3,32	0,51	X_8	–0,28	2,48
X_3	3,23	0,64	X_9	3,15	1,97
X_4	3,15	0,82	X_{10}	2,78	3,12
X_5	3,84	0,50			

2. Рассчитывается вектор-строка коэффициентов корреляции ряда Y с рядами X_j :

$$(r_{0j}) = (0,03 \ 0,37 \ 0,25 \ 0,25 \ 0,38 \ 0,30 \ 0,43 - 0,61 \ 0,23 - 0,05). \quad (7.91)$$

3. Рассчитываются парные коэффициенты корреляции между предикторами r_{hj} и строится матрица A_{00} :

$$A_{00} = \begin{vmatrix} 1 & 0,07 & 0,08 & 0,23 & 0,04 & 0,09 & 0,12 & 0,18 & 0,15 & -0,01 \\ & 1 & 0,19 & 0,05 & 0,66 & 0,03 & -0,01 & -0,13 & -0,20 & -0,10 \\ & & 1 & 0,14 & 0,20 & 0,52 & 0,05 & -0,19 & -0,01 & 0,14 \\ & & & 1 & 0,10 & 0,07 & 0,66 & -0,14 & -0,02 & -0,12 \\ & & & & 1 & 0,03 & 0,03 & 0,00 & 0,08 & -0,17 \\ & & & & & 1 & 0,13 & -0,26 & 0,32 & 0,21 \\ & & & & & & 1 & -0,26 & -0,04 & 0,01 \\ & & & & & & & 1 & 0,03 & 0,15 \\ & & & & & & & & 1 & -0,05 \\ & & & & & & & & & 1 \end{vmatrix} \quad (7.92)$$

4. Из всего набора предикторов, во избежание плохой обусловленности системы (7.57), определяются пары дублирующих аргументов, один из которых затем исключается. Для этого по формуле (3.25) рассчитывается средняя квадратическая ошибка максимального по абсолютной величине коэффициента корреляции $r_{jh} - \sigma_{j\max} = 0,09$. Затем по формуле (3.71) определяется, что дублирующими считаются предикторы, если коэффициент корреляции между ними $|r_{jh}| > 0,82$. В данном случае максимальная величина r_{jh} равна 0,66, и, следовательно, среди потенциальных предикторов дублирующих нет.

5. Производится отбор оптимального числа m наиболее эффективных предикторов и отсеиваются неэффективные. Для этого вначале из всей совокупности предикторов отбирается предиктор, имеющий наибольшую величину коэффициента корреляции r_{0h} . Коэффициент корреляции R_0 равен в данном случае r_{0h} , а σ_R рассчитывается по формуле (3.64). Затем из оставшихся девяти предикторов выбирается второй предиктор, который в совокупности с первым отобранным предиктором дает наибольший сводный коэффициент корреляции. Затем из оставшихся восьми предикторов выбирается предиктор, который в сочетании с двумя ранее выбранными дает наибольший сводный коэффициент корреляции. И так далее, до тех пор, пока не начнет возрастать средняя квадратическая ошибка коэффициента корреляции σ_R . В результате такого перебора получаем сочетание наиболее эффективных предикторов при различном их числе (табл. 7.2).

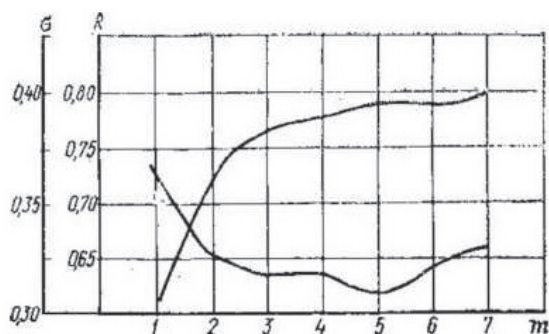


Рис. 7.10. График зависимости сводного коэффициента корреляции (r) и его средней квадратической ошибки σ_r от числа наиболее эффективных аргументов m

Обычно по результатам такого перебора строятся графики зависимости $R = f(m)$ и $\sigma_{R_0} = f(m)$ (рис. 7.10), по которым производится выбор оптимального числа предикторов. В данном случае, по-видимому, наилучшая регрессионная зависимость будет для случая $m = 5$ при наборе предикторов X_8, X_5, X_7, X_9, X_2 . Действительно, при $m = 5$ величина R_0 уже практически не возрастает, а величина σ_{R_0} начинает возрастать.

Таблица 7.2. Сочетание наиболее эффективных предикторов при различном их числе m_0

m_0	Предикторы	R_0	σ_R
1	X_8	0,61	0,37
2	X_8, X_5	0,72	0,33
3	X_8, X_5, X_7	0,77	0,32
4	X_8, X_5, X_7, X_9	0,78	0,32
5	X_8, X_5, X_7, X_9, X_2	0,79	0,31
6	$X_8, X_5, X_7, X_9, X_2, X_{10}$	0,79	0,32
7	$X_8, X_5, X_7, X_9, X_2, X_{10}, X_{11}$	0,80	0,33

Возможны и другие методы отбора эффективных предикторов. В частности, в соответствии с формулой (3.70) предварительно можно было исключить неэффективные предикторы.

6. Для оптимального числа наиболее эффективных предикторов строится матрица A_{00} :

$$A_{00} = \begin{vmatrix} 1 & 0,0 & -0,26 & 0,03 & -0,13 \\ 0,0 & 1 & 0,03 & 0,08 & 0,66 \\ -0,26 & 0,03 & 1 & -0,04 & -0,01 \\ 0,03 & 0,08 & -0,04 & 1 & -0,20 \\ -0,13 & 0,66 & -0,01 & -0,20 & 1 \end{vmatrix} \quad (7.93)$$

7. Рассчитывается определитель D_{00} . В данном случае $D_{00} = 0,08$.
8. Путем последовательной замены соответствующего столбца в матрице A_{00} столбцом свободных членов из коэффициентов корреляции r_{0j} строятся матрицы A_{nj} и рассчитываются определители D_{0j} .
9. По формуле (3.52) рассчитываются весовые коэффициенты:
 $a_2 = 0,11, a_5 = 0,11, a_7 = 0,14, a_8 = -0,23, a_9 = 0,09$.
10. Подставляя значения весовых коэффициентов в формулу (3.40), получаем:

$$\delta y_i = -0,23\delta x_{8i} + 0,11\delta x_{5i} + 0,14\delta x_{7i} + 0,09\delta x_{9i} + 0,11\delta x_{2i}.$$

Таким образом, в результате идентификации регрессионной математической модели притока воды в водохранилище определена структура модели и ее параметры. При построении модели выяснилось, что приток воды в водохранилище зависит от: температуры воздуха в два предшествующих месяца, осадков по данным второй станции в течение 3 и 1 предшеспредшесующих месяцев и осадков по данным первой станции в течение первого предшествующего месяца.

7.3.10. Преобразование исходной информации в целях нормализации и линеаризации взаимосвязей

Как отмечалось выше, использование аппарата множественной линейной корреляции для анализа экологических процессов связано с рядом ограничений, среди которых важнейшими являются требования о нормальном законе распределения исходной информации и линейном характере связи между отдельными рядами (см. разд. 4.3.1). Так как многие или даже большинство рядов наблюдений подчиняются асимметричным законам распределения, а связи между ними далеко не всегда линейные, то механическое использование

аппарата линейной множественной корреляции часто приводит к серьезным просчетам. Поэтому естественны попытки в какой-то мере учесть названные ограничения, например, путем преобразования исходной информации.

Недостатком использования изложенных ранее методов (см. разд. 4.2) преобразования в множественной корреляции является то, что для каждой пары сопоставляемых рядов может возникнуть необходимость в своем преобразовании, отличающемся от вида преобразования взаимосвязи других рядов. При этом очевидно, что путем субъективного подбора весьма трудно найти такие m функций $f_1(x_1), f_2(x_2), \dots, f_m(x_m)$, которые бы выравняли все возможные попарные корреляционные связи.

Для учета этих обстоятельств Г.А. Алексеевым был предложен [2] свой метод нормализации и линеаризации связей. Как следует из названия, этот метод может применяться одновременно и для нормализации асимметричных распределений и для спрямления монотонно возрастающих или монотонно убывающих корреляционных связей.

Монотонными связи называются тогда, когда все попарные кривые регрессии $\delta y = f(\delta x_j)$, $j = 1, 2, \dots$ представляют собой либо возрастающие, либо убывающие функции (без максимумов и минимумов) (рис. 7.11).

Нормализация исходных переменных заключается в двух последовательно выполняемых монотонных преобразованиях.

1. Замена исходных значений каждого ряда Y и X их эмпирическими вероятностями неперевышения, рассчитанными по формуле (3.90)¹. Эти эмпирические вероятности называются ранговыми переменными.

2. Замена ранговых переменных нормализованными переменными. Для этого по таблице значений нормальной нормированной интегральной функции распределения по известным значениям P_m

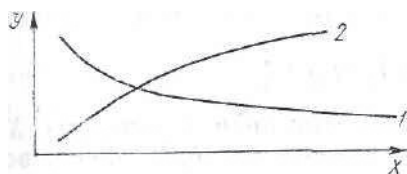


Рис. 7.11. Монотонно убывающая (1) и монотонно возрастающая (2) связь двух переменных величин

¹ Г.А. Алексеев предлагает формулу $p = \frac{(m - 0,25)}{(n + 0,5)100 \%}$, как более обоснован-

ную. Но, поскольку результаты расчетов по этой и формуле (3.90) мало различаются, в данном случае принимается последняя.

определяются нормированные ординаты U_m . Графическая иллюстрация первого и второго преобразований представлена на рис. 7.12.

Нормализованные переменные обладают следующим замечательными свойствами [2].

1. Нормализованные переменные $U(x_i)$ не изменяются (инварианты) при функциональном преобразовании исходных переменных x_i по любой возрастающей функции $f(x_i)$, например, $x_i - \ln x_i$, так как при этом не изменяются вероятности превышения $p(x_i) = p(\ln x_i)$ и, следовательно (см. рис. 7.11), не изменяются нормализованные переменные:

$$U_i = F[p(x_i)] = F[p(\ln x_i)].$$

Благодаря этому свойству при нормализации переменных X_j , т. е. в результате их двойного монотонного преобразования, все попарные криволинейные монотонные корреляционные связи преобразуются в попарно линейные между нормализованными переменными.

2. Для всех нормализованных переменных U_j , математическое ожидание $M[U_j]$ и среднее квадратическое отклонение σ_{U_j} представляют постоянные величины:

$$M[U_i] = 0, \sigma_{U_j} = 1, j = 1, 2, \dots, m. \quad (7.95)$$

Благодаря этому эмпирический коэффициент линейной корреляции $\hat{r}_{jk}$ между двумя нормализованными переменными U_j и U_k вычисляется по формуле:

$$\hat{r}_{jk} = \frac{\sum_{i=1}^n U_{ji} U_{ki}}{\left[(n-1) \hat{\sigma}_U^2 \right]}, \quad (7.96)$$

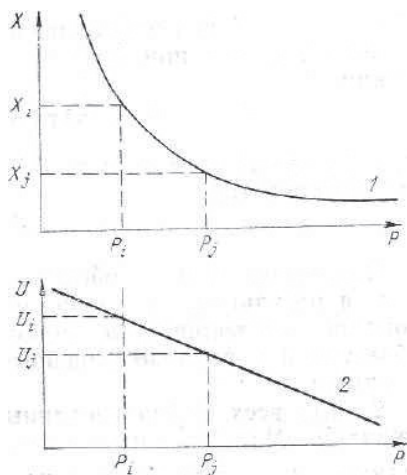


Рис. 7.12. Схема преобразования исходной переменной X в ранговую P и нормализованную U : 1 – кривая обеспеченности исходного ряда; 2 – кривая обеспеченности нормального закона распределения

где эмпирическая дисперсия любой нормализованной переменной

$$\sigma_U^2 = \left[\frac{1}{(n-1)} \right] \sum_{i=1}^n U_i^2 \quad (7.97)$$

зависит только от числа членов ряда и всегда меньше теоретической дисперсии $\sigma_U^2 = 1$, соответствующей генеральной совокупности.

Эмпирический коэффициент ковариации:

$$\widehat{\text{cov}}(U_j, V_k) = \left[\frac{1}{(n-1)} \right] \sum_{i=1}^n U_{ji} V_{ki} \quad (7.98)$$

также всегда отрицательно смещен, т. е. меньше по абсолютной величине $\text{cov}(U_j, V_k)$, соответствующей генеральной совокупности. Поэтому эмпирический коэффициент корреляции r_{jk} , определяемый как отношение двух одностронне смещенных величин $\widehat{\text{cov}}$ и σ , не имеет смещения из-за ограниченности объема выборки.

3. Коэффициент корреляции r_{jk} между нормализованными рядами U_j и V_k вполне характеризует тесноту связи при любом виде монотонной криволинейной корреляции между исходными переменными величинами X_j и X_k . Между тем как коэффициент корреляции непреобразованных рядов вполне характеризует тесноту связи только в случае линейной корреляции. Отсюда:

$$|r_U| \geq |r_x|, \quad (7.99)$$

причем знак равенства имеет место только при линейной корреляционной связи между рассматриваемыми исходными переменными.

4. При данном объеме выборок за совместный период наблюдений n эмпирические коэффициенты корреляции нормализованных рядов $\widehat{r_U}$ более устойчивы, чем эмпирические коэффициенты корреляции исходных рядов.

Действительно, в ряду исходных наблюдаемых величин $(x_{j1}, x_{k1}), (x_{j2}, x_{k2}), \dots, (x_{jn}, x_{kn})$ могут оказаться отдельные «выскакивающие» точки (x_{jl}, x_{kl}) , которые искажают величину $\widehat{r_x}$.

При этом, если «выскакивающая» точка располагается на одном из концов графика связи, т. е. x_{jl} и x_{kl} являются наибольшими или наименьшими членами в ранжированных рядах, то соответствующая нормализованная точка располагается непосредственно на линии нормализованной регрессии как точка с одинаковыми ранговыми номерами у обеих переменных, либо 1 или n . Это значит,

что крайние отскакивающие точки на концах кривых связей не искажают r_U .

Вместе с тем надо учитывать и недостатки названного метода. Прежде всего, к ним следует отнести некоторое сглаживание исходной информации за счет нормализации. Этот фактор может достаточно существенно сказаться в случае, если по одному из сопоставляемых рядов эмпирические точки на концевых участках сильно отклоняются от теоретической кривой обеспеченности.

7.3.11. Графические методы построения уравнения множественной корреляции

В экологических исследованиях часто возникает необходимость в графическом представлении взаимосвязи изучаемых процессов. Ранее (см. разд. 7.2.5) рассматривались методы построения графиков связи двух переменных величин (парная корреляция). При большем количестве переменных построение графиков значительно сложнее.

Обычно для этого используются три метода: метод контуров, метод остаточных отклонений, метод коаксиальной (соосной) корреляции.

Метод контуров

Метод заключается в построении графика связи трех переменных величин $Y = f(X_1, X_2)$. В поле координат Y и X наносятся точки (y, x_i) . Около каждой точки выписываются сами значения X_1 . Как те, так и другие показатели могут задаваться как в числовой, так и в нечисловой форме.

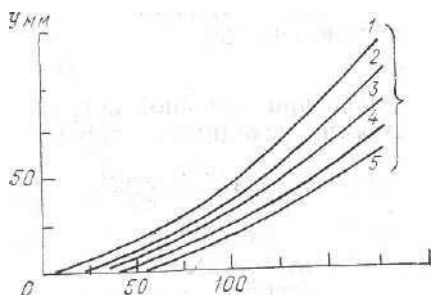


Рис. 7.13. Связь слоя ливневого стока Y со слоем осадков X_1 и увлажненностью почвы X_2

Если ряд наблюдений X_2 задан в нечисловой форме или в числовой, но количество его возможных значений m ограничено, т. е. $m \ll n$, где n — число наблюдений, то в поле графика проводятся линии связи для каждого отдельного числового или нечислового значения X_2 . Например, на рис. 7.13 представлен график зависимости слоя ливневого стока (Y) от

слоя осадков (X_1) и влажности почвы (X_2). В данном случае ряд X_2 , состоящий из показателей влажности, имеет всего 5 возможных значений, каждое из которых наблюдалось не менее 10–15 раз, поэтому для каждого значения оказалось возможным провести свою линию связи. В графике на рис. 7.13 ничего не изменится, если X_2 задать в нецифровой форме. Например, влажность почвы можно обозначить не в баллах, а соответственно через $a, б, в, г, д$ или просто задать ей словесную характеристику.

Несколько иначе производится построение графиков связи методом контуров если число возможных значений X_2 велико (X_2 — непрерывная величина) и точки взаимосвязи $Y = f(X_1)$ имеют большой разброс относительно линии связи. В этом случае значения X_2 разбиваются на интервалы, количество и размер которых определяется числом точек в поле графика. Чем больше число точек, тем больше можно взять интервалов. Интервалы $Y = f(X_1, X_2)$ и имеющимися данными о возможных значениях X_2 . Интервалы X_2 нумеруются и у каждой точки (y, x_1) выписывается номер соответствующего интервала X_2 . Затем по точкам, имеющим равные номера интервалов, проводятся линии связи $Y = f_j(X_1)$, своя для каждого j -го интервала. Число линий связи будет равно количеству интервалов X_2 . В результате получается семейство кривых $Y = f(X_1)$ для различных значений X_2 .

Значения X_2 семейства кривых в поле графика могут задаваться не только номером интервала, когда значение Y в данном интервале значений X_2 принимается независимым от X_1 и меняется только с изменением номера интервала, но, при большом количестве наблюдений, каждой линии связи может быть приписано среднее из значений X_2 в данном интервале. Это позволяет перейти к интерполяции значений Y по X_2 . Например, на представленной на рис. 7.14 зависимости $Y = f(X_1, X_2)$ по известному значению X_1 (суммарный слой максимального запаса воды в снежном покрове и осадков в период снеготаяния) и X_2 (показатель осеннего увлажнения бассейна) требуется определить слой весеннего стока Y . Пусть $X_1 = 150$ мм, $X_2 = 75$ мм, тогда Y (см. рис. 7.14) равно 88 мм.

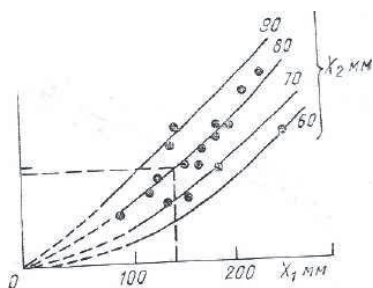


Рис. 7.14. Зависимость слоя весеннего стока Y от максимального запаса воды в снежном покрове X_1 и показателя осеннего увлажнения бассейна X_2

Метод остаточных отклонений

Теоретически число переменных не ограничено. В действительности при графических построениях обычно рассматривается не более 4–5 переменных.

Алгоритм расчетов [21] состоит из нескольких последовательных приближений. Предварительно ряды-предикторы, исходя из имеющихся представлений о формировании процесса, располагаются в порядке уменьшения их важности, и в соответствии с этим вводится индексация этих рядов $X_1, X_2, \dots, X_m$.

Первое приближение заключается в следующем. Строится график связи $Y = f(X_1)$, причем линия связи в поле точек (y, x_1) проводится приближенно и может быть как прямолинейной, так и криволинейной.

Затем по линии связи по каждому данному значению X_1 определяются значения $Y(X_1)$ и рассчитывается их отклонение от действительных значений $Y - \Delta_1 Y$ (т. е. отклонение точек связи (y, x_1) от линии связи). Далее строится график связи $\Delta_1 Y = f_1(X_2)$ и определяются отклонения точек $(\Delta_1 y, x_2)$ от линии связи $\Delta_2 Y$. Последние связываются с X_3 и т.д. На этом завершается первое приближение.

Второе приближение. Отклонение точек последнего графика от линии связи $\Delta_m Y$ принимается в качестве поправки к линии связи $Y = f_1(X_1)$, т. е. координаты каждой точки этого графика по оси ординат исправляют на $\Delta_m Y$. После исправления проводится новая линия связи $Y = f_2(X_1)$, определяется отклонение от нее $\Delta_1 Y$ и строится график зависимости $\Delta_1 Y = f_2(X_2)$ и т.д. На этом завершается второе приближение. Если оно оказалось эффективным, что определяется путем сравнения погрешностей расчета, то производится третье приближение и т.д.

Расчетные значения Y на каждом i -м приближении определяют по формуле:

$$\bar{Y}(X) = f_i(X_1) + f_i(X_2) + \dots + f_i(X_m), \quad (7.97)$$

где i – номер приближения.

В качестве примера рассмотрим построение связи $Y = f(X_1, X_2, X_3)$, где Y – максимальная влажность почвы, X_1 – потери при прокаливании, X_2 – глина, X_3 – ил (рис. 7.15).

В поле декартовых координат наносятся точки (y, x_1) и проводится линия связи $Y = f(X_1)$ (рис. 7.15.1.a). По полученной линии

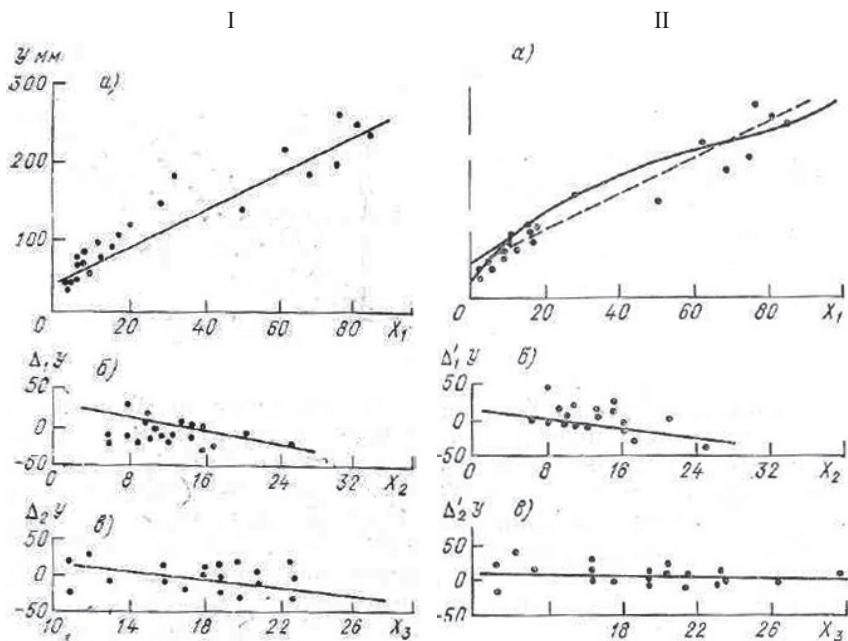


Рис. 7.15. Построение графиков связи методом остаточных отклонений:
 Y – максимальная влажность почвы; X_1 – потеря при прокаливании; X_2 – глина,
 X_3 – ил; I – первое приближение; II – второе приближение

связи определяются расчетные значения $Y - \bar{Y}(X_1)$ и разность Y и $\bar{Y}(X_1) - \Delta_1 Y$. Затем строится график зависимости $\Delta_1 Y = f(X_2)$ (рис. 7.15.1.б). По линии связи определяются расчетные значения $\Delta_1 Y - \bar{\Delta_1 Y}(X_2)$ и их разность $\Delta_2 Y$. По полученным значениям $\Delta_2 Y$ строится график зависимости $\Delta_2 Y = f(X_3)$ (см. рис. 7.15.1.в), определяется $\Delta_3 Y$ и по формуле (7.97) рассчитывается Y .

Для второго приближения расчеты производятся аналогично. Отличие заключается только в том, что вместо Y берутся исправленные значения $Y_1 = Y - \Delta_3 Y$. Графики второго приближения представлены на рис. 7.15.II.

Для отличия от первого приближения значения $\Delta_1 Y$ и $\Delta_2 Y$ приводятся здесь со штрихом, т. е. $\Delta'_1 Y, \Delta'_2 Y$.

Расчеты методом остаточных отклонений могут производиться и на основе построения графиков связей трех переменных величин.

В этом случае для построения каждого графика используется метод контуров.

Метод остаточных отклонений может быть поставлен в аналитической форме на ЭВМ [21].

Метод коаксиальной (соосной) корреляции

Метод заключается в представлении контурным способом на одном поле графика трех переменных величин и комбинированной корреляции этих величин с другими, причем в каждом случае графических построений используются общие оси.

Алгоритм расчетов, как и в предыдущем методе, состоит из нескольких последовательных приближений.

На первом приближении функции Y связывается с рядами X_1 и X_2 контурным способом (рис. 7.16), т. е. путем построения семейства зависимостей $Y = f(X_1)$ для различных X_2 . Затем по этой

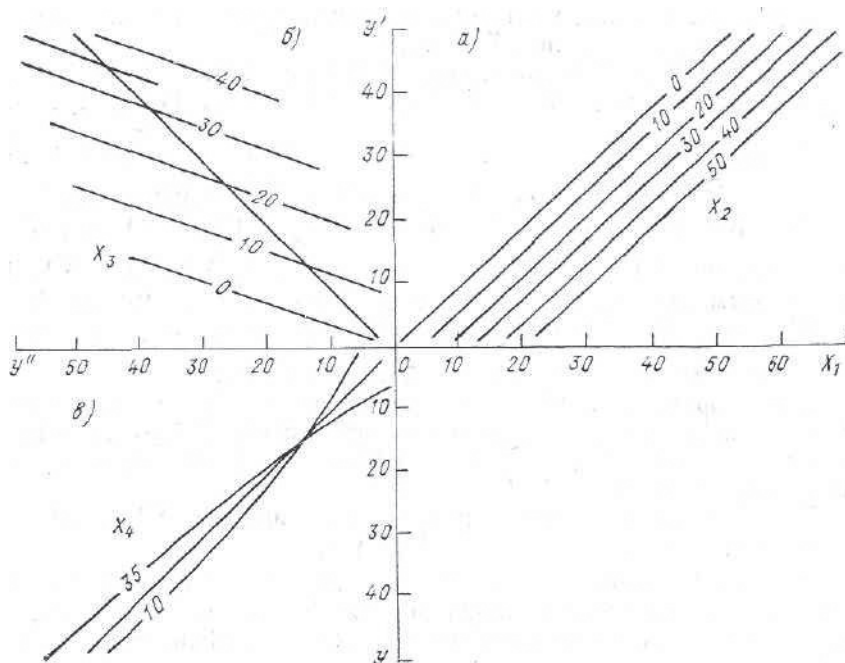


Рис. 7.16. Графики связи методом соосной корреляции.

а) $Y' = f(X_1, X_2)$, б) $Y'' = f(Y', X_3)$, в) $Y = f(Y'', X_4)$

зависимости определяются расчетные значения $\bar{Y}(X_1, X_2)$ – обозначим Y_{1p} , которые связываются на новом графике со значениями Y . Точки этого графика помечаются значениями X_3 и строится второе семейство кривых $Y = f(Y_{1p})$ для различных значений X_3 . По этому семейству кривых определяется $Y_{2p} = f(Y_{1p}, X_3)$, которое на новом графике связывается со значениями Y . Точки этого графика помечаются значениями X_4 и т.д. Такие построения продолжаются до тех пор, пока в расчет не будут введены все переменные, число которых обычно не превышает 5–6. Второе приближение $Y = f(X_1, X_2)$ строится на основании анализа связи Y с Y_{m-1} и т.д.

Использование графической корреляции

Графические методы корреляции широко применялись в расчетах и прогнозах еще до внедрения аналитических методов. Это было связано со следующими обстоятельствами:

- а) графические методы не требуют большого числа ручных расчетов, необходимых для аналитических методов;
- б) для применения графических методов можно лишь немного знать или вовсе не знать статистические методы;
- в) графическая корреляция наглядна и поэтому легче понимается;
- г) графические методы могут применяться для криволинейных связей, тогда как аналитические методы предполагают задание вида кривой.

Большинство из этих преимуществ (а, б, в) методов графической корреляции в настоящее время перестали быть существенными, что связано с повышением общего математического уровня специалистов и широким внедрением электронных вычислительных машин. Зато более существенными стали недостатки методов графической корреляции, заключающиеся в следующем:

- а) графическая корреляция не учитывает корреляцию между предикторами, что может привести к ложной корреляции и получению неустойчивых результатов. При получении результатов путем графической корреляции, особенно при большом количестве предикторов, возможны крупные просчеты и необоснованные выводы;
- б) не позволяет оценить значимость отдельных предикторов;
- в) не дает возможности достаточно полной оценки полученных результатов.

Между тем, в случае линейной связи, точность результатов, полученных графическими методами, в лучшем случае, приближается к точности результатов, полученных аналитическими методами, а чаще всего, значительно ниже. Зато в случае криволинейной связи, если она не носит монотонно возрастающий или монотонно убывающий характер, метод графических построений часто остается единственным методом анализа. Но и в этом случае необходимо иметь в виду, что точность результатов из-за ограниченности информации ($n \leq 60 \dots 100$), очень низкая, что, к сожалению, не учитывается многими исследователями.

Глава 8. Численные методы анализа колебаний геоэкологических процессов во времени и пространстве

В предыдущих разделах учебника рассматривались методы обработки и анализа геоэкологических данных, основанные на теории случайной величины. Эти методы получили наибольшее распространение в практике исследований. Вместе с тем все в большей степени начинают проявляться некоторые ограничения этих методов. Поэтому постепенно круг возможных математических моделей геоэкологических процессов стал расширяться за счет использования аппарата теории случайных функции.

8.1. Математические модели колебания геоэкологических процессов во времени и пространстве

В географических исследованиях большое внимание уделяется анализу изменений рассматриваемых характеристик окружающей среды и факторов их формирования во времени и пространстве. Назначение этого анализа выявить основные свойства исследуемых процессов – изменчивость во времени и пространстве, характеристики периодических и непериодических колебаний, преемственность в развитии – и предсказать их поведение в будущем.

Известно, что колебаниям природных процессов присуща тенденция к образованию группировок с одноименными значениями исследуемой характеристики. В некоторых случаях такие колебания имеют почти периодический характер, в частности колебания, определяемые сменой времен года (средние месячные и средние квартальные значения стока и др.), в других случаях явной периодичности не прослеживается, но группирование одноименных значений исследуемого процесса проявляется достаточно наглядно (например, значения солнечной активности в индексах Вольфа).

Часто, тенденцию к образованию группировок по тому или иному признаку принято называть цикличностью.

По мнению многих исследователей, это определение не достаточно информативно. Так, если взять чисто случайный ряд, т. е. ряд, в котором внутрирядная связь отсутствует вообще, то и в этом случае наблюдаются группировки одноименных величин. Вероятность группировок различной продолжительности p_T в этом случае может быть рассчитана по формуле умножения вероятностей:

$$p_T = p^T \alpha^2, \quad (8.1)$$

где p – вероятность значений исследуемой величины, меньших заданного значения; T – продолжительность группировок; $\alpha = 1 - p$.

Однако прогностической информации эти группировки не имеют и какого-либо влияния на результаты исследования, основанные на теории случайной величины, не оказывают. Поэтому, исходя из чисто практических соображений, тенденцию к образованию группировок следует учитывать только в том случае, если группировки рассматриваемого ряда отличаются от группировок чисто случайной величины.

По-видимому, более точным, определяющим суть вопроса, является следующее определение, данное О.А. Дроздовым [14]:

«Циклические колебания или цикличность – колебания связанных величин различной степени регулярности при условии существования математических ожиданий параметров этих колебаний.

Можно привести и другое достаточно полное определение, часто используемое в метеорологии:

Циклические колебания или цикличность – колебания значений исследуемого процесса, спектральная плотность которых существенно отличаются от спектральной плотности колебаний значений случайной величины.

В зависимости от принятых представлений о характере колебаний (цикличности) процессов природы во времени или пространстве существуют различные подходы к их описанию.

Так, в 30-х годах, начиная с работ Д.Л. Соколовского [41], С.Н. Крицкого и М.Ф. Менкеля [23], широкое распространение получило представление о том, что колебания многих характеристик режима рек, в частности среднего годового и максимального стока, являются «случайными». В данном случае это означало, что каждое последующее значение не зависит от предшествующих. Это представление для своего времени было прогрессивным и по

имеющимся немногочисленным материалам наблюдений достаточно обоснованным. Оно позволило использовать мощный аппарат математической статистики на уровне теории случайной величины. Все водохозяйственные расчеты 30–50-х годов основывались именно на использовании этого аппарата. Ничего равноценного другие представления о процессах стока, в частности об их периодичности или динамической обусловленности, не дали.

Постепенно, по мере накопления материалов, выявился ряд новых фактов, не объяснимых с позиций теории чисто случайной величины. В частности, трудно объяснимым с этих позиций было затяжное маловодье тридцатых годов в бассейне р. Волги. Не подходила теория случайной величины и к описанию колебания стока естественно зарегулированных рек, таких как Нева, Ангара и т.д. В связи с этим появилась новая гипотеза о характере колебаний стока во времени. Впервые ее предложил в 1936 г. П.А. Ефимович [16]. Впоследствии в 1946 г. С.Н. Крицкий и М.Ф. Менкель [22] развили эту гипотезу до уровня практического применения. Она заключалась в том, что на случайные колебания процессов стока накладывается под влиянием внешних факторов динамическая составляющая, вызывающая некоторую зависимость будущего значения годового стока от предшествующего. В результате между смежными значениями стока появляется связь, которая может быть определена с помощью коэффициента корреляции r_1 . На этой основе для математического описания стока была предложена модель в виде простой цепи Маркова, которая, начиная с 60-х годов, широко использовалась в водохозяйственных расчетах.

В начале 50-х годов выяснилось, что математические модели в виде «чисто случайной величины» и простой цепи Маркова не подходят для описания многих процессов, в частности, для описания процессов внутригодового распределения стока. В 1956 г. для описания этих процессов Н.А. Картвелишвили была предложена математическая модель в виде случайной функции [19].

В 1961 г. Ю.М. Алехиным было впервые установлено, что внутригодичная связь последовательности значений годового стока и ряда других процессов распространяется не только на смежные значения [4]. В результате выполненных им расчетов выяснилось, что значения годового стока на многих реках скоррелированы на достаточно большом временном интервале, превышающем двадцать-тридцать лет. В качестве теоретического обоснования широкой скоррелированности Ю.М. Алехиным была выдвинута гипотеза о том, что

развитие природных макропроцессов, в частности процессов среднего годового стока, подчиняется динамико-статистическим закономерностям: т. е. будущее природных процессов в какой-то степени заложено в их прошлом и в какой-то степени определяется предшествующими состояниями.

Обнаружение ближних и дальних внутрирядных связей в некоторых процессах привело к внедрению в гидрологических математических понятий, связанных с теорией случайных функций и сложных цепей Маркова.

Таким образом, для описания изменения процессов окружающей среды во времени используется ряд математических моделей. Далее представлены наиболее распространенные из них.

1) Математическая модель исходного ряда в виде последовательности некоррелированных значений случайной величины. Эта последовательность значений часто называется случайным рядом, а сама случайная величина – чисто случайной величиной.

2) Математическая модель в виде последовательности значений простой цепи Маркова. В этом случае исследуемый процесс представляется в качестве случайного процесса без последствия, в котором каждое последующее значение зависит только от предыдущего и не зависит от предшествующих значений.

3) Математическая модель в виде последовательности значений сложной цепи Маркова. В этом случае принимается, что каждое последующее значение в какой-то степени зависит от m предшествующих.

Вторая и третья математические модели являются достаточно широкими представителями более общей математической модели по схеме случайной функции.

В настоящее время, как правило, принимается, что процессы природы являются стационарными и для описания их применяются так называемые авторегрессионные модели (АР). В последнее время все чаще приходится считаться с фактом нестационарности геоэкологических процессов. Для их анализа широкое распространение в СССР и за рубежом получили стохастические модели авторегрессии – проинтегрированного скользящего среднего (АРПСС) [7].

Выбор той или иной модели для описания природных процессов в основном определяется известными свойствами исследуемых процессов, а также зависит от целей исследований и взглядов исследователя. Выше, в гл. 3 достаточно детально рассматривались вопросы, связанные с применением теории вероятностей и

математической статистики на уровне понятий случайной величины. Далее приводятся основные понятия, связанные с теорией случайных функций.

8.2. Случайные функции и их характеристики

8.2.1. Определение случайной функции

Случайная функция – функция, которая в результате опыта может принять тот или иной конкретный вид; неизвестно заранее, какой именно.

Случайная функция обычно обозначается большими буквами с указанием аргумента $X(s)$, $Y(f)$ и т.д., где s, f и т.д. – аргументы случайной функции.

Реализация случайной функции – конкретный вид, принимаемый случайной функцией в результате опыта.

При каждом повторении опыта получается новая реализация. Таким образом, случайную функцию можно рассматривать как множество всех ее реализаций. Реализацию случайной функции обычно обозначают малыми буквами $x_1(t)$, $x_2(t)$, ..., $x_n(t)$ с опущенным индексом, указывающим номер опыта, в котором данная реализация получена.

Если аргументом случайной функции является время, то она обычно называется вероятностным или случайным процессом и обозначается $X(t)$, $Y(t)$ и т.д.

Так, например, имеются многолетние данные по ежедневным расходам воды какой-либо реки с четко выраженным половодьем. В каждый i -й год получены нормированные гидрографы суточных расходов воды $Q_i(t)$, описывающие их изменения в течение года. Эти гидрографы (рис. 8.1) при сохранении общего вида будут довольно существенно отличаться друг от друга. Каждый гидрограф $Q_i(t)$, полученный в i -м году, (при i -м опыте) можно рассматривать как отдельную реализацию случайного процесса, а множество имеющихся гидрографов – как совокупность или ансамбль реализации случайного процесса колебаний стока в разные годы.

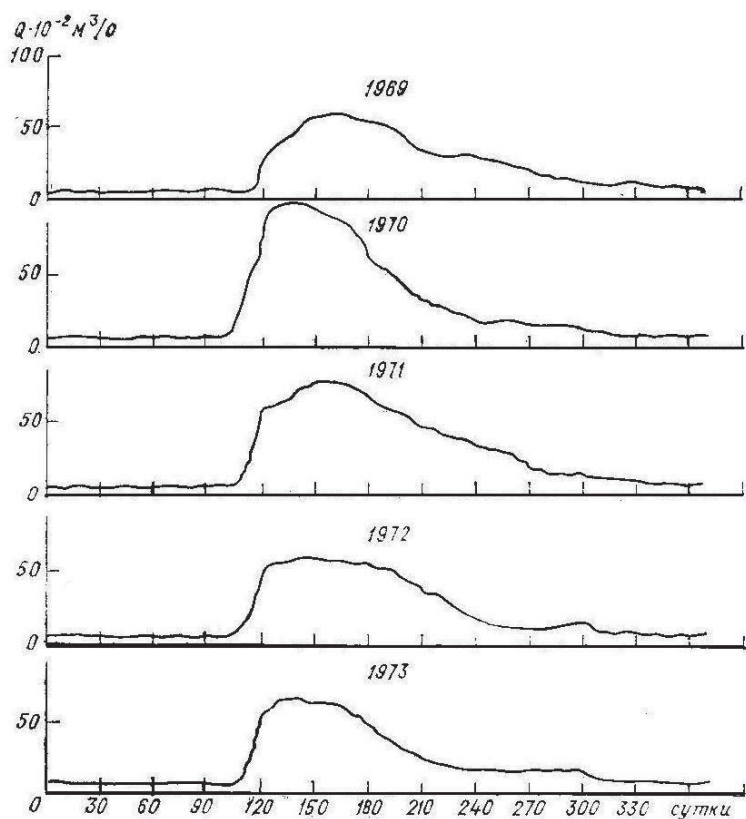


Рис. 8.1. Ансамбль реализаций суточных гидрографов стока за отдельные годы (р. Иртыш, г. Тобольск)

Аналогично реализациями случайного процесса колебаний стока, например, в каком-либо замкнутом районе, обособленном по выбранному признаку, могут служить нормированные ряды стока отдельных рек. Так, на рис. 8.2 представлен ансамбль гидрографов стока за многолетний период в бассейне р. Камы. Каждый гидрограф здесь, как отмечалось выше, представляет собой одну из реализаций случайной функции.

Если зафиксировать значения аргумента случайной функции, например $t = t_0$, и провести прямую, перпендикулярную оси абсцисс, то она пересечет каждую реализацию в одной точке.

Сечение случайной функции – совокупность значений множества реализаций при фиксированных значениях аргумента.

Сечения случайной функции, отвечающие значению аргумента $t_0, t_1, \dots$, будем обозначать соответственно $X(t_0), X(t_1), \dots$

Аргумент случайной функции может принимать либо любые вещественные значения в заданном интервале конечном или бесконечном, либо только определенные дискретные значения. В первом случае $X(t)$ называется **случайным процессом**, во втором – **случайной последовательностью** или случайным процессом с дискретным временем.

Случайная функция может зависеть от нескольких аргументов.

Случайная функция нескольких аргументов называется случайным полем.

Случайное поле может быть скалярным (поле температур воздуха, поле давления, карта модулей стока) или векторным (поле ветра, поле скорости течения).

Обычно гидрометеорологические процессы являются функциями непрерывных аргументов, однако, чаще всего, характеристики этих процессов даются либо за определенные интервалы осреднения, либо через определенный период времени.

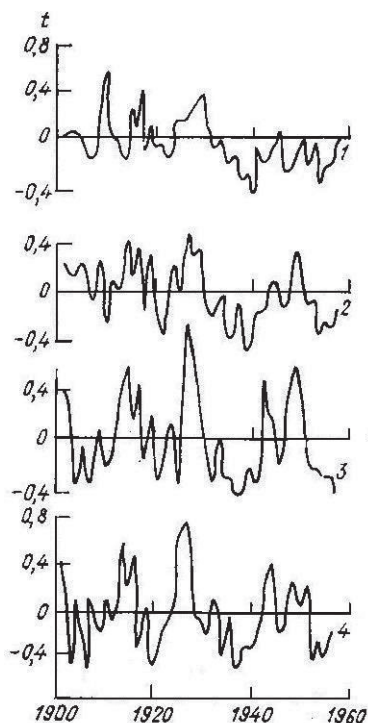


Рис. 8.2. Ансамбль гидрографов нормированного стока рек в бассейне р. Камы: 1 – р. Вишера, с. Митраково; 2 – р. Вятка, г. Киров; 3 – р. Белая, г. Уфа; 4 – р. Чусовая, г. Нижние Шалыги

8.2.2. Законы распределения случайной функции

Фиксируя значения аргумента $t_1, t_2, \dots, t_n$, можно получить n -сечений случайного процесса $X(t) : X_1 = X(t_1), X_2 = X(t_2), \dots, X_n = X(t_n)$, каждое из которых представляет собой случайную величину.

Совокупность сечений случайных функций представляет собой систему случайных величин. Следовательно, случайный процесс

можно охарактеризовать функцией распределения системы случайных величин:

$$F(x_1, x_2, \dots, x_n) = p(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n). \quad (8.2)$$

Очевидно, что функция распределения тем полнее будет характеризовать случайный процесс, чем ближе друг к другу расположены значения аргумента t_i и чем большее число их взято.

Важной характеристикой случайного процесса является **одномерная случайная функция** распределения случайной величины в каждом сечении t_i :

$$F_{1i}(x) = F(x, t_i) = p(X_i < x_i). \quad (8.3)$$

Эта функция в общем случае не описывает всех вероятностных свойств случайного процесса, так как, например, остается неясным как связаны между собой сечения процесса, допустим, $X(t_i)$, $X(t_j)$, $X(t_k)$ и т.д.

Более полную характеристику случайного процесса дает двухмерная функция распределения, т. е.

$$F_{2i}(x_i) = F_2(x_i, x_{i+1}; t_i, t_{i+1}) = p(X_i < x_i; X_{i+1} < x_{i+1}) \quad (8.4)$$

и т.д.

Случайный процесс $X(t)$ считается заданным, если для каждого значения t определена одномерная функция распределения случайной величины (8.2), для каждой пары значений t_i и t_{i+1} аргумента t определена двухмерная функция распределения (8.4) и для любых n значений аргумента t определена n -мерная функция распределения системы случайных величин:

$$\begin{aligned} F_n(x) &= F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \\ &= p(X_1 < x_1; X_2 < x_2, \dots, X_n < x_n). \end{aligned} \quad (8.5)$$

Как отмечалось выше, случайная функция может рассматриваться как система случайных величин. В то же время известно, что из функций распределения системы случайных величин может быть получена функция распределения любой ее подсистемы. Аналогично, если известна n -мерная функция распределения или плотность распределения случайного процесса, то тем самым заданы и все функции и плотности распределения более низкого порядка.

В практических приложениях характеристика случайной функции на основе многомерных законов распределения обычно не

используется как вследствие сложности экспериментального определения многомерных законов распределения, так и вследствие громоздкости алгоритмов расчета при решении прикладных задач. Поэтому в большинстве случаев ограничиваются заданием отдельных характеристик этих законов, аналогично тому, как в теории случайных величин вместо законов распределения часто используются их числовые характеристики. Кроме того, числовые характеристики необходимы для задания законов распределения и их сочетаний в числовой форме.

8.2.3. Числовые характеристики случайных функций

Числовые характеристики случайных функций в отличие от числовых характеристик случайных величин, представляющих собой определенные числа, в общем случае представляются в виде функций.

Исходя из определения случайной функции в виде последовательности случайных величин, числовые характеристики случайных функций могут быть представлены в виде последовательности значений числовых характеристик случайных величин, допустим, по аргументу t .

Математическое ожидание случайной функции – неслучайная функция $m_x(t)$, которая при каждом значении аргумента t равна математическому ожиданию от соответствующего сечения случайной функции:

$$m_x(t) = m[x(t)] = \int_{-\infty}^{\infty} x(t)p(x,t)dx, \quad (8.6)$$

где $p(x,t)$ – одномерная плотность распределения рассматриваемого случайного процесса.

Если сечения представлены в виде дискретной случайной величины, то

$$m_x(t) = \sum_{j=1}^n \zeta_j(t)p_j(t), \quad (8.7)$$

где $\zeta_j(t)$ – j -ое значение дискретной случайной величины в t -том сечении, $p_j(t)$ – вероятность значения ζ_j в данном сечении.

Дисперсия случайной функции – неслучайная функция $D_x(t)$, значение которой для каждого t равно дисперсии соответствующего сечения случайной функции:

$$D_x(t) = M[X(t)^2] = \int_{-\infty}^{\infty} [X(t) - m_x(t)]^2 p(x, t) dx, \quad (8.8)$$

или для дискретного процесса:

$$D_x(t) = \sum_{j=1}^n [\zeta_j(t) - m_x(t)]^2 p_j(t). \quad (8.9)$$

Среднее квадратическое отклонение случайной функции:

$$\sigma_x(t) = \sqrt{D_x(t)}; \quad (8.10)$$

коэффициент вариации:

$$C_{v_x}(t) = \frac{\sigma_x(t)}{m_x(t)}. \quad (8.11)$$

Характеристики $m_x(t)$, $D_x(t)$, $\sigma_x(t)$ и $C_{v_x}(t)$ отражают статические, внешние свойства рассматриваемых процессов. Однако, обладая одними и теми же характеристиками положения и рассеяния, случайные процессы могут существенно отличаться друг от друга по характеру колебаний во времени или по своим динамическим свойствам (рис. 8.3). Представление о динамических свойствах процесса дает его корреляционная, или, как часто ее называют, автокорреляционная функция.

Эта характеристика является новой, не встречавшейся ранее при изучении случайных величин. Она показывает степень

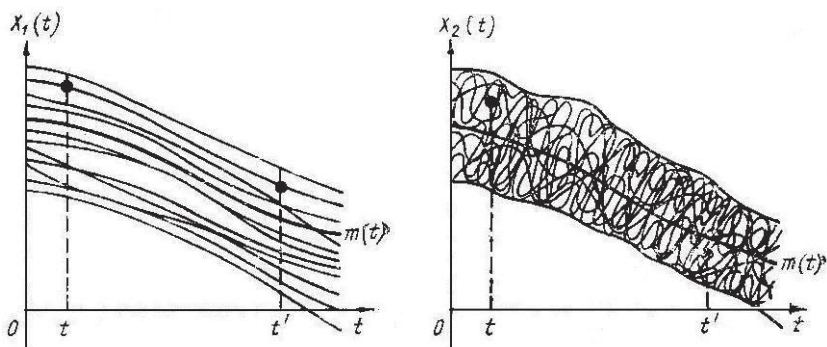


Рис. 8.3. Реализация случайных процессов $X_1(t)$ и $X_2(t)$

зависимости между значениями случайной функции, относящимися к различным сечениям.

Корреляционная, или автокорреляционная, функция $K_x(t, t')$ – неслучайная функция двух аргументов, которая при каждой паре значений t, t' равна коэффициенту корреляции соответствующих сечений случайной функции:

$$K_x(t, t') = M \left[[X(t) - m_x(t)] [X(t') - m_x(t')] \right]. \quad (8.12)$$

При $t = t'$ корреляционная функция обращается в дисперсию:

$$K_x(t, t) = M \left[X(t) - m_x(t) \right]^2 = D_x(t).$$

В практике чаще используется относительный показатель внутренней взаимосвязи, называемый **нормированной корреляционной функцией**:

$$r_x(t, t') = \frac{K_x(t, t')}{\left[\sigma_x(t) \sigma_x(t') \right]}. \quad (8.13)$$

В дальнейшем для краткости, если это не оговорено иначе, нормированную корреляционную функцию будем называть просто корреляционной функцией.

8.2.4. Классификация математических моделей случайных процессов

Случайные функции или случайные процессы – очень широкое, емкое понятие, включающее в себя весь континуум процессов между строгой периодичностью, или определенностью последующих значений процесса по отношению к предшествующим, и полной случайностью или независимостью последующих значений от предшествующих, полной устойчивостью характеристик во времени и функциональным или случайным изменением их с изменением аргумента.

Выбор того или иного вида случайных процессов в качестве математической модели определяется соответствием его признаков (граничных условий) признакам изучаемого процесса. В качестве этих признаков, по которым обычно и классифицируются случайные процессы, могут выступать следующие.

1) Зависимость характеристик случайных процессов от начала отсчета аргумента. По этому признаку случайные процессы делятся на **стационарные и нестационарные**.

2) Наличие или отсутствие связей между характеристиками по аргументу и характеристиками по множеству реализаций. По этому признаку стационарные случайные процессы делятся на *эргодические и неэргодические*.

3) Закон распределения значений случайной функции в каждом сечении. Обычно рассматриваются гауссовские (нормальные), пуассоновские, Пирсона и другие случайные функции.

4) Вид корреляционных функций. По этому признаку существует ряд классификаций. В науках о Земле чаще всего используется классификация по максимальному сдвигу между связанными сечениями случайной функции.

Рассмотрим различные типы математических моделей по этим признакам.

Стационарные и нестационарные случайные процессы.

Стационарность в широком и узком смысле

По зависимости характеристик от начала отсчета аргумента случайные процессы делятся на стационарные и нестационарные.

В свою очередь стационарные процессы могут быть строго стационарными или стационарными в узком смысле и стационарными в широком смысле.

***Строго стационарные случайные процессы** – процессы в которых все конечномерные законы распределения $F_1(x)$, $F_2(x)$, ..., $F_n(x)$ не зависят от аргумента t , а зависят только от разности значений t , т. е.*

$$\begin{aligned} F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \\ = F_n(x_1, x_2, \dots, x_n; t_1 + t_0, t_2 + t_0, \dots, t_n + t_0). \end{aligned} \quad (8.14)$$

То есть, функции распределения строго стационарных случайных процессов инвариантны по отношению к сдвигу начала отсчета аргумента t_0 . В этом случае одномерная плотность распределения не зависит от t , она одна и та же для всех сечений. Действительно, полагая $t_0 = -t$, из выражения (8.2) получаем:

$$F_1(x; t) = F_1(x; t - t_0) = F_1(x; 0) = F_1(x). \quad (8.15)$$

Двухмерная плотность распределения зависит не от двух аргументов t_1 и t_2 , а только от одного аргумента – их разности $\tau = t_2 - t_1$. Так, при $t_0 = -t$ из формулы (8.3) следует, что

$$\begin{aligned} F_2(x_1, x_2; t_1, t_2) &= F_2(x_1, x_2; 0, t_1 - t_1) = \\ &= F_2(x_1, x_2; t_2 - t_1) = F_2(x_1, x_2; \tau). \end{aligned} \quad (8.16)$$

И, согласно выражениям (8.6) и (8.12),

$$m_x(t) = \int_{-\infty}^{\infty} xp(x)dx = m_x = \text{const}; \quad (8.17)$$

$$K_x(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_x)(x_2 - m_x)p(x_1, x_2; \tau) = K_x(\tau). \quad (8.18)$$

Таким образом, следствием строгой стационарности является постоянство во времени математического ожидания и дисперсии и зависимость корреляционной функции только от одного аргумента τ .

Эти следствия строгой стационарности являются достаточными условиями нестрогой стационарности или стационарности в широком смысле [11,19].

При прочих равных условиях для расчета числовых характеристик стационарных функций при той же точности требуется меньший объем информации (в данном случае длина реализаций), а сам расчет намного проще. Поэтому многие теоретические исследования последнего времени исходят из аппроксимации рассматриваемых процессов в качестве стационарных. В некоторых случаях это может быть обосновано относительностью понятия стационарный процесс, которая заключается в том, что в отдельные ограниченные периоды времени случайный процесс общего типа обладает всеми признаками стационарного.

Эргодические и неэргодические стационарные случайные процессы

По наличию или отсутствию связи между средними значениями процессов по аргументу (по одной реализации) и средними по множеству реализаций случайные процессы делятся на *эргодические и неэргодические*.

В большинстве случаев в гидрометеорологических и геоэкологических исследованиях принимается, что рассматриваемые процессы обладают эргодическим свойством. Это свойство состоит в том, что каждая отдельная реализация случайной функции является как бы полномочным представителем всей совокупности возможных

реализаций; одна реализация достаточной продолжительности может заменить при обработке множество реализаций той же общей продолжительности. Если случайная функция обладает свойством эргодичности, то все ее характеристики можно приближенно определить по одной достаточно длинной реализации, и наоборот, все характеристики случайной функции за длительный период можно определить по множеству наблюдений за сравнительно короткий период.

Свойство эргодичности представляет особый интерес, так как на практике большинство исследований выполняются или по одной, или по крайне ограниченному числу реализаций. Между тем результаты исследования желательно, а часто и необходимо, распространить на весь исследуемый процесс в целом. Свойство эргодичности позволяет вполне обоснованно подойти к таким обобщениям. Именно поэтому в гидрометеорологии стараются или доказать свойство эргодичности, или использовать его за счет каких-либо приближений. В частности, широко известны исследования Г.П. Калинина по глобальной геоэкологии, основанные на понятии «условно-эргодичности» [18].

Известно, что ряд случайных процессов не обладает эргодическим свойством. В частности, неэргодичность случайного процесса может быть вызвана наличием в его составе слагаемого в виде обычной случайной величины.

Достаточным условием эргодичности стационарной случайной функции является приближение ее корреляционной функции к нулю при возрастании промежутка τ между сечениями, т. е. $r(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$. Более строгое определение эргодичности стационарной случайной функции с дискретным временем дано в специальных работах [7, 11, 36].

Простые и сложные цепи Маркова

В ряде прикладных дисциплин случайные процессы разделяются по величине максимального сдвига или радиуса корреляции $\tau_{\max}$ между связанными сечениями случайного процесса на процессы без последствия $\tau_{\max} = 1$ и с последствием $\tau_{\max} > 1$. В дальнейшем для учета $\tau_{\max}$ в основном используются понятия простой и сложной цепи Маркова.

Простая цепь Маркова – случайная последовательность, в которой для каждого момента времени вероятность любого состояния системы в будущем зависит только от состояния системы

в настоящий момент (t_0) и не зависит от того, каким образом система пришла в это состояние.

В такой последовательности имеется автокорреляция лишь между смежными членами ряда.

Часто процесс, описываемый простой цепью Маркова, называют марковским случайным процессом с дискретным временем. В настоящее время математическая модель по схеме простой цепи Маркова нашла широкое применение в описании многолетних колебаний речного стока при расчете многолетнего регулирования.

Сложная цепь Маркова – случайная последовательность $X(t)$, в которой для каждого момента времени t вероятность любого состояния системы в будущем зависит от состояния системы в предшествующие моменты времени $t_i (i = 1, 2, \dots)$.

8.3. Оценка общего характера колебаний геоэкологических процессов

Возможность использования той или иной математической модели для описания исходных данных определяются, рядом признаков, среди которых главными являются однородность, стационарность и характер внутрирядных связей.

Как отмечалось выше, случайные процессы делятся на стационарные (в широком и узком смысле) и нестационарные. В рамках корреляционной теории условиями стационарности в широком смысле являются постоянство математического ожидания и дисперсии во времени и зависимость корреляционной функции только от разности аргументов $\tau = t_2 - t_1$.

Если внутрирядные связи отсутствуют, то для описания процесса используется математическая модель в виде последовательности значений чисто случайной величины. В гидрометеорологии принято называть такие ряды случайными. Существует большое количество тестов для проверки на «случайность», некоторые из которых рассматриваются ниже. Если ряд не является «случайным», то он характеризуется наличием ближних или ближних и дальних внутрирядных связей. Характер внутрирядных связей определяется непосредственно по корреляционной функции и связывается с вопросами точности и достоверности ее значений, рассчитанных

по коротким выборкам. Вопрос о характере внутрирядных связей является одним из наиболее сложных, и решение его во многих случаях носит дискуссионный характер.

8.3.1. Анализ стационарности, однородности и наличия тренда

Стационарность процессов (характеристика процессов) и однородность (часто отражает качество наблюдений) имеющегося ряда наблюдений контролируются одинаковыми статистическими показателями, и различие их может констатироваться во многом только на основе физического анализа.

Для оценки стационарности и однородности рядов наблюдений по среднему и дисперсии используются рассмотренные статистические критерии Стьюдента и Фишера, при общей оценке однородности – критерии Уилкоксона и Манна–Уитни (см. гл. 6). В последнее время для этого применяются также критерии Диксона и Смирнова–Граббса.

В некоторых случаях нестационарность процесса может определяться наличием тренда.

Тренд – постепенное направленное изменение случайной переменной X в течение анализируемого периода.

Иногда тренд является частью колебаний с периодами, длительность которых сравнима с продолжительностью периода наблюдений. В некоторых случаях: продолжительность такого тренда может превышать имеющуюся продолжительность наблюдений.

Для оценки наличия тренда часто используется коэффициент корреляции рассматриваемого ряда X со временем или порядковым номером членов ряда или критерий Спирмэна.

Оценка наличия тренда по коэффициенту корреляции

Для оценки тренда используется формула коэффициента корреляции значений ряда X с ранговым номером членов ряда I :

$$\hat{r}_{xi} = \frac{\sum_{i=1}^n (x_i - \bar{x})(i - \bar{i})}{\left[(n-1) \hat{\sigma}_x \sigma_i \right]}, \quad (8.19)$$

где $\bar{i} = \frac{(n+1)}{2}$; $\hat{\sigma}_x$ и σ_i – средние квадратические отклонения ряда X и ряда порядковых номеров i .

Если значение r_{xi} окажется значимым (см. разд. 7.3), то гипотеза о наличии тренда не опровергается, и если r_{xi} на много больше $r_{кр}$, то скорее всего, в развитии процесса имеется тренд.

Оценка тренда по критерию Спирмэна

Одним из наиболее распространенных при исследовании зависимости процесса от времени является ранговый критерий Спирмэна:

$$\hat{\rho} = \frac{1 - 6 \sum_{i=1}^n d_i^2}{(n^3 - n)}, \quad (8.20)$$

где d_i – разность между порядковым номером и рангом каждого значения ряда длиной n . Если связь отсутствует, то $\rho = 0$; при прямой связи, т. е. при постепенном нарастании X , $\rho = 1$; при обратной связи, т. е. постепенном уменьшении X , $\rho = -1$. Математическое ожидание ρ при отсутствии тренда $M[\rho] = 0$, дисперсия $D_\rho = \frac{1}{(n-1)}$.

Определение даты нарушения однородности

Во многих случаях нам известна та точка отсчета (дата), начиная с которой по той или иной причине произошло нарушение однородности ряда. Такими точками отсчета могут быть: дата начала заполнения водохранилища, дата начала работы крупных водозаборных сооружений и т.д. Однако в некоторых случаях эта дата неизвестна. Для ее установления в ряде случаев может использоваться график связи нарастающих сумм членов исследуемого ряда с порядковым номером или с нарастающими суммами другого ряда, относительно которого известно, что он является однородным.

Разностные интегральные кривые получили широкое распространение с тридцатых годов прошлого столетия, когда они активно начали использоваться в водохозяйственных расчетах для расчетов многолетнего регулирования стока. С выходом переводной книги «Прикладная геоэкология» [27] широкое распространение в целях анализа однородности получили интегральные кривые [50].

Интегральные кривые представляют собой графическое отображение зависимости:

$$U_j = \sum_{i=1}^j x_i, \quad j = 1, 2, \dots, n, \quad (8.21)$$

где U_j – j -тое значение ряда U , i – порядковый номер значений ряда X ; j – порядковый номер нарастающих сумм U_j .

Как следует из формулы (8.21):

$$U_j = U_{j-1} + x_j.$$

Зная U_j , можно на каждый данный момент j определить среднее значение ряда X от 1-ого до j -ого члена ряда включительно:

$$\bar{x} = \frac{U_j}{j}. \quad (8.22)$$

Из определения интегральной кривой вытекают ее основные свойства:

- каждая ордината кривой зависимости (8.21) характеризует объем стока через данный створ от начала наблюдений (или принятого начала) до j -ого момента времени;
- разность ординат интегральной кривой при различных значениях аргумента j характеризует объем стока за интервал времени между этими значениями;
- при постоянном значении рассматриваемого элемента ($X = \text{const}$) интегральная кривая представляет собой прямую линию;
- с увеличением j и возрастанием суммы U_j роль отдельных значений ряда X в сумме U_j уменьшается и, поэтому, отклонения отдельных точек от линии связи на графике уменьшаются, становятся все менее существенными. Поэтому (см. ниже) при $j \geq 5$ или 6 в крайнем случае 10, при однородности ряда наблюдений практически все точки графика зависимости (8.21) ложатся на линию связи;
- вместе с тем, при наличии какой-либо тенденции в изменении рассматриваемых процессов, то есть накапливании отклонений в ту или иную сторону, происходит изменение направления линии связи.

Именно исходя из этих свойств, в настоящее время интегральные кривые используются не только при расчетах регулирования стока, но и для выявления изменений тенденции развития процесса и времени начала этих изменений, то есть для определения переломных точек в многолетних колебаниях стока, если такие есть [22, 50]. Это позволяет выявить наличие тенденции в развитии процесса,

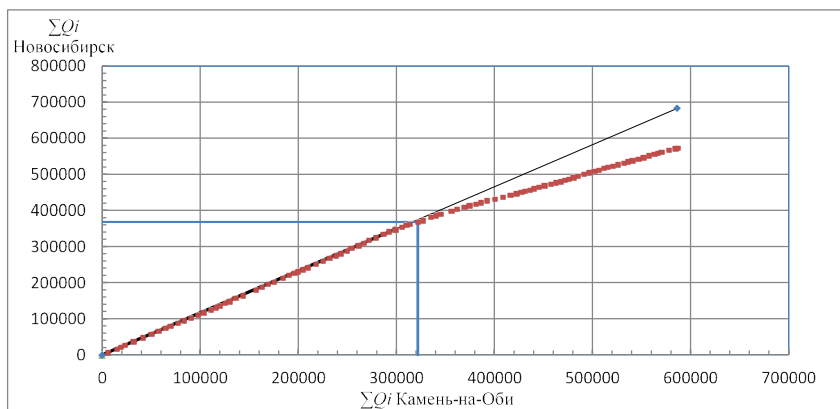


Рис. 8.4. График связи интегральных кривых максимального годового стока р. Обь у г. Новосибирск и г. Камень-на-Оби за 1915–2012 гг. (Новосибирское водохранилище)

даже если другими методами оценки однородности и стационарности она не выявляется.

Последнее определяется тем, что, как отмечалось выше, с увеличением j значимость отклонений средних значений в отдельные годы от средних многолетних значений в сумме W_j уменьшается и общий вид графика связи (5.3) определяется общей тенденцией нарастания суммы значений с увеличением периода наблюдений.

На рис. 8.4 представлены интегральные кривые среднего годового стока р. Обь у г. Новосибирск и г. Камень-на-Оби. Очевидно, что до 1959 или 1960 г., когда ввели в строй новосибирского водохранилища, все точки зависимости хорошо ложатся на прямую линию связи. После этих лет произошло постепенное отклонение интегральной кривой стока в сторону уменьшения.

8.3.2. Проверка выборочных рядов на случайность

Для решения вопроса о наличии или отсутствии внутрирядных связей производится проверка рассматриваемых рядов наблюдений на случайность, под которой подразумевается отсутствие внутрирядных связей. С учетом выборочности рядов наблюдений такая проверка имеет статистический характер и основана на нулевой гипотезе об отсутствии внутрирядных связей. Если гипотеза опровергается, то рассматриваемый ряд по-видимому не является

случайным, то есть, скорее всего, он обладает некоторыми свойствами, связанными главным образом с характером группировок (серий) членов ряда по тем или иным признакам.

Проверка случайности по критерию длин и числа серий

Серия – последовательность элементов (членов ряда) одинаковых по какому-либо признаку, непосредственно перед которой и после которой стоят элементы другого вида.

Длина серии – число элементов в серии.

Чаще всего в качестве признака серий берется знак отклонений, положительный или отрицательный, значений ряда от медианы.

Проверка случайности по данному критерию основана на нулевой гипотезе об отсутствии внутрирядных связей:

$$H_0 : r(\tau) = 0 \text{ при } \tau \neq 0$$

Она проводится путем сопоставления длин и числа серий исследуемого ряда с длиной и числом серий выборок случайной величины той же продолжительности.

Число серий различной длины в выборках случайной величины различной продолжительности при заданном уровне значимости, устанавливается теоретически [60]. Так, математическое ожидание числа серий с длиной, большей или равной k , состоящих из элементов больших ($MR_{1,k}$) и меньших ($MR_{2,k}$) медианы¹, определяется соответственно по формулам:

$$\begin{aligned} MR_{1,k} &= \frac{(n_2 + 1) \prod_{i=1}^k (n_1 - i + 1)}{\prod_{i=1}^k (n - i + 1)}, \\ MR_{2,k} &= \frac{(n_1 + 1) \prod_{i=1}^k (n_2 - i + 1)}{\prod_{i=1}^k (n - i + 1)}, \end{aligned} \quad (8.22)$$

где n_1 и n_2 – число членов ряда соответственно больших и меньших медианы; n – число членов ряда, $\prod$ – знак произведения. Можно

¹ Если закон распределения близок к нормальному, то вместо медианы можно взять среднее значение.

вместо формул (8.22) воспользоваться вытекающими из них рекуррентными формулами:

$$\begin{aligned} MR_{1,k} &= \frac{MR_{1,k-1} [n_1 - (k-1)]}{[n - (k-1)]}, \\ MR_{2,k} &= \frac{MR_{2,k-1} [n_2 - (k-1)]}{[n - (k-1)]}. \end{aligned} \quad (8.23)$$

Зная математические ожидания числа серий, длина которых не меньше k , можно рассчитать математическое ожидание числа серий данной продолжительности k :

$$\begin{aligned} Mr_{1,k} &= MR_{1,k} - MR_{1,k+1}, \\ Mr_{2,k} &= MR_{2,k-1} - MR_{2,k+1}. \end{aligned} \quad (8.24)$$

Математическое ожидание общего числа серий в исходном ряду с длиной, не меньшей k , можно рассчитать путем суммирования $MR_{1,k}$ и $MR_{2,k}$, т. е.

$$MR_k = MR_{1,k} + MR_{2,k}. \quad (8.25)$$

Соответственно математическое ожидание общего числа серий продолжительности k :

$$Mr_k = Mr_{1,k} + Mr_{2,k}. \quad (8.26)$$

Зная математическое ожидание и дисперсию общего числа серий Mr_k , можно получить функцию распределения общего числа серий при различной длительности выборок случайной величины и, исходя из заданного уровня значимости α , определить доверительные границы и критическую область общего числа серий (табл. 8.1).

Таблица 8.1. Критические границы общего числа серий R в случайных рядах различной продолжительности при двухстороннем уровне значимости

Уровень значимо- сти, %	Число членов ряда													
	10	20	30	40	50	60	80	100	120	140	160	180	200	
5	3	6	11	15	19	24	33	42	51	60	70	79	88	
95	8	15	20	26	32	37	48	59	70	81	91	102	113	
2,5	2	6	10	14	18	22	31	40	49	58	68	77	86	
97,5	9	15	21	27	33	39	50	61	72	83	93	104	115	

Если по исследуемому ряду наблюдений общее число серий окажется меньше критических границ при 5%-м или 2,5%-м уровне

значимости, то можно считать, что нулевая гипотеза о случайности ряда опровергается.

Для подсчета числа серий в исследуемом ряду при ручном счете члены исходного ряда больше медианы обычно обозначаются через «а», меньше медианы – через «в»; при машинном счете – через «0» и «1». Незамкнутые серии в начале и конце ряда наблюдений обязательно исключаются.

Проверка случайности по числу повышений и понижений значений ряда

Сущность проверки состоит в следующем. Пусть имеется выборка $x_1, x_2, \dots, x_n$. Переход $x_{i-1} < x_i$ называется повышением (+), а $x_i > x_{i+1}$ – понижением (–). Общее число повышений (или понижений) ряда значений случайной величины распределено асимптотически нормально с математическим ожиданием:

$$m_+ = m_- = \frac{n}{2}, \quad (8.27)$$

и дисперсией:

$$D_+ = D_- = \frac{(n+1)}{12}. \quad (8.28)$$

Зная m и D , можно рассчитать по данному ряду нормированную величину числа повышений или понижений:

$$\hat{t}_+ = \frac{(\hat{n}_+ - m_+)}{\sqrt{D_+}}, \quad \hat{t}_- = \frac{(\hat{n}_-^* - m_-)}{\sqrt{D_-}}, \quad (8.29)$$

где n_+ и n_- – число повышений и понижений в исследуемом ряду соответственно.

Дальше, учитывая асимптотически нормальный закон распределения числа повышений и понижений, сравниваем t_+ или t_- со значениями нормированных ординат таблицы нормального закона распределения (см. например, работу _). Если вероятность рассчитанных значений t_+ или t_- по таблице окажется меньше заданного уровня значимости, то гипотеза о случайности исследуемого ряда опровергается и считается, что ряд имеет устойчивую тенденцию к повышению или понижению. С помощью этого критерия хорошо обнаруживаются имеющиеся систематические изменения уровня ряда (тренда).

Проверка случайности по числу экстремумов

Экстремумом называется любой элемент, принадлежащий последовательности $x_1, x_2, \dots, x_n$, для которого выполняется одно из неравенств $x_{i-1} < x_i > x_{i+1}$ или $x_{i-1} > x_i < x_{i+1}$. В первом случае x_i – максимум, во втором – минимум. Общее число экстремумов случайного ряда распределено асимптотически нормально с математическим ожиданием:

$$m_3 = \frac{2n}{3}, \quad (8.30)$$

и дисперсией:

$$D_3 = \frac{(16n - 29)}{90}. \quad (8.31)$$

Для проверки гипотезы случайности ряда X достаточно рассчитать по этому ряду фактическое нормированное число экстремумов:

$$\hat{t}_3 = \frac{(\hat{n}_3 - m_3)}{\sqrt{D_3}}, \quad (8.32)$$

где $\hat{n}_3$ – число экстремумов, и сравнить его с нормированным значением t_α нормального закона распределения при заданном уровне значимости α . При $t_3^* > t_\alpha$ гипотеза о случайности ряда, по-видимому, неверна.

С помощью этого критерия хорошо обнаруживаются характерные циклы в ряду наблюдений.

При подсчете числа повышений (понижений) и числа экстремумов может обнаружиться, что два соседних элемента x_j и x_{j+1} равны между собой. В этом случае один из них следует исключить, а общее число членов ряда уменьшить на 1.

8.4. Корреляционные функции геоэкологических процессов

8.4.1. Общие сведения

Если в результате проверки случайности имеющегося ряда наблюдений выяснилось наличие внутрирядных связей, то для описания изменений данного процесса наряду с ранее рассмотренными характеристиками используется корреляционная функция. При этом предполагается, что внутрирядные связи являются линейными.

В общем случае корреляционная функция¹ показывает степень зависимости между сечениями случайной функции, относящимися к различным значениям аргумента. При этом в качестве аргумента чаще всего выступают время или какие-либо пространственные характеристики.

По одной реализации стационарного эргодического процесса корреляционная функция показывает степень зависимости значений процесса от предшествующих значений, при различных сдвигах τ . Чем больше абсолютная величина ординат корреляционной функции, тем больше связаны отдельные сечения (множество реализаций) или значения (одна реализация) и тем в большей степени выявляется какая-то тенденция в развитии процесса. Следовательно, корреляционная функция характеризует внутреннюю структуру или динамику развития процесса во времени (пространстве).

Зная оценку корреляционных функций, можно решать широкий круг задач, связанных с расчетом и прогнозом геофизических процессов. К числу таких задач относятся определение продолжительности и частоты циклов колебаний геофизических процессов, выявление скрытых периодичностей, краткосрочные и долгосрочные прогнозы. При наличии внутрирядной связи по-иному ставится и решается задача оценки числовых характеристик распределения.

Для расчетов корреляционных функций по выборкам неограниченной продолжительности используется ряд формул.

Корреляционная функция непрерывных случайных процессов согласно выражению (8.13) рассчитывается по формуле:

¹ Ранее отмечалось, что под этим названием как здесь, так и дальше будет пониматься нормированная корреляционная функция (формула 8.13). В дальнейшем в качестве аргумента, если это специально не оговорено, будет выступать время.

$$r(t, t') = \frac{1}{[\sigma(t)\sigma(t')]} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [X(t) - m_x(t)][X(t') - m_x(t')] \times \\ \times f[X(t)]f[X(t')]d[X(t)]d[X(t')], \quad (8.33)$$

где $f[X(t)]$ и $f[X(t')]$ – плотность распределения случайной функции в сечениях t и t' .

В практических расчетах временные и пространственные ряды наблюдений часто являются или сводятся к дискретным. Корреляционная функция таких рядов в общем случае рассчитывается по формуле:

$$r(t, t') = \left\{ \frac{1}{[l\sigma(t)\sigma(t')]} \right\} \sum_{i=1}^l [x_i(t) - m_x(t)][x_i(t') - m_x(t')], \quad (8.34)$$

где l – число реализаций.

Расчет корреляционных функций стационарных процессов с учетом выражения (8.18) производится по формуле:

$$r(\tau) = \frac{\sum_{i=1}^l [x_i(t) - m_x][x_i(t') - m_x(t')]}{(l\sigma_x^2)}, \quad (8.35)$$

где m_x и σ_x в соответствии с условиями стационарности постоянны во времени.

Расчет корреляционных функций стационарных эргодических процессов, согласно определению эргодичности, может производиться по одной реализации:

$$r(\tau) = \frac{\sum_{i=1}^{N-\tau} [x_i(t) - m_x][x_{i+\tau} - m_x]}{[(N-\tau)\sigma_x^2]}, \quad (8.36)$$

где X_i и $X_{i+\tau}$ – значения исследуемого ряда X соответственно в момент времени i и $i + \tau$.

8.4.2. Корреляционные функции стационарных процессов

Как отмечалось выше при анализе рядов наблюдений над теми или иными характеристиками окружающей среды, часто

принимается, что исследуемые процессы являются стационарными и эргодическими. Это предположение позволяет приближенно определять все характеристики процесса по одной достаточно длинной или по множеству коротких реализаций.

Рассмотрим основные свойства корреляционных функций стационарных процессов.

1. Корреляционная функция стационарного случайного процесса является функцией только одного аргумента $\tau = t' - t$, т. е.

$$r(t, t') = r(t, t + \tau) = r(\tau). \quad (8.37)$$

Это свойство следует непосредственно из определения стационарного процесса (см. разд. 8.3.2).

2. Значение корреляционной функции при сдвиге $\tau = 0$ равно 1, т. е.

$$r(\tau = 0) = 1. \quad (8.38)$$

3. Корреляционная функция стационарного случайного процесса является четной, т. е.

$$r(\tau) = r(-\tau). \quad (8.39)$$

4. Корреляционная функция стационарного случайного процесса является положительно определенной, т. е. корреляционной матрицы, построенной из последовательности значений корреляционной функции:

$$D_{00} = \begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{n-1} \\ r_1 & 1 & r_1 & \dots & r_{n-2} \\ r_2 & r_1 & 1 & \dots & r_{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n-1} & r_{n-2} & r_{n-3} & \dots & 1 \end{vmatrix}, \quad (8.40)$$

и все его главные миноры неотрицательны. (В формуле 8.40 для краткости $r(\tau)$ обозначено через r_τ .) Матрица такого вида, симметричная, с одними и теми же элементами на любой диагонали, параллельной главной, называется автокорреляционной матрицей.

5. Для случайного стационарного эргодического процесса всегда выполняется условие:

$$r(\tau) \rightarrow 0 \text{ при } \tau \rightarrow \infty. \quad (8.41)$$

8.4.3. Корреляционные функции различных временных рядов

В зависимости от характера изменений во времени различные процессы имеют различную степень внутрирядной связанности, а, следовательно, и различные корреляционные функции.

Так, корреляционная функция чисто случайного ряда (рис. 8.5.а) имеет только один пик при $\tau = 0$, при остальных значениях сдвига τ ординаты корреляционной функции равны нулю.

Общим для корреляционных функций динамических процессов, в которых каждое последующее значение однозначно зависит от предшествующих, является четкая тенденция изменений ординат по τ . Например, на рис. 8.5.б представлена корреляционная функция динамического процесса $x(t) = a \sin \omega t$ с дисперсией $D_x = 1$. Как следует из формулы (8.36), в этом случае:

$$r(\tau) = \frac{\sum_t \sin \omega t \sin \omega(t - \tau)}{(n - \tau)} = b \cos \omega \tau. \quad (8.42)$$

Таким образом, корреляционная функция синусоидального процесса представляет собой косинусоиду, изменяющуюся от 1 до -1 с периодом, равным периоду исходного ряда наблюдений.

Если ряд представляется в виде суммы синусоид:

$$x(t) = \sum_{k=1}^m a_k \sin \omega_k t, \quad (8.43)$$

то по аналогии с выражением (8.42) его корреляционная функция будет определяться формулой:

$$r(\tau) = \sum_{k=1}^m b_k \cos \omega_k \tau. \quad (8.44)$$

В случае, если развитие процессов во времени подчиняется динамико-статистическим закономерностям, корреляционные функции являются более сложными и лишь в простейших случаях могут быть аппроксимированы какой-либо аналитической формулой.

Примером реализации такого простейшего процесса может служить ряд:

$$x(t) = a \sin \omega t + \alpha, \quad (8.45)$$

где α – случайная величина.

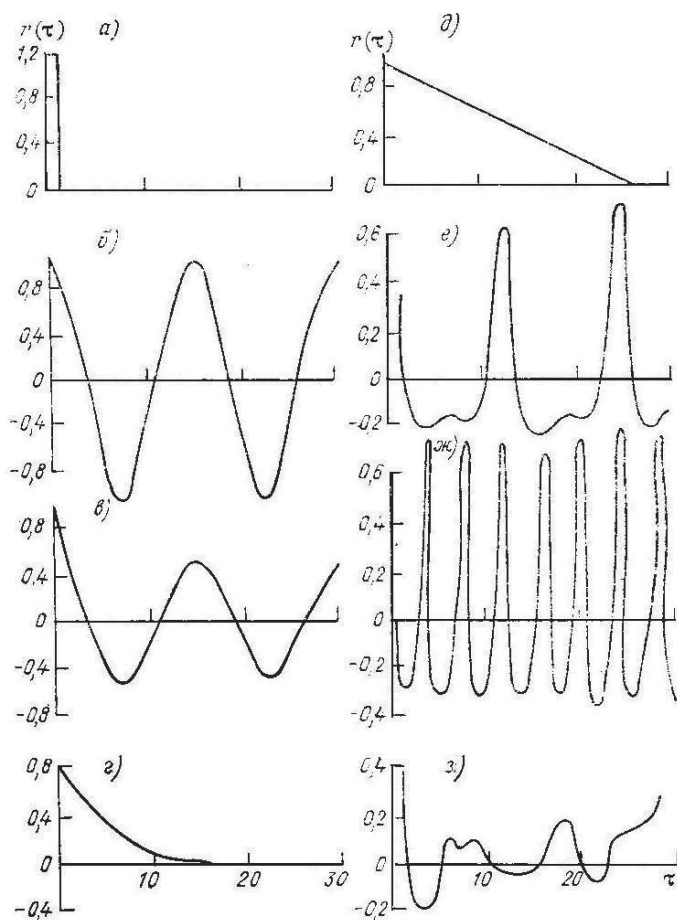


Рис. 8.5. Корреляционные функции различных временных рядов: чисто случайного ряда (а); синусоиды (б); синусоиды с наложенным «белым шумом» (в); ряда значений простой цепи Маркова (г); ряда, полученного скользящим сглаживанием чисто случайной величины (д) и последовательностей средних месячных, средних квартальных и средних годовых значений стока соответственно (е-з)

Пусть, например, критерий случайности, равный отношению дисперсии случайной составляющей к дисперсии исходного ряда

$$\delta = \frac{D_{\alpha}}{D_x}, \quad (8.46)$$

равен 0,5, тогда корреляционная функция $X(t)$ (см. рис. 8.5.в) будет определяться формулой:

$$r(\tau) = \frac{b \cos \omega \tau}{(1 + D_a)} = \frac{b \cos \omega \tau}{2}, \quad (8.47)$$

т. е. все значения $r(\tau)$ процесса, определяемого по формуле (8.47), будут по абсолютной величине в 2 раза меньше $r(\tau)$ аналогичного динамического процесса без случайной составляющей.

Сравнительно простой вид имеет корреляционная функция ряда значений простой цепи Маркова (красный шум) (см. рис. 8.5.з). Она описывается формулой:

$$r(\tau) = r(\tau = 1)^\tau = r_1^\tau. \quad (8.48)$$

Большой интерес представляют корреляционные функции временных рядов, сглаженных скользящим образом. Так, на рис. 8.5.д представлена корреляционная функция чисто случайного ряда, сглаженного скользящим образом периодом в m лет. Можно показать, что ее значения определяются по формуле:

$$r(\tau) = \frac{(m - \tau)}{m}. \quad (8.49)$$

На рис. 8.5.е приведены корреляционные функции средних месячных, средних квартальных (рис. 8.5.ж) и средних годовых (см. рис. 8.5.з) расходов воды. Следует отметить, что в отличие от предыдущих примеров значения этих корреляционных функций определялись по выборкам и имеют выборочный характер, поэтому их визуальное сопоставление носит условный характер. Вместе с тем можно отметить, что представленные корреляционные функции геоэкологических рядов меняются примерно в тех же пределах, что и, например, представленная здесь корреляционная функция синусоиды со случайной составляющей.

8.4.4. Оценки корреляционных функций по выборочным данным

Расчеты корреляционных функций, как правило, проводятся по выборочным данным. В результате таких расчетов получают не истинные, а выборочные значения или, иначе говоря, оценки корреляционных функций:

$$\hat{r}(\tau) = \frac{\sum_{i=1}^l [X_i(t) - \hat{m}_x(t)][X_i(t') - \hat{m}_x(t')]}{[l\hat{\sigma}(t)\hat{\sigma}(t')]}, \quad (8.50)$$

где [см. формулу (8.35)] вместо действительных значений $m_x(t)$, $m_x(t')$, $\sigma(t)$, $\sigma(t')$ вводятся выборочные характеристики: $\hat{m}_x(t)$, $\hat{m}_x(t')$, $\hat{\sigma}(t)$, $\hat{\sigma}(t')$.

Для стационарных случайных процессов формула выборочной корреляционной функции может быть представлена в виде:

$$\hat{r}(\tau) = \frac{\sum_{i=1}^l [X_i(t) - \hat{m}_x][X_i(t+\tau) - \hat{m}_x]}{(l\hat{\sigma}_x^2)}. \quad (8.51)$$

При расчетах корреляционной функции стационарных эргодических процессов по одной реализации обычно используется формула:

$$\hat{r}(\tau) = \frac{\sum_{i=1}^{n-\tau} (x_i - \hat{m}_x)(x_{i+\tau} - \hat{m}_x)}{[(n-\tau-1)\hat{\sigma}_x^2]}, \quad (8.52)$$

которая может приниматься в качестве состоятельной и несмещенной оценки действительного значения $r(\tau)$ [65].

В практических приложениях вместо формулы (8.52) часто используется формула:

$$\hat{r}(\tau) = \frac{\sum_{i=1}^{n-\tau} (x_i - \hat{m}_{x1})(x_{i+\tau} - \hat{m}_{x2})}{[(n-\tau-1)\hat{\sigma}_1\hat{\sigma}_2]}, \quad (8.53)$$

где $\hat{m}_{x1}$, $\hat{\sigma}_1$ и $\hat{m}_{x2}$, $\hat{\sigma}_2$ – средние значения и средние квадратические отклонения значений ряда X соответственно от 1-го до $(n-\tau)$ -го и от τ -го до n -го члена.

Различие формул (8.52) и (8.53) состоит в том, что при расчетах по формуле (8.35) линии регрессии $X_{i+\tau}$ по X_i и X_i по $X_{i+\tau}$ проходят через центр тяжести не всей возможной совокупности точек,

а только тех из них, которые непосредственно использовались для расчета коэффициентов корреляции. При этом возможные смещения линий регрессии относительно центра тяжести всей совокупности точек могут нарастать с ростом запаздывания τ , так как чем больше x , тем сильнее могут различаться m_x , σ_x и $\bar{x}_1$, $\bar{x}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$. В то же время, как показал опыт исследований, смещение линий регрессии позволяет в некоторой степени учесть возможную нестационарность при анализе внутрирядных связей.

Понятие корреляционной функции широко используется при решении ряда научных и практических вопросов. Вместе с тем во многих случаях имеются сомнения в достоверности отличия от нуля ее отдельных значений, особенно при больших сдвигах τ . Это вызвано тем, что до настоящего времени не существует достаточно строгих оценок ординат корреляционных функций по выборкам ограниченного объема.

Имеющиеся приближенные приемы оценок достоверности разделяются по объекту исследований и количественным характеристикам корреляционных функций, выбранных для оценки. Так, по объекту исследований выделяют методы оценок по конкретному ряду наблюдений и по группе рядов, объединенных в один ансамбль по какому-либо признаку. В качестве количественных характеристик используются величины отдельных ординат, обобщенные характеристики ординат всех значений корреляционной функции, средний период и т.д.

Технически наиболее простой является оценка отдельных ординат корреляционной функции по формуле средней квадратической погрешности парного коэффициента корреляции [6, 12, 33]:

$$\hat{\sigma}_{r(\tau)} \cong \frac{(1 - r^2(\tau))}{\sqrt{n - \tau - 1}}, \quad (8.54)$$

где вместо неизвестного значения $r(\tau)$ принимается его выборочная оценка.

Рассчитанные значения $\sigma_{r(\tau)}$ используются для оценки погрешности расчета отдельных ординат $r(\tau)$. Если значения $\hat{r}(\tau)$ и $\hat{\sigma}(\tau)$ соизмеримы, то делается вывод о незначимости отличия $\hat{r}(\tau)$ от нуля и, как следствие, о равенстве действительного значения $r(\tau)$ нулю.

Следует отметить, что такая оценка корреляционных функций не всегда правомерна. В частности, если рассматриваемый процесс

является циклическим, то его корреляционная функция также должна быть циклической, т. е. значения ее ординат должны переходить от положительных значений к отрицательным и наоборот. Значения корреляционных функций, переходные от максимальных положительных к максимальным отрицательным, могут быть близкими или равными нулю. Достоверность проверки этих значений по $\sigma_{r(\tau)}$ очень низка. Кроме того, статистически незначимость отличия ординат от нуля еще не означает равенство их нулю (см. гл. 6).

Оценка достоверности ординат корреляционных функций при заданных значениях τ нередко проводится также с помощью доверительных интервалов, построенных на основе нулевой гипотезы, что действительные значения корреляционной функции равны нулю, т. е. $H_0 : r(\tau) = 0$ при $\tau \neq 0$. При этом определение доверительных интервалов производится по формуле:

$$|r_0| = t_\alpha \sigma_{r(\tau)}, \quad (8.55)$$

где t_α – критерий Стьюдента с $n - \tau - 1$ степенями свободы при заданном уровне значимости α ; $\sigma_{r(\tau)}$ – среднее квадратическое отклонение ординат корреляционных функций, рассчитанное по формуле (8.54) в предположении, что $r(\tau) = 0$.

Исходя из нулевой гипотезы производится проверка отдельных ординат выборочной корреляционной функции $\hat{r}(\tau)$. Если окажется, что при данном τ ордината $\hat{r}(\tau)$ выходит за пределы доверительного интервала и попадает в критическую область гипотезы об отсутствии внутрирядной связи, т. е. отличается от нуля более, чем это можно разумно приписать случайной вариации, то отличие ординаты от нуля считается значимым и H_0 для данного τ опровергается.

Так же, как и предыдущий, этот прием не учитывает возможную цикличность колебаний исследуемого процесса, когда по выборкам ограниченного объема в критическую область гипотезы о равенстве значений корреляционных функций нулю могут попасть лишь некоторые значения корреляционных функций. Сколько должно быть таких значений и как они должны распределяться по τ – неизвестно. Попытка компенсации этого недостатка предпринята в работе [44]. При этом для учета возможной цикличности процессов было предложено производить оценку достоверности по доверительным уровням значимости только экстремумов пиков корреляционных функций, если каждому такому экстремуму предшествуют и следуют за

ним 2–3 переходных значения. Однако эта оценка не получила распространения и как следует не разработана.

Как отмечалось, рассмотренные оценки ординат корреляционных функций не учитывают возможную цикличность рассматриваемого процесса. Поэтому Ю.М. Алехиным была предложена новая характеристика корреляционных функций – так называемый средний период [4]:

$$t_r = \frac{2(\tau_{\text{посл}} - \tau_1)}{(N-1)}, \quad (8.56)$$

где $\tau_{\text{посл}}$ и τ_1 – абсциссы соответственно последней и первой точек пересечения оси абсцисс линий в поле графика $r(\tau)$, осредняющей смежные ординаты; N – общее число точек пересечения или касания.

Если $t_r \geq 7$ или 8, то корреляционная функция считается достоверной. Недостатком оценки по t_r является некоторая неопределенность нижних и верхних границ t_r для различных процессов.

Изложенные приемы оценки корреляционных функций по конкретным рядам наблюдений достаточно эффективны при расчетах по очень длинным выборкам, особенно по выборкам простой цепи Маркова, во всех других случаях результаты анализа являются сомнительными и попытки их использования носят дискуссионный характер.

В связи с этим для увеличения объема информации в последнее время разрабатываются приемы оценки, основанные на объединении некоторого числа рядов в один ансамбль. Один из таких приемов предложен в 1969 г. [34] и заключается в расчете средней из множества корреляционных функций, полученных по рядам наблюдений за процессами в районе с одинаковыми физико-географическими условиями. Другой прием, предложенный в том же году, заключается в расчете корреляционной функции по математическому ожиданию процесса стока, представленному множеством реализаций в однородном районе [48].

В результате объединения рядов наблюдений с помощью первого и второго способа происходит существенное увеличение объема информации, и оценка полученных корреляционных функций становится более надежной.

Наконец, в последнее время предложен прием обобщенной оценки достоверности корреляционных функций по множеству рядов наблюдений [48]. Он предназначен для решения вопроса об общей временной структуре какого-либо процесса, например, процесса среднего годового стока.

Прием заключается в расчете корреляционных функций отдельных рядов стока по всему периоду наблюдений N и нарастающим периодам n ($n = 40, \dots, N$). Градиент нарастания n может быть, например, принят равным двум годам. В качестве числовых характеристик корреляционных функций принимается среднее квадратическое отклонение от оси абсцисс по всем τ ($\tau = 1, 2, \dots, \tau_{\max}$) — σ_{r_0} , отдельно для первой ($\tau = 1, 2, \dots, 15$) — σ_{r_1} и второй ее части ($\tau = 16, 17, \dots, 30$) — σ_{r_2} :

$$\sigma_{r_{0,1,2}} = \sqrt{\frac{\sum_{\tau} r^2(\tau)}{k}}. \quad (8.57)$$

Затем результаты расчетов для каждого n осредняются по всем рядам наблюдений: $\bar{\sigma}_{r_0,n}$; $\bar{\sigma}_{r_1,n}$; $\bar{\sigma}_{r_2,n}$, и строится график зависимости $\sigma_{r,n} = f(n)$ (рис. 8.6). Одновременно на этом же графике строятся

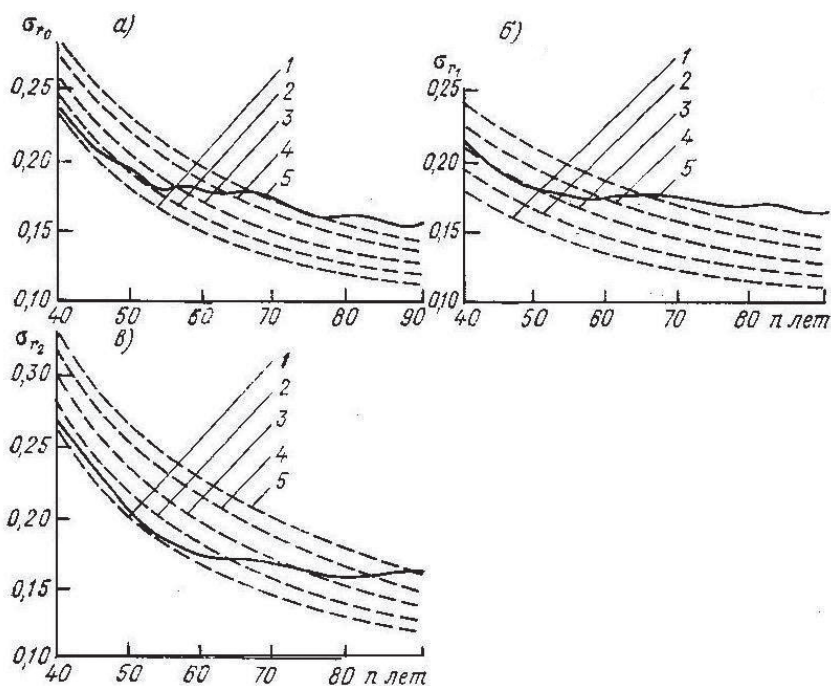


Рис. 8.6. Зависимость σ_{r_0} (а), σ_{r_1} (б), σ_{r_2} (в) от продолжительности наблюдений в поле доверительных интервалов гипотезы H_{01} при различных уровнях значимости: 1) 1%, 2) 2%, 3) 5%, 4) 20%, 5) 50%

доверительные интервалы $\sigma_r = f(n)$ при различных уровнях значимости исходя из нулевой гипотезы $H_0 : r(\tau) = 0$ при $\tau = 0$, и следовательно, и $H_0 : \sigma_r = 0$. Если окажется, что значимость рассчитанных величин повышается с увеличением n , то это говорит об отличии исследуемых рядов от случайного ряда. Несколько сложнее оказалась проверка предположения о том, что исходные ряды аппроксимируются простой цепью Маркова, так как обоснованное построение доверительных интервалов здесь представляет серьезные трудности. В данном случае построение доверительных границ производилось методом Монте-Карло (см. гл. 9).

8.5. Спектральный анализ геоэкологических процессов

С помощью корреляционных функций может быть решено достаточно много задач по исследованию колебаний геофизических процессов. Однако при анализе сложных процессов корреляционные функции часто не поддаются интерпретации. Кроме того, непосредственно по корреляционным функциям нельзя количественно установить выраженность колебаний с тем или иным периодом. В связи с этим при анализе колебаний геофизических процессов наряду с корреляционным широко используется спектральный анализ, позволяющий оценить распределение энергии колебаний исследуемого процесса в частотной области.

8.5.1. Гармонический анализ

Понятие спектра применяется в различных областях науки и техники, особенно в физике и механике.

Как известно, любую функцию, заданную в каждой отдельной точке, можно представить бесконечным рядом синусоидальных и косинусоидальных функций. Такой ряд называется *рядом Фурье*, а метод нахождения функций – *анализом Фурье* [5,32]. Исходные данные обычно представлены в виде дискретных значений, заданных через равные промежутки времени или пространства, высоты и др. Поэтому на заданном интервале все наблюдения могут быть учтены конечным числом синусов и косинусов.

Так, пусть данный дискретный ряд $z(t)$ представляет непрерывный процесс $Z(t)$ за период длительностью T . Члены ряда заданы через равные интервалы времени Δ . Тогда общее число отсчетов или объем ряда наблюдений $N = T/\Delta$. Каждое значение ряда $z(t) = z(t = r\Delta) = z_r$, где r – порядковый номер отсчета, «+», если вправо от точки отсчета, и «-», если влево. Будем считать, что N четное и равно $2n$, так что r может изменяться от $-n$ до n . Тогда ряд $z(t) = z_r$ может быть представлен в виде суммы гармонических колебаний различных частот:

$$z_r = A_0 + \sum_{i=1}^{n-1} (A_i \cos 2\pi f_i r \Delta + B_i \sin 2\pi f_i r \Delta) + A_n \cos 2\pi f_1 n r \Delta, \quad (8.58)$$

где $r = -n, \dots, 0, 1, \dots, n-1$, f_1 – основная частота колебаний; A_i и B_i – коэффициенты i -й гармоники, подбираемые таким образом, чтобы рассчитанные значения z_r совпадали с фактическими.

Обычно основная частота f_1 выбирается равной $1/(N\Delta)$. При этом решение системы уравнений (8.58) сильно упрощается, так как синусы и косинусы будут ортогональны, т. е. будут удовлетворять следующим соотношениям:

$$\begin{aligned} \sum_{r=-n}^{n-1} \sin\left(\frac{2\pi kr}{N}\right) \cos\left(\frac{2\pi mr}{N}\right) &= 0, \quad k, m - \text{целый;} \\ \sum_{r=-n}^{n-1} \sin\left(\frac{2\pi kr}{N}\right) \cos\left(\frac{2\pi mr}{N}\right) &= \begin{cases} 0, k \neq m, \\ \frac{N}{2}, k = m \neq 0, n, \\ 0, k = m = 0, n; \end{cases} \\ \sum_{r=-n}^{n-1} \sin\left(\frac{2\pi kr}{N}\right) \cos\left(\frac{2\pi mr}{N}\right) &= \begin{cases} 0, k \neq m, \\ \frac{N}{2}, k = m \neq 0, n, \\ N, k = m = 0, n. \end{cases} \end{aligned} \quad (8.59)$$

Частоты всех гармоник кратны основной частоте f_1 , т. е. являются гармониками основной частоты. Наивысшей из присутствующих частот является частота $f = f_1 n \Delta$, где $\frac{n\Delta}{\Delta N} = \frac{1}{2}$, что соответствует периоду, равному двум интервалам отсчета, т. е. 2Δ .

Коэффициенты A_i и B_i можно найти, умножая обе части выражения (8.58) сначала на $\cos(2\pi f_i r \Delta)$, затем на $\sin(2\pi f_i r \Delta)$. Суммируя

обе части по r и учитывая соотношения ортогональности (8.59), в окончательном виде, получаем:

$$A_i = \left(\frac{1}{N} \right) \sum_{r=-n}^{n-1} z_r \cos \left(\frac{2\pi m r}{N} \right), \quad (8.60)$$

$$B_i = \left(\frac{1}{N} \right) \sum_{r=-n}^{n-1} z_r \sin \left(\frac{2\pi m r}{N} \right), \quad (8.61)$$

где $i = 0, 1, \dots, n$. Из формулы (8.58) следует, что $A_0 = \bar{z}$.

Уравнение (8.58) часто удобнее представлять в виде:

$$z_r = \bar{z} + 2 \sum_{i=1}^{n-1} \left[C_i \cos(2\pi f_i t - \varphi_i) \right] + C_n \cos(2\pi f_1 t). \quad (8.62)$$

Возможность такого представления можно доказать, преобразовав уравнения (8.58) в соответствии с формулой косинуса разности:

$$z_r = \bar{z} + 2 \sum_{i=1}^{n-1} \left[C_i \sin(2\pi f_i t) + C_i \cos(2\pi f_i t) \cos \varphi_i \right] + C_n \cos(2\pi f_1 t).$$

Сравнив полученную формулу с формулой (8.58), видим:

$$A_i = C_i \cos \varphi_i; \quad B_i = C_i \sin \varphi_i, \quad (8.63)$$

и отсюда:

$$C_i = \sqrt{A_i^2 + B_i^2}, \quad (8.64)$$

$$\varphi_i = \arctg \left(\frac{B_i}{A_i} \right). \quad (8.65)$$

Переменная C_i называется **амплитудой i -й гармоники**, φ_i – **фазой этой гармоники**. Переменная φ_i показывает время, при котором i -я гармоника имеет максимум. В приведенных выше формулах начало отсчета времени бралось в точке, расположенной примерно посередине между первым и последним отсчетом z_r . Если бы мы изменили начало отсчета, то амплитуда C_i осталась бы прежней, а значения A_i , B_i и φ_i изменились бы соответствующим образом.

Дисперсия значений z_r определяется по формуле:

$$D_z = \left(\frac{1}{N} \right) \sum_{r=-n}^{n-1} \left(z_r - \bar{z} \right)^2. \quad (8.66)$$

Используя свойство ортогональности (8.59) из формулы (8.66), можно получить:

$$D_z = 2 \sum_{r=-n}^{n-1} C_i^2 + C_n^2, \quad (8.67)$$

что является частным случаем теоремы Парсеваля, в которой утверждается, что дисперсия ряда z_r может быть разложена на составляющие, даваемые каждой i -й гармоникой.

Разложение дисперсии по различным частотам можно представить, если нанести на график значения $C_i = f(i)$ или $C_i = f(2\pi if_1)$. Обычно этот график называется **спектром наблюдений**. Он показывает участие колебаний различной частоты в дисперсии временного ряда. График $\frac{C_i^2}{D_z}$ называется **нормализованным спектром**

(рис. 8.4).

Начиная с 1920–30-х годов, в геоэкологии и других смежных науках было предпринято много попыток выделить отдельные гармоники с фиксированными периодами для анализа и прогноза колебаний гидрометеорологических процессов. Однако, несмотря на длительную историю поисков, попытки прогнозирования на этой основе оказались безуспешными. Поэтому в последние годы анализ гармоник оказался сильно скомпрометированным и уступил место спектральному анализу.

Во многом причиной неудачи является то, что для динамических рядов спектр является пределом выборочного спектра C_i при безграничном увеличении длины записи. Однако для стохастических рядов оценка C_i является настолько неустойчивой, что она становится бесполезной. При этом эксперименты со значениями, чисто случайного ряда [17] показали, что выборочный спектр не сходится

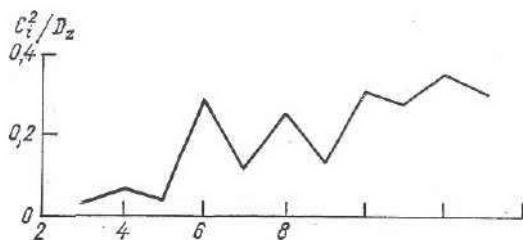


Рис. 8.7. Нормализованный спектр на основе разложения в ряд Фурье

с увеличением объема выборки, т. е. оценка спектра на основе гармонического анализа не является эффективной. Основная причина, по которой анализ Фурье неприменим к геофизическим рядам, заключается в том, что он основан на предположении, что амплитуды, частоты и фазы фиксированы [17], что не соответствует структуре этих рядов.

8.5.2. Спектральный анализ

При анализе стационарных случайных процессов обычно исходят из канонического разложения случайных процессов:

$$z_t = \bar{z} + \sum_{i=1}^m v_i \varphi_i(t), \quad (8.68)$$

где $v_1, v_2, \dots, v_m$ – некоррелированные случайные величины с математическим ожиданием, равным нулю; $\varphi_1(t), \varphi_2(t), \dots, \varphi_m(t)$ – координатные функции.

Спектральной плотностью или спектром этого процесса называется функция:

$$S(\omega) = \left(\frac{1}{\pi} \right) \int_0^{\infty} r(\tau) \cos \omega \tau d\tau, \quad (8.69)$$

где $\omega = \frac{2\pi}{T}$ – круговая частота.

Спектр представляет собой Фурье-преобразование корреляционной функции и, следовательно, полностью определяется ее значениями. Так же, как и выше при гармоническом анализе (см. формулу 8.67):

$$\sum S(\omega) = D_z, \quad (8.70)$$

где D_z – дисперсия исходного ряда.

Если значения z заданы в нормированном виде, то $D_z = 1$ и $\sum S(\omega) = 1$. Таким образом, формула (8.69) описывает распределение дисперсии по различным частотам. Если отложить по оси ординат плотность дисперсии, т. е. дисперсию, приходящуюся на единицу длины данного интервала частот (рис. 8.8), то полученная кривая называется **спектральной плотностью**.

С помощью спектральных плотностей дается описание исследуемого процесса в частотной области. Это описание может служить

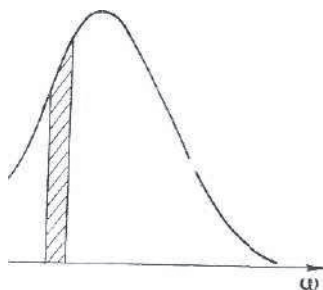


Рис. 8.8. Спектральная плотность

основой для выяснения происхождения и механизма колебаний процессов тех или иных временных масштабов.

При анализе спектральной плотности иногда оценивается часть дисперсии, соответствующая конечному интервалу частот $(0, \omega)$. Функция, характеризующая долю дисперсии D , приходящуюся на заданный интервал частот, называется **спектральной функцией**:

$$F(\omega) = \int_0^{\omega} S(\omega) d\omega. \quad (8.71)$$

Функция $F(\omega)$ является положительной неубывающей функцией, причем $F(0) = 0$, $F(\infty) = 1$. В некотором смысле понятия нормированная спектральная плотность и спектральная функция аналогичны плотности и функции распределения случайной величины.

8.5.3. Численные методы расчета спектральной плотности

Если известно аналитическое выражение корреляционной функции, то определение спектральной плотности не вызывает затруднений. Так, если корреляционная функция является экспоненциальной (см., например, корреляционную функцию простой цепи Маркова):

$$r(\tau) = \exp[-a(\tau)], \quad (8.72)$$

то ее спектральная плотность, как следует из формулы (8.69), определяется выражением:

$$S(\omega) = \frac{a}{\pi(\omega^2 + a^2)}. \quad (8.73)$$

В случае, если аналитическое выражение корреляционной функции неизвестно или это выражение имеет сложный характер, расчет спектральной плотности производится численными методами. Наибольшее распространение среди этих методов получили

методы трапецеидальных характеристик, метод прямоугольников, приближенный метод с учетом неравноточности определения значений корреляционных функций, формула, рекомендованная Всемирной метеорологической организацией (ВМО).

Метод трапецеидальных характеристик сводится к аппроксимации подынтегральной функции в формуле (8.69) ломаной кривой с соответствующим переходом от точного значения ее площади к приближенному значению, определенному как сумма площадей конечного числа трапеций:

$$S(\omega) \cong \frac{2}{\pi} \sum_{i=1}^{m-1} r_i t_i \frac{\sin \omega t_i}{\omega t_i} \frac{\sin \omega \Delta_i}{\omega \Delta_i}, \quad (8.74)$$

где $r_i = r_{\tau} - r_{\tau-1}$ – разности ординат соседних переломных точек ломаной кривой функции $r(\tau)$; $\Delta_i = \frac{(\tau_{i+1} - \tau_i)}{2}$ – полуразность абсцисс соседних точек; $t_i = \Delta_i + \tau_{i+1}$ – абсциссы середин между этими точками; $\omega = \frac{2\pi}{T}$ – круговая частота; m – максимальный сдвиг.

На конечном интервале времени, если интервал между наблюдениями равен 1, можно на основе перехода к конечной сумме по способу прямоугольника воспользоваться приближенным выражением:

$$S(\omega) \cong \left(\frac{1}{\pi} \right) \left[1 + \sum_{\tau=1}^m r(\tau) \cos \omega \tau \right]. \quad (8.75)$$

При расчетах спектральной плотности по выражениям (8.74) и (8.75) предполагаются известными действительные значения корреляционной функции $r(\tau)$. Однако, как отмечалось выше, на практике мы имеем дело с выборочными оценками корреляционной функции $\hat{r}(\tau)$ в пределах от 0 до r_m . Поэтому оценка спектра производится лишь на конечном промежутке времени:

$$\hat{S}(\omega) = \left(\frac{1}{\pi} \right) \int_0^{T_0} \hat{r}(\tau) \cos \omega \tau d\tau, \quad (8.76)$$

где $\hat{S}(\omega)$ – оценка спектральной плотности.

Математическое ожидание оценки $\hat{S}(\omega)$ имеет вид:

$$M[\hat{S}(\omega)] = \left(\frac{1}{\pi}\right) \int_0^{T_0} \left(\frac{1-\tau}{T_0}\right) r(\tau) \cos \omega \tau d\tau. \quad (8.77)$$

При увеличении интервала интегрирования математическое ожидание оценки стремится к ее истинному значению и, следовательно, оценка (8.77) является несмещенной, т. е.

$$\lim_{T \rightarrow \infty} M[\hat{S}(\omega)] = S(\omega). \quad (8.78)$$

Дисперсия этой оценки при $T_0 \rightarrow \infty$ стремится к нулю, следовательно, эта оценка несостоятельна.

Для получения состоятельной оценки применяется способ предварительного сглаживания выборочной корреляционной функции, т. е. принимается $\tilde{r}(\tau) = \lambda(\tau) \hat{r}(\tau)$, тогда

$$\tilde{S}(\omega) = \left(\frac{1}{\pi}\right) \int_0^{T_0} \tilde{r}(\tau) \cos \omega \tau d\tau. \quad (8.79)$$

Функция $\lambda(\tau)$ обычно называется временным окном, а ее преобразование Фурье – спектральным окном.

В гидрометеорологии в настоящее время чаще всего используются функции $\lambda(\tau)$, предложенные Бартлеттом, Хеммингом и Парзенном.

Весовая функция Бартлетта:

$$\lambda(\tau) = \begin{cases} 1, & \text{при } 0 \leq \tau \leq T_0, \\ 0, & \text{при } \tau > T_0. \end{cases} \quad (8.80)$$

Весовая функция Хемминга:

$$\lambda(\tau) = \begin{cases} \frac{0,54 + 0,46\pi}{T_0} & \text{при } 0 \leq \tau \leq T_0, \\ 0, & \text{при } \tau > T_0. \end{cases} \quad (8.81)$$

Однако применение функции Хемминга может привести к появлению отрицательных ординат спектральной плотности, которая в действительности всегда является положительной. Поэтому на частотах с отрицательными значениями $\tilde{S}(\omega)$ принимается равным нулю. Отрицательных оценок можно избежать, если использовать весовую функцию Парзена:

$$\lambda(\tau) = \begin{cases} 1 - 6\tau^2 \left(\frac{1-\tau}{T_0} \right) T_0^2 & \text{при } 0 \leq \tau \leq \frac{T_0}{2}, \\ 2 \left(\frac{1-\tau}{T_0} \right)^3 & \text{при } \frac{T_0}{2} \leq \tau \leq T_0, \\ 0, & \text{при } \tau > T_0. \end{cases} \quad (8.82)$$

При дискретном представлении исходных данных или невозможности взять интеграл (8.69) из-за сложности подынтегрального выражения расчет оценки спектральной плотности $\hat{S}(\omega)$ производится по формулам (8.74) или (8.75), а затем $\hat{S}(\omega)$ уточняется с помощью какого-либо из указанных выше весовых коэффициентов, что аналогично введению весовой функции в подынтегральное выражение.

В настоящее время в практике используются две численные схемы оценки спектральной плотности.

1. Схема Блекмана–Тьюки, которая заключается в нахождении первых r_m оценок корреляционной функции $\hat{r}(\tau)$ ($\tau = 0, 1, \dots, \tau_m$), последующего вычисления оценок спектральной плотности $\hat{S}(\omega)$ по формулам (8.74) или (8.75) и в сглаживании их по формуле (8.80), (8.81) или (8.82).

2. Методика Кули–Тьюки, которая заключается в нахождении периодограммы путем преобразования Фурье имеющегося ряда наблюдений [см. формулу (8.62)] и дальнейшей оценке корреляционной функции по формуле (8.70).

8.5.4. Выборочная оценка спектральной плотности

При оценках спектральной плотности по конкретным рядам приходится считаться с тем, что ординаты ее рассчитываются со значительными погрешностями. Это вызывается, во-первых, погрешностями значений корреляционных функций и, во-вторых,

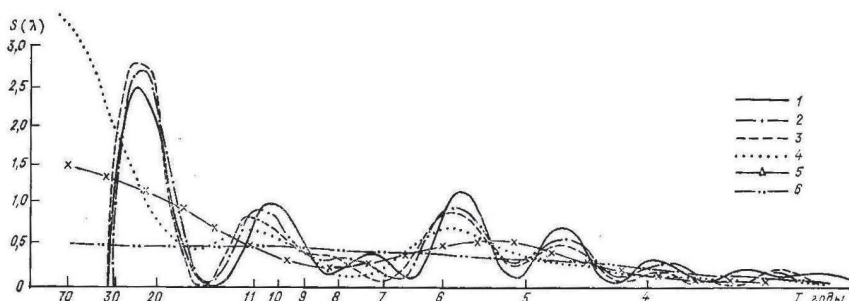


Рис. 8.9. Спектральная плотность
при различных максимальных временных сдвигах:
1) $m = 30$; 2) $m = 25$; 3) $m = 19$; 4) $m = 15$; 5) $m = 9$; 6) $m = 1$

неопределенностью значения максимального сдвига m . Для иллюстрации последнего обстоятельства на рис. 8.9 представлены спектральные плотности для различных значений m . Очевидно, что особенно при небольших m (до 10) значения спектра меняются очень резко, даже при изменении m на 1–2 года.

Спектральная плотность, рассчитанная по рядам наблюдений, обычно имеет много пиков, больших или меньших. Часть или все пики являются чисто случайными, вызванными выборочностью исходных данных. Очевидно, что случайные пики должны исключаться из анализа.

Достаточно надежных оценок значений спектральной плотности по коротким рядам нет. В качестве первого приближения обычно используется метод Тьюки.

Согласно исследованиям Тьюки, если выборки случайные и распределение случайной переменной нормально, то распределение выборочных спектральных оценок по отношению к спектру соответствующей генеральной совокупности соответствует распределению χ^2 , деленному на число степеней свободы ν :

$$\nu = \frac{\left(\frac{2n - m}{2} \right)}{m}, \quad (8.83)$$

где n — число членов ряда, m — максимальный сдвиг.

Для определения доверительного интервала каждой оценки спектра можно воспользоваться значениями распределения χ^2/ν (табл. 8.2).

**Таблица 8.2. Доверительные границы χ^2
в зависимости от числа степеней свободы и уровня значимости α**

γ	α					ν	α				
	1	5	10	95	99		1	5	10	95	99
1	6,64	3,84	2,71	0,07	—	4	3,30	2,37	1,94	0,18	0,07
2	4,60	3,00	2,30	0,05	0,01	5	3,02	2,21	1,85	0,23	0,11
3	3,78	2,61	2,08	0,12	0,04	6	2,60	2,10	1,77	0,27	0,15

Значения доверительного интервала определяются относительно среднего спектра:

$$S_0 = \frac{\sum_{\omega=0,01}^{2\pi} S(\omega)}{n_1} \quad (8.84)$$

или средней интенсивности спектра на определенном участке:

$$S_1 = \frac{\sum_{\omega=T_1}^{T_2} S(\omega)}{n_1} \quad (8.85)$$

по формуле:

$$S_0 - \left(\frac{\chi^2}{\nu} \right) S_0 \leq S \leq S_0 + \left(\frac{\chi^2}{\nu} \right) S_0. \quad (8.86)$$

В общем, теория показывает, что выборочные флуктуации спектров рядов стока достаточно велики и что пики и ложбины значимы только тогда, когда они экстремальны или если несколько соседних точек тоже подтверждают этот экстремум.

Средний спектр («нулевой континуум») может определяться также относительно простой цепи Маркова (рекомендация ВМО). В этом случае рассчитывается теоретический спектр простой цепи Маркова («красный шум»):

$$S_{0i} = \frac{S_0 (1 - r_i^2)}{\left[1 + r_i^2 - 2r_1 \cos\left(\frac{\pi_i}{m}\right) \right]}, \quad (8.87)$$

где S_0 — определяется по формуле (8.84), r_1 — значение корреляционной функции при сдвиге $\tau = 1$.

Остальные расчеты производятся так же, как и при оценке относительного спектра случайного ряда.

8.5.5. Спектральная плотность различных временных рядов

В зависимости от характера внутрирядной связи спектральные плотности имеют различный вид.

Так, нормированный спектр чисто случайного ряда ($r(\tau) = 0$ при $\tau > 0$) является прямой линией, параллельной оси абсцисс (рис. 8.10.а1). Эта форма спектра по аналогии со спектром видимого света называется «белым шумом».

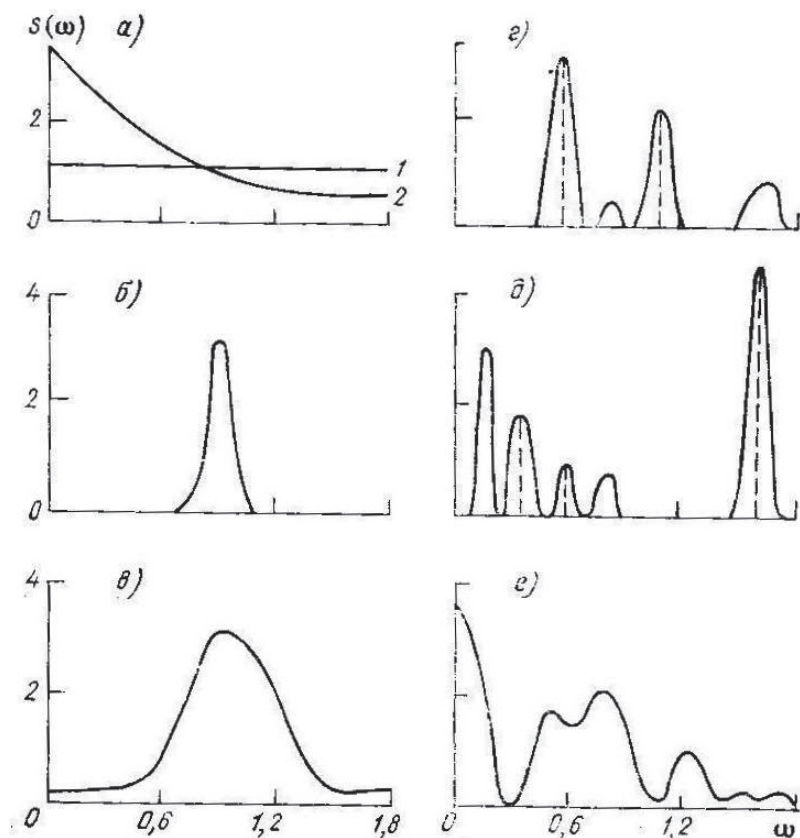


Рис. 8.10. Спектральная плотность различных временных рядов: чисто случайного ряда и простой цепи Маркова (а); синусоиды (б); синусоиды с наложенным «белым шумом» (в) и последовательностей средних месячных, средних квартальных и средних годовых значений стока соответственно (г–е)

Спектр временного ряда, в котором каждое последующее значение зависит только от предшествующего (простая цепь Маркова), представляется в виде линии, плавно или более резко (в зависимости от тесноты связи смежных значений) снижающейся к оси абсцисс (см. рис. 8.10.а2). Эта форма спектра часто называется «красным шумом».

Если во временных рядах содержится составляющая синусоидального вида, то спектр ряда будет содержать относительно резкий пик на частоте, соответствующей периоду синусоиды.

В случае, если ряд является квазипериодическим, то на спектре появляется относительно широкий бугор (горб), распространяющийся на соответствующий широкий диапазон длин волн (см. рис. 8.10.б).

Очевидно, что спектральные плотности различны для различных геоэкологических процессов и их характеристик. Например, для последовательностей средних месячных значений стока они имеют достаточно выраженный максимум при $T = 12$ месяцев (см. рис. 8.10.г), для последовательностей средних квартальных значений – максимум при $T = 4$ квартала (см. рис. 8.10.д) и т.д. Более сложными являются спектры средних годовых значений расхода воды. Они характеризуются наличием широких и узких бугров, относящихся к различным частотам (см. рис. 8.10.е).

8.6. Математические модели линейных стационарных процессов¹

Во многих случаях оказывается полезным представление геоэкологических процессов в виде выхода линейного фильтра, на вход которого поступает белый шум a_t , т. е.

$$x_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \neq a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}, \quad (8.88)$$

где $x_t = x_t - \mu$ – отклонение процесса от некоторого начального уровня или, если процесс стационарен, от среднего значения.

Другая форма представления (8.88) имеет вид:

$$x_t = \psi(B)a_t, \quad (8.89)$$

¹ Используются данные монографии [12].

где

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots, \quad (8.90)$$

линейный оператор, преобразующий a_t в x_t , называемый передаточной функцией фильтра; B – оператор сдвига назад:

$$Bx_t = x_{t-1}, \quad B^2 x_t = x_{t-2}, \quad \dots, \quad B^m x_t = x_{t-m}. \quad (8.91)$$

В математической литературе процесс (8.89) называется линейным, если a_t независимы, одинаково распределены, имеют конечную дисперсию и, кроме того, $\sum_{j=1}^{\infty} \psi_j < \infty$. Если последовательность

$\psi_1, \psi_2, \dots$ – сходящаяся, фильтр называется устойчивым, а процесс x_t будет стационарным. В других случаях процесс x_t нестационарный.

Частным видом модели линейного фильтра является модель авторегрессии (АР), которая представляется в виде регрессии текущих значений X на предшествующие:

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} + a_t. \quad (8.92)$$

В этой модели текущее значение процесса x_t линейно выражается через предыдущие и последовательность случайных величин a_t , образующую процесс типа «белого шума». Можно показать [10], что веса π в уравнении (8.93) и ψ в уравнении линейного фильтра [см. формулу (8.89)] связаны следующим соотношением:

$$\pi(B) = \psi^{-1}(B), \quad (8.93)$$

где B – оператор сдвига [см. формулу (8.92)], $\pi(B)$ – оператор авторегрессии.

Если известно, что в уравнении (8.93) только первые p весов ненулевые, то мы получаем процесс авторегрессии p порядка или сокращенно АР(P):

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \dots + \Phi_p x_{t-p} + a_t. \quad (8.94)$$

Если определить оператор авторегрессии порядка P как

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p, \quad (8.95)$$

то в сжатой форме модель АР(P) имеет вид:

$$\Phi_p(B) x_t = a_t \quad (8.96)$$

или

$$x_t = \Phi_p^{-1}(B)a_t. \quad (8.97)$$

Таким образом, процесс АР представим в двух принципиально разных формах (8.95) и (8.98). В первой – в виде регрессии текущих значений X на предшествующие, во второй – в виде выхода линейного фильтра с передаточной функцией $\Phi^{-1}(B)$, на входе которого генерируется «белый шум» a_t .

В приложениях для описания геоэкологических процессов довольно часто используются процессы авторегрессии первого ($p = 1$) и второго ($p = 2$) порядка:

$$x_t = \Phi_1 x_{t-1} + a_t; \quad (8.98)$$

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t. \quad (8.99)$$

Процесс (8.99), представляющий собой по существу простую цепь Маркова, более подробно рассмотрен в разделе 9.5.3. Процесс (8.99), представляющий собой простейший случай сложной цепи Маркова, получил распространение в геоэкологических исследованиях сравнительно недавно. Некоторые сведения по его использованию можно почерпнуть в работах [7, 17].

В общем виде модель АР(P) содержит $p + 2$ неизвестных параметра ($\mu, \sigma^2, \Phi_1, \Phi_2, \dots, \Phi_p$), которые определяются по эмпирическим данным.

Другим видом линейных моделей, имеющих большое значение при описании временных рядов, являются модели скользящего среднего (СС). Если x_t линейно зависит от конечного числа q предыдущих значений a , то такой процесс

$$x_t = a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q} \quad (8.100)$$

называется процессом скользящего среднего порядка q $CC(q)$. Оператор скользящего среднего порядка q определяется как:

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q. \quad (8.101)$$

Тогда в сжатой форме модель $CC(q)$ имеет вид:

$$x_t = \Theta(B)a_t. \quad (8.102)$$

Она содержит $q + 2$ неизвестных параметра ($\mu, \sigma^2, \Theta_1, \Theta_2, \dots, \Theta_q$), которые на практике должны оцениваться по данным наблюдений.

Процесс скользящего среднего обычно трактуется как выход линейного фильтра с передаточной функцией $\Theta(B)$, на вход которого поступает «белый шум» a_t . Можно показать, что процесс $CC(q)$ представляется в виде бесконечной взвешенной суммы предшествующих значений.

На практике довольно часто используются процессы скользящего среднего первого и второго порядка:

$$\overset{\circ}{x}_t = a_t - \Theta_1 a_{t-1}, \quad (8.103)$$

$$\overset{\circ}{x}_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2}. \quad (8.104)$$

В некоторых случаях более гибкое описание исследуемых процессов позволяет дать смешанные модели авторегрессии – скользящего среднего (АР СС), в которых в одной модели объединены и авторегрессия и скользящее среднее. Такая комбинированная модель записывается в виде:

$$\overset{\circ}{x}_t = \Phi_1 x_{t-1} + \dots + \Phi_p x_{t-p} + a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q} \quad (8.105)$$

или

$$\Phi(B) \overset{\circ}{x}_t = \Theta(B) a_t, \quad (8.106)$$

где $\Phi(B)$ и $\Theta(B)$ – операторы авторегрессии порядка p [см. формулу (8.97)] и скользящего среднего [см. формулу (8.102)]. Они имеют $p + q + 2$ неизвестных параметра ($\mu, \sigma^2, \Phi_1, \dots, \Phi_p, \Theta_1, \Theta_2, \dots, \Theta_q$), которые оцениваются по наблюдениям.

Из выражения (8.105) следует, что

$$\overset{\circ}{x}_t = \left[\frac{\Theta(B)}{\Phi(B)} \right] a_t, \quad (8.107)$$

т. е. смешанный процесс авторегрессии – скользящего среднего – может трактоваться как выход линейного фильтра с передаточной функцией в виде отношения двух полиномов и «белым шумом» a_t на входе.

Для описания нестационарных процессов, имеющих однородные флуктуации относительно некоторого изменяющегося во времени уровня, может быть использована модификация предыдущей

модели: процесс авторегрессии – проинтегрированного скользящего среднего (АР ПСС).

Нестационарный однородный процесс может быть представлен обобщенным оператором авторегрессии $\varphi(B)$, в котором один или несколько корней полинома $\varphi(B)$ равны единице, т. е.

$$\varphi(B) = \Phi(B)(1 - B)^d, \quad (8.108)$$

где $\Phi(B)$ – оператор авторегрессии стационарного процесса (8.107).

Тогда обобщенная модель однородного нестационарного процесса [см. формулу (8.107)] примет вид:

$$\varphi(B)x_t = \Phi(B)(1 - B)^d x_t = \Theta(B)a_t, \quad (8.109)$$

где

$$\Phi(B)\omega_t = \Theta(B)a_t, \quad (8.110)$$

или

$$\omega_t = \nabla^d x_t, \quad (8.111)$$

где ∇x_t – разностный оператор со сдвигом назад:

$$\nabla x_t = x_t - x_{t-1} = (1 - B)x_t. \quad (8.112)$$

При $d = 0$ и $\omega_t = x_t$ – процесс [см. формулу (8.110)] будет содержать как частные случаи стационарную комбинированную модель [см. формулу (8.105)], модель авторегрессии [см. формулу (8.94)] и скользящего среднего [см. формулу (8.100)].

Процесс АР ПСС можно получить из «белого шума» a_t с помощью трех последовательных фильтров. Первый фильтр имеет вход a_t , передаточную функцию $\Theta(B)$ и выход l_t ; второй фильтр имеет вход l_t , передаточную функцию $\Phi^{-1}(B)$ и выход W_t , третий фильтр – вход W_t , передаточную функцию $S = \nabla^{-1}$ (S – оператор суммирования) и выход x_t .

Более детально линейные процессы описаны в монографиях [12, 36, 47].

Глава 9. Численные эксперименты в геоэкологии

9.1. Методы экспериментальных исследований в экологии

В первых страницах учебника отмечалось, что до недавнего времени [см. гл. 1] в науках о Земле использовались почти исключительно два метода экспериментальных исследований: метод пассивного и метод активного эксперимента. В последние годы широкое распространение получил метод численных экспериментов.

Среди различных направлений использования численных экспериментов в исследовании природных процессов наибольшее распространение получил метод статистических испытаний или метод Монте-Карло. Во многом это объясняется универсальностью метода. Действительно, численные эксперименты методом статистических испытаний позволяют:

- дать глубокий анализ процессов, на протекание которых влияют случайные факторы;
- решать многие математические задачи, не связанные с какими-либо случайностями.

Применение этого метода в практических расчетах заключается, например, во всестороннем анализе будущего режима работы проектируемых сложных систем, невозможном при использовании только календарного ряда наблюдений (продолжительность и вероятность циклов колебаний водности, возможные экстремальные значения стока на будущий период ограниченной продолжительности и т.д.).

В научной работе метод Монте-Карло используется для оценки различных методов анализа цикличности и расчетов стока, для решения вопроса о точности тех или иных статистических характеристик по выборкам для построения математических моделей и т.д.

В геоэкологии метод статистических испытаний используется как в практических расчетах, так и в научных исследованиях пока явно недостаточно.

Поэтому в данном случае основное внимание будет уделено простейшим вопросам моделирования рядов наблюдений. Вопросы применения этих моделей для решения других задач рассматриваются в соответствующих курсах.

9.2. Общее представление о методе Монте-Карло

9.2.1. Происхождение метода

Под методом Монте-Карло обычно понимается совокупность приемов, позволяющих получать решения различных задач с помощью многократных случайных испытаний.

Название метода происходит от города в княжестве Монако, знаменитого своим игорным домом. Это объясняется тем, что необходимым элементом метода Монте-Карло является последовательность случайных чисел (см. ниже), а одним из простейших механических приборов для получения этой последовательности являлась рулетка.

Подсознательно человечество использовало идеи метода Монте-Карло с незапамятных времен, когда на основании тех или иных экспериментов (как в области азартных игр, так и в области естественных наук) с последовательностью случайных явлений делались выводы о вероятности тех или иных значений. Первые попытки использовать построения, по смыслу аналогичные методу Монте-Карло, в экологии и водохозяйственных расчетах относятся к 1927 г. Они были сделаны в работе С. Садлера (в 1946 г. С.Н. Крицкий и М.Ф. Менкель на этой же основе смоделировали годовой ход уровня Каспийского моря). Однако годом рождения метода Монте-Карло считается 1949 г., когда появилась статья Метрополиса и Улама, в которой была сформулирована принципиальная сущность метода. В Советском Союзе первые данные о методе Монте-Карло были опубликованы в 1955–1956 гг. (В.В. Чавчавадзе, Ю.А. Шрейдер, В.С. Владимиров). Некоторое отставание начала исследований в нашей стране объяснялось общим отставанием в развитии ЭВМ, а развитие метода Монте-Карло, как универсального численного метода стало возможным только благодаря появлению ЭВМ.

Впервые в применении к наукам о Земле методические основы использования метода Монте-Карло в достаточно полном виде были разработаны Г.Г. Сванидзе [36]. С появлением первых работ Г.Г. Сванидзе значимость и число исследований с использованием этого метода непрерывно стали возрастать.

9.2.2. Схема и некоторые особенности применения метода Монте-Карло

Рассмотрим схему и некоторые особенности применения метода Монте-Карло или, иначе, метода статистических испытаний, на простейшем примере вычисления площадей плоской фигуры.

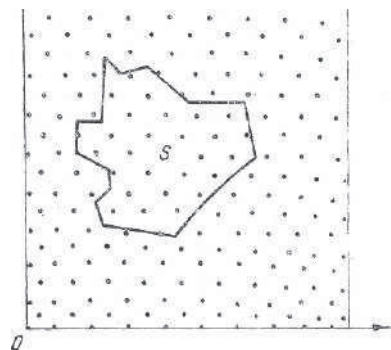


Рис. 9.1. Произвольная плоская фигура в площади единичного квадрата

Пусть произвольная плоская фигура площадью S располагается внутри единичного квадрата площадью 1 (рис. 9.1). Требуется определить площадь этой плоской фигуры S . Имеется несколько способов решения этой задачи от планиметрирования до взвешивания. Сейчас чаще используется другой оригинальный способ.

В единичном квадрате выбирается N случайных равномерно распределенных по площади квадрата точек. Обозначим через N' число точек, попавших внутрь рассматриваемой фигуры. Геометрически очевидно, что

$$\frac{N'}{N} = \frac{S}{1}, \quad (9.1)$$

где 1 – площадь единичного квадрата. Отсюда искомая площадь плоской фигуры:

$$S = \frac{N'}{N}. \quad (9.2)$$

Для практической реализации данного расчета необходимо выполнить следующий набор действий.

1. В площади единичного квадрата (рис. 9.1) случайно выбирается какая-то точка. При этом, так как точка выбирается случайно,

координаты этой точки, допустим y_1 и x_1 , являются значениями случайных величин Y и X , изменяющихся в пределах от 0 до 1. Затем случайно выбирается вторая точка, затем третья и т.д. Таким образом, можно сказать, что первое действие для решения задачи о площади плоской фигуры заключается в задании двух значений случайной величины определяющих положение точки в единичном квадрате.

2. Проверяется, принадлежит ли эта точка площади S заданной фигуры.

Первое и второе действие составляют одно случайное испытание. Перед тем как перейти к третьему действию – определению исследуемого признака – случайное испытание повторяется N раз.

3. Находится отношение числа точек, принадлежащих данной фигуре, к общему числу точек в единичном квадрате и по формуле (9.2) определяется искомая площадь S .

Изложенный алгоритм состоит из серии случайных испытаний, подвергнутых последующей статистической обработке. Отсюда другое название метода – *метод статистических испытаний*.

На приведенном примере можно выявить характерные особенности и условия применения метода Монте-Карло.

Первой его особенностью является простая структура вычислительного алгоритма.

Второй особенностью является медленная сходимость метода. Так, погрешность вычислений площадей (экспериментов вообще) методом Монте-Карло медленно уменьшается с увеличением числа испытаний. Можно показать, что эта погрешность пропорциональна корню квадратному из N [5], т. е.

$$\sigma = \sqrt{\frac{D}{N}}, \quad (9.3)$$

где D – некоторая постоянная, N – число испытаний.

Отсюда следует, что для того, чтобы уменьшить погрешность в 10 раз, нужно увеличить число испытаний в 100 раз.

Именно поэтому для применения метода статистических испытаний обычно используют ЭВМ.

Третьей особенностью метода Монте-Карло является его исключительная универсальность. Как следует из приведенного примера, он может быть использован для вычисления площадей и, по аналогичной схеме, объемов, т. е. в общем случае метод может быть использован для вычисления двойных, тройных интегралов и т.д.

Рассмотрим теперь необходимые условия применения метода. Вернемся еще раз к примеру на вычисление площадей. Пусть фигура в площади единичного квадрата на рис. 9.1 представляет собой мишень. Стрелок, находящийся на некотором расстоянии от мишени, стреляет N раз, целясь в центр мишени. Конечно, не все пули будут ложиться в центр мишени. Они пробьют на мишени N' случайных точек, распределенных относительно центра по нормальному закону. Можно ли по этим точкам оценить площадь S ? Конечно, нет, так как в этом случае получим значительное завышение площади, причем завышение будет тем больше, чем выше квалификация стрелка.

Отсюда следует важный вывод, что площади определяются правильно только в том случае, когда случайные точки (значения случайных величин) равномерно разбросаны по всей площади единичного квадрата, в котором находится рассматриваемая фигура.

Таким образом, для применения стандартной, принятой в экологии схемы метода моделирования Монте-Карло необходимо выполнение двух обязательных условий.

1) Наличие случайных чисел, равномерно распределенных в заданном интервале. Обычно принимается интервал от 0 до 1. Если границы интервала возможных значений отличаются от заданного, то они легко приводятся к нему.

2) Разработка приема определения той области (площади, отрезка кривой, точки), куда попадают случайные числа.

Числа, удовлетворяющие первому условию, называются **случайными числами, равномерно распределенными в интервале от 0 до 1**. Обычно они обозначаются α (могут быть и другие обозначения).

Пример определения площади плоской фигуры был использован для уяснения общей схемы и некоторых особенностей метода Монте-Карло. Покажем теперь, что этот же метод – метод статистических испытаний – может быть использован для моделирования различных экологических процессов. С этой целью, в качестве введения к более сложным схемам, рассмотрим моделирование наиболее простых процессов, значения которых могут быть представлены в виде последовательности значений дискретной случайной величины. Предваряя последующее изложение, заметим также, что в тех случаях, когда моделирование производится не по какому-либо конкретному ряду, а по заданным характеристикам распределения процесс моделирования часто называется **розыгрышем**.

9.3. Разыгрывание последовательности значений дискретной случайной величины

Пусть, например, перед нами поставлена задача получить ряд значений дискретной случайной величины X с распределением

$$X = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_j & \dots & \xi_k \\ p_1 & p_2 & \dots & p_j & \dots & p_k \end{pmatrix},$$

где $(\xi_j = 1, 2, \dots, k)$ – возможные значения случайной величины X , расположенные в убывающем порядке; p_j – вероятности этих значений, $\sum_{j=1}^k p_j = 1$.

Для решения этой задачи представим себе (см. пример в начале главы), что единичный квадрат, площадь которого $S_0 = 1$, разделен на k площадок, размеры которых $S_1, S_2, \dots, S_k$ заданы в долях единицы и равны соответственно вероятностям $p_1, p_2, \dots, p_k$. Выберем в единичном квадрате N случайных, равномерно распределенных точек, каждая из которых задана координатами (x, y) , представляющими собой значения случайных величин X и Y , равномерно распределенных в интервале от 0 до 1.

Если i -я точка ($i = 1, 2, \dots, N$) попала в какую-то j -ю площадку, то будем считать, что мы получили значение X , равное ξ_j , т. е. $x_i = \xi_j$. Если $i+1$ -я точка попала в какую-то ζ -ю площадку, то будем считать, что мы получили значение X , равное ξ_ζ , т. е. $x_{i+1} = \xi_\zeta$. И так далее.

В пределе при достаточно большом N распределение полученных значений $X(x_1, x_2, \dots, x_N)$ будет сходиться по вероятности к заданному распределению. Это с очевидностью следует из того, что вследствие равномерного распределения случайных точек в площади единичного квадрата число попаданий в каждую площадку при $N \rightarrow \infty$ со будет определяться ее размерами, в свою очередь равными вероятности j -го значения случайной величины.

В данном случае двумерные координаты (x, y) использовались только для уяснения аналогии и общности алгоритма метода Монте-Карло при решении различных задач. Вообще же для решения задачи розыгрыша дискретной случайной величины достаточно иметь одну числовую ось.

Подготовка к розыгрышу при этом заключается в том, что на числовой оси Y (рис. 9.2) откладывается интервал от 0 до 1, $(0 \leq Y \leq 1)$, который разбивается на k интервалов длиной, равной



Рис. 9.2. Вероятности значений случайной величины на числовой оси

соответственно $p_1, p_2, \dots, p_k$. Полученные интервалы нумеруются цифрами $j = 1, 2, 3, \dots, k$.

Сам розыгрыш заключается в следующем. Каким-либо способом (например, по таблице случайных чисел, равномерно распределенных (см. разд. 9.4) в интервале от 0 до 1) последовательно считываются значения a_i ($i = 1, 2, \dots, N$). Затем на оси Y определяется, в какой интервал попадает заданное значение точки, т. е. где $y_j = a_i$.

Если точка a_i попадает в интервал с номером j , то считается, что данное значение $x_i = \xi_j$, и т.д.

Разыгрывание дискретной случайной величины, состоящее из множества испытаний, обычно производится на ЭВМ. При этом значения случайной величины a могут быть получены различными путями (см. разд. 9.4).

Пусть распределение разыгрываемой дискретной случайной величины задано в памяти машины в виде табл. 9.1.

Таблица 9.1. Распределение дискретной случайной величины

Значения X	ξ_1	ξ_2	...	ξ_i	...	ξ_k
Вероятность значений	p_1	p_2	...	p_i	...	p_k
Обеспеченность	P_1	P_2	...	P_i	...	P_k

В этой таблице i – порядковый номер значений случайной величины X ; ξ_i – значения случайной величины, расположенные в убывающем порядке; p_i – вероятность значений ξ_i ; $P_j = P(X \geq \xi_j)$ – обеспеченность значений ξ_i .

Разыгрывание производится по следующей схеме (рис. 9.3). Задается номер члена ряда ($i = 1, 2, \dots, n$). Затем по таблице случайных чисел находится a_i , далее a_i сравнивается со значениями обеспеченности P_j ($j = 1, 2, \dots, k - 1$), и если $P_j \leq a_i \leq P_{j+1}$, то i -му члену моделируемого ряда присваивается значение ξ_j . Затем проверяется $i = n$, и если равенство выполняется, т. е. получены все n значений, то розыгрыш прекращается, если нет, то i увеличивается на 1 и весь расчет, начиная со 2-го оператора (см. рис. 9.3), повторяется.

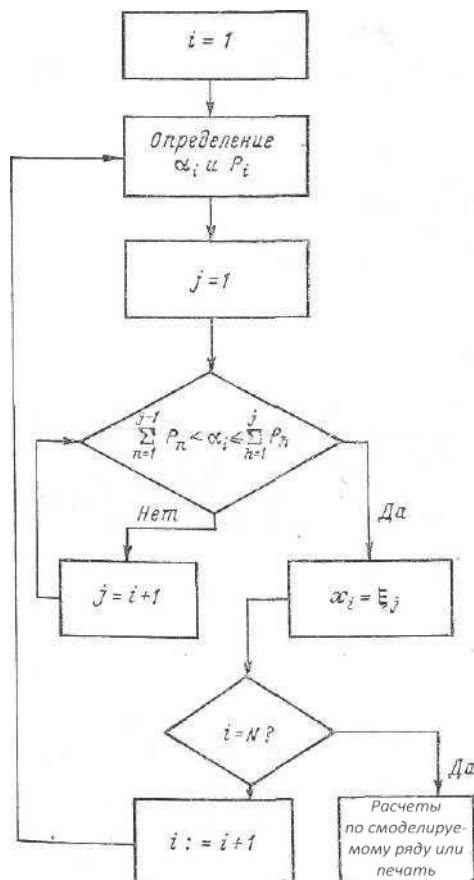


Рис. 9.3. Блок-схема розыгрыша ряда значений дискретной случайной величины

9.4. Воспроизведение последовательности значений случайной величины, равномерно распределенной в интервале $[0, 1]$

Успех применения метода Монте-Карло во многом определяется качеством случайных чисел, используемых в процессе расчета. Случайные числа a , равномерно распределенные в интервале $(0, 1)$, могут быть получены одним из трех методов: табличный метод,

метод генерирования с помощью специальных датчиков, метод генерирования программным способом. Каждый из перечисленных методов имеет свои достоинства и недостатки, и свою область применения.

Табличный метод заключается в считывании в определенном порядке (например, подряд) случайных чисел α , равномерно распределенных в интервале $(0, 1)$, из специально составленной таблицы. Имеется несколько таблиц a , разработанных как в СССР, так и за рубежом. Качество этих таблиц, т. е. соответствие требованиям, предъявляемым к случайным, равномерно распределенным числам, достаточно высокое. До последнего времени таблицы использовались главным образом при ручных вспомогательных расчетах или в учебных целях. Сейчас, в связи с увеличением оперативной памяти машин и увеличением скорости передачи данных из внешних запоминающих устройств, интерес к табличному методу снова возродился. Стоит отметить, что на первых порах сами таблицы получались с помощью рулетки, отсюда и название: метод Монте Карло.

Метод датчиков заключается в использовании специальных датчиков, формирующих случайные величины путем физического моделирования некоторых случайных процессов. В качестве датчиков могут выступать различные механические или электронные устройства. На первом этапе использования метода статистических испытаний случайные числа генерировались с помощью различных механических устройств. Например, первые таблицы равномерно распределенных случайных чисел были получены с помощью диска, приводимого во вращательное движение мотором и действующего подобно колесу рулетки.

В настоящее время равномерные случайные числа получают на некоторых ЭВМ с помощью специального приспособления – «датчика» случайных чисел, который преобразует результаты случайного физического процесса с ЭВМ (шумы электронных ламп и др.) в последовательность двоичных разрядов.

Использование такого датчика повышает скорость вычислений, так как в каждый такт работы ЭВМ в фиксированную стандартную ячейку памяти выдается новое случайное число.

Недостатком генерирования с помощью датчиков является некоторая неустойчивость случайных чисел, связанная, например, с изменением напряжения (случайные числа начинают «течь»). Вследствие этого датчики нуждаются в постоянной профилактической проверке.

Программный метод получил наибольшее распространение. Он заключается в воспроизведении случайных чисел на ЭВМ по одной из специально разработанных программ. Алгоритм этих программ представляет собой некоторое рекуррентное соотношение, по которому каждое последующее значение a_{i+1} образуется из предыдущих a и a_{i-1} путем преобразования, состоящего из арифметических и логических операций. Такая последовательность чисел может удовлетворять различным статистическим критериям случайности. В то же время каждое из чисел последовательности находится с помощью детерминированного математического выражения по предшествующему числу, т. е. не является случайным. С учетом этих двух обстоятельств для чисел, полученных программным способом, было введено название «псевдослучайные». К основным достоинствам программного способа генерирования случайных чисел помимо обстоятельств, связанных с объемом памяти ЭВМ, относится возможность контроля работы ЭВМ в процессе решения задачи (возможность проведения двойного счета), а также простота алгоритма и устойчивость результата счета.

Однако и этот метод имеет некоторые недостатки. Главным недостатком является периодичность «псевдослучайных» чисел. Правда, длина отрезка аperiodичности очень велика (10^8 – 10^9), но, тем не менее, в некоторых случаях даже с практической точки зрения очень длинные последовательности не будут случайными.

Один из наиболее простых способов генерирования псевдослучайных чисел, известный под названием способа средних квадратов, разработан Нейманом. Этот способ состоит в том, что исходное случайное число возводится в квадрат. В качестве следующего случайного числа берется последовательность цифр, стоящих по нумерации разрядов на среднем месте в вычисленном квадрате, содержащая столько же разрядов, сколько и исходное число. Затем это полученное число возводится в квадрат для получения следующего числа и т.д. Этот способ довольно прост и легко реализуется на машине. Вместе с тем он имеет следующие недостатки:

- 1) если не заменять появившиеся нулевые значения цифр в числах каким-либо другим, то вся дальнейшая последовательность быстро вырождается и в дальнейшем будет состоять из нулей;
- 2) проверка по статистическим критериям показала, что этот алгоритм дает больше чем нужно малых значений.

Поэтому были разработаны новые алгоритмы, в той или иной мере учитывающие названные недостатки.

9.5. Моделирование рядов наблюдений по различным математическим схемам

9.5.1. Понятие о моделировании рядов наблюдений

Моделирование рядов методом статистических испытаний – процесс определения последовательности возможных значений ряда по его заданным числовым характеристикам или функции распределения путем преобразования значений случайной величины a , равномерно распределенной в интервале $(0, 1)$.

Модель того или иного ряда представляет собой ряд заданной продолжительности N , характеристики которого при увеличении числа испытаний стремятся к характеристикам исходного ряда.

Характеристики исходного ряда могут задаваться разными способами. В одном случае в качестве заданных могут выступать характеристики закона распределения $\hat{F}(x)$, например, $\hat{m}_x$, $\hat{D}_x$, $\hat{C}_s$, $\hat{r}(\tau)$, и т.д., рассчитанные по конкретному числовому ряду X . При моделировании эти характеристики принимаются в качестве генеральных, т. е. принимается, что $F(x) = \hat{F}(x)$, $m_x = \hat{m}_x$, $D_x = \hat{D}_x$, $C_s = \hat{C}_s$, $r(\tau) = \hat{r}(\tau)$. Полученный в результате моделирования ряд будет представлять собой модель заданной продолжительности конкретного исходного ряда. В другом случае в качестве заданных могут выступать искусственно назначаемые, исходя из тех или иных соображений, характеристики закона распределения. В этом случае получается модель абстрактного ряда с теми или иными заданными свойствами, а сам процесс моделирования называется розыгрышем.

В зависимости от принятой математической модели исходного ряда закон распределения может быть одномерным $F_1(x)$ (случайный ряд), двумерным $F_2(x)$ (простая цепь Маркова), n -мерным $F_n(x)$ (например, сложная цепь Маркова) и т.д. Принципиальная схема моделирования во всех перечисленных случаях одинакова, однако практическая реализация ее существенно различается.

Моделирование основано на известном положении о том, что если случайная величина X имеет плотность распределения $f(x)$, то

вне зависимости от закона распределения X , распределение соответствующих X значений функции распределения

$$F(x) = \int_{-\infty}^x f(x) dx \quad (9.4)$$

является равномерным в интервале $(0, 1)$.

Отсюда процедура преобразования равномерно распределенных в интервале $(0, 1)$ случайных чисел a в случайные числа X с заданным законом распределения сводится к решению уравнения

$$a = \int_{-\infty}^x f(x) dx \quad (9.5)$$

относительно x по заданному a .

В большинстве случаев уравнение (9.5) аналитически не решается. Поэтому в практических приложениях обычно используются приближенные приемы определения X . К ним относятся:

- решение уравнения (9.5) посредством аппроксимации подынтегральной функции полиномами или другими функциями, обеспечивающими удобство определения X ;
- определение значений функции $x = f(a)$ по таблицам, содержащим заранее рассчитанные решения уравнения (9.5).

Наибольшее распространение в практике моделирования в геоэкологии получил второй прием. При этом в оперативную память ЭВМ вводится, например, таблица нормированных ординат кривых обеспеченности Пирсона III типа $t_i = f(P_i, C_s)$ или таблица модульных коэффициентов какого либо другого закона распределения $k_i = f(C_s, P, C_v)$ и т.д. Далее одним из изложенных выше методов (см. разд. 5.4) производится генерирование значений случайной величины a . Значения a принимаются за значения обеспеченности P в долях единицы ($P = a$) или в процентах (в последнем случае $P = 100a$). По значениям P_i из таблицы заданного закона распределения, например, $t_i = f(P_i, C_s)$ последовательно считываются значения t_i , которые затем по формуле (3.35) переводятся в x_i . Полученный ряд случайных значений $x_1, x_2, x_3, \dots$ будет подчиняться заданному закону распределения тем точнее, чем больше его объем.

Таким образом, имея некоторый ограниченный объем экспериментальных данных n , при заданном известном или предполагаемом законе распределения, можно получить практически неограниченную искусственную выборку объемом N , содержащую значительно

большую информацию о возможных вариантах чередования и значениях случайной величины, отвечающей данной функции распределения.

При использовании таблиц для определения значений X следует иметь в виду следующее обстоятельство. Значения P и t_i или k в таблицах заданы в дискретном виде, в то время как в действительности эти величины часто являются непрерывными. Поэтому в ходе моделирования обычно требуется интерполяция (чаще всего линейная или квадратичная) между заданными в таблице значениями, которая в общем не вызывает трудностей. Сложности возникают в том случае, когда полученные в ходе моделирования значения случайной величины выходят за интервал заданных в таблице. Так, в таблице нормированных ординат биномиальной кривой обеспеченности Пирсона III типа значения P заданы в пределах от 0,01 до 99,9 %, а при моделировании значения P могут быть получены, например, равными 0,001 %. В этом случае нет общепринятых решений. Однако имеющий опыт показал, что достаточную точность (с учетом того, что значения $P > 99,9$ и $P < 0,01$ встречаются крайне редко) можно получить путем линейной и нелинейной экстраполяции по двум-трем предшествующим табличным значениям P и X . Другое предложение [45] сводится к преобразованию интервала рассчитанных значений P (0–100 %), – обозначим через P' , в табличный интервал от 0,01 до 99,9 % – обозначим через P , с помощью линейного равенства:

$$P = \left[\frac{(d - c)}{(b - a)} \right] P' + c - a = 0,99899P' + 0,001, \quad (9.6)$$

где $(b - a)$ – исходный интервал (0, 1); $(d - c)$ – преобразованный интервал.

Однако это преобразование приводит к некоторому завышению малых обеспеченностей и уменьшению больших, а следовательно, занижению больших значений стока и завышению малых.

Возможно, более правильным было бы при моделировании принимать, что при всех $P < 0,01$ % значение $x_p = x_p = 0,01$. Частично это обосновывается тем, что значения максимальных, средних годовых и т.д. расходов для каждой реки ограничены сверху. Известно также, что на земле не наблюдался максимум обеспеченностью (по кривой Пирсона III типа) меньше 0,3 % [47].

В то же время теоретические кривые обеспеченности, используемые в экологических расчетах (см. гл. 4), за исключением кривой Джонсона, не ограничены сверху, и при очень малых P по ним

можно получить очень большие значения, не оправданные ни физическими соображениями, ни фактическими данными.

Есть еще ряд обстоятельств, которые не учитываются в современных методах моделирования. Это связано с тем, что влияние их сказывается очень редко. Но так как при моделировании методом Монте-Карло, как правило, стремятся получить выборки большого объема, названные редкие явления становятся вполне вероятными. В этой связи интересным, например, является вопрос о числе знаков после запятой в значении a и, следовательно, в обеспеченности P . Например, если в значении P , измеренного в процентах, учитывать пять знаков после запятой, то вполне вероятно, что при моделировании потребуется определять значение X обеспеченностью в одну стотысячную долю процента. Если же после запятой взять три знака, то при моделировании длинной выборки потребуется найти значение X обеспеченностью в одну тысячную долю процента. Разница между первым и вторым случаем весьма существенная.

9.5.2. Моделирование экологического ряда по схеме случайной величины

Алгоритм счета и программа моделирования случайного ряда по схеме случайной величины чрезвычайно просты и во многом согласуются со схемой розыгрыша дискретной случайной величины (см. разд. 9.2.3). Так, пусть закон распределения моделируемого ряда задан в виде таблицы значений кривой обеспеченности Пирсона III типа $k_i = f(P_i, C_s)$.

Статистические характеристики m_x , C_v и соотношение C_s/C_v моделируемого ряда известны. Тогда алгоритм моделирования по заданному ряду состоит в следующих операциях.

1) Определение последовательности a_i ($i = 1, 2, \dots, N$) (см. разд. 5.4).

2) Перевод значений a_i в значения обеспеченности P_i . Если в таблице обеспеченность задана в долях единицы, то $P_i = a_i$, если в процентах, то $P_i = 100a_i$.

3) По известным P_i , C_s , C_v с таблицы $k_i = f(P_i, C_s)$ считывается значение k_i , которое затем пересчитывается в x_i .

Эти операции повторяются N раз по заданному объему модели. Изложенный алгоритм моделирования по схеме случайного ряда достаточно легко ставится на ЭВМ.

9.5.3. Моделирование геоэкологического ряда по схеме простой цепи Маркова

В результате моделирования по схеме простой цепи Маркова должен быть получен ряд X заданной продолжительности, закон распределения которого (одномерная $\hat{F}_1(x)$ и двухмерная $F_2(x)$ функция распределения) и числовые характеристики $\hat{m}_x$, $\hat{D}_x$, $\hat{r}_1$ при увеличении числа членов ряда стремятся к заданным законам распределения $F_1(x)$, $F_2(x)$ и числовым характеристикам m_x , D_x , r_1 .

В настоящее время существует почти три десятка различных способов моделирования по схеме простой цепи Маркова. Обзор, обоснование и систематизация этих способов дан в работе Г.Г. Сванидзе [48].

В основном способы моделирования основаны на разложении членов ряда X на две составляющие:

$$x_{i+1} = \bar{x}_{i+1}(x_i) + \beta_{i+1} \quad (9.7)$$

или в модульных коэффициентах,

$$k_{i+1} = \bar{k}_{i+1}(k_i) + k_{\beta_{i+1}}, \quad (9.8)$$

где $\bar{x}_{i+1}(x_i)$ и $\bar{k}_{i+1}(k_i)$ – средние из возможных значений x_{i+1} при данном x_i или k_{i+1} при данном k_i , рассчитанные по уравнению регрессии x_{i+1} по x_i или k_{i+1} по k_i (см. разд. 7.2); β_{i+1} и $k_{\beta_{i+1}}$ случайные составляющие соответственно в исходных значениях или в модульных коэффициентах, не зависящие от предшествующих значений исследуемого ряда.

Выше отмечалось, что характеристики разложений (9.7) или (9.8) задаются на основе тех или иных соображений или рассчитываются по какому-либо исходному ряду наблюдений. В последнем случае возникают серьезные затруднения, связанные с тем, что объем имеющихся выборочных данных наблюдений недостаточен для точного определения параметров разложений.

Так, при имеющихся объемах выборок, практически невозможно установить истинную двухмерную функцию распределения $F_2(x)$, определяющую закон распределения $x_{i+1}(x_i)$ и β_{i+1} . Действительно, из анализа корреляционного поля $k_{i+1} = f(k_i)$, (рис. 9.5) следует, что при малой тесноте связи, естественной, например, для рядов среднего годового и, тем более, максимального стока, выбор какой-нибудь

поверхности плотности вероятности практически невозможен. В связи с этим остается неизвестной и форма связи $k_{i+1} = f(k_i)$.

Единственный выход из создавшейся ситуации заключается в подборе подходящих гипотез о законе распределения и форме связи:

1) двумерное распределение коррелированных величин k_{i+1}, k_i , одномерные функции распределения которых являются гамма-распределением, также являются гамма-распределением;

2) зависимость $k_{i+1} = f(k_i)$ определяется с помощью линейного уравнения регрессии.

Принимаемые гипотезы о законе распределения и форме связи, как отмечает Г.Г. Сванидзе, не могут быть непосредственно проверены на имеющемся статистическом материале. «Однако здесь на помощь приходит замечательное свойство метода Монте-Карло – возможность самопроверки»¹ [38].

При определении уравнения регрессии k_{i+1} по k_i следует иметь в виду, что аппарат корреляции двух переменных величин разработан для случая, когда их закон распределения является нормальным. Для асимметричных законов распределения этот аппарат не всегда является достаточно эффективным. Поэтому в последнее время началась разработка аппарата гамма-корреляции для связи случайных величин, подчиняющихся гамма-распределению. Имеются также предложения об использовании гамма-корреляции при моделировании рядов стока по схеме простой цепи Маркова.

Графоаналитический способ моделирования

Рассмотрим графоаналитический способ моделирования по схеме простой цепи Маркова. Он заключается в построении графика связи $k_{i+1} = f(k_i)$. Полученное поле точек разбивается на m интервалов по оси k_i . Границы интервалов выбираются таким образом, чтобы в каждый интервал входило приблизительно $(n-1)/m$ точек. Для каждого j -интервала рассчитываются статистические характеристики значений k_{i+1} , попавших в этот интервал, $\bar{k}_{i+1}$, C_s , C_v . По этим статистическим характеристикам в поле клетчатки вероятностей строятся m кривых обеспеченности (рис. 9.4). Предварительно ординаты кривых обеспеченности приводятся к единому модульному

¹ По смоделированным рядам строятся функции распределения основных параметров и сравниваются с параметрами исходного ряда наблюдений. Чем ближе выборочные характеристики к параметрам ряда наблюдений и чем меньше их рассеяние, тем лучше (в определенном смысле) метод моделирования [38].

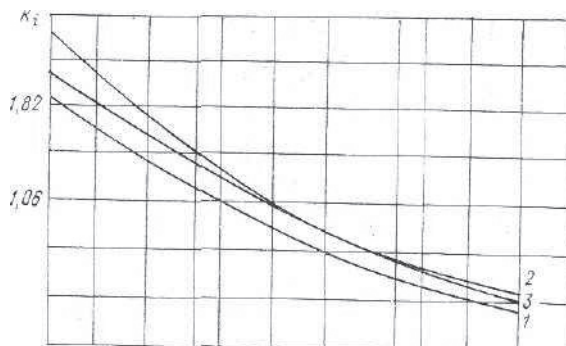


Рис. 9.4. Кривые обеспеченности значений k_{i+1} стока р. Волги у г. Старица для различных интервалов (1–3) значений k_i

коэффициенту. Для этого ординаты из таблицы нормированных ординат закона распределения Пирсона III типа умножаются на $k_{(i+1)j}$.

На основе полученных построений производится следующее моделирование.

- 1) Задается произвольное k_0 . Например, $k_0 = 0$.
- 2) По оси ординат (см. рис. 9.4) определяется, к какому интервалу значений k_i относится k_0 . Естественно, что для определения k_1 используется кривая обеспеченности этого интервала.
- 3) Считывается случайное число α_1 и умножается на 100. Полученное произведение принимается за обеспеченность первого числа P_1 .
- 4) По определенной выше кривой обеспеченности снимается значение k_1 .
- 5) По k_1 определяется (см. п. 2) интервал значений k_1 , и следовательно, номер кривой обеспеченности для следующего члена.
- 6) Находится (см. п. 3) обеспеченность следующего числа – P_α .
- 7) По определенной выше кривой обеспеченности снимается значение k_2 и т.д.

Названный прием моделирования достаточно удобен при ручном счете.

Аналитические способы моделирования

Рассмотрим способ моделирования в случае нормальной корреляции, тогда составляющая $\bar{k}_{i+1}(k_i)$ определяется по уравнению регрессии:

$$\bar{k}_{i+1}(k_i) = \frac{\bar{k}_{i+1} + r_1 \sigma_{k_{i+1}}}{\sigma_{k_i} (k_i - \bar{k}_i)}, \quad (9.9)$$

где r_1 – коэффициент корреляции смежных значений ряда; $\bar{k}_{i+1}$ и $\bar{k}_i$ – средние значения ряда X соответственно от первого до $n - 1$ -го члена и от второго до n -го члена; $\sigma_{k_{i+1}}$ и σ_{k_i} – средние квадратические отклонения ряда соответственно от первого до $n - 1$ -го члена и от второго до n -го члена.

В большинстве случаев с достаточной точностью можно принять:

$$\begin{aligned} \bar{k}_{i+1} &\approx \bar{k}_i \approx \bar{k} \approx 1, \\ \sigma_{k_{i+1}} &\approx \sigma_{k_i} \approx \sigma_k. \end{aligned} \quad (9.10)$$

Тогда:

$$\bar{k}_{i+1}(k_i) = 1 + r_1 (k_i - 1). \quad (9.11)$$

В полученном уравнении регрессии параметры k_i и r_i известны, поэтому расчет $k_{i+1}(k_i)$ не представляет трудности.

Несколько сложнее определить случайную составляющую β . Относительно нее известно, что среднее значение $\bar{\beta} = 0$, а дисперсия:

$$D_\beta = D_\Delta = D(1 - r_1^2), \quad (9.12)$$

где D – дисперсия исходного ряда, D_Δ – по существу дисперсия погрешности расчета x_{i+1} по x_i по уравнению регрессии (см. разд. 7.2.3).

Отсюда:

$$\sigma_\beta = \sigma \sqrt{1 - r_1^2} \quad (9.13)$$

и

$$C_{\nu_\beta} = \frac{\sigma_\beta}{x} = C_\nu \sqrt{1 - r_1^2}. \quad (9.14)$$

Зная числовые характеристики $\bar{\beta}_{i+1}$, C_{ν_β} , $\frac{C_s}{C_\nu}$ и значение случайного числа α , можно определить $i + 1$ -ое значение случайной составляющей:

$$\beta_{i+1} = t_{\beta_{i+1}} C_{v_{\beta}} = t_{\beta_{i+1}} C_v \sqrt{1-r^2} = k_{\beta_{i+1}} - 1, \quad (9.15)$$

где $t_{\beta_{i+1}}$ и $k_{\beta_{i+1}}$ – соответственно нормированная ордината и модульный коэффициент случайной составляющей $i + 1$ -го члена моделируемого ряда, которые определяются обычно по таблицам заданного закона распределения по полученным P_{i+1} . Например, для распределения Крицкого–Менкеля $t_{\beta_{i+1}} = f(C_{s_{\beta}}, P_{i+1})$ и $k_{\beta_{i+1}} = f(C_{v_{\beta}}, \beta_{i+1})$ в соответствии с формулами (9.8), (9.11), (9.15) получаем:

$$k_{i+1} = 1 + r_1 (k_i - 1) + k_{\beta_{i+1}} = 1 + r_1 (k_i - 1) + t_{i+1} C_v \sqrt{1-r_1^2}. \quad (9.16)$$

Эта формула обычно и используется для моделирования по схеме простой цепи Маркова при расчетах на ЭВМ. При этом P_{i+1} определяется по случайному числу α_{i+1} , полученному одним из известных способов.

Закон распределения β_{i+1} определяется исходя из гипотезы о том, что отношение $\frac{C_{s_{\beta}}}{C_{v_{\beta}}}$ равно отношению $\frac{C_s}{C_v}$. Эта гипотеза достаточно справедлива для законов распределения, близких к нормальному.

Литература

1. *Алексеев Г.А.* Методы оценки случайных погрешностей гидрометеорологической информации. – Л.: Гидрометеоиздат, 1975. – 94 с.
2. *Алексеев Г.А.* Объективные методы выравнивания и нормализации корреляционных связей. – Л.: Гидрометеоиздат, 1971. – 362 с.
3. *Алексеев Д.К., Зуева Н.В., Розенкова И.В., Урусова Е.С., Шелутко В.А.* Экологические и гидрометеорологические проблемы больших городов и промышленных зон: новые горизонты / Метеорологический вестник. – 2017. – Т. 9, № 2. – С. 1–8.
4. *Алехин Ю.М.* Статистические прогнозы в геофизике. — Л.: Изд-во ЛГУ, 1963. – 86 с.
5. *Бендат Дж., Пирсол А.* Применение корреляционного и спектрального анализа. – М.: Мир, 1983. – 312 с.
6. *Блохинов Е.Г.* Распределение вероятностей величин речного стока. – М.: Наука, 1974. – 173 с.
7. *Бокс Д., Дженкинс Г.* Анализ временных рядов, прогноз и управление. – Вып. 1. – М.: Мир, 1974. – 406 с.
8. *Болгов М.В., Сарманов И.О., Сарманов О.В.* Марковские процессы в гидрологии. – М.: ИВП РАН, 2009. – 210 с.
9. *Виноградов Ю.Б.* Прикладная гидрология. – СПб.: СПбГТУ, 2014. – 196 с.
10. Водноэнергетические расчеты методом Монте-Карло / под ред. А.М. Резниковского. – М.: Энергия, 1969. – 304 с.
11. *Гнеденко Б.В.* Курс теории вероятностей. – М.: Наука, 1969. – 400 с.
12. *Дженкинс Г., Ваттс Д.* Спектральный анализ и его приложение. – Вып. 1. – М.: Мир, 1971. – 316 с.
13. *Дмитриев В.В., Шелутко В.А.* Методологические аспекты развития прикладной экологии в системе наук о земле // Ученые записки Российского государственного гидрометеорологического университета. – 2015. – № 39. – С. 189–200.
14. *Дроздов О.А., Григорьев А.С.* Многолетние циклические колебания атмосферных осадков на территории СССР. – Л.: Гидрометеоиздат, 1971. – 151 с.
15. *Евстигнеев В.М., Магрицкий Д.В.* Речной сток. Методические основы современной практики гидрологических расчетов. – М.: Триумф, 2016. – 224 с.

16. *Ефимович П.П.* Вопросы водохозяйственных расчетов в гидрологии. – М.–Л.: Изд. ОНТИ НКТП, 1936. – 320 с.
17. *Кайсл Ч.* Анализ временных рядов гидрологических данных. – Л.: Гидрометеиздат, 1972. – 138 с.
18. *Калинин Г.П.* Проблемы глобальной гидрологии. – Л.: Гидрометеиздат, 1968. – 378 с.
19. *Картвелишвили Н.А.* Стохастическая гидрология. – Л.: Гидрометеиздат, 1975. – 164 с.
20. *Кондратьев С.А., Шмакова М.В.* Математическое моделирование в системе водосбор – водоток – водоем. – СПб.: Нестор-История, 2019. – 248 с.
21. *Константинов А.Р., Химин Н.М.* Применение сплайнов и метода остаточных отклонений в гидрометеорологии. – Л.: Гидрометеиздат, 1983. – 184 с.
22. *Крицкий С.Н., Менкель М.Ф.* Гидрологические основы управления речным стоком. – М.: Наука, 1981. – 255 с.
23. *Крицкий С.Н., Менкель М.Ф.* Расчеты речного стока. – М.–Л.: Госстройиздат, 1934. – 200 с.
24. *Крюков В.Ф.* Статистические критерии проверки гипотезы однородности двух выборок // Труды ГГИ. – 1973. – Вып. 196. – С. 97–123.
25. *Кучмент Л.С.* Математическое моделирование речного стока. – Л.: Гидрометиздат, 1972. – 190 с.
26. *Кучмент Л.С.* Речной сток (генезис, моделирование, предвычисление). – М.: Наука, 2008. – 394 с.
27. *Линслей Р.К., Колер М.А., Паулюс Д.Л.* Прикладная гидрология. – Л.: Гидрометеиздат, 1962. – 759 с.
28. *Музылев С.В., Привальсий В.Е., Раткович Д.Д.* Стохастические модели в инженерной гидрологии. – М.: Наука, 1982. – 255 с.
29. *Мякишев Т.Я.* Динамические и статистические закономерности в физике. – М.: Наука, 1973. – 272 с.
30. *Никитин С.П., Земцов В.А.* Изменчивость полей гидрологических характеристик в Западной Сибири. – Новосибирск: Наука, 1986. – 203 с.
31. Определение основных расчетных гидрологических характеристик. СП 33-101-2003. – М.: Стройиздат, 2004. – 72 с.
32. *Пановский Г.А., Брайер Г.В.* Статистические методы в метеорологии. – Л.: Гидрометеиздат, 1967. – 242 с.

33. Раткович Д.Я., Болгов М.В. Стохастические модели колебаний составляющих водного баланса речного бассейна. – М.: ИВП РАН, 1997. – 261 с.
34. Рождественский А.В., Ежов А.В., Сахарюк А.В. Оценка точности гидрологических расчетов. – Л.: Гидрометеиздат, 1990. – 273 с.
35. Рождественский А.В., Чеботарев А.И. Статистические методы в гидрологии. – Л.: Гидрометеиздат, 1974. – 424 с.
36. Рожков В.А. Теория и методы статистического оценивания вероятностных характеристик случайных величин и функций с гидрометеорологическими приложениями. – СПб.: Гидрометеиздат, 2002. – Кн. 1. – 340 с.; Кн. 2. – 440 с.
37. Румянцев В.А., Сулимов В.С. Об устойчивости эмпирических значений статистических параметров гидрологических процессов // Труды ГГИ. – 1973. – Вып. 196. – С. 3–40.
38. Сванидзе Г.Г. Математическое моделирование гидрологических рядов. – Л.: Гидрометеиздат, 1977. – 296 с.
39. Сикан А.В. Методы статистической обработки гидрометеорологической информации. – СПб.: РГГМУ, 2007. – 279 с.
40. Слуцкий Е.Е. Сложение случайных причин как источник циклических процессов. Избранные труды. – М.: Изд. АН СССР, 1960.
41. Соколовский Д.Л. Применение кривых распределения к установлению вероятных колебаний годового стока рек европейской части СССР. – Л.: Гостехиздат, 1930. – 77 с.
42. Соколовский Д.Л. Речной сток. – Л.: Гидрометеиздат, 1968. – 536 с.
43. Соколовский Д.Л., Шелутко В.А. О применимости кривой распределения Гумбела к оценке вероятности максимальных расходов воды // Труды ЛГМИ. – 1969. – Вып. 35. – С. 32–38.
44. Статистические методы в гидрологии / пер.с англ. – Л.: Гидрометеиздат, 1970. – 270 с.
45. Христофоров А.В. Надежность расчетов речного стока. – М.: Изд-во МГУ, 1993. – 168 с.
46. Чичасов Г.Н. Численные методы обработки и анализа гидрометеорологической информации. – М.: Боргес, 2013. – 235 с.
47. Шелутко В.А. Методы обработки и анализа гидрологической информации. – СПб, СПбГУ, 2007. – 192 с.

48. *Шелутко В.А.* Статистические модели и методы исследования многолетних колебаний стока. – Л.: Гидрометеиздат, 1984. – 159 с.
49. *Шелутко В.А., Колесников Е.В.* Анализ погрешности расчета средних годовых концентраций загрязняющих веществ в реках за счет неучета водности // Вестник СПбГУ. Геология. География. – Сер. 7. Вып. 3. – 2008. – С. 75–78.
50. *Шикломанов И.А.* Влияние хозяйственной деятельности на речной сток. – Л.: Гидрометеиздат, 1989. – 334 с.
51. *Шитиков В.К., Розенберг Г.С., Зинченко Г.Д.* Количественная гидроэкология: методы, критерии, решения. Кн. 1. – 2005, 281 с.

Содержание

Введение	3
--------------------	---

Глава 1. Геоэкологическая информация и ряды наблюдений

1.1. Общие понятия	10
1.2. Источники геоэкологической информации	10
1.3. Методы изучения геоэкологических процессов	12
1.4. Геоэкологические процессы и ряды наблюдений	14
1.5. Особенности геоэкологической информации	16

Глава 2. Математические модели геоэкологических процессов

2.1. Общие понятия о математических моделях	18
2.2. Структура и параметры модели	22
2.3. Идентификация математических моделей	23
2.4. Методы оптимизации	26
2.5. Типы математических моделей в геоэкологии	29
2.6. Общее обоснование применений статистических методов в геоэкологии	31

Глава 3.

Теория вероятностей и математическая статистика в экологических исследованиях

3.1. Случайные величины и их характеристики	38
3.1.1. Определения и обозначения	38
3.1.2. Дискретные и непрерывные случайные величины	40
3.1.3. Общая характеристика законов распределения	41
3.1.4. Числовые характеристики случайных величин	46
Моменты статистического распределения	47
Числовые характеристики распределения	50
3.1.5. Свойства числовых характеристик случайных величин	54
Свойства математического ожидания	54
Свойства дисперсии случайной величины	56

3.1.6. Нормированная (стандартизованная) случайная величина.	57
3.2. Системы случайных величин и их характеристики.	59
3.2.1. Общая характеристика законов распределения системы случайных величин.	59
3.2.2. Зависимость между случайными величинами	60
3.3. Математическая статистика в экологических исследованиях	62
3.3.1. Основные задачи математической статистики	62
3.3.2. Генеральная совокупность и выборка	63
3.3.3. Статистические характеристики.	65
Статистический ряд	65
Статистическая совокупность. Гистограмма (статистический многоугольник распределения).	66
Эмпирическая (статистическая) функция распределения и функция обеспеченности	68

Глава 4. Теоретические законы распределения в геоэкологических исследованиях (экстраполяция и интерполяция рядов наблюдений)

4.1. Назначение теоретических законов распределения.	73
4.2. Дифференциальное уравнение кривых плотности вероятности Пирсона	76
4.3. Нормальный закон распределения	79
4.3.1. Уравнение кривой плотности вероятности нормального закона распределения	79
4.3.2. Функции распределения и обеспеченности	82
4.4. Распределение Пирсона III типа	84
4.4.1. Уравнение кривой распределения	84
4.4.2. Запись уравнения кривой распределения через гамма-функцию	87
4.4.3. Свойства кривой распределения Пирсона III типа.	89
4.4.4. Интегральный закон распределения	90
4.5. Трехпараметрическое гамма-распределение (распределение Крицкого–Менкеля).	91
4.5.1. Уравнение кривой распределения	91
4.5.2. Свойства распределения	93
4.6. Логарифмически нормальное распределение	94

4.7. Распределение Гумбела	98
4.8. Распределение Джонсона	104
4.9. Области и особенности применения различных законов распределения	108

Глава 5. Оценка параметров законов распределения в геоэкологических исследованиях

5.1. Постановка задачи	111
5.2. Метод моментов	115
5.2.1. Оценка математического ожидания	115
5.2.2. Оценка дисперсии	116
5.2.3. Особенности применения метода моментов	118
5.2.4. Оценка случайных погрешностей выборочных числовых характеристик	119
5.3. Метод наибольшего правдоподобия	121
5.3.1. Общие положения	121
5.3.2. Оценка параметров нормального закона распределения	122
5.3.3. Оценка параметров распределения Пирсона III типа.	124
5.3.4. Особенности применения метода наибольшего правдоподобия.	125
5.4. Метод квантилей (графоаналитический метод Алексеева)	126

Глава 6. Статистическая проверка гипотез и оценка однородности геоэкологической информации

6.1. Постановка задачи	128
6.1.1. Общие соображения	128
6.1.2. Основные обозначения и определения	131
6.1.3. Уровень значимости.	133
6.2. Проверка гипотез о законе распределения	137
6.2.1. Постановка задачи	137
6.2.2. Критерий согласия Колмогорова	137
6.2.3. Критерий согласия Пирсона $P[\chi^2]$	140
6.2.4. Критерий согласия $n\omega$	142
6.3. Проверка гипотез о параметрах распределения.	144
6.3.1. Постановка задачи	144

6.3.2. Проверка гипотез о математических ожиданиях	145
Постановка задачи	145
Определение доверительных границ математического ожидания	147
Оценка значимости среднего значения.	148
Оценка расхождения между средними значениями	148
6.3.3. Оценка дисперсии	150
Постановка задачи	150
Определение доверительных границ дисперсии	151
Оценка равенства дисперсий	152
6.4. Непараметрические методы проверки гипотез	153
6.4.1. Постановка задачи	153
6.4.2. Критерий Уилкоксона	154
6.4.3. Критерий Манна–Уитни	155
6.4.4. Критерий знаков.	156

Глава 7. Статистические зависимости и линейная корреляция в геоэкологии

7.1. Статистические зависимости	157
7.1.1. Задачи изучения взаимосвязей природных процессов.	157
7.1.2. Функциональные и стохастические связи.	158
Функционально зависимые величины	158
Независимые величины	159
Вероятностная (стохастическая) зависимость.	160
7.1.3. Различные виды связи природных процессов.	162
7.2. Регрессионная математическая модель взаимосвязи двух переменных величин (парная корреляция).	163
7.2.1. Линейное уравнение регрессии двух переменных величин	163
7.2.2. Коэффициент корреляции	166
7.2.3. Оценка коэффициента корреляции и уравнения регрессии. Преобразование Фишера	168
Оценка коэффициентов корреляции и регрессии при $r < 0,4$ и $n > 30$	169
Оценка коэффициента корреляции при $r > 0,4$. Преобразование Фишера.	171
7.2.4. Анализ погрешностей расчетов по уравнению регрессии	172
Расчетные формулы	172

7.2.5. Проверка адекватности математической модели линейной корреляции на основе анализа погрешностей расчетов	175
7.2.6. Граничные условия математической модели парной корреляции.	178
7.3. Регрессионная математическая модель взаимосвязи ряда переменных величин (множественная корреляция)	182
7.3.1. Задачи исследования многофакторных связей	182
7.3.2. Уравнение множественной линейной корреляции	185
7.3.3. Сводные параметры множественной линейной корреляции.	188
7.3.4. Частный коэффициент корреляции	190
7.3.5. Оценка параметров уравнения множественной корреляции по данным наблюдений	191
7.3.6. Определение оптимального числа аргументов. Исключение дублирующих и неэффективных аргументов.	193
7.3.7. Граничные условия регрессионной математической модели взаимосвязи ряда процессов.	195
7.3.8. Алгоритм расчетов на основе регрессионной математической модели	199
7.3.9. Пример расчета на основе регрессионной математической модели	201
7.3.10. Преобразование исходной информации в целях нормализации и линеаризации взаимосвязей	204
7.3.11. Графические методы построения уравнения множественной корреляции.	208
Метод контуров.	208
Метод остаточных отклонений	210
Метод коаксиальной (соосной) корреляции.	212
Использование графической корреляции.	213

Глава 8. Численные методы анализа колебаний геоэкологических процессов во времени и пространстве

8.1. Математические модели колебания геоэкологических процессов во времени и пространстве	215
8.2. Случайные функции и их характеристики	219

8.2.1. Определение случайной функции	219
8.2.2. Законы распределения случайной функции	221
8.2.3. Числовые характеристики случайных функций	223
8.2.4. Классификация математических моделей случайных процессов.	225
Стационарные и нестационарные случайные процессы. Стационарность в широком и узком смысле	226
Эргодические и неэргодические стационарные случайные процессы.	227
Простые и сложные цепи Маркова	228
8.3. Оценка общего характера колебаний геоэкологических процессов.	229
8.3.1. Анализ стационарности, однородности и наличия тренда.	230
Оценка наличия тренда по коэффициенту корреляции.	230
Оценка тренда по критерию Спирмэна	231
Определение даты нарушения однородности	231
8.3.2. Проверка выборочных рядов на случайность	233
Проверка случайности по критерию длин и числа серий	234
Проверка случайности по числу повышений и понижений значений ряда	236
Проверка случайности по числу экстремумов.	237
8.4. Корреляционные функции геоэкологических процессов	238
8.4.1. Общие сведения	238
8.4.2. Корреляционные функции стационарных процессов	239
8.4.3. Корреляционные функции различных временных рядов.	241
8.4.4. Оценки корреляционных функций по выборочным данным.	243
8.5. Спектральный анализ геоэкологических процессов	249
8.5.1. Гармонический анализ	249
8.5.2. Спектральный анализ	253
8.5.3. Численные методы расчета спектральной плотности	254
8.5.4. Выборочная оценка спектральной плотности.	257
8.5.5. Спектральная плотность различных временных рядов.	260
8.6. Математические модели линейных стационарных процессов.	261

Глава 9. Численные эксперименты в геоэкологии

9.1. Методы экспериментальных исследований в экологии	266
9.2. Общее представление о методе Монте-Карло	267
9.2.1. Происхождение метода	267
9.2.2. Схема и некоторые особенности применения метода Монте-Карло	268
9.3. Разыгрывание последовательности значений дискретной случайной величины	271
9.4. Воспроизведение последовательности значений случайной величины, равномерно распределенной в интервале $[0, 1]$	273
9.5. Моделирование рядов наблюдений по различным математическим схемам	276
9.5.1. Понятие о моделировании рядов наблюдений	276
9.5.2. Моделирование экологического ряда по схеме случайной величины.	279
9.5.3. Моделирование геоэкологического ряда по схеме простой цепи Маркова	280
Графоаналитический способ моделирования	281
Аналитические способы моделирования	282
Литература	285

Учебное издание

В.А. Шелутко
доктор геогр. наук, профессор

МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА
ГЕОЭКОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Начальник РИО А.В. Ляхтейнен
Редактор Л.Ю. Кладова
Верстка М.В. Ивановой

Подписано в печать 30.12.2020. Формат 60×90 ¹/₁₆. Гарнитура Times New Roman.
Печать цифровая. Усл. печ. л. 18,5. Тираж 30 экз. Заказ № 1023.
РГГМУ, 192007, Санкт-Петербург, Воронежская ул., 79.
