

Королев Виктор Юрьевич
Шестаков Олег Владимирович

ВЕРОЯТНОСТНЫЕ МОДЕЛИ

Введение

В основу этой книги лег курс лекций, уже много лет читаемый авторами студентам МГУ имени М.В. Ломоносова. Идея же курса лекций возникла в результате частого общения авторов с инженерами, экономистами, медиками, юристами, физиками и специалистами в других областях, которые используют аппарат теории вероятностей и математической статистики для решения своих специфических задач.

Среди математиков за теорией вероятностей прочно закрепился статус прикладной науки. Благодаря этому, к сожалению, несмотря на большое число выдающихся математических результатов, полученных специалистами в области теории вероятностей и совершенных ими действительно фундаментальных математических открытий, ни один из них не удостоен (и, по-видимому, не будет удостоен) премии Филдса, которая вручается за выдающийся вклад в фундаментальную математику. Тем не менее, несмотря на этот статус прикладной математической дисциплины, между “академической” теорией вероятностей и практическим применением ее результатов и методов возникла огромная пропасть.

В классической теории вероятностей и математической статистике много очень математически красивых и мощных результатов. Однако каждый такой результат начинается со слов “*предположим, что...*” после которых идет перечень условий, которые должны выполняться, чтобы этот результат был справедлив. Очень часто чрезвычайно сложно или просто невозможно проверить, выполнены ли эти условия в конкретной реальной ситуации. Поэтому правильное применение результатов теории вероятностей и математической статистики на практике сопряжено с глубоким осмыслением предпосылок и реальных возможностей применения тех или иных методов.

К сожалению, этот важнейший этап очень часто игнорируется. Вместо такого анализа, во многих случаях исследователи-

нематематики просто принимают на веру многие рекомендации математиков, находясь в плену стереотипов о всемогуществе “точнойшей из наук”.

Этим обстоятельством часто вольно или невольно пользуются авторы некоторых книг или статей, в которых приводятся результаты статистической обработки тех или иных данных. Почти все наверняка неоднократно встречали в различных публикациях “заклинания” типа

- *“вычисления на компьютере показывают, что...”*,
 - *“применяя методы математической статистики, мы убеждаемся, что...”*,
 - *“согласно закону больших чисел,...”*
- или
- *“основываясь на центральной предельной теореме теории вероятностей, мы заключаем, что...”*,

где вместо многоточий приводятся удобные авторам выводы.

Такой прием рассчитан на читателя, привыкшего слепо доверять печатному слову и/или слепо верящего в непогрешимость компьютера, результатов теории вероятностей или методов статистического анализа. Для такого читателя, по сути являющегося пользователем моделей и методов теории вероятностей на уровне любителя, перечень источников моделей сводится к следующему:

- “от фонаря” (из справочника);
- “с потолка” (из Википедии);
- “из пальца” (когда-то слышал);
- “кто-то сказал” (где-то читал)

и далее в таком же роде.

В действительности компьютер является лишь орудием в руках его пользователя, равно как и подготовка данных, выбор статистического метода для их анализа и интерпретация результатов

применения этого метода всегда была, есть и будет прерогативой исследователя.

Во многом сложность выбора правильной (адекватной) вероятностной модели обусловлена тем, что в отличие от большинства объектов математического моделирования, вероятность нельзя “пощупать” или измерить непосредственно

(«– Видишь суслика?

– Нет...

– И я нет. А он есть!»).

В данной книге сделана попытка, не углубляясь в дебри математических выкладок, но сохраняя логическую строгость, ознакомить читателей с основными принципами построения математических моделей в условиях стохастической неопределенности. При этом особый акцент сделан на особенностях применения результатов и методов теории вероятностей и математической статистики к анализу случайных явлений, наблюдаемых на практике. Мы постараемся по возможности разрушить некоторые вредные стереотипы, связанные с практическим применением результатов и методов этих наук.

Почему необходимо знать принципы построения вероятностных моделей?

1. Значения параметров адекватных математических моделей статистических закономерностей можно использовать как дополнительные признаки при применении методов машинного обучения. При этом объем информации, изначально доступной исследователю, не увеличивается, а качество обучения повышается.
2. Выбор адекватной вероятностной модели крайне важен при анализе рисков, связанных с редкими событиями (отказами, сбоями, переполнениями, авариями, катастрофами). “Стандартные” модели (например, нормальное распределение) как правило недооценивают вероятности таких событий. Поэтому в таких ситуациях крайне важно уметь делать правиль-

ный выбор между “классическими” вероятностными моделями и моделями с “тяжелыми хвостами”.

3. Наконец, в большинстве задач выбор адекватной вероятностной модели важен сам по себе для принятия правильного решения.

Можно предложить следующую условную классификацию подходов к формированию вероятностных моделей.

1. **Исторический (эмпирический).** В рамках такого подхода в качестве вероятностной модели используется ее непараметрическая статистическая оценка (например, эмпирическая функция распределения).
2. **Асимптотический.** При таком подходе в модель вводится некоторый условный бесконечно малый или бесконечно большой параметр и в качестве модели берется асимптотическая аппроксимация рассматриваемой модели. Пример – нормальное распределение, являющееся предельным для сумм независимых случайных величин при неограниченном увеличении числа слагаемых.
3. **Энтропийный (информационный).** В рамках такого подхода в качестве вероятностной модели используется наиболее неблагоприятная модель, обладающая свойствами наибольшей неопределенности при заданных предположениях. По сути этот подход реализует минимаксный принцип выбора наилучшего метода в наихудших условиях.

По ходу изложения мы повторим и обсудим базовые положения теории вероятностей, основанной на аксиоматике Колмогорова, и подробно остановимся на каждом из перечисленных выше принципов построения вероятностных моделей.

Данная книга, естественно, не дает готовых рецептов на все случаи жизни. Она предназначена для заинтересованных читателей, которым может помочь самим разобраться в том, что теория вероятностей может и чего не может в их конкретных задачах, связанных с теми или иными конкретными ситуациями.

Глава 1

Основные понятия прикладной теории вероятностей

1.1 Стохастические ситуации

Окружающая нас действительность постоянно порождает неопределенные ситуации, исходы которых невозможно заранее предсказать с исчерпывающей точностью. Иногда это связано просто с недостатком информации. В таких случаях получение дополнительной информации может существенно уменьшить неопределенность и даже совсем ее устранить. Однако иногда неопределенность принципиально нельзя устранить совсем. Но даже в тех ситуациях, в которых неопределенность принципиально не устранима полностью, ее часто можно существенно уменьшить за счет лучшего понимания, уточнения самих механизмов проявления неопределенности. В частности, для этих целей можно использовать математические методы.

Математика предоставляет средства описания окружающей действительности, которые являются универсальными в том смысле, что они с одинаковым успехом могут быть использованы в самых разных областях – от физики, техники, биологии и медицины до страхования, финансов и юриспруденции. К разделам математики, изучающим механизмы проявления принципиально неустранимой неопределенности, можно отнести и теорию вероятностей.

Теория вероятностей изучает свойства математических моделей случайных явлений или процессов. Такие модели называются вероятностными. С помощью понятий и утверждений теории вероятностей можно описать сами механизмы проявления случайности и выявить закономерности в ее проявлениях.

Под *случайностью* мы будем понимать принципиально неустранимую неопределенность. Источники такой неопределенности могут носить разный характер. Принципиальность может заключаться в нежелании устранять неопределенность (например, из-за неоправданных затрат). Скажем, в игре “орел и решка” можно учесть все возможные факторы, начиная от силы и начального угла подбрасывания и заканчивая направлением ветра и магнитным полем Земли, и с точностью предсказать сторону, на которую упадет монета. Однако игроки так не поступают, и поэтому возникает неопределенность, хотя теоретически результат предreshен. Или возьмем, к примеру, социальные опросы: для выяснения отношения населения страны к какому-либо вопросу теоретически можно опросить всех жителей этой страны. Но такой подход неоправданно затратен, поэтому опрашивают некоторую выборку, и, предполагая, что эта выборка дает адекватный срез общественного мнения, делают некоторый вывод. В этих примерах мы ограничиваем себя в наборе параметров, которые измеряем.

В некоторых ситуациях эти ограничения могут быть связаны не с нежеланием идти на неоправданные затраты, а с объективным отсутствием возможностей учесть все факторы. Например, при игре на бирже мы не можем знать мотивы поведения всех участников и того, как на них влияют события, происходящие в мире. А в казино или лотереях сами организаторы прилагают все усилия, чтобы участники не могли уменьшить непредсказуемость исхода игры.

Другая причина неопределенности носит природный характер. В XIX веке господствовала теория детерминизма: считалось, что на самом деле в мире ничего случайного нет, все предreshено, а наша неспособность с точностью предсказывать события связана с тем, что мы не знаем всех начальных данных. Одним из первых ударов по этой парадигме стало появление теории относительности, из которой следует, что из процесса наблюдения нельзя устранить наблюдателя. Сам факт наличия измерения каких-либо параметров вносит изменения в процесс, и поэтому точно предсказать исход эксперимента невозможно. На это последователи теории детерминизма возражали тем, что если бы существовал “идеальный

наблюдатель”, который никак не вмешивается в процесс, то исход эксперимента был бы предрешен.

Однако с появлением квантовой механики выяснилось, что частицы могут одновременно находиться в нескольких состояниях, если над ними не производится наблюдение. То есть на этом уровне существует настоящая “природная” неопределенность. Самой известной попыткой перенести эти эффекты в макромир является мысленный эксперимент с котом Шредингера: в непрозрачной коробке находится кот, очень небольшое количество радиоактивного вещества, счетчик Гейгера, и ампула с ядом. Ампула разбивается, если счетчик регистрирует распад радиоактивного атома. Поскольку атом, если за ним не производится наблюдение, находится в суперпозиции (смешении) двух состояний, кот, сидящий в коробке, и жив, и мертв одновременно.

В действительности, этот мысленный эксперимент лишь обнаружил некоторую неполноту квантовой теории и необходимость дополнить ее более четкими определениями таких процессов, как “наблюдение”. Однако на сегодняшний день считается, что природная случайность на квантовом уровне все же существует, хотя создатель теории относительности Альберт Эйнштейн был с этим не согласен и считал, что “Бог не играет в кости”.

Любая математическая теория устроена следующим образом. Фундаментом каждой такой теории является набор аксиом, то есть не противоречащих друг другу утверждений или принципов, заведомо считающихся верными и принимаемых без доказательств. Из этих аксиом с помощью логических переходов конструируются понятия и утверждения соответствующей теории. Разные наборы аксиом ведут к разным математическим теориям, которые могут описывать одни и те же процессы и явления. При этом практическая полезность или эффективность той или иной математической теории определяется удобством ее применения и ее адекватностью, то есть степенью согласованности получаемых с ее помощью выводов со свойствами описываемой ею реальности, наблюдаемыми на практике.

Имеется довольно много математических теорий, описываю-

щих свойства математических моделей случайных явлений или процессов. В каждой из этих теорий так или иначе присутствует понятие вероятности как числового выражения меры возможности осуществления того или иного события, связанного с неопределенной ситуацией. Другими словами, имеется несколько теорий вероятностей. Мы будем иметь дело с теорией вероятностей, основанной на системе аксиом, которая была предложена в 20-х – 30-х годах XX столетия великим русским математиком Андреем Николаевичем Колмогоровым. Как правило, именно эта теория и называется собственно теорией вероятностей. За другими теориями вероятностей закреплены особые названия, например, теория субъективных вероятностей, интервальная теория вероятностей и т. п.

Не останавливаясь пока на формулировке аксиом колмогоровской теории вероятностей, опишем те свойства, которые должны быть присущи реальной неопределенной ситуации, чтобы ее можно было успешно математически описать на языке теории вероятностей. Другими словами, мы выделим те неопределенные ситуации, описание которых с помощью теории вероятностей ведет к адекватным выводам.

Назовем *стохастической* такую ситуацию, которая характеризуется следующими свойствами или условиями:

- **непредсказуемость**: исход ситуации невозможно заранее предсказать с абсолютной точностью;
- **воспроизводимость**: имеется по крайней мере теоретическая возможность воспроизвести рассматриваемую ситуацию как угодно много раз в остающихся неизменными условиях;
- **устойчивость частот**: каким бы ни было интересующее нас событие, связанное с рассматриваемой ситуацией, при многократном воспроизведении этой ситуации частота события (то есть отношение числа случаев, в которых наблюдалось рассматриваемое событие, к общему числу воспроизведений ситуации) колеблется возле некоторого числа, приближаясь

к нему все больше и больше по мере увеличения числа воспроизведений ситуации.

Свойство непредсказуемости довольно очевидно. Если исход ситуации прогнозируем однозначно, то вообще нет никакой необходимости в привлечении аппарата теории вероятностей.

Свойство воспроизводимости ситуации является ключевым для того, чтобы быть уверенным в успехе применения аппарата теории вероятностей к ее описанию. Именно это свойство имеют в виду, когда говорят, что теория вероятностей и математическая статистика направлены на изучение *массовых* явлений. В связи с условием воспроизводимости следует весьма осторожно относиться к попыткам применения теории вероятностей к анализу *уникальных* явлений или систем. Например, известны многочисленные попытки дать количественный ответ на вопрос о том, какова вероятность существования во Вселенной других планет, населенных разумными существами. Однако пока нет достаточных оснований считать, что наличие на планетах жизни, тем более разумной, является массовым явлением. Поэтому существующие прогнозы весьма разноречивы и потому неадекватны.

Наконец, свойство устойчивости частот позволяет связать математическое определение вероятности события с интуитивным представлением о ней как о понимаемом в определенном смысле пределе частоты осуществления события при неограниченном воспроизведении соответствующей ситуации.

1.2 Классические вероятностные модели

Рассмотрим следующую классическую задачу.

Задача 1.1. *Предположим, что в некотором сосуде (урне) имеется N шаров, различающихся лишь цветом: из этих шаров M – красные ($M \leq N$) и $N - M$ – белые. Из сосуда случайно извлекаются n шаров так, что каждый вынутый шар не возвращается обратно. Пусть m – некоторое целое число, $0 \leq m \leq n$. Какова вероятность того, что среди n вынутых шаров окажется ровно m красных?*

По смыслу задачи ясно, что порядок шаров в выборке безразличен. Поэтому общее число разных способов, которыми мы можем извлечь n шаров из совокупности, содержащей N шаров, то есть общее число возможных исходов ситуации, равно C_N^n . При этом все эти исходы равновероятны. Найдем число исходов, в результате каждого из которых наступит интересующее нас событие. Выбрать m шаров из имеющихся M красных можно C_M^m способами. При этом остальные $n - m$ шаров в выборке должны быть белыми. Выбрать $n - m$ шаров из имеющихся $N - M$ белых можно C_{N-M}^{n-m} способами. Ясно, что каждому набору из m красных шаров в выборке может соответствовать любой из C_{N-M}^{n-m} наборов белых шаров. Таким образом, общее число “благоприятных” исходов равно $C_M^m \cdot C_{N-M}^{n-m}$. Поскольку все возможные исходы равновероятны, в качестве искомой вероятности разумно взять отношение числа благоприятных исходов к общему числу исходов, а именно,

$$p_m = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}. \quad (1.1)$$

Казалось бы, данная задача в том виде, как она сформулирована, имеет чисто познавательный интерес. Однако, как только мы зададим конкретные значения для параметров N , M , n и m , то сразу можем заметить, что эта задача очень важна для практики. Пусть, например, $N = 49$, $M = 6$, $n = 6$. Тогда при $m = 4$, $m = 5$ и $m = 6$ с помощью формулы (1.1) мы получаем правило вычисления вероятностей угадать соответственно 4, 5 и 6 номеров в лотерее типа Спортлото “6 из 49”.

Набор чисел p_m , определяемых по формуле (1.1) для $m = 0, 1, \dots, n$, называется *гипергеометрическим распределением* вероятностей.

Задача 1.1 является примером использования так называемой *классической вероятностной модели*, которая характеризуется тем, что рассматриваемая стохастическая ситуация имеет конечное число равновероятных исходов. В такой модели под вероятностью события понимается отношение числа исходов, каждый из которых влечет осуществление этого события, к общему числу возможных исходов.

1.3 Дискретные вероятностные модели

Классические вероятностные модели приводят к самому простому и хорошо согласующемуся с интуитивными представлениями определению вероятности. Однако далеко не каждая стохастическая ситуация имеет *равновероятные* исходы, не говоря уже о том, что исходов может быть бесконечно много. Поэтому нужны более общие и более гибкие подходы к определению вероятности. Рассмотрим еще одну классическую задачу, известную как *задача о справедливом дележе ставки*. Решение этой задачи было найдено в ходе знаменитой переписки великих ученых XVII столетия Блеза Паскаля (1623-1662) и Пьера Ферма (1601-1665), в которой, как считает большинство историков науки, и были заложены основы теории вероятностей.



Blais Pascal



Pierre Fermat

Задача 1.2. Два игрока (скажем, A и B), поставив одинаковые суммы денег, играют в игру, состоящую из отдельных последовательных партий. В каждой партии возможны лишь два исхода: выигрыш игрока A или выигрыш игрока B . Вся игра считается выигранной тем из игроков, кто первым выиграет 6 партий (не обязательно подряд). Однако при счете 5:3 в пользу игрока A игра прервана и не может быть продолжена. В какой пропорции следует разделить ставку между A и B ?

Ясно, что речь идет о наиболее справедливом способе деле-

жа, ведь всего способов довольно много. Например, поровну – 1:1. Однако такой способ не устраивает первого игрока, который к моменту прекращения игры уже выиграл больше партий, чем игрок B . Можно было бы разделить ставку в отношении 1:0, отдав все игроку A , который выиграл больше партий. Но такой способ, очевидно, не устраивает игрока B , ибо такой дележ предусмотрен лишь для случая, когда весь матч закончен, в то время как матч не закончен, и игрок B все-таки выиграл 3 партии. Кажется, что дележ ставки в отношении 5:3 будет справедливым, ведь такой дележ пропорционален количеству выигранных партий. Однако и он не является идеальным, поскольку на самом деле ничем не лучше дележа в отношении 3:1, обратно пропорциональном числу партий, которые осталось выиграть игрокам до окончательной победы.

Паскаль и Ферма пришли к обоснованному выводу, что делить ставку следует пропорционально шансам игроков на окончательную победу, то есть пропорционально вероятностям того, что матч закончится победой соответствующего игрока, если бы он был продолжен. Найдем эти вероятности. Сначала опишем все возможные исходы данной стохастической ситуации. Победу игрока A в отдельной партии условно обозначим a , победу игрока B в отдельной партии обозначим b . Если первую после возобновления игры партию выигрывает игрок A , уже выигравший до этого пять партий, то весь матч (играющийся до шестой победы одного из игроков) заканчивается победой игрока A , выигравшего нужные шесть партий. Если эту партию выигрывает игрок B , то матч продолжается. Если следующую партию выигрывает игрок A , то весь матч заканчивается победой игрока A . Если же следующую партию выигрывает B , то в матче должна быть сыграна еще одна, решающая, партия. Если эту партию выигрывает A , то он побеждает в матче. Но если же эту партию выигрывает B , то весь матч заканчивается победой игрока B . Таким образом, возможные пути развития событий в матче, то есть возможные исходы рассматриваемой стохастической ситуации имеют вид a , ba , bba и bbb .

Термин “исход” здесь используется не для *результата* партии или матча, но для описания того *процесса*, который привел к тому или иному результату матча. Будем считать, что игроки одинаково

сильны в этой игре и, отложив математическую конкретизацию сказанного до следующих разделов, предположим, что результат каждой последующей партии не зависит от того, как закончились предыдущие. Тогда, очевидно,

- вероятность исхода a равна $\frac{1}{2}$,
- вероятность исхода ba равна $\frac{1}{4}$, а
- вероятности каждого из исходов bba и bbb равны $\frac{1}{8}$.

Событие “игра выиграна игроком A ” состоит из исходов a , ba и bba . В качестве вероятности этого события следует взять сумму вероятностей элементарных исходов, в результате каждого из которых наступает рассматриваемое событие. Таким образом, вероятность события “игра выиграна игроком A ” равна $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$.

В то же время, вероятность события “игра выиграна игроком B ” равна $\frac{1}{8}$. Следовательно, оптимальная пропорция, в которой следует поделить ставку, как это ни странно, составляет $7:1$.

Задача 1.2 приводит нас к следующей конструкции *дискретной вероятностной модели*. Пусть Ω – совокупность всех исходов рассматриваемой стохастической ситуации, элементарных в том смысле, что каждый из них нельзя разбить на более мелкие исходы. Под *событием* мы будем понимать любое подмножество A элементов совокупности Ω . При этом вероятность события A определим как суммарную вероятность элементарных исходов, его составляющих.

Более формально, пусть $\Omega = \{\omega_1, \dots, \omega_n\}$ и вероятность исхода ω_i равна p_i , $i = 1, \dots, n$. Пусть A – некоторое событие. Тогда вероятность события A , обозначаемая $P(A)$, определяется как

$$P(A) = \sum_{i: \omega_i \in A} p_i. \quad (1.2)$$

В (1.2) суммирование идет по всем номерам i элементарных исходов, составляющих событие A .

При этом должно выполняться естественное условие нормировки

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

Заметим, что в этой модели допускается возможность бесконечно-го числа элементарных исходов стохастической ситуации ($n = \infty$), а вероятности p_i совсем не обязаны быть равны между собой.

Классическая вероятностная модель является частным случаем дискретной модели, в котором $p_i = \frac{1}{n}$, $i = 1, \dots, n$.

1.4 Геометрические вероятностные модели

Обе вероятностные модели, рассмотренные выше, характеризуются тем, что *все* элементарные исходы соответствующих стохастических ситуаций можно занумеровать, поставив каждому исходу в соответствие натуральное число так, что разным исходам будут соответствовать разные числа. Именно это свойство позволяет определить вероятность события по формуле (1.2). Однако зачастую исходы стохастической ситуации не могут быть перенумерованы разумным образом. В качестве примера рассмотрим следующую задачу.

Задача 1.3. *Два человека, назовем их, скажем, X и Y , договорились встретиться в условленном месте между 12⁰⁰ и 13⁰⁰. Согласно обоюдной договоренности, пришедший первым ждет второго лишь в течение 20 минут (например, чтобы не привлекать излишнее внимание окружающих). Считая, что каждый из X и Y может прийти на место встречи в любой момент указанного интервала времени с одинаковой возможностью, найти вероятность того, что встреча состоится.*

Чтобы решить эту задачу, сначала построим модель множества возможных элементарных исходов данной стохастической ситуации. Совокупность возможных моментов появления X на месте встречи – это отрезок длиной 1 час. Аналогично, другой точно такой же отрезок является совокупностью моментов появления Y на

месте встречи. Отложим эти отрезки на соответствующих перпендикулярных координатных осях, резервируя горизонтальную ось за временем прихода X , а вертикальную – за временем прихода Y . Пусть x – момент появления X , y – момент появления Y . Тогда каждый исход стохастической ситуации характеризуется парой чисел (x, y) . Если мы отождествим каждую такую пару с точкой, имеющей координаты (x, y) на плоскости с указанными координатными осями, то станет ясно, что совокупность исходов данной стохастической ситуации можно представить в виде квадрата с вершинами $(0, 0)$, $(0, 1)$, $(1, 0)$ и $(1, 1)$.

Теперь найдем совокупность тех исходов, в результате каждого из которых наступает встреча X и Y . Если x отстоит от y более чем на $\frac{1}{3}$ (20 минут составляют $\frac{1}{3}$ часа), то встреча X с Y не может состояться и наоборот. Таким образом, совокупность исходов, “благоприятствующих” встрече X с Y , содержит те и только те точки с координатами (x, y) , которые, во-первых, попадают внутрь указанного квадрата и, во-вторых, удовлетворяют условию $|x - y| \leq \frac{1}{3}$. С учетом того, что по условию все исходы равновозможны, искомая вероятность будет равна отношению площади шестиугольника, задаваемого условиями $0 \leq x \leq 1$, $0 \leq y \leq 1$, $|x - y| \leq \frac{1}{3}$, и равной, как несложно подсчитать, $\frac{5}{9}$, к площади всего квадрата, равной, очевидно, единице. Таким образом, искомая вероятность равна $\frac{5}{9}$.

Подобные рассуждения вполне применимы и в других задачах, где оказывается возможным отождествить совокупность всех элементарных исходов стохастической ситуации с каким-либо одно-, дву- или трехмерным множеством и все исходы считаются *равновозможными*. Тогда вероятность события, отождествляемого с каким-либо подмножеством совокупности всех элементарных исходов, определяется как отношение длин в первом случае, площадей во втором и объемов в третьем случае. Такие вероятностные модели принято называть *геометрическими*. Эти модели являются частными случаями так называемых непрерывных вероятностных моделей, связанных с описанием поведения непрерывных случайных величин.

1.5 Структура вероятностных моделей

Во всех примерах, рассмотренных выше, каждая *вероятностная модель* той или иной стохастической ситуации имеет вид тройки объектов $(\Omega, \mathcal{A}, P)$, где Ω – совокупность всех *элементарных исходов* стохастической ситуации, каждое возможное событие является некоторым множеством элементарных исходов, то есть подмножеством множества Ω , так что $\mathcal{A}$ – совокупность всех возможных *событий* (которая в первых двух примерах совпадала с совокупностью всех подмножеств совокупности Ω), P – неотрицательная функция, областью определения которой является множество $\mathcal{A}$, а областью значений – отрезок $[0, 1]$. Каждому событию $A \in \mathcal{A}$ функция P ставит в соответствие число $P(A)$, понимаемое как вероятность события A . При этом $P(\Omega) = 1$.

Каждая такая тройка является математическим описанием (математической моделью) соответствующей стохастической ситуации. Условимся в дальнейшем отождествлять термин *вероятностная модель* с конкретной тройкой $(\Omega, \mathcal{A}, P)$.

Выбор правильной, адекватной вероятностной модели в конкретном случае очень важен и иногда очень не прост. Дело в том, что иногда стохастическая ситуация, которая, казалось бы, вполне конкретно, однозначно описывается словами, допускает несколько разных вероятностных моделей, как показывает следующий пример.

Пример 1.1. (парадокс Бертрана, см. J. Bertrand. *Calcul des Probabilités*. Paris, 1899) Рассмотрим следующую задачу. Пусть имеется окружность, в которую вписан равносторонний треугольник с вершинами, скажем, A , B и C . Спрашивается: какова вероятность того, что наугад (случайно) выбранная хорда данной окружности окажется длиннее стороны вписанного равностороннего треугольника?

Этот пример, на первый взгляд кажущийся модельным, на самом деле имеет исторические предпосылки. При кустарном производстве глиняной посуды с использованием гончарного круга на поверхности этого круга образовывались трещины, имеющие вид



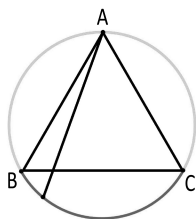
Joseph Louis François
Bertrand (1822-1900)

хорды. Если длина трещины не превосходила длины стороны вписанного равностороннего треугольника, то мастер мог ее заделать и вновь использовать гончарный круг. Если же длина оказывалась больше, то круг необходимо было менять, иначе посуда получалась бы с браком.

Рассмотрим несколько возможных способов решения этой задачи.

Первое решение. Так как хорда выбирается случайно, то каждый из ее концов может попасть в любую точку окружности. Предположим, что один из концов хорды попал в точку A , являющуюся одной из вершин треугольника. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его вершина A совместилась с одним из концов хорды (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы другой ее конец (напомним, что один из ее концов совмещен с точкой A) попал на дугу BC , соединяющую две другие вершины треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то каждый ее конец может с равными возможностями попасть в любую точку окружности. Следовательно, искомая ве-

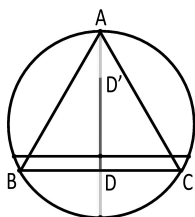
роятность будет равна отношению длины дуги BC , каждая точка которой “благоприятствует” наступлению интересующего нас события, к длине всей окружности. А так как треугольник ABC – равносторонний, то это отношение равно $\frac{1}{3}$.



Выбор точки на
окружности

Второе решение. Так как хорда выбирается случайно, то она может быть сориентирована произвольным образом по отношению к стороне BC треугольника. Предположим, что хорда оказалась параллельна этой стороне. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его сторона BC оказалась параллельна хорде (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Середина всех хорд, параллельных стороне BC , обязательно попадает на диаметр окружности, перпендикулярный стороне BC . Более того, так как треугольник ABC – правильный, то точка пересечения этого диаметра со стороной BC , скажем, точка D , делит радиус пополам. Пусть D' – точка, симметричная точке D относительно центра окружности. Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь отрезка DD' . Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку диаметра, перпендикулярного стороне BC . Следовательно, искомая вероятность будет равна отношению длины отрезка DD' , каждая точка которого “благоприятствует” наступлению интересующего нас со-

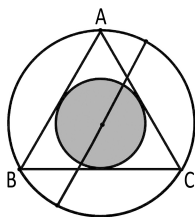
бытия, к длине всего диаметра. Но из сказанного выше вытекает, что длина отрезка DD' составляет ровно половину длины диаметра окружности. Таким образом, искомая вероятность равна $\frac{1}{2}$.



Выбор точки на диаметре

Третье решение. Так как хорда выбирается случайно, то ее середина может попасть в любую точку круга, ограниченного рассматриваемой окружностью. Используя построения, аналогичные проведенным в ходе второго решения, легко убедиться, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь окружности, вписанной в треугольник ABC , радиус которой, как несложно видеть, равен половине радиуса окружности, описанной около этого же треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку круга, ограничиваемого описанной окружностью. Следовательно, искомая вероятность будет равна отношению площади круга, ограниченного вписанной окружностью, каждой точка которого “благоприятствует” наступлению интересующего нас события, к площади круга, ограничиваемого описанной окружностью. А так как это отношение равно отношению квадратов радиусов окружностей, то искомая вероятность равна $\frac{1}{4}$.

Налицо явный парадокс: одна задача имеет сразу три решения, и каждое из этих решений, будучи полученным логически непротиворечивым путем, является верным! Однако с математической точки зрения здесь нет парадокса. Дело в том, что исходная задача была сформулирована довольно нестрого, неоднозначно, так ска-



Выбор середины
хорды

зать, некорректно. Это проявилось в том, что понятие “случайного выбора хорды” допускает несколько математических конкретизаций.

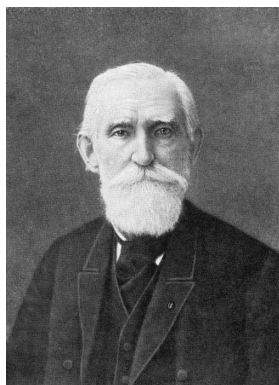
Данная стохастическая ситуация может быть описана несколькими разными вероятностными моделями.

При первом решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с окружностью, совокупность событий $\mathcal{A}$ включает в себя все дуги, а функция P , определенная на $\mathcal{A}$, каждой дуге приписывает число, равное отношению длины этой дуги к длине всей окружности.

При втором решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с диаметром окружности, совокупность событий $\mathcal{A}$ включает в себя все отрезки, лежащие на этом диаметре, а функция P , определенная на $\mathcal{A}$, каждому отрезку приписывает число, равное отношению длины этого отрезка к диаметру.

При третьем же решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с кругом, ограниченным описанной окружностью, совокупность событий $\mathcal{A}$ включает в себя все подмножества этого круга, площадь которых можно измерить, а функция P , определенная на $\mathcal{A}$, каждому подмножеству указанного круга ставит в соответствие число, равное отношению площади этого подмножества к площади круга.

В данной задаче самым сложным является адекватный выбор конкретной вероятностной модели. Такая ситуация типична



П. Л. Чебышев

для практики. Большинство задач, выдвигаемых жизнью, являются некорректными, допускают неоднозначную математическую интерпретацию.

Можно сказать, что настоящий парадокс заключается в том, что выбор адекватной *математической*, вероятностной модели часто связан с *нематематическими* обстоятельствами. Еще в XIX веке великий русский математик П. Л. Чебышев говорил, что *правильно поставить математическую задачу – значит наполовину ее решить*.

В реальной ситуации с гончарным кругом появление разного рода трещин зависит от технологии изготовления круга и материала, из которого он сделан. Интересно отметить, что древнегреческие купцы назначали цены для разного типа гончарных кругов в пропорции $3 : 2 : 4 = (1/3)^{-1} : (1/2)^{-1} : (1/4)^{-1}$.

Задачи типа парадокса Бертрана стимулировали поиск удобных аксиоматизаций теории вероятностей. Наиболее известными стали те попытки, которые связаны с именами Р. Фон Мизеса и С. Н. Бернштейна, которые в качестве отправного пункта рассматривали свойство устойчивости частот.

Однако такие системы аксиом оказались неудобными и сложно интерпретируемыми.

Наиболее удачной и удобной стала система аксиом, предложен-

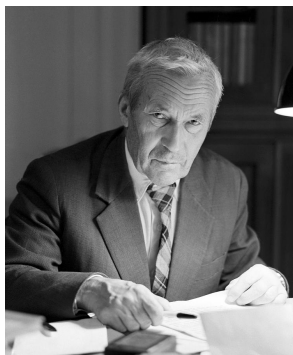


R. Von Mises



С. Н. Бернштейн

ная А. Н. Колмогоровым, который заметил, что вероятность можно трактовать как некий аналог объема, и предложил использовать систему аксиом, принятую в хорошо к тому времени разработанной теории меры, дополнив ее требованием нормированности вероятности как меры множества (события).



А. Н. Колмогоров

Пусть Ω – непустое множество, элементы которого будут обозначаться ω . Мы будем отождествлять элементы Ω с возможными элементарными, то есть неделимыми исходами некоторой стохастической ситуации. В связи с этим множество Ω будет называться *множеством элементарных исходов*.

Пусть $\mathfrak{A}$ – множество подмножеств множества Ω элементарных исходов, обладающее свойствами

- $\Omega \in \mathfrak{A}$;
- если $B \in \mathfrak{A}$, то $B^c \in \mathfrak{A}$;
- если $B_i \in \mathfrak{A}$, $i = 1, 2, \dots$, то

$$\bigcup_{i=1}^{\infty} B_i \in \mathfrak{A}, \quad \bigcap_{i=1}^{\infty} B_i \in \mathfrak{A}.$$

Множество $\mathfrak{A}$ называется σ -алгеброй событий, а его элементы (являющиеся подмножествами множества Ω) называются *событиями*.

Множество Ω вместе с σ -алгеброй его подмножеств образуют *измеримое пространство* $(\Omega, \mathfrak{A})$.

Мера, то есть неотрицательная σ -аддитивная функция множеств, P , определенная на $\mathfrak{A}$ и нормированная условием $P(\Omega) = 1$, называется *вероятностной мерой* или *вероятностью*. Напомним, что свойство σ -аддитивности, также называемое счетной аддитивностью, заключается в следующем: если $A_1, A_2, \dots$ – события ($A_i \in \mathfrak{A}$), причем $A_i \cap A_j = \emptyset$ при $i \neq j$, то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Для $B \in \mathfrak{A}$ значение $P(B)$ называется *вероятностью события B* . Тройка $(\Omega, \mathfrak{A}, P)$ называется *вероятностным пространством* или *вероятностной моделью*.

1.6 Независимость событий

Одним из основных инструментов теории вероятностей является понятие условной вероятности. Приведем пример, приводящий к этому понятию.

Пример 1.2. Пусть из общего числа N человек N_A больны, и N_B – женщины. Пусть A обозначает событие, что случайно выбранный

человек болен, а B – событие, что случайно выбранный человек – женщина. Тогда

$$P(A) = \frac{N_A}{N}, \quad P(B) = \frac{N_B}{N}.$$

Может случиться, что интерес представляет не все множество N людей, а подмножество одних женщин, и нужно найти вероятность того, что случайно выбранная женщина больна. Эта вероятность равна $\frac{N_{AB}}{N_B}$, где N_{AB} – число больных женщин.

Здесь нужны новые обозначения. Наиболее часто используется символ $P(A|B)$. В нашем примере его следует читать: “вероятность события A (болезнь) при условии, что произошло событие B (выбрана женщина)”. Итак, имеем формулу

$$P(A|B) = \frac{N_{AB}}{N_B} = \frac{N_{AB}}{N} \frac{N}{N_B} = \frac{P(A \cap B)}{P(B)}.$$

Введем формальное определение. Пусть A и B – события, причем $P(B) \neq 0$. *Условной вероятностью A при условии B* называется величина

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Можно переписать эту формулу в виде

$$P(A \cap B) = P(A|B)P(B).$$

Это так называемая теорема умножения вероятностей. Ее можно обобщить на большее число событий. Например, для трех событий A, B, C сначала теорема умножения применяется для двух событий A и $H = BC$, а затем еще раз для событий B и C . В результате получается

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C).$$

Дальнейшие обобщения на случай четырех и более событий получаются аналогично.

Понятие условной вероятности позволяет естественным образом определить понятие независимости событий. Если события A

и B независимы, то условная вероятность $P(A|B)$ равна безусловной вероятности $P(A)$, то есть наступление события B не влияет на вероятность наступления события A . Это определение можно переписать в следующем виде: события A и B , связанные с некоторой вероятностной моделью $(\Omega, \mathfrak{A}, P)$, называются *независимыми*, если вероятность их одновременного осуществления равна произведению вероятностей каждого из них:

$$P(A \cap B) = P(A)P(B).$$

В данной формуле уже не требуется, чтобы $P(B) \neq 0$.

Для большего числа событий вводится понятие независимости в совокупности.

События $A_1, A_2, \dots, A_n$ в вероятностной модели $(\Omega, \mathfrak{A}, P)$ ($n \geq 2$) называются *независимыми в совокупности*, если для любого $k \leq n$ и произвольных индексов $i_1, \dots, i_k$ ($i_p \neq i_q$ при $p \neq q$ и $1 \leq i_p \leq n$, $p = 1, \dots, k$)

$$P\left(\bigcap_{p=1}^k A_{i_p}\right) = \prod_{p=1}^k P(A_{i_p}).$$

Заметим, что понятие независимости событий тесно связано с выбранной вероятностной моделью, и независимость событий необходимо проверять исходя из приведенных выше определений, а не интуитивных представлений. Одни и те же события при разном выборе вероятностной модели могут быть как зависимыми так и независимыми. Приведем пример.

Пример 1.3. Имеется колода карт. При случайном выборе одной карты требуется проверить, являются ли события A – “выбрана карта масти пик” и B – “выбран туз” независимыми?

Ответ зависит от состава колоды.

1. Если в колоде 52 карты: по 13 карт каждой масти от двойки до туза, то $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{13}$ и $P(A \cap B) = \frac{1}{52} = P(A)P(B)$, то есть события являются независимыми.

2. Если же в колоду добавить карту “Джокер”, то $P(A) = \frac{13}{53}$, $P(B) = \frac{4}{53}$ и $P(A \cap B) = \frac{1}{53} \neq P(A)P(B)$, то есть события зависимы.

Приведем еще две важнейшие формулы, связанные с понятием условной вероятности.

В некоторой вероятностной модели $(\Omega, \mathcal{A}, P)$ рассмотрим события $A_1, A_2, \dots, A_n$ ($n \geq 2$), которые обладают следующими свойствами:

- a) события $A_1, A_2, \dots, A_n$ несовместны, то есть никакие два из них не могут произойти одновременно;
- b) одно из событий $A_1, A_2, \dots, A_n$ обязательно произойдет, то есть $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$, причем $P(A_i) > 0$, $i = 1, \dots, n$.

Если события $A_1, A_2, \dots, A_n$ обладают свойствами a) и b), то говорят, что они образуют *полную группу*.

Пусть B – некоторое событие, а события $A_1, A_2, \dots, A_n$ образуют полную группу. В таком случае справедлива формула

$$P(B) = \sum_{k=1}^n P(B|A_k)P(A_k),$$

называемая *формулой полной вероятности*.

Пусть B – некоторое событие, имеющее положительную вероятность ($P(B) > 0$), а события $A_1, A_2, \dots, A_n$ образуют полную группу. В таком случае справедлива формула

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{k=1}^n P(B|A_k)P(A_k)}, \quad k = 1, \dots, n,$$

называемая *формулой Байеса*. Формула Байеса позволяет уточнить представление о вероятности любого из событий, составляющих полную группу, с учетом информации об осуществлении некоторого события.

Рассмотрим весьма примечательную задачу.



Thomas Bayes (1702-1761)

Задача 1.4. При контроле правдивости показаний подозреваемого на “детекторе лжи” вероятность признать ложью ответ, не соответствующий действительности, равна 0.99, вероятность ошибочно признать ложью правдивый ответ равна 0.01. Известно, что ответы, не соответствующие действительности, составляют 1% всех ответов подозреваемого. Какова вероятность того, что ответ, признанный ложью, и в самом деле не соответствует действительности?

Решение. Пусть α – вероятность признать ложью ответ, соответствующий действительности, β – вероятность ошибочно признать ложью правдивый ответ, γ – доля ответов, не соответствующих действительности, среди всех ответов подозреваемого. Тогда по формуле Байеса искомая условная вероятность p равна

$$p = \frac{\alpha\gamma}{\alpha\gamma + \beta(1 - \gamma)} = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99} = \frac{1}{2}.$$

Другими словами, на самом деле лишь *половина* признанных ложью ответов не соответствуют действительности, хотя, казалось бы, выбранные значения параметров контроля α и β являются довольно жесткими.

1.7 Случайные величины и их распределения

В большинстве стохастических ситуаций каждому элементарному исходу соответствует некоторое число. Например, число угаданных номеров в лотерее, количество пострадавших при пожаре и тому подобное. Даже тогда, когда фактически числа не участвуют

в определении исходов (например, диагноза, который врач поставит своему очередному пациенту), исходам можно искусственно приписать числа (например, занумеровав все известные болезни).

Любая вещественная функция $X = X(\omega)$, определенная на Ω , отображает множество элементарных исходов во множество вещественных чисел $\mathbb{R}$. Пусть A – произвольное множество вещественных чисел: $A \subseteq \mathbb{R}$. Определим множество

$$X^{-1}(A) = \{\omega : X(\omega) \in A\},$$

являющееся подмножеством множества Ω , называемое *прообразом множества A* (при преобразовании X).

Определим *борелевскую σ -алгебру $\mathfrak{B}$* подмножеств вещественной прямой $\mathbb{R}$ как наименьшую σ -алгебру, содержащую все интервалы. Любой элемент борелевской σ -алгебры называется *борелевским множеством*. Таким образом, любое борелевское множество можно построить из интервалов с помощью операций дополнения, счетного объединения или счетного пересечения.

Если $X^{-1}(A) \in \mathfrak{A}$ для любого борелевского множества $A \in \mathfrak{B}$, то функция $X(\omega)$ называется *измеримой*.

Конечная вещественная измеримая функция называется *случайной величиной*.

Простейшим примером нетривиальной случайной величины является индикатор $\mathbf{1}_B(\omega)$ множества $B \in \mathfrak{A}$,

$$\mathbf{1}_B(\omega) = \begin{cases} 1, & \text{если } \omega \in B, \\ 0, & \text{если } \omega \notin B. \end{cases}$$

Рассмотрим события $B_1, B_2, \dots$ такие, что $B_i \cap B_j = \emptyset$ при $i \neq j$ и $\bigcup_i B_i = \Omega$. Пусть $\{x_1, x_2, \dots\}$ – вещественные числа. Случайная величина

$$X(\omega) = \sum_j x_j \mathbf{1}_{B_j}(\omega)$$

называется *дискретной*. Заметим, что свойства событий B_j гарантируют, что для каждого ω в последней сумме один и только один индикатор отличен от нуля. При этом $B_i = \{\omega : X(\omega) = x_i\}$.

Обозначим

$$p_i = P(B_i) \quad (= P(\{\omega : X(\omega) = x_i\})).$$

Набор $\{(x_i, p_i)\}_{i \geq 1}$ называется *распределением вероятностей* или просто *распределением* дискретной случайной величины X .

В дальнейшем для краткости вместо $P(\{\omega : X(\omega) \in A\})$ мы будем писать $P(X \in A)$.

Распределение дискретной случайной величины X полностью определяет вероятности попадания случайной величины X в любое борелевское множество: если $A \in \mathfrak{B}$, то

$$P(X \in A) = \sum_{i: x_i \in A} p_i.$$

Требование измеримости случайной величины как функции элементарного исхода гарантирует возможность рассмотрения множеств вида $\{\omega : X(\omega) \in A\}$ в качестве событий, какими бы ни были борелевские множества A , и, следовательно, определены вероятности попадания случайной величины в любые борелевские множества.

Таким образом, мы можем рассмотреть вероятностную меру P_X , определенную на σ -алгебре $\mathfrak{B}$ с помощью соотношения

$$P_X(A) = P(\{\omega : X(\omega) \in A\}), \quad A \in \mathfrak{B}.$$

Эта вероятностная мера называется *распределением* случайной величины X . Итак, любая случайная величина X порождает новое вероятностное пространство $(\mathbb{R}, \mathfrak{B}, P_X)$.

Пусть X – случайная величина. Рассмотрим $P(X \in A)$ в том случае, когда множество A является бесконечным интервалом вида $(-\infty, x)$, $x \in \mathbb{R}$. В таком случае положим

$$F_X(x) \equiv F(x) = P(X < x).$$

Функция $F(x)$ определена для каждого вещественного x и называется *функцией распределения* случайной величины X . Если X – дискретная случайная величина, для которой $P(X = x_i) = p_i$, то

$$F_X(x) = \sum_{i: x_i < x} p_i. \quad (1.3)$$

Функция распределения вида (1.3) называется *дискретной*.

Функция распределения $F(x)$ любой случайной величины обладает следующими свойствами:

- $F(x)$ не убывает и непрерывна слева;
- $\lim_{x \rightarrow -\infty} F(x) = 0$,
- $\lim_{x \rightarrow +\infty} F(x) = 1$.

Обратное утверждение также верно: для любой функции $F(x)$, удовлетворяющей этим трем условиям, существуют вероятностное пространство и заданная на нем случайная величина такая, что $F(x)$ является ее функцией распределения.

Каждой случайной величине соответствует одна и только одна функция распределения.

Обратное неверно. Например, на вероятностном пространстве $(\Omega, \mathfrak{A}, \mathbf{P})$, в котором $\Omega = [0, 1]$, $\mathfrak{A}$ – совокупность всех борелевских подмножеств отрезка $[0, 1]$, а $\mathbf{P}$ – мера, каждому интервалу приписывающая его длину, случайные величины $X(\omega) \equiv \omega$ и $Y(\omega) \equiv 1 - \omega$ имеют одну и ту же функцию распределения

$$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ x & \text{при } 0 \leq x \leq 1, \\ 1 & \text{при } x > 1. \end{cases}$$

Число x называется *точкой роста* функции распределения $F(x)$, если $F(x + \delta) - F(x - \delta) > 0$ для любого $\delta > 0$.

Среди всех мер, заданных на $(\mathbb{R}, \mathfrak{B})$, особую роль играет *мера Лебега*, которая каждому интервалу (a, b) приписывает его длину (очевидно, равную $b - a$), так что любое одноточечное множество имеет нулевую меру Лебега, что вполне естественно для непрерывных моделей. Кстати, практически ни в одном учебнике по теории вероятностей не обсуждается один важный вопрос, а именно, вопрос о том, почему для того, чтобы определить вероятностную модель (вероятностное пространство), непременно нужно рассматривать специальную σ -алгебру событий. Казалось бы,

было бы намного проще, если бы в качестве совокупности событий *всегда* рассматривалось множество всех подмножеств множества Ω (или $\mathbb{R}$, если вероятностная мера задается на подмножествах вещественной прямой). Однако, вообще говоря, такой подход оказывается невозможным. Еще в 1930 выдающийся польский математик С. М. Улам доказал теорему, устанавливающую, что конечная мера, заданная на множестве всех подмножеств некоторого множества мощности континуума и приписывающая нулевую меру каждому множеству, содержащему ровно один элемент (одну точку), обязана приписывать нулевую меру любым другим множествам, то есть является тривиальной (краткое доказательство теоремы Улама можно найти, например, в книге (Окстоби, 1971), глава 5).

Пусть μ – мера, заданная на измеримом пространстве $(\mathbb{R}, \mathfrak{B})$. Распределение P_X называется *абсолютно непрерывным относительно* μ , если существует неотрицательная функция $p(x)$ такая, что

$$P_X(A) = \int_A p(y) \mu(dy)$$

для любого $A \in \mathfrak{B}$ (здесь интеграл понимается в смысле Лебега). При этом функция $p(x)$ называется *плотностью распределения* или *плотностью вероятностей* или просто *плотностью* случайной величины X . Дискретное распределение $\{(x_i, p_i)\}_{i \geq 1}$ не является абсолютно непрерывным относительно меры Лебега, но является абсолютно непрерывным относительно *считающей меры*, которая каждому множеству $A \in \mathfrak{B}$ приписывает число, равное количеству тех точек $\{x_i\}$, которые попадают в A . Более того, в последнем случае

$$p(x) = P(X = x) = \begin{cases} p_i, & \text{если } x = x_i, \\ 0, & \text{если } x \neq x_i \text{ для любого } i. \end{cases}$$

Распределение, абсолютно непрерывное относительно меры Лебега, мы будем ниже называть просто *абсолютно непрерывным*. Можно показать, что распределение случайной величины X абсолютно непрерывно, если $P(X \in A) = 0$ для любого множества

$A \in \mathfrak{B}$ нулевой меры Лебега. Для абсолютно непрерывного распределения $p(x) = F'(x)$. Случайная величина, распределение которой абсолютно непрерывно, и соответствующая функция распределения также называются абсолютно непрерывными.

Дискретные и абсолютно непрерывные распределения не исчерпывают все возможные виды распределений. Существуют также функции распределения, не являющиеся дискретными, множество точек роста которых имеет лебегову меру нуль. Такие функции распределения и соответствующие им случайные величины и распределения вероятностей называются *сингулярными*. Сингулярные распределения представляют собой довольно экзотические объекты и практически не используются в прикладной теории вероятностей, теории риска или актуарной математике.

Приведем несколько общих свойств функций распределения, доказательство которых можно найти в стандартных курсах теории вероятностей или теории функций вещественной переменной.

Теорема 1.1. *Для любого $\delta > 0$ функция распределения $F(x)$ имеет не более чем конечное число точек скачков, в которых скачок превышает δ , и, следовательно, не более чем счетное число точек разрыва. Производная $F'(x)$ функции распределения $F(x)$ существует почти во всех точках x .*

Теорема 1.2. *Любая функция распределения $F(x)$ может быть однозначно представлена в виде взвешенной суммы двух компонент:*

$$F(x) = a_1 F_1(x) + a_2 F_2(x),$$

где a_1, a_2 – неотрицательные числа, сумма которых равна единице, $F_1(x)$ – непрерывная функция распределения, $F_2(x)$ – дискретная функция распределения, в точке x равная сумме величин всех скачков функции $F(x)$ в точках разрыва, не превосходящих x .

Теорема 1.3. *Любая функция распределения $F(x)$ может быть однозначно представлена в виде взвешенной суммы трех компонент:*

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x),$$

где a_1, a_2, a_3 – неотрицательные числа, сумма которых равна единице, $F_1(x)$ – абсолютно непрерывная функция распределения, $F_2(x)$ – дискретная функция распределения, в точке x равная сумме величин всех скачков функции $F(x)$ в точках разрыва, не превосходящих x , $F_3(x)$ – сингулярная функция распределения.

Если X – дискретная случайная величина и $P(X = x) > 0$, то x называется *возможным значением* случайной величины X . Случайная величина X имеет *решетчатое распределение*, если все ее возможные значения имеют вид $\{b + nh, n = 0, \pm 1, \pm 2, \dots\}$, где b и $h > 0$ – фиксированные числа. Для решетчатого распределения максимальное из чисел h называется *шагом распределения*.

Говорят, что случайная величина X имеет *симметричное распределение*, если $P(X < x) = P(-X < x)$ для любого $x \in \mathbb{R}$ или, что эквивалентно, $F_X(x) \equiv 1 - F_X(x + 0)$. Если при этом случайная величина X абсолютно непрерывна, то ее плотность $p_X(x)$ четна: $p_X(x) = p_X(-x)$, $x \in \mathbb{R}$.

Случайные величины $X_1, X_2, \dots, X_n$ называются *независимыми в совокупности*, если для любых борелевских множеств $B_1, B_2, \dots, B_n$ события $\{\omega : X_1(\omega) \in B_1\}, \{\omega : X_2(\omega) \in B_2\}, \dots, \{\omega : X_n(\omega) \in B_n\}$ независимы в совокупности.

Говорят, что случайные величины $\{X_j\}_{j \geq 1}$ образуют *последовательность независимых случайных величин*, если для любого $n \geq 1$ случайные величины $X_1, X_2, \dots, X_n$ независимы в совокупности.

Следует помнить, что независимость случайных величин нужно понимать только в смысле введенного выше определения. Часто, если между случайными величинами есть какая-то простая функциональная связь, ошибочно делается вывод об их зависимости. Это не всегда так.

Пример 1.4. Рассмотрим дискретные случайные величины X и Y , заданные на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, в котором $\Omega = [0, 1]$, $\mathfrak{A}$ – совокупность всех борелевских подмножеств отрезка

$[0, 1]$, а P – мера Лебега:

$$X(\omega) = \begin{cases} 1, & \text{при } \omega \in [0, \frac{1}{2}], \\ -1, & \text{при } \omega \in (\frac{1}{2}, 1], \end{cases}$$

$$Y(\omega) = \begin{cases} 1, & \text{при } \omega \in [0, \frac{1}{4}] \cup (\frac{1}{2}, \frac{3}{4}], \\ -1, & \text{при } \omega \in (\frac{1}{4}, \frac{1}{2}] \cup (\frac{3}{4}, 1]. \end{cases}$$

Очевидно, $X^2 + Y^2 = 2$, то есть эти две случайные величины связаны простой формулой. Однако

$$\begin{aligned} P(X = 1, Y = 1) &= P(X = 1)P(Y = 1) = 1/4, \\ P(X = 1, Y = -1) &= P(X = 1)P(Y = -1) = 1/4, \\ P(X = -1, Y = 1) &= P(X = -1)P(Y = 1) = 1/4, \\ P(X = -1, Y = -1) &= P(X = -1)P(Y = -1) = 1/4, \end{aligned}$$

то есть X и Y являются независимыми.

Если $X_1 = X_1(\omega), \dots, X_n = X_n(\omega)$ – случайные величины, определенные на одном и том же вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, то вектор $\mathbf{X} = (X_1, \dots, X_n)$ называется случайным вектором, или n -мерной случайной величиной.

Областью значений случайного вектора $\mathbf{X}$ является n -мерное евклидово пространство $\mathbb{R}^n$.

Борелевской σ -алгеброй $\mathfrak{B}_n$ подмножеств $\mathbb{R}^n$ называется минимальная σ -алгебра, содержащая все n -мерные параллелепипеды. Элементы борелевской σ -алгебры, как и ранее, будем называть борелевскими множествами.

Для каждого борелевского множества B пространства $\mathbb{R}^n$ определена вероятность $P(\mathbf{X} \in B)$. Набор $\{P(\mathbf{X} \in B) : B \in \mathfrak{B}_n\}$ называется *распределением случайного вектора $\mathbf{X}$* . В частности, для любых действительных чисел $x_1, \dots, x_n$ определена функция

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n),$$

называемая *функцией распределения случайного вектора $\mathbf{X}$* .

Случайные величины $X_1, \dots, X_n$ независимы тогда и только тогда, когда

$$F(x_1, \dots, x_n) = \prod_{k=1}^n F_k(x_k)$$

для любых действительных $x_1, \dots, x_n$. Здесь

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n) \quad \text{и} \quad F_k(x) = P(X_k < x).$$

Для любой последовательности функций распределения $F_1(x), F_2(x), \dots$ существует вероятностное пространство $(\Omega, \mathfrak{A}, P)$ и определенная на нем последовательность независимых случайных величин $X_1, X_2, \dots$ такая, что для любого n функция распределения случайной величины X_n есть $F_n(x)$.

1.8 Характеристики центра случайных величин

Часто при анализе той или иной стохастической ситуации основной интерес представляет даже не распределение вероятностей наблюдаемой случайной величины, а лишь такое число, вокруг которого концентрируются возможные значения этой случайной величины при многократном воспроизведении стохастической ситуации. Примером может служить непосредственное измерение некоторого параметра, когда вследствие случайных погрешностей каждое новое измерение отличается от предыдущего, но вместе с тем ясно, что все результаты измерения группируются вокруг некоторого “центрального” значения. Возможными синонимами понятия “центрального” значения случайной величины являются “среднее”, “ожидаемое”, “наиболее вероятное” значение. Несмотря на кажущуюся простоту, понятие “центрального” значения случайной величины является довольно тонким, и каждому из словесных описаний, приведенных выше, можно придать свое строгое математическое определение. Эти определения, вообще говоря, приводят к *разным* понятиям и, соответственно, к *разным* числам.

Определение 1.1. Пусть X – дискретная случайная величина, принимающая значения $x_1, x_2, \dots$ с вероятностями соответственно $p_1, p_2, \dots$. Математическим ожиданием дискретной случайной величины называется число

$$EX = x_1 p_1 + x_2 p_2 + \dots \quad (1.4)$$

Если ряд (1.4) расходится, то говорят, что математическое ожидание не существует.

Пример дискретной случайной величины, у которой математическое ожидание не существует:

$$P(X = \pm k) = \frac{3}{\pi^2 k^2}, \quad k = 1, 2, \dots$$

$$\left(\frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} = 1, \text{ но ряд } \frac{3}{\pi^2} \sum_{k=-\infty}^{\infty} \frac{1}{k} \text{ расходится.} \right)$$

Смысл математического ожидания становится ясным, если мы предположим, что случайная величина может с равными вероятностями принять одно из конечного числа значений $x_1, \dots, x_n$. Тогда $p_k = \frac{1}{n}$, $k = 1, \dots, n$, и математическое ожидание такой случайной величины совпадает со средним арифметическим чисел $x_1, \dots, x_n$:

$$EX = \frac{1}{n} \sum_{k=1}^n x_k.$$

Таким образом, мы можем заключить, что в общем случае математическое ожидание дискретной случайной величины, определенное формулой (1.4), является обобщением понятия среднего арифметического на случай (возможно, бесконечного числа) неравновероятных возможных значений.

Определение 1.2. Пусть X – абсолютно непрерывная случайная величина, имеющая плотность $p_X(x)$. Математическим ожиданием абсолютно непрерывной случайной величины называется число

$$EX = \int_{-\infty}^{\infty} x p_X(x) dx. \quad (1.5)$$

Если в (1.5) интеграл расходится, то говорят, что математическое ожидание не существует.

Пример абсолютно непрерывной случайной величины, у которой математическое ожидание не существует: пусть $a \in \mathbb{R}$,

$$p_X(x) = \frac{1}{\pi(1+(x-a)^2)}, \quad x \in \mathbb{R}.$$

Тогда плотность $p_X(x)$ имеет максимум в точке a , но интеграл

$$\int_{-\infty}^{\infty} \frac{x dx}{\pi(1+(x-a)^2)}$$

расходится.

В общем случае математическое ожидание произвольной случайной величины X с функцией распределения $F_X(x)$ можно записать в виде *интеграла Лебега–Стилтьеса*

$$EX = \int_{-\infty}^{\infty} x dF_X(x).$$

Основным свойством математического ожидания можно считать его линейность: если a и b – два числа, а X и Y – две случайные величины, то

$$E(aX + bY) = aEX + bEY$$

при условии, что любые два из участвующих математических ожиданий существуют.

Отсюда, в частности, вытекает, что

- ★ математическое ожидание суммы случайных величин равно сумме математических ожиданий этих случайных величин;
- ★ неслучайный коэффициент при случайной величине можно выносить за знак математического ожидания.

Перечислим некоторые другие свойства математического ожидания.

- Пусть X – случайная величина, A – некоторое подмножество ее значений. Пусть $\mathbf{1}_A(x)$ – индикаторная функция множества A . Тогда $\mathbf{1}_A(X)$ – случайная величина, принимающая значение 1, если $X \in A$, и значение 0, если $X \notin A$, причем

$$P(\mathbf{1}_A(X) = 1) = P(X \in A) = 1 - P(\mathbf{1}_A(X) = 0).$$

Тогда в соответствии с определением математического ожидания дискретной случайной величины мы имеем

$$E\mathbf{1}_A(X) = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) = P(X \in A).$$

- Если случайные величины X и Y независимы, то

$$EXY = EX \cdot EY.$$

Обратное утверждение верно не всегда.

Смысл математического ожидания как центрального значения случайной величины поясняет следующая задача.

Задача 1.5. *Предположим, что некий человек на перемещение из точки a в точку b затрачивает усилия, равные $(b - a)^2$. Это означает, что он легко перемещается на малые расстояния, но большие перемещения даются ему с большим трудом. Скажем, мы рассматриваем пожилого немощного человека. Пусть X – случайная величина, характеризующая координату точки, в которую необходимо переместиться этому пожилому человеку (скажем, он может пойти или поехать в поликлинику, в собес, в магазин, к внукам и т. п.). Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?*

Решение. Интуитивно ясно, что если этот человек будет жить в “средней”, “центральной” точке между возможными конечными пунктами своих перемещений, то его усилия в среднем будут невелики. Подкрепим эту догадку строгими рассуждениями. Средние

усилия, затрачиваемые этим человеком на перемещение из случайной точки x в случайную точку X , имеют вид

$$E(X - x)^2 = E(X^2 - 2xX + x^2) = x^2 - 2xE(X) + E(X^2).$$

Правая часть этого выражения как функция от x представляет собой хорошо знакомую квадратичную параболу. Ее минимум находится в точке $x_0 = EX$. Таким образом, математическое ожидание характеризует центр распределения случайной величины X (или, в терминах данной задачи, определяет “среднюю” или “центральную” точку населенного пункта) в смысле квадратичной функции потерь.

Определение 1.3. Пусть q – некоторое число из интервала $(0, 1)$. Квантилью порядка q (или q -квантилью) случайной величины X называется число x_q такое, что одновременно

$$P(X \leq x_q) \geq q \text{ и } P(X \geq x_q) \geq 1 - q.$$

Определение 1.4. Медианой случайной величины X называется ее квантиль порядка $\frac{1}{2}$.

Медиана случайной величины X обозначается $\text{med}X$. По определению медианы, одновременно выполняются неравенства

$$P(X \leq \text{med}X) \geq \frac{1}{2} \text{ и } P(X \geq \text{med}X) \geq \frac{1}{2}.$$

Другими словами, медиана случайной величины X характеризует “центр” ее распределения в том смысле, что случайная величина X может с *равными возможностями* быть как меньше медианы, так и больше нее.

Для $q \in (0, 1)$ квантиль x_q *непрерывной* случайной величины можно вычислить как корень уравнения

$$F_X(x) = q,$$

где $F_X(x)$ – функция распределения случайной величины X .

Таким образом, медиана *непрерывной* случайной величины является корнем уравнения

$$F_X(x) = \frac{1}{2}.$$

Смысл понятия медианы поясняет следующая задача, являющаяся в некотором смысле дополнением к Задаче 1.5.

Задача 1.6. *Предположим, что некий человек на перемещение из точки a в точку b затрачивает усилия, равные $|b - a|$. Это означает, что его затраты на перемещение пропорциональны длине пути. Скажем, мы рассматриваем делового человека, для которого фактор времени имеет решающее значение и который может перемещаться в любую точку с постоянной скоростью. Пусть X – случайная величина, характеризующая координату точки, куда необходимо переместиться этому человеку. Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?*

Сведем **решение** этой задачи к доказательству неравенства

$$E|X - \text{med}X| \leq E|X - d|, \quad (1.6)$$

справедливого для любого числа d . Это будет означать, что решением задачи является медиана случайной величины X .

Сначала рассмотрим случай $d > \text{med}X$. Тогда

$$|X - d| - |X - \text{med}X| = \begin{cases} \text{med}X - d & \text{при } X \in [d, \infty), \\ d + \text{med}X - 2X & \text{при } X \in (\text{med}X, d), \\ d - \text{med}X & \text{при } X \in (-\infty, \text{med}X]. \end{cases}$$

Так как $d + \text{med}X - 2X > \text{med}X - d$ при $\text{med}X < X < d$, то

$$\begin{aligned} E|X - d| - E|X - \text{med}X| &= E[|X - d| - |X - \text{med}X|] \geq \\ &\geq (\text{med}X - d)E\mathbf{1}_{[d, \infty)}(X) + \\ &\quad + (\text{med}X - d)E\mathbf{1}_{(\text{med}X, d)}(X) + \end{aligned}$$

$$\begin{aligned}
& +(d - \text{med}X)E\mathbf{1}_{(-\infty, \text{med}X]}(X) = \\
& = (\text{med}X - d)P(X \geq d) + \\
& \quad + (\text{med}X - d)P(\text{med}X < X < d) + \\
& \quad + (d - \text{med}X)P(X \leq \text{med}X) = \\
& = (d - \text{med}X)[P(X \leq \text{med}X) - P(X > \text{med}X)] \geq 0.
\end{aligned}$$

Последнее неравенство вытекает из определения медианы. Но это и означает справедливость (1.6) для случая $d > \text{med}X$. Аналогично рассматривается случай $d < \text{med}X$.

Таким образом, медиана характеризует центр распределения случайной величины X (определяет центр населенного пункта в терминах Задачи 1.6) в смысле функции потерь, имеющей вид абсолютной величины.

В отличие от математического ожидания, медиана всегда определена, но не всегда единственна. Поясним сказанное на примере случайной величины X , принимающей значения 0 и 1 с равными вероятностями. Тогда, как легко видеть, определению медианы удовлетворяет любая точка отрезка $[0, 1]$ (включая концы). В то же время, очевидно, $EX = \frac{1}{2}$.

Определение 1.5. Модой абсолютно непрерывной случайной величины X , имеющей плотность распределения $p_X(x)$, называется число $\text{mod}X$, удовлетворяющее условию

$$p_X(x) \leq p_X(\text{mod}X)$$

для любого x .

Другими словами, мода абсолютно непрерывной случайной величины X – это такая точка, в которой плотность распределения случайной величины X принимает наибольшее значение.

Определение 1.6. Модой дискретной случайной величины X , принимающей значения $x_1, x_2, \dots$, называется такое ее значение x_{k_0} , которое удовлетворяет условию

$$P(X = x_k) \leq P(X = x_{k_0})$$

для любого $k = 1, 2, \dots$

В общем случае распределение P_X случайной величины X называется *унимодальным*, если существует значение $x = a$ такое, что при $x < a$ функция распределения $F_X(x)$ выпукла, а при $x > a$ – вогнута. При этом число a называется *модой* распределения.

Из этих определений вытекает, что моду случайной величины X можно интерпретировать как наиболее вероятное ее значение.

Отметим, что мода не всегда единственна. Если распределение случайной величины мультимодально, то, как правило, это говорит о том, что описываемые такой случайной величиной данные неоднородны.

Каждая из введенных выше числовых характеристик – математическое ожидание, медиана и мода – описывает “центральный” тенденцию в поведении случайной величины. Однако эти характеристики совсем не обязаны совпадать. Чтобы в этом убедиться, рассмотрим следующую весьма поучительную задачу.

Задача 1.7. *Предположим, что вероятность того, что среднегодовой доход наугад выбранного жителя некоторой страны превосходит уровень x , равна $e^{-\lambda x}$, где $\lambda > 0$. Другими словами, доля индивидуумов, чей среднегодовой доход превосходит уровень x , во всем населении страны, составляет $100 \cdot e^{-\lambda x}\%$. Каков средний доход жителей этой страны?*

Эта кажущаяся очень простой задача весьма знакома. Однако на самом деле эта задача не является тривиальной, но все затруднения не являются математическими. Главная особенность связана с интерпретацией понятия среднего дохода.

Решение. Пусть X – случайная величина, равная среднегодовому доходу наугад выбранного жителя. Тогда $F_X(x) = P(X < x) = 1 - e^{-\lambda x}$. Такая функция распределения соответствует абсолютно непрерывной случайной величине с плотностью вероятностей $p_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Параметр λ влияет лишь на выбор единиц измерения. Поэтому, не ограничивая общности, выберем такую “условную единицу”, которая соответствует значению $\lambda = 1$. Тогда

$$EX = \int_0^{\infty} x e^{-x} dx = 1.$$

Медиана случайной величины X находится как решение уравнения $1 - e^{-x} = \frac{1}{2}$ и оказывается равной

$$\text{med}X = \ln 2 \approx 0.693.$$

В то же время, мода случайной величины X , то есть та точка, в которой плотность $p_X(x) = e^{-x}$ максимальна, равна нулю, то есть

$$\text{mod}X = 0.$$

Таким образом, если мы по какой-либо причине желаем приукрасить ситуацию (скажем, готовя отчет правительства перед парламентом), но *не хотим подтасовывать данные*, то, конечно же, мы будем характеризовать средний доход как математическое ожидание, которое в данном случае почти в полтора раза (весьма существенно!) превосходит медиану. Если же мы, к примеру, представляем ультра-радикальную оппозицию, то на митингах мы будем, опять-таки, *не подтасовывая данные*, говорить о среднем доходе, имея в виду моду случайной величины X .

В определении математического ожидания в равной степени учитываются как вероятности значений случайной величины, так и *сами значения*. Поэтому наличие одного-двух мультимиллиардеров довольно заметно повышает средний доход граждан страны, характеризуемый математическим ожиданием. В определении же медианы, грубо говоря, величина больших значений рассматриваемой характеристики не играет большой роли. Важно лишь их количество. Поэтому наличие нескольких мультимиллиардеров практически не влияет на средний доход граждан страны, характеризуемый медианой.

К примеру, если в некотором населенном пункте живет 10001 человек, из которых 10000 имеет доход 1000 условных единиц, а один житель имеет доход 10^9 условных единиц, тогда медиана дохода наугад выбранного жителя этого населенного пункта, очевидно, равна 1000 условных единиц, тогда как математическое ожидание дохода наугад выбранного жителя этого городка равно

$$\frac{10000}{10001} \cdot 10^3 + \frac{1}{10001} \cdot 10^9 \approx 10^3 + 10^5$$

условных единиц, то есть практически в 100 раз больше дохода наугад выбранного жителя этого городка, характеризующего его медианой. Поэтому если мы хотим дать более взвешенный и объективный ответ (скажем, готовя аналитическую записку для администрации президента), то будем характеризовать средний доход медианой.

Помимо введенных выше характеристик “центра” распределения случайной величины (математическое ожидание, медиана, мода), которые мы будем условно называть *структурными* средними, в эконометрических задачах традиционно используют так называемые *Q-средние значения*, примерами которых являются среднее арифметическое, среднее гармоническое и среднее геометрическое.

Все *Q-средние значения* определяются единообразно. А именно, пусть X – положительная случайная величина, а $Q(x)$ – некоторая функция, имеющая обратную функцию $Q^{-1}(x)$ (если $y = Q(x)$, то $x = Q^{-1}(y)$). С помощью функции $Q(x)$ и операции вычисления математического ожидания *Q-среднее значение* случайной величины X , обозначаемое $M_Q X$, определяется как

$$M_Q X = Q^{-1}(EQ(X)).$$

В частности, если $Q(x) \equiv x$, то $Q^{-1}(x) \equiv x$ и мы получаем “обобщенное среднее арифметическое”

$$M_Q X \equiv m_A(X) = EX.$$

Если $Q(x) \equiv 1/x$, то $Q^{-1}(x) \equiv 1/x$ и мы получаем “обобщенное среднее гармоническое”

$$M_Q X \equiv m_H(X) = \frac{1}{E[1/X]}.$$

Если $Q(x) \equiv \log_a x$, то $Q^{-1}(x) \equiv a^x$ и мы получаем “обобщенное среднее геометрическое”

$$M_Q X \equiv m_G(X) = a^{E \log_a X}.$$

Несложно убедиться, что, если случайная величина X принимает положительные значения $x_1, \dots, x_n$ с равными вероятностями $P(X = x_k) = \frac{1}{n}$, $k = 1, \dots, n$, то мы соответственно получаем традиционные средние арифметическое, гармоническое и геометрическое:

$$m_A(X) = \frac{1}{n} \sum_{k=1}^n x_k, \quad m_H(X) = \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{x_k}}, \quad m_G(X) = \sqrt[n]{x_1 \cdot \dots \cdot x_n}.$$

Справедливы следующие соотношения между средним арифметическим, средним гармоническим и средним геометрическим:

$$m_H(X) \leq m_G(X) \leq m_A(X).$$

1.9 Характеристики разброса случайных величин

Наряду с характеристикой “центра” распределения вероятностей случайной величины, определенную информацию об этом распределении несет параметр, характеризующий степень разброса возможных значений случайной величины вокруг “центрального” значения. Как и “центр”, разброс могут описывать различные характеристики.

Определение 1.7. Дисперсией случайной величины X называется число DX , определяемое как

$$DX = E(X - EX)^2.$$

Дисперсия определена только для тех случайных величин X , для которых EX^2 существует.

Дисперсия характеризует “центр” распределения квадратов отклонения случайной величины X от ее “центра” при многократном воспроизведении стохастической ситуации, с которой связана случайная величина X .

Дисперсию иногда бывает удобно вычислять по формуле

$$DX = EX^2 - (EX)^2.$$

Перечислим основные свойства дисперсии.

- Неслучайный коэффициент можно выносить из-под знака дисперсии, возведя его в квадрат:

$$D(cX) = c^2 \cdot DX.$$

- Дисперсия любой случайной величины неотрицательна.
- Дисперсия случайной величины равна нулю тогда и только тогда, когда эта случайная величина всегда (с вероятностью единица) принимает одно и то же значение (естественно, равное ее математическому ожиданию).

Для унификации единиц измерения степени разброса с единицами измерения самой случайной величины используют понятие *среднеквадратичного отклонения*, определяемого как квадратный корень из дисперсии: $\sigma = \sqrt{DX}$.

Несложно видеть, что

$$DX = E(X - EX)^2 = \min_a E(X - a)^2$$

(см. Задачу 1.5). Поэтому, вспоминая Задачу 1.6, мы заключаем, что наряду с дисперсией степень разброса значений случайной величины может характеризовать также такой параметр как

$$E|X - \text{med}X| = \min_a E|X - a|.$$

Однако такая характеристика используется редко из-за отсутствия у нее удобных аналитических свойств.

В описательной статистике довольно часто используется еще одна характеристика степени разбросанности возможных значений случайной величины.

Определение 1.8. Пусть $x_{1/4}$ и $x_{3/4}$ – квантили случайной величины X порядков $\frac{1}{4}$ и $\frac{3}{4}$ соответственно. Разность $R_X = x_{3/4} - x_{1/4}$ называется интерквартильным размахом случайной величины X .

Можно сказать, что интерквартильный размах в той же мере характеризует степень разброса возможных значений случайной величины, в какой медиана характеризует ее центральное значение: внутри интервала $[x_{1/4}, x_{3/4}]$ случайная величина X попадает с вероятностью $\frac{1}{2}$. При этом среди всех возможных интервалов с таким свойством (легко видеть, что все такие интервалы имеют своими концами квантили случайной величины порядков $q - \frac{1}{4}$ и $q + \frac{1}{4}$ при $\frac{1}{4} \leq q \leq \frac{3}{4}$) интервал $[x_{1/4}, x_{3/4}]$ занимает “срединное” положение, поскольку вероятность того, что случайная величина X попадет левее этого интервала равна вероятности того, что случайная величина X попадет правее этого интервала. Поэтому длина этого интервала вполне может быть принята в качестве меры разброса случайной величины.

1.10 Неравенства для характеристик центра и разброса случайных величин

Неравенство Маркова. Пусть X – неотрицательная случайная величина. Тогда для произвольного положительного числа x мы можем записать:

$$\begin{aligned} E \min\{x, X\} &= E[X\mathbf{1}_{[0,x)}(X) + x\mathbf{1}_{[x,\infty)}(X)] = \\ &= xE\mathbf{1}_{[x,\infty)}(X) + E[X\mathbf{1}_{[0,x)}(X)] \geq xE\mathbf{1}_{[x,\infty)}(X) = xP(X \geq x), \end{aligned}$$

откуда, сравнивая левую и правую части этого соотношения, мы получаем неравенство

$$P(X \geq x) \leq \frac{E \min\{x, X\}}{x}, \quad (1.7)$$

справедливое при любом $x > 0$.

Вытекающее из (1.7) неравенство

$$P(X \geq x) \leq \frac{EX}{x}$$

называется *неравенством Маркова*.

Неравенство Чебышева. Пусть X – произвольная случайная величина, у которой существует EX^2 . Применим неравенство Маркова к случайной величине $Z = (X - EX)^2$ и получим

$$P(|X - EX| \geq x) = P((X - EX)^2 \geq x^2) \leq \frac{E(X - EX)^2}{x^2} = \frac{DX}{x^2}.$$

Соотношение между левой и правой частями этого неравенства называют *неравенством Чебышева*:

$$P(|X - EX| \geq x) \leq \frac{DX}{x^2}. \quad (1.8)$$

С помощью неравенств Маркова и Чебышева можно оценивать вероятности больших значений неотрицательных случайных величин или вероятности больших отклонений случайных величин от своего среднего значения.

Неравенство Гаусса. При дополнительных условиях неравенство Чебышева можно уточнить. А именно, рассмотрим абсолютно непрерывную случайную величину X с унимодальным распределением. В таком случае существует такая точка $v \in \mathbb{R}$, что плотность $p_X(x)$ случайной величины X не убывает при $x \leq v$ и не возрастает при $x \geq v$. При этом число v является *модой* случайной величины X . Предположим, что $EX^2 < \infty$. Обозначим

$$\tau^2 = E(X - v)^2.$$

В 1821 году К. Гаусс доказал следующий результат.

Теорема 1.4. (*Gauss, 1821*) Пусть X – абсолютно непрерывная случайная величина с унимодальным распределением и модой v . Тогда для любого $r > 0$ справедливо неравенство

$$P(|X - v| \geq r) \leq \frac{4\tau^2}{9r^2}.$$

Вообще говоря, такая характеристика как τ^2 нетрадиционна для практических задач. Намного привычнее иметь дело с дисперсией, то есть со средним квадратом отклонения случайной величины X не от моды (которая может быть не единственной), а

от математического ожидания. Если распределение случайной величины X унимодально и симметрично, то математическое ожидание (равное нулю) совпадает с модой (по крайней мере с одной из мод). В таком случае неравенство Гаусса позволяет убедиться в справедливости следующего «правила трех сигм» для случайных величин с симметричным унимодальным распределением.

Следствие 1.1. (*«правило трех сигм» для симметричных унимодальных распределений*) Если распределение случайной величины X симметрично и унимодально, то

$$P(|X| \leq 3\sigma) \geq \frac{77}{81} > 0.95,$$

где $\sigma^2 = DX$.

Неравенство Иенсена и его частные случаи.

Определение 1.9. Функция $g(x)$ называется выпуклой, если ее график лежит выше касательной к нему, проведенной в любой его точке. Функция $h(x)$ называется вогнутой, если ее график лежит ниже касательной к нему, проведенной в любой его точке.

Примерами выпуклых функций служат функции $g(x) = x^2$ ($-\infty < x < \infty$), $g(x) = 1/x$ ($x > 0$). Примерами вогнутых функций служат функции $h(x) = \sqrt{x}$ ($x > 0$), $h(x) = \log_a x$ ($a > 1, x > 0$).

Пусть X – случайная величина, $g(x)$ – произвольная выпуклая функция, а $h(x)$ – произвольная вогнутая функция. Неравенство Иенсена устанавливает, что

$$Eg(X) \geq g(EX), \quad Eh(X) \leq h(EX).$$

В частности, применяя неравенство Иенсена для функции $g(x) = |x|$, мы получаем

$$|EX| \leq E|X| \tag{1.9}$$

С помощью неравенства Иенсена можно получить следующие полезные соотношения между характеристиками “центра” случайной величины и характеристиками ее разброса.

Пусть X – произвольная случайная величина. В свою очередь, применим неравенство (1.9) к случайной величине $Y = X - \text{med}X$. Вследствие линейности математического получаем

$$|EX - \text{med}X| \leq E|X - \text{med}X|. \quad (1.10)$$

Пусть X – произвольная случайная величина. Применим теперь неравенство Иенсена для выпуклых функций к случайной величине $Y = |X - EX|$ и функции $g(x) = x^2$. Получим

$$(E|X - EX|)^2 \leq E(X - EX)^2$$

или, что то же самое,

$$E|X - EX| \leq \sqrt{DX}. \quad (1.11)$$

Соотношение (1.11) принято называть *неравенством Ляпунова*.

Вспомним Задачу 1.6, согласно которой $E|X - \text{med}X| \leq E|X - a|$ для любого a . В частности, при $a = EX$ получаем

$$E|X - \text{med}X| \leq E|X - EX|.$$

Используя последнее неравенство как связующее звено между (1.10) и (1.11), мы получаем неравенство

$$|EX - \text{med}X| \leq \sqrt{DX}.$$

Таким образом, математическое ожидание не может отличаться от медианы больше чем на среднеквадратичное отклонение.

Неравенства для дисперсии и интерквартильного размаха.

Пусть $\mathcal{F}$ – класс всех распределений с конечным вторым моментом. Пусть $\mathcal{F}_0$ – подкласс класса $\mathcal{F}$, содержащий все унимодальные абсолютно непрерывные распределения с конечным вторым моментом и только их. Для простоты, без принципиального ограничения общности последующих рассуждений будем считать, что у распределений класса $\mathcal{F}$ математическое ожидание равно нулю, а у распределений класса $\mathcal{F}_0$ мода совпадает с математическим ожиданием и также равна нулю.

Символом $\mathcal{L}(X)$ будем обозначать распределение случайной величины X .

С помощью неравенств Чебышева и Гаусса можно получить следующее утверждение.

Теорема 1.5. Пусть $q \in [\frac{1}{2}, 1)$. Справедливы соотношения

$$\sup_{\mathcal{L}(X) \in \mathcal{F}} \frac{\sqrt{1-q}(x_q - x_{1-q})}{\sqrt{DX}} = \sqrt{2} < 1.4142,$$

$$\sup_{\mathcal{L}(X) \in \mathcal{F}_0} \frac{\sqrt{1-q}(x_q - x_{1-q})}{\sqrt{DX}} \leq \frac{2\sqrt{2}}{3} < 0.9429.$$

Следствие 1.2. Справедливы соотношения

$$\sup_{\mathcal{L}(X) \in \mathcal{F}} \frac{R_X}{\sqrt{DX}} = 2\sqrt{2} < 2.8285,$$

$$\sup_{\mathcal{L}(X) \in \mathcal{F}_0} \frac{R_X}{\sqrt{DX}} \leq \frac{4\sqrt{2}}{3} < 1.8857.$$

Этот результат позволяет оценить, насколько могут отличаться друг от друга такие характеристики разброса, как дисперсия и интерквартильный размах, в соответствующих классах распределений.

1.11 Характеристики формы распределения случайных величин

Пусть X – случайная величина, а EX и DX – соответственно ее математическое ожидание и дисперсия. Рассмотрим случайную величину

$$\tilde{X} = \frac{X - EX}{\sqrt{DX}}.$$

Несложно убедиться, что

$$E\tilde{X} = 0, \quad D\tilde{X} = E\tilde{X}^2 = 1.$$

Эти соотношения выполняются для любой случайной величины X , у которой существует (конечна) дисперсия. Однако математические ожидания более высоких степеней случайной величины $\tilde{X}$ уже будут разными для разных случайных величин X и несут в себе некоторую информацию о форме распределения X .

Определение 1.10. Предположим, что существует $E|X|^3$. Число

$$\kappa_3 = E\tilde{X}^3 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^3$$

называется коэффициентом асимметрии случайной величины X .

Если $\kappa_3 < 0$, то левый “хвост” распределения случайной величины X тяжелее правого. Если при этом случайная величина X абсолютно непрерывна, то в этом случае левый “склон” ее плотности более пологий, нежели правый. Если $\kappa_3 > 0$, то левый “хвост” распределения случайной величины X легче правого. Если при этом случайная величина X абсолютно непрерывна, то в этом случае левый “склон” ее плотности более крут, нежели правый. Если случайная величина X симметрична, то есть распределения величин X и $-X$ совпадают, то $\kappa_3 = 0$.

Определение 1.11. Предположим, что существует EX^4 . Число

$$\kappa_4 = E\tilde{X}^4 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^4 = \frac{E(X - EX)^4}{(DX)^2}$$

называется коэффициентом эксцесса или коэффициентом островершинности случайной величины X .

Чем больше коэффициент островершинности, тем острее вершина плотности соответствующей случайной величины и тяжелее “хвосты” распределения. Наоборот, если коэффициент островершинности мал, то график плотности соответствующей случайной величины имеет тупую вершину, а “хвосты” – легкие.

В качестве эталона “островершинности” традиционно принимают нормальное распределение, задаваемое плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Коэффициент эксцесса случайной величины, имеющей “эталонную” стандартную нормальную плотность $\varphi(x)$, равен 3.

1.12 Ковариация и коэффициент корреляции

Наряду с числовыми характеристиками центра, разброса и формы можно ввести числовые характеристики взаимозависимости случайных величин.

Определение 1.12. Пусть X и Y – случайные величины с конечными вторыми моментами. Число

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)]$$

называется ковариацией случайных величин X и Y .

Ковариацию удобно вычислять по формуле

$$\text{cov}(X, Y) = EXY - EX \cdot EY.$$

Ковариация линейна по обоим аргументам: если a и b – числа, а X , Y и Z – случайные величины, то

- $\text{cov}(aX + bY, Z) = a\text{cov}(X, Z) + b\text{cov}(Y, Z).$
- $\text{cov}(Z, aX + bY) = a\text{cov}(Z, X) + b\text{cov}(Z, Y).$

Также ковариация не зависит от сдвига случайной величины на число. Например, $\text{cov}(X + a, Y) = \text{cov}(X, Y).$

Если ковариация случайных величин X и Y положительна, то в среднем, при многократном воспроизведении стохастической ситуации, отклонение одной из этих случайных величин от своего среднего значения в какую-либо сторону влечет одновременное отклонение другой случайной величины от своего среднего значения в ту же сторону. В таком случае говорят, что случайные величины X и Y *положительно коррелированы*.

Если же ковариация случайных величин X и Y отрицательна, то эти случайные величины в среднем, при многократном воспроизведении стохастической ситуации, отклоняются от своих средних значений в разные стороны. В таком случае говорят, что случайные величины X и Y *отрицательно коррелированы*.

Если $\text{cov}(X, Y) = 0$, то говорят, что случайные величины X и Y *некоррелированы*.

Некоррелированность является более слабым свойством случайных величин, чем их независимость. Действительно, легко видеть, что, если случайные величины X и Y независимы, то $\text{cov}(X, Y) = 0$.

Однако обратное утверждение неверно: из равенства нулю ковариации случайных величин не следует их независимость. В этом нас убеждает следующий пример.

Пример 1.5. Рассмотрим случайные величины X и Y , совместное распределение вероятностей которых задается следующим соотношением

$$\begin{aligned} P(X = 0; Y = 1) &= P(X = 0; Y = -1) = \\ &= P(X = -1; Y = 0) = P(X = 1; Y = 0) = \\ &= P(X = \sqrt{2}/2; Y = \sqrt{2}/2) = P(X = -\sqrt{2}/2; Y = \sqrt{2}/2) = \\ &= P(X = \sqrt{2}/2; Y = -\sqrt{2}/2) = P(X = -\sqrt{2}/2; Y = -\sqrt{2}/2) = \frac{1}{8}. \end{aligned}$$

Несложно видеть, что при этом

$$\text{cov}(X, Y) = 0,$$

то есть случайные величины X и Y некоррелированы.

Однако, например,

$$0 = P(X = 0; Y = 0) \neq P(X = 0) \cdot P(Y = 0) = \frac{1}{16},$$

то есть случайные величины X и Y не являются независимыми.

Кроме того, случайные величины X и Y связаны формулой

$$X^2 + Y^2 = 1$$

(вспомним Пример 1.4, в котором случайные величины были связаны похожей формулой, но при этом являлись независимыми).

Для дисперсии суммы двух случайных величин справедлива формула

$$D(X \pm Y) = DX + DY \pm 2\text{cov}(X, Y).$$

Отсюда вытекает, что дисперсия суммы *или разности* независимых случайных величин равна *сумме* дисперсий слагаемых.

Ковариация не является безразмерной и ее величина зависит от того, какими единицами измеряются рассматриваемые случайные величины. Другими словами, по величине ковариации нельзя судить о том, насколько тесна взаимосвязь (взаимозависимость) случайных величин. Чтобы обойти это препятствие, рассмотрим еще одну характеристику.

Определение 1.13. Число

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX \cdot DY}}$$

называется коэффициентом корреляции случайных величин X и Y .

Справедливо неравенство

$$|\rho(X, Y)| \leq 1 \quad (1.12)$$

Чтобы в этом убедиться, для произвольных чисел a и b рассмотрим математическое ожидание $E(aX + bY)^2$, которое всегда неотрицательно. Поскольку ковариация не зависит от сдвига случайной величины на число, можно считать, что $EX = 0$ и $EY = 0$. Тогда $\text{cov}(X, Y) = EXY$, $DX = EX^2$ и $DY = EY^2$. Далее

$$E(aX + bY)^2 = a^2 EX^2 + 2ab EXY + b^2 EY^2 \geq 0$$

для всех вещественных a и b и, следовательно, дискриминант этого квадратного трехчлена неположителен:

$$(EXY)^2 - EX^2 EY^2 \leq 0.$$

Отсюда и следует (1.12). Более того, если $|\rho(X, Y)| = 1$, то случайные величины X и Y связаны линейной зависимостью: существуют

числа a и b такие, что $P(Y = aX + b) = 1$. При этом, если $\rho = 1$, то $a > 0$, а если $\rho = -1$, то $a < 0$. Если случайные величины X и Y независимы, то, очевидно, $\rho = 0$. Таким образом, можно сказать, что величина коэффициента корреляции характеризует степень *линейной* зависимости.

Близость коэффициента корреляции к единице означает лишь наличие *статистической* взаимосвязи, но никак не может быть использована для выявления направления *причинно-следственной* взаимосвязи.

Например, коэффициент корреляции между числом пожаров в наугад выбранном районе большого города за год и числом пожарных частей, расположенном в этом районе, как правило, довольно велик (скажем, для Нью-Йорка он приблизительно равен 0.8). Но это совсем не означает, что чем больше пожарных частей, тем больше пожаров.

Другой пример. В Голландии значение коэффициента корреляции между числом новорожденных за год и числом гнезд аистов в соответствующем районе довольно близко к единице (≈ 0.8). Однако мы не можем использовать это обстоятельство для обоснования теории о том, что детей приносят аисты. Объяснение здесь, увы, другое. Дело в том, что, в Голландии вообще рождаемость низка и, как правило, дети рождаются в молодых семьях, а молодые семьи поселяются в отдельных новых домах, на крышах которых и вьют свои гнезда аисты.

Следует также помнить, что иногда из-за неоднородности исследуемых данных может возникать так называемая ложная корреляция, обусловленная наличием некоторого третьего фактора. Например, если взять группу людей и вычислить корреляцию между ростом и длиной волос человека, то можно обнаружить, что она отлична от нуля. Это связано с тем, что в группе есть и мужчины, и женщины, а мужчины в среднем выше женщин, и волосы у мужчин в среднем короче, чем у женщин. Здесь фактором, влияющим на появление корреляции, является пол. Если разбить группу на две подгруппы по половому признаку и искать корреляцию уже в них, то она не обнаружится.

Глава 2

Испытания Бернулли

Как уже отмечалось, теория вероятностей имеет дело с такими экспериментами, которые можно повторять, по крайней мере теоретически, неограниченное число раз. Пусть некоторый эксперимент повторяется n раз, причем результаты каждого повторения не зависят от предыдущих испытаний. Такие серии повторений часто называют **независимыми испытаниями**. Частным случаем таких испытаний являются независимые испытания Бернулли.

Математическая модель испытаний Бернулли используется во многих прикладных областях. В частности, она часто применяется для решения задач, связанных с контролем качества продукции и надежности различных механизмов.



Jacob Bernoulli
(1655-1705)

Рассмотрим последовательность однотипных случайных экспериментов (испытаний), в каждом из которых возможны лишь два взаимоисключающих исхода. Условимся называть эти исходы “успех” и “неудача”. Пусть вероятность успеха одинакова для каждого испытания и равна p , $0 \leq p \leq 1$.

С этими испытаниями свяжем случайные величины $X_1, \dots, X_n$ по следующему правилу. Если в k -ом испытании зафиксирован успех, то $X_k = 1$, а если в k -ом испытании зафиксирована неудача, то $X_k = 0$, $k = 1, \dots, n$. При этом, очевидно, все случайные величины $X_1, \dots, X_n$ имеют одинаковое распределение:

$$P(X_k = 1) = p = 1 - P(X_k = 0), \quad k = 1, \dots, n. \quad (2.1)$$

Предположим, что рассматриваемые дихотомические испытания независимы в том смысле, что случайные величины $X_1, \dots, X_n$ независимы в совокупности. Испытания, удовлетворяющие указанным условиям, называются *испытаниями Бернулли*.

2.1 Биномиальное распределение

Рассмотрим стохастическую ситуацию, заключающуюся в осуществлении n испытаний Бернулли ($n \geq 1$). Построим соответствующую вероятностную модель. С этой целью заметим, что каждый элементарный исход *последовательности* n испытаний Бернулли представляет собой *цепочку* вида $\omega = (\omega_1, \dots, \omega_n)$, где ω_k равно либо 1 либо 0 в зависимости от того, какое значение случайной величины X_k наблюдается в k -ом испытании, то есть в зависимости от того, закончилось ли k -е испытание успехом или неудачей.

Несложно убедиться, что всего имеется $N = 2^n$ разных цепочек вида $\omega = (\omega_1, \dots, \omega_n)$, в которых ω_k равно либо 1 либо 0. Отсюда мы соответственно заключаем, что совокупность всех элементарных исходов рассматриваемой стохастической ситуации состоит из $N = 2^n$ элементов указанного выше вида.

Так как испытания независимы, то вероятность каждой элементарной цепочки $\omega = (\omega_1, \dots, \omega_n)$ оказывается равной произведению вероятностей соответствующих исходов дихотомических испытаний, составляющих эту цепочку:

$$P(\{\omega\}) = \prod_{k=1}^n p^{\omega_k} (1-p)^{1-\omega_k} = p^{\sum_{k=1}^n \omega_k} (1-p)^{n - \sum_{k=1}^n \omega_k} \quad (2.2)$$

(число единиц в цепочке $\omega = (\omega_1, \dots, \omega_n)$ равно $\sum_{k=1}^n \omega_k$ и, соответственно, число нулей в такой цепочке равно $n - \sum_{k=1}^n \omega_k$).

Пусть X – случайная величина, равная числу успехов в n испытаниях Бернулли. Найдем ее распределение.

Очевидно, что возможными значениями случайной величины X являются числа $0, 1, \dots, n$. Несложно видеть, что для каждого элементарного исхода $\omega = (\omega_1, \dots, \omega_n)$ случайная величина X равна

$$X(\omega) = \sum_{k=1}^n \omega_k.$$

Пусть m – целое число, $0 \leq m \leq n$. Существует C_n^m разных цепочек $\omega = (\omega_1, \dots, \omega_n)$, удовлетворяющих условию

$$\sum_{k=1}^n \omega_k = m. \quad (2.3)$$

С учетом соотношения (2.2) мы замечаем, что каждая из цепочек $\omega = (\omega_1, \dots, \omega_n)$, удовлетворяющих условию (2.3), имеет вероятность $P(\{\omega\}) = p^m(1-p)^{n-m}$.

Событие $\{\omega : X(\omega) = m\}$ состоит из всех цепочек, удовлетворяющих условию (2.3). Поэтому

$$p_m = P(X = m) = \sum_{\substack{\omega = (\omega_1, \dots, \omega_n): \\ \omega_1 + \dots + \omega_n = m}} P(\{\omega\}) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, \dots, n.$$

Используя формулу бинома Ньютона несложно убедиться, что

$$p_0 + p_1 + \dots + p_n = 1.$$

Набор пар $\{(m, p_m), m = 0, \dots, n\}$ называется *биномиальным распределением вероятностей*.

Случайную величину X , равную числу успехов в n испытаниях Бернулли, можно представить в виде

$$X = X_1 + \dots + X_n, \quad (2.4)$$

где $X_1, \dots, X_n$ – введенные в самом начале этого раздела независимые случайные величины. На основании (2.1) мы заключаем, что

$$EX_k = 1 \cdot p + 0 \cdot (1 - p) = p,$$

$$DX_k = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p), \quad k = 1, \dots, n.$$

Поэтому в силу справедливости представления (2.4) мы заключаем, что если случайная величина X имеет биномиальное распределение, то

$$EX = EX_1 + \dots + EX_n = n \cdot EX_1 = np,$$

$$DX = DX_1 + \dots + DX_n = n \cdot DX_1 = np(1 - p).$$

Вычисление биномиальных вероятностей $C_n^k p^k (1 - p)^{n-k}$ при больших n представляет собой очень сложную задачу. Однако оказывается возможным использовать приближенные формулы, которые являются следствиями предельных теорем, о которых пойдет речь ниже.

2.2 Распределение Пуассона

Пусть X – случайная величина, имеющая биномиальное распределение с параметрами n и p , то есть

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Справедлив следующий результат.

Теорема 2.1. Пусть $np = \lambda > 0$, $k - 1 \leq \frac{n}{4}$, $p \leq \frac{1}{4}$. Тогда

$$P(X = k) = \frac{\lambda^k}{k!} \exp\{-\lambda + r_n(k)\}, \quad (2.5)$$

где

$$\frac{k\lambda}{n} + 2 \log \frac{4}{3} \cdot \frac{k(1-k)}{n} - \frac{2\lambda^2}{3n} \leq r_n(k) \leq \frac{k\lambda}{n} + \frac{k(1-k)}{2n}.$$

Доказательство. Представим биномиальную вероятность $P(X = k)$ в следующем виде:

$$\begin{aligned} P(X = k) &= C_n^k p^k (1-p)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} p^k (1-p)^{n-k} = \\ &= \frac{(np)^k}{k!} e^{-np} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{np} = \\ &= \frac{\lambda^k}{k!} \exp\{-\lambda + r_n(k)\}. \end{aligned}$$

Таким образом, достаточно показать, что остаточный член

$$\begin{aligned} r_n(k) &= \log \left[\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{np} \right] = \\ &= \sum_{i=1}^{k-1} \log \left(1 - \frac{i}{n}\right) + (n-k) \log(1-p) + \lambda \end{aligned}$$

удовлетворяет условиям теоремы. Из неравенства

$$\log(1-x) \leq -x,$$

справедливого при $0 \leq x \leq 1$, следует соотношение

$$r_n(k) \leq -\sum_{i=1}^{k-1} \frac{i}{n} - (n-k)p + \lambda = \frac{k(1-k)}{2n} + kp = \frac{k(1-k) + 2k\lambda}{2n}.$$

Далее, так как при $0 \leq x \leq \frac{1}{4}$ справедливы неравенства

$$\log(1-x) \geq -4 \log \frac{4}{3} \cdot x, \quad \log(1-x) + x \geq -\frac{x^2}{2(1-x)} \geq -\frac{2x^2}{3}$$

(первое устанавливается элементарно, второе вытекает из разложения функции $\log(1-x)$ в ряд Тейлора), то

$$r_n(k) \geq -4 \log \frac{4}{3} \sum_{i=1}^{k-1} \frac{i}{n} + kp - \frac{2np^2}{3} = 2 \log \frac{4}{3} \cdot \frac{k(1-k)}{n} + \frac{k\lambda}{n} - \frac{2\lambda^2}{3n}.$$

Теорема доказана.

Теорема 2.1 устанавливает естественность приближения

$$P(X = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad (2.6)$$

которое обладает высокой точностью даже при умеренных n . Однако из этой же теоремы вытекает, что аппроксимация

$$P(X = k) \approx \frac{\lambda^k}{k!} \exp \left\{ -\lambda + \alpha \cdot \frac{k(1-k)}{n} + \frac{k\lambda}{n} \right\},$$

где $0.5 \leq \alpha \leq 0.5754$, более точна.

Условия Теоремы 2.1 можно интерпретировать следующим образом. Если число испытаний Бернулли велико, а вероятность успеха в одном отдельном испытании мала и при этом ожидаемое число успехов умеренно, то биномиальные вероятности можно вычислять по приближенной формуле (2.6):

$$C_n^k p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.$$

Просуммировав числа в правой части соотношения (2.6) по всем k , получим

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1,$$

то есть набор пар $\{(k, e^{-\lambda} \lambda^k / k!), k = 0, 1, \dots\}$ представляет собой распределение вероятностей. Это распределение называется *распределением Пуассона*.

Если случайная величина X имеет распределение Пуассона, то

$$EX = DX = \lambda.$$

2.3 Теорема Муавра–Лапласа. Нормальное распределение

Если нет оснований считать вероятность успеха в отдельном испытании Бернулли малой, то можно воспользоваться следующим утверждением.

Теорема 2.2. Пусть X – число успехов в n испытаниях Бернулли. Если число испытаний неограниченно увеличивается, а вероятность успеха в одном испытании остается той же самой, то, какими бы ни были числа a и b , $a \leq b$,

$$P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Это утверждение называется *теоремой Муавра–Лапласа*.



Abraham de Moivre
(1667-1756)



Pierre-Simon de
Laplace (1749-1827)

Можно убедиться, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

то есть функция $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ представляет собой плотность распределения. Распределение вероятностей, соответствующее плотности $\varphi(x)$, называется *стандартным нормальным*. Соответствующую функцию распределения обозначают $\Phi(x)$ и называют *стандартной нормальной*. Значения функции $\Phi(x)$ можно найти в специальных таблицах.

Если случайная величина X имеет стандартное нормальное распределение, то $EX = 0$, $DX = 1$. Можно показать, что если слу-

чаянная величина X имеет стандартное нормальное распределение, то случайная величина $Y = \sigma X + a$ имеет плотность

$$\varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и функцию распределения $\Phi\left(\frac{x-a}{\sigma}\right)$. Тогда $EY = a$, $DY = \sigma^2$ и говорят, что случайная величина Y имеет *нормальное распределение с параметрами a и σ^2* .

Приведем пример применения теоремы Муавра–Лапласа, поясняющий некоторые ее свойства.

Задача 2.1. *Общеизвестны книги американского врача Бенджамена Спока о воспитании детей. Доктор Спок известен не только своими книгами, но и своими антивоенными убеждениями. Он неоднократно участвовал в антивоенных акциях, в том числе во время войны, которую США вели во Вьетнаме, и был судим по обвинению в антиконституционной деятельности. Однажды он был арестован в Бостоне за участие в очередной антивоенной демонстрации. Его дело должен был рассматривать суд присяжных. Присяжные отбираются из дееспособного населения Бостона с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек (из числа которых потом и должны были быть выбраны 12 присяжных). Однако среди этих 300 человек оказалось лишь 90 женщин. Это обстоятельство послужило причиной того, что адвокаты доктора Спока подали протест (воспитанием детей в США, как и в России, в основном занимаются женщины и поэтому ясно, что в рассматриваемой ситуации при прочих равных женщины с большим сочувствием отнесутся к подсудимому). Насколько обоснован протест?*

Решение. Доля женщин в дееспособном населении Бостона составляет примерно 50%. Поэтому если вероятность случайного выбора женщины в число 300 кандидатов обозначить p , то при абсолютно беспристрастной процедуре выбора должно быть $p = \frac{1}{2}$. Проверим, так ли это на самом деле. Для этого, предполагая, что $p = \frac{1}{2}$, найдем вероятность того, что будет отобрано 90 или еще

меньше женщин. Другими словами, найдем вероятность того, что наблюдаемое отклонение числа отобранных женщин от ожидаемого их числа 150 обусловлено чистой случайностью. Если эта вероятность окажется исчезающе малой, то мы придем к выводу о том, что осуществилось событие, которое в сделанном предположении произойти практически не может. Это будет означать, что сделанное предположение не верно. (Такая логика рассуждений характерна для проверки *статистических гипотез*.)

Итак, пусть X – количество отобранных женщин. Мы можем считать X числом успехов в 300 испытаниях Бернулли с вероятностью успеха в одном испытании, равной p . В предположении полной случайности выбора ($p = \frac{1}{2}$) искомая вероятность имеет вид

$$P(X \leq 90) = \sum_{k=0}^{90} C_{300}^k p^k (1-p)^{n-k} = \frac{1}{2^{300}} \sum_{k=0}^{90} C_{300}^k.$$

Вычисление этой вероятности затруднено. Чтобы приближенно оценить ее, применим теорему Муавра–Лапласа и получим

$$\begin{aligned} P(X \leq 90) &= P\left(\frac{X - 150}{\sqrt{75}} \leq \frac{90 - 150}{\sqrt{75}}\right) \approx \\ &\approx \Phi\left(\frac{90 - 150}{\sqrt{75}}\right) = \Phi(-6.9282). \end{aligned}$$

Чтобы оценить, насколько мала эта вероятность, зафиксируем произвольное число $x > 0$ и рассмотрим значение $\Phi(-x)$. Так как плотность стандартного нормального распределения $\varphi(x)$ четна, то $\Phi(-x) = 1 - \Phi(x)$. Поэтому

$$\Phi(-x) = 1 - \Phi(x) = \int_x^\infty \varphi(y) dy = \int_x^\infty \frac{y}{y} \cdot \varphi(y) dy \leq \frac{1}{x} \int_x^\infty y \cdot \varphi(y) dy.$$

Но $y dy = d(y^2/2)$. Поэтому

$$\Phi(-x) \leq \frac{1}{\sqrt{2\pi}x} \int_x^\infty e^{-y} dy = \frac{\varphi(x)}{x}.$$

Следовательно, при $|x| \rightarrow \infty$ “хвосты” нормального распределения убывают быстрее, чем функция $\varphi(x)/|x| = e^{-x^2/2}/(\sqrt{2\pi}|x|)$.

Можно показать, что для $x > 0$

$$\frac{1}{x} - \frac{1}{x^3} \leq \frac{1 - \Phi(x)}{\varphi(x)} \leq \frac{1}{x}.$$

Таким образом, $\Phi(-6.9282) < 0.06e^{-24} < 6 \cdot 10^{-12}$. То есть вероятность того, что среди 300 кандидатов окажется не более 90 женщин, вычисленная в предположении беспристрастности выбора, исчезающе мала. Следовательно, это предположение, не верно. При этом вероятность того, что этот вывод ошибочен, тоже меньше $6 \cdot 10^{-12}$.

Теорема Муавра–Лапласа является частным случаем *центральной предельной теоремы*, которая будет подробно обсуждаться ниже.

2.4 Геометрическое распределение

Рассмотрим стохастическую ситуацию, состоящую в проведении испытаний Бернулли до первого успеха. В такой ситуации совокупность элементарных исходов состоит из всех цепочек вида $\omega = (\omega_1, \dots, \omega_{n+1})$, в которых $\omega_1 = \dots = \omega_n = 0$, а $\omega_{n+1} = 1$, $n = 0, 1, \dots$ Таких цепочек бесконечно много. Так как испытания независимы, то цепочка $\omega = (\omega_1, \dots, \omega_{n+1})$ указанного вида имеет вероятность

$$p_n = P(\{\omega\}) = p(1-p)^n. \quad (2.7)$$

По формуле суммы членов бесконечно убывающей геометрической прогрессии можно убедиться, что

$$\sum_{n=0}^{\infty} p_n = p \sum_{n=0}^{\infty} (1-p)^n = 1.$$

Набор пар $\{(n, p_n), n = 0, 1, \dots\}$, где p_n имеет вид (2.7), называется *геометрическим распределением вероятностей*.

С рассматриваемой стохастической ситуацией свяжем случайную величину X , определив ее для каждого элементарного исхода $\omega = (\omega_1, \dots, \omega_{n+1})$ указанного выше вида как $X(\omega) = n$. Случайная величина X принимает значения $0, 1, \dots$ и имеет смысл количества испытаний Бернулли, осуществленных до первого успеха. В таком случае говорят, что случайная величина X имеет геометрическое распределение.

Можно показать, что если случайная величина X имеет геометрическое распределение, то

$$EX = \frac{1-p}{p}, \quad DX = \frac{1-p}{p^2}.$$

Пусть X – случайная величина с геометрическим распределением

$$P(X = n) = p(1-p)^n, \quad n = 0, 1, \dots,$$

$0 < p < 1$. Если $p \rightarrow 0$, то $EX \rightarrow \infty$. Однако если одновременно со стремлением величины p к нулю производить нормировку рассматриваемой случайной величины, деля ее на ее математическое ожидание, то можно убедиться, что, каково бы ни было $x > 0$,

$$P\left(\frac{X}{EX} < x\right) = P\left(\frac{p}{1-p}X < x\right) \rightarrow 1 - e^{-x}.$$

Легко видеть, что предельная функция является функцией распределения. Такое распределение называется *стандартным показательным*. В общем случае, вводя параметр масштаба $\lambda > 0$, можно определить *показательное распределение с параметром λ* как задаваемое соответствующей функцией распределения $F(x)$, равной 0 при $x \leq 0$ и равной $1 - e^{-\lambda x}$ при $x > 0$.

2.5 Показательное распределение

Если Y – случайная величина с показательным распределением с параметром λ , то

$$EY = \frac{1}{\lambda}, \quad DY = \frac{1}{\lambda^2}.$$

Показательное распределение часто используется в качестве модели распределения продолжительности жизни или длительности безотказной работы. Однако такие модели весьма грубы и, вообще говоря, являются идеальными, поскольку не учитывают эффект старения. Действительно, в предположении о том, что время X безотказной работы некоторого устройства имеет показательное распределение с некоторым параметром λ , рассмотрим условную вероятность того, что после того как устройство проработало в течение времени t , устройство проработает еще как минимум время τ . Эта вероятность имеет вид

$$\begin{aligned} P(X > t + \tau | X > t) &= \frac{P(X > t + \tau; X > t)}{P(X > t)} = \frac{P(X > t + \tau)}{P(X > t)} = \\ &= \frac{e^{-(t+\tau)}}{e^{-t}} = e^{-\tau} = P(X > \tau), \end{aligned}$$

то есть эта условная вероятность не зависит от t и совпадает с безусловной вероятностью того, что устройство проработает как минимум время τ .

Итак, мы рассмотрели удобную и простую схему испытаний Бернулли, которую по сути можно описать в терминах случайного процесса с дискретным временем. Этот процесс в момент времени $t = n \in \mathbb{N}$ принимает одно из двух значений 0 или 1 независимо от того, какие значения он принял в другие моменты времени.

Со схемой испытаний Бернулли оказались тесно связанными следующие распределения вероятностей:

- биномиальное,
- геометрическое,
- Пуассона,
- показательное,
- нормальное.

Глава 3

Основные предельные теоремы теории вероятностей

При реализации *асимптотического подхода* к построению вероятностных моделей статистических закономерностей, присущих большим массивам данных, в качестве модели принимается асимптотическая аппроксимация – предельное распределение в некоторой предельной теореме теории вероятностей, условия которой можно считать выполненными в рассматриваемой ситуации.

При этом главную роль играет *правильный* (адекватный) выбор структуры наблюдаемой случайной величины, например, аддитивной, мультипликативной или экстремальной. То есть при моделировании первый шаг заключается в осознании того, можно ли считать наблюдаемую случайную величину

- суммой,
- произведением,
- максимумом или минимумом

каких-либо других “элементарных” или “атомических” случайных величин.

Например, во многих ситуациях можно считать, что наблюдаемое значение обусловлено *суммарным* воздействием случайных величин. В частности, приращение случайного процесса на некотором интервале времени может быть представлено в виде суммы его приращений на подынтервалах, на которые делится исходный интервал.

Мультипликативная структура данных возникает, когда наблюдения являются не абсолютными, но относительными приращениями. С помощью перехода к логарифмам мультипликативная схема может быть сведена к аддитивной.

При исследовании данных типа времени жизни, например, в теории надежности, могут возникнуть ситуации, когда агрегат выходит из строя при отказе хотя бы одной из его подсистем. Тогда время его безотказной работы – это наименьшее из времен безотказной работы его подсистем.

Мы сосредоточимся на предельных теоремах для вероятностных моделей данных с аддитивной структурой.

Предельные теоремы для сумм случайных величин занимают центральное место в классической теории вероятностей.

Примерами таких теорем являются

- законы больших чисел,
- центральная предельная теорема,
- теорема Пуассона.

Для их правильного применения и, соответственно, получения адекватных результатов необходимо четко понимать, при каких условиях справедливы их утверждения. Описанием этих условий мы и займемся.

3.1 Законы больших чисел

Пусть X – число успехов в n испытаниях Бернулли с вероятностью успеха в отдельном испытании, равной p . Случайная величина X имеет биномиальное распределение с параметрами n и p . Тогда под *частотой успеха* следует понимать дробь X/n . Еще очень давно было замечено, что при увеличении n частота успеха сближается с вероятностью p .

Формально это сближение можно установить, например, с помощью неравенства Чебышева. Зафиксируем какое-либо произвольно малое положительное число ε и рассмотрим вероятность

$$P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right)$$

того, что отклонение частоты успеха от вероятности успеха превзойдет ε по абсолютной величине.

Так как $EX = np$ и $DX = np(1-p)$, то

$$E\frac{X}{n} = p \quad \text{и} \quad D\frac{X}{n} = \frac{DX}{n^2} = \frac{p(1-p)}{n}.$$

Тогда по неравенству Чебышева

$$P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} D\left(\frac{X}{n}\right) = \frac{p(1-p)}{\varepsilon^2 n} \rightarrow 0$$

при $n \rightarrow \infty$.

Это утверждение называется *законом больших чисел Бернулли* и является самым простым видом закона больших чисел, который устанавливает сближение среднего арифметического случайных величин со средним арифметическим их математических ожиданий.

Пусть теперь $X_1, X_2, \dots$ – независимые не обязательно одинаково распределенные случайные величины. Обозначим

$$EX_j = a_j, \quad DX_j = \sigma_j^2, \quad j = 1, 2, \dots$$

Тогда по неравенству Чебышева для любого $\varepsilon > 0$ имеем

$$P\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^n a_j\right| > \varepsilon\right) \rightarrow 0,$$

если

$$\frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 \rightarrow 0.$$

Это утверждение называется *законом больших чисел Чебышева*.

В формулировке закона больших чисел Чебышева в *утверждение* входят лишь математические ожидания случайных величин, тогда как *условие* сформулировано в терминах их дисперсий, существование которых является более сильным условием, нежели существование математических ожиданий.

В случае одинаково распределенных слагаемых условие существования дисперсий оказывается излишним. А именно, справедлив *усиленный закон больших чисел Колмогорова*.

Теорема 3.1. Пусть $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины.

I. Если существует $EX_1 \equiv a$, то с вероятностью единица

$$\frac{1}{n} \sum_{j=1}^n X_j \rightarrow a \quad (n \rightarrow \infty). \quad (3.1)$$

II. Если при некотором $a \in \mathbb{R}$ с вероятностью единица имеет место сходимость (3.1), то существует $EX_1 \equiv a$.

Условие существования моментов определенного порядка играет критическую роль для справедливости тех или иных предельных теорем теории вероятностей. Для их адекватного применения на практике необходимо быть уверенным в его справедливости.

Усиленный закон больших чисел Колмогорова может быть применен для *практической* проверки конечности моментов.

Пусть, к примеру, нас интересует, конечен ли момент порядка α случайной величины X , реализации $X_1, X_2, \dots$ которой доступны наблюдению. Строим график функции

$$f(n) = \frac{1}{n} \sum_{j=1}^n X_j^\alpha, \quad n \in \mathbb{N},$$

натурального аргумента n . Если при увеличении n наблюдается стабилизация этой функции на некотором уровне, то можно сделать вывод о том, что момент EX^α конечен.

3.2 Центральная предельная теорема

Термин *центральная предельная теорема* означает любое утверждение о том, что при выполнении определённых условий функция распределения суммы малых случайных величин с ростом числа слагаемых сходится к нормальной функции распределения. Важность центральной предельной теоремы объясняется тем, что она дает теоретическое объяснение следующему многократно подтвержденному практикой наблюдению: если исход случайного эксперимента определяется большим числом случайных факторов,

влияние каждого из которых пренебрежимо мало, то распределение результата такого эксперимента хорошо аппроксимируется нормальным законом с соответствующим образом подобранными математическим ожиданием и дисперсией.

Одинаково распределенные слагаемые. Пусть $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины, удовлетворяющие условиям

$$EX_1 = a, \quad 0 < DX_1 = \sigma^2 < \infty. \quad (3.2)$$

Функцию распределения случайной величины X_1 обозначим $F(x)$. Функцию распределения центрированной и нормированной суммы

$$S_n = \frac{\sum_{k=1}^n X_k - na}{\sigma\sqrt{n}}$$

обозначим $F_n(x)$.

Функцию распределения и плотность стандартного нормального закона как и ранее будем обозначать $\Phi(x)$ и $\varphi(x)$ соответственно,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

Центральная предельная теорема утверждает, что последовательность функций распределения центрированных и нормированных сумм S_n случайных величин, удовлетворяющих условию (3.2), при $n \rightarrow \infty$ равномерно сходится к стандартной нормальной функции распределения:

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \rightarrow 0.$$

Из центральной предельной теоремы вытекает, что при больших n и любых x можно пользоваться приближением

$$P\left(\sum_{k=1}^n X_k < x\right) \approx \Phi\left(\frac{x - na}{\sigma\sqrt{n}}\right).$$

Другими словами, при больших n распределение суммы n независимых случайных величин можно аппроксимировать нормальным распределением с параметрами na и $\sigma\sqrt{n}$.

Центральную предельную теорему часто используют для обоснования того, что распределение погрешностей измерений предполагается нормальным. Действительно, погрешность измерения является результатом суммарного воздействия большого числа случайных факторов. Поэтому вследствие центральной предельной теоремы погрешность можно считать случайной величиной, распределенной по нормальному закону.

С помощью центральной предельной теоремы можно получить представление о том, какова скорость сходимости в законе больших чисел, то есть о том, как быстро средние арифметические независимых одинаково распределенных случайных величин сближаются с математическим ожиданием. Действительно, пусть, как и ранее, $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины с $EX_1 = a$ и $DX_1 = \sigma^2$, причем $0 < \sigma^2 < \infty$. В силу усиленного закона больших чисел при этом последовательность неотрицательных случайных величин

$$Z_n = \left| \frac{1}{n} \sum_{j=1}^n X_j - a \right| \quad (3.3)$$

с вероятностью единица стремится к нулю при $n \rightarrow \infty$. Под *скоростью* сходимости последовательности случайных величин Z_n при $n \rightarrow \infty$ условимся понимать такую неслучайную функцию $r(n)$, которая ведет себя при $n \rightarrow \infty$ в определенном смысле так же, как *случайная* последовательность $\{Z_n\}$. Построим функцию $r(n)$ для последовательности $\{Z_n\}$, определяемой соотношением (3.3). Пусть $r(n)$ – *некоторая* последовательность положительных чисел. Рассмотрим поведение случайных величин

$$Y_n = \frac{Z_n}{r(n)}$$

при $n \rightarrow \infty$. Интуитивно ясно, что функция $r(n)$ характеризует скорость сходимости последовательности $\{Z_n\}$ к нулю, если последовательность $\{Y_n\}$ имеет отличный от нуля и от бесконечности

предел. Но мы рассматриваем последовательность *случайных* величин, а случайные величины являются не числами, но *функциями* элементарных исходов, и потому понятие предела такой последовательности отличается от понятия предела числовой последовательности. Предел последовательности случайных величин можно трактовать по-разному. Не ставя цели перечислить все подходы, мы будем следовать такому, при котором рассматриваются не сами случайные величины, а их распределения.

Под упомянутым выше желательным отличием случайной величины, предельной для последовательности $\{Y_n\}$, от нуля или бесконечности мы будем понимать то, что ее функция распределения отлична от вырожденной в нуле или бесконечности.

Положим

$$r(n) = \frac{1}{\sqrt{n}}.$$

Пусть $x > 0$ – произвольное число. Тогда согласно центральной предельной теореме

$$\begin{aligned} P(Y_n < x) &= P(\sqrt{n}Z_n < x) = P\left(\sqrt{n}\left|\frac{1}{n}\sum_{j=1}^n X_j - a\right| < x\right) = \\ &= P\left(\left|\frac{\sum_{j=1}^n X_j - na}{\sqrt{n}}\right| < x\right) = P\left(-\frac{x}{\sigma} < \frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} < \frac{x}{\sigma}\right) = \\ &= P\left(\frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} < \frac{x}{\sigma}\right) - P\left(\frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} \leq -\frac{x}{\sigma}\right) \rightarrow \\ &\rightarrow \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right). \end{aligned} \quad (3.4)$$

Так как функция $e^{-x^2/2}$, стоящая под интегралом в определении функции распределения $\Phi(x)$, четна, то $\Phi(-x) = 1 - \Phi(x)$ для любого $x > 0$. Поэтому выражение в правой части (3.4) оказывается равным

$$G_\sigma(x) = \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right) = \Phi\left(\frac{x}{\sigma}\right) - \left[1 - \Phi\left(\frac{x}{\sigma}\right)\right] = 2\Phi\left(\frac{x}{\sigma}\right) - 1.$$

Эта функция распределения, очевидно, не является вырожденной (она соответствует случайной величине $Z = |Y|$, где Y – случайная величина, имеющая нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2).

Таким образом, мы можем заключить, что разность средних арифметических независимых одинаково распределенных случайных величин и математического ожидания стремится к нулю примерно как $1/\sqrt{n}$. Другими словами, скорость сходимости в законе больших чисел равна $1/\sqrt{n}$.

Обратим внимание, что такое заключение мы можем сделать только если в дополнение к условию справедливости закона больших чисел (существованию математического ожидания) выполнено условие справедливости центральной предельной теоремы (существование дисперсии случайной величины X_1). Можно показать, что из существования дисперсии вытекает существование математического ожидания, но не наоборот. Ответ на вопрос о скорости сходимости в законе больших чисел, когда математическое ожидание случайной величины X_1 существует, а дисперсия – нет, требует привлечения более тонкой техники.

На практике центральная предельная теорема применяется в качестве обоснования возможности аппроксимации распределения случайной величины, имеющей аддитивную структуру, то есть допускающей представление в виде суммы (независимых) случайных величин, нормальным законом. При этом для практических нужд, особенно в задачах оценивания риска, большое значение имеет оценка точности нормальной аппроксимации для распределения суммы $S_n = X_1 + \dots + X_n$ независимых случайных величин. Параметрами такой оценки разумно считать число слагаемых n и простейшие интегральные характеристики распределения слагаемых, например, моменты. Известно, что без дополнительных предположений сходимость распределений нормированных сумм к нормальному закону имеет место, но может быть как угодно медленной (Мацкявичюс, 1983). Поэтому для построения нетривиальных оценок скорости сходимости в центральной предельной теореме необходимы дополнительные условия и дополнительная

информация.

Простейшим дополнительным условием такого рода может стать требование существования абсолютного момента

$$\beta_{2+\delta} = E|X_1|^{2+\delta} < \infty$$

с $0 < \delta \leq 1$, а дополнительной информацией – конкретное значение $\beta_{2+\delta}$.

Обозначим

$$\rho(F_n, \Phi) = \sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)|.$$

Для случая $\delta = 1$ справедливо *неравенство Берри–Эссеена*

$$\rho(F_n, \Phi) \leq C \frac{\beta_3}{\sigma \sqrt{n}},$$

где C – абсолютная положительная константа. История поиска числовых оценок этой константы насчитывает около восьмидесяти лет и изобилует замечательными результатами, коллизиями и курьезами. В 1956 г. Эссеен доказал, что

$$C \geq \frac{\sqrt{10} + 3}{6\sqrt{2\pi}} \approx 0.409732.$$

И. Г. Шевцова показала, что $C \leq 0.4690$ (Шевцова, 2013).

Для случая $0 < \delta < 1$ еще в 1967 г. Л. В. Осипов и В. В. Петров показали, что если $0 < \delta < 1$, то при $n \rightarrow \infty$

$$\rho(F_n, \Phi) = o(n^{-\delta/2}).$$

Разнораспределенные слагаемые. Рассмотрим теперь ситуацию, в которой распределения слагаемых различны. Тогда сходимость распределений центрированных и нормированных сумм независимых случайных слагаемых к нормальному закону имеет место, если вклад каждого слагаемого в сумму мал по сравнению

с самой суммой, то есть ни одно из слагаемых не играет доминирующей роли. Чтобы формализовать сказанное, обозначим

$$EX_j = a_j, \quad a_1 + \dots + a_n = A_n; \quad DX_j = \sigma_j^2, \quad \sigma_1^2 + \dots + \sigma_n^2 = B_n^2,$$

$$F_n(x) = P(X_1 + \dots + X_n - A_n < xB_n), \quad j \geq 1, \quad n \geq 1.$$

Наиболее хорошо известной версией центральной предельной теоремы для сумм неодинаково распределенных независимых слагаемых является *теорема Линдберга–Феллера*, которая формулируется следующим образом.

Теорема 3.2. (*Lindeberg, 1922*), (*Feller, 1935*) Пусть случайные величины $X_1, X_2, \dots$ независимы. Для того чтобы

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \rightarrow 0 \quad (n \rightarrow \infty)$$

и при каждом $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} P(|X_j - a_j| > \varepsilon B_n) = 0,$$

необходимо и достаточно, чтобы было выполнено условие Линдберга: для любого $\tau > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) = 0.$$

Заметим, что все приводимые утверждения формально корректны, если $\sigma_j^2 > 0$, $j = 1, \dots, n$.

Условие Линдберга означает, что ни одно слагаемое (ни одна конечная группа слагаемых) в сумме не играет доминирующей роли. Во-первых, это следует из того, что

$$\begin{aligned} \max_{1 \leq j \leq n} P(|X_j - a_j| > \varepsilon B_n) &\leq \sum_{j=1}^n P(|X_j - a_j| > \varepsilon B_n) = \\ &= \sum_{j=1}^n \int_{|x-a_j| > \varepsilon B_n} dP(X_j < x) = \sum_{j=1}^n \int_{|x-a_j| > \varepsilon B_n} \frac{(x-a_j)^2}{(x-a_j)^2} dP(X_j < x) \leq \end{aligned}$$

$$\leq \frac{1}{\varepsilon^2 B_n^2} \sum_{j=1}^n \int_{|x-a_j| > \varepsilon B_n} (x-a_j)^2 dP(X_j < x) \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Во-вторых, дисперсия каждого слагаемого мала по сравнению с дисперсией всей суммы: при произвольно малом положительном ε для всех j , начиная с некоторого n , выполнено $\frac{\sigma_j^2}{B_n^2} \leq \varepsilon$:

$$\begin{aligned} \frac{\sigma_j^2}{B_n^2} &= \frac{\int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) + \int_{|x-a_j| \leq \tau B_n} (x-a_j)^2 dP(X_j < x)}{B_n^2} \leq \\ &\leq \frac{\int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) + \int_{|x-a_j| \leq \tau B_n} \tau^2 B_n^2 dP(X_j < x)}{B_n^2} \leq \\ &\leq \frac{\int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) + \tau^2 B_n^2}{B_n^2}. \end{aligned}$$

Пусть τ такое, что $\tau^2 \leq \frac{\varepsilon}{2}$. Тогда в силу условия Линдеберга, начиная с некоторого n ,

$$\begin{aligned} &\frac{\int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) + \tau^2 B_n^2}{B_n^2} \leq \\ &\leq \frac{\sum_{j=1}^n \int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x)}{B_n^2} + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

Предположим теперь, что $\beta_j^3 \equiv E|X_j - a_j|^3 < \infty$, $j \geq 1$, и обозначим $\beta_1^3 + \dots + \beta_n^3 = M_n^3$. Тогда, как несложно видеть,

$$\begin{aligned} &\frac{1}{B_n^2} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} (x-a_j)^2 dP(X_j < x) \leq \\ &\leq \frac{1}{\tau B_n^3} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} |x-a_j|^3 dP(X_j < x) \leq \frac{M_n^3}{\tau B_n^3}. \end{aligned}$$

Поэтому в случае существования третьих моментов слагаемых условие Линдеберга вытекает из условия Ляпунова:

$$\lim_{n \rightarrow \infty} \frac{M_n^3}{B_n^3} = 0,$$

то есть для случая конечных третьих абсолютных моментов слагаемых условие Ляпунова является достаточным для равномерной сходимости функций распределения центрированных и нормированных сумм независимых случайных слагаемых к стандартной нормальной функции распределения. Это утверждение, доказанное на рубеже XIX–XX веков в работах (Liapunoff, 1900), (Liapunoff, 1901), принято называть *теоремой Ляпунова*.

3.3 Суммы случайных индикаторов. Теорема Пуассона

Рассмотрим последовательность однотипных независимых испытаний (испытаний Бернулли), в каждом из которых некоторое событие A наступает с вероятностью p . Пусть случайная величина X_k равна 1, если в k -ом испытании событие A произошло, и 0 в противном случае. Тогда случайные величины $X_1, X_2, \dots$ независимы и

$$P(X_k = 1) = p, \quad P(X_k = 0) = 1 - p = q.$$

Образует сумму

$$S_n = X_1 + \dots + X_n,$$

которая равна числу появлений события A в первых n испытаниях. Тогда $ES_n = np$, $DS_n = npq$. Центральная предельная теорема устанавливает, что если дисперсия DS_n достаточно велика (что наблюдается при большом n , а p и q существенно отличных от нуля и единицы), то распределение случайной величины S_n близко к нормальному. Однако зачастую, например, в страховой деятельности, приходится рассматривать “редкие” события A , вероятность p появления которых в каждом конкретном испытании мала (то есть близка к нулю). При этом значение n может быть умеренным. Например, если $p = 0.02$ и $n = 200$, то $np = 2$ и $npq = 1.96$. В этом

случае можно пользоваться пуассоновской аппроксимацией (2.6) для распределения случайной величины S_n .

Из оценки остаточного члена в Теореме 2.1 вытекает, что предельной схемы, в которой лишь n стремится к бесконечности, недостаточно для строгого обоснования формулы (2.6), то есть для получения предельной теоремы, в которой $P(S_n = k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$. Необходимо, чтобы одновременно и $p = P(X_k = 1) \rightarrow 0$. Но этого нельзя добиться, рассматривая схему “нарастающих” сумм, в которой последовательность $X_1, X_2, \dots$ остается фиксированной.

Рассмотрим *схему серий* случайных величин

$$\begin{array}{l} X_{1,1} \\ X_{2,1}, X_{2,2} \\ \dots\dots\dots \\ X_{n,1}, X_{n,2}, \dots, X_{n,n} \\ \dots\dots\dots \end{array}$$

Первый индекс указывает на номер серии. Предположим, что в каждой серии случайные величины $X_{n,1}, X_{n,2}, \dots, X_{n,n}$, $n = 1, 2, \dots$, независимы и

$$P(X_{n,i} = 1) = p_n = 1 - q_n.$$

Из теоремы 2.1 вытекает следующий результат, известный как *теорема Пуассона*. Положим

$$S_n = X_{n,1} + \dots + X_{n,n}.$$

Теорема 3.3. Если $np_n \rightarrow \lambda > 0$ при $n \rightarrow \infty$, то при каждом фиксированном $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Естественным обобщением рассмотренной ситуации является та, в которой случайные величины $X_{n,j}$ могут иметь разное распределение даже в рамках одной и той же серии. Предположим теперь, что

$$P(X_{n,j} = 1) = p_{n,j}, \quad j = 1, \dots, n.$$

Подобную схему серий неодинаково распределенных независимых индикаторов принято называть *схемой Пуассона*. Распределение случайной величины S_n в такой схеме иногда называют *пуассон-биномиальным*.

Теорема 3.4. *Если*

$$\max_{1 \leq j \leq n} p_{n,j} \rightarrow 0, \quad \sum_{j=1}^n p_{n,j} \rightarrow \lambda \in (0, \infty)$$

при $n \rightarrow \infty$, то при каждом фиксированном $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Доказательство. Докажем теорему методом производящих функций. Символ $f_X(s)$ будет обозначать производящую функцию случайной величины X в точке s :

$$f_X(s) = Es^X, \quad |s| \leq 1.$$

Тогда, очевидно,

$$f_{X_{n,j}}(s) = Es^{X_{n,j}} = 1 + p_{n,j}(s - 1).$$

Поскольку случайные величины $X_{n,1}, \dots, X_{n,n}$ независимы, мы имеем

$$f_{S_n}(s) = Es^{S_n} = \prod_{j=1}^n f_{X_{n,j}}(s) = \prod_{j=1}^n (1 + p_{n,j}(s - 1)).$$

Далее, так как

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n p_{n,j}^2 \leq \lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} p_{n,j} \cdot \lim_{n \rightarrow \infty} \sum_{j=1}^n p_{n,j} = 0$$

и при любом фиксированном s , $|s| \leq 1$,

$$\log f_{S_n}(s) = \sum_{j=1}^n \log(1 + p_{n,j}(s - 1)) \sim$$

$$\sim \sum_{j=1}^n p_{n,j}(s-1) \sim \lambda(s-1) \quad (n \rightarrow \infty),$$

то при любом фиксированном s , $|s| \leq 1$,

$$f_{S_n}(s) \rightarrow e^{\lambda(s-1)}.$$

Но последнее выражение представляет собой производящую функцию распределения Пуассона с параметром $\lambda > 0$. Теорема доказана.

Приведенные выше версии теоремы Пуассона описывают поведение числа редко наблюдаемых событий при большом числе испытаний. Их иногда называют *теоремами о редких событиях* или *законами малых чисел*. Именно эти теоремы могут рассматриваться как математическое обоснование использования пуассоновского распределения во многих прикладных задачах.

Рассмотрим теперь вопрос о скорости сходимости в законах малых чисел, являющийся ключевым для понимания адекватности пуассоновской аппроксимации.

Для $k = 0, 1, 2, \dots$ обозначим

$$\pi_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad b_k = P(S_n = k),$$

$$\Pi = \{\pi_0, \pi_1, \pi_2, \dots\}, \quad B = \{b_0, b_1, b_2, \dots\}.$$

Определим *расстояние по вариации* между распределением случайной величины S_n и распределением Пуассона по формуле

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |b_k - \pi_k|.$$

Следующие результаты приведем без доказательств, ограничившись лишь соответствующими ссылками.

Теорема 3.5. (*Barbour, Hall, 1984*) Если $p_{n,1} = \dots = p_{n,n} = p$, $\lambda = np$, то справедливо неравенство

$$\|B - \Pi\| \leq 2p \min\{1, \lambda\}.$$

Обозначим

$$\lambda = p_{n,1} + \dots + p_{n,n}, \quad \lambda_2 = p_{n,1}^2 + \dots + p_{n,n}^2.$$

Легко видеть, что

$$ES_n = \lambda, \quad DS_n = \lambda - \lambda_2. \quad (3.5)$$

Теорема 3.6. *(LeCam, 1960) Справедливо неравенство*

$$\|B - \Pi\| \leq 2 \min\{9, \lambda\} \max_{1 \leq j \leq n} p_{n,j}.$$

Теорема 3.7. *(Пресман, 1985) Справедливо неравенство*

$$\|B - \Pi\| \leq 2.08 \frac{\lambda_2}{\lambda}.$$

Теорема 3.8. *(Deheuvels and Pfeifer, 1987), (Deheuvels and Pfeifer, 1988) Если $\lambda \rightarrow \infty$ и $\lambda_2/\lambda \rightarrow 0$, то*

$$\|B - \Pi\| \sim \frac{\lambda_2}{\sqrt{2\pi e\lambda}}.$$

Из последней теоремы вытекает, что условие $\max_{1 \leq j \leq n} p_{n,j} \rightarrow 0$ не является необходимым для справедливости пуассоновской аппроксимации для пуассон-биномиального распределения. На самом же деле в качестве *необходимого и достаточного* условия нужно рассматривать условие $\lambda_2/\lambda \rightarrow 0$, которое в силу (3.5) равносильно тому, что

$$\frac{DS_n}{ES_n} \rightarrow 1.$$

3.4 Устойчивые и безгранично делимые распределения

Пусть случайные величины $X_1, X_2, \dots$ независимы и одинаково распределены. Из теоремы Линдберга–Феллера вытекает, что для

сходимости распределений их последовательных сумм, нормированных и центрированных соответствующим образом, к нормальному закону необходимо и достаточно существования конечной дисперсии слагаемых. Возникает вполне резонный вопрос: а к каким распределениям могут сходиться распределения нормированных сумм независимых и одинаково распределенных случайных величин, если у слагаемых нет моментов второго порядка? Оказывается, что класс предельных законов для таких сумм совпадает с классом *устойчивых распределений*.

Функция распределения $G(x)$ и соответствующая ей характеристическая функция $g(t)$ называются *устойчивыми*, если для любых $a_1 > 0$ и $a_2 > 0$ найдутся числа $a > 0$ и $b \in \mathbb{R}$ такие, что

$$g(a_1 t)g(a_2 t) \equiv e^{ibt} g(at). \quad (3.6)$$

Несложно убедиться в том, что условие (3.6) эквивалентно тому, что для любых $a_1 > 0$, $a_2 > 0$, $b_1 \in \mathbb{R}$ и $b_2 \in \mathbb{R}$ существуют числа $a > 0$ и $b \in \mathbb{R}$ такие, что

$$(G_1 * G_2)(x) \equiv G(ax + b),$$

где $G_1(x) = G(a_1 x + b_1)$, $G_2(x) = G(a_2 x + b_2)$ (здесь символ $*$ обозначает свертку).

Характеристическая функция $g(t)$ устойчива тогда и только тогда, когда она может быть представлена в виде

$$g(t) = \exp \left\{ iat - c|t|^\alpha \left(1 + ibQ(t, \alpha) \operatorname{sign} t \right) \right\},$$

где

$$Q(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \text{если } \alpha \neq 1, \\ \frac{2}{\pi} \log |t|, & \text{если } \alpha = 1. \end{cases}$$

Параметр α при этом называется *характеристическим показателем*.

Все невырожденные устойчивые распределения абсолютно непрерывны. Примерами устойчивых распределений являются

нормальное ($\alpha = 2$), Коши ($\alpha = 1$), Леви ($\alpha = 1/2$) с плотностью

$$p_{1/2}(x) = \begin{cases} 0, & x < 0, \\ \frac{\sigma}{\sqrt{2\pi x^3}} \exp\left\{-\frac{\sigma^2}{2x}\right\}, & x > 0 \end{cases}$$

и распределение с плотностью $\bar{p}(x) = p_{1/2}(|x|)$ при $x < 0$ и $\bar{p}(x) = 0$ при $x > 0$. Другие примеры устойчивых законов, плотности которых выражаются через элементарные функции, неизвестны.

В 1925 г. П. Леви доказал следующую фундаментальную теорему.

Теорема 3.9. *Для того чтобы функция распределения $F(x)$ могла быть предельной при $n \rightarrow \infty$ для распределений сумм*

$$S_n = \frac{X_1 + \dots + X_n - a_n}{b_n}$$

независимых одинаково распределенных случайных величин при некотором выборе чисел $a_n \in \mathbb{R}$ и $b_n > 0$, необходимо и достаточно, чтобы она была устойчивой.

Полное доказательство этой теоремы см., например, в (Хинчин, 1938) и (Гнеденко и Колмогоров, 1949).

Устойчивые законы являются хорошо известными примерами распределений с тяжелыми хвостами. Если $G_\alpha(x)$ – устойчивая функция распределения с характеристическим показателем $\alpha \in (0, 2)$, то

$$G_\alpha(-x) + 1 - G_\alpha(x) \sim \frac{c}{x^\alpha}$$

при $x \rightarrow \infty$, где $c > 0$. Более того, если Z_α – случайная величина с функцией распределения $G_\alpha(x)$, $\alpha \in (0, 2)$, то $E|Z_\alpha|^\gamma < \infty$ для любого $\gamma < \alpha$, но моменты величины Z_α порядков, больше или равных α , не существуют. Таким образом, дисперсии всех устойчивых законов за исключением нормального не определены («бесконечны»).

Устойчивые законы, а также отрицательное биномиальное, гамма-распределения и распределение Пуассона являются примерами безгранично делимых законов. Функция распределения $F(x)$

с характеристической функцией $f(t)$ называются *безгранично делимыми*, если для любого $n \geq 1$ существует характеристическая функция $f_n(t)$ такая, что

$$f(t) \equiv (f_n(t))^n$$

Безгранично делимые характеристические функции нигде не обращаются в нуль.

Рассмотрим предельную схему серий, обобщающую рассмотренную выше *схему нарастающих сумм* S_n случайных величин $X_1, X_2, \dots$, образующих одну последовательность. А именно, предположим, что распределения слагаемых могут изменяться вместе с изменением числа слагаемых в сумме. Пусть $\{m_n\}_{n \geq 1}$ – последовательность натуральных чисел. Пусть $\{X_{n,j}\}_{j \geq 1, n = 1, 2, \dots}$ – последовательность серий независимых в каждой серии случайных величин. Говорят, что случайные величины $\{X_{n,j}\}$ удовлетворяют *условию равномерной предельной малости*, если для любого положительного числа ε

$$\lim_{n \rightarrow \infty} \sup_{1 \leq j \leq m_n} P(|X_{n,j}| > \varepsilon) = 0.$$

В 1937 г. А. Я. Хинчин доказал следующую фундаментальную теорему.

Теорема 3.10. *Для того чтобы функция распределения $F(x)$ могла быть предельной при $n \rightarrow \infty$ для распределений сумм*

$$S_{n,m_n} = X_{n,1} + \dots + X_{n,m_n}$$

независимых случайных величин, удовлетворяющих условию равномерной предельной малости, необходимо и достаточно, чтобы она была безгранично делимой.

Доказательство см., например, в (Гнеденко и Колмогоров, 1949).

Глава 4

Энтропийный подход к построению вероятностных моделей

Как уже было сказано, в рамках энтропийного (информационного) подхода в качестве вероятностной модели используется наиболее неблагоприятная модель, обладающая свойствами наибольшей неопределенности при заданных предположениях. По сути этот подход реализует минимаксный принцип выбора наилучшего метода в наихудших условиях.

Для строгой формализации базовых принципов такого подхода нам необходимо подробно обсудить понятия (модели) информации и энтропии. Интуитивно ясно, что понятие информации тесно связано с понятием неопределенности. В свою очередь, в 20-30-х годах XX столетия Хартли и Шеннон предложили связать понятия информации и вероятности. Такой подход оказался вполне разумным и плодотворным.

4.1 Информация и энтропия

Пусть A и B – события, вероятности которых положительны.

Определение 4.1. Информацией (по Шеннону), содержащейся в событии B относительно события A , называется

$$I(A|B) = \log \frac{P(A|B)}{P(A)}.$$

Если $B = A$, то, очевидно, $I(A|A) = -\log P(A)$. Таким образом, мы приходим к следующему важному определению.

Определение 4.2. Информацией (по Шеннону), содержащейся в событии A , называется

$$I(A) = -\log P(A). \quad (4.1)$$

Смысл этих определений легко пояснить, рассмотрев простейшие свойства введенных понятий. Эти свойства оказываются аналогичными тем, которые должны быть присущи информации с точки зрения здравого смысла.

1. Осуществление события, вероятность которого невелика, как правило, несет в себе больше информации, нежели осуществление события, вероятность которого значительна. Например, в начале нашего столетия представлялось практически невероятно обнаружить живой экземпляр считавшегося давно вымершим вида кистеперых рыб. Когда же такая рыба – целакант – была поймана, фактически произошла революция в ихтиологии. В то же время, поимка любого экземпляра рыбы такого распространенного вида как, скажем, треска, может нести в себе информацию разве что о месте прохождения косяка этих рыб и никакой научной революции не вызывает.
2. Если события A и B независимы, то, очевидно, осуществление события B не дает никакой информации о событии A . Действительно, в этом случае мы имеем

$$I(A|B) = \log 1 = 0.$$

3. Если события A и B независимы, то их одновременное осуществление несет в себе столько же информации, сколько содержится в каждом из них в отдельности. Действительно, в этом случае мы имеем

$$\begin{aligned} I(AB) &= -\log P(AB) = -\log(P(A)P(B)) = \\ &= -\log P(A) - \log P(B) = I(A) + I(B). \end{aligned}$$

Логарифмическое определение информации (4.1) восходит к Хартли (Hartley, 1928). Основание логарифма не играет определяющей роли и существенно лишь для выбора единицы измерения информации (для того чтобы в определении 4.2 информация всегда была неотрицательна, это основание должно быть больше единицы). Обычно используют логарифмы по основанию 2, для которых единица информации содержится в событии, вероятность которого 1/2. Такая единица информации называется бит (от английского bit (кусочек) или как аббревиатура термина Binary digiT (двоичный разряд)). Единица информации, порожденная натуральными логарифмами, называется нат (nat).

Пусть $\mathcal{E}$ – эксперимент, в котором может осуществиться лишь один из n исходов $A_1, \dots, A_n$. Обозначим $P(A_i) = p_i$ (очевидно, что $p_1 + \dots + p_n = 1$). Тогда мы можем считать информацию, полученную в результате этого эксперимента, случайной величиной, принимающей значения $I(A_1), \dots, I(A_n)$ соответственно с вероятностями $p_1, \dots, p_n$. Обозначим эту случайную величину $Q(\mathcal{E})$. Введем следующую интегральную информационную характеристику $\mathcal{E}$.

Определение 4.3. Энтропией $H(\mathcal{E})$ эксперимента $\mathcal{E}$ называется величина

$$H(\mathcal{E}) = EQ(\mathcal{E}) = - \sum_{i=1}^n p_i \log p_i. \quad (4.2)$$

Энтропия эксперимента может служить мерой его неопределенности, что подтверждается совпадением свойств формально введенной величины $H(\mathcal{E})$, приводимых в следующей теореме, с ожидаемыми с позиций здравого смысла свойствами разумной меры неопределенности.

Теорема 4.1. Величина $H(\mathcal{E})$ обладает следующими свойствами.

1°. $H(\mathcal{E}) \geq 0$, причем равенство достигается тогда и только тогда, когда существует $i \in \{1, \dots, n\}$ такое, что $p_i = 1$.

2°. Пусть $\mathcal{E}_0$ – эксперимент с n равновероятными исходами. Тогда $H(\mathcal{E}) \leq H(\mathcal{E}_0)$, каким бы ни был эксперимент $\mathcal{E}$ с таким же числом n возможных исходов.

3°. Пусть $\mathcal{E}_1$ – эксперимент с $n - 1$ исходом, построенный из эксперимента $\mathcal{E}$ с помощью объединения двух исходов, скажем, A_i и A_j ($i \neq j$), и пусть $\mathcal{E}_2$ – эксперимент с исходами A_i и A_j , вероятности которых (в рамках $\mathcal{E}_2$) соответственно равны $p_i/(p_i + p_j)$ и $p_j/(p_i + p_j)$. Тогда

$$H(\mathcal{E}) = H(\mathcal{E}_1) + (p_i + p_j)H(\mathcal{E}_2).$$

4°. Энтропия $H(\mathcal{E})$ зависит не от $A_1, \dots, A_n$, а от $p_1, \dots, p_n$, будучи симметрической функцией переменных $p_1, \dots, p_n$.

5°. Энтропия $H(\mathcal{E})$ – непрерывная функция от $p_1, \dots, p_n$.

Доказательство. 1°. Доопределив функцию $f(p) = -p \log p$ по непрерывности нулем при $p = 0$, заметим, что $f(p) \geq 0$, причем $f(p) = 0$ тогда и только тогда, когда $p = 0$ или $p = 1$.

2°. Обозначим $g(x) = x \log x$. Легко убедиться, что $g''(x) \geq 0$ при $x \in [0, 1]$, то есть $g(x)$ выпукла при указанных x . Это означает, что для любых неотрицательных $\alpha_1, \dots, \alpha_n$ таких, что $\alpha_1 + \dots + \alpha_n = 1$, при любых неотрицательных $y_1, \dots, y_n$

$$g\left(\sum_{i=1}^n \alpha_i y_i\right) \leq \sum_{i=1}^n \alpha_i g(y_i).$$

Положим $\alpha_i = \frac{1}{n}$, $y_i = p_i$. Тогда в силу последнего неравенства

$$H(\mathcal{E}_0) = -\log \frac{1}{n} = -n \left[\sum_{i=1}^n \frac{p_i}{n} \log \left(\sum_{i=1}^n \frac{p_i}{n} \right) \right] \geq -n \sum_{i=1}^n \frac{p_i}{n} \log p_i = H(\mathcal{E}),$$

что и требовалось доказать.

3°. Во-первых, покажем, что $H(\mathcal{E}_1) \leq H(\mathcal{E})$. Действительно,

$$\begin{aligned} H(\mathcal{E}_1) &= - \sum_{k \neq i, k \neq j} p_k \log p_k - (p_i + p_j) \log(p_i + p_j) = \\ &= - \sum_{k \neq i, k \neq j} p_k \log p_k - p_i \log(p_i + p_j) - p_j \log(p_i + p_j) \leq \\ &\leq - \sum_{k \neq i, k \neq j} p_k \log p_k - p_i \log p_i - p_j \log p_j = H(\mathcal{E}). \end{aligned}$$

Во-вторых, вычислим

$$\begin{aligned}
 H(\mathcal{E}) - H(\mathcal{E}_1) &= -p_i \log p_i - p_j \log p_j + (p_i + p_j) \log(p_i + p_j) = \\
 &= (p_i + p_j) \left[-\frac{p_i}{p_i + p_j} \log p_i - \frac{p_j}{p_i + p_j} \log p_j + \right. \\
 &\quad \left. + \left(\frac{p_i}{p_i + p_j} + \frac{p_j}{p_i + p_j} \right) \log(p_i + p_j) \right] = \\
 &= (p_i + p_j) \left[-\frac{p_i}{p_i + p_j} \log \frac{p_i}{p_i + p_j} - \frac{p_j}{p_i + p_j} \log \frac{p_j}{p_i + p_j} \right] = \\
 &\quad (p_i + p_j) H(\mathcal{E}_2),
 \end{aligned}$$

что и требовалось доказать.

Наконец, пункты 4° и 5° очевидны. Теорема доказана.

Из свойства 2° энтропии вытекает, что энтропия эксперимента с максимальной неопределенностью максимальна, а из свойства 1° видно, что энтропия минимальна в полностью “определенном” эксперименте. Свойство 3° показывает, что энтропия в определенном смысле аддитивна: при увеличении неопределенности за счет увеличения числа исходов энтропия возрастает, причем прирост энтропии пропорционален вероятности дополнительных исходов. Эти свойства показывают, что энтропия, определяемая соотношением (4.2), является вполне разумной мерой неопределенности стохастического эксперимента с конечным числом исходов. Более того, Д. К. Фаддееву удалось показать, что система свойств 1° – 5° однозначно определяет функционал (4.2) (Фаддеев, 1956) (см. также (Яглом и Яглом, 1973)). Другими словами, если мы захотим сконструировать какую-либо меру неопределенности, которая должна обладать вполне естественно ожидаемыми от такой характеристики свойствами 1° – 5°, то мы неизбежно придем к энтропии.

Очевидно, вместо экспериментов с n исходами мы можем рассматривать дискретные случайные величины, принимающие какие-либо значения $x_1, \dots, x_n$ с вероятностями $p_1, \dots, p_n$, считая, что значение x_i случайной величины X наблюдается в результате

осуществления исхода A_i эксперимента $\mathcal{E}$. Поэтому по аналогии с (4.2) мы можем определить энтропию простой случайной величины X как

$$H(X) = - \sum_{i=1}^n p_i \log p_i,$$

или, вводя плотность $p(x)$ распределения X относительно считающей меры

$$p(x) = \begin{cases} p_i, & x = x_i, \ i = 1, \dots, n; \\ 0, & x \notin \{x_1, \dots, x_n\}, \end{cases}$$

как

$$H(X) = -E \log p(X). \quad (4.3)$$

По формальной аналогии с (4.3), если X – абсолютно непрерывная случайная величина с лебеговой плотностью $p(x)$, то определим энтропию $H(X)$ величины X так же, как (4.3). Однако, аналогия таких определений энтропии дискретной и абсолютно непрерывной случайных величин оказывается чисто формальной. Как известно, каждая случайная величина X как функция элементарного исхода может быть представлена в виде поточечного предела последовательности простых случайных величин X_n . Тогда, конечно же, $X_n \Rightarrow X$. Но если мы попытаемся использовать предельный переход при неограниченно увеличивающемся числе значений аппроксимирующих простых случайных величин для того, чтобы получить энтропию предельной абсолютно непрерывной случайной величины X в виде (4.3), мы неизбежно потерпим неудачу, так как предел энтропий аппроксимирующих простых случайных величин оказывается бесконечным. Тем не менее, выражение в правой части (4.3) оказывается пределом для “стандартизованных” энтропий аппроксимирующих простых случайных величин в следующем смысле. По сути энтропия (4.3) абсолютно непрерывной случайной величины определяет среднюю информацию, содержащуюся в X по сравнению с “бесконечно большой аддитивной постоянной”.

Чтобы пояснить сказанное, разобьем область значений величины X на непересекающиеся интервалы Δ_i равной длины δ . Опре-

делим соответствующую дискретную аппроксимацию X_δ случайной величины X , полагая X_δ равной некоторому фиксированному элементу интервала Δ_i , если X попадает в этот интервал. Тогда $X_\delta \Rightarrow X$ ($\delta \rightarrow 0$), и при некоторых условиях регулярности на плотность $p(x)$ величины X для некоторых точек $x_i^* \in \Delta_i$ мы имеем

$$\begin{aligned} H(X_\delta) &= -\sum_i P(X \in \Delta_i) \log P(X \in \Delta_i) = -\sum_i p(x_i^*) \delta \log(p(x_i^*) \delta) = \\ &= -\sum_i p(x_i^*) \delta \log p(x_i^*) - \sum_i p(x_i^*) \delta \log \delta = \\ &\quad -\sum_i p(x_i^*) \delta \log p(x_i^*) - \log \delta. \end{aligned} \quad (4.4)$$

Первое слагаемое в правой части (4.4) представляет собой интегральную сумму Дарбу для величины $H(X)$, определяемой соотношением (4.3). Поэтому мы можем записать

$$H(X) = \lim_{\delta \rightarrow 0} [H(X_\delta) + \log \delta]. \quad (4.5)$$

Таким образом, величину $H(X)$, определенную для абсолютно непрерывной случайной величины X соотношением (4.3) и являющуюся пределом нормированных мер неопределенности, саму можно считать мерой неопределенности стохастического эксперимента, в результате которого наблюдается случайная величина X , то есть мерой неопределенности ее распределения.

В соответствии с определением 4.2, величину $-\log \delta$ в соотношении (4.5) можно интерпретировать как информацию, содержащуюся в событии, вероятность которого равна δ . Таким образом, эта величина может характеризовать рост неопределенности распределения случайной величины X_δ , вызванный квантованием. Поэтому можно считать, что $H(X)$ характеризует неопределенность абсолютно непрерывной случайной величины, обусловленную формой ее распределения. При этом, однако, соотношение (4.3) имеет разный смысл для дискретных и абсолютно непрерывных случайных величин.

Определение 4.4. Величина $H(X)$, определенная для абсолютно непрерывной случайной величины X соотношением (4.3), называется дифференциальной энтропией случайной величины X .

Рассмотрим теперь вопрос о том, какие абсолютно непрерывные распределения имеют наибольшую дифференциальную энтропию. Ответ на него укажет на наиболее неопределенные (непредсказуемые) абсолютно непрерывные случайные величины.

В вариационном исчислении хорошо известен следующий метод решения так называемой изопериметрической задачи.

Пусть $p(x)$ – некоторая функция вещественного аргумента x , a и b – некоторые числа, $a < b$. Предположим, что заданы функции $F(x, p(x))$, $\varphi_i(x, p(x))$ и числа a_i , $i = 1, \dots, n$. Известно, что функция $p(x)$, обращающая в максимум функционал

$$J(p) = \int_a^b F(x, p) dx$$

при условиях

$$\int_a^b \varphi_i(x, p) dx = a_i, \quad i = 1, \dots, n, \quad (4.6)$$

находится из уравнения

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \dots + \lambda_n \frac{\partial \varphi_n}{\partial p} = 0. \quad (4.7)$$

Здесь $\lambda_1, \dots, \lambda_n$ – неопределенные коэффициенты, определяемые подстановкой решения уравнения (4.7) в условия (4.6) (см., например, (Эльсгольц, 1969)).

Теорема 4.2. 1°. Пусть случайная величина X имеет равномерное распределение на отрезке $[-a, a]$. Тогда $H(X) \geq H(Y)$ для любой случайной величины Y , удовлетворяющей условию $P(|Y| \leq a) = 1$.

2°. Пусть случайная величина X имеет показательное распределение с параметром $\mu > 0$. Тогда $H(X) \geq H(Y)$ для любой случайной величины Y , удовлетворяющей условиям $P(Y \geq 0) = 1$ и $EY = \mu^{-1}$.

3°. Пусть случайная величина X имеет нормальное распределение с параметрами a и σ^2 . Тогда $H(X) \geq H(Y)$ для любой случайной величины Y , удовлетворяющей условиям $EY = a$, $DY = \sigma^2$.

Доказательство. 1°. Положим $F(x, p) = -p \log p$, $\varphi_1(x, p) = p$. Тогда мы имеем задачу

$$J(p) = - \int_{-a}^a p \log p dx \rightarrow \max_p$$

при условии

$$\int_{-a}^a \varphi_1(x, p) dx = 1. \quad (4.8)$$

Уравнение (4.7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda \frac{\partial \varphi_1}{\partial p} = -(1 + \log p) + \lambda = 0.$$

Решением этого уравнения является функция $p(x) = e^{\lambda-1}$. Эта функция постоянна по $x \in [-a, a]$. С учетом условия (4.8) мы заключаем, что $p(x) = \frac{1}{2a} \mathbf{1}(-a \leq x \leq a)$, то есть $p(x)$ – это плотность распределения, равномерного на $[-a, a]$.

2°. Положим $F(x, p) = -p \log p$, $\varphi_1(x, p) = p$, $\varphi_2(x, p) = xp$. Тогда мы имеем задачу

$$J(p) = - \int_0^{\infty} p \log p dx \rightarrow \max_p$$

при условиях

$$\int_0^{\infty} \varphi_1(x, p) dx = 1, \quad \int_0^{\infty} \varphi_2(x, p) dx = \mu^{-1}. \quad (4.9)$$

Уравнение (4.7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \lambda_2 \frac{\partial \varphi_2}{\partial p} = -(1 + \log p) + \lambda_1 + \lambda_2 x = 0.$$

Решением этого уравнения является функция $p(x) = e^{\lambda_1-1+\lambda_2 x}$. Подставляя ее в условия (4.9), находим, что $e^{\lambda_1-1} = -\lambda_2$ и $\lambda_2 = -\mu$, то есть $p(x) = \mu e^{-\mu x} \mathbf{1}(x \geq 0)$.

3°. Не ограничивая общности, будем считать, что $a = 0$. Положим $F(x, p) = -p \log p$, $\varphi_1(x, p) = p$, $\varphi_2(x, p) = x^2 p$. Тогда мы имеем задачу

$$J(p) = - \int_{-\infty}^{\infty} p \log p dx \rightarrow \max_p$$

при условиях

$$\int_{-\infty}^{\infty} \varphi_1(x, p) dx = 1, \quad \int_{-\infty}^{\infty} \varphi_2(x, p) dx = \sigma^2. \quad (4.10)$$

Уравнение (4.7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \lambda_2 \frac{\partial \varphi_2}{\partial p} = -(1 + \log p) + \lambda_1 + \lambda_2 x^2 = 0.$$

Решением этого уравнения является функция $p(x) = e^{\lambda_1 - 1 + \lambda_2 x^2}$. Подставляя ее в условия (4.10), находим, что $e^{\lambda_1 - 1} = \sqrt{-\lambda_2 / \pi}$ и $\lambda_2 = -\frac{1}{2} \sigma^{-2}$, откуда $e^{\lambda_1 - 1} = \frac{1}{\sigma \sqrt{2\pi}}$, то есть $p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\}$, $-\infty < x < \infty$. Теорема доказана.

4.2 Случайные процессы

Определение 4.5. Случайным процессом называется семейство случайных величин $X(t) = X(\omega, t)$, заданных на одном (базовом) вероятностном пространстве $(\Omega, \mathcal{A}, P)$ и зависящих от параметра t , принимающего значения из некоторого множества T .

Из определения 4.5 следует, что случайный процесс имеет двоякую сущность. С одной стороны, если $\omega_0 \in \Omega$ фиксировано, то $X(t) = X(\omega_0, t)$ есть некоторая функция от $t \in T$. С другой стороны, если $t_0 \in T$ фиксировано, то $X(t_0) = X(\omega, t_0)$ – случайная величина, определенная на $(\Omega, \mathcal{A}, P)$. Эту случайную величину иногда называют *проекцией* случайного процесса в точке t_0 .

Очевидно, что последовательности случайных величин $X_1, X_2, \dots$ являются случайными процессами с $T = \{1, 2, \dots\}$.

Процессы, определенные на множестве $T = \{\dots, -1, 0, 1, \dots\}$ или его подмножестве, обычно называются *процессами с дискретным временем* или *случайными последовательностями*. Если же множество T представляет собой интервал (конечный или бесконечный), то семейство случайных величин $X(t) = X(\omega, t)$ принято называть *случайным процессом с непрерывным временем*. Тем не менее, параметр t в определении случайного процесса, конечно, не обязан иметь смысл времени.

Рассмотрим случайный процесс $X(t) = X(\omega, t)$. Как уже было отмечено, при фиксированном $\omega_0 \in \Omega$ мы получаем функцию $X(t) = X(\omega_0, t)$, $t \in T$, которую принято называть *траекторией* случайного процесса $X(t)$. Пусть $\mathcal{S}$ – функциональное пространство всех возможных траекторий случайного процесса $X(t)$. В этом случае случайный процесс $X(t)$ можно рассматривать как случайный элемент, принимающий значения в пространстве $\mathcal{S}$. Остановимся на данной трактовке определения 4.5 более подробно.

Итак, пусть $\mathcal{S}$ – некоторое метрическое пространство. Рассмотрим вероятностные меры, определенные на классе Σ всех борелевских подмножеств пространства $\mathcal{S}$. Класс Σ представляет собой σ -алгебру, порожденную всеми открытыми подмножествами пространства $\mathcal{S}$, то есть минимальную σ -алгебру, содержащую все открытые подмножества. Вероятностная мера P , заданная на σ -алгебре Σ , – это неотрицательная счетно-аддитивная функция множеств $P : \Sigma \rightarrow [0, 1]$ такая, что $P(\mathcal{S}) = 1$.

Пусть X есть отображение базового вероятностного пространства $(\Omega, \mathcal{A}, P)$ в метрическое пространство $\mathcal{S}$. Если отображение X является измеримым, то есть $\{\omega \mid X(\omega) \in B\} \in \mathcal{A}$ для любого множества $B \in \Sigma$, то X называется *случайным элементом*. Причем, если пространство $\mathcal{S}$ является пространством вещественнозначных функций, то X принято называть *случайной функцией* или *случайным процессом*.

Определение 4.6. Распределением случайного процесса X называется вероятностная мера P_X , заданная на измеримом пространстве $(\mathcal{S}, \Sigma)$, определенная для всех множеств $A \in \Sigma$ следующим

образом:

$$P_X(A) = P(\{\omega \mid X(\omega) \in A\}) \equiv P(X \in A).$$

При фиксированных значениях $t_1, \dots, t_k$, $t_i \in T$, $i = 1, \dots, k$, мы получаем k -мерный случайный вектор $(X(\omega, t_1), \dots, X(\omega, t_k))$. Распределения этих случайных векторов для различных $k \geq 1$ и $t_1, \dots, t_k$ называются *конечномерными распределениями* процесса $\{X(t), t \in T\}$.

Определение 4.7. Случайный процесс $\{X(t), t \in T \subseteq \mathbb{R}\}$ называется процессом с независимыми приращениями, если для любых $n \geq 1$ и любых $t_i \in T$, $i = 0, \dots, n$, таких, что $t_0 < t_1 < \dots < t_n$, случайные величины

$$X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$$

независимы в совокупности.

Определение 4.8. Случайный процесс $\{X(t), t \in T\}$ называется однородным (по времени), если

$$X(t+h) - X(t) \stackrel{d}{=} X(s+h) - X(s)$$

для всех t, s и $h \geq 0$ таких, что $t \in T, t+h \in T, s \in T, s+h \in T$ (знак $\stackrel{d}{=}$ обозначает равенство распределений случайных величин).

Определение 4.9. Случайный процесс $\{X(t), t \in T\}$ называется стохастически непрерывным, если для любых $t \in T$ и $\varepsilon > 0$

$$\lim_{s \rightarrow t, s \in T} P(|X(s) - X(t)| > \varepsilon) = 0.$$

Определение 4.10. Однородный стохастически непрерывный случайный процесс с независимыми приращениями, определенный на $T = [0, \infty)$, P -почти все траектории которого выходят из нуля, непрерывны справа и имеют конечные пределы слева, называется процессом Леви.

Важным примером случайного процесса с непрерывными траекториями, который можно рассматривать как модель непрерывного хаотического движения, является винеровский процесс, служащий для описания координаты частицы, подверженной броуновскому движению.

Определение 4.11. Процесс Леви $W_{a,\sigma^2}(t)$, $a \in \mathbb{R}$, $\sigma > 0$, такой, что при $t_0 > 0$ проекция $W_{a,\sigma^2}(t_0)$ имеет нормальное распределение с параметрами at_0 и $\sigma^2 t_0$, называется винеровским процессом или процессом броуновского движения с коэффициентами сноса a и диффузии σ^2 . Винеровский процесс $W_{0,1}(t)$ называется стандартным винеровским процессом.

Еще одним случайным процессом, широко используемым в качестве математической модели хаотических потоков событий, является пуассоновский процесс.

Определение 4.12. Процесс Леви $N_\lambda(t)$ называется пуассоновским процессом с параметром $\lambda > 0$, если при $t_0 > 0$ проекция $N_\lambda(t_0)$ имеет пуассоновское распределение с параметром λt_0 .

Ниже эти два типа процессов будут рассмотрены подробнее в качестве базовых моделей многих реальных явлений.

4.3 Пуассоновский процесс как модель дискретного хаоса

Здесь и в дальнейшем мы делаем акцент на тех свойствах пуассоновского процесса, которые позволяют считать его основной математической моделью потока событий, абсолютно хаотично рассредоточенных во времени.

Приведем еще одно определение пуассоновского процесса (по Хинчину).

Определение 4.13. Случайный процесс $\{N(t), t \geq 0\}$, называется пуассоновским, если он обладает следующими свойствами:

1. $N(t)$ – процесс с независимыми приращениями, то есть для любых натуральных n и $t_0, \dots, t_n \in \mathbb{R}$ ($0 \leq t_0 < t_1 < \dots < t_n$) случайные величины $N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1})$ независимы в совокупности;
2. $N(t)$ – однородный процесс, то есть для любых неотрицательных s, t и $h > 0$ случайные величины $N(t+h) - N(t)$ и $N(s+h) - N(s)$ одинаково распределены;
3. $N(0) = 0$ почти наверное;
4. при $h \downarrow 0$ и некотором $\lambda > 0$

$$P(N(h) = 0) = 1 - \lambda h + o(h),$$

$$P(N(h) = 1) = \lambda h + o(h),$$

$$P(N(h) \geq 2) = o(h).$$

Найдем распределение случайной величины $N(t)$ при произвольном $t > 0$. С этой целью рассмотрим ее производящую функцию $\psi_t(s) = E s^{N(t)}$, определенную при $|s| \leq 1$. По свойствам 1, 2 и 3 для произвольного $h > 0$ мы имеем

$$\begin{aligned} \psi_{t+h}(s) &= E s^{N(t+h)} = E s^{N(t+h)-N(t)+N(t)} = \\ &= E s^{N(t)} E s^{N(t+h)-N(t)} = E s^{N(t)} E s^{N(h)} = \psi_t(s) \psi_h(s). \end{aligned} \quad (4.11)$$

По свойству 4 при $h \downarrow 0$ в силу ограниченности s

$$\psi_h(s) = (1 - \lambda h + o(h)) + s(\lambda h + o(h)) + o(h) = 1 + \lambda h(s - 1) + o(h).$$

Подставляя это выражение в (4.11), мы получаем

$$\psi_{t+h}(s) = \psi_t(s)(1 + \lambda h(s - 1) + o(h)),$$

откуда

$$\frac{\psi_{t+h}(s) - \psi_t(s)}{h} = \lambda(s - 1)\psi_t(s) + o(1). \quad (4.12)$$

Устремляя в (4.12) $h \downarrow 0$, мы приходим к дифференциальному уравнению

$$\frac{d\psi_t(s)}{dt} = \lambda(s - 1)\psi_t(s),$$

решением которого, удовлетворяющим начальному условию $\psi_0(s) \equiv 1$ (см. свойство 3), является функция

$$\psi_t(s) = \exp\{\lambda t(s-1)\}. \quad (4.13)$$

Раскладывая функцию (4.13) в ряд по степеням s , мы получаем

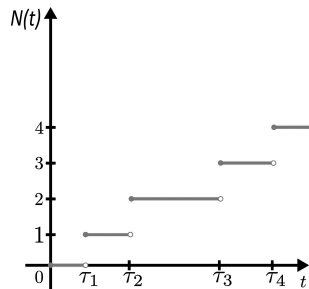
$$\psi_t(s) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} s^k. \quad (4.14)$$

Функция (4.14), будучи степенным рядом, однозначно определяется коэффициентами при степенях s . Но с учетом определения производящей функции это означает, что

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \dots$$

Другими словами, случайная величина $N(t)$ имеет распределение Пуассона с параметром λt .

Итак, пуассоновский процесс принимает только целые неотрицательные значения. При этом свойство 4 означает, что траектории пуассоновского процесса не убывают, кусочно-постоянны, непрерывны справа, а их скачки имеют одинаковую величину, равную единице.



Траектория
пуассоновского процесса

Мы сразу же замечаем, что

$$EN(t) \equiv DN(t) \equiv \lambda t.$$

Поэтому λ имеет смысл среднего числа скачков пуассоновского процесса за единицу времени. Параметр λ называется *интенсивностью пуассоновского процесса*. Пуассоновский процесс с единичной интенсивностью договоримся называть *стандартным пуассоновским процессом*.

Пусть $\tau_0 = 0, \tau_1, \dots$ – точки скачков пуассоновского процесса. Большой интерес представляет вопрос о том, каковы распределения случайных величин $\tau_n - \tau_{n-1}, n \geq 1$, то есть о распределении длин интервалов времени между скачками пуассоновского процесса. Вначале мы дадим не вполне строго обоснованный ответ на этот вопрос.

Итак, пуассоновский процесс однороден, поэтому величины $\tau_n - \tau_{n-1}, n \geq 1$, распределены одинаково. Пуассоновский процесс имеет независимые приращения, поэтому эти величины независимы. Вследствие однородности, чтобы определить тип распределения этих величин, нам достаточно рассмотреть лишь одну из них, скажем, $\tau_1 - \tau_0 = \tau_1$. Событие $\{\tau_1 > t\}$ эквивалентно событию $\{N(t) = 0\}$. Поэтому

$$P(\tau_1 \leq t) = 1 - P(\tau_1 > t) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t}.$$

Другими словами, случайные величины $\tau_n - \tau_{n-1}, n \geq 1$, независимы и имеют одно и то же показательное распределение с параметром λ .

Если точки скачков пуассоновского процесса отождествить с моментами регистрации некоторых одностипных событий, то $N(t)$ принимает смысл общего количества событий, зарегистрированных до момента t . В этом смысле оказывается очень полезно рассматривать пуассоновский процесс как точечный процесс на полупрямой $t \geq 0$.

Определение 4.14. Последовательность случайных величин $\{\tau_n\}_{n \geq 1}$ называется *точечным процессом*, если она удовлетворяет следующим условиям:

- 1° если $\tau_n < \infty$, то $\tau_{n+1} > \tau_n$;
- 2° для всякого $t < \infty$ найдется такое n , что $\tau_n > t$.

Со всяким точечным процессом $\{\tau_n\}_{n \geq 1}$ можно связать случайную целочисленную неотрицательную (читающую) меру $\nu(A)$, определенную на борелевских множествах A , положив

$$\nu(A) = \sum_{k=1}^{\infty} \mathbf{1}(\tau_k \in A).$$

Реализацией случайной считающей меры является обычная считающая мера. Каждая траектория точечного процесса однозначно определяет реализацию случайной считающей меры и наоборот. Поэтому иногда точечные процессы определяют как случайные меры. Например, если $A = [0, t)$, а $\{\tau_n\}_{n \geq 1}$ – точки скачков пуассоновского процесса $N(t)$, то, очевидно,

$$\nu([0, t)) = N(t).$$

Пусть $n \geq 1$ – произвольное целое число, $[a, b]$ – произвольный непустой отрезок, $\{\tau_j\}_{j \geq 1}$ – точки скачков пуассоновского процесса с некоторой интенсивностью $\lambda > 0$. Найдем условное совместное распределение точек скачков пуассоновского процесса, попавших в отрезок $[a, b]$, при условии, что на этом отрезке пуассоновский процесс имеет ровно n скачков. Перенумеруем точки скачков так, чтобы τ_1 оказалась наименьшей из точек скачков, попавших в $[a, b]$.

Теорема 4.3. *Условное совместное распределение случайных величин $\tau_1, \dots, \tau_n$ при условии $N(b) - N(a) = n$ совпадает с совместным распределением вариационного ряда, построенного по выборке объема n из равномерного распределения на $[a, b]$.*

Доказательство. Если $(\eta_1, \dots, \eta_n)$ – случайный вектор с совместной плотностью $f(x_1, \dots, x_n)$, $h > 0$, то при $h \downarrow 0$

$$P(\eta_j \in [x_j, x_j + h), j = 1, \dots, n) = f(x_1, \dots, x_n)h^n + o(h^n).$$

Мы будем использовать это соотношение. Пусть

$$a = t_0 + h < t_1 < \dots < t_n < t_{n+1} = b,$$

а $h > 0$ столь мало, что $h < \min_{0 \leq j \leq n} (t_{j+1} - t_j)$. Найдем условную вероятность $P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid N(b) - N(a) = n)$. Очевидно, что

$$\begin{aligned} & P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid N(b) - N(a) = n) = \\ &= \frac{P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n; N(b) - N(a) = n)}{P(N(b) - N(a) = n)}. \end{aligned} \quad (4.15)$$

Рассмотрим событие, стоящее под знаком вероятности в числителе правой части соотношения (4.15). Оно эквивалентно тому, что в интервалы $[t_{j-1} + h, t_j), j = 1, \dots, n+1$, не попала ни одна точка скачков, а в каждый из интервалов $[t_j, t_j + h), j = 1, \dots, n$, попало ровно по одной точке скачков. Поэтому, используя однородность пуассоновского процесса и независимость его приращений мы будем иметь

$$\begin{aligned} & P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n; N(b) - N(a) = n) = \\ &= P(N(t_j) - N(t_{j-1} + h) = 0, j = 1, \dots, n+1; \\ & \quad N(t_j + h) - N(t_j) = 1, j = 1, \dots, n) = \\ &= (\lambda h)^n \exp\{-n\lambda h\} \times \\ & \quad \times \exp\{-\lambda[t_1 - a + t_2 - t_1 - h + \dots + t_n - t_{n-1} - h + b - t_n - h]\} = \\ & \quad = (\lambda h)^n \exp\{-\lambda(b - a)\}. \end{aligned} \quad (4.16)$$

Далее,

$$P(N(b) - N(a) = n) = \exp\{-\lambda(b - a)\} \frac{\lambda^n (b - a)^n}{n!} \quad (4.17)$$

Подставляя (4.16) и (4.17) в (4.15), получаем

$$P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid N(b) - N(a) = n) = \frac{n!}{(b - a)^n} h^n.$$

С учетом сказанного в самом начале доказательства мы заключаем, что совместная условная плотность случайных величин $\tau_1, \dots, \tau_n$ при условии $N(b) - N(a) = n$ равна

$$\frac{n!}{(b - a)^n} \mathbf{1}(a \leq t_1 < \dots < t_n \leq b).$$

Но, как известно, именно такой вид имеет совместная плотность порядковых статистик, построенных по выборке объема n из равномерного распределения на $[a, b]$. Теорема доказана.

Вернемся к вопросу о совместном распределении случайных величин $\zeta_j = \tau_j - \tau_{j-1}$, $j = 1, \dots, n$, в пуассоновском точечном процессе. В соотношении (4.16) положим $a = h$, $b = t_n + h$. Тогда

$$\begin{aligned} P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n; N(b) - N(a) = n) = \\ = P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n) = \lambda^n \exp\{-\lambda t_n\} h^n. \end{aligned}$$

Как и при доказательстве Теоремы 4.3, отсюда мы заключаем, что для совместной плотности $p_{\tau_1, \dots, \tau_n}(t_1, \dots, t_n)$ случайных величин $\tau_1, \dots, \tau_n$ справедливо соотношение

$$p_{\tau_1, \dots, \tau_n}(t_1, \dots, t_n) = \lambda^n \exp\{-\lambda t_n\}.$$

Но якобиан преобразования $\tau_j \mapsto \zeta_j = \tau_j - \tau_{j-1}$, $j = 1, \dots, n$, равен единице ($\tau_0 = 0$). Поэтому, обозначая $s_j = t_j - t_{j-1}$, $j = 1, \dots, n$, для совместной плотности $p_{\zeta_1, \dots, \zeta_n}(s_1, \dots, s_n)$ случайных величин $\zeta_1, \dots, \zeta_n$ при $s_j \geq 0$, $j = 1, \dots, n$, мы получаем представление

$$p_{\zeta_1, \dots, \zeta_n}(s_1, \dots, s_n) = \lambda^n \exp\{-\lambda(s_1 + \dots + s_n)\} = \prod_{j=1}^n \lambda e^{-\lambda s_j},$$

что означает, что случайные величины $\zeta_1, \dots, \zeta_n$ независимы и одинаково показательно распределены с параметром λ .

Как мы уже установили, равномерное распределение является “наиболее непредсказуемым” среди всех распределений, сосредоточенных на конечном интервале (оно имеет максимальную дифференциальную энтропию среди всех таких распределений). Другими словами, равномерное распределение лучше других соответствует представлению об абсолютно хаотичном расположении точек на отрезке. Поэтому доказанная выше Теорема 4.3 убедительно свидетельствует о том, что пуассоновский точечный процесс является как нельзя более подходящей моделью потока случайных событий, абсолютно хаотично рассредоточенных во времени.

В Теореме 4.3 концы отрезка $[a, b]$ выбирались неслучайно. Оказывается, что если концы рассматриваемого отрезка отождествить с какими-либо точками скачков пуассоновского процесса, сделав их тем самым случайными, то свойство равномерности распределения точек скачков, расположенных внутри такого случайного отрезка, сохранится. Действительно, очень легко убедиться, что условная плотность $p_{\zeta_j | \zeta_j + \zeta_{j+1} = s}(x)$ случайной величины ζ_j при условии $\zeta_j + \zeta_{j+1} = s$, где $s > 0$ произвольно, равна $s^{-1} \mathbf{1}(0 \leq x \leq s)$, что соответствует равномерному распределению на отрезке $[0, s]$.

Из пункта 2° Теоремы 4.2 вытекает, что пуассоновский процесс является наиболее неопределенным, наиболее хаотичным среди всех процессов восстановления с конечными математическими ожиданиями и абсолютно непрерывными распределениями длин ζ_j , $j \geq 1$, промежутков времени между последовательными “восстановлениями”, так как его характеризует показательное распределение случайных величин ζ_j , $j \geq 1$.

Таким образом, установлена связь пуассоновского процесса с двумя наиболее неопределенными распределениями – равномерным и показательным.

Подводя промежуточный итог, можно сделать следующий вывод. Теоремы 4.2 и 4.3 являются убедительными аргументами в пользу того, что пуассоновский процесс является самой подходящей математической моделью однородных и абсолютно хаотичных потоков однотипных событий.

Оказывается, есть и асимптотическая связь этого процесса с третьим наиболее непредсказуемым распределением – нормальным.

Пусть $N(t)$, $t \geq 0$, – пуассоновский процесс с интенсивностью $\lambda > 0$. Как мы видели выше, $EN(t) \equiv DN(t) \equiv \lambda t$. Пуассоновский процесс является асимптотически нормальным в том смысле, что при любом $x \in \mathbb{R}$

$$P\left(\frac{N(t) - \lambda t}{\sqrt{\lambda t}} < x\right) \rightarrow \Phi(x) \text{ при } \lambda t \rightarrow \infty. \quad (4.18)$$

На самом деле справедливо более сильное утверждение, из ко-

торого вытекает (4.18). Оказывается, что сходимость функций распределения, участвующих в (4.18), равномерна по x и справедлива следующая оценка скорости этой равномерной сходимости.

Теорема 4.4. (Korolev, Shevtsova, 2012) При любых $\lambda > 0$, $t > 0$

$$\sup_{x \in \mathbb{R}} \left| P \left(\frac{N(t) - \lambda t}{\sqrt{\lambda t}} < x \right) - \Phi(x) \right| \leq \frac{0.3041}{\sqrt{\lambda t}}.$$

Обратим внимание, что пуассоновский процесс обладает свойством асимптотической нормальности при $\lambda t \rightarrow \infty$, что выполнено, например, при фиксированной интенсивности, но при неограниченно растущем времени или при фиксированном времени, но при неограниченно растущей интенсивности.

4.4 Принципы выбора моделей в рамках энтропийного подхода

Распределения, участвующие в теореме 2 в качестве распределений с максимальной дифференциальной энтропией, то есть в качестве наиболее неопределенных, возникают в предельных теоремах теории вероятностей.

- **Нормальное распределение** возникает в центральной предельной теореме и ее модификациях, описывающих асимптотическое поведение распределений случайных величин с аддитивной структурой при неограниченном увеличении числа случайных аддитивных составляющих.
- **Показательное распределение** возникает в предельных теоремах для случайных величин типа времени жизни, в зависимости которых от “атомических случайных величин” операция суммирования заменена операцией взятия максимума (или минимума). Также оно, будучи тесно связанным с пуассоновским процессом, появляется и в предельных теоремах для сумм случайных величин, описывающих потоки редких событий.

- **Равномерное распределение** возникает в предельных теоремах для цензурированных данных, описывая поведение погрешностей округления.

Формальное совпадение энтропии по Шеннону, описанной выше, с хорошо известной из курса физики больцмановской энтропией наводит на заключение о том, что предельные теоремы теории вероятностей сами по себе являются математическими (вероятностными) моделями проявления универсального принципа неубывания энтропии в энергетически или информационно замкнутых системах.

Из этого немедленно вытекает, что три типа распределений, упомянутых выше, отнюдь не исчерпывают все распределения с условно максимальной энтропией. Ближайшими примерами служат устойчивые законы, возникающие в предельных теоремах Леви–Хинчина для сумм независимых одинаково распределенных случайных величин с бесконечными дисперсиями.

Поэтому, если какая-либо вероятностная модель, отличная от упомянутых в теореме 4.2, оказалась адекватной, следует исследовать те условия, при которых она обладает свойствами максимума энтропии. Описание таких условий может дать дополнительное понимание сути наблюдаемых закономерностей.

Универсальный принцип неубывания энтропии описывает поведение замкнутых систем. Информационная или энергетическая “замкнутость” системы, поведение которой моделируется на языке теории вероятностей, характеризуется тем, что в используемой модели рассматривается только один источник случайности – элементарные (атомические) случайные величины. На практике ни одна из реальных систем замкнутой не является. Поэтому в модели разумно вводить дополнительные источники случайности, “отвечающие” за влияние внешней среды на рассматриваемую систему. При этом естественно возникают смешанные вероятностные модели, основанные на уже описанных предельных распределениях с максимальной энтропией, но параметры которых становятся случайными. Далее мы будем рассматривать именно такие смешанные (или, на современном языке, *байесовские*) модели.

Глава 5

Случайные суммы

5.1 Элементарные свойства случайных сумм

В некоторых прикладных задачах информация накапливается не до достижения какого-то заданного числа наблюдений, а в течение фиксированного промежутка времени. Например, отчет страховой компании может содержать данные о клиентах, заключивших контракты в течение определенного месяца. В таких ситуациях естественно считать, что число наблюдений тоже является случайной величиной, и тогда характеристики исследуемой системы описываются суммами со случайным числом случайных слагаемых.

Пусть $X_1, X_2, \dots$ – последовательность независимых одинаково распределенных случайных величин. Пусть N – неотрицательная целочисленная случайная величина. Предположим, что случайные величины $N, X_1, X_2, \dots$ определены на одном и том же вероятностном пространстве $(\Omega, \mathcal{A}, P)$ и независимы. Под случайной суммой $S_N = X_1 + \dots + X_N$ мы будем понимать случайную величину, которая при каждом $\omega \in \Omega$ принимает значение $X_1(\omega) + \dots + X_{N(\omega)}(\omega)$.

Для определенности мы полагаем $\sum_{k=1}^0 = 0$.

Обозначим $p_n = P(N = n)$, $n = 0, 1, \dots$. Производящую функцию случайной величины N обозначим $\psi(s)$, $|s| \leq 1$. Функцию распределения, плотность (если она существует) и характеристическую функцию случайной величины X_1 обозначим соответственно $F(x)$, $p(x)$ и $f(t)$. Основные сведения о распределении случайной суммы содержатся в следующей теореме.

Теорема 5.1. *Справедливы следующие утверждения:*

1°. Функция распределения $F_{S_N}(x)$ случайной суммы S_N имеет вид

$$F_{S_N}(x) = \sum_{n=0}^{\infty} p_n F^{*n}(x),$$

где $F^{*n}(x)$ – n -кратная свертка функции распределения F , причем $F^{*0}(x)$ – функция распределения с единственным единичным скачком в нуле.

2°. Если $p_0 > 0$, то $F_{S_N}(x)$ не является абсолютно непрерывной, даже если X_1 абсолютно непрерывна. Если $p_0 = 0$ и X_1 имеет плотность $p(x)$, то плотность $p_{S_N}(x)$ случайной суммы существует и равна

$$p_{S_N}(x) = \sum_{n=1}^{\infty} p_n p^{*n}(x),$$

где $p^{*n}(x)$ – n -кратная свертка $p(x)$.

3°. Характеристическая функция $f_{S_N}(t)$ случайной суммы S_N имеет вид

$$f_{S_N}(t) = \psi(f(t)).$$

Математическое ожидание и дисперсию случайной суммы можно вычислить по формулам

$$ES_N = ENEX_1, \quad DS_N = ENDX_1 + DN(EX_1)^2$$

при условии, что все величины в правых частях существуют.

Доказательство. Два первых утверждения следуют из формулы полной вероятности и малой теоремы Фубини. Докажем пункт 3°. Прежде всего заметим, что для $n \geq 1$ мы имеем

$$\int_{-\infty}^{\infty} e^{itx} dF^{*n}(x) = E \exp\{it(X_1 + \dots + X_N)\} = f^n(t), \quad t \in \mathbb{R}.$$

Поэтому по пункту 1° мы получаем

$$f_{S_N}(t) = \int_{-\infty}^{\infty} e^{itx} dF_{S_N}(x) = \int_{-\infty}^{\infty} e^{itx} d \left(\sum_{n=0}^{\infty} p_n F^{*n}(x) \right) =$$

$$= \sum_{n=0}^{\infty} p_n \int_{-\infty}^{\infty} e^{itx} dF^{*n}(x) = \sum_{n=0}^{\infty} p_n f^n(t) = \psi(f(t)).$$

Теперь докажем пункт 4°. Чтобы получить ES_N , воспользуемся хорошо известным свойством характеристических функций и только что доказанным утверждением. Мы получаем

$$\begin{aligned} ES_N &= \frac{1}{i} \cdot \frac{df_{S_N}(t)}{dt} \Big|_{t=0} = \frac{1}{i} \cdot \frac{d\psi(f(t))}{dt} \Big|_{t=0} = \frac{1}{i} \cdot \frac{d\psi(f(t))}{df(t)} \cdot \frac{df(t)}{dt} \Big|_{t=0} = \\ &= \frac{1}{i} EN \cdot iEX_1 = ENEX_1. \end{aligned}$$

Таким же образом вычислим и дисперсию случайной величины S_N . Имеем

$$\begin{aligned} ES_N^2 &= -\frac{d^2 f_{S_N}(t)}{dt^2} \Big|_{t=0} = -\frac{d}{dt} \left(\frac{d\psi(f(t))}{df(t)} \cdot \frac{df(t)}{dt} \right) \Big|_{t=0} = \\ &= -\frac{d^2 \psi(f(t))}{(df(t))^2} \cdot \left(\frac{df(t)}{dt} \right)^2 \Big|_{t=0} - \frac{d\psi(f(t))}{df(t)} \cdot \frac{d^2 f(t)}{dt^2} \Big|_{t=0} = \\ &= EN(N-1)(EX_1)^2 + ENEX_1^2 = EN^2(EX_1)^2 + ENDX_1. \end{aligned}$$

Поэтому

$$\begin{aligned} DS_N &= ES_N^2 - (ES_N)^2 = EN^2(EX_1)^2 + ENDX_1 - (EN)^2(EX_1)^2 = \\ &= ENDX_1 + DN(EX_1)^2. \end{aligned}$$

Теорема доказана.

5.2 Пуассоновские и геометрические случайные суммы

Рассмотрим два частных типа случайных сумм: пуассоновские (случайные) суммы и геометрические (случайные) суммы. Пуассоновские суммы могут использоваться при моделировании ситуаций, в которых наблюдения накапливаются в течение фиксированного временного интервала и моменты их регистрации хаотично рассредоточены во времени. Геометрические случайные суммы

моделируют ситуацию, в которой наблюдения накапливаются до наступления некоторого события, в результате которого сбор данных прекращается.

Пусть случайная величина N имеет распределение Пуассона. Тогда случайная сумма S_N называется *пуассоновской случайной суммой* или просто *пуассоновской суммой*. Распределение пуассоновской случайной суммы называется *обобщенным пуассоновским*. Из теоремы 5.1 мы получаем следующие элементарные свойства пуассоновских сумм.

Теорема 5.2. Пусть случайная величина N имеет распределение Пуассона с параметром $\lambda > 0$.

1°. Характеристическая функция пуассоновской случайной суммы S_N имеет вид

$$f_{S_N}(t) = \exp\{\lambda(f(t) - 1)\}, \quad t \in \mathbb{R}.$$

Следовательно, распределение пуассоновской случайной суммы безгранично делимо.

2°. Если $EX_1^2 < \infty$ то

$$ES_N = \lambda EX_1, \quad DS_N = \lambda EX_1^2$$

Сформулируем также теорему об асимптотической нормальности пуассоновских случайных сумм. Как обычно, $\Phi(x)$ обозначает стандартную нормальную функцию распределения. Символ $\Rightarrow$ будет обозначать сходимость по распределению (слабую сходимость).

Теорема 5.3. Предположим, что $EX_1^2 < \infty$. Обозначим $EX_1 = a$, $DX_1 = \sigma^2$. Тогда при $\lambda \rightarrow \infty$

$$P\left(\frac{S_N - a\lambda}{\sqrt{\lambda(a^2 + \sigma^2)}} < x\right) \Rightarrow \Phi(x)$$

равномерно по $x \in \mathbb{R}$.

Теперь рассмотрим геометрические случайные суммы. Если случайная величина N имеет геометрическое распределение

$$P(N = n) = p(1 - p)^n, \quad n = 0, 1, \dots, \quad (5.1)$$

то случайная величина S_N называется *геометрической случайной суммой* или просто *геометрической суммой*.

Несложно видеть, что производящая функция геометрического распределения имеет вид

$$\psi_N(s) = Es^N = \frac{p}{1 - (1 - p)s}, \quad (5.2)$$

тогда как

$$EN = \frac{1 - p}{p}, \quad DN = \frac{1 - p}{p^2}.$$

В соответствии с Теоремой 5.1 характеристическая функция геометрической случайной суммы имеет вид

$$f_{S_N}(t) = \frac{p}{1 - (1 - p)f(t)}.$$

Пусть $G(x)$ – функция распределения стандартного показательного распределения, то есть $G(x) = 1 - e^x$ при $x \geq 0$ и $G(x) = 0$ при $x < 0$. Асимптотическое поведение геометрических случайных сумм описано в следующей теореме.

Теорема 5.4. Пусть случайная величина N имеет геометрическое распределение (5.1).

1° (теорема Реньи). Пусть случайная величина X_1 неотрицательна и $0 < a = EX_1 < \infty$. Тогда при $p \rightarrow 0$

$$\sup_{x \in \mathbb{R}} P \left| (pa^{-1}S_N < x) - G(x) \right| \rightarrow 0.$$

2°. Пусть случайная величина X_1 неотрицательна, $a = EX_1 > 0$ и $b^2 = EX_1^2 < \infty$. Тогда

$$\sup_{x \in \mathbb{R}} P \left| (pa^{-1}S_N < x) - G(x) \right| \leq \frac{pb^2}{(1 - p)a^2}.$$

Рассмотрим теперь взаимосвязь пуассоновских и геометрических случайных сумм.

Лемма 5.1. Пусть M – целочисленная случайная величина с обобщенным пуассоновским распределением, $X_1, X_2, \dots$ – одинаково распределенные случайные величины с общей характеристической функцией $f(t)$. Предположим, что случайные величины $M, X_1, X_2, \dots$ независимы. Тогда случайная величина

$$S_M = X_1 + \dots + X_M$$

является пуассоновской случайной суммой.

Доказательство. Производящая функция $\psi_M(s)$ случайной величины M имеет вид $\psi_M(s) = \exp\{\lambda(\psi(s) - 1)\}$, где $\psi(s)$ – некоторая производящая функция. В соответствии с пунктом 3 Теоремы 5.1 характеристическая функция $f_{S_M}(t)$ случайной суммы S_M имеет вид

$$f_{S_M}(t) = \psi_M(f(t)) = \exp\{\lambda(\psi(f(t)) - 1)\}. \quad (5.3)$$

Но из пункта 1 Теоремы 5.2 и пункта 3 Теоремы 5.1 вытекает, что (5.3) является характеристической функцией пуассоновской случайной суммы $Y_1 + \dots + Y_N$, в которой случайная величина N имеет распределение Пуассона с параметром λ , $Y_j \stackrel{d}{=} X_1 + \dots + X_L$, $j = 1, 2, \dots$, L – случайная величина, производящая функция которой равна $\psi(s)$, и случайные величины $N, Y_1, Y_2, \dots$ независимы, как и случайные величины $L, X_1, X_2, \dots$. Другими словами, $S_M \stackrel{d}{=} Y_1 + \dots + Y_N$.

Лемма 5.2. Пусть M – случайная величина с геометрическим распределением (5.1). Тогда M является обобщенной пуассоновской случайной величиной.

Доказательство. Производящая функция случайной величины M имеет вид (5.2). Положим

$$\lambda = \log \frac{1}{p}, \quad \psi(s) = \frac{1}{\lambda} \log \frac{1}{1 - (1-p)s}.$$

Тогда несложно видеть, что $\psi_M(s) = \exp\{\lambda(\psi(s) - 1)\}$. Остается убедиться, что $\psi(s)$ – производящая функция. Но, используя разложение логарифмической функции в ряд Тейлора, мы видим, что

$$\psi(s) = \frac{1}{\lambda} \log \frac{1}{1 - (1-p)s} = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{[(1-p)s]^k}{k},$$

тогда как

$$\frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{(1-p)^k}{k} = 1,$$

то есть набор пар $\{(k, \lambda^{-1}(1-p)^k/k), k = 1, 2, \dots\}$ задает дискретное распределение вероятностей (это распределение обычно называют логарифмическим или распределением логарифмического ряда).

Из лемм 5.1 и 5.2 вытекает следующая теорема.

Теорема 5.5. *Любая геометрическая случайная сумма является пуассоновской случайной суммой. Более точно, если $M, X_1, X_2, \dots$ – независимые случайные величины такие, что M имеет геометрическое распределение (5.1) и $X_1, X_2, \dots$ одинаково распределены с общей характеристической функцией $f(t)$, то*

$$X_1 + \dots + X_M \stackrel{d}{=} Y_1 + \dots + Y_N,$$

где N – пуассоновская случайная величина с параметром $\log \frac{1}{p}$, случайные величины $N, Y_1, Y_2, \dots$ независимы, причем $Y_1, Y_2, \dots$ имеют одно и то же распределение с характеристической функцией

$$f_{Y_1}(t) = \frac{1}{\log \frac{1}{p}} \log \frac{1}{1 - (1-p)f(t)},$$

то есть $Y_j \stackrel{d}{=} X_1 + \dots + X_L$, $j = 1, 2, \dots$, где случайная величина L имеет логарифмическое распределение

$$P(L = k) = \frac{1}{\log \frac{1}{p}} \cdot \frac{(1-p)^k}{k}, \quad k = 1, 2, \dots$$

и независима от $X_1, X_2, \dots$

Следствие 5.1. Пусть S_N – пуассоновская случайная сумма с характеристической функцией $\exp\{\lambda(f(t) - 1)\}$. Если функция

$$g(t) = \frac{1 - e^{-\lambda f(t)}}{1 - e^{-\lambda}}$$

является характеристической, то S_N – геометрическая случайная сумма, $S_N \stackrel{d}{=} Y_1 + \dots + Y_M$, где случайные величины $M, Y_1, Y_2, \dots$ независимы, причем M имеет геометрическое распределение (5.1) с $p = e^{-\lambda}$, а $Y_1, Y_2, \dots$ одинаково распределены с общей характеристической функцией $g(t)$.

Следствие 5.2. Распределение любой геометрической случайной суммы безгранично делимо.

5.3 Теорема переноса

Чтобы иметь возможность проследить изменение распределения случайных сумм, когда распределение слагаемых может изменяться вместе с изменением распределения числа слагаемых, мы рассмотрим простейший вариант так называемой теоремы переноса для случайных сумм независимых одинаково распределенных случайных величин в схеме серий.

Пусть $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$ – последовательность серий независимых и одинаково в каждой серии распределенных случайных величин, а N_n , $n = 1, 2, \dots$ – положительные целочисленные случайные величины такие, что при каждом n случайная величина N_n независима от последовательности $\{X_{n,j}\}_{j \geq 1}$. Для натуральных k обозначим

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}.$$

Теорема 5.6. Предположим, что существуют неограниченно возрастающая последовательность натуральных чисел $\{m_n\}_{n \geq 1}$ и функции распределения $H(x)$ и $A(x)$ такие, что при $n \rightarrow \infty$

$$P(S_{n,m_n} < x) \implies H(x), \quad (5.4)$$

и

$$P(N_n < m_n x) \implies A(x). \quad (5.5)$$

Тогда существует функция распределения $F(x)$ такая, что при $n \rightarrow \infty$

$$P(S_{n,N_n} < x) \implies F(x). \quad (5.6)$$

При этом функция распределения $F(x)$ соответствует характеристической функции

$$f(t) = \int_0^{\infty} h^u(t) dA(u), \quad t \in \mathbb{R},$$

где $h(t)$ – характеристическая функция, соответствующая функции распределения $H(x)$.

Доказательство. Обозначим $g_n(t) = Ee^{itX_{n,1}}$, $h_n(t) = g_n^{m_n}(t)$, $A_n(x) = P(N_n < m_n x)$. По формуле полной вероятности имеем

$$\begin{aligned} f_n(t) &\equiv E \exp \left\{ it \sum_{j=1}^{N_n} X_{n,j} \right\} = \sum_{k=1}^{\infty} P(N_n = k) E \exp \left\{ it \sum_{j=1}^k X_{n,j} \right\} = \\ &= \sum_{k=1}^{\infty} P(N_n = k) g_n^k(t) = \int_0^{\infty} g_n^u(t) dP(N_n < u) = \\ &= \int_0^{\infty} g_n^{m_n u}(t) dA_n(u) = \int_0^{\infty} h_n^u(t) dA_n(u). \end{aligned}$$

Соотношение (5.6) будет доказано, если мы убедимся, что при каждом $t \in \mathbb{R}$

$$|f_n(t) - f(t)| \rightarrow 0 \quad (n \rightarrow \infty).$$

Фиксируем произвольное $t \in \mathbb{R}$. Имеем

$$|f_n(t) - f(t)| = \left| \int_0^{\infty} h_n^u(t) dA_n(u) - \int_0^{\infty} h^u(t) dA(u) \right| \leq$$

$$\begin{aligned} &\leq \left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h^u(t) dA_n(u) \right| + \\ &+ \left| \int_0^\infty h^u(t) dA_n(u) - \int_0^\infty h^u(t) dA(u) \right| \equiv I_1(n) + I_2(n). \end{aligned}$$

Вначале рассмотрим $I_2(n)$. Пусть N – случайная величина с функцией распределения $A(x)$. В силу условия (5.5) по определению слабой сходимости для любой непрерывной и ограниченной функции $\varphi(u)$ имеет место соотношение

$$E\varphi(N_n/m_n) \rightarrow E\varphi(N) \quad (n \rightarrow \infty).$$

Следовательно, так как функция $\varphi_z(u) = z^u$ ограничена и непрерывна по u при $|z| \leq 1$, то, полагая $z = h(t)$ (при таком z имеем $\varphi_z(u) = \varphi_{h(t)}(u) = h^u(t)$), мы замечаем, что в силу условия (5.5) имеет место сходимость $I_2(n) \rightarrow 0$ при $n \rightarrow \infty$. Таким образом, нам достаточно убедиться, что с помощью выбора большого n можно сделать $I_1(n)$ произвольно малым. Пусть $\varepsilon > 0$ – произвольное малое число. Для любого положительного M имеем

$$\begin{aligned} I_1(n) &\leq \int_0^M |h_n^u(t) - h^u(t)| dA_n(u) + \int_M^\infty |h_n^u(t) - h^u(t)| dA_n(u) \equiv \\ &\equiv I_{11}(n) + I_{12}(n). \end{aligned} \quad (5.7)$$

В силу слабой компактности семейства функций распределения $\{A_n(x)\}_{n \geq 1}$, обусловленной соотношением (5.5), существует такое $M = M(\varepsilon) > 0$, что

$$I_{12}(n) \leq 2 \int_M^\infty dA_n(u) \leq 2 \sup_{n \geq 1} [1 - A_n(M)] < \varepsilon. \quad (5.8)$$

Рассмотрим $I_{11}(n)$. Нам понадобится следующий аналог формулы Лагранжа для комплекснозначных функций

Лемма 5.3. Если комплекснозначная функция $\Psi(z)$ дифференцируема на множестве U и отрезок с концами a и b содержится в U , то

$$|\Psi(a) - \Psi(b)| \leq |b - a| \cdot \sup_{0 \leq \alpha \leq 1} |\Psi'(\alpha a + (1 - \alpha)b)|.$$

Полагая в Лемме 5.3 $\Psi(z) = z^u$, при каждом $u > 0$ мы получаем неравенство

$$\begin{aligned} |h_n''(t) - h''(t)| &\leq u |h_n(t) - h(t)| \sup_{0 \leq \alpha \leq 1} |\alpha h_n(t) + (1 - \alpha)h(t)|^{u-1} \leq \\ &\leq \frac{u |h_n(t) - h(t)|}{\min\{|h_n(t)|, |h(t)|\}}. \end{aligned}$$

Поскольку выполнено условие (5.4) и $m_n \rightarrow \infty$, характеристическая функция $h(t)$ безгранично делима, и стало быть, ни при каком $t \in \mathbb{R}$ не обращается в нуль. В силу условия (5.4) $h_n(t) \rightarrow h(t)$ при $n \rightarrow \infty$, и стало быть, найдутся такие $\delta > 0$ и n_0 , что $\min\{|h_n(t)|, |h(t)|\} \geq \delta$ для всех $n \geq n_0$. Поэтому для всех n , начиная с n_0 , справедливо неравенство

$$I_{11}(n) \leq \frac{M}{\delta} |h_n(t) - h(t)|.$$

Поэтому условие (5.4) позволяет выбрать $n_1 = n_1(\varepsilon) \geq n_0$ столь большим, чтобы

$$I_{11}(n) < \varepsilon \tag{5.9}$$

для всех $n \geq n_1$. Из (5.7), (5.8) и (5.9) следует, что $I_1(n) < 2\varepsilon$ при $n \geq n_1$. Таким образом, теорема доказана.

Теорема 5.6 впервые была доказана в статье (Гнеденко и Фахим, 1969). В той же работе впервые использован термин “теорема переноса”, подчеркивающий, что в этой теореме описывается перенос свойства сходимости с сумм неслучайного числа случайных слагаемых на случайные суммы и сопутствующая трансформация предельного закона. Здесь мы привели доказательство этого результата, отличающееся как от приведенного в работе (Гнеденко и Фахим, 1969), так и от доказательств в книгах (Круглов и Королев, 1990), (Gnedenko and Korolev, 1996).

В классической теории суммирования случайных величин безразлично, центрируются суммы или отдельные слагаемые, так как константы, центрирующие сумму, могут быть “распределены” между слагаемыми и наоборот. При случайном суммировании, когда число слагаемых в сумме случайно, центрирование слагаемых приводит к тому, что сама сумма центрируется случайной величиной (случайной суммой констант, центрирующих слагаемые). В то же время, для построения предельных или асимптотических аппроксимаций для распределения случайных сумм необходимо центрировать суммы константами. А это уже новая, более общая схема по сравнению с рассмотренной в Теореме 5.6. Следующую более общую теорему переноса, в которой рассматриваются центрированные константами случайные суммы, мы приведем без доказательства. Заинтересованный читатель может найти его в книге (Gnedenko and Korolev, 1996).

Теорема 5.7. Пусть для некоторых случайных величин Y , U и V и последовательностей $\{m_n\}_{n \geq 1}$ натуральных чисел, $\{a_n\}_{n \geq 1}$ и $\{c_n\}_{n \geq 1}$ действительных чисел при $n \rightarrow \infty$ выполнены условия

$$\begin{aligned} S_{n,m_n} - a_n &\Longrightarrow Y, \\ \frac{N_n}{m_n} &\Longrightarrow U, \\ a_n \frac{N_n}{m_n} - c_n &\Longrightarrow V. \end{aligned} \quad (5.10)$$

Тогда при $n \rightarrow \infty$

$$S_{n,N_n} - c_n \Longrightarrow Z,$$

где Z – случайная величина с характеристической функцией

$$f(t) = \mathbb{E}[h^U(t)e^{itV}], \quad t \in \mathbb{R}. \quad (5.11)$$

Будем говорить, что пара случайных величин (U, V) принадлежит к классу $\mathcal{H}_0$, если, во-первых, $P(U \geq 0) = 1$ и, во-вторых, либо хотя бы одна из двух случайных величин U и V вырождена, либо для некоторых действительных чисел α и β

$$P(V = \alpha U + \beta) = 1.$$

Если слагаемые $X_{n,j}$ предельно малы, то есть для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_{n,1}| > \varepsilon) = 0,$$

то, согласно теореме Хинчина о сходимости типов (Хинчин, 1938), с. 87 (см. также (Феллер, 1984), т. 2, с. 291, лемма 1), класс предельных законов для центрированных случайных сумм независимых одинаково распределенных слагаемых состоит из тех распределений, характеристические функции которых представимы в виде (5.11), где характеристическая функция h безгранично делима, а пара случайных величин (U, V) принадлежит к классу $\mathcal{H}_0$.

Рассмотрим структуру предельных законов (5.11) более подробно. Если случайная величина V вырождена, то для некоторого $\beta \in \mathbb{R}$

$$f(t) = e^{it\beta} E h^U(t), \quad t \in \mathbb{R}, \quad (5.12)$$

то есть с точностью до неслучайного сдвига, предельная характеристическая функция представляет собой степенную смесь безгранично делимых характеристических функций. Подобная ситуация рассмотрена в Теореме 5.6.

В частности, если для сумм S_{n,m_n} выполнена центральная предельная теорема, то есть $H(x) \equiv \Phi(x)$, тогда в Теореме 5.6 предельная функция распределения $F(x)$ имеет вид

$$F(x) = \int_0^\infty \Phi(x/\sqrt{u}) dA(u),$$

то есть является масштабной смесью нормальных законов (подобные смеси распределений ниже будут рассматриваться более подробно).

Если обе случайные величины U и V невырождены, то для некоторых α и β

$$f(t) = E[h^U(t) \exp\{it\alpha U\} \exp\{it\beta\}] = \exp\{it\beta\} E h_\alpha^U(t),$$

где $h_\alpha = e^{it\alpha} h(t)$, и так как h_α – безгранично делимая характеристическая функция, то мы опять находимся в рамках ситуации (5.12).

При рассмотрении неслучайно центрированных случайных сумм ситуация, когда величина U вырождена, занимает особое положение. В таком случае для некоторого $\gamma \geq 0$

$$f(t) = h^\gamma(t) \mathbb{E} \exp\{itV\} \quad (5.13)$$

и функция распределения, соответствующая характеристической функции из (5.11), принимает вид

$$F(x) = \int_{-\infty}^{\infty} H_\gamma(x-z) dG(z) = (H_\gamma * G)(x), \quad (5.14)$$

где H_γ – функция распределения, соответствующая характеристической функции $h^\gamma(t)$, а $G(z) = P(V < z)$, то есть $F(x)$ является сдвиговой смесью безгранично делимой функции распределения $H_\gamma(x)$.

Таким образом, если слагаемые $\{X_{n,j}\}$ предельно малы, то предельные распределения центрированных случайных сумм независимых одинаково распределенных слагаемых соответствуют характеристическим функциям вида либо $e^{it\beta} \mathbb{E} g^U(t)$, либо $q(t)g(t)$, где $P(U \geq 0) = 1$, $\beta \in \mathbb{R}$, g – безгранично делимая, а q – произвольная характеристические функции.

Особое положение ситуации (5.13) (или (5.14)) заключается в следующем. Если случайная величина U вырождена, то из (5.13) вытекает возможность представления предельной случайной величины Z в виде $Z = Y_\gamma + V$, где Y_γ – случайная величина с характеристической функцией h^γ , независимая от V . Следовательно, если на массив $\{X_{n,j}\}$ или последовательность $\{N_n\}$ не накладываются никакие другие условия, то предельной для случайных величин $S_{n,N_n} - c_n$ может быть *любая* функция распределения, поскольку вырожденная случайная величина $Y_\gamma = 0$ безгранично делима, и любую функцию распределения можно представить в виде слабого предела решетчатых распределений в (5.10).

5.4 Случайные суммы случайных индикаторов. Аналог теоремы Пуассона

В этом разделе мы докажем аналог теоремы Пуассона для случайных сумм случайных индикаторов.

Рассмотрим семейство последовательностей случайных величин $\{X_{p,j}, j \geq 1, 0 < p < 1\}$, определенных на одном вероятностном пространстве, такое, что при каждом фиксированном p случайные величины $X_{p,1}, X_{p,2}, \dots$ имеют одно и то же распределение Бернулли

$$X_{p,j} = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1 - p. \end{cases}$$

Пусть $\{N_p, 0 < p < 1\}$ – семейство положительных целочисленных случайных величин. Предположим, что при каждом фиксированном p случайные величины $N_p, X_{p,1}, X_{p,2}, \dots$ независимы. Положим

$$S_p = \sum_{j=1}^{N_p} X_{p,j}.$$

Рассмотрим асимптотическое поведение распределения случайной величины S_p , когда N_p неограниченно возрастает при $p \rightarrow 0$.

Подобные проблемы типичны для изучения так называемых редящих потоков событий. Они могут возникать в страховании. Действительно, пусть N_p – объем страхового портфеля, то есть число договоров страхования, заключенных в течение некоторого периода времени, скажем, в течение месяца. Если p – вероятность неблагоприятного события, от которого страхуется клиент, так что $X_{p,j}$ – индикатор такого события, то S_p равно количеству страховых выплат в рамках данного портфеля.

Теорема 5.8. *Предположим, что существует случайная величина N такая, что при $p \rightarrow 0$*

$$pN_p \implies N. \quad (5.15)$$

Тогда при $p \rightarrow 0$

$$S_p \implies S, \quad (5.16)$$

где S – дискретная случайная величина с распределением

$$P(S = k) = \frac{1}{k!} \int_0^{\infty} e^{-z} z^k dP(N < z), \quad k = 0, 1, \dots \quad (5.17)$$

Доказательство. Доказательство этой теоремы основано на Теореме 5.6.

Пусть $\{p_n\}_{n \geq 1}$ – произвольная последовательность такая, что $0 < p_n < 1$ и $p_n \rightarrow 0$ при $n \rightarrow \infty$. Положим $m_n = [1/p_n]$, где символом $[a]$ обозначается целая часть числа a . Тогда по теореме Пуассона при $n \rightarrow \infty$

$$\sum_{j=1}^{m_n} X_{p_n, j} \Longrightarrow Y, \quad (5.18)$$

где Y – случайная величина со стандартным пуассоновским распределением, которому соответствует характеристическая функция $h(t) = e^{e^{it} - 1}$, $t \in \mathbb{R}$. Несложно убедиться, что в силу (5.15) мы имеем

$$\frac{N_{p_n}}{m_n} = p_n N_{p_n} \cdot \frac{1}{p_n [1/p_n]} \Longrightarrow N \quad \text{при } n \rightarrow \infty \quad (5.19)$$

так как

$$\lim_{n \rightarrow \infty} p_n \left[\frac{1}{p_n} \right] = 1.$$

Поэтому по Теореме 5.6 из (5.18) и (5.19) мы получаем

$$S_{p_n} \Longrightarrow S \quad \text{при } n \rightarrow \infty,$$

где S – случайная величина с характеристической функцией

$$f(t) = \int_0^{\infty} e^{z(e^{it} - 1)} dP(N < z).$$

Но $f(t)$ представляет собой характеристическую функцию, соответствующую распределению (5.17), поэтому имеет место (5.16). Поскольку предельные распределения в (5.18) и (5.19) не зависят от выбора последовательности $\{p_n\}_{n \geq 1}$, лишь бы $0 < p_n < 1$ и $p_n \rightarrow 0$ при $n \rightarrow \infty$, теорема доказана.

Распределения вида (5.17) называются *смешанными пуассоновскими распределениями*. Теорема 5.8 может считаться обоснованием использования смешанных пуассоновских законов в качестве математических моделей распределений редких событий в неоднородной ситуации. Она является естественным обобщением классической теоремы Пуассона, иначе называемой законом малых чисел, которая часто используется для обоснования использования распределения Пуассона. Как мы увидим ниже, одномерные распределения процессов Кокса (дважды стохастических пуассоновских процессов) являются смешанными пуассоновскими.

Класс смешанных пуассоновских распределений довольно широк. В частности, согласно Теореме 5.8, если N имеет гамма-распределение (например, это так в случае, когда N_p имеет отрицательное биномиальное распределение), тогда S имеет отрицательное биномиальное распределение.

Глава 6

Вероятностные модели типа случайных блужданий

Рассмотрим принципы формирования смешанных вероятностных моделей на примере смешанных нормальных (гауссовских) моделей. Попутно мы коснемся и смешанных пуассоновских моделей.

Такой выбор обусловлен несколькими причинами.

Во-первых, классы смешанных гауссовских или смешанных пуассоновских моделей очень широки.

Во-вторых, подобные модели весьма адекватно описывают многие реально наблюдаемые закономерности. Примечательно, что для некоторых из них сначала статистически устанавливалось высокое согласие с теми или иными данными, и лишь спустя много лет доказывалась возможность их представления в виде смешанной гауссовской (или смешанной пуассоновской) модели. Пример – обобщенные гиперболические распределения, которым ниже будет посвящен один из разделов.

В-третьих, такие модели имеют серьезное теоретическое обоснование в виде специальных предельных теорем.

6.1 Случайные блуждания с дискретным временем

Во многих областях – от финансовой математики до физики – повсеместно наблюдается отклонение эмпирических закономерностей от ожидаемых в соответствии с классической теорией нормальных моделей.

Эмпирические распределения (плотности) практически всегда имеют более острую вершину, нежели нормальное распределение, и, соответственно, более тяжелые хвосты.

Чтобы объяснить отклонения наблюдаемых распределений от нормального (гауссовского), сначала подробнее обсудим математические обоснования самих гауссовских моделей. В основе гауссовских вероятностных моделей многих физических процессов лежит понятие однородного случайного блуждания. Суть этого понятия такова. Рассмотрим частицу, которая в *равноотстоящие* моменты времени испытывает мгновенные перемещения на случайные расстояния. Формально координата частицы после $n \geq 1$ перемещений может быть записана в виде

$$S_n = X_1 + \dots + X_n, \quad (6.1)$$

где X_i – i -е перемещение частицы. В предположении о том, что $X_1, X_2, \dots$ – независимые случайные величины с конечными дисперсиями, удовлетворяющие *условию Линдеберга*, суть которого, как уже говорилось, в том, что все слагаемые в суммах (6.1) в определенном смысле равноправны (условие Линдеберга, например, запрещает ситуацию, когда при $n \rightarrow \infty$ итоговое перемещение является результатом лишь конечного числа скачков), из центральной предельной теоремы теории вероятностей вытекает, что при $n \gg 1$ случайная величина S_n имеет приближенно нормальное (гауссово) распределение.

Доказаны так называемые *функциональные предельные теоремы*, устанавливающие сходимость при $n \rightarrow \infty$ (и одновременном соответствующем изменении масштаба) траекторий случайного блуждания (6.1) к траекториям винеровского процесса, являющегося математической моделью процесса броуновского движения (см., например, (Прохоров, 1956), (Леви, 1972), (Биллингсли, 1977)).

Винеровский процесс является гауссовским, то есть все его приращения имеют нормальное распределение. Винеровские процессы могут считаться в определенном смысле асимптотическими моделями и тесно связаны с центральной предельной теоремой, применение которой к анализу распределений приращений

реальных процессов с непрерывным временем сводится к следующим рассуждениям. Деля интервал времени $[t, t + \tau]$ ($t \geq 0$, $\tau > 0$) на некоторое число подынтервалов равной длины, мы замечаем, что приращение $P(t + \tau) - P(t)$ рассматриваемого физического процесса может быть представлено в виде суммы приращений процесса $P(t)$ на подынтервалах. В предположении, что приращения на подынтервалах независимы, в соответствии с центральной предельной теоремой распределение приращения $P(t + \tau) - P(t)$ приближается к нормальному при увеличении числа подынтервалов, если выполнено условие Линдеберга. Как уже говорилось, это условие означает, что никакая конечная группа слагаемых не играет доминирующей роли в сумме. Так как подынтервалы имеют одинаковую длину, то естественно считать, что распределения приращений на подынтервалах одинаковы. В этом случае условие Линдеберга сводится к требованию конечности дисперсий элементарных приращений.

Однако, как уже было отмечено выше, распределения приращений регистрируемых процессов заметно отличны от нормальных. Поэтому винеровские процессы оказываются отнюдь не бесспорными математическими моделями рассматриваемых процессов. Здесь следует также отметить, что не-нормальность распределений приращений регистрируемых процессов также делает весьма проблематичной адекватность таких модных ныне вероятностных моделей как фрактальные броуновские движения, которые, хотя и учитывают возможную зависимость приращений процесса (обладают “памятью”), но остаются гауссовыми – все их конечномерные распределения нормальны.

Как уже было сказано, отмеченная не-нормальность распределений приращений проявляется в том, что в действительности наблюдается заметно больше очень больших и очень маленьких по абсолютной величине значений приращений, нежели их должно быть в соответствии с нормальным распределением. Другими словами, наблюдаемые распределения приращений реальных процессов на небольших интервалах времени являются более островершинными, нежели нормальные, имея заметно более тяжелые хвосты.

Важная попытка объяснить замеченную островершинность этих распределений была предпринята в 60-х годах XX столетия. Обычно эту попытку ассоциируют с исследованиями Бенуа Мандельброта (Mandelbrot, 1963), (Mandelbrot, 1967) и Юджина Фамы (Fama, 1963), (Fama, 1965) в области моделирования процессов эволюции биржевых цен. В частности, Мандельброт в статье (Mandelbrot, 1963) обратил внимание на то, что “хвосты распределений приращений биржевых цен столь исключительно тяжелы, что [эмпирические] вторые моменты обычно изменяются беспорядочным образом”. Как известно, беспорядочное изменение выборочных моментов при изменении объема выборки свидетельствует о том, что соответствующий теоретический момент отсутствует (в отличие от сближения эмпирических моментов с некоторым числом в случае существования теоретических моментов).

Основная идея в рассуждениях Б. Мандельброта может быть кратко изложена так: отклонение распределений приращений от нормального означает, что классическая центральная предельная теорема здесь не применима, так как нарушаются ее условия. Если не выходить за рамки общей структурной модели, согласно которой итоговое приращение процесса $P(t)$ на каждом интервале является суммой независимых приращений этого же процесса на подынтервалах одинаковой длины, и предполагать, что элементарные приращения имеют одинаковые распределения, то единственным условием, которое должно нарушаться, остается требование конечности дисперсии элементарных приращений. Поэтому вместо классической центральной предельной теоремы в качестве основы для построения моделей распределений приращений биржевых цен Б. Мандельброт предложил использовать предельные теоремы, описывающие сходимость сумм независимых слагаемых с бесконечными дисперсиями.

Как уже обсуждалось, если слагаемые одинаково распределены, то в таких теоремах в качестве предельных распределений выступают так называемые устойчивые законы. Напомним, что функция распределения $G(x)$ называется *устойчивой*, если для любых $a_1 > 0$, $a_2 > 0$, $b_1 \in \mathbb{R}$ и $b_2 \in \mathbb{R}$ существуют числа $a > 0$ и $b \in \mathbb{R}$

такие, что

$$(G_1 * G_2)(x) \equiv G(ax + b),$$

где $G_1(x) = G(a_1x + b_1)$, $G_2(x) = G(a_2x + b_2)$. Характеристическая функция любого устойчивого распределения допускает представление в виде

$$g(t) = \exp \left\{ iat - c|t|^\alpha \left(1 + ibQ(t, \alpha) \operatorname{sign} t \right) \right\},$$

где

$$Q(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \text{если } \alpha \neq 1, \\ \frac{2}{\pi} \log |t|, & \text{если } \alpha = 1. \end{cases}$$

где $a \in \mathbb{R}$, $c \geq 0$, $-1 \leq b \leq 1$, $0 < \alpha \leq 2$ – параметры, определяющие тип распределения. Параметр α , определяющий скорость убывания хвостов устойчивого закона, называется *характеристическим*.

Устойчивые законы являются хорошо известными примерами распределений с тяжелыми хвостами. Напомним, что дисперсии всех устойчивых законов за исключением нормального не определены (“бесконечны”).

Предположения Мандельброта были подтверждены Ю. Фамой, который в статье (Fama, 1965) привел результаты статистического анализа цен акций больших компаний. Согласно этой статье, распределения приращений цен акций хорошо согласуются с устойчивой моделью, в которой характеристический показатель хоть и близок к двум, но все же не равен двум.

Исследования Б. Мандельброта и Ю. Фамы вызвали взрыв интереса как к аналитическим свойствам устойчивых законов (см., например, монографии (Золотарев, 1983), (Samorodnitsky and Taqqu, 1994), (Uchaikin and Zolotarev, 1999)) так и к статистическим процедурам, связанным с устойчивыми семействами.

В функциональных предельных теоремах для однородных случайных блужданий с дискретным временем и скачками с бесконечной дисперсией в качестве предельных выступают так называемые устойчивые процессы Леви (процессы с независимыми и стационарными приращениями, имеющими устойчивые распределения,

см., например, (Samorodnitsky and Taqqu, 1994)). Подобные модели обладают весьма привлекательным аналитическим свойством масштабной инвариантности, которое, по мнению некоторых исследователей, должно быть присуще описываемым этими моделями сложным системам (Mandelbrot, 1997).

Традиционная аргументация в пользу применения устойчивых законов в качестве моделей распределений (логарифмических) приращений тех или иных процессов, рассматриваемых в эконометрике или финансовой математике, заключается в следующем. Погрешность любой детерминированной эконометрической модели обусловлена суммарным воздействием тех факторов, которые не учитываются моделью. А поскольку устойчивые законы (и только они) являются предельным для распределений сумм независимых одинаково распределенных величин, случайные составляющие модели (сосредоточенные в членах, соответствующих погрешностям) должны иметь устойчивое распределение. Обзоры применения устойчивых распределений в качестве моделей процессов, возникающих в финансовых и других эконометрических приложениях, можно найти, например, в работах (McCulloch, 1996) и (Rachev, 2003).

Однако наряду с большим числом благоприятных аналитических качеств, устойчивые законы обладают одним неприятным свойством: за исключением четырех случаев (нормальное распределение, распределение Коши, распределение Леви (распределение времени первого достижения процессом броуновского движения единичного уровня) и распределение, симметричное к распределению Леви), отсутствуют явные выражения устойчивых плотностей (которые всегда существуют) в терминах элементарных функций. Немного больше известно устойчивых плотностей, выражающихся через наиболее популярные специальные функции (лишь относительно недавно доказано, что все устойчивые плотности допускают представление в терминах так называемых функций Фокса или обобщенных J -функций Мейера, см. (Schneider, 1986), (Uchaikin and Zolotarev, 1999)). Это обстоятельство затрудняет статистический анализ в рамках устойчивых моделей.

Невозможность применять классические статистические процедуры, основанные на использовании функции правдоподобия, не является единственным обстоятельством, заставляющим с некоторым сомнением относиться к целесообразности использования устойчивых законов для моделирования распределений приращений биржевых цен. Напомним, что основным предположением, обуславливающим их применение в подобных задачах, является отсутствие конечных дисперсий у приращений биржевых цен за интервалы времени (бесконечно) малой длины. Но для отсутствия дисперсии необходимо, чтобы упомянутые элементарные приращения могли с положительной вероятностью принимать сколь угодно большие по абсолютной величине значения, что на практике невозможно, так как регистрируемые процессы ограничены в пространстве. К тому же, как было отмечено, например, в работах (Blattberg and Gonedes, 1974), (Akgiray and Booth, 1987), (Jansen and de Vries, 1991), (Lux, 1996), многие статистические тесты отвергают гипотезу о согласии распределений реальных данных с устойчивыми законами, поскольку, все же, слишком больших по абсолютной величине наблюдений в реальных данных меньше, чем должно быть при не-нормальных устойчивых распределениях с бесконечной дисперсией, хвосты которых спадают при $|x| \rightarrow \infty$ как $O(|x|^{-\alpha})$, $0 < \alpha < 2$. В частности, как показано в работах (Jansen and de Vries, 1991), (Lux, 1996), при статистической подгонке к финансовым данным распределений с хвостами, убывающими степенным образом, оцененное значение показателя α оказывается существенно больше двух.

Еще одним веским аргументом против устойчивых моделей является следующее обстоятельство. Все устойчивые модели обладают свойством фрактальности (самоподобия) в том смысле, что согласно таким моделям распределения приращений биржевых цен на интервалах времени разной длины должны принадлежать к одному и тому же устойчивому типу с одним и тем же характеристическим показателем и отличаться лишь параметром масштаба. В то же время еще в работе (Officer, 1972) отмечен эффект “унормализации” реальных распределений приращений биржевых цен при увеличении длины промежутка времени, приращения за ко-

торый рассматриваются: в таких условиях “тяжесть” хвостов этих распределений уменьшается, а сами эти распределения становятся все более и более похожими на нормальное. Другими словами, в реальной практике распределения приращений не обладают свойством фрактальности (самоподобия) в указанном смысле.

6.2 Случайные блуждания с непрерывным временем

В поисках удовлетворительной элементарной стохастической модели, обобщающей (6.1), которая могла бы объяснить наблюдаемую не-нормальность распределений приращений регистрируемых хаотических процессов, многие исследователи уделяют внимание однородным случайным блужданиям с непрерывным временем, в которых случайными являются не только перемещения частицы на очередном шаге, но и интервалы времени между последовательными (мгновенными) перемещениями (см., например, (Zaslavsky, 2002)).

Модель однородных случайных блужданий с непрерывным временем была введена Монтроллом и Вейссом в 1965 г. (Montroll and Weiss, 1965). С тех пор эта модель интенсивно изучалась и применялась, например, для описания турбулентных потоков (Shlesinger, West and Klafter, 1987), (Klafter, Blumen and Shlesinger, 1987), (Pedrosa et al., 1996), (Connor et al., 1999) (Batanov et al., 1998), переноса в неоднородной фрактальной среде (Scher and Lax, 1973), (Weiss and Rubin, 1983), (Bouchaud and Georges, 1990), (Metzler, Barkai and Klafter, 1999), (West and Deering, 1994), (Isichenko, 1992), динамики неустойчивых хаотических систем (Klafter, Zumofen and Shlesinger, 1993) недебаевской релаксации (Bendler and Shlesinger, 1985), (Shlesinger, 1984), проблем, связанных с потоками в пористых средах (Benson, Meerschaert, Schumer and Wheatcraft, 2001a), (Benson, Meerschaert, Schumer and Wheatcraft, 2001b), (Benson, Meerschaert and Wheatcraft, 2000), (Baeumer, Benson, Meerschaert and Wheatcraft 2001).

В работе (Kotulski, 1995) в модели однородных случайных

блужданий с непрерывным временем впервые возникли новые негауссовские предельные (при $t \rightarrow \infty$) распределения. Специфическим свойством этих распределений является наличие тяжелых хвостов. Позднее (и независимо) эти же распределения возникли в работах (Saichev and Zaslavsky, 1997), (Korolev and Uchaikin, 1999) и (Королев и Учайкин, 2000).

Однородное случайное блуждание с непрерывным временем, которое также можно назвать *обобщенным (сotrround) процессом восстановления*, формально определяется следующим образом. Пусть $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины и пусть далее $T_1, T_2, \dots$ – неотрицательные одинаково распределенные случайные величины, независимые друг от друга и от последовательности $X_1, X_2, \dots$. Рассмотрим случайный процесс

$$S(t) = \sum_{i=1}^{N(t)} X_i, \quad t > 0,$$

где $N(t)$ – считающий процесс, определяемый соотношением

$$\sum_{i=1}^{N(t)} T_i \leq t < \sum_{i=1}^{N(t)+1} T_i, \quad t > 0,$$

(для определенности предполагается, что $\sum_{i=1}^0 = 0$). С физической точки зрения X_i означает величину i -го мгновенного скачка частицы, T_i имеет смысл случайного времени ожидания между $(i-1)$ -м и i -м скачками, а $S(t)$ имеет смысл координаты блуждающей частицы в момент времени t .

Наряду с условием $ET_1 < \infty$ мы также рассмотрим случай, когда распределение случайной величины T_i имеет тяжелый хвост (исходя из сказанного выше, мы всюду ниже будем предполагать, что $DX_1 = \sigma^2 < \infty$). Под “тяжелохвостостью” распределения времени ожидания T_1 будем понимать, что $P(T_1 > t) = O(t^{-\beta})$ при некотором $\beta \in (0, 1)$. Более формально, предположим, что распределение случайной величины T_1 принадлежит области нормального притяжения строго устойчивого закона $G(x; \beta)$, определяемого

характеристической функцией

$$\tilde{g}(s; \beta) = \exp \left\{ -|s|^\beta \exp \left(-i \frac{\pi \beta}{2} \operatorname{sign} s \right) \right\}, \quad 0 < \beta < 1, s \in \mathbb{R}.$$

Этот факт будет обозначаться как $\mathcal{L}(T_1) \in DNA(G(x; \beta))$.

Сходимость одномерных распределений процесса $S(t)$ изучалась во многих работах. В частности, в работах (Kotulski, 1995), (Korolev and Uchaikin, 1999), (Королев и Учайкин, 2000), (Kolokoltsov, Korolev and Uchaikin, 2001) были сформулированы и доказаны различные версии результата, частным случаем которого является следующее утверждение.

Теорема 6.1. *Предположим, что распределение случайной величины X_1 удовлетворяет условиям $EX_1 = 0$ и $DX_1 = \sigma^2 < \infty$. Тогда:*

1°. *если $ET_1 < \infty$, то существует такая постоянная $c > 0$, что при $t \rightarrow \infty$*

$$P\left(\frac{S(t)}{c\sqrt{t}} < x\right) \Rightarrow \Phi(x);$$

2°. *если $\mathcal{L}(T_1) \in DNA(G(x; \beta))$, $0 < \beta < 1$, то существует такая постоянная $c > 0$, что при $t \rightarrow \infty$*

$$P\left(\frac{S(t)}{ct^{\beta/2}} < x\right) \Rightarrow \int_0^\infty \Phi(xu^{\beta/2}) dG(u; \beta) \equiv Q(x; \beta). \quad (6.2)$$

При этом для плотности $q(x; \beta) = Q'(x; \beta)$ справедливо представление

$$q(x; \beta) = \frac{g(|x|^{-2/\beta}; \beta/2)}{\beta|x|^{1+2/\beta}}, \quad x \in \mathbb{R},$$

где $g(y; \beta/2)$, $y \geq 0$, – плотность одностороннего устойчивого распределения с характеристическим показателем $\beta/2$. Распределение $Q(x; \beta)$ обладает моментами всех порядков. Более того, для $n = 1, 2, \dots$ справедливы соотношения

$$\int_{-\infty}^{\infty} x^{2n} q(x; \beta) dx = \frac{4n! \Gamma(n + \frac{1}{2})}{\sqrt{\pi} \Gamma(n\beta + 1)}, \quad \int_{-\infty}^{\infty} x^{2n-1} q(x; \beta) dx = 0,$$

где $\Gamma(\cdot)$ – эйлерова гамма-функция.

Распределение $Q(x; \beta)$, стоящее в правой части (6.2), является представителем семейства так называемых *дробно-устойчивых* законов. Разнообразные аналитические свойства дробно-устойчивых законов (в том числе и распределения $Q(x; \beta)$) описаны в статье (Бенинг, Королев и др., 2003). Обратим особое внимание на то, что распределение $Q(x; \beta)$ является масштабной смесью нормальных законов с нулевым средним (см. ниже). Это обстоятельство в дальнейшем будет играть важнейшую роль в контексте обсуждаемого подхода.

Однако статистический анализ экспериментальных данных показывает, что далеко не всегда наблюдаемые распределения приращений регистрируемых процессов удовлетворительно согласуются с моделью, описываемой правой частью (6.2). К тому же негауссовы распределения возникают в модели однородного случайного блуждания с непрерывным временем, только если среднее время ожидания очередного скачка бесконечно, что возможно, только если время ожидания очередного скачка может с положительной вероятностью быть произвольно большим, что плохо согласуется с традиционными представлениями, так как регистрируемые процессы ограничены во времени.

Таким образом, учитывая сказанное в этом и предыдущем разделах, можно прийти к выводу, что *однородные случайные блуждания с дискретным или непрерывным временем, вообще говоря, не могут быть адекватными элементарными моделями многих реальных процессов.*

6.3 Обобщенные процессы Кокса

Естественным обобщением рассмотренной выше модели является неоднородное случайное блуждание с непрерывным временем. От однородного блуждания оно отличается тем, что распределения случайных величин $T_1, T_2, \dots$ являются, вообще говоря, различными. Предположение неоднородности (различия распределений времен между последовательными скачками) случайного блуждания хорошо согласуется с представлением о том, что интенсив-

ность изменений координаты частицы, испытывающей броуновское движение в турбулентной среде, существенно непостоянна. Это непостоянство может быть вызвано многими причинами, проявляющимися, например, как непериодические или периодические компоненты (тренды), связанные с изменениями локальных (во времени) тенденций, например, панического характера на биржах. Кроме того, участки нестационарности могут быть вызваны некоторыми случайно возникающими (не поддающимися абсолютно надежному прогнозированию) причинами.

Пусть, как и выше, $N(t)$ – число скачков случайно блуждающей частицы за период времени $[0, t]$, $t \geq 0$. Моменты скачков образуют хаотический точечный случайный процесс на оси времени. Однако этот хаотический случайный процесс не является однородным в силу указанных выше представлений. Как известно, наиболее разумными стохастическими моделями неоднородных хаотических точечных процессов являются *дважды стохастические пуассоновские процессы*, иначе называемые *процессами Кокса* (см., например, (Bening and Korolev, 2002)). Определение процессов Кокса следует предварить кратким экскурсом в описание некоторых свойств однородных и неоднородных пуассоновских процессов. В полном объеме свойства однородных и неоднородных пуассоновских процессов описаны, например, в книгах (Kingman, 1993) и (Кингман, 2007), где также можно найти многочисленные примеры математических и прикладных задач, в которых возникают такие процессы.

Мы уже убедились, что пуассоновский процесс, характеризующий тем обстоятельством, что интервалы времени между событиями потока стохастически независимы и имеют одинаковое показательное распределение, является наилучшей моделью *однородного* хаотического потока событий. Напомним, что привлекательность пуассоновского процесса в качестве модели однородного дискретного хаоса обусловлена как минимум двумя обстоятельствами. Во-первых, показательное распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на всей положительной полуоси и имеющих конечное математическое ожида-

ние, а энтропия, как известно, является очень удобной численной характеристикой неопределенности. Во-вторых, точки (события) пуассоновского потока равномерно распределены на оси времени в том смысле, что для любого конечного интервала времени $[t_1, t_2]$, $t_1 < t_2$, условное совместное распределение точек пуассоновского потока, попавших в интервал $[t_1, t_2]$, при условии, что в этот интервал попало фиксированное число, скажем, n точек, совпадает с совместным распределением вариационного ряда, построенного по независимой однородной выборке объема n из равномерного на $[t_1, t_2]$ распределения. Равномерное же распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на конечных интервалах, и очень хорошо соответствует общепринятому представлению об абсолютно непредсказуемой ограниченной случайной величине.

Вспомним, что если $N_\lambda(t)$ – однородный пуассоновский процесс с некоторой интенсивностью $\lambda > 0$, то параметр λ имеет смысл среднего числа скачков процесса (количества событий наблюдаемого потока) в единицу времени. При этом

$$P(N_\lambda(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \dots \quad (6.3)$$

и

$$EN_\lambda(t) = DN_\lambda(t) = \lambda t, \quad t > 0,$$

так что для любых $s, t > 0$

$$\frac{EN_\lambda(t+s) - EN_\lambda(t)}{s} = \lambda = \text{const.}$$

Из соотношения (6.3) легко видеть, что для любого $k = 0, 1, \dots$

$$P(N_\lambda(t) = k) = P(N_1(\lambda t) = k), \quad t \geq 0, \quad (6.4)$$

то есть процессы $N_\lambda(t)$ и $N_1(\lambda t)$ *стохастически эквивалентны*.

Однако, как уже было отмечено, в реальной практике хаотические потоки не бывают однородными. Для наглядности, перед тем как определить дважды стохастический пуассоновский процесс,

рассмотрим неоднородный пуассоновский поток событий $N^*(t)$, определяемый следующим образом. Пусть $\lambda(t)$ – некоторая положительная функция. Предположим, что функция $\lambda(t)$ интегрируема, и обозначим

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0.$$

Неоднородный пуассоновский процесс $N^*(t)$, $t \geq 0$, определяется как случайный процесс с независимыми приращениями, траектории которого стартуют из нуля ($N^*(0) = 0$ почти наверное) и

$$P(N^*(t) = k) = e^{-\Lambda(t)} \frac{(\Lambda(t))^k}{k!}, \quad k = 0, 1, \dots \quad (6.5)$$

Тогда

$$EN^*(t) = \Lambda(t), \quad t \geq 0,$$

так что

$$\lim_{s \downarrow 0} \frac{EN^*(t+s) - EN^*(t)}{s} = \lim_{s \downarrow 0} \frac{\Lambda(t+s) - \Lambda(t)}{s} = \frac{d\Lambda(t)}{dt} = \lambda(t), \quad t \geq 0.$$

Другими словами, при таком определении функция $\lambda(t)$ играет роль *мгновенной интенсивности* процесса в точке t . При этом по аналогии с (6.4) из (6.5) вытекает, что

$$P(N^*(t) = k) = P(N_1(\Lambda(t)) = k), \quad k = 0, 1, \dots, \quad t \geq 0,$$

то есть процессы $N^*(t)$ и $N_1(\Lambda(t))$ стохастически эквивалентны.

Наконец, рассмотрим предложенное Д. Коксом (Сох, 1955) понятие *дважды стохастического пуассоновского процесса*, являющееся естественным обобщением неоднородного пуассоновского процесса.

Пусть $N_1(t)$, $t \geq 0$, – однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t)$, $t \geq 0$, – независимый от $N_1(t)$ случайный процесс, обладающий следующими свойствами: $\Lambda(0) = 0$ почти наверное, $P(\Lambda(t) < \infty) = 1$ для любого $t > 0$, траектории $\Lambda(t)$

не убывают и непрерывны справа. Дважды стохастический пуассоновский процесс $N(t)$, называемый теперь также процессом Кокса, определяется как суперпозиция $N_1(t)$ и $\Lambda(t)$:

$$N(t) = N_1(\Lambda(t)), \quad t \geq 0.$$

В этом случае будем говорить, что процесс Кокса $N(t)$ управляется процессом $\Lambda(t)$. В частности, если процесс $\Lambda(t)$ допускает представление

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0,$$

в котором $\lambda(t)$ – положительный случайный процесс с интегрируемыми траекториями, то $\lambda(t)$ можно интерпретировать как мгновенную стохастическую интенсивность процесса $N(t)$. Поэтому иногда процесс $\Lambda(t)$, управляющий процессом Кокса $N(t)$, будет называться *накопленной интенсивностью* процесса $N(t)$.

Несложно убедиться, что для процесса Кокса $N(t)$, управляемого процессом $\Lambda(t)$, справедливы соотношения

$$EN(t) = E\Lambda(t), \quad DN(t) = E\Lambda(t) + D\Lambda(t).$$

При каждом фиксированном t распределение случайной величины $N(t) = N_1(\Lambda(t))$ является смешанным пуассоновским. Класс смешанных пуассоновских распределений весьма широк и содержит, в частности, само пуассоновское распределение (ему соответствует вырожденное смешивающее распределение), отрицательное биномиальное распределение (ему соответствует смешивающее гамма-распределение), распределение Зихеля, которому соответствует обобщенное обратное нормальное смешивающее распределение, бета-пуассоновское распределение при смешивающем бета-распределении, распределение Делaporte со сдвинутым гамма-распределением в качестве смешивающего, обобщенное распределение Варинга, инфекционное распределение Неймана типа А (иначе называемое пуассон-пуассоновским). Эти и другие примеры смешанных пуассоновских распределений подробно

рассмотрены в книгах (Grandell, 1997), (Королев, Бенинг и Шоргин, 2011). В этих книгах также можно найти детальное описание аналитических и асимптотических свойств процессов Кокса.

В основе конструкции базовых математических моделей неоднородных хаотических процессов, рассматриваемых в этой книге, лежит асимптотическая схема, основанная на принципе, который может быть наглядно проиллюстрирован на примере простейшей задачи из теории измерений. Погрешность измерения является результатом суммарного воздействия большого числа случайных факторов, ни один из которых не является доминирующим, и потому, согласно центральной предельной теореме, должна иметь нормальное распределение. Однако на *разные* измерения воздействует, вообще говоря, *разное* число случайных факторов, то есть число случайных факторов, определяющих погрешность, само является случайным фактором. Поэтому вместо классической центральной предельной теоремы здесь более уместно пользоваться предельными теоремами для сумм случайного числа независимых случайных величин.

Теория случайного суммирования довольно хорошо развита (см., например, монографии (Круглов и Королев, 1990), (Gnedenko and Korolev, 1996)). Некоторые ее результаты обсуждались в предыдущей главе. Не ставя перед собой цели ознакомить читателя с этой теорией во всей полноте, сосредоточимся лишь на очень частном конкретном варианте постановки задачи, когда число слагаемых в суммах формируется в соответствии с дважды стохастическим пуассоновским процессом (процессом Кокса). Этот случай имеет чрезвычайно важное практическое значение.

Наша цель – описание задачи моделирования неоднородных хаотических потоков событий с помощью обобщенных процессов Кокса. Мы увидим, что отклонения распределений наблюдаемых процессов от нормального могут быть успешно объяснены наличием существенной изменчивости интенсивности неоднородных хаотических потоков событий, описываемых процессами Кокса.

Пусть $X_1, X_2, \dots$ – одинаково распределенные случайные величины. Предположим, что при каждом $t \geq 0$ случайные величины

$N(t), X_1, X_2, \dots$ независимы. Процесс

$$S(t) = \sum_{j=1}^{N(t)} X_j, \quad t \geq 0, \quad (6.6)$$

назовем обобщенным процессом Кокса (при этом для определенности считаем, что $\sum_{j=1}^0 = 0$). Процессы вида (6.6) играют чрезвычайно важную роль во многих прикладных задачах. Достаточно сказать, что при $\Lambda(t) \equiv \lambda t$ с $\lambda > 0$ процесс $S(t)$ превращается в классический обобщенный пуассоновский процесс, широко используемый при моделировании многих явлений в физике, теории надежности, финансовой и актуарной деятельности, биологии и т. д. Большое число разнообразных прикладных задач, приводящих к обобщенным пуассоновским процессам, описано в книгах (Gnedenko and Korolev, 1996) и (Bening and Korolev, 2002).

Приведем примеры некоторых задач, сводящихся к обобщенным процессам Кокса.

Пример 6.1. Как уже говорилось, пуассоновский точечный процесс описывает абсолютно хаотическое распределение точек на прямой. Рассмотрим следующее вполне естественное обобщение этого свойства применительно к задаче о суммарной массе тел. Пусть $\rho(x)$, $x \in \mathbb{R}^3$, – плотность (концентрация) каких-либо материальных частиц в пространстве. Предположим, что частицы – это материальные точки, распределенные в пространстве в соответствии с неоднородным трехмерным пуассоновским точечным процессом, так что если $N(B)$ – число точек в некотором (измеримом) множестве $B \subset \mathbb{R}^3$, то

$$P(N(B) = k) = \exp\{-\lambda(B)\} \frac{(\lambda(B))^k}{k!}, \quad k = 0, 1, 2, \dots,$$

где $\lambda(B) = \int_B \rho(x) dx$. Положим $\Lambda(t) = \int_{|x| \leq t} \rho(x) dx$. Пусть X_j – масса, сосредоточенная в j -ой точке. Тогда процесс $S(t)$, определяемый соотношением (6.6), описывает суммарную массу, содержащуюся в сфере радиуса t . Подобные объекты представляют традиционный интерес в космологии.

Пример 6.2. Рассмотрим следующую модель одномерного броуновского движения – теплового движения некоторой частицы, испытывающей соударения с молекулами вещества, заполняющего среду, в которой движется частица. Пусть $N(t)$ – число соударений частицы с молекулами до момента t . Если бы среда и свойства частицы не изменялись во времени, то было бы естественно предполагать, что $N(t)$ является однородным пуассоновским процессом. Однако если допустить, что среда или свойства частицы изменяются во времени, то интенсивность столкновений перестает быть постоянной. Поэтому, обозначив перемещение частицы в результате j -го столкновения через X_j , можно прийти к выводу, что процесс $S(t)$, определяемый соотношением (6.6), описывает координату частицы в момент t , если она испытывает броуновское движение в неоднородной (случайной) среде.

Всюду в дальнейшем будем считать, что у случайных величин $\{X_j\}_{j \geq 1}$ имеется, по крайней мере, два первых момента. Обозначим $EX_1 = a$, $DX_1 = \sigma^2$, $0 < \sigma^2 < \infty$. Как будет показано, даже в таких предположениях предельные распределения обобщенных процессов Кокса могут иметь произвольно тяжелые хвосты. Для большей наглядности изложение в основном будет сосредоточено на ситуации, в которой $a = 0$.

Несложно убедиться, что если $E\Lambda(t) < \infty$, то

$$ES(t) = aE\Lambda(t)$$

и, если дополнительно $E(\Lambda(t))^2 < \infty$, то

$$DS(t) = (a^2 + \sigma^2)E\Lambda(t) + a^2 D\Lambda(t).$$

6.4 Центральная предельная теорема для обобщенных процессов Кокса

Мы приведем необходимые и достаточные условия сходимости одномерных распределений обобщенных процессов Кокса со скачками, обладающими указанными выше свойствами, без каких бы то

ни было моментных условий на управляющий процесс. Мы продемонстрируем, что асимптотическое поведение процесса $S(t)$ полностью определяется асимптотическим поведением накопленной интенсивности $\Lambda(t)$. Более того, тяжелые (например, типа Парето) хвосты распределений, предельных для сумм (6.6), могут быть обусловлены не “плохим” поведением слагаемых (например, отсутствием у них моментов), а чрезмерно большим разбросом значений управляющего процесса $\Lambda(t)$.

Пусть $d(t) > 0$ – функция, неограниченно возрастающая при $t \rightarrow \infty$. Роль функции $d(t)$ – нормировка или масштабирование во времени (или в пространстве, если “бесконечно большой” параметр t не является астрономическим временем).

Теорема 6.2. *Предположим, что $\Lambda(t) \rightarrow \infty$ по вероятности при $t \rightarrow \infty$. Для того чтобы одномерные распределения нормированного обобщенного процесса Кокса слабо сходились к распределению некоторой случайной величины Z ,*

$$\frac{S(t)}{\sigma \sqrt{d(t)}} \Rightarrow Z \quad (t \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала неотрицательная случайная величина U такая, что

$$\begin{aligned} 1^\circ \quad & \mathbb{P}(Z < x) = \int_0^\infty \Phi(xy^{-1/2}) d\mathbb{P}(U < y) \quad (= \mathbb{E}\Phi(xU^{-1/2})), \quad x \in \mathbb{R}; \\ 2^\circ \quad & \frac{\Lambda(t)}{d(t)} \Rightarrow U \quad (t \rightarrow \infty). \end{aligned}$$

Доказательство этой теоремы основано на теореме переноса и мы его опустим.

Обратим внимание на то, что условие 2° Теоремы 6.2 может быть интерпретировано как требование статистической регулярности накопленной интенсивности: предел отношения $\Lambda(t)/d(t)$ при $t \rightarrow \infty$ может быть случайным, но он должен существовать. Еще одна интерпретация этого условия заключается в том, что при больших t распределение случайной величины $\Lambda(t)/d(t)$ практически не зависит от t .

При этом условие 1° означает, что, вообще говоря, предельным распределением обобщенного процесса Кокса является масштабная смесь нормальных законов, имеющая, как будет видно ниже, (вообще говоря, произвольно) более тяжелые хвосты, нежели само нормальное распределение. Идея использования масштабных смесей нормальных законов в качестве островершинных (и обладающих тяжелыми хвостами) альтернатив нормальному закону в задачах обработки измерений отнюдь не является новой. В частности, такую идею высказал еще в 1882 г. американский астроном С. Ньюкомб (Newcomb, 1882), см. также (Stigler, 1973). Обсуждение возможности применения таких смесей к описанию распределения скоростей в турбулентной среде можно найти в работе (Dutton, 1970), где в качестве физического обоснования этого подхода используется неоднородный, перемежающийся характер турбулентности. Другие ссылки на работы из разных областей, в которых обсуждается применимость масштабных смесей нормальных законов при обработке реальных данных, можно найти, например, в статье (Barndorff-Nielsen, 1977).

Из Теоремы 6.2 и свойства идентифицируемости (см. ниже) семейства масштабных смесей нормальных законов немедленно вытекает следующее утверждение.

Следствие 6.1. *В условиях Теоремы 6.2*

$$P\left(\frac{S(t)}{\sigma\sqrt{d(t)}} < x\right) \Rightarrow \Phi(x) \quad (t \rightarrow \infty),$$

тогда и только тогда, когда

$$\frac{\Lambda(t)}{d(t)} \Rightarrow 1 \quad (t \rightarrow \infty).$$

Другими словами, предельное распределение обобщенного процесса Кокса может быть нормальным только в том случае, когда случайная величина $\Lambda(t)/d(t)$ асимптотически (при $t \rightarrow \infty$) неслучайна.

В качестве еще одного следствия Теоремы 6.2 приведем критерий сходимости одномерных распределений обобщенных процессов Кокса с нулевым средним и *конечными дисперсиями* к

устойчивым законам. Будет показано, что одномерные распределения обобщенных процессов Кокса с описанными выше свойствами асимптотически строго устойчивы тогда и только тогда, когда асимптотически строго устойчивы их управляющие процессы.

Пусть $G_{\alpha,\theta}(x)$ – строго устойчивая функция распределения с показателем α и параметром θ , которой соответствует характеристическая функция

$$\tilde{g}_{\alpha,\theta}(s) = \exp \left\{ -|s|^\alpha \exp \left\{ -i \frac{\pi\theta\alpha}{2} \text{sign}s \right\} \right\}, \quad s \in \mathbb{R},$$

где $0 < \alpha \leq 2$, $|\theta| \leq \theta_\alpha = \min(1, 2/\alpha - 1)$.

Теорема 6.3. *Предположим, что $\Lambda(t) \rightarrow \infty$ по вероятности при $t \rightarrow \infty$. Для того чтобы*

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{S(t)}{\sigma \sqrt{d(t)}} < x \right) = G_{\alpha,0}(x), \quad x \in \mathbb{R},$$

необходимо и достаточно, чтобы

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{\Lambda(t)}{d(t)} < x \right) = G_{\alpha/2,1}(x), \quad x \in \mathbb{R}.$$

Доказательство этого утверждения заключается в последовательном применении Теоремы 6.2 и теоремы 3.3.1 из (Золотарев, 1983), согласно которой

$$G_{\alpha,0}(x) = \int_0^\infty \Phi(x/\sqrt{u}) dG_{\alpha/2,1}(u), \quad x \in \mathbb{R},$$

с использованием свойства идентифицируемости масштабных смесей нормальных законов (см. ниже) и абсолютной непрерывности устойчивых распределений.

Пример 6.3. Для иллюстрации того, насколько изменяется характер предельных законов для суммарных воздействий при стохастическом характере интенсивности потока отдельных воздействий, рассмотрим ситуацию, в которой случайная величина U ,

предельная в Теореме 6.2 для нормированного управляющего процесса, то есть нормированной накопленной интенсивности потока экстремальных событий, имеет стандартное показательное распределение. Это может быть реализовано, к примеру, если случайная величина $\Lambda(t)$ при каждом t имеет геометрическое распределение с параметром (вероятностью “успеха”) t^{-1} , или же если эта случайная величина при каждом t имеет показательное распределение с таким же параметром t^{-1} . В таком случае $P(U < x) = (1 - e^{-x})\mathbf{1}(x \geq 0)$, и, следовательно, можно показать, что

$$E\Phi\left(\frac{x}{\sqrt{U}}\right) = \begin{cases} \frac{1}{2}e^{\sqrt{2}x}, & x \leq 0, \\ 1 - \frac{1}{2}e^{-\sqrt{2}x}, & x > 0. \end{cases}$$

Это есть не что иное как двойное показательное распределение (распределение Лапласа) с параметром $\sqrt{2}$. Как несложно видеть, ему соответствует плотность

$$\ell(x) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x|\}, \quad x \in \mathbb{R}.$$

Хвост нормальной функции распределения $\Phi(x)$ убывает при $x \rightarrow \infty$ как $\sqrt{2/\pi}x^{-1}e^{-x^2/2}$, в то время как хвост функции распределения Лапласа, естественно, убывает существенно медленнее.

Это приводит к существенному различию в оценках вероятностей катастрофических воздействий и размеров самих катастрофических воздействий, получаемых на основе этих двух функций распределения. К примеру, квантиль порядка 0.99 функции распределения $G(x) = 2\Phi(x) - 1$ модуля стандартной нормальной случайной величины примерно равна 2.576. В то же время квантиль того же порядка показательной функции распределения с параметром $\sqrt{2}$ примерно равна 3.256. Вероятность превышения порога 2.576, “критического” при функции распределения $G(x)$, максимальным суммарным воздействием, имеющим указанную показательную функцию распределения, превышает 0.026, то есть оказывается более чем в два с половиной раза выше, чем предполагаемая согласно классической модели. Этот пример наглядно

иллюстрирует, насколько можно недооценить риск катастроф, вызванных накапливающимися эффектами неблагоприятных воздействий, если не принимать во внимание стохастический характер интенсивности потока экстремальных событий.

6.5 Закон больших чисел для обобщенных процессов Кокса

Как уже говорилось, обобщенные процессы Кокса играют важную роль при моделировании характеристик неоднородных хаотических стохастических потоков случайных событий. При этом особую важность при использовании обобщенных процессов Кокса вида (6.6), скажем, в страховой математике при моделировании неоднородных потоков страховых выплат или в теории управления запасами при моделировании неоднородных потоков заявок на поставку некоторого продукта имеет тот случай, когда математическое ожидание слагаемых X_j в сумме (6.6) отлично от нуля (в частности, в условиях приведенных примеров – положительно). В терминах моделирования хаотических процессов с помощью неоднородных случайных блужданий отличие математического ожидания элементарного скачка X_j от нуля означает, что имеется систематический тренд: например, блуждающая частица дрейфует благодаря систематическому смещению среды, в которой осуществляется блуждание. Наличие такого тренда (дрейфа) существенно изменяет как условия существования нетривиального предельного распределения координаты блуждающей частицы, так и сами предельные распределения.

Для полноты изложения и лучшего понимания того, насколько изменяются основные статистические закономерности при непостоянной интенсивности потока информативных событий, здесь будет приведена простейшая формулировка закона больших чисел для обобщенных процессов Кокса с ненулевым средним.

Теорема 6.4. *Предположим, что $\Lambda(t) \rightarrow \infty$ по вероятности при $t \rightarrow \infty$ и $EX_1 = a \neq 0$. Для того чтобы существовала случайная*

величина Z такая, что

$$\frac{S(t)}{t} \Rightarrow Z \quad (t \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала неотрицательная случайная величина U такая, что

$$\frac{\Lambda(t)}{t} \Rightarrow U \quad (t \rightarrow \infty).$$

Более того, при этом $Z \stackrel{d}{=} aU$.

Теорема 6.4 представляет собой усиление одного результата О. Лундберга (Lundberg, 1964). Ее доказательство можно найти, например, в книге (Bening and Korolev, 2002).

Таким образом, как видно из Теоремы 6.4, в законе больших чисел для обобщенных процессов Кокса (как и для других более общих случайных сумм) предел “средних арифметических” может быть случайным. Это никак не противоречит традиционному представлению о сути законов больших чисел. Действительно, в классических законах больших чисел при увеличении числа слагаемых в средних арифметических информация о конкретном виде распределения отдельного слагаемого как бы затухает, сжимаясь в информацию о значении лишь одной характеристики распределения – математического ожидания. Именно этот эффект и описывается Теоремой 6.4, в которой (случайный) предел не зависит от вида распределения слагаемого, а зависит лишь от его математического ожидания. При этом тип предельного распределения полностью определяется статистическими закономерностями поведения накопленной интенсивности.

6.6 Моделирование распределений приращений финансовых индексов

Воспользуемся схемой неоднородных случайных блужданий с непрерывным временем для моделирования эволюции финансовых индексов и, прежде всего, биржевых цен. Эта предметная

область позволяет наглядно продемонстрировать основные принципы моделирования многих стохастических хаотических процессов, поскольку наблюдается явная аналогия между механизмами формирования процессов эволюции биржевых цен и других не менее важных хаотических процессов в других областях, например, траффика в сложных коммуникационных сетях (таких как Интернет) или процессов, протекающих в турбулентной плазме. Здесь будет построена конкретная модель, являющаяся основой СРС-метода анализа компонент волатильности.

Многие исследователи обращали внимание на то, что реальная интенсивность биржевых торгов крайне изменчива, и предпринимали попытки объяснить островершинность распределений приращений биржевых цен, основываясь на этой изменчивости. По-видимому, первым, кто систематически исследовал неоднородность операционного времени на биржах, был П. Кларк (Clark, 1970), (Clark, 1973). Вместо обычных винеровских процессов П. Кларк предложил в качестве моделей динамики биржевых цен использовать подчиненные винеровские процессы со случайным временем вида $W(X(t))$, где $W(t)$, $t \geq 0$ – винеровский процесс, а $X(t)$ – процесс с неубывающими траекториями, начинающимися в нуле.

По Кларку, при использовании центральной предельной теоремы в качестве базы при построении моделей для распределений приращений биржевых цен из-за неоднородности времени нарушается не конечность дисперсий, а другое структурное предположение, а именно, предположение о том, что число слагаемых в сумме неслучайно.

На основе догадки Кларка сейчас будет показано, как построить более реалистичную и довольно простую модель процесса эволюции биржевых цен на интервалах времени малой длины (так сказать, элементарную модель на микроуровне), а затем перейти к моделям типа $W(X(t))$ на макроуровне с помощью центральной предельной теоремы для случайных сумм. Эти модели, по сути, являются удобными аппроксимациями реальных процессов.

Рассмотрим процесс изменения цены акции на интервале вре-

мени малой длины. Основная посылка будет заключаться в том, что *формальное выражение цены акции существует только в заключенном на бирже контракте (сделке)*. Эта посылка не предполагает существования некоего “объективно существующего”, “основного” или “истинного” процесса эволюции биржевой цены и прямо противоречит принимаемой во многих исследованиях концепции, согласно которой проблемы участников рынка обусловлены незнанием этого процесса, характеризующего “истинную” цену. Таким образом, в соответствии с нашим основным предположением на интервалах времени между последовательными заключениями контрактов цена акции остается постоянной и равной той цене, которая указана в последнем заключенном до рассматриваемого момента времени контракте. Тогда изменение цены акции может быть описано маркированным точечным процессом $\{(T_j, P_j)\}_{j \geq 1}$, где T_j – момент заключения j -го контракта, а P_j – цена акции, указанная в j -ом контракте (контракты нумеруются в хронологическом порядке). Таким образом, если $P(\tau)$ – цена акции в момент τ , то

$$P(\tau) = P_j \text{ при } T_j \leq \tau < T_{j+1}.$$

Пусть $N(\tau)$ – число контрактов, заключенных к моменту τ (мы считаем, что отсчет начинается с $\tau = 0$). Как отмечено, например, в (Hull, 2003), ключевым моментом формирования цен акций и стратегии игры на бирже является то, что инвесторы, главным образом, заинтересованы в прибыли, выраженной в процентах от стоимости акций (то есть в относительной прибыли) и не зависящей от самой цены. Поэтому с учетом сказанного выше будем рассматривать следующее представление процесса $P(\tau)$:

$$\frac{P(\tau)}{P(0)} = \prod_{j=1}^{N(\tau)} \frac{P_j}{P_{j-1}}, \quad P_0 = P(0). \quad (6.7)$$

(Если бы нас интересовали абсолютные прибыли, то мультипликативное представление (6.7) следовало бы заменить его аддитив-

ным аналогом

$$P(\tau) - P(0) = \sum_{j=1}^{N(\tau)} (P_j - P_{j-1}), \quad P_0 = P(0),$$

что просто привело бы к тому, что в нижеследующих выкладках надо было бы опустить знак логарифма.) Обозначим

$$S(\tau) = \log P(\tau) - \log P(0), \quad X_j = \log P_j - \log P_{j-1}.$$

Тогда из (6.7) вытекает, что

$$S(\tau) = \sum_{j=1}^{N(\tau)} X_j \quad \text{и} \quad P(\tau) = P(0) \exp\{S(\tau)\}.$$

Предположим, что $X_1, X_2, \dots$ – независимые (и пока, вообще говоря, необязательно одинаково распределенные) случайные величины. Рассмотрим поведение распределения случайных величин $P(\tau)$ при неограниченном возрастании $N(\tau)$. Отметим, что это не означает, что обязательно $\tau \rightarrow \infty$. Отрезок $[0, \tau]$ (пока) остается фиксированным. Для удобства построения аппроксимаций формально рассмотрим последовательность целочисленных неотрицательных случайных величин $N_1, N_2, \dots$ и предположим, что при каждом $k \geq 1$ случайные величины N_k и $X_1, X_2, \dots$ независимы.

Пусть $f_k(s)$ – характеристическая функция случайной величины $S(\tau)$, в которой $N(\tau)$ заменено на N_k :

$$f_k(s) = \mathbb{E} \exp \left\{ is \sum_{j=1}^{N_k} X_j \right\}, \quad s \in \mathbb{R}.$$

Справедлив следующий результат.

Теорема 6.5. Пусть распределения слагаемых $X_1, X_2, \dots$ таковы, что для некоторых $b_k > 0$, $b_k \rightarrow \infty$ ($k \rightarrow \infty$) имеет место сходимость

$$\mathbb{P} \left(\frac{1}{b_k} \sum_{j=1}^k X_j < x \right) \Longrightarrow \Phi(x) \quad (k \rightarrow \infty).$$

Предположим, что при некоторых $d_k > 0$, $d_k \rightarrow \infty$ ($k \rightarrow \infty$) семейство случайных величин $\{b_{N_k}/d_k\}$ слабо компактно. Тогда при каждом $s \in \mathbb{R}$

$$\lim_{k \rightarrow \infty} \left| f_k \left(\frac{s}{d_k} \right) - \int_0^\infty \exp \left\{ -\frac{x^2 s^2}{2} \right\} dA_k(x) \right| = 0,$$

где $A_k(x) = P(b_{N_k} < xd_k)$.

Доказательство этого результата можно найти в (Gnedenko and Korolev, 1996).

Очевидно, что, если обозначить $U_k = b_{N_k}/d_k$, то

$$\int_0^\infty \exp \left\{ -\frac{x^2 s^2}{2} \right\} dA_k(x) = E \exp \{ is(XU_k) \}, \quad s \in \mathbb{R},$$

где $P(X < x) = \Phi(x)$; более того, X и U_k независимы при каждом $k \geq 1$. Таким образом, с помощью Теоремы 6.5 мы приходим к следующему заключению: если число контрактов, заключенных в течение интервала времени $[0, \tau]$, довольно велико и последовательные изменения цены акции от контракта к контракту удовлетворяют условиям центральной предельной теоремы, то для некоторой неотрицательной случайной величины U справедливо приближенное равенство

$$P(S(\tau) < x) \approx P(XU < x) = E \Phi \left(\frac{x}{U} \right), \quad x \in \mathbb{R},$$

где $P(X < x) = \Phi(x)$ и случайные величины X и U независимы. Другими словами, законы, аппроксимирующие распределения приращений логарифмов процессов эволюции цен акций следует искать среди масштабных смесей нормальных законов.

Моменты заключения контрактов на бирже хаотически рассредоточены на оси времени. Однако этот хаос не является однородным, так как интенсивность торгов изменяется в течение биржевого дня: сразу после начала торгов совершают свои сделки те трейдеры, которые получили заказ до открытия торгов, затем интенсивность несколько спадает. В течение дня наблюдается несколько систематических всплесков интенсивности торгов, связанных с

поступлением информации о положении дел на открывающихся биржах, расположенных в других часовых поясах. Ближе к концу торгов интенсивность возрастает, так как совершают свои сделки те трейдеры, которые выжидали в течение дня. Наряду с “систематическими”, то есть наблюдаемыми день ото дня, всплесками интенсивности, конечно же, возможны и случайные всплески, связанные, например, с теми или иными политическими событиями. Поэтому разумными моделями для реальных потоков заключения контрактов на биржах являются процессы Кокса.

Таким образом, будем считать, что введенная выше случайная величина $S(\tau)$ является значением обобщенного процесса Кокса в момент τ . Одними из основных характеристик, интересующих тех, кто играет на бирже, являются максимальная и минимальная цены за определенный интервал времени, например, биржевой день. Информация об этих характеристиках существенно определяет стратегию поведения того или иного игрока на следующем интервале времени. Обозначим

$$\underline{P}(t) = \min_{0 \leq \tau \leq t} P(\tau), \quad \bar{P}(t) = \max_{0 \leq \tau \leq t} P(\tau) \quad (6.8)$$

Будем считать, что для управляющего процесса $E(\Lambda(\tau)) = \lambda \tau$ при некотором $\lambda > 0$. Мы построим асимптотические аппроксимации для распределений случайных величин $\underline{P}(t)$ и $\bar{P}(t)$. Но по смыслу задачи разумно считать фиксированной длину t интервала времени, на котором рассматриваются экстремумы процесса эволюции цен. Поэтому поведение случайных величин $\underline{P}(t)$ и $\bar{P}(t)$ будем изучать при неограниченно растущей средней интенсивности торгов, то есть при $\lambda \rightarrow \infty$. Такой подход вполне оправдан, поскольку реально на крупных биржах за интервалы времени $[0, t]$, представляющие разумный интерес (часы, дни), заключается огромное число контрактов. Таким образом, в задачу вводится еще один параметр λ и поэтому фактически $\Lambda(\tau) = \Lambda(\tau, \lambda)$, $P(\tau) = P(\tau, \lambda)$, $\underline{P}(t) = \underline{P}(t, \lambda)$ и $\bar{P}(t) = \bar{P}(t, \lambda)$. Функция распределения максимума стандартного винеровского процесса на отрезке $[0, 1]$ будет обозначаться $G(x)$,

$$G(x) = 2\Phi(\max\{0, x\}) - 1, \quad x \in \mathbb{R}.$$

Предположим, что случайные величины

$$X_j = \log P_j - \log P_{j-1}, \quad j \geq 1,$$

независимы и одинаково распределены с

$$EX_j = 0 \text{ и } 0 < \sigma^2 = DX_j < \infty.$$

Теорема 6.6. *Предположим, что $\Lambda(t, \lambda) \rightarrow \infty$ по вероятности при $\lambda \rightarrow \infty$. Тогда следующие условия эквивалентны:*

$$1^\circ. \ P \left(\frac{\log P(t, \lambda) - \log P(0)}{\sigma \sqrt{\lambda t}} < x \right) \Rightarrow F(x) \quad (\lambda \rightarrow \infty);$$

$$2^\circ. \ P \left(\frac{\log \bar{P}(t, \lambda) - \log P(0)}{\sigma \sqrt{\lambda t}} < x \right) \Rightarrow \bar{F}(x) \quad (\lambda \rightarrow \infty);$$

$$3^\circ. \ P \left(\frac{\log \underline{P}(t, \lambda) - \log P(0)}{\sigma \sqrt{\lambda t}} < x \right) \Rightarrow \underline{F}(x) \quad (\lambda \rightarrow \infty);$$

4°. *существует неотрицательная случайная величина U такая,*

что

$$\frac{\Lambda(t, \lambda)}{t\lambda} \Rightarrow U \quad (\lambda \rightarrow \infty).$$

При этом

$$F(x) = E\Phi\left(x/\sqrt{U}\right), \quad \bar{F}(x) = EG\left(x/\sqrt{U}\right), \quad \underline{F}(x) = 1 - EG\left(-x/\sqrt{U}\right).$$

Доказательство см. в (Королев, 2000).

Обратим внимание, что предельные законы в Теореме 6.6 имеют существенно более тяжелые хвосты по сравнению с лог- нормальным или лог- G распределениями.

Основываясь на Теореме 6.6, можно высказать предположение, что, скорее всего, не существует одной универсальной случайной величины U , точнее, одного разумным образом параметризованного семейства таких случайных величин, которое удовлетворительно описывает поведение всех акций на всех биржах. Действительно, условие 4° теоремы 6.6 напрямую связывает случайную

величину U с изменением интенсивностей торгов или волатильностями, которые, разумеется, сильно зависят от объекта торговли, времени и места торгов. По-видимому, самое большее, на что можно надеяться, – это более или менее удачная подгонка “локально универсальной” параметрической модели, то есть для каждого конкретного объекта торговли, каждой конкретной биржи можно попробовать подобрать свою случайную величину U , то есть ее распределение, параметры которого все же заметно меняются во времени.

С помощью Теоремы 6.6 можно в определенном смысле уточнить утверждения, объясняющие наблюдаемую не-нормальность распределений логарифмов приращений биржевых цен на интервалах времени умеренной длины, которая проявляется в более тяжелых хвостах у реальных распределений. В то же время распределения приращений процессов за интервалы времени большей длины намного ближе к нормальным, нежели распределения приращений процессов за интервалы меньшей длины. А именно, при увеличении длины интервала времени, приращения за который рассматриваются, наблюдается эффект сближения распределений логарифмов относительных приращений финансовых индексов, в частности, биржевых цен, с нормальным законом. При рассмотрении экстремумов процессов эволюции цен аналогом этого феномена должно быть сближение распределения максимума процесса эволюции цен с функцией распределения $G(x)$, а распределения минимума этого процесса – с функцией распределения $1 - G(-x)$.

Построенные модели эволюции биржевых цен довольно естественно объясняют этот эффект. В противоположность тем рассуждениям, которые привели к Теореме 6.6, сейчас длина интервала времени t не будет считаться фиксированной, а напротив, в предположении, что она измеряется в каких-либо целых единицах (например, биржевых днях), она будет устремляться к бесконечности при постоянной средней интенсивности торгов λ . Итак, пусть $t = n = 1, 2, \dots$. Рассмотрим поведение распределений случайных величин $\underline{P}(n) = \underline{P}(n, \lambda)$ и $\bar{P}(n) = \bar{P}(n, \lambda)$, определяемых соотношением (6.8) с заменой t на n , при $n \rightarrow \infty$. В этой ситуации разумно

считать, что управляющий процесс $\Lambda(n, \lambda)$ имеет вид

$$\Lambda(n, \lambda) = \sum_{i=1}^n \Lambda_i,$$

где $\Lambda_1, \Lambda_2, \dots$ – независимые одинаково распределенные случайные величины с $E\Lambda_i = \lambda, i \geq 1$. Тогда в силу закона больших чисел при $n \rightarrow \infty$

$$\frac{\Lambda(n, \lambda)}{n\lambda} = \frac{1}{n\lambda} \sum_{i=1}^n \Lambda_i \Rightarrow 1,$$

то есть распределение случайной величины U в Теореме 6.6 вырождено в единицу. Поэтому в этом случае в Теореме 6.6

$$F(x) = \Phi(x), \quad \bar{F}(x) = G(x), \quad \underline{F}(x) = 1 - G(-x).$$

Глава 7

Смешанные вероятностные модели

7.1 Смеси вероятностных распределений

В центральной предельной теореме для обобщенных процессов Кокса предельными законами оказываются масштабные смеси нормальных законов с нулевыми средними. Класс таких смесей очень богат и содержит много типов распределений с существенно различным поведением хвостов. К примеру, помимо самого нормального распределения с очень быстро (как $|x|^{-1}e^{-x^2}$) убывающими (при $|x| \rightarrow \infty$) хвостами, этот класс содержит распределение Лапласа (двойное показательное распределение), плотность которого имеет вид $l(x) = \frac{1}{2}\mu e^{-\mu|x|}$, $x \in \mathbb{R}$ ($\mu > 0$ – параметр) и хвосты которого убывают при $|x| \rightarrow \infty$ как $e^{-\mu|x|}$; распределение Стюдента с хвостами, убывающими как $|x|^{-\gamma}$ с произвольным $\gamma > 0$, в частности, распределение Коши, соответствующее случаю $\gamma = 1$; симметричные строго устойчивые законы с таким же степенным убыванием хвостов с $\gamma \in (0, 2)$; симметризованные гамма-распределения, хвосты которых ведут себя как $|x|^a e^{-b|x|}$ с $-1 < a < \infty$, $b > 0$; симметризованное распределение Вейбулла–Гнеденко с хвостами, убывающими как $e^{-|x|^d}$ при $0 < d < 1$; а также много других типов, см., например, (Королев, 2011). Этот пример показывает, что свойства смесей вероятностных распределений могут существенно отличаться от свойств смешиваемых законов.

Дадим определение смеси вероятностных распределений. Рассмотрим функцию $F(x, \mathbf{y})$, определенную на множестве $\mathbb{R} \times \mathcal{Y}$. Для простоты мы будем предполагать, что $\mathcal{Y}$ – это некоторое подмножество m -мерного евклидова пространства, $\mathcal{Y} \subseteq \mathbb{R}^m$ при некотором $m \geq 1$, причем множество $\mathcal{Y}$ снабжено борелевской

σ -алгеброй Σ . Более того, предположим, что при каждом фиксированном y функция $F(x, y)$ является функцией распределения по x , а при каждом фиксированном x функция $F(x, y)$ измерима по y , то есть для любых $x \in \mathbb{R}$ и $c \in \mathbb{R}$ выполнено условие $\{y : F(x, y) < c\} \in \Sigma$. Пусть Q – вероятностная мера, определенная на измеримом пространстве $(\mathcal{Y}, \Sigma)$. Функция распределения

$$H(x) = \int_{\mathcal{Y}} F(x, y) Q(dy), \quad x \in \mathbb{R},$$

называется *смесью* функции распределения $F(x, y)$ по y относительно Q . Распределение $F(x, y)$ называется *смешиваемым*, в то время как мера Q задает *смешивающее* распределение.

Если Y – m -мерная тождественная случайная величина (то есть $Y(y) \equiv y$, $y \in \mathcal{Y}$), определенная на вероятностном пространстве $(\mathcal{Y}, \Sigma, Q)$, то функция распределения $H(x)$ может быть записана в виде

$$H(x) = EF(x, Y), \quad x \in \mathbb{R}.$$

Если $f(x, y)$ – плотность распределения, соответствующая функции распределения $F(x, y)$,

$$f(x, y) = \frac{d}{dx} F(x, y),$$

то смеси $H(x)$ соответствует плотность

$$h(x) = Ef(x, Y) = \int_{\mathcal{Y}} f(x, y) Q(dy), \quad x \in \mathbb{R}.$$

Если случайный вектор Y имеет дискретное распределение и принимает значения $y_1, y_2, \dots$ соответственно с вероятностями $p_1, p_2, \dots$, то мы получаем смесь вида

$$H(x) = EF(x, Y) = \sum_{j \geq 1} p_j F(x, y_j), \quad x \in \mathbb{R},$$

называемую *дискретной*. В таком случае функции распределения $F(x, \mathbf{y}_j)$ называются *компонентами* смеси $H(x)$, а числа p_j называются *весами* соответствующих компонент, $j \geq 1$. Если в дискретной смеси число ненулевых весов конечно (то есть случайный вектор $\mathbf{Y}$ принимает конечное число значений), то дискретная смесь называется *конечной*.

Если в таком случае функции распределения $F(x, \mathbf{y})$ соответствует плотность $f(x, \mathbf{y})$, то дискретной смеси $H(x)$ соответствует плотность

$$h(x) = \mathbb{E}f(x, \mathbf{Y}) = \sum_{j \geq 1} p_j f(x, \mathbf{y}_j), \quad x \in \mathbb{R}.$$

Особо подчеркнем, что при разных значениях параметра $\mathbf{y}$ функции распределения $F(x, \mathbf{y})$ могут относиться к *разным* типам распределения. Например, если смесь дискретна, то есть мера $\mathbb{Q}$ приписывает вероятности p_j точкам $\mathbf{y}_j$, $j = 1, 2, \dots$, причем $F(x, \mathbf{y}_j) = F_j(x)$, где $F_j(x)$ – функции распределения, возможно, соответствующие разным типам при разных j , то

$$H(x) = \mathbb{E}F(x, \mathbf{Y}) = \sum_{j \geq 1} p_j F_j(x), \quad x \in \mathbb{R}.$$

Более того, если в таком случае компоненты F_j абсолютно непрерывны и имеют плотности f_j , то смесь $H(x)$ также будет абсолютно непрерывной с плотностью

$$h(x) = \mathbb{E}f(x, \mathbf{Y}) = \sum_{j \geq 1} p_j f_j(x), \quad x \in \mathbb{R}.$$

В дальнейшем особую роль будут играть сдвиг/масштабные смеси. Формально они определяются следующим образом. Пусть в определении смеси, сформулированном выше, $m = 2$. Предположим, что вектор $\mathbf{y}$ имеет вид

$$\mathbf{y} = (u, v),$$

где $u > 0$ и $v \in \mathbb{R}$, так что функция распределения $F(x, \mathbf{y})$ допускает представление

$$F(x, \mathbf{y}) = F\left(\frac{x - v}{u}\right), \quad x \in \mathbb{R}.$$

Тогда $\mathcal{Y}$ – это положительная полуплоскость, то есть $\mathcal{Y} = \mathbb{R}^+ \times \mathbb{R}$, и функция распределения

$$H(x) = \int_{\mathcal{Y}} F\left(\frac{x-v}{u}\right) Q(du, dv), \quad x \in \mathbb{R}, \quad (7.1)$$

называется *сдвиг/масштабной смесью* функции распределения F относительно Q . Здесь u – это параметр масштаба, а v – параметр сдвига (положения). Если функция распределения F имеет плотность f , то функции распределения $F((x-v)/u)$ соответствует плотность

$$f(x, \mathbf{y}) = \frac{1}{u} f\left(\frac{x-v}{u}\right),$$

так что смеси (7.1) соответствует плотность

$$h(x) = \int_{\mathcal{Y}} \frac{1}{u} f\left(\frac{x-v}{u}\right) Q(du, dv), \quad x \in \mathbb{R}.$$

Если X , U и V – случайные величины, заданные на одном и том же достаточно богатом вероятностном пространстве, так что при каждом фиксированном значении (u, v) пары случайных величин (U, V) случайная величина X имеет функцию распределения $F((x-v)/u)$, то смесь (7.1) может быть записана в виде

$$H(x) = \mathbb{E}F\left(\frac{x-V}{U}\right), \quad x \in \mathbb{R}.$$

Более того, в таком случае из теоремы Фубини вытекает, что функция распределения $H(x)$ соответствует случайной величине $X \cdot U + V$, где случайная величина X и случайный вектор $\mathbf{Y} = (U, V)$ стохастически независимы. Кстати, легко убедиться, что в таком контексте чисто сдвиговая смесь $H(x) = \mathbb{E}F(x - V)$ является не чем иным как функцией распределения суммы двух независимых случайных величин X и V , то есть сверткой их функций распределения. В то же время, чисто масштабная смесь $H(x) = \mathbb{E}F(x/U)$ является не чем иным как функцией распределения произведения двух независимых случайных величин X и U .

Чтобы определить дискретную сдвиг/масштабную смесь функции распределения $F(x)$, положим $\mathbf{y}_j = (\sigma_j, a_j)$, где $a_j \in \mathbb{R}$, $\sigma_j > 0$, $j = 1, \dots, k$, и в качестве специального случая приведенного выше определения дискретной смеси получим

$$H(x) = \sum_{j=1}^k p_j F\left(\frac{x - a_j}{\sigma_j}\right), \quad x \in \mathbb{R}. \quad (7.2)$$

Если при этом $a_j = 0$, $j = 1, \dots, k$, то мы получаем чисто масштабную конечную смесь

$$H(x) = \sum_{j=1}^k p_j F\left(\frac{x}{\sigma_j}\right), \quad x \in \mathbb{R}.$$

Если функция распределения F абсолютно непрерывна и имеет плотность $f = F'$, то смеси (7.2) функций распределения соответствует смесь плотностей

$$h(x) = \sum_{j=1}^k \frac{p_j}{\sigma_j} f\left(\frac{x - a_j}{\sigma_j}\right), \quad x \in \mathbb{R}.$$

Физический смысл понятия смеси вероятностных распределений может быть проиллюстрирован на примере дискретной смеси. Рассмотрим некоторую популяцию, которая не является однородной и в свою очередь состоит из некоторого числа, скажем, k суб-популяций. Предположим, что наблюдаемая характеристика или наблюдаемый признак внутри j -й суб-популяции распределен в соответствии с функцией распределения $F_j(x) \equiv F(x, \mathbf{y}_j)$, которую можно интерпретировать как условную вероятность того, что значение наблюдаемого признака у случайно выбранного индивидуума будет меньше, чем x , при условии, что случайно выбранный индивидуум является представителем j -й суб-популяции. Пусть вероятность того, что при случайном выборе индивидуума из всей (генеральной) популяции будет выбран представитель именно j -й суб-популяции, равна p_j ($p_j \geq 0$, $p_1 + \dots + p_k = 1$). Тогда по формуле полной вероятности безусловная вероятность того, что значение наблюдаемого признака у индивидуума, случайно выбранного из

всей генеральной популяции, будет меньше, чем x , окажется равной

$$H(x) = \sum_{j=1}^k p_j F_j(x).$$

В определенном смысле операция смешивания вероятностных распределений обеспечивает возможность формально интерпретировать популяции, реально являющиеся неоднородными, как однородные.

7.2 Идентифицируемость смесей вероятностных распределений

Очень часто нельзя непосредственно определить тип суб-популяции, к которой принадлежит очередное наблюдение. Вследствие этого вся (генеральная) популяция вынужденно считается однородной, хотя на самом деле она таковой не является и содержит индивидуумов, принадлежащих к существенно различным типам. Именно такая ситуация типична для анализа многих хаотических стохастических процессов. Поэтому чрезвычайно важно иметь возможность осуществить операцию, в некотором смысле обратную операции смешивания, а именно, операцию разделения (расщепления) смесей. Статистические процедуры, реализующие эту операцию, в значительной степени зависят от свойства идентифицируемости смесей вероятностных распределений.

Понятие идентифицируемой смеси интенсивно используется в прикладных задачах, связанных с декомпозицией (разделением, разложением, расщеплением) совокупностей (популяций). В качестве примеров можно упомянуть задачи классификации, распознавания образов или идентификации вероятностных распределений. Библиография по этим вопросам обширна, см., например, обзоры (Исаенко и Урбах, 1976), (Круглов, 1991), или книги (Titterington, Smith and Makov, 1987), (Айвазян, Бухштабер, Енюков и Мешалкин, 1989) и списки литературы в указанных источниках.

Дадим определение идентифицируемых семейств смесей вероятностных распределений. Оно было предложено в работе

(Teicher, 1961).

Пусть функция $F(x, \mathbf{y})$ определена на множестве $\mathbb{R} \times \mathcal{Y}$, где $\mathcal{Y} \subseteq \mathbb{R}^m$ при некотором $m \geq 1$. Пусть множество $\mathcal{Y}$ снабжено борелевской σ -алгеброй Σ . Как и ранее, мы предполагаем, что функция $F(x, \mathbf{y})$ измерима по $\mathbf{y}$ при каждом фиксированном x и является функцией распределения как функция аргумента x при каждом фиксированном $\mathbf{y}$. Пусть $\mathcal{Q}$ – семейство случайных величин, принимающих значения во множестве $\mathcal{Y}$. Обозначим

$$\mathcal{H} = \{H_Q(x) = EF(x, Q), x \in \mathbb{R} : Q \in \mathcal{Q}\}. \quad (7.3)$$

Определение 7.1. Семейство $\mathcal{H}$, определяемое ядром F и множеством $\mathcal{Q}$, называется *идентифицируемым*, если из равенства

$$EF(x, Q_1) = EF(x, Q_2), \quad x \in \mathbb{R},$$

с $Q_1 \in \mathcal{Q}$, $Q_2 \in \mathcal{Q}$ вытекает, что $Q_1 \stackrel{d}{=} Q_2$.

К примеру, идентифицируемыми являются конечные смеси нормальных распределений, показательных распределений, пуассоновских распределений и распределений Коши. Однако, легко привести очень простые примеры неидентифицируемых семейств. В частности, в качестве ядра рассмотрим равномерное распределение и связанные с ним смеси

$$\frac{1}{3} \cdot 3 \cdot \mathbf{1}_{[0, \frac{1}{3})}(x) + \frac{2}{3} \cdot \frac{3}{2} \cdot \mathbf{1}_{[\frac{1}{3}, 1)}(x) = \frac{1}{2} \cdot 2 \cdot \mathbf{1}_{[0, \frac{1}{2})}(x) + \frac{1}{2} \cdot 2 \cdot \mathbf{1}_{[\frac{1}{2}, 1)}(x).$$

Если X , U и V – случайные величины, определенные на одном и том же достаточно богатом вероятностном пространстве так, что случайная величина X стохастически независима от пары (U, V) и для любых фиксированных значений $u \in \mathbb{R}^+$ случайной величины U и $v \in \mathbb{R}$ случайной величины V случайная величина X имеет функцию распределения $F((x - v)/u)$, то приведенное выше определение идентифицируемости применительно к сдвиг-/масштабным смесям сводится к следующему.

Сдвиг/масштабная смесь $H(x) = EF((x - V)/U)$, порожденная ядром F (сдвиг/масштабная смесь функции распределения F)

идентифицируема, если из соотношения

$$X_1 \cdot U_1 + V_1 \stackrel{d}{=} X_2 \cdot U_2 + V_2,$$

где (X_i, U_i, V_i) , $i = 1, 2$, – тройки случайных величин, обладающих точно такими же свойствами, что присущи описанным выше случайным величинам X, U, V , вытекает, что

$$(U_1, V_1) \stackrel{d}{=} (U_2, V_2).$$

Сужение определения идентифицируемости на класс конечных сдвиг/масштабных смесей сводится к следующему.

Семейство смесей

$$\mathcal{H} = \left\{ \sum_{j=1}^k p_j F\left(\frac{x-a_j}{\sigma_j}\right) : \right. \\ \left. k \geq 1; p_j \geq 0, p_1 + \dots + p_k = 1; a_j \in \mathbb{R}, \sigma_j > 0, j = \overline{1, k} \right\},$$

порожденное ядром F , идентифицируемо, если из равенства

$$\sum_{j=1}^k p_j F\left(\frac{x-a_j}{\sigma_j}\right) = \sum_{i=1}^m q_i F\left(\frac{x-b_i}{\delta_i}\right)$$

вытекает, что

- (i) $k = m$;
- (ii) для каждого индекса $j \in \{1, \dots, k\}$ существует такой индекс $i \in \{1, \dots, k\}$, что

$$p_j = q_i, \quad a_j = b_i, \quad \sigma_j = \delta_i.$$

Теперь мы перейдем к рассмотрению условий идентифицируемости. Сначала рассмотрим условия идентифицируемости смесей, в которых смешивание производится либо по параметру сдвига, либо по параметру масштаба. Другими словами, сначала мы рассмотрим семейства однопараметрических смесей. Условия идентифицируемости таких семейств хорошо известны. Приведем некоторые из них.

Семейство функций распределения $\{F(x, y) : y > 0\}$ называется *аддитивно замкнутым*, если для любых $y_1 > 0, y_2 > 0$ справедливо соотношение

$$(F_1 * F_2)(x) \equiv F(x, y_1 + y_2), \quad (7.4)$$

где $F_1(x) = F(x, y_1)$, $F_2(x) = F(x, y_2)$.

Иногда свойство (7.4) семейств распределений вероятностей называется *воспроизводимостью* по параметру y .

Семейство нормальных законов с нулевым математическим ожиданием $\mathcal{N}_0 = \{\Phi(x/\sqrt{s}), s > 0\}$ является очевидным примером аддитивно замкнутого семейства (относительно дисперсии s).

Следующие результаты принадлежат Г. Тейчеру (Teicher, 1961).

Теорема 7.1. Семейство смесей (7.3) функций распределения $F(x, \cdot)$ из аддитивно замкнутого семейства является идентифицируемым.

Отсюда немедленно вытекает, что семейство $\mathcal{N}_0$ масштабных смесей нормальных законов с нулевым средним идентифицируемо.

Смеси, порожаемые ядрами из аддитивно замкнутых семейств, конечно же не исчерпывают все примеры идентифицируемых смесей. Рассмотрим масштабные смеси распределений, сосредоточенных на неотрицательной полупрямой.

Теорема 7.2. Пусть $F(x, y) = F_0(xy)$, $y \geq 0$, где F_0 – функция распределения такая, что $F_0(0) = 0$. Предположим, что преобразование Фурье–Стилтьеса функции $G^*(y) = F_0(e^y)$, $y \geq 0$, нигде не обращается в нуль. Тогда семейство смесей

$$\mathcal{H} = \{H_Q(x) = EF(x, Q), x \geq 0 : P(Q > 0) = 1\}$$

идентифицируемо.

Аналогичное свойство присуще некоторым семействам сдвиговых смесей вероятностных распределений.

Теорема 7.3. Пусть $\mathcal{Q}$ – множество всех случайных величин. Семейство сдвиговых смесей

$$\mathcal{H} = \{H_Q(x) = EF(x - Q), x \in \mathbb{R} : Q \in \mathcal{Q}\},$$

идентифицируемо, если характеристическая функция, соответствующая функции распределения $F(x)$, нигде не обращается в нуль.

Теорема 7.1 гарантирует, что наряду со смесями нормальных законов с нулевым средним, идентифицируемыми являются семейства масштабных смесей любых строго устойчивых законов.

Теорема 7.3 гарантирует идентифицируемость семейств сдвиговых смесей любых безгранично делимых (в том числе устойчивых) законов.

Здесь мы упомянули только некоторые идентифицируемые семейства, которые так или иначе рассматриваются в данной книге. Многочисленные примеры других идентифицируемых семейств можно найти в работах (Medgyessy, 1961), (Teicher, 1961), (Teicher, 1963), (Yakowitz and Spragins, 1968).

Некоторые критерии (то есть необходимые и достаточные условия) идентифицируемости семейств смесей доказаны в работе (Tallis, 1969), также см. обзор (Kruglov, 1991).

К сожалению, примеры ядер, порождающих идентифицируемые семейства сдвиг/масштабных смесей в общей ситуации (то есть без каких-либо дополнительных условий на смешивающие распределения) неизвестны. Такие примеры известны лишь для дискретных смешивающих законов. В частности, справедливо следующее утверждение, доказанное Г. Тейчером (Teicher, 1963).

Теорема 7.4. Семейство конечных сдвиг/масштабных смесей нормальных законов идентифицируемо.

Покажем, что семейство сдвиг/масштабных смесей нормальных законов при произвольном (двумерном) смешивающем законе не является идентифицируемым. Пусть X_1 , X_2 , U_1 и U_2 – независимые случайные величины такие, что X_1 и X_2 имеют одно и

то же стандартное нормальное распределение, $P(U_1 = 1) = 1$, а случайная величина U_2 невырождена, то есть ни при каком u не имеет места соотношение $P(U_2 = u) = 1$. Более того, предположим, что $P(U_2 > 0) = 1$. Тогда распределение случайной величины $Z = X_1 U_2 + X_2$ может быть записано как в виде

$$P(Z < x) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left(\frac{x-v}{u}\right) dP(X_2 < v) dP(U_2 < u), \quad (7.5)$$

так и в виде

$$P(Z < x) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left(\frac{x-v}{u}\right) dP(X_1 U_2 < v) dP(U_1 < u). \quad (7.6)$$

Легко видеть, что смешивающие двумерные распределения в представлениях (7.5) и (7.6) различны.

7.3 Островершинность масштабных смесей нормальных законов

В связи с упомянутой выше островершинностью эмпирических закономерностей вполне уместно обсудить важное с точки зрения рассматриваемых моделей свойство масштабных смесей нормальных законов, а именно, их островершинность и “тяжелохвостость”.

Понятие островершинности распределения можно определять по-разному. К примеру, в прикладных исследованиях в качестве численной характеристики островершинности часто рассматривается коэффициент эксцесса $\varkappa(Z)$, который для случайной величины Z с $EZ^4 < \infty$ определяется как

$$\varkappa(Z) = E\left(\frac{Z - EZ}{\sqrt{DZ}}\right)^4.$$

Если $P(X < x) = \Phi(x)$, то $\varkappa(X) = 3$. Плотности с более острыми вершинами (и, соответственно, более тяжелыми хвостами), чем у

нормальной плотности, имеют $\kappa > 3$, а для плотностей с менее острой вершиной $\kappa < 3$.

Следующее утверждение устанавливает, что смеси $E\Phi(x/\sqrt{U})$ всегда являются более островершинными и, следовательно, имеют более тяжелые хвосты, нежели нормальный закон, если в качестве характеристики островершинности рассматривается коэффициент эксцесса.

Лемма 7.1. Пусть X и Y – независимые случайные величины с конечными четвертыми моментами. Предположим, что $EX = 0$ и $P(Y \geq 0) = 1$. Тогда

$$\kappa(XY) \geq \kappa(X).$$

Более того,

$$\kappa(XY) = \kappa(X)$$

тогда и только тогда, когда

$$P(Y = \text{const}) = 1.$$

Доказательство. Из независимости X и Y следует, что

$$\begin{aligned} \kappa(XY) &= E\left(\frac{XY - EXY}{\sqrt{DXY}}\right)^4 = \frac{E(XY - EXY)^4}{(E(XY - EXY)^2)^2} = \\ &= \frac{E(XY - EXEY)^4}{(E(XY - EXEY)^2)^2} = \frac{EX^4EY^4}{(EX^2)^2(EY^2)^2} = \kappa(X) \cdot \frac{EY^4}{(EY^2)^2}. \end{aligned} \quad (7.7)$$

Но согласно неравенству Иенсена, $EY^4 \geq (EY^2)^2$. Поэтому правая часть соотношения (7.7) всегда не меньше, чем $\kappa(X)$. Более того, она равна $\kappa(X)$ тогда и только тогда, когда $EY^4 = (EY^2)^2$, что возможно лишь, только если $P(Y = \text{const}) = 1$. Лемма доказана.

Таким образом, если X – стандартная нормальная случайная величина, а U – неотрицательная случайная величина с конечным вторым моментом, которая не зависит от X , то по теореме Фубини $P(X \cdot \sqrt{U} < x) = E\Phi(x/\sqrt{U})$. При этом из Леммы 7.1 следует, что $\kappa(X \cdot \sqrt{U}) \geq 3$ и $\kappa(X \cdot \sqrt{U}) = 3$, если и только если U вырождена, то есть неслучайна.

К слову сказать, равенство коэффициента эксцесса минимально возможному значению (равному трем) характеризует нормальный закон не только в классе масштабных смесей нормальных законов (см. вышеприведенную лемму), но и в классе всех безгранично делимых законов с конечным четвертым моментом (Круглов, 2003).

Понятие “распределения с тяжелыми хвостами”, естественно, тесно связанное с понятием островершинности, также можно определить по-разному. В этой книге мы не придерживаемся какого-либо единого формального определения распределения с тяжелыми хвостами, понимая под таковым распределение, хвосты которого имеют более низкую скорость убывания по сравнению с нормальным законом при неограниченном росте аргумента. В статье (Birnbaum, 1948) предложено вместо “абсолютной” тяжести хвостов распределений рассматривать “относительную”, сравнивая вероятности больших отклонений. Следуя этому подходу и используя неравенство Иенсена, легко получить неравенства, напрямую связывающие хвосты масштабных смесей нормальных законов с хвостами самого нормального распределения. Пусть, как и ранее, X – стандартная нормальная случайная величина, а U – неотрицательная случайная величина, независимая от X . Плотность случайной величины $Z = X \cdot \sqrt{U}$ (всегда существующую, симметричную и одновершинную) обозначим $p_Z(x)$. Легко видеть, что $P(Z > x) = 1 - E\Phi(x/\sqrt{U})$ для $x > 0$.

Лемма 7.2. *Для $x > 0$ справедливо неравенство*

$$P(Z > x) \geq 1 - \Phi(\sqrt{2\pi x p_Z(0)}).$$

Если случайная величина U удовлетворяет условию нормировки

$$EU^{-1/2} = 1,$$

то

$$P(Z > x) \geq 1 - \Phi(x), \quad x > 0.$$

Из Леммы 7.2 вытекает, что, если $EU^{-1/2} = 1$, то для любого $x \geq 0$

$$P(|X \cdot \sqrt{U}| \geq x) \geq P(|X| \geq x) \quad (= 2[1 - \Phi(x)]),$$

то есть масштабные смеси нормальных законов всегда имеют более тяжелые хвосты, нежели само нормальное распределение.

7.4 Устойчивость смесей нормальных законов относительно смешивающего распределения

В этом разделе будет показано, что семейство масштабных смесей нормальных законов обладает свойством устойчивости относительно смешивающего распределения. Несмотря на то что задача разделения общих масштабных смесей нормальных законов имеет единственное решение, на практике невозможно ее решить с абсолютной точностью, поскольку сама исходная смесь не известна. Даны лишь результаты наблюдений, о которых известно, что они представляют собой реализации случайных величин, которые распределены в соответствии с этой неизвестной смесью. На практике задача статистического разделения смесей совпадает с задачей статистической реконструкции самой смеси. Таким образом, свойство устойчивости смесей относительно смешивающего распределения становится решающим для реалистичности и адекватности результатов статистического анализа, получаемых с помощью метода скользящего разделения смесей (CPC-метода).

Прямая задача

Пусть X_i и Y_i – независимые случайные величины и пусть, более того, $P(Y_i \geq 0) = 1$, $i = 1, 2$. Функции распределения случайных величин X_i и Y_i будут, соответственно, обозначаться $F_i(x)$ и $G_i(x)$. Равномерное расстояние между случайными величинами X и Y с соответствующими функциями распределения $F_X(x)$ и $F_Y(x)$ будет обозначаться $\rho(X, Y)$,

$$\rho(X, Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|.$$

Наша ближайшая цель – оценить величину

$$\Delta = \rho\left(\frac{X_1}{Y_1}, \frac{X_2}{Y_2}\right)$$

через $\rho(X_1, X_2)$ и $\rho(Y_1, Y_2)$. Обозначим

$$\Delta(x) \equiv \left| P\left(\frac{X_1}{Y_1} < x\right) - P\left(\frac{X_2}{Y_2} < x\right) \right|.$$

Легко видеть, что для $-\infty < x < +\infty$ справедливо соотношение

$$\begin{aligned} P\left(\frac{X}{Y} < x\right) &= P(X < 0)P(Y = 0) + P(X < xY; Y > 0) = \\ &= F_X(0)[F_Y(0+) - F_Y(0)] + \int_{0+}^{\infty} F_X(xy) dF_Y(y). \end{aligned}$$

Поэтому

$$\begin{aligned} \Delta(x) &\leq F_1(0)[G_1(0+) - G_1(0)] + F_2(0)[G_2(0+) - G_2(0)] + \\ &+ \left| \int_{0+}^{\infty} F_1(xy) dG_1(y) - \int_{0+}^{\infty} F_2(xy) dG_2(y) \right| \equiv Q_1 + Q_2. \end{aligned}$$

Очевидно,

$$Q_1 = P(X_1 < 0)P(Y_1 = 0) + P(X_2 < 0)P(Y_2 = 0). \quad (7.8)$$

Оценим величину Q_2 следующим образом:

$$\begin{aligned} Q_2 &\leq \left| \int_{0+}^{\infty} [F_1(xy) - F_2(xy)] dG_1(y) \right| + \left| \int_{0+}^{\infty} F_2(xy) d[G_1(y) - G_2(y)] \right| \equiv \\ &\equiv Q_{21} + Q_{22}. \end{aligned}$$

Имеем

$$Q_{21} \leq \rho(X_1, X_2) \int_{0+}^{\infty} dG_1(y) = \rho(X_1, X_2)P(Y_1 > 0). \quad (7.9)$$

Если $x = 0$, то

$$Q_{22} = P(X_2 < 0) |P(Y_1 = 0) - P(Y_2 = 0)|. \quad (7.10)$$

Рассмотрим случай, когда $x \neq 0$. Интегрируя по частям, получаем

$$\begin{aligned} Q_{22} &= \left| F_2(xy) [G_1(y) - G_2(y)] \Big|_{y=0+}^{\infty} - \int_{0+}^{\infty} [G_1(y) - G_2(y)] d_y F_2(xy) \right| \leq \\ &\leq F_2(xy) [G_1(y) - G_2(y)] \Big|_{y=0+}^{\infty} + \left| \int_{0+}^{\infty} [G_1(y) - G_2(y)] d_y F_2(xy) \right| \equiv \\ &\equiv Q_{221} + Q_{222}. \end{aligned}$$

Очевидно,

$$Q_{221} \leq F_2(0+) |G_1(0+) - G_2(0+)| = P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)|. \quad (7.11)$$

В то же время

$$Q_{222} \leq \rho(Y_1, Y_2) \left| \int_{0+}^{\infty} d_y F_2(xy) \right| = \rho(Y_1, Y_2) I_2(x), \quad (7.12)$$

где

$$I_2(x) = \begin{cases} P(X_2 > 0), & \text{если } x > 0, \\ P(X_2 < 0), & \text{если } x < 0. \end{cases} \quad (7.13)$$

Объединяя (7.8), (7.9), (7.10), (7.11) и (7.12), мы приходим к следующему утверждению.

Лемма 7.3. При указанных выше условиях, налагаемых на случайные величины X_1 , X_2 , Y_1 и Y_2 , справедлива оценка

$$\begin{aligned} \Delta(x) &\leq P(X_1 < 0)P(Y_1 = 0) + P(X_2 < 0)P(Y_2 = 0) + \\ &+ \rho(X_1, X_2)P(Y_1 > 0) + P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)| + \\ &+ \rho(Y_1, Y_2) I_2(x), \end{aligned}$$

где функция $I_2(x)$ определена в (7.13).

Поскольку в приведенных рассуждениях случайные величины X_1 и X_2 играют абсолютно равносильные роли, из Леммы 7.3 можно получить следующее утверждение.

Следствие 7.1. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то

$$\Delta \leq \rho(X_1, X_2) + \rho(Y_1, Y_2) \min_{i=1,2} \max \{P(X_i < 0), P(X_i > 0)\}.$$

Следствие 7.2. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$ и, по крайней мере, одна из случайных величин X_i , $i = 1, 2$, удовлетворяет условию $P(X_i < 0) = P(X_i > 0)$, то

$$\Delta \leq \rho(X_1, X_2) + \frac{1}{2}\rho(Y_1, Y_2).$$

И конечно же, простейшим следствием Леммы 7.3 является

Следствие 7.3. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то

$$\Delta \leq \rho(X_1, X_2) + \rho(Y_1, Y_2).$$

Так как в случае $P(Y_1 = 0) = P(Y_2 = 0) = 0$ справедливы соотношения

$$\begin{aligned} \rho(Y_1^{-1}, Y_2^{-1}) &= \sup_{x \in \mathbb{R}} \left| P\left(\frac{1}{Y_1} < x\right) - P\left(\frac{1}{Y_2} < x\right) \right| = \\ &= \sup_{x \in \mathbb{R}} \left| P\left(Y_1 > \frac{1}{x}\right) - P\left(Y_2 > \frac{1}{x}\right) \right| = \\ &= \sup_{x \in \mathbb{R}} \left| \left[1 - P\left(Y_1 \leq \frac{1}{x}\right)\right] - \left[1 - P\left(Y_2 \leq \frac{1}{x}\right)\right] \right| = \\ &= \sup_{z \in \mathbb{R}} \left| P(Y_1 < z) - P(Y_2 < z) \right| = \rho(Y_1, Y_2), \end{aligned}$$

то из Леммы 7.3 также можно получить следующие оценки равномерного расстояния между произведениями независимых случайных величин.

Следствие 7.4. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2) \min_{i=1,2} \max \{P(X_i < 0), P(X_i > 0)\}.$$

Следствие 7.5. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$ и, по крайней мере, одна из случайных величин X_i , $i = 1, 2$, удовлетворяет условию $P(X_i < 0) = P(X_i > 0)$, то

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \frac{1}{2} \rho(Y_1, Y_2).$$

Следствие 7.6. Если в дополнение к условиям Леммы 7.3 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2).$$

Лемма 7.3 и Следствия 7.1–7.6 содержат простые оценки близости масштабных смесей в терминах расстояний между соответствующими смешиваемыми и смешивающими распределениями.

Основная идея регуляризации общей некорректно поставленной задачи разделения смеси (на практике не известной точно) нормальных законов основана на замене неизвестного “истинного” смешивающего распределения его конечной дискретной аппроксимацией.

Рассмотрим некоторые вопросы, связанные с оцениванием потери точности, сопутствующей такой подмене, на примере масштабных смесей нормальных законов.

Пусть Y_1 и Y_2 – некоторые положительные случайные величины. Из Следствия 7.5 вытекает, что в условиях Леммы 7.3 справедлива оценка

$$\sup_{x \in \mathbb{R}} |E\Phi(x/\sqrt{Y_1}) - E\Phi(x/\sqrt{Y_2})| \leq \frac{1}{2} \sup_{y>0} |P(Y_1 < y) - P(Y_2 < y)|.$$

Таким образом, если известна оценка равномерного расстояния между “истинным” смешивающим распределением и его конечной дискретной аппроксимацией, то оценка равномерного расстояния между соответствующими “истинной” и приближенной смесью оказывается вдвое меньшей.

Обратная задача

На самом деле намного интереснее найти решение задачи, которая является в некотором смысле обратной к только что рассмот-

ренной: по заданному расстоянию (например, равномерному) между смесями (которое может быть оценено статистически, например, с использованием модификации методов оценивания квантильных мер риска (VaR), основанных на статистике Колмогорова–Смирнова) найти оценку расстояния между “истинным” и статистически подогаанным смешивающим распределением.

Один из возможных подходов к решению этой задачи таков. Расстояние между смесями может быть охарактеризовано расстоянием между соответствующими характеристическими функциями. Характеристическая функция $f_j(t)$, соответствующая смешанной функции распределения $E\Phi(x/\sqrt{Y_j})$, имеет вид

$$f_j(t) = E \exp \left\{ -\frac{1}{2} t^2 Y_j \right\}, \quad t \in \mathbb{R},$$

$j = 1, 2$. В правой части этого соотношения стоит преобразование Лапласа–Стилтьеса $\psi_j(s) = E e^{-sY_j}$ случайной величины Y_j в точке $s = \frac{1}{2}t^2$, $j = 1, 2$. Как известно, преобразование Лапласа–Стилтьеса однозначно определяет распределение неотрицательной случайной величины. Таким образом, расстояние (невязка) между смесями связано с “расстоянием” между смешивающими распределениями соотношением

$$|f_1(t) - f_2(t)| = \left| \psi_1 \left(\frac{1}{2} t^2 \right) - \psi_2 \left(\frac{1}{2} t^2 \right) \right|, \quad t \in \mathbb{R}.$$

Более того, по формуле обращения преобразования Лапласа–Стилтьеса (см., например, (Феллер, 1984), том 2) имеем

$$P(Y_j < x) = \lim_{t \rightarrow \infty} \sum_{n \leq \sqrt{2tx}} \frac{(-1)^n (2t)^{n/2}}{n!} f_j^{(n)}(\sqrt{2t}), \quad j = 1, 2.$$

Поэтому

$$|P(Y_1 < x) - P(Y_2 < x)| \leq \lim_{t \rightarrow \infty} \sum_{n \leq \sqrt{2tx}} \frac{(2t)^{n/2}}{n!} \left| f_1^{(n)}(\sqrt{2t}) - f_2^{(n)}(\sqrt{2t}) \right|,$$

где, как и ранее, $f_j(t)$ – характеристическая функция, соответствующая смешанной функции распределения $E\Phi(x/\sqrt{Y_j})$, $j = 1, 2$.

Как видно из последнего соотношения, для получения приемлемой оценки равномерного расстояния между смешивающими распределениями при таком подходе нужна существенно более подробная информация о расстоянии между смесями, нежели значение равномерной метрики. Вообще говоря, равномерное расстояние – “неправильная” метрика для решения обсуждаемой “обратной” задачи во всей ее полноте. Действительно, рассмотрим следующий пример.

Пусть $\varepsilon > 0$ – произвольно малое положительное число, U_1 и U_2 – две “случайные” величины, имеющие вырожденные распределения, соответственно, в точках 1 и $1 + \varepsilon'$, где $\varepsilon' = \varepsilon\sqrt{2\pi e}$. Очевидно, что при этом по формуле Лагранжа

$$\begin{aligned} \left| E\Phi\left(\frac{x}{U_1}\right) - E\Phi\left(\frac{x}{U_2}\right) \right| &= \left| \Phi(x) - \Phi\left(\frac{x}{1 + \varepsilon'}\right) \right| = \\ &= \left| \frac{x\varepsilon'}{1 + \varepsilon'} \varphi\left(x\left[\delta + (1 - \delta)\frac{1}{1 + \varepsilon'}\right]\right) \right| \end{aligned}$$

для некоторого $\delta \in [0, 1]$, где $\varphi(x)$ – стандартная нормальная плотность.

Легко видеть, что

$$\delta + (1 - \delta)\frac{1}{1 + \varepsilon'} = \frac{\delta\varepsilon' + 1}{1 + \varepsilon'} \geq \frac{1}{1 + \varepsilon'}.$$

Поэтому

$$\left| E\Phi\left(\frac{x}{U_1}\right) - E\Phi\left(\frac{x}{U_2}\right) \right| \leq \varepsilon' \sup_{x \in \mathbb{R}} |x| \varphi(x) = \frac{\varepsilon'}{\sqrt{2\pi e}} = \varepsilon,$$

в то время как

$$\rho(U_1, U_2) = 1.$$

По-видимому, для решения “обратной” задачи в общем случае следует использовать метрики (расстояния), метризирующие слабую сходимость, например, метрику Леви или сглаженное равномерное расстояние.

Дадим определение метрики Леви. Пусть X и Y – две случайные величины с функциями распределения, соответственно, $F(x)$

и $G(x)$, $x \in \mathbb{R}$. Расстояние (метрика) Леви $L(X, Y)$ между случайными величинами X и Y определяется как

$$L(X, Y) = \inf\{h : G(x-h) - h \leq F(x) \leq G(x+h) + h \text{ для всех } x \in \mathbb{R}\}.$$

Расстояние Леви допускает наглядную геометрическую интерпретацию: оно равно длине стороны наибольшего квадрата, который можно вписать между графиками функций распределения $F(x)$ и $G(x)$ так, чтобы его стороны были параллельны координатным осям. Так как в определении метрики Леви задействованы не сами случайные величины X и Y , а только их функции распределения $F(x)$ и $G(x)$, то выражения $L(X, Y)$ и $L(F, G)$ не будут различаться:

$$L(X, Y) \equiv L(F, G).$$

Для равномерного расстояния между функциями распределения $F(x)$ и $G(x)$, как и ранее, будем использовать обозначение $\rho(X, Y) \equiv \rho(F, G)$,

$$\rho(X, Y) \equiv \rho(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|.$$

По крайней мере, если рассматривается задача оценки устойчивости идентифицируемости *конечных* смесей нормальных законов, то в терминах метрики Леви можно получить вполне наглядные результаты.

Проиллюстрируем сказанное на примере обратной задачи оценивания устойчивости нормального закона по отношению к так называемому “загрязнению” (контаминации). А именно, рассмотрим схему Дж. Тьюки, в рамках которой вместо “чистого” нормального распределения $\Phi(x)$ имеется масштабная смесь нормальных законов

$$F_{p,\sigma}(x) = p\Phi(x) + (1-p)\Phi(\sigma x), \quad x \in \mathbb{R},$$

где $p \in (0, 1]$, $\sigma \in (0, 1]$ (Tukey, 1960). Если выборка представляет собой набор независимых реализаций случайной величины с функцией распределения $F_{p,\sigma}(x)$, то компонента $\Phi(\sigma x)$ смеси

$F_{p,\sigma}(x)$ соответствует “загрязняющим” (нетипичным) наблюдениям, искажающим статистические выводы, полученные в предположении “чистой” нормальности распределения выборки.

Несложно видеть, что смесь $F_{p,\sigma}(x)$ – это функция распределения произведения стандартно нормально распределенной случайной величины Z и независимой от нее случайной величины $W_{p,\sigma}$, принимающей значения 1 и σ^{-1} с вероятностями p и $(1-p)$, соответственно. Обозначим $U_{p,\sigma} = W_{p,\sigma}^{-1}$.

Обратная задача оценивания устойчивости нормального закона по отношению к “загрязнению” может быть сформулирована следующим образом. Пусть $\varepsilon > 0$ – произвольно малое число. Предположим, что

$$L(F_{p,\sigma}, \Phi) \leq \varepsilon. \quad (7.14)$$

Что можно сказать о том, насколько близкими к единице в таком случае должны быть параметры p и σ ? Другими словами, насколько близким к вырожденному в единице должно быть распределение случайной величины $U_{p,\sigma}$?

Следует отметить, что “прямая” задача оценивания устойчивости нормального закона по отношению к “загрязнению” в схеме Тьюки является частным случаем задачи оценивания расстояния между масштабной смесью и порождающим ее смешиваемым законом. Пусть U – неотрицательная случайная величина. Пусть $F(x)$ – некоторая функция распределения. Задача оценивания расстояния между масштабной смесью и порождающим ее смешиваемым законом заключается в оценивании расстояния между функциями распределения $F(x)$ и $EF(U^{-1}x)$ в терминах расстояния между распределением случайной величины U и распределением, вырожденным в единице. Из результатов на эту тему для $F = \Phi$ можно выделить следующие. Обозначим $F_U(x) = E\Phi(xU^{-1})$.

Теорема 7.5. Пусть $E \max\{U, U^{-1}\} < \infty$. Тогда

$$\rho(\Phi, F_U) \leq 0.242 E(\max\{U, U^{-1}\} - 1).$$

Доказательство. Для некоторого $\delta = \delta(U) \in [0, 1]$ по формуле Лагранжа имеем

$$\Phi(x) - E\Phi(xU^{-1}) = E[\Phi(x) - \Phi(xU^{-1})] =$$

$$\begin{aligned}
&= E[x(1 - U^{-1})\varphi(x(\delta + (1 - \delta)U^{-1}))] \leq \\
&\leq E[|1 - U^{-1}| \sup_{x \geq 0} x\varphi(x(\delta + (1 - \delta)U^{-1}))] = \\
&= \frac{1}{\sqrt{2\pi e}} E[|1 - U^{-1}| U(U\delta + 1 - \delta)^{-1}] \leq \frac{1}{\sqrt{2\pi e}} E(\max\{U, U^{-1}\} - 1),
\end{aligned}$$

что и требовалось доказать.

Теорема 7.6. Пусть $EU^{-2} < \infty$. Тогда

$$\rho(\Phi, F_U) \leq E|U^{-2} - 1|.$$

Доказательство см. в работе (Shaked, 1981).

То обстоятельство, что расстояние между представителями семейства масштабных смесей нормальных законов можно естественно охарактеризовать с помощью моментов смешивающих величин, по-видимому, впервые отмечено в статье (Keilson and Steutel, 1974). Пусть X – случайная величина со стандартной нормальной функцией распределения Φ . Тогда функция распределения $E\Phi(U^{-1}x)$ соответствует случайной величине $Z = X \cdot U$, где сомножители независимы. В работе (Keilson and Steutel, 1974), в частности, замечено, что близость представителя подмножества указанного семейства с $EU^4 < \infty$ к “чистому” нормальному закону удобно характеризовать с помощью величины

$$\gamma(Z) = \varkappa(Z) - 3,$$

где

$$\varkappa(Z) = E\left(\frac{Z - EZ}{\sqrt{DZ}}\right)^4$$

– коэффициент эксцесса случайной величины Z , $\varkappa(X) = 3$. Справедлив следующий результат.

Теорема 7.7. Пусть $EU^2 = 1$ и $EU^4 < \infty$. Тогда

$$\rho(\Phi, F_U) \leq 0.648 \gamma(XU).$$

Доказательство см. в работе (Heyde and Leslie, 1976). Константа уточнена в статье (Hall, 1979).

Замечание 7.1. Легко убедиться, что из условия $EU^2 = 1$ вытекают соотношения

$$\gamma(XU) = 3E(U^4 - 1) = 3E(U^2 - 1)^2 = 3DU^2.$$

Поэтому в условиях Теоремы 7.7 справедливы оценки

$$\rho(\Phi, F_U) \leq 1.944E(U^4 - 1) = 1.944E(U^2 - 1)^2 = 1.944DU^2.$$

Для масштабных смесей, в которых случайные величины U имеют моменты меньших порядков, справедлив следующий результат. Напомним, что символом $L(\cdot, \cdot)$ обозначается метрика Леви.

Теорема 7.8. Пусть $EU^\delta < \infty$ для некоторого $\delta > 0$. Тогда

$$L(\Phi, F_U) \leq \left(2 + \frac{2}{\delta\sqrt{2\pi e}}\right) L(U^\delta, 1).$$

Доказательство см. в работе (Hall, 1979).

Вернемся к обратной задаче оценивания устойчивости нормального закона по отношению к “загрязнению”. Справедлив следующий результат.

Теорема 7.9. Пусть $p \in (0, 1]$, $\sigma \in (0, 1]$ и для некоторого $\varepsilon > 0$ выполнено неравенство (7.14). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma}, 1) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять

$$C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405.$$

Доказательство. Справедливо неравенство

$$\rho(\Phi, F_{p,\sigma}) \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) L(\Phi, F_{p,\sigma})$$

(см., например, (Золотарев, 1986), с. 107). Поэтому из неравенства (7.14) вытекает, что

$$\rho(\Phi, F_{p,\sigma}) \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \varepsilon.$$

Несложно видеть, что

$$\rho(\Phi, F_{p,\sigma}) = (1-p) \sup_{x \in \mathbb{R}} |\Phi(\sigma x) - \Phi(x)| \geq (1-p) |\Phi(\sigma) - \Phi(1)|.$$

Поэтому, если выполнено неравенство (7.14), то

$$(1-p) |\Phi(\sigma) - \Phi(1)| \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \varepsilon. \quad (7.15)$$

Вместе с тем по формуле Лагранжа в силу того, что $0 < \sigma \leq 1$, имеем

$$|\Phi(\sigma) - \Phi(1)| = (1-\sigma) \varphi(\sigma') \geq (1-\sigma) \varphi(1) = \frac{1-\sigma}{\sqrt{2\pi e}}, \quad (7.16)$$

где $\varphi(x)$ – стандартная нормальная плотность, $\sigma' \in [\sigma, 1]$. Из (7.15) и (7.16) получаем, что, если выполнено неравенство (7.14), то

$$(1-p)(1-\sigma) \leq \sqrt{2\pi e} \cdot \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \varepsilon = \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon < 5.7815 \varepsilon. \quad (7.17)$$

Таким образом, если “загрязненная” смесь $F_{p,\sigma}(x)$ близка к “чистому” нормальному закону в смысле метрики Леви или равномерного расстояния, то параметры p и σ *должны* быть близки к единице в том смысле, что должно быть выполнено соотношение (7.17).

Но с другой стороны, используя геометрическую интерпретацию расстояния Леви, несложно убедиться, что

$$L(U_{p,\sigma}, 1) = \min\{1-p, 1-\sigma\}.$$

Следовательно, из (7.17) окончательно получаем, что

$$L^2(U_{p,\sigma}, 1) = [\min\{1-p, 1-\sigma\}]^2 \leq (1-p)(1-\sigma) \leq$$

$$\leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon < 5.7815\varepsilon.$$

Теорема доказана.

Попутно была получена следующая характеристика нормального распределения в классе смесей типа $F_{p,\sigma}$. Пусть $\varepsilon \rightarrow 0$. Пусть одновременно параметры $p = p(\varepsilon)$ и $\sigma = \sigma(\varepsilon)$ изменяются так, что неравенство (7.14) остается справедливым. При этом из Теоремы 7.9 вытекает, что это возможно, только если выполнено условие

$$L(U_{p,\sigma}, 1) \rightarrow 0.$$

Но это означает, что смесь $F_{p,\sigma}$ может быть «чистым» нормальным законом тогда и только тогда, когда $p = 1$ или $\sigma = 1$.

В силу имеющих место в предположениях Теоремы 7.9 неравенств

$$\sigma L(U_{p,\sigma}^{-1}, 1) \leq L(U_{p,\sigma}, 1) \leq L(U_{p,\sigma}^{-1}, 1)$$

из Теорем 7.8 и 7.9 вытекает следующее утверждение.

Следствие 7.7. Пусть $p \in (0, 1]$ и $\sigma \in (0, 1]$. Тогда

$$\frac{1}{\sqrt{e}(1 + \sqrt{2\pi})} L^2(U_{p,\sigma}, 1) \leq L(F_{p,\sigma}, \Phi) \leq \frac{\sqrt{2}(1 + \sqrt{2\pi e})}{\sigma \sqrt{\pi e}} L(U_{p,\sigma}, 1).$$

Аналогичную задачу можно рассмотреть и для сдвиговых смесей нормальных законов вида

$$G_{p,a}(x) = p\Phi(x) + (1 - p)\Phi(x - a), \quad a \in \mathbb{R}, x \in \mathbb{R}.$$

Несложно заметить, что смесь $G_{p,a}(x)$ – это функция распределения суммы стандартно нормально распределенной случайной величины Z и независимой от нее случайной величины $V_{p,a}$, принимающей значения 0 и a с вероятностями p и $(1 - p)$, соответственно.

Пусть $|a| \leq 1$. Так как

$$\rho(G_{p,a}, \Phi) = (1 - p) \sup_{x \in \mathbb{R}} |\Phi(x - a) - \Phi(x)| \geq (1 - p) \cdot |\Phi(a) - \Phi(0)| =$$

$$= (1-p)|a|\varphi(a') \geq (1-p)|a|\varphi(1) \geq \frac{(1-p)|a|}{\sqrt{2\pi e}},$$

где $a' \in [0, a]$, то, используя ту же самую логику доказательства, что и в Теореме 7.9, можно доказать следующее утверждение.

Теорема 7.10. Пусть $p \in (0, 1]$, $|a| \leq 1$ и для некоторого $\varepsilon > 0$ выполнено неравенство

$$L(G_{p,a}, \Phi) \leq \varepsilon.$$

Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a}, 0) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

При этом смесь $G_{p,a}$ может быть «чистым» нормальным законом тогда и только тогда, когда $p = 1$ или $a = 0$.

Теоремы 7.9 и 7.10 допускают простое обобщение. Рассмотрим две смеси:

$$F_{p,\sigma}(x) = p\Phi(x) + (1-p)\Phi(\sigma x)$$

и

$$F_{q,\sigma}(x) = q\Phi(x) + (1-q)\Phi(\sigma x),$$

где $p \in [0, 1]$, $q \in [0, 1]$, $\sigma \in (0, 1]$. Так как

$$\begin{aligned} \rho(F_{p,\sigma}, F_{q,\sigma}) &= |p - q| \sup_{x \in \mathbb{R}} |\Phi(x) - \Phi(\sigma x)| \geq \\ &\geq |p - q| \cdot |\Phi(\sigma) - \Phi(1)| \geq \frac{|p - q| \cdot |1 - \sigma|}{\sqrt{2\pi e}}, \end{aligned}$$

то, используя ту же логику доказательства, что и в Теореме 7.9, можно доказать, что, если

$$L(F_{p,\sigma}, F_{q,\sigma}) \leq \varepsilon, \tag{7.18}$$

то

$$|p - q| \cdot |1 - \sigma| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon,$$

и справедливо следующее утверждение.

Теорема 7.11. Пусть $p \in [0, 1]$, $q \in [0, 1]$, $\sigma \in (0, 1]$ и для некоторого $\varepsilon > 0$ выполнено неравенство (7.18). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma}, U_{q,\sigma}) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Аналогично, если рассмотреть смеси

$$G_{p,a}(x) = p\Phi(x) + (1-p)\Phi(x-a)$$

и

$$G_{q,a}(x) = q\Phi(x) + (1-q)\Phi(x-a),$$

где $p \in [0, 1]$, $q \in [0, 1]$, $|a| \leq 1$, то, используя ту же логику доказательства, что и в Теореме 7.10, можно доказать, что, если

$$L(G_{p,a}, G_{q,a}) \leq \varepsilon, \quad (7.19)$$

то

$$|a| \cdot |p - q| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon,$$

и справедливо следующее утверждение.

Теорема 7.12. Пусть $p \in [0, 1]$, $q \in [0, 1]$, $|a| \leq 1$ и для некоторого $\varepsilon > 0$ выполнено неравенство (7.19). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a}, V_{q,a}) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Пусть теперь $\sigma_1 \in (0, 1]$ и $\sigma_2 \in (0, 1]$. Так как

$$\begin{aligned} \rho(F_{p,\sigma_1}, F_{p,\sigma_2}) &= (1-p) \sup_{x \in \mathbb{R}} |\Phi(\sigma_1 x) - \Phi(\sigma_2 x)| \geq \\ &\geq (1-p) \cdot |\Phi(\sigma_1) - \Phi(\sigma_2)| \geq \frac{(1-p) \cdot |\sigma_1 - \sigma_2|}{\sqrt{2\pi e}}, \end{aligned}$$

то, используя ту же логику доказательства, что и в теореме 7.9, можно доказать, что, если

$$L(F_{p,\sigma_1}, F_{p,\sigma_2}) \leq \varepsilon, \quad (7.20)$$

то

$$(1-p) \cdot |\sigma_1 - \sigma_2| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon,$$

и справедливо следующее утверждение.

Теорема 7.13. Пусть $p \in [0, 1]$, $\sigma_1 \in (0, 1]$ и $\sigma_2 \in (0, 1]$ и для некоторого $\varepsilon > 0$ выполнено неравенство (7.20). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma_1}, U_{p,\sigma_2}) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Наконец, пусть теперь $|a_1| \leq 1$ и $|a_2| \leq 1$. Так как

$$\begin{aligned} \rho(G_{p,a_1}, G_{p,a_2}) &= (1-p) \sup_{x \in \mathbb{R}} |\Phi(x-a_1) - \Phi(x-a_2)| \geq \\ &\geq (1-p) \cdot |\Phi(a_1) - \Phi(a_2)| \geq \frac{(1-p) \cdot |a_1 - a_2|}{\sqrt{2\pi e}}, \end{aligned}$$

то, используя ту же логику доказательства, что и в теореме 7.10, можно доказать, что, если

$$L(G_{p,a_1}, G_{p,a_2}) \leq \varepsilon, \quad (7.21)$$

то

$$(1-p) \cdot |a_1 - a_2| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \varepsilon,$$

и справедливо следующее утверждение.

Теорема 7.14. Пусть $p \in [0, 1]$, $|a_1| \leq 1$, $|a_2| \leq 1$ и для некоторого $\varepsilon > 0$ выполнено неравенство (7.21). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a_1}, V_{p,a_2}) \leq C\sqrt{\varepsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Указанные обобщения играют ключевую роль при обосновании возможности приближенного разделения смесей с помощью сеточных методов (см. (Королев, 2011)).

7.5 Дисперсионно-сдвиговые смеси нормальных законов

Во многих практических задачах в качестве базовых рассматриваются вероятностные модели типа дисперсионно-сдвиговых смесей нормальных законов. Подобные модели уже хорошо себя зарекомендовали во многих исследованиях, где они продемонстрировали очень высокую адекватность. Последнее обстоятельство можно легко объяснить довольно большим числом настраиваемых параметров в указанных моделях. Однако на самом деле их адекватность имеет гораздо более глубокие теоретические обоснования. А именно, дисперсионно-сдвиговые смеси нормальных законов являются предельными законами в довольно простых предельных теоремах для случайно остановленных случайных блужданий. Такие теоремы позволяют однозначно связать конкретный смешивающий закон в дисперсионно-сдвиговых смесях с поведением интенсивности потока информативных событий, в результате которых накапливаются данные, характеризующие анализируемый случайный процесс. Тем самым, эти теоремы как бы позволяют разделить вклады внешних и внутренних факторов в случайность поведения анализируемого процесса.

Для оценивания параметров упомянутых выше моделей недавно разработан, реализован и теоретически и экспериментально исследован новый комбинированный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов, превосходящий известные методы по точности и эффективности. Последнее обстоятельство особенно важно при применении этого метода для анализа реальных процессов в режиме реального времени. На первом этапе данного метода исследуемая непрерывная дисперсионно-сдвиговая смесь нормальных законов приближается конечной смесью нормальных законов с, возможно, очень большим числом известных узлов. Для обоснования сходимости этого метода необходимо иметь оценки погрешности при подмене непрерывного смешивающего распределения его дискретным аналогом. Эта задача обсуждается ниже.

Понятие дисперсионно-сдвиговой смеси нормальных законов (normal variance-mean mixture) введено в 70–80-х годах прошлого столетия в работах О.-Е. Барндорфф-Нильсена и его коллег (Barndorff-Nielsen, 1977), (Barndorff-Nielsen, 1978), (Barndorff-Nielsen, Kent, Sørensen, 1982) как довольно гибкое обобщение нормального распределения.

Пусть $\beta \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $0 < \sigma < \infty$, $G(x)$ – функция распределения, все точки роста которой сосредоточены на $\mathbb{R}_+$. Дисперсионно-сдвиговой смесью нормальных законов называется функция распределения

$$F(x) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sigma \sqrt{z}}\right) dG(z), \quad x \in \mathbb{R}. \quad (7.22)$$

Обратим внимание, что в соотношении (7.22) смешивание происходит одновременно и по параметру сдвига, и по параметру масштаба, но так как эти параметры в (7.22) связаны жесткой зависимостью, при которой параметры положения (*сдвига*) смешиваемых нормальных законов пропорциональны их *дисперсиям* то фактически смесь (7.22) является однопараметрической. Именно поэтому смеси вида (7.22) называются *дисперсионно-сдвиговыми*.

Если Y и U – независимые случайные величины с функциями распределения $\Phi(x)$ и $G(x)$ (при этом, очевидно, $P(U \geq 0) = 1$), а Z – случайная величина, удовлетворяющая соотношению

$$Z \stackrel{d}{=} \sigma Y \sqrt{U} + \alpha U + \beta,$$

то $P(Z < x) = F(x)$, $x \in \mathbb{R}$.

Легко убедиться, что если $EU < \infty$, то

$$EZ = \beta + \alpha EU, \quad (7.23)$$

а если при этом и $EU^2 < \infty$, то

$$EZ^2 = \beta^2 + (\sigma^2 + 2\alpha\beta)EU + \alpha^2 EU^2, \quad DZ = \alpha^2 DU + \sigma^2 EU.$$

При вычислении моментов более высоких порядков можно использовать следующее соотношение между характеристическими функциями f_Z и f_U случайных величин Z и U :

$$f_Z(t) = e^{it\beta} f_U\left(\alpha t + \frac{1}{2}i\sigma^2 t^2\right), \quad t \in \mathbb{R}.$$

Чтобы убедиться в справедливости последнего соотношения, достаточно заметить, что по теореме Фубини

$$\begin{aligned} f_Z(t) &= \mathbb{E}e^{itZ} = \mathbb{E}\exp\{it(\beta + \sigma\sqrt{U} \cdot Y + \alpha U)\} = \\ &= \int_0^\infty \mathbb{E}\exp\{it(\beta + \sigma\sqrt{u} \cdot Y + \alpha u)\} dG(u) = \\ &= e^{it\beta} \int_0^\infty \exp\left\{it\alpha u - \frac{1}{2}\sigma^2 t^2 u\right\} dG(u) = \\ &= e^{it\beta} \int_0^\infty \exp\left\{iu\left(\alpha t + \frac{1}{2}i\sigma^2 t^2\right)\right\} dG(u) = e^{it\beta} f_U\left(\alpha t + \frac{1}{2}i\sigma^2 t^2\right), \quad t \in \mathbb{R}. \end{aligned}$$

В качестве примеров дисперсионно-сдвиговых смесей нормальных распределений рассмотрим два семейства: обобщенные гиперболические распределения и обобщенные дисперсионные гамма-распределения.

7.6 Обобщенные гиперболические распределения

Плотность обобщенного обратного гауссовского распределения обозначим $p_{GIG}(x; \nu, \mu, \lambda)$,

$$p_{GIG}(x; \nu, \mu, \lambda) = \frac{\lambda^{\nu/2}}{2\mu^{\nu/2}K_\nu(\sqrt{\mu\lambda})} x^{\nu-1} \exp\left\{-\frac{1}{2}\left(\frac{\mu}{x} + \lambda x\right)\right\}, \quad x > 0.$$

Здесь $\nu \in \mathbb{R}$,

$$\mu > 0, \quad \lambda \geq 0, \quad \text{если } \nu < 0,$$

$$\mu > 0, \quad \lambda > 0, \quad \text{если } \nu = 0,$$

$$\mu \geq 0, \quad \lambda > 0, \quad \text{если } \nu > 0,$$

$K_\nu(z)$ – модифицированная бесселева функция третьего рода порядка ν ,

$$K_\nu(z) = \frac{1}{2} \int_0^\infty y^{\nu-1} \exp \left\{ -\frac{z}{2} \left(y + \frac{1}{y} \right) \right\} dy, \quad z \in \mathbb{C}, \operatorname{Re} z > 0.$$

Соответствующую функцию распределения обозначим $P_{GIG}(x; \nu, \mu, \lambda)$,

$$P_{GIG}(x; \nu, \mu, \lambda) = \begin{cases} 0, & x < 0, \\ \int_0^x p_{GIG}(z; \nu, \mu, \lambda) dz, & x \geq 0. \end{cases}$$

Как сказано в статье (Seshadri, 1997), обобщенное обратное гауссовское распределение введено в 1946 г. Этьеном Альфеном (Étienne Halphen), который использовал его для описания объема воды, проходящего ежемесячно через гидроэлектростанции. В (Seshadri, 1997) обобщенное обратное гауссовское распределение названо *распределением Альфена*. В 1973 г. это распределение было переоткрыто Гербертом Зихелем, который использовал его в качестве смешивающего закона при рассмотрении специальных смешанных пуассоновских распределений (*распределений Зихеля*) как дискретных распределений с тяжелыми хвостами. В 1977 г. эти распределения снова переоткрыл О.-Э. Барндорфф-Нильсен, который, в частности, использовал их для описания распределения размеров частиц.

Класс обобщенных обратных гауссовских распределений довольно обширен и содержит, в частности, как распределения с экспоненциально убывающими хвостами (гамма-распределение ($\mu = 0$, $\nu > 0$)), так и распределения с хвостами, убывающими степенным образом (обратное гамма-распределение ($\lambda = 0$, $\nu < 0$), обратное гауссовское распределение ($\nu = -\frac{1}{2}$) и его предельный при $\lambda \rightarrow 0$ случай – распределение Леви).

В 1977–78 гг. О.-Э. Барндорфф-Нильсен ввел класс *обобщенных гиперболических распределений* как класс специальных

дисперсионно-сдвиговых смесей нормальных законов. А именно, пусть $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$. Если функцию обобщенного гиперболического распределения с параметрами α , β , ν , μ , λ обозначить $P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda)$, то по определению

$$P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sqrt{z}}\right) p_{GIG}(z; \nu, \mu, \lambda) dz, \quad x \in \mathbb{R}. \quad (7.24)$$

Обратим внимание, что в (7.24) смешивание происходит одновременно и по параметру сдвига, и по параметру масштаба, но так как эти параметры в (7.24) связаны жесткой зависимостью, то фактически смесь (7.24) является *однопараметрической*.

Несложно убедиться, что плотность $p_{GH}(x; \alpha, \beta, \nu, \mu, \lambda)$ обобщенного гиперболического распределения имеет вид

$$\begin{aligned} p_{GH}(x; \alpha, \beta, \nu, \mu, \lambda) &= \int_0^{\infty} \frac{1}{\sqrt{z}} \varphi\left(\frac{x - \beta - \alpha z}{\sqrt{z}}\right) p_{GIG}(z; \nu, \mu, \lambda) dz = \\ &= \frac{\lambda^{\nu/2} (\lambda + \alpha^2)^{\nu/2 - 1/4}}{2\sqrt{2\pi} \mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} \cdot (\mu + (x - \beta)^2)^{1/4 - \nu/2} \times \\ &\times \exp\{\alpha(x - \beta)\} K_{\nu - 1/2}\left(\sqrt{(\mu + (x - \beta)^2)(\lambda + \alpha^2)}\right). \end{aligned}$$

Класс обобщенных гиперболических распределений очень широк и содержит, в частности, (a) симметричные и скошенные (skew) распределения Стьюдента (в том числе распределение Коши), которым в представлении (7.24) соответствуют смешивающие обратные гамма-распределения; (b) дисперсионные гамма-распределения (Variance Gamma (VG)-distributions) (в том числе симметричные и несимметричные распределения Лапласа), которым в представлении (7.24) соответствуют смешивающие гамма-распределения; (c) нормальные\обратные нормальные (NIG) распределения, которым в представлении (7.24) соответствуют смешивающие обратные нормальные распределения; и многие другие типы.

Обобщенные гиперболические распределения продемонстрировали высочайшую адекватность при их применении для описания статистических закономерностей поведения различных характеристик сложных открытых систем, в частности, турбулентных систем и финансовых рынков. Публикации, посвященные моделям, основанным на обобщенных гиперболических распределениях, исчисляются сотнями. Согласно расхожему мнению, столь высокая адекватность обобщенных гиперболических моделей может быть формально объяснена большим числом настраиваемых параметров, позволяющим подогнать какую угодно модель к каким угодно данным. Среди статистиков хорошо известно высказывание Ж. Бертрана «Give me four parameters and I shall describe an elephant; with five, it will wave its trunk» (цитируется по статье Л. ЛеКама [67]). Это обстоятельство, конечно же, играет свою роль, однако на самом деле модели типа (7.24) в большинстве случаев адекватны по гораздо более естественным глубоким причинам.

В прикладной теории вероятностей ту или иную модель можно считать в достаточной мере обоснованной (адекватной) только тогда, когда она является *асимптотической аппроксимацией*, то есть когда существует довольно простая предельная схема (например, схема суммирования) и соответствующая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения (Гнеденко, Колмогоров, 1949). В первоисточниках упомянутые выше обобщенные гиперболические модели вводились чисто умозрительно как распределения процесса броуновского движения со случайным временем, в каждый момент имеющим то или иное обобщенное обратное гауссовское распределение. Лишь в 1982 г. Барндорфф-Нильсен со ссылкой на работу А. Реньи (Renyi, 1960) дал довольно расплывчатый намек на то, что смеси нормальных законов могут быть предельными для сумм случайного числа случайных величин.

Однако, как ни удивительно, несмотря на то, что свойства обобщенных гиперболических распределений изучены довольно полно, до недавнего времени не было дано корректного доказательства того факта, что обобщенные гиперболические распреде-

ления выступают в качестве предельных в простейшей схеме случайно остановленных случайных блужданий, и, стало быть, приводимая в некоторых работах аргументация, связывающая смешивание в модели (7.24) со случайным характером волатильности при применении обобщенных гиперболических распределений в финансовой математике, не имела строгого формального обоснования. Возможно, причина этого лежит в том, что в схеме «нарастающих» сумм, рассматривавшейся А. Реньи, полное решение указанной задачи невозможно. Его можно получить лишь, рассматривая случайные суммы в рамках асимптотической схемы серий. Основопологающей работой в этом направлении стала работа Б. В. Гнеденко и Х. Фахима 1969 г.

«Асимптотическое» обоснование некоторых из упомянутых выше моделей было дано лишь недавно в статьях (Королев, Соколов, 2012), (Королев, 2012), где показано, что скошенные распределения Стюдента и дисперсионные гамма-распределения могут выступать в качестве предельных в довольно простых предельных теоремах для регулярных статистик, построенных по выборкам случайного объема, в частности, в схеме случайного суммирования случайных величин и, следовательно, могут считаться *естественными* асимптотическими аппроксимациями для распределений многих процессов, например, сходных с неоднородными случайными блужданиями.

Мы приведем общую теорему о необходимых и достаточных условиях сходимости распределений сумм случайного числа независимых одинаково распределенных случайных величин к однопараметрическим сдвиг/масштабным смесям нормальных законов, и в качестве следствия из нее будут получены необходимые и достаточные условия сходимости распределений случайных сумм независимых одинаково распределенных случайных величин к обобщенным гиперболическим распределениям.

В дальнейшем для простоты без существенного ограничения общности мы будем полагать, что параметр сдвига β рассматриваемых обобщенных гиперболических распределений равен нулю.

Будем считать, что все случайные величины, о которых пойдет

речь ниже, заданы на одном вероятностном пространстве. Пусть $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$ – семейство последовательностей (серий) одинаково распределенных в каждой последовательности (при каждом фиксированном n) случайных величин. Пусть $\{N_n\}_{n \geq 1}$ – последовательность целочисленных неотрицательных случайных величин таких, что при каждом $n \geq 1$ случайные величины $N_n, X_{n,1}, X_{n,2}, \dots$ независимы. Положим

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}$$

(во избежание недоразумений полагаем $\sum_{j=1}^0 = 0$).

Каждой паре функций распределения (F, H) поставим в соответствие множество $\mathcal{M}(F|H)$, содержащее все функции распределения $Q(x)$ с $Q(0) = 0$, обеспечивающие представление характеристической функции, соответствующей функции распределения F , в виде степенной смеси характеристической функции, соответствующей функции распределения H :

$$\int_{-\infty}^{\infty} e^{itx} dF(x) = \int_0^{\infty} h^x(t) dQ(x), \quad t \in \mathbb{R},$$

где

$$h(t) = \int_{-\infty}^{\infty} e^{itx} dH(x), \quad t \in \mathbb{R}.$$

Следующее утверждение обобщает теорему переноса Гнеденко–Фохима.

Лемма 7.4. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и функция распределения $H(x)$ такие, что при $n \rightarrow \infty$*

$$P(S_{n,k_n} < x) \implies H(x).$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость

$$P(S_{n,N_n} < x) \implies F(x),$$

распределений случайных сумм к некоторой функции распределения $F(x)$, необходимо и достаточно, чтобы существовала слабо компактная последовательность функций распределения $\{Q_n^*(x)\}_{n \geq 1}$ такая, что выполняются условия

$$(i) \quad Q_n^*(x) \in \mathcal{M}(F|H), \quad n = 1, 2, \dots,$$

$$(ii) \quad L(Q_n^*, Q_n) \rightarrow 0, \quad (n \rightarrow \infty)$$

где $Q_n(x) = P(N_n < k_n x)$, $x \in \mathbb{R}$.

Данное утверждение является частным случаем теоремы 4.2.1 в книге (Gnedenko, Korolev, 1996).

Теорема 7.15. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что при $n \rightarrow \infty$*

$$P(S_{n,k_n} < x) \implies \Phi(x - \alpha). \quad (7.25)$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к некоторой функции распределения $F(x)$:

$$P(S_{n,N_n} < x) \implies F(x),$$

необходимо и достаточно, чтобы существовала функция распределения $Q(x)$ такая, что $Q(0) = 0$,

$$F(x) = \int_0^\infty \Phi\left(\frac{x - \alpha z}{\sqrt{z}}\right) dQ(z), \quad (7.26)$$

и

$$P(N_n < x k_n) \implies Q(x). \quad (7.27)$$

Доказательство. Зафиксируем вещественное число α и положим

$$H_\alpha(x; y) = \Phi\left(\frac{x - \alpha y}{\sqrt{y}}\right). \quad (7.28)$$

Убедимся, что семейство так определенных функций распределения аддитивно замкнуто по $y > 0$. Действительно, функции распределения (7.28) соответствует характеристическая функция

$$\int_{-\infty}^{\infty} e^{itx} d_x H_{\alpha}(x; y) = \exp\{y(it\alpha - t^2/2)\}.$$

Следовательно, свертке функций распределения $H_{\alpha}(x; y_1)$ и $H_{\alpha}(x; y_2)$ соответствует характеристическая функция

$$\exp\{y_1(it\alpha - t^2/2)\} \cdot \exp\{y_2(it\alpha - t^2/2)\} =$$

$$\exp\{(y_1 + y_2)(it\alpha - t^2/2)\} = \int_{-\infty}^{\infty} e^{itx} d_x H_{\alpha}(x; y_1 + y_2),$$

что и означает аддитивную замкнутость по $y > 0$ семейства функций распределения (7.28).

Таким образом, согласно Теореме 7.1 семейство однопараметрических сдвиг/масштабных смесей нормальных законов

$$\mathcal{F}_{\alpha} = \left\{ \int_0^{\infty} H_{\alpha}(x; y) dQ(y), x \in \mathbb{R} : Q \in \mathcal{Q} \right\},$$

где $\mathcal{Q}$ – множество всех функций распределения $Q(y)$ с $Q(0) = 0$, идентифицируемо.

В свою очередь, это означает, что для любой функции распределения $F(x)$ множество $\mathcal{M}(F|H_{\alpha})$ содержит не более одного элемента. Действительно, обозначим $h_{\alpha}(t) = \exp\{it\alpha - t^2/2\}$ (характеристическая функция $h_{\alpha}(t)$ соответствует функции распределения $\Phi(x - \alpha)$). Тогда $h_{\alpha}^y(t) = \exp\{y(it\alpha - t^2/2)\}$, и, стало быть, характеристические функции

$$\int_0^{\infty} h_{\alpha}^y(t) dQ(y),$$

фигурирующие в определении множества $\mathcal{M}(F|H_\alpha)$, соответствуют функциям распределения

$$\int_0^\infty H_\alpha(x; y) dQ(y),$$

составляющим семейство $\mathcal{F}_\alpha$. При этом условие (7.26) означает, что $Q \in \mathcal{M}(F(x)|\Phi(x - \alpha))$. Теперь остается сослаться на Лемму 7.4, в которой роль условий (i) и (ii) играют соответственно (7.26) и (7.27). Теорема доказана.

Замечание 7.2. Условие (7.25) выполняется в следующей довольно общей ситуации. Предположим, что случайные величины $X_{n,j}$ имеют конечные дисперсии. Также предположим, что величины $X_{n,j}$ могут быть представлены в виде

$$X_{n,j} = X_{n,j}^* + \alpha_n,$$

где $\alpha_n \in \mathbb{R}$, а $X_{n,j}^*$ — случайные величины с $EX_{n,j}^* = 0$, $DX_{n,j}^* = \sigma_n^2 < \infty$, так что $EX_{n,1} = \alpha_n$ и $DX_{n,1} = \sigma_n^2$. Предположим, что $\alpha_n k_n \rightarrow a$ и $k_n \sigma_n^2 \rightarrow 1$ при $n \rightarrow \infty$. Тогда вследствие хорошо известного результата о необходимых и достаточных условиях сходимости к нормальному закону распределений сумм независимых случайных величин с конечными дисперсиями в схеме серий (см., например, (Гнеденко, Колмогоров, 1949)), можно заметить, что соотношение (7.25) имеет место тогда и только тогда, когда выполнено условие Линдеберга: для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} k_n E(X_{n,1}^*)^2 \mathbf{1}(|X_{n,1}^*| \geq \varepsilon) = 0,$$

то есть квадратичные хвосты распределений слагаемых должны убывать достаточно быстро.

Следствие 7.8. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что при $n \rightarrow \infty$ имеет место сходимость (7.25). Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость*

$$P(S_{n,N_n} < x) \implies P_{GH}(x; \alpha, 0, \nu, \mu, \lambda), \quad (7.29)$$

необходимо и достаточно, чтобы

$$P(N_n < xk_n) \implies P_{GIG}(x; \nu, \mu, \lambda). \quad (7.30)$$

Доказательство. Из определения обобщенного гиперболического распределения вытекает, что множество $\mathcal{M}(P_{GH}(x; \alpha, 0, \nu, \mu, \lambda) | \Phi(x - \alpha))$ состоит из *единственного* элемента $P_{GIG}(x; \nu, \mu, \lambda)$. Поэтому, применяя Теорему 7.15 с $F(x) = P_{GH}(x; \alpha, 0, \nu, \mu, \lambda)$, замечаем, что в рассматриваемом случае условие (7.27) принимает вид (7.30). Следствие доказано.

Замечание 7.3. В соотношениях (7.25), (7.29) и (7.30) предельные функции распределения непрерывны. Поэтому в этих соотношениях сходимость по распределению эквивалентна равномерной сходимости функций распределения.

7.7 Обобщенные дисперсионные гамма-распределения

В данном разделе в качестве более гибкой альтернативы часто (и успешно) применяемым в практических исследованиях обобщенным гиперболическим распределениям мы рассмотрим класс обобщенных дисперсионных гамма-распределений и дадим теоретическое обоснование использованию его представителей в качестве асимптотических аппроксимаций при решении практических задач.

Как и обобщенные гиперболические распределения, обобщенные дисперсионные гамма-распределения являются специальными дисперсионно-сдвиговыми смесями нормальных законов, но вместо обобщенных обратных гауссовских распределений смешивающими здесь являются обобщенные гамма-распределения.

Обобщенные гамма-распределения (ОГ-распределения) были впервые описаны как единое семейство в 1962 г. в работе (Stacy, 1962) в качестве семейства вероятностных моделей, включающего в себя одновременно гамма-распределения и распределения Вейбулла. Важная роль ОГ-распределений в моделировании и анали-

зе стохастической структуры информационных потоков описана в книге (Королев, Шоргин, 2011).

Обобщенным гамма-распределением называется распределение, определяемое плотностью вероятностей вида

$$f(x; \nu, \kappa, \delta) = \begin{cases} \frac{|\nu|}{\delta^{\kappa|\nu|}\Gamma(\kappa)} x^{\kappa\nu-1} \exp\left\{-\frac{x^\nu}{\delta^{|\nu|}}\right\}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (7.31)$$

с параметрами $\nu \in \mathbb{R}$, $\kappa, \delta \in \mathbb{R}^+$, отвечающими соответственно за *степень, форму и масштаб*, где

$$\Gamma(\kappa) = \int_0^\infty x^{\kappa-1} e^{-x} dx$$

– эйлера гамма-функция. Семейство ОГ-распределений включает в себя практически все наиболее популярные абсолютно непрерывные распределения. В частности, семейство ОГ-распределений содержит следующие распределения.

1. Гамма-распределение ($\nu = 1$):

$$f(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)} \theta^\kappa x^{\kappa-1} e^{-\theta x}, \quad x \geq 0, \quad \kappa > 0, \quad \theta > 0;$$

1.1. Показательное распределение ($\nu = 1, \kappa = 1$):

$$f(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0, \quad \theta > 0;$$

1.2. Распределение Эрланга ($\nu = 1, \kappa \in \mathbb{N}$):

$$f(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)} \theta^\kappa x^{\kappa-1} e^{-\theta x}, \quad x \geq 0, \quad \kappa > 0, \quad \theta > 0;$$

1.3. Распределение хи-квадрат ($\nu = 1, \delta = 2$):

$$f(x; n) = \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{x}{2}\right)^{n/2-1} e^{-x/2}, \quad x \geq 0, \quad n \in \mathbb{N};$$

2. Распределение Накагами ($\nu = 2$):

$$f(x; \mu, \lambda) = \frac{2(\lambda\mu)^\mu}{\Gamma(\mu)} x^{2\mu-1} e^{-\lambda\mu x^2}, \quad x \geq 0, \mu > 0, \lambda > 0;$$

2.1. Полуноормальное распределение (распределение максимума винеровского процесса на отрезке $[0, 1]$) ($\nu = 2, \kappa = \frac{1}{2}$):

$$f(x; \delta) = \sqrt{\frac{2}{\pi\delta}} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0;$$

2.2. Распределение Рэлея ($\nu = 2, \kappa = 1$):

$$f(x; \delta) = \frac{x}{\delta^2} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0;$$

2.3. Хи-распределение ($\nu = 2, \delta = \sqrt{2}$):

$$f(x; n) = \frac{1}{2^{n/2-1}\Gamma(\frac{n}{2})} x^{n-1} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \geq 0, n \in \mathbb{N};$$

2.4. Распределение Максвелла (распределение модулей скоростей движения молекул в разреженном газе) ($\nu = 2, \kappa = 3/2$):

$$f(x; \delta) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\delta^3} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0;$$

3. Распределение Вейбулла–Гнеденко ($\kappa = 1$):

$$f(x; \eta, \mu) = \frac{\eta x^{\eta-1}}{\mu^\eta} \exp\left\{-\left(\frac{x}{\mu}\right)^\eta\right\}, \quad x \geq 0, \eta > 0, \mu > 0;$$

4. Обратное гамма-распределение ($\nu = -1$):

$$f(x; \mu, \lambda) = \frac{1}{\mu\lambda\Gamma(\lambda)} \left(\frac{\mu\lambda}{x}\right)^{\lambda+1} \exp\left\{-\frac{\mu\lambda}{x}\right\},$$

$$x \geq 0, \lambda > 0, \mu > 0;$$

4.1. Распределение Фреше (обратное экспоненциальное распределение) ($\nu = -1$):

$$f(x; \mu) = \frac{1}{\mu} \left(\frac{\mu}{x} \right)^2 \exp \left\{ -\frac{\mu}{x} \right\}, \quad x \geq 0, \mu > 0;$$

4.2. Распределение Леви ($\nu = -1, \kappa = 1/2$):

$$f(x; \mu) = \sqrt{\frac{\mu}{2\pi}} \frac{1}{x^{3/2}} \exp \left\{ -\frac{\mu}{2x} \right\}, \quad x \geq 0, \mu > 0;$$

5. Логнормальное распределение ($\kappa \rightarrow \infty$):

$$f(x; \mu, \delta) = \frac{1}{\delta x \sqrt{2\pi}} \exp \left\{ -\frac{(\log x - \mu)^2}{2\delta^2} \right\},$$

$$x \geq 0, \mu \in \mathbb{R}, \delta > 0.$$

Широкая применимость ОГ-распределений обусловлена возможностью их использования в качестве адекватных асимптотических аппроксимаций, поскольку практически все они выступают в качестве предельных в различных предельных теоремах теории вероятностей. А именно:

- показательное распределение выступает в качестве предельного как в схеме максимума (минимума) (см., например, (Гумбель, 1965), (Галамбош, 1984); более подробно эта схема будет рассмотрена в следующей главе), так и в схеме геометрического суммирования, описывая распределение времени восстановления в прореженных процессах восстановления, выступающих моделями потоков редких событий (см., например, (Kalashnikov, 1997));
- гамма-распределение является безгранично делимым и поэтому выступает в качестве предельного для распределений сумм независимых равномерно предельно малых случайных величин. При этом распределение Эрланга возникает как

допредельное распределение суммы независимых экспоненциально распределенных случайных величин, что в терминах случайной интенсивности может означать, что если случайная интенсивность потока поступления запросов имеет гамма-распределение со значимым параметром формы, то при обработке этих запросов в основном задействованы механизмы последовательной обработки информации;

- распределение Вейбулла–Гнеденко принадлежит к так называемому второму типу предельных распределений экстремальных порядковых статистик (минимума или максимума) (см., например, (Гумбель, 1965), (Галамбош, 1984));
- полунормальное распределение (распределение модуля стандартной нормальной случайной величины) возникает как предельное для максимальных частичных сумм независимых случайных величин (см., например, (Королев, Соколов, 2008));
- распределение Фреше принадлежит к так называемому первому типу предельных распределений экстремальных порядковых статистик (минимума или максимума) (см., например, (Гумбель, 1965), (Галамбош, 1984));
- распределение Леви принадлежит к классу устойчивых законов и потому является предельным для сумм независимых одинаково распределенных случайных величин. Оно также является распределением времени достижения стандартным винеровским процессом (процессом броуновского движения) фиксированного уровня;
- логнормальное распределение выступает в качестве предельного для распределения размера частиц при дроблении (см., например, (Колмогоров, 1941), (Королев, 2009)).

Данные свойства ОГ-распределений обосновывают, в частности, целесообразность моделирования с их помощью распределения случайной интенсивности потока запросов в информацион-

ных системах. Это семейство также широко используется в других прикладных задачах в самых разных областях, см., например, (Королев, Шоргин, 2011).

Упомянутые выше четырехпараметрические семейства скошенных распределений Стьюдента и дисперсионных гамма-распределений являются подклассами введенного в работе (Королев, Соколов, 2012) пятипараметрического семейства распределений

$$W(x; a, \sigma, v, \kappa, \delta) = \int_0^{\infty} \Phi\left(\frac{x - au}{\sigma\sqrt{u}}\right) f(u; v, \kappa, \delta) du, \quad (7.32)$$

где $f(u; v, \kappa, \delta)$ – плотность ОГ-распределения (7.31). В статье (Королев, Соколов, 2012) распределения вида (7.32) названы *обобщенными дисперсионными гамма-распределениями*.

Задача поиска универсальной модели статистических закономерностей во многих областях, в частности в финансовой математике или в физике плазмы, подобна задаче отыскания «философского камня» в алхимии и поэтому не имеет точного решения. Однако, основываясь на вышеперечисленных аналитических и асимптотических свойствах представителей семейства ОГ-распределений и предельных теоремах для сумм независимых случайных величин как теоретико-вероятностной формализации принципа неубывания неопределенности в сложных системах (Gnedenko, Korolev, 1996), можно утверждать, что семейство обобщенных дисперсионных гамма-распределений является *практически* универсальным для многих задач. Оно представляется еще более гибкой моделью, нежели обобщенные гиперболические распределения, так как класс обобщенных гамма-распределений в определенном смысле шире класса обобщенных обратных гауссовских распределений, поскольку, в отличие от последнего, он содержит распределения вейбулловского (экспоненциально-степенного) типа с произвольным показателем степени в экспоненте.

Критерий сходимости распределений случайных сумм к обобщенным дисперсионным гамма-распределениям аналогичен кри-

терию сходимости распределений случайных сумм к обобщенным гиперболическим распределениям, приведенному в предыдущем разделе, и доказывается точно так же. Поэтому мы приводим его без доказательства.

Теорема 7.16. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что при $n \rightarrow \infty$ имеет место сходимость (7.25). Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к обобщенным дисперсионным гамма-распределениям:*

$$P(S_{n, N_n} < x) \implies W(x; \alpha, \sigma, \nu, \kappa, \delta),$$

необходимо и достаточно, чтобы

$$P(N_n < xk_n) \implies F(x; \nu, \kappa, \delta),$$

где $F(x; \nu, \kappa, \delta)$ – функция распределения обобщенного гамма-распределения, соответствующая плотности $f(x; \nu, \kappa, \delta)$ (см. (7.31)).

Предельные функции распределения, фигурирующие в Теореме 7.16 непрерывны. Поэтому в соответствующих предельных соотношениях сходимость по распределению эквивалентна равномерной сходимости функций распределения.

7.8 Двухэтапный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов

При практическом решении задачи моделирования и исследования волатильности (изменчивости) хаотических стохастических процессов ключевым этапом является статистическое разделение смесей вероятностных распределений. Задача разделения смесей – статистического оценивания параметров смесей вероятностных

распределений – в деталях разобрана, например, в книге (Королев, 2011).

Для решения задачи разделения смесей вероятностных распределений традиционно используются итерационные процедуры типа ЕМ-алгоритма. К сожалению, классический ЕМ-алгоритм обладает рядом серьезных недостатков при его применении к смесям нормальных законов. В частности, он демонстрирует крайнюю неустойчивость по отношению к исходным данным и начальным приближениям. Для преодоления этих недостатков предложено много модификаций ЕМ-алгоритма, см., например, (Королев, 2011). Вместе с тем, в указанной книге предложен и исследован принципиально новый «сеточный» метод приближенного решения задачи разделения смесей. В работе (Назаров, 2013) подробно исследованы вопросы сходимости сеточных методов разделения смесей.

В соответствии с подходом к статистическому анализу хаотических стохастических процессов, в частности, к решению задачи декомпозиции волатильности таких процессов, развитом в книге (Королев, 2011), в общем случае на практике приходится решать задачу разделения конечных смесей нормальных законов с произвольно большим числом неизвестных (параметров отдельно взятых компонент и их весов). И хотя в большинстве приложений возникают смеси не более чем с пятью-семью компонентами, даже при использовании таких смесей, скажем, в задачах анализа и прогнозирования финансовых рисков, приходится моделировать траекторию движения точки в пространствах, размерность которых, соответственно, лежит в пределах от 14 (для пятикомпонентных смесей) до 20 (для семикомпонентных смесей), что существенно увеличивает вычислительные и временные ресурсы, необходимые для практического решения указанных задач. Поскольку во многих ситуациях, например, при прогнозировании на основе высокочастотных данных, эти задачи необходимо решать в режиме, близком к реальному времени, для создания эффективных методов статистического анализа на основе смешанных моделей на первый план выходит проблема снижения размерности решаемой задачи, т. е. параметрического пространства.

Одним из возможных подходов к снижению размерности является априорное сужение классов допустимых смесей. Как уже отмечалось выше, при решении многих задач, например, связанных с анализом процессов атмосферной или плазменной турбулентности, а также процессов, описывающих эволюцию различных финансовых индексов, высочайшую адекватность продемонстрировали модели, основанные на обширном классе дисперсионно-сдвиговых смесей нормальных законов, в частности, на обобщенных гиперболических законах.

В указанном семействе смесей число неизвестных параметров равно пяти или шести, если учитывать неслучайный сдвиг. Вместе с тем, как было отмечено выше, у подобных моделей имеются довольно серьезные теоретические обоснования: указанные модели являются асимптотическими аппроксимациями в простой предельной схеме случайного суммирования и потому могут успешно применяться для анализа процессов по типу остановленных случайных блужданий. Эти выводы подтверждены статистическим анализом высокочастотных финансовых данных, в результате которого выявлен синхронизированный характер изменения интенсивностей потоков заявок в системах электронных торгов, что естественно приводит к синхронизированному поведению параметров сдвига и диффузии в соответствующих моделях вида смесей нормальных законов (Королев, Черток, Корчагин, Горшенин, 2013).

В работе (Королев, Корчагин, 2014) предложен принципиально новый метод разделения дисперсионно-сдвиговых смесей нормальных законов, в частности, на примере обобщенных гиперболических и обобщенных дисперсионных гамма-распределений. Оказывается, сеточные методы разделения смесей достаточно эффективны не только при разделении конечных смесей нормальных законов, но и при разделении произвольных дисперсионно-сдвиговых смесей нормальных законов. Поясним сказанное на примере задачи оценивания параметров обобщенных гиперболических распределений.

Для решения задачи оценивания параметров обобщенных гиперболических распределений традиционно используется метод,

предложенный в статье (Protassov, 2004), и по сути являющийся классическим ЕМ-алгоритмом, приспособленным к конкретной задаче, и соответственно, наследующий присущие ЕМ-алгоритмам недостатки.

Рассмотрим следующий альтернативный двухэтапный метод, предложенный в (Королев, Корчагин, 2014). На первом этапе на положительной полупрямой выделим основную часть носителя смешивающего распределения, то есть ограниченный интервал, вероятностная мера которого, вычисленная в соответствии со смешивающим распределением, практически равна единице. На этот интервал накинem конечную сетку, содержащую, возможно, очень много *известных* узлов $u_1, \dots, u_K$. Считая параметр сдвига β равным нулю, приблизим искомое обобщенное гиперболическое распределение конечной смесью нормальных законов:

$$P_{GH}(x; \alpha, 0, \nu, \mu, \lambda) \approx \sum_{i=1}^K p_i \Phi\left(\frac{x - \alpha u_i}{\sqrt{u_i}}\right), \quad x \in \mathbb{R}. \quad (7.33)$$

В смеси, стоящей в правой части соотношения (7.33), неизвестными являются только параметры $p_1, \dots, p_{K-1}, \alpha$. Пусть $x_1, \dots, x_n$ – анализируемая выборка значений случайной величины с оцениваемым обобщенным гиперболическим распределением. Итерационный процесс, определяющий сеточный ЕМ-алгоритм для данной задачи, определяется следующим образом. Пусть $p_1^{(m)}, \dots, p_{K-1}^{(m)}, \alpha^{(m)}$ – оценки параметров $p_1, \dots, p_{K-1}$ и α на m -й итерации, $p_K^{(m)} = 1 - p_1^{(m)} - \dots - p_{K-1}^{(m)}$. Обозначим

$$\varphi_{ij}^{(m)} = \frac{1}{\sqrt{u_i}} \varphi\left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}}\right), \quad g_{ij}^{(m)} = \frac{p_i^{(m)} \varphi_{ij}^{(m)}}{\sum_{r=1}^K p_r^{(m)} \varphi_{rj}^{(m)}},$$

$$i = 1, \dots, K; \quad j = 1, \dots, n.$$

Тогда, используя стандартные рассуждения, определяющие вычислительные формулы ЕМ-алгоритма для параметров конечной смеси нормальных законов (см, например, (Королев, 2011), разделы

5.3.7-5.3.8), следует положить

$$p_i^{(m+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m)}, \quad i = 1, \dots, K. \quad (7.34)$$

Обозначим $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Используя соотношение (5.3.24) в [46],

с учетом очевидного равенства $\sum_{i=1}^K g_{ij}^{(m)} = 1$ можно заметить, что уточненная оценка параметра α имеет вид

$$\alpha^{(m+1)} = \frac{\bar{x}}{\sum_{i=1}^K u_i p_i^{(m+1)}}, \quad (7.35)$$

то есть равна отношению генерального выборочного среднего и текущего эмпирического среднего смешивающего распределения, что вполне согласуется с соотношением (7.22) при $\beta = 0$.

В силу монотонности классического ЕМ-алгоритма справедливо следующее утверждение (Королев, Корчагин, 2014).

Теорема 7.17. Пусть узлы $u_1, \dots, u_K$ сетки различны, неотрицательны и известны. Итерационный процесс (7.34) – (7.35) является монотонным, то есть каждая его итерация не уменьшает целевую сеточную функцию правдоподобия

$$L(p_1, \dots, p_{K-1}, \alpha; x_1, \dots, x_n) = \prod_{j=1}^n \left[\sum_{i=1}^K \frac{p_i}{\sqrt{u_i}} \varphi \left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}} \right) \right].$$

Замечание 7.4. В разделе 5.7.4 книги (Королев, 2011) показано, что при каждом фиксированном значении параметра α сеточная функция правдоподобия $L(p_1, \dots, p_{K-1}, \alpha; x_1, \dots, x_n)$ вогнута по аргументам $p_1, \dots, p_{K-1}$. Поэтому на каждом шаге итерационного процесса вместо соотношения (7.34) можно использовать любой более быстрый алгоритм максимизации функции $L(p_1, \dots, p_{K-1}, \alpha^{(m)}; x_1, \dots, x_n)$ по переменным $p_1, \dots, p_{K-1}$. Например, оценки весов $p_1, \dots, p_K$ можно искать методом условного градиента (Королев, 2011), (Королев, Назаров, 2010).

Таким образом, на первом этапе получают оценки параметра α и весов всех узлов u_i конечной сетки, накинута на носитель смешивающего обобщенного обратного гауссовского распределения $P_{GIG}(z; \nu, \mu, \lambda)$.

На втором этапе остается применить какой-либо стандартный метод подгонки обобщенного обратного гауссовского распределения $P_{GIG}(z; \nu, \mu, \lambda)$ к эмпирическим данным типа гистограммы $(u_1, p_1), \dots, (u_K, p_K)$.

Например, параметры ν , μ и λ можно оценить, минимизируя соответствующую статистику хи-квадрат. Или же, например, можно решить задачу наименьших квадратов

$$(\nu^*, \mu^*, \lambda^*) = \arg \min_{\nu, \mu, \lambda} \sum_{i=1}^K \left[p_i - \int_{\frac{1}{2}(u_{i-1} + u_i)}^{\frac{1}{2}(u_i + u_{i+1})} P_{GIG}(u; \nu, \mu, \lambda) du \right]^2,$$

где $u_0 = 0$, $u_{K+1} = \infty$.

На практике подход с применением метода наименьших квадратов показал хорошие результаты. Для поиска параметров подбираемых распределений использовалось несколько алгоритмов, в частности, алгоритм ns2sol, описанный в книге (Dennis, Schnabel, 1983). Данный алгоритм, доступный во многих статистических пакетах, отличается высоким быстродействием и возможностью при желании задавать априорные разумные интервалы для поиска параметров.

Также хорошие результаты показал метод поиска наилучшего распределения в смысле минимизации расстояния Кульбака-Лейблера, который в данном случае эквивалентен максимизации правдоподобия полученной гистограммы в выбранных классах распределений.

7.9 Устойчивость представления вероятностных распределений в виде дисперсионно-сдвиговых смесей нормальных законов

Как видно из описания двухэтапного метода, приведенного в предыдущем разделе, ключевую роль в нем играет выбор конечной сетки на первом этапе. Эту сетку надо выбрать так, чтобы конечная аппроксимирующая смесь мало отличалась от исходного распределения. В данном разделе рассмотрены вопросы устойчивости представления вероятностных распределений в виде дисперсионно-сдвиговых смесей нормальных законов. Будут рассматриваться задачи, сходные с теми, которые рассмотрены в разделе 7.4. Нас будут интересовать оценки в метрике Леви, метризующей слабую сходимость.

Напомним, что расстояние Леви между случайными величинами X и Y с функциями распределения F_X и F_Y соответственно (равносильно будем использовать понятие расстояния Леви между функциями распределения) определяется, как

$$L(X, Y) = L(F_X, F_Y) = \inf\{h : F_Y(x-h) - h \leq F_X(x) \leq F_Y(x+h) + h \text{ для всех } x \in \mathbb{R}\}. \quad (7.36)$$

Также можно убедиться, что

$$L(X, Y) = \max\{\Delta(X, Y), \Delta(Y, X)\}, \quad (7.37)$$

где $\Delta(X, Y) = \max\{\min\{x - y, F_X(x) - F_Y(y)\} : x, y \in \mathbb{R}\}$. С геометрической точки зрения соотношение (7.37) означает, что метрика Леви имеет смысл длины стороны наибольшего квадрата со сторонами, параллельными координатным осям, который может быть вписан между графиками функций $F_X(x)$ и $F_Y(x)$.

В доказательствах нам понадобятся неравенства, связывающие метрику Леви с равномерной метрикой ρ , которая определяется, как

$$\rho(X, Y) = \rho(F_X, F_Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|.$$

Ясно, что $L(X, Y) \leq \rho(X, Y)$, т. к. для $\rho(X, Y)$ выполняются неравенства, стоящие под знаком инфимума в (7.36). Если одна из функций распределений F_X, F_Y , например, F_Y является гладкой и, кроме того, имеет ограниченную плотность $p_Y(x)$, то

$$L(X, Y) \leq \rho(X, Y) \leq \left(1 + \sup_{x \in \mathbb{R}} p_Y(x)\right) L(X, Y).$$

Мы будем рассматривать смеси вероятностных распределений, характеристические функции которых представимы в виде $E f^U(t)$, где $f(t)$ – некоторая безгранично делимая характеристическая функция. Особое внимание будет уделено случаю, когда $f(t) = \varphi_{\alpha, \sigma}(t) = \exp\{i\alpha t - \frac{1}{2}t^2\sigma^2\}$ – характеристическая функция нормального распределения с параметрами $\alpha \in \mathbb{R}$ и $\sigma > 0$, а $U \geq 0$ – случайная величина. Такие смеси будем называть степенными. Введем обозначение

$$\{f, A\}(t) = \int_0^\infty f^u(t) dA(u), \quad t \in \mathbb{R},$$

где $f(t)$ – безгранично делимая характеристическая функция, а $A(x)$ – функция распределения положительной с вероятностью единица случайной величины.

Если $\Phi_{\alpha, \sigma}(x) = \Phi\left(\frac{x-\alpha}{\sigma}\right)$ – функция распределения, соответствующая характеристической функции $\varphi_{\alpha, \sigma}$, то функцию распределения дисперсионно-сдвиговой смеси нормальных законов, соответствующую характеристической функции $\{\varphi_{\alpha, \sigma}, A\}$, будем обозначать, как $[\Phi_{\alpha, \sigma}, A]$.

Предположим, что A и B – функции распределения с точками роста, расположенными на неотрицательной полуоси, и что по крайней мере одна из них (пусть это будет A) имеет плотность, ограниченную некоторым числом $0 < a < \infty$. В данном разделе будет решена следующая задача: как оценить расстояние Леви между функциями распределения $[\Phi_{\alpha, \sigma}, A]$ и $[\Phi_{\alpha, \sigma}, B]$ через расстояние Леви между функциями распределения A и B ? Эта задача важна для обоснования метода разделения дисперсионно-сдвиговых смесей, описанного в предыдущем разделе. Поскольку здесь одно из

двух смешивающих распределений дискретно, то именно метрика Леви в бóльшей мере соответствует задаче, нежели, к примеру, равномерное расстояние.

Лемма 7.5. Если $F(x)$ и $G(x)$ – две дифференцируемые функции распределения, то $\rho(F, G)$ реализуется ($\sup_{x \in \mathbb{R}} |F(x) - G(x)|$ достигается) в одной из точек, в которых $F'(x) = G'(x)$.

Доказательство. Очевидно, что в рассматриваемой ситуации

$$\begin{aligned} \rho(F, G) &= \sup_{x \in \mathbb{R}} |F(x) - G(x)| = \\ &= \max \left\{ \max_{x \in \mathbb{R}} [F(x) - G(x)], \max_{x \in \mathbb{R}} [G(x) - F(x)] \right\}, \end{aligned}$$

а экстремум каждого из выражений в фигурных скобках в правой части достигается в точке, где производная соответствующего выражения равна нулю, что эквивалентно равенству производных функций распределения F и G , то есть равенству соответствующих плотностей.

Теорема 7.18. Предположим, что A и B – функции распределения с точками роста, расположенными на неотрицательной полуоси, и что по крайней мере одна из них (пусть это будет A) имеет плотность, ограниченную некоторым числом $0 < a < \infty$. Тогда

$$L([\Phi_{\alpha, \sigma}, A], [\Phi_{\alpha, \sigma}, B]) \leq 2(1 + a)L(A, B).$$

Доказательство. Пусть x_0 точка, в которой реализуется $\rho([\Phi_{\alpha, \sigma}, A], [\Phi_{\alpha, \sigma}, B])$. Тогда

$$\begin{aligned} \rho([\Phi_{\alpha, \sigma}, A], [\Phi_{\alpha, \sigma}, B]) &= \\ &= \left| \int_0^\infty \Phi\left(\frac{x_0 - \alpha u}{\sigma \sqrt{u}}\right) dA(u) - \int_0^\infty \Phi\left(\frac{x_0 - \alpha u}{\sigma \sqrt{u}}\right) dB(u) \right| = \\ &= \left| \int_0^\infty \Phi\left(\frac{x_0 - \alpha u}{\sigma \sqrt{u}}\right) d[A(u) - B(u)] \right|. \end{aligned}$$

Интегрируя по частям, получим

$$\begin{aligned}
 & \left| \int_0^\infty \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) d[A(u) - B(u)] \right| = \\
 & = \left| \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) [A(u) - B(u)] \Big|_{u=0}^\infty - \int_0^\infty [A(u) - B(u)] d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right| = \\
 & = \left| \int_0^\infty [A(u) - B(u)] d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right| \leq \rho(A, B) \cdot \int_0^\infty \left| d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right|.
 \end{aligned}$$

Рассмотрим поведение аргумента функции $\Phi((x_0 - \alpha u)/(\sigma\sqrt{u}))$ как функции u . Имеем

$$\frac{d}{du} \left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}} \right) = -\frac{1}{2\sqrt{u}} \left(\frac{x_0}{\sigma u} + \alpha \right),$$

откуда видно, что эта производная меняет знак не более чем в одной точке $u_0 = -x_0/(\alpha\sigma)$. При этом функция

$$g(u) = \frac{x_0 - \alpha u}{\sigma\sqrt{u}}$$

монотонна при $u \leq u_0$ и при $u \geq u_0$. Стало быть, поскольку Φ – монотонная функция распределения, функция $\Phi((x_0 - \alpha u)/(\sigma\sqrt{u}))$ монотонна как функция u на тех же множествах. Поэтому

$$\int_0^\infty \left| d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right| \leq \int_0^{u_0} \left| d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right| + \int_{u_0}^\infty \left| d_u \Phi\left(\frac{x_0 - \alpha u}{\sigma\sqrt{u}}\right) \right| \leq 2,$$

так как суммарное приращение любой функции распределения не превосходит единицы. Таким образом,

$$\rho([\Phi_{\alpha,\sigma}, A], [\Phi_{\alpha,\sigma}, B]) \leq 2\rho(A, B),$$

откуда с учетом соотношений между метрикой Леви и равномерной метрикой мы получаем

$$\begin{aligned}
 L([\Phi_{\alpha,\sigma}, A], [\Phi_{\alpha,\sigma}, B]) & \leq \rho([\Phi_{\alpha,\sigma}, A], [\Phi_{\alpha,\sigma}, B]) \leq \\
 & \leq 2\rho(A, B) \leq 2(1 + a)L(A, B),
 \end{aligned}$$

что завершает доказательство.

Глава 8

Вероятностные модели типа распределений времени жизни

В предыдущих разделах рассматривались вероятностные модели, теоретической основой которых служат предельные теоремы для (нестационарных) случайных блужданий. В таких моделях в качестве “допредельной” структуры изучаемого процесса (или объекта) фигурируют *суммы* случайных величин. Вместе с тем существует много задач, в которых вместо схемы суммирования случайных величин естественно возникает схема максимума или минимума случайных величин. Прежде всего, это задачи, связанные с изучением времени жизни различных объектов, например, технических систем (в последнем случае вместо термина “время жизни” специалисты в области теории надежности используют термин “время безотказной работы” или “наработка на отказ”).

Представим себе, что изучаемая система состоит из многих блоков. Возможны две противоположные ситуации: 1) вся система выходит из строя, как только перестает правильно функционировать хотя бы один ее блок, и 2) система продолжает функционировать, пока работает хотя бы один блок. В первом случае время жизни системы – это наименьшее из времен жизни ее блоков, во втором случае время жизни системы – это наибольшее из времен жизни ее блоков. Конечно, если система имеет сложную структуру, то ее время жизни определяется менее простым способом, то тем не менее, все равно при определении соответствующего алгоритма возникает необходимость иметь модели типа максимума или минимума случайных величин в качестве составных частей более сложной общей модели. Данная глава посвящена как раз рассмотрению моделей типа экстремума (максимума или минимума). При этом с учетом сказанного в предыдущих разделах мы сразу изу-

чим модели распределений экстремальных значений в выборках случайного объема.

8.1 Определение и примеры макс-обобщенных дважды стохастических пуассоновских процессов

Пусть $N(t)$, $t \geq 0$, – процесс Кокса, управляемый процессом $\Lambda(t)$ (см. раздел 6.3). Пусть $T_1, T_2, \dots$ – точки скачков процесса $N(t)$. Рассмотрим *маркированный процесс Кокса* $\{(T_i, X_i)\}_{i \geq 1}$, где $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины, независимые от процесса $N(t)$. Основным объектом изучения в предыдущих главах был обобщенный процесс Кокса $S(t)$, являющийся случайной функцией от маркированного процесса Кокса и определявшийся в каждый момент времени $t \geq 0$ как *сумма* всех маркирующих случайных величин маркированного процесса Кокса до момента времени t . Как мы видели, такие случайные функционалы играют важную роль при моделировании многих процессов, а *их* экстремумы могут успешно использоваться как модели катастрофически накопленных неблагоприятных воздействий во многих областях. Не менее интересным и важным объектом является так называемый макс-обобщенный процесс Кокса, также являющийся случайной функцией от маркированного процесса Кокса и отличающийся от $S(t)$ тем, что в нем обобщающая операция суммирования маркирующих случайных величин заменена операцией *максимума*.

Определение 8.1. Макс-обобщенным процессом Кокса называется процесс $M(t)$, определяемый как

$$M(t) = \begin{cases} -\infty, & \text{если } N(t) = 0, \\ \max_{1 \leq k \leq N(t)} X_k, & \text{если } N(t) \geq 1, \end{cases} \quad t \geq 0.$$

Помимо задач описания времени жизни, макс-обобщенные процессы Кокса могут найти самое серьезное применение при по-

строении математических моделей потоков катастрофических событий. Поясним сказанное примерами.

Пример 8.1. Максимальная страховая выплата. Естественно, что для увеличения прибыли резерв страховой компании не должен лежать без дела, а какая-то часть его должна быть вложена в какие-либо ценные бумаги или какое-либо другое прибыльное предприятие. Для определения оптимальной части резерва, не подлежащего инвестированию, необходимо иметь информацию о размере максимальной страховой выплаты. Если не инвестированная (свободная) часть резерва оказывается меньше максимальной страховой выплаты, то это может повлечь определенные сложности при осуществлении этой выплаты.

Пример 8.2. При проектировании строений в сейсмоопасных зонах необходимо иметь информацию о максимально возможной силе землетрясения в таком районе. Моменты землетрясений представляют собой хаотический точечный процесс. Если бы этот процесс был однородным, то наилучшей его моделью, конечно же, был бы пуассоновский точечный процесс. Однако, довольно очевидно, что реальная интенсивность катастрофических событий непостоянна. Многие исследователи напрямую связывают эту интенсивность с солнечной активностью, а она, как хорошо известно, обладает цикличностью. Поэтому существенно более разумной моделью потока землетрясений является пуассоновский процесс со стохастической интенсивностью, то есть процесс Кокса. Прочностные требования к проектируемым строениям определяются поведением хвостов распределения максимально возможной силы землетрясения с учетом допустимого риска, то есть допустимых вероятностях практически невозможных событий.

Аналогичные примеры могут быть связаны с проектированием дамб с учетом максимально возможных уровней подъемов воды и с другой деятельностью, связанной с противодействием риску катастрофических событий. При этом, как показано ниже, по аналогии с экстремумами дважды стохастических (маркированных) пуассоновских процессов, обобщенных в традиционном смысле,

то есть с помощью операции суммирования маркирующих величин (см. (Королев, Соколов, 2008)), асимптотические распределения макс-обобщенных процессов Кокса (то есть дважды стохастических пуассоновских процессов, обобщенных в смысле операции максимума, применяемой к маркирующим величинам) могут иметь хвосты, отличающиеся от хвостов предельных законов в классической теории. А это приводит к тому, что *предположение о стохастичности интенсивности потока нежелательных событий ведет к выводу о том, что вероятности катастрофических событий реально оказываются не такими (например, большими), чем предсказывает классическая асимптотическая теория экстремальных значений*. Другими словами, классическая теория не всегда правильно оценивает риск катастрофических событий.

Целью данной главы является изучение асимптотического поведения макс-обобщенных процессов Кокса. Представленные здесь результаты относятся к теории экстремальных значений в выборках случайного объема. Основы этой теории изложены в работах (Berman, 1964), (Barndorff-Nielsen, 1964), (Mogyorodi, 1967), (Гнеденко и Сенукси-Берекси, 1982), (Гнеденко и Сенукси-Берекси, 1982), (Сенукси-Берекси и Янич, 1984), (Б. В. Гнеденко и Д. Б. Гнеденко, 1982). В книге (Галамбош, 1984) можно найти обзор соответствующих результатов. Во всех упомянутых работах рассматриваются так называемые теоремы переноса, которые устанавливают сходимость распределений экстремумов выборок случайного объема в предположении, что экстремумы в выборках неслучайного объема, будучи надлежащим образом центрированы и нормированы, имеют предельные распределения. Мы по сути рассматриваем частный случай общей схемы, рассмотренной в упомянутых работах. Однако при этом удастся довольно подробно проследить взаимосвязь констант, нормирующих экстремумы, и констант, нормирующих индексы (в нашем случае, нормирующих управляющий процесс, что эквивалентно нормированию самого процесса Кокса).

Рассмотрим распределение вероятностей макс-обобщенных дважды стохастических пуассоновских процессов.

Пусть сначала $N_{\lambda}(t)$ – однородный пуассоновский процесс с

интенсивностью $\lambda > 0$. Обозначим

$$M^{(\lambda)}(t) = \begin{cases} -\infty, & \text{если } N_\lambda(t) = 0, \\ \max_{1 \leq k \leq N_\lambda(t)} X_k, & \text{если } N_\lambda(t) \geq 1, \end{cases} \quad t \geq 0.$$

Процесс $M^{(\lambda)}(t)$ назовем макс-обобщенным пуассоновским процессом. Функцию распределения случайной величины X_1 обозначим $F(x)$:

$$F(x) = P(X_1 < x), \quad x \in \mathbb{R}.$$

Лемма 8.1. *Для произвольного $t \geq 0$ функция распределения макс-обобщенного пуассоновского процесса имеет вид*

$$P(M^{(\lambda)}(t) < x) = \exp \{ \lambda t [F(x) - 1] \}, \quad x \in \mathbb{R}. \quad (8.1)$$

Доказательство. С учетом соотношения

$$P\left(\max_{1 \leq k \leq n} X_k < x\right) = F^n(x), \quad x \in \mathbb{R},$$

справедливого при любом натуральном n , для произвольных $x \in \mathbb{R}$ и $t \geq 0$ по формуле полной вероятности имеем

$$\begin{aligned} P(M^{(\lambda)}(t) < x) &= \sum_{k=0}^{\infty} P(N_\lambda(t) = n) P\left(\max_{1 \leq k \leq n} X_k < x\right) = \\ &= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{[\lambda t F(x)]^n}{n!} = \exp \{ \lambda t [F(x) - 1] \}. \end{aligned}$$

Лемма доказана.

Очевидно, что утверждение леммы 8.1 является частным случаем следующего более общего утверждения.

Лемма 8.2. *Пусть $X_1, X_2, \dots$ – одинаково распределенные случайные величины с общей функцией распределения $F(x)$. Пусть N – целочисленная неотрицательная случайная величина с производящей функцией $\psi(s)$,*

$$\psi(s) = E s^N = \sum_{n=0}^{\infty} s^n P(N = n), \quad |s| \leq 1.$$

Предположим, что случайные величины $N, X_1, X_2, \dots$ независимы в совокупности. Тогда

$$P\left(\max_{0 \leq k \leq N} X_k < x\right) = \psi(F(x)), \quad x \in \mathbb{R}.$$

Видно, что если $x \in \mathbb{R}$ фиксировано, причем $F(x) < 1$, то при $\lambda t \rightarrow \infty$ вероятность (8.1) стремится к нулю экспоненциально быстро. Соответственно, можно легко заметить, что при фиксированном $x \in \mathbb{R}$ функция

$$B(t) \equiv P(M^{(\lambda)}(t) \geq x) = 1 - \exp\{-\lambda t[1 - F(x)]\}, \quad t \geq 0,$$

совпадает с показательной функцией распределения с параметром $\lambda[1 - F(x)]$. Таким образом, если $\tau_\lambda = \tau_\lambda(x)$ – случайная величина, определяемая соотношением

$$\tau_\lambda \equiv \tau_\lambda(x) = \inf\{t : M^\lambda(t) \geq x\}$$

и имеющая смысл момента первого достижения процессом $M^\lambda(t)$ уровня x , то, как легко видеть, события $\{\tau_\lambda < t\}$ и $\{M^{(\lambda)}(t) \geq x\}$ эквивалентны ($t \geq 0$). Поэтому распределение случайной величины $\tau_\lambda = \tau_\lambda(x)$ является показательным с параметром $\lambda[1 - F(x)]$.

Если же $x \rightarrow -\infty$, то $F(x) \rightarrow 0$, так что

$$\lim_{x \rightarrow -\infty} P(M^{(\lambda)}(t) < x) = e^{-\lambda t} = P(N^{(\lambda)}(t) = 0),$$

то есть конечномерные распределения макс-обобщенных пуассоновских процессов являются несобственными.

Предположим, что функция распределения $F(x)$ абсолютно непрерывна. Обозначим соответствующую ей плотность через $f(x)$. Тогда условная функция распределения

$$P(M^{(\lambda)}(t) < x | N_\lambda(t) > 0) = \frac{e^{-\lambda t[1 - F(x)]} - e^{-\lambda t}}{1 - e^{-\lambda t}} = \frac{e^{\lambda t F(x)} - 1}{e^{\lambda t} - 1}$$

абсолютно непрерывна, причем ей соответствует плотность

$$p(x) = \frac{\lambda t e^{\lambda t F(x)} f(x)}{e^{\lambda t} - 1}.$$

Пусть теперь $M(t)$ – макс-обобщенный процесс Кокса, порожденный последовательностью $X_1, X_2, \dots$ и управляемый случайным процессом $\Lambda(t)$. Тогда, очевидно, из леммы 8.1 мы получаем, что

$$\begin{aligned} P(M(t) < x) &= E \exp \{ -\Lambda(t)[1 - F(x)] \} = \\ &= \int_0^{\infty} \exp \{ -\lambda[1 - F(x)] \} dP(\Lambda(t) < \lambda), \quad x \in \mathbb{R}. \end{aligned}$$

Замечание 8.1. Приведенное выше соотношение означает, что функция распределения макс-обобщенного процесса Кокса, управляемого случайным процессом $\Lambda(t)$, в произвольной точке $x \in \mathbb{R}$ равна преобразованию Лапласа–Стилтьеса случайной величины $\Lambda(t)$ в точке $z = 1 - F(x)$.

При решении практических задач полезным может оказаться следующее утверждение, очевидным образом вытекающее из приведенного выше представления функции распределения макс-обобщенного процесса Кокса.

Утверждение 8.1. Пусть $t > 0$ и $x \in \mathbb{R}$ произвольны. Пусть Z – случайная величина со стандартным показательным распределением, независимая от случайной величины $\Lambda(t)$. Тогда

$$P(M(t) \geq x) = P(Z < \Lambda(t)[1 - F(x)]).$$

Из этого утверждения, в частности, вытекает, что если $M^{(\lambda)}(t)$ – макс-обобщенный пуассоновский процесс с интенсивностью λ , то

$$P(M^{(\lambda)}(t) \geq x) = P(Z < \lambda t[1 - F(x)]),$$

где, как и в утверждении 8.1, Z – случайная величина со стандартным показательным распределением.

Если $\tau = \tau(x)$ – случайная величина, определяемая соотношением

$$\tau \equiv \tau(x) = \inf\{t : M(t) \geq x\}$$

и имеющая смысл первого достижения уровня x процессом $M(t)$, то, как легко видеть, события $\{\tau < t\}$ и $\{M(t) \geq x\}$ эквивалентны ($t \geq 0$). Поэтому распределение случайной величины $\tau = \tau(x)$

является смешанным показательным, в котором смешивающее распределение совпадает с распределением случайной величины $\Lambda(t)[1 - F(x)]$.

Пусть, к примеру, накопленная интенсивность $\Lambda(t)$ имеет показательное распределение с параметром t^{-1} . Тогда

$$\begin{aligned} P(\tau(x) < t) &= P(M(t) \geq x) = \\ &= \int_0^{\infty} (1 - \exp\{-\lambda[1 - F(x)]\}) t^{-1} e^{-\lambda t^{-1}} d\lambda = \\ &= 1 - \frac{1}{t} \int_0^{\infty} e^{-\lambda[1 - F(x) + t^{-1}]} d\lambda = 1 - \frac{1}{t(t^{-1} + 1 - F(x))} \int_0^{\infty} e^{-\lambda} d\lambda = \\ &= 1 - \frac{1}{1 + t[1 - F(x)]}. \end{aligned}$$

Другими словами, в этом случае момент первого достижения уровня x макс-обобщенным процессом Кокса имеет распределение Парето с масштабным параметром $[1 - F(x)]$ и единичным параметром формы. У такого распределения отсутствуют все моменты порядков $\delta \geq 1$, в том числе отсутствует математическое ожидание. При этом медиана этого распределения равна

$$\text{med}\tau(x) = \frac{1}{1 - F(x)}.$$

Этот пример довольно наглядно иллюстрирует эффект существенного изменения характера распределения временных характеристик экстремальных событий при случайной интенсивности их потока по сравнению с той ситуацией, в которой интенсивность постоянна.

Приведем несколько простых и полезных неравенств, связывающих распределение макс-обобщенных процессов Кокса с распределением накопленной интенсивности и распределением случайных величин $X_1, X_2, \dots$

Заметим, что функция $g(\lambda) = e^{-r\lambda}$ выпукла при $\lambda > 0$ для каждого фиксированного $r > 0$. Поэтому, применяя неравенство Йенсена, в случае когда $E\Lambda(t) < \infty$, из приведенного выше представления для функции распределения макс-обобщенного процесса Кокса вытекает, что при каждом $x \in \mathbb{R}$

$$P(M(t) < x) = E \exp \{ -\Lambda(t)[1 - F(x)] \} \geq \exp \{ -E\Lambda(t)[1 - F(x)] \},$$

откуда, в частности, получаем оценку

$$E\Lambda(t)[1 - F(x)] \geq -\log P(M(t) < x).$$

Далее, используя хорошо известное представление математического ожидания неотрицательных случайных величин (см., например, (Зубков, Севастьянов и Чистяков, 1989), задача 3.137), в случае когда случайные величины $X_1, X_2, \dots$ положительны и имеют конечное математическое ожидание, из последней оценки получаем неравенство, связывающее параметры распределения обобщенного (в обычном смысле, см. раздел 6.3) процесса Кокса $S(t)$ и соответствующего макс-обобщенного процесса Кокса:

$$ES(t) = E\Lambda(t) \cdot EX_1 \geq - \int_0^{\infty} \log P(M(t) < x) dx.$$

Еще одним следствием приведенной выше оценки для функции распределения $P(M(t) < x)$ является следующее неравенство. Пусть, как и ранее, $\tau(x)$ – момент первого достижения уровня x макс-обобщенным процессом Кокса $M(t)$. Тогда

$$\begin{aligned} P(\tau(x) < t) &= P(M(t) \geq x) = 1 - E \exp \{ -\Lambda(t)[1 - F(x)] \} \leq \\ &\leq 1 - \exp \{ -E\Lambda(t)[1 - F(x)] \}. \end{aligned}$$

Отсюда вытекает экстремальное положение макс-обобщенных пуассоновских процессов в классе макс-обобщенных дважды стохастических пуассоновских процессов: для макс-обобщенных пуассоновских процессов распределение случайной величины $\tau(x) = \tau_\lambda(x)$ имеет самый легкий хвост среди всех макс-обобщенных дважды стохастических пуассоновских процессов.

Этот факт проиллюстрирован приведенным выше примером, в котором случайная величина $\tau(x)$ имеет распределение Парето.

Утверждение 8.2. Для любых $t > 0$, $x \in \mathbb{R}$ и $z > 0$ справедливы неравенства

$$P(\tau(x) < t) = P(M(t) \geq x) \geq (1 - e^{-z[1-F(x)]})P(\Lambda(t) \geq z),$$

$$P(\tau(x) \geq t) = P(M(t) < x) \geq e^{-z[1-F(x)]}P(\Lambda(t) < z).$$

Доказательство этого утверждения элементарно. Первое неравенство получается так:

$$\begin{aligned} P(M(t) \geq x) &= 1 - P(M(t) < x) = \int_0^{\infty} (1 - e^{-\lambda[1-F(x)]})dP(\Lambda(t) < \lambda) \geq \\ &\geq \int_z^{\infty} (1 - e^{-\lambda[1-F(x)]})dP(\Lambda(t) < \lambda) \geq (1 - e^{-z[1-F(x)]})P(\Lambda(t) \geq z). \end{aligned}$$

Второе неравенство доказывается аналогично:

$$\begin{aligned} P(M(t) < x) &= \int_0^{\infty} e^{-\lambda[1-F(x)]}dP(\Lambda(t) < \lambda) \geq \\ &\geq \int_0^z e^{-\lambda[1-F(x)]}dP(\Lambda(t) < \lambda) \geq e^{-z[1-F(x)]}P(\Lambda(t) < z). \end{aligned}$$

Из утверждения 8.2 несложно получить несколько следствий.

Во-первых, предполагая, что x таково, что $F(x) \neq 1$, и полагая $z = [1 - F(x)]^{-1}$, из утверждения 8.2 получим неравенство

$$P(M(t) \geq x) \geq \frac{e-1}{e} \cdot P\left(\Lambda(t) \geq \frac{1}{1-F(x)}\right).$$

Во-вторых, обозначив $m(t) = \text{med}\Lambda(t)$, из утверждения 8.2 получим неравенство

$$2P(M(t) \geq x) \geq 1 - \exp\{-m(t)[1 - F(x)]\}.$$

В-третьих, если $x > 0$, то полагая $z = x$, получаем неравенство, напрямую связывающее “хвосты” распределений случайных величин X_1 , $M(t)$ и $\Lambda(t)$:

$$P(M(t) \geq x) \geq (1 - e^{-x[1-F(x)]})P(\Lambda(t) \geq x).$$

В частности, если известно, что функция распределения $F(x)$ имеет правый хвост, убывающий не быстрее степенной функции, то есть

$$1 - F(x) \geq \frac{C}{x^\alpha}$$

для $x \geq x_0$ при некоторых $x_0 > 0$, $C > 0$ и $\alpha > 0$, то получим соотношение

$$P(M(t) \geq x) \geq (1 - e^{-Cx^{1-\alpha}})P(\Lambda(t) \geq x),$$

справедливое для $x \geq x_0$.

Наконец, из второго неравенства, приведенного в утверждении 8.2, для любого $\alpha > 0$ при $z = x^\alpha > 0$ получаем оценку

$$\log \frac{P(\Lambda(t) < x^\alpha)}{P(M(t) < x)} \leq x^\alpha [1 - F(x)].$$

В частности, используя уже упомянутое представление моментов неотрицательных случайных величин, в случае когда случайные величины $X_1, X_2, \dots$ положительны и имеют конечный момент порядка α при некотором $\alpha \geq 0$, из последней оценки получаем неравенство

$$\frac{1}{\alpha + 1} \int_0^\infty \log \frac{P(\Lambda(t) < x^\alpha)}{P(M(t) < x)} dx \leq EX_1^{\alpha+1}.$$

При решении практических задач полезным может оказаться следующее утверждение, очевидным образом вытекающее из приведенного выше представления функции распределения максимизированного процесса Кокса.

Утверждение 8.3. Пусть $t > 0$ и $x \in \mathbb{R}$ произвольны. Пусть Z – случайная величина со стандартным показательным распределением, независимая от случайной величины $\Lambda(t)$. Тогда

$$P(M(t) < x) = P\left(\Lambda(t) < \frac{Z}{1-F(x)}\right).$$

Из этого утверждения, в частности, вытекает, что если $M^{(\lambda)}(t)$ – макс-обобщенный пуассоновский процесс с интенсивностью λ , то

$$P(M^{(\lambda)}(t) < x) = P(Z > \lambda t [1 - F(x)]),$$

где, как и в утверждении 8.1, Z – случайная величина со стандартным показательным распределением.

8.2 Предельные распределения для макс-обобщенных дважды стохастических пуассоновских процессов

В этом разделе мы докажем несколько предельных теорем, описывающих асимптотическое поведение макс-обобщенных процессов Кокса и опишем класс предельных распределений.

Общую функцию распределения случайных величин X_j будем обозначать $F(x)$. Мы также будем использовать обозначения

$$\text{lex}(F) = \inf\{x : F(x) > 0\}, \quad \text{rex}(F) = \sup\{x : F(x) < 1\}.$$

Теорема 8.1. *Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что*

$$\frac{\Lambda(t)}{d(t)} \Longrightarrow \Lambda \quad (t \rightarrow \infty). \quad (8.2)$$

Предположим также, что $\text{rex}(F) = \infty$ и существует положительное число γ такое, что для любого $x > 0$

$$\lim_{y \rightarrow \infty} \frac{1 - F(yx)}{1 - F(y)} = x^{-\gamma}. \quad (8.3)$$

Тогда существуют положительная функция $b(t)$ и функция распределения $H_1(x)$ такие, что

$$P\left(\frac{1}{b(t)} \max_{1 \leq k \leq N(t)} X_k < x\right) \Longrightarrow H_1(x) \quad (t \rightarrow \infty).$$

При этом

$$H_1(x) = \begin{cases} 0, & x \leq 0, \\ \int_0^\infty e^{-\lambda x^{-\gamma}} dP(\Lambda < \lambda), & x > 0, \end{cases}$$

а функция $b(t)$ может быть определена как

$$b(t) = \inf \left\{ x : 1 - F(x) \leq \frac{1}{d(t)} \right\}. \quad (8.4)$$

Доказательство. Пусть $\varphi_{N(t)}(z)$ – производящая функция случайной величины $N(t)$,

$$\varphi_{N(t)}(z) = \sum_{n=0}^{\infty} P(N(t) = n) z^n = \int_0^\infty e^{\lambda(z-1)} dP(\Lambda(t) < \lambda), \quad |z| \leq 1.$$

Тогда по формуле полной вероятности для любого $x \in \mathbb{R}$ мы имеем

$$\begin{aligned} P \left(\max_{1 \leq k \leq N(t)} X_k < x \right) &= \sum_{n=0}^{\infty} P(N(t) = n) P \left(\max_{1 \leq k \leq n} X_k < x \right) = \\ &= \sum_{n=0}^{\infty} P(N(t) = n) F^n(x) = \varphi_{N(t)}(F(x)) = \int_0^\infty e^{\lambda(F(x)-1)} dP(\Lambda(t) < \lambda). \end{aligned}$$

Поэтому

$$\begin{aligned} P \left(\frac{1}{b(t)} \max_{1 \leq k \leq N(t)} X_k < x \right) &= \\ &= \int_0^\infty \exp \{ -\lambda d(t) [1 - F(b(t)x)] \} dP(\Lambda(t) < \lambda d(t)). \end{aligned} \quad (8.5)$$

Покажем, что для произвольного x

$$\lim_{t \rightarrow \infty} d(t) [1 - F(b(t)x)] = \begin{cases} x^{-\gamma}, & x > 0, \\ \infty, & x \leq 0. \end{cases} \quad (8.6)$$

В силу условий (8.4) и $\text{rext}(F) = \infty$ функция $b(t)$ неограниченно возрастает при $t \rightarrow \infty$. Поэтому $F(b(t)x) \rightarrow 0$ при $t \rightarrow \infty$ для любого фиксированного $x < 0$, то есть $1 - F(b(t)x) \rightarrow 1$. Поскольку функция $d(t)$ неограниченно возрастает, справедливо второе из соотношений (8.6). По той же причине это соотношение тривиально верно и при $x = 0$. Пусть теперь $x > 0$. В силу условия (8.3) мы имеем

$$\begin{aligned} \lim_{t \rightarrow \infty} d(t)[1 - F(b(t)x)] &= \lim_{t \rightarrow \infty} d(t)[1 - F(b(t))] \cdot \frac{1 - F(b(t)x)}{1 - F(b(t))} = \\ &= x^{-\gamma} \lim_{t \rightarrow \infty} d(t)[1 - F(b(t))]. \end{aligned}$$

Поэтому соотношение (8.6) будет доказано, если мы убедимся, что

$$\lim_{t \rightarrow \infty} d(t)[1 - F(b(t))] = 1. \quad (8.7)$$

Следуя идеям, использованным при доказательстве Теоремы 1.1 в (Berman, 1964), заметим, что из определения (8.4) вытекает, что

$$1 - F(b(t) + 0) \leq \frac{1}{d(t)} \leq 1 - F(b(t))$$

или

$$1 \leq d(t)[1 - F(b(t))] \leq \frac{1 - F(b(t))}{1 - F(b(t) + 0)}. \quad (8.8)$$

В силу того, что функция $1 - F(x)$ убывает, для произвольного $y > 1$ имеем

$$1 - F(b(t) + 0) \geq 1 - F(b(t)y).$$

Используя это неравенство в соотношении (8.8) и условие (8.3), для произвольного $y > 1$ получаем

$$1 \leq d(t)[1 - F(b(t))] \leq \frac{1 - F(b(t))}{1 - F(b(t)y)} \leq (1 - \varepsilon(t))y^\gamma, \quad (8.9)$$

где $\varepsilon(t) \rightarrow 0$ при $t \rightarrow \infty$. Так как $y > 1$ выбрано произвольно, из (8.9) вытекает (8.7), и следовательно, соотношение (8.6) доказано. Обозначим

$$h_t(x) = \exp\{-d(t)[1 - F(b(t)x)]\}, \quad h(x) = \exp\{-x^{-\gamma}\}.$$

Несложно видеть, что $0 \leq h_t(x) \leq 1$ и $0 < h(x) \leq 1$ при $x > 0$. Из (8.6) при этом следует, что

$$h_t(x) \rightarrow h(x) \quad (t \rightarrow \infty) \quad (8.10)$$

при любом фиксированном $x > 0$. Обозначим

$$A_t(\lambda) = P(\Lambda(t) < \lambda d(t)), \quad A(\lambda) = P(\Lambda < \lambda).$$

Фиксируем произвольное $x > 0$. В силу (8.5) имеем

$$\begin{aligned} & \left| P \left(\max_{1 \leq k \leq N(t)} X_k < xb(t) \right) - \int_0^\infty \exp \{ -\lambda x^{-\gamma} \} dP(\Lambda < \lambda) \right| = \\ & = \left| \int_0^\infty h_t^\lambda(x) dA_t(\lambda) - \int_0^\infty h^\lambda(x) dA(\lambda) \right| \leq \\ & \leq \left| \int_0^\infty h_t^\lambda(x) dA_t(\lambda) - \int_0^\infty h^\lambda(x) dA_t(\lambda) \right| + \\ & + \left| \int_0^\infty h^\lambda(x) dA_t(\lambda) - \int_0^\infty h^\lambda(x) dA(\lambda) \right| \equiv I_1(t) + I_2(t). \end{aligned}$$

Вначале рассмотрим $I_2(t)$. Поскольку функция $\psi_z(\lambda) = z^\lambda$ ограничена и непрерывна по λ при $|z| \leq 1$, в силу условия (8.2) $I_2(t) \rightarrow 0$ при $t \rightarrow \infty$ по определению слабой сходимости, так как

$$\int_0^\infty h^\lambda(x) dA_t(\lambda) = E\psi(\Lambda(t)/d(t)) \rightarrow E\psi(\Lambda) = \int_0^\infty h^\lambda(x) dA(\lambda)$$

при $t \rightarrow \infty$ для любой непрерывной и ограниченной функции $\psi(\lambda)$, в том числе и для $\psi(\lambda) = \psi_{h(x)}(\lambda) = h^\lambda(x)$. Таким образом, нам достаточно убедиться, что с помощью выбора большого t можно сделать $I_1(t)$ произвольно малым. Пусть $\varepsilon > 0$ – произвольное малое число. Для любого положительного M имеем

$$I_1(t) \leq \int_0^M |h_t^\lambda(x) - h^\lambda(x)| dA_t(\lambda) +$$

$$+ \int_M^\infty |h_t^\lambda(x) - h^\lambda(x)| dA_t(\lambda) \equiv I_{11}(t) + I_{12}(t). \quad (8.11)$$

В силу слабой компактности семейства функций распределения $\{A_t(x)\}_{t>0}$ на бесконечности, обусловленной соотношением (8.2), существует такое $M = M(\varepsilon) > 0$, что

$$I_{12}(t) \leq 2 \int_M^\infty dA_t(\lambda) \leq 2 \sup_{t>0} [1 - A_t(M)] < \varepsilon. \quad (8.12)$$

Рассмотрим $I_{11}(t)$. Согласно Лемме 5.3, мы имеем

$$|\Psi(a) - \Psi(b)| \leq |b - a| \cdot \sup_{0 \leq \alpha \leq 1} |\Psi'(\alpha a + (1 - \alpha)b)|,$$

где $\Psi(z)$ – функция, непрерывная на отрезке $[a, b]$ и дифференцируемая во всех внутренних точках этого отрезка. Полагая $\Psi(z) = z^\lambda$, при каждом $\lambda > 0$ мы получаем неравенство

$$\begin{aligned} |h_t^\lambda(x) - h^\lambda(x)| &\leq \lambda |h_t(x) - h(x)| \sup_{0 \leq \alpha \leq 1} |\alpha h_t(x) + (1 - \alpha)h(x)|^{\lambda-1} \leq \\ &\leq \frac{\lambda |h_t(x) - h(x)|}{\min\{|h_t(x)|, |h(x)|\}}. \end{aligned}$$

Функция $h(x)$ ни при каком $x \in \mathbb{R}$ не обращается в нуль. В силу (8.10) найдутся такие $\delta > 0$ и t_0 , что $\min\{h_t(x), h(x)\} \geq \delta$ для всех $t \geq t_0$. Поэтому для всех t начиная с t_0 справедливо неравенство

$$I_{11}(t) \leq \frac{M}{\delta} |h_t(x) - h(x)|.$$

Таким образом, (8.10) позволяет выбрать $t_1 = t_1(\varepsilon) \geq t_0$ столь большим, чтобы

$$I_{11}(t) < \varepsilon \quad (8.13)$$

для всех $t \geq t_1$. Из (8.11), (8.12) и (8.13) следует, что $I_1(t) < 2\varepsilon$ при $t \geq t_1$. Теорема доказана.

Пример 8.3. (Королев, Горшенин, 2017). Предположим, что в (8.2) случайная величина Λ имеет гамма-распределение с параметром формы $r > 0$ и параметром масштаба $\mu > 0$:

$$P(\Lambda < x) = \frac{\mu^r}{\Gamma(r)} \int_0^x y^{r-1} e^{-\mu y} dy, \quad x \geq 0.$$

Это может быть, например, если $N(t)$ имеет отрицательное биномиальное распределение с параметрами r и $p_t = \min\{1, \mu/t\}$. Тогда $H_1(x) = 0$ при $x \leq 0$, а при $x > 0$

$$\begin{aligned} H_1(x) &\equiv H_1(x; r, \gamma, \mu) = \frac{\mu^r}{\Gamma(r)} \int_0^\infty e^{-z(\mu+x^{-\gamma})} z^{r-1} dz = \\ &= \frac{\mu^r}{\Gamma(r)(\mu+x^{-\gamma})^r} \int_0^\infty e^{-z} z^{r-1} dz = \left(\frac{\mu x^\gamma}{1 + \mu x^\gamma} \right)^r. \end{aligned}$$

Необходимо заметить, что в таком случае предельное распределение с хвостами, убывающими степенным образом, получено как масштабная смесь распределения Фреше (предельное распределение экстремальных значений первого типа), в которой смешивающим является гамма-распределение. Пусть $Q_{r,1}$ – случайная величина, имеющая распределение Снедекора–Фишера с параметрами (“степенями свободы”) $r, 1$, определяемое плотностью

$$f_{r,1}(x) = \frac{r^{r+1} x^{r-1}}{(1+rx)^{r+1}}, \quad x \geq 0$$

(см., например, (Большев, Смирнов, 1965), раздел 27). В статьях (Королев, Горшенин, 2017), (Korolev, Gorshenin, 2018) продемонстрировано, что

$$H_1(x; r, \gamma, \mu) = P(Q_{r,1} < \mu r x^\gamma), \quad x \geq 0,$$

то есть предельная функция распределения $H_1(x; r, \gamma, \mu)$ с точностью до степенного преобразования аргумента x совпадает с функцией распределения Снедекора–Фишера. В статистике распределения с аргументами, подвергнутыми степенному преобразованию,

принято называть *темперированными* (tempered). Поэтому в рамках данного примера имеются вполне серьезные основания называть распределение $H_1(x; r, \gamma, \mu)$ *темперированным распределением Снедекора–Фишера*.

Плотность, соответствующая предельной функции распределения $H_1(x; r, \gamma, \mu)$, имеет вид

$$p(x; r, \gamma, \mu) = \frac{r\gamma\mu^r x^{\gamma r - 1}}{(1 + \mu x^\gamma)^{r+1}} = \frac{\gamma r \mu^r}{x^{1+\gamma}(\mu + x^{-\gamma})^{r+1}}, \quad x > 0.$$

Некоторые свойства распределения $H_1(x; r, \gamma, \mu)$ были рассмотрены в работе (Королев, Горшенин, 2017). В частности, в этой работе показано, что моменты случайной величины, имеющей распределение $H_1(x; r, \gamma, \mu)$, конечны, только если $\delta < \gamma$. Более того, в (Королев, Горшенин, 2017) приведены явные выражения для моментов функции распределения $H_1(x; r, \gamma, \mu)$: если $0 < \delta < \gamma < \infty$, то

$$\int_0^\infty x^\delta p(x; r, \gamma, \mu) dx = \frac{\Gamma(r + \frac{\delta}{\gamma}) \Gamma(1 - \frac{\delta}{\gamma})}{\mu^{\delta/\gamma} \Gamma(r)}.$$

Теорема 8.2. *Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что имеет место сходимость (8.2). Предположим также, что $\text{rtext}(F) < \infty$ и функция распределения $G_F(x) = F(\text{rtext}(F) - x^{-1})$ удовлетворяет условию (8.3), то есть существует положительное число γ такое, что для любого $x > 0$*

$$\lim_{y \rightarrow \infty} \frac{1 - G_F(yx)}{1 - G_F(y)} = x^{-\gamma}. \quad (8.14)$$

Тогда существуют функции $a(t)$ и $b(t)$ и функция распределения $H_2(x)$ такие, что

$$\mathbb{P} \left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < x \right) \implies H_2(x) \quad (t \rightarrow \infty).$$

При этом

$$H_2(x) = \begin{cases} 1, & x > 0, \\ \int_0^\infty e^{-\lambda|x|^\gamma} dP(\Lambda < \lambda), & x \leq 0, \end{cases}$$

а функции $a(t)$ и $b(t)$ могут быть определены как

$$a(t) = \text{rext}(F), \quad b(t) = \text{rext}(F) - \inf \left\{ x : 1 - F(x) \leq \frac{1}{d(t)} \right\}.$$

Доказательство. Применим Теорему 8.1 к макс-обобщенному процессу Кокса

$$M(t) = \max_{1 \leq k \leq N(t)} X_k^*,$$

в котором $P(X_1^* < x) = G_F(x)$, $x > 0$. Учитывая, что $\text{rext}(G_F) = \infty$ и G_F удовлетворяет соотношению (8.14), по Теореме 8.1 для $x > 0$ мы получаем соотношение

$$\begin{aligned} & \lim_{t \rightarrow \infty} \varphi_{N(t)}(G_F(b_F(t)x)) = \\ & = \lim_{t \rightarrow \infty} \varphi_{N(t)} \left(F \left(\text{rext}(F) - \frac{1}{b_F(t)x} \right) \right) = H_1(x), \end{aligned} \quad (8.15)$$

где

$$\begin{aligned} b_F(t) &= \inf \left\{ x : 1 - G_F(x) \leq \frac{1}{d(t)} \right\} = \\ &= \inf \left\{ x : 1 - F \left(\text{rext}(F) - \frac{1}{x} \right) \leq \frac{1}{d(t)} \right\} = \\ &= \left[\text{rext}(F) - \inf \left\{ x : 1 - F(x) \leq \frac{1}{d(t)} \right\} \right]^{-1}, \end{aligned}$$

а функция $H_1(x)$ определена в Теореме 8.1. Следовательно, взяв $a(t) = \text{rext}(F)$ и $b(t) = (b_F(t))^{-1}$, из (8.15) получим, что

$$\lim_{t \rightarrow \infty} P \left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < \frac{1}{-x} \right) =$$

$$= \lim_{t \rightarrow \infty} \varphi_{N(t)} \left(F \left(a(t) - \frac{b(t)}{x} \right) \right) = H_1(x),$$

если $x > 0$, или, что то же самое,

$$\begin{aligned} \lim_{t \rightarrow \infty} P \left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < x \right) = \\ = \lim_{t \rightarrow \infty} \varphi_{N(t)} (F(a(t) + b(t)x)) = H_1 \left(\frac{1}{-x} \right) = H_2(x) \end{aligned}$$

для $x < 0$. Для завершения доказательства осталось заметить, что $F(a(t) + b(t)x) \equiv 1$ при $x > 0$, поскольку $a(t) = \text{rext}(F)$ и $b(t) > 0$. Поэтому при $x > 0$

$$\lim_{t \rightarrow \infty} \varphi_{N(t)} (F(a(t) + b(t)x)) = 1.$$

Теорема доказана.

Пример 8.4. Пусть, как и в примере 8.3, случайная величина Λ имеет гамма-распределение с параметром формы $r > 0$ и параметром масштаба $\mu > 0$. Тогда $H_2(x) = 1$ при $x > 0$, тогда как при $x \leq 0$

$$\begin{aligned} H_2(x) = H_2(x; r, \gamma, \mu) = \\ = \frac{\mu^r}{\Gamma(r)} \int_0^\infty \exp\{-\lambda(|x|^\gamma + \mu)\} \lambda^{r-1} d\lambda = \left(\frac{\mu}{\mu + |x|^\gamma} \right)^r. \end{aligned}$$

Как и в примере 8.3, у случайной величины с распределением $H_2(x; r, \gamma, \mu)$ отсутствуют моменты порядков $\delta \geq \gamma$.

Теорема 8.3. *Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что имеет место сходимость (8.2). Пусть*

$$\int_a^{\text{rext}(F)} [1 - F(z)] dz < \infty$$

при некотором конечном a и для $y \in (\text{lex}(F), \text{rxt}(F))$ функция $R(y)$ определена как

$$R(y) = \frac{1}{1 - F(y)} \int_y^{\text{rxt}(F)} [1 - F(z)] dz.$$

Предположим, что для любого $x \in \mathbb{R}$ существует предел

$$\lim_{y \rightarrow \text{rxt}(F)} \frac{1 - F(y + xR(y))}{1 - F(y)} = e^{-x}. \quad (8.16)$$

Тогда существуют функции $a(t)$, $b(t) > 0$ и $H_3(x)$ такие, что

$$\mathbb{P} \left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < x \right) \Rightarrow H_3(x) \quad (t \rightarrow \infty).$$

При этом

$$H_3(x) = \int_0^{\infty} \exp\{-\lambda e^{-x}\} d\mathbb{P}(\Lambda < \lambda), \quad x \in \mathbb{R},$$

а функции $a(t)$ и $b(t)$ могут быть выбраны в виде

$$a(t) = \inf \left\{ x : 1 - F(x) \leq \frac{1}{d(t)} \right\}, \quad b(t) = R(a(t)). \quad (8.17)$$

Доказательство. Как и при доказательстве Теоремы 8.1, убеждаемся, что

$$\begin{aligned} & \mathbb{P} \left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < x \right) = \\ &= \int_0^{\infty} \exp\{-\lambda d(t)[1 - F(a(t) + b(t)x)]\} d\mathbb{P}(\Lambda(t) < \lambda d(t)). \end{aligned}$$

Заметим, что функции $a(t)$, определенные в (8.17), стремятся к $\text{rxt}(F)$ при $t \rightarrow \infty$. Поэтому из условия (8.16) вытекает, что

$$\lim_{t \rightarrow \infty} \frac{1 - F(a(t) + b(t)x)}{1 - F(a(t))} = e^{-x} \quad (8.18)$$

для всех $x \in \mathbb{R}$. Поэтому для произвольного фиксированного $x \in \mathbb{R}$ справедливы соотношения

$$\begin{aligned} & \lim_{t \rightarrow \infty} d(t)[1 - F(a(t) + b(t)x)] = \\ &= \lim_{t \rightarrow \infty} d(t)[1 - F(a(t))] \cdot \frac{1 - F(a(t) + b(t)x)}{1 - F(a(t))} = \\ &= e^{-x} \lim_{t \rightarrow \infty} d(t)[1 - F(a(t))]. \end{aligned}$$

Покажем, что

$$\lim_{t \rightarrow \infty} d(t)[1 - F(a(t))] = 1. \quad (8.19)$$

Из определения функции $a(t)$ вытекает, что

$$1 - F(a(t) + 0) \leq \frac{1}{d(t)} \leq 1 - F(a(t)).$$

С другой стороны, для любого $\varepsilon > 0$ справедливо неравенство

$$1 - F(a(t) + \varepsilon b(t)) \leq 1 - F(a(t) + 0).$$

Таким образом, из (8.18) вытекает, что

$$1 \leq d(t)[1 - F(a(t))] \leq \frac{1 - F(a(t))}{1 - F(a(t) + \varepsilon b(t))} \rightarrow e^\varepsilon \quad (8.20)$$

при $t \rightarrow \infty$. Поскольку соотношение (8.20) справедливо для как угодно малого $\varepsilon > 0$, соотношение (8.19) доказано. Дальнейшее доказательство практически дословно совпадает с заключительной частью доказательства Теоремы 8.1. Теорема доказана.

Пример 8.5. Снова рассмотрим ситуацию, в которой случайная величина Λ имеет гамма-распределение с параметром формы $r > 0$ и параметром масштаба $\mu > 0$. Тогда

$$\begin{aligned} H_3(x) &= H_3(x; r, \mu) = \\ &= \frac{\mu^r}{\Gamma(r)} \int_0^\infty \exp\{-\lambda(e^{-x} + \mu)\} \lambda^{r-1} d\lambda = \left(\frac{\mu e^x}{\mu e^x + 1} \right)^r. \end{aligned}$$

Если $r = 1$, то $H_3(x; 1, \mu)$ является “чистым” логистическим распределением. Тот факт, что логистическое распределение может быть предельным для экстремальных порядковых статистик в выборках случайного объема, имеющего геометрическое распределение, был отмечен почти сорок лет назад в работах Б.В. Гнеденко и его учеников (Гнеденко, Гнеденко, 1982) (Гнеденко, Сенуси-Берекси, 1982).

Согласно классической асимптотической теории экстремальных значений, функции распределения

$$H_{1,\gamma}(x) = \begin{cases} \exp\{-x^{-\gamma}\}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0, \end{cases}$$

$$H_{2,\gamma}(x) = \begin{cases} \exp\{-|x|^\gamma\}, & \text{если } x \leq 0, \\ 1, & \text{если } x > 0, \end{cases}$$

$$H_{3,0}(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R},$$

исчерпывают все возможные типы предельных законов для нормированных максимумов независимых одинаково распределенных случайных величин. Следуя Мизесу (Mises, 1936) и Дженкинсону (Jenkinson, 1955), эти три типа распределений можно записать единообразно. Действительно, рассмотрим функцию

$$H_\tau(x) = \exp\left\{-\frac{1}{(1 + \tau x)^{1/\tau}}\right\}.$$

Для $\tau = 0$ доопределим функцию $H_\tau(x)$ ее предельным значением при $\tau \rightarrow 0$. Тогда с точностью до изменения параметров сдвига и масштаба,

$$H_\tau(x) = \begin{cases} H_{1,\gamma}(x), & \text{при } \tau > 0, \\ H_{2,\gamma}(x), & \text{при } \tau < 0, \\ H_{3,0}(x), & \text{при } \tau = 0, \end{cases}$$

причем $\gamma = |1/\tau|$, если $\tau \neq 0$. (Галамбош в своей известной монографии (Галамбош, 1984) приписывает эту единую форму записи

предельных законов Мизесу в то время как Гумбель в своей не менее известной книге (Гумбель, 1965) указывает на то, что эта форма предложена Дженкинсоном. Справедливости ради надо отметить, что в отличие от записи Мизеса, Дженкинсон использует два параметра (помимо τ еще и параметр масштаба, который может, однако, принимать отрицательные значения)).

Определение 8.2. Будем говорить, что функция распределения $F(x)$ принадлежит к области тах-притяжения закона $H_\tau(x)$, если существует последовательность центрирующих и нормирующих констант, обеспечивающих сходимость последовательных максимумов независимых случайных величин с общей функцией распределения $F(x)$ к функции распределения $H_\tau(x)$.

Принадлежность F к области тах-притяжения закона H_τ будет обозначаться записью $F \in \text{max-DA}(H_\tau)$.

Согласно классической асимптотической теории экстремумов, $F \in \text{max-DA}(H_{1,\gamma})$ тогда и только тогда, когда F удовлетворяет условиям Теоремы 8.1, $F \in \text{max-DA}(H_{2,\gamma})$ тогда и только тогда, когда F удовлетворяет условиям Теоремы 8.2, $F \in \text{max-DA}(H_{3,0})$ тогда и только тогда, когда F удовлетворяет условиям Теоремы 8.3. Более того, если функция распределения F не принадлежит ни к одной из упомянутых областей тах-притяжения, то распределения экстремумов в выборке из такого распределения не сходятся ни к одной невырожденной функции распределения ни при каком выборе центрирующих и нормирующих констант, обеспечивающих линейную нормировку экстремумов (см., например, (Галамбош, 1984)).

Лемма 8.3. Если произвольная положительная функция $d(t)$, $t \geq 0$, неограниченно возрастает и $F \in \text{max-DA}(H_\tau)$, то найдутся функции $a(t)$ и $b(t) > 0$ такие, что

$$P\left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq d(t)} X_k - a(t) \right) < x\right) \implies H_\tau(x) \quad (t \rightarrow \infty).$$

Это утверждение доказывается точно так же, как теоремы 1.1 – 1.3 в (Барлоу и Прошан, 1969). При этом взаимосвязь функций $a(t)$

и $b(t)$ с $F(x)$ и $d(t)$ полностью аналогична описанной в Теоремах 8.1 – 8.3.

Из Теорем 8.1 – 8.3 мы получаем следующее общее утверждение, которое является теоремой переноса для тах-обобщенных процессов Кокса.

Теорема 8.4. Пусть $F \in \text{тах-DA}(H_\tau)$. Предположим, что существуют неограниченно возрастающая функция $d(t)$ и случайная величина Λ такие, что имеет место сходимость (8.2). Тогда существуют функции $a(t)$ и $b(t) > 0$, обеспечивающие сходимость

$$P\left(\frac{1}{b(t)} \left(\max_{1 \leq k \leq N(t)} X_k - a(t) \right) < x\right) \Longrightarrow \int_0^\infty H_\tau^\lambda(x) dP(\Lambda < \lambda) \quad (t \rightarrow \infty).$$

8.3 Предельные теоремы для мин-обобщенных дважды стохастических пуассоновских процессов

В этом разделе мы сформулируем утверждения об асимптотическом поведении мин-обобщенных процессов Кокса, аналогичные предельным теоремам, приведенным выше и описывающим асимптотическое поведение макс-обобщенных процессов Кокса. Доказательства утверждений данного раздела аналогичны доказательствам соответствующих результатов для макс-обобщенных процессов Кокса, поскольку $\min\{X_1, \dots, X_{N(t)}\} = -\max\{-X_1, \dots, -X_{N(t)}\}$.

Теорема 8.5. Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что

$$\frac{\Lambda(t)}{d(t)} \Longrightarrow \Lambda \quad (t \rightarrow \infty). \quad (8.21)$$

Предположим также, что $\text{lex}(F) = -\infty$ и существует положительное число γ такое, что для любого $x > 0$

$$\lim_{y \rightarrow -\infty} \frac{F(yx)}{F(y)} = x^{-\gamma}. \quad (8.22)$$

Тогда существуют положительная функция $b(t)$ и функция распределения $L_1(x)$ такие, что

$$P\left(\frac{1}{b(t)} \min_{1 \leq k \leq N(t)} X_k < x\right) \Longrightarrow L_1(x) \quad (t \rightarrow \infty).$$

При этом

$$L_1(x) = \begin{cases} 1, & x > 0, \\ \int_0^\infty (1 - e^{-\lambda|x|^{-\gamma}}) dP(\Lambda < \lambda), & x \leq 0, \end{cases}$$

а функция $b(t)$ может быть определена как

$$b(t) = \sup \left\{ x : F(x) \leq \frac{1}{d(t)} \right\}.$$

Пример 8.6. Предположим, что в (8.21) случайная величина Λ имеет стандартное показательное распределение. Тогда $L_1(x) = 1$ для $x > 0$, а для $x \leq 0$

$$L_1(x) = 1 - \int_0^\infty \exp \left\{ -\lambda \left(1 + \frac{1}{|x|^\gamma} \right) \right\} d\lambda = \frac{1}{1 + |x|^\gamma}.$$

Этой функции распределения соответствует плотность

$$l_1(x) = \begin{cases} \frac{\gamma|x|^{\gamma-1}}{(1 + |x|^\gamma)^2}, & x \leq 0 \\ 0, & x > 0. \end{cases}$$

Легко видеть, что у такого распределения $L_1(x)$ отсутствуют моменты порядка $\beta \geq \gamma + 1$.

Теорема 8.6. Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что имеет место сходимость (8.21). Предположим также, что $\text{lex}(F) > -\infty$ и функция распределения

$W_F(x) = F(\text{lex}(F) - x^{-1})$ удовлетворяет условию (8.22), то есть существует положительное число γ такое, что для любого $x > 0$

$$\lim_{y \rightarrow \infty} \frac{W_F(yx)}{W_F(y)} = x^{-\gamma}.$$

Тогда существуют функции $a(t)$ и $b(t)$ и функция распределения $L_2(x)$ такие, что

$$P\left(\frac{1}{b(t)} \left(\min_{1 \leq k \leq N(t)} X_k - a(t)\right) < x\right) \Rightarrow L_2(x) \quad (t \rightarrow \infty).$$

При этом

$$L_2(x) = \begin{cases} 0, & x \leq 0, \\ \int_0^\infty (1 - e^{-\lambda x^\gamma}) dP(\Lambda < \lambda), & x > 0, \end{cases}$$

а функции $a(t)$ и $b(t)$ могут быть определены как

$$a(t) = \text{lex}(F), \quad b(t) = \sup \left\{ x : F(x) \leq \frac{1}{d(t)} \right\} - \text{lex}(F).$$

Пример 8.7. Пусть как и в примере 8.6, случайная величина Λ имеет стандартное показательное распределение. Тогда $L_2(x) = 0$ для $x \leq 0$, а для $x > 0$

$$L_2(x) = 1 - \int_0^\infty \exp\{-\lambda(x^\gamma + 1)\} d\lambda = \frac{x^\gamma}{1 + x^\gamma}.$$

Как и в примере 8.6, у такого распределения отсутствуют моменты порядка $\beta \geq \gamma + 1$.

Теорема 8.7. Предположим, что существуют положительная неограниченно возрастающая функция $d(t)$ и неотрицательная случайная величина Λ такие, что имеет место сходимость (8.21). Пусть

$$\int_{\text{lex}(F)}^a F(z) dz < \infty$$

при некотором конечном a и для $y \in (\text{lex}(F), \text{rex}(F))$ функция $r(y)$ определена как

$$r(y) = \frac{1}{F(y)} \int_{\text{lex}(F)}^y F(z) dz.$$

Предположим, что для любого $x \in \mathbb{R}$ существует предел

$$\lim_{y \rightarrow \text{lex}(F)} \frac{F(y + xr(y))}{F(y)} = e^x.$$

Тогда существуют функции $a(t)$, $b(t) > 0$ и $L_3(x)$ такие, что

$$P\left(\frac{1}{b(t)} \left(\min_{1 \leq k \leq N(t)} X_k - a(t)\right) < x\right) \implies L_3(x) \quad (t \rightarrow \infty).$$

При этом

$$L_3(x) = \int_0^\infty (1 - \exp\{-\lambda e^x\}) dP(\Lambda < \lambda), \quad x \in \mathbb{R},$$

а функции $a(t)$ и $b(t)$ могут быть выбраны в виде

$$a(t) = \sup \left\{ x : F(x) \leq \frac{1}{d(t)} \right\}, \quad b(t) = R(a(t)).$$

Пример 8.8. Снова рассмотрим случай, когда Λ имеет стандартное показательное распределение. В такой ситуации

$$L_3(x) = 1 - \int_0^\infty \exp\{-\lambda(e^x + 1)\} d\lambda = \frac{1}{1 + e^{-x}}, \quad x \in \mathbb{R},$$

то есть распределение, предельное для мин-обобщенного процесса Кокса, является логистическим и совпадает с предельным распределением для макс-обобщенного процесса Кокса из примера 8.5.

Согласно классической асимптотической теории экстремальных значений, функции распределения

$$L_{1,\gamma}(x) = \begin{cases} 1, & \text{если } x > 0, \\ 1 - \exp\{-|x|^{-\gamma}\}, & \text{если } x \leq 0, \end{cases}$$

$$L_{2,\gamma}(x) = \begin{cases} 1 - \exp\{-x^\gamma\}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0, \end{cases}$$

$$L_{3,0}(x) = 1 - \exp\{-e^x\}, \quad x \in \mathbb{R},$$

исчерпывают все возможные типы предельных законов для нормированных минимумов независимых одинаково распределенных случайных величин. По аналогии с тем, как это было сделано в предыдущем разделе, эти три типа распределений можно записать единообразно. Действительно, рассмотрим функцию

$$L_\tau(x) = 1 - \exp\left\{-\frac{1}{(1-\tau x)^{1/\tau}}\right\}.$$

Для $\tau = 0$ доопределим функцию $L_\tau(x)$ ее предельным значением при $\tau \rightarrow 0$. Тогда с точностью до изменения параметров сдвига и масштаба,

$$L_\tau(x) = \begin{cases} L_{1,\gamma}(x), & \text{при } \tau > 0, \\ L_{2,\gamma}(x), & \text{при } \tau < 0, \\ L_{3,0}(x), & \text{при } \tau = 0, \end{cases}$$

причем $\gamma = |1/\tau|$, если $\tau \neq 0$. Легко видеть, что

$$L_\tau(x) = 1 - H_\tau(-x),$$

где семейство $\{H_\tau(x)\}$ определено в разделе 8.2.

Определение 8.3. Будем говорить, что функция распределения $F(x)$ принадлежит к области min-притяжения закона $L_\tau(x)$, если существует последовательность центрирующих и нормирующих констант, обеспечивающих сходимость последовательных минимумов независимых случайных величин с общей функцией распределения $F(x)$ к функции распределения $L_\tau(x)$.

Принадлежность F к области min-притяжения закона L_τ будет обозначаться записью $F \in \min\text{-}DA(L_\tau)$.

Согласно классической асимптотической теории экстремумов, $F \in \min\text{-}DA(L_{1,\gamma})$ тогда и только тогда, когда F удовлетворяет условиям Теоремы 8.5, $F \in \min\text{-}DA(L_{2,\gamma})$ тогда и только тогда, когда

F удовлетворяет условиям Теоремы 8.6, $F \in \min\text{-}DA(L_{3,0})$ тогда и только тогда, когда F удовлетворяет условиям Теоремы 8.7. Более того, если функция распределения F не принадлежит ни к одной из упомянутых областей $\min$ -притяжения, то распределения *линейно* нормированных минимумов в выборке из такого распределения не сходятся ни к одной невырожденной функции распределения ни при каком выборе центрирующих и нормирующих констант.

Из Теорем 8.5 – 8.7 мы получаем следующее общее утверждение, которое является теоремой переноса для $\min$ -обобщенных процессов Кокса.

Теорема 8.8. Пусть $F \in \min\text{-}DA(L_\tau)$. Предположим, что существуют неограниченно возрастающая функция $d(t)$ и случайная величина Λ такие, что имеет место сходимостъ (8.21). Тогда существуют функции $a(t)$ и $b(t) > 0$, обеспечивающие при $t \rightarrow \infty$ сходимостъ

$$P\left(\frac{1}{b(t)} \left(\min_{1 \leq k \leq N(t)} X_k - a(t) \right) < x\right) \implies 1 - \int_0^\infty H_\tau^\lambda(-x) dP(\Lambda < \lambda).$$

Литература

- [1] *Sichel H. S.* Statistical evaluation of diamondiferous deposits // *Journal of South Afr. Inst. Min. Metall.*, 1973. Vol. 76. P. 235–243.
- [2] *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. 2-е изд., перераб. и дополн. – М.: ФИЗМАТЛИТ, 2011. 620 с.
- [3] *Barndorff-Nielsen O. E.* Exponentially decreasing distributions for the logarithm of particle size // *Proc. Roy. Soc. London. Ser. A*, 1977. Vol. A(353). P. 401–419.
- [4] *Barndorff-Nielsen O. E.* Hyperbolic distributions and distributions of hyperbolae // *Scand. J. Statist.*, 1978. Vol. 5. P. 151–157.
- [5] *Barndorff-Nielsen O. E.* Models for non-Gaussian variation, with applications to turbulence // *Proc. Roy. Soc. London. Ser. A*, 1979. Vol. A(368). P. 501–520.
- [6] *Madan D. B., Seneta E.* The variance gamma (V.G.) model for share market return // *Journal of Business*, 1990. Vol. 63. P. 511–524.
- [7] *Eberlein E., Keller U.* Hyperbolic Distributions in Finance // *Bernoulli*, 1995. Vol. 1, No. 3. P. 281–299.

- [8] *Prause K.* Modeling Financial Data Using Generalized Hyperbolic Distributions. – Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1997. Preprint No. 48.
- [9] *Carr P. P., Madan D. B., Chang E. C.* The Variance Gamma process and option pricing // *European Finance Review*, 1998. Vol. 2. P. 79–105.
- [10] *Eberlein E., Keller U., Prause K.* New insights into smile, mispricing and value at risk: the hyperbolic model // *Journal of Business*, 1998. Vol. 71. P. 371–405.
- [11] *Barndorff-Nielsen O. E.* Processes of normal inverse Gaussian type // *Finance and Stochastics*, 1998. Vol. 2. P. 41–18.
- [12] *Eberlein E., Prause K.* The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures. – Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1998. Preprint No. 56.
- [13] *Ширяев А. Н.* Основы стохастической финансовой математики. Т. 1. Факты. Модели. – М.: Фазис, 1998.
- [14] *Eberlein E.* Application of Generalized Hyperbolic Lévy Motions to Finance. – Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1999. Preprint No. 64.
- [15] *Barndorff-Nielsen O. E., Blæsild P., Schmiegel J.* A parsimonious and universal description of turbulent velocity increments // *European Physical Journal*, 2004. Vol. B. 41. P. 345–363.
- [16] *LeCam L.* Maximum likelihood; an introduction // *International Statistical Review*, 1990. Vol. 58. P. 153–171.
- [17] *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. – М.–Л.: ГИТТЛ, 1949.

- [18] *Barndorff-Nielsen O. E., Kent J., Sørensen M.* Normal variance-mean mixtures and z-distributions // *International Statistical Review*, 1982. Vol. 50. No. 2. P. 145–159.
- [19] *Rényi A.* On the central limit theorem for the sum of a random number of independent random variables // *Acta Math. Acad. Sci. Hung.*, 1960. Vol. 11. P. 97–102.
- [20] *Гнеденко Б. В., Фахим Х.* Об одной теореме переноса // *ДАН СССР*, 1969. Т. 187. № 1. С. 15–17.
- [21] *Королев В. Ю., Соколов И. А.* Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // *Информатика и ее применения*, 2012. Т. 6. Вып. 1. С. 2–10.
- [22] *Королев В. Ю.* О взаимосвязи обобщенного распределения Стьюдента и дисперсионного гамма-распределения при статистическом анализе выборок случайного объема // *ДАН*, 2012 (в печати).
- [23] *Королев В. Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // *Теория вероятностей и ее применения*, 2013. Т. 58. Вып. 1 (в печати).
- [24] *Королев В. Ю., Шоргин С. Я.* Математические методы анализа стохастической структуры информационных потоков. – М.: ИПИ РАН, 2011. 130 с.
- [25] *Stacy E. W.* A generalization of the gamma distribution // *Annals of Mathematical Statistics*, 1962. Vol. 33. P. 1187–1192.
- [26] *Гумбель Э.* Статистика экстремальных значений. – М.: Мир, 1965.
- [27] *Kalashnikov V. V.* Geometric Sums: Bounds for Rare Events with Applications. – Dordrecht: Kluwer Academic Publishers, 1997.
- [28] *Королев В. Ю.* О распределении размеров частиц при дроблении // *Информатика и ее применения*, 2009. Т. 3. Вып. 3. С. 60–68.

- [29] *Gnedenko B. V., Korolev V. Yu.* Random Summation: Limit Theorems and Applications. – Boca Raton: CRC Press, 1996.
- [30] *Teicher H.* Identifiability of mixtures // *Ann. Math. Stat.*, 1961. Vol. 32. P. 244–248.
- [31] *Королев В. Ю.* Построение моделей распределений биржевых цен при помощи методов асимптотической теории случайного суммирования // *Обозрение промышленной и прикладной математики. Сер. Финансовая и страховая математика*, 1997. Т. 4. Вып. 1. С. 86–102.
- [32] *Королев В. Ю.* Асимптотические свойства экстремумов обобщенных процессов Кокса и их применение к некоторым задачам финансовой математики // *Теория вероятностей и ее применения*, 2000. Т. 45. Вып. 1. С. 182–194.
- [33] *Bening V., Korolev V.* Generalized Poisson Models and their Applications in Insurance and Finance. – Utrecht: VSP, 2002.
- [34] *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. – М.: Торус Пресс, 2008.
- [35] *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. – М.: Изд-во Московского университета, 2011.
- [36] *Korolev V., Skvortsova N. (Eds.)* Stochastic Models of Structural Plasma Turbulence. – Utrecht: VSP, 2006.
- [37] *Королев В. Ю., Шевцова И. Г., Шоргин С. Я.* О неравенствах типа Берри–Эссеена для пуассоновских случайных сумм // *Информатика и ее применения*, 2011. Т. 5. Вып. 3. С. 64–66.
- [38] *Korolev V., Shevtsova I.* An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // *Scandinavian Actuarial Journal*, 2012. No. 2. P. 81–105. DOI:10.1080/03461238.2010.485370.

- [39] *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. М.-Л.: ГИТТЛ, 1949.
- [40] *Гумбель Э.* Статистика экстремальных значений. – М.: Мир, 1965.
- [41] *Закс Л. М., Королев В. Ю.* Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // Информатика и ее применения, 2013. Т. 7. Вып. 1. С. 105–115.
- [42] *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий // М.: Торус, 2008.
- [43] *Королев В. Ю.* О распределении размеров частиц при дроблении // Информатика и ее применения, 2009. Т. 3. Вып. 3. С. 60–68.
- [44] *Королев В. Ю., Назаров А. Л.* Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // Автоматика и телемеханика, 2010. Вып. 3. С. 98–116.
- [45] *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. 2-е издание, переработанное и дополненное // М.: ФИЗМАТЛИТ, 2011. 620 С.
- [46] *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов // М.: изд-во Московского университета, 2011. 510 С.
- [47] *Королев В. Ю., Соколов И. А.* Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // Информатика и ее применения, 2012. Т. 6. Вып. 1. С. 2–10.

- [48] *Королев В. Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // Теория вероятностей и ее применения, 2013. Т. 58. Вып. 1. С. 117–132.
- [49] *Королев В. Ю., Черток А. В., Корчагин А. Ю., Горшенин А. К.* Вероятностно-статистическое моделирование информационных потоков в сложных финансовых системах на основе высокочастотных данных // Информатика и ее применения, 2013. Т. 7. Вып. 1. С. 12–21.
- [50] *Королев В. Ю., Корчагин А. Ю.* Модифицированный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов // Информатика и ее применения, 2014. Т. 8. Вып. 4. С. 11–19.
- [51] *Назаров А. Л.* Приближенные методы разделения смесей вероятностных распределений: Дисс. ... канд. физ.-мат. наук. – М.: Московский государственный университет им. М. В. Ломоносова, 2013.
- [52] *Ширяев А. Н.* Основы стохастической финансовой математики. Том 1. Факты. Модели. М.: «Фазис», 1998.
- [53] *Barndorff-Nielsen O. E.* Exponentially decreasing distributions for the logarithm of particle size // Proc. Roy. Soc. London, Ser. A, 1977. Vol. A(353). P. 401–419.
- [54] *Barndorff-Nielsen O. E.* Hyperbolic distributions and distributions of hyperbolae // Scand. J. Statist., 1978. Vol. 5. P. 151–157.
- [55] *Barndorff-Nielsen O. E.* Models for non-Gaussian variation, with applications to turbulence // Proc. Roy. Soc. London, Ser. A, 1979. Vol. A(368). P. 501–520.
- [56] *Barndorff-Nielsen O. E., Kent J., Sørensen M.* Normal variance-mean mixtures and z-distributions // International Statistical Review, 1977. Vol. 50. No. 2 (Aug., 1982), P. 145–159.

- [57] *Barndorff-Nielsen O. E.* Processes of normal inverse Gaussian type // *Finance and Stochastics*, 1998. Vol. 2. P. 41–18.
- [58] *Barndorff-Nielsen O. E., Blæsild P., Schmiegel J. A.* parsimonious and universal description of turbulent velocity increments // *European Physical Journal*, 2004. Vol. B 41. P. 345–363.
- [59] *Carr P. P., Madan D. B., Chang E. C.* The Variance Gamma process and option pricing // *European Finance Review*, 1998. Vol. 2. P. 79–105.
- [60] *Dennis J. E., Schnabel R. B.* Numerical Methods for Unconstrained Optimization and Nonlinear Equations. – Englewood Cliffs: Prentice-Hall, 1983.
- [61] *Eberlein E., Keller U.* Hyperbolic Distributions in Finance // *Bernoulli*, 1995. Vol. 1, No. 3. P. 281–299.
- [62] *Eberlein E., Keller U., Prause K.* New insights into smile, mispricing and value at risk: the hyperbolic model // *Journal of Business*, 1998. Vol. 71. P. 371–405.
- [63] *Eberlein E., Prause K.* The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures // Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, Preprint N 56, 1998.
- [64] *Eberlein E.* Application of Generalized Hyperbolic Lévy Motions to Finance // Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, Preprint No. 64, 1999.
- [65] *Gnedenko B. V., Korolev V. Yu.* Random Summation: Limit Theorems and Applications // Boca Raton: CRC Press, 1996.
- [66] *Kalashnikov V. V.* Geometric Sums: Bounds for Rare Events with Applications. – Dordrecht: Kluwer Academic Publishers, 1997.
- [67] *LeCam L.* Maximum likelihood; an introduction // *International Statistical Review*, 1990. Vol. 58. P. 153–171.

- [68] *Madan D. B., Seneta E.* The variance gamma (V.G.) model for share market return // Journal of Business, 1990. Vol. 63. P. 511–524.
- [69] *Prause K.* Modeling Financial Data Using Generalized Hyperbolic Distributions // Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, Preprint N 48, 1997.
- [70] *Protassov R. S.* EM-based maximum likelihood parameter estimation for a multivariate generalized hyperbolic distribution with fixed λ // Statistics, Computing, 2004. Vol. 14. P. 67–77.
- [71] *Rényi A.* On the central limit theorem for the sum of a random number of independent random variables // Acta Math. Acad. Sci. Hung., 1960. Vol. 11. P. 97–102.
- [72] *Seshadri V.* Halphen's laws // Kotz, S., Read, C. B., Banks, D. L. (Eds.). Encyclopedia of Statistical Sciences, Update Volume 1. New York: Wiley, 1997. P. 302–306.
- [73] *Stacy E. W.* A generalization of the gamma distribution // Annals of Mathematical Statistics, 1962. Vol. 33. P. 1187–1192.
- [74]
- [75]

Оглавление

Введение	3
1 Основные понятия прикладной теории вероятностей	7
1.1 Стохастические ситуации	7
1.2 Классические вероятностные модели	11
1.3 Дискретные вероятностные модели	13
1.4 Геометрические вероятностные модели	16
1.5 Структура вероятностных моделей	18
1.6 Независимость событий	25
1.7 Случайные величины и их распределения	29
1.8 Характеристики центра случайных величин	37
1.9 Характеристики разброса случайных величин	47
1.10 Неравенства для характеристик центра и разброса случайных величин	49
1.11 Характеристики формы распределения случайных величин	53
1.12 Ковариация и коэффициент корреляции	55
2 Испытания Бернулли	59
2.1 Биномиальное распределение	60
2.2 Распределение Пуассона	62
2.3 Теорема Муавра–Лапласа. Нормальное распределение	64
2.4 Геометрическое распределение	68
2.5 Показательное распределение	69

3	Основные предельные теоремы теории вероятностей	71
3.1	Законы больших чисел	72
3.2	Центральная предельная теорема	74
3.3	Суммы случайных индикаторов. Теорема Пуассона .	82
3.4	Устойчивые и безгранично делимые распределения .	86
4	Энтропийный подход к построению вероятностных моделей	90
4.1	Информация и энтропия	90
4.2	Случайные процессы	99
4.3	Пуассоновский процесс как модель дискретного хаоса	102
4.4	Принципы выбора моделей в рамках энтропийного подхода	110
5	Случайные суммы	112
5.1	Элементарные свойства случайных сумм	112
5.2	Пуассоновские и геометрические случайные суммы .	114
5.3	Теорема переноса	119
5.4	Случайные суммы случайных индикаторов. Аналог теоремы Пуассона	126
6	Вероятностные модели типа случайных блужданий	129
6.1	Случайные блуждания с дискретным временем . . .	129
6.2	Случайные блуждания с непрерывным временем . .	136
6.3	Обобщенные процессы Кокса	139
6.4	Центральная предельная теорема для обобщенных процессов Кокса	146
6.5	Закон больших чисел для обобщенных процессов Кокса	151
6.6	Моделирование распределений приращений финансовых индексов	152
7	Смешанные вероятностные модели	161
7.1	Смеси вероятностных распределений	161
7.2	Идентифицируемость смесей вероятностных распределений	166

7.3	Острове́ршинность масшта́бных смесей нормаль- ных законов	171
7.4	Устойчи́вость смесей нормальных законов относи- тельно смешивающего распределения	174
7.5	Дисперсионно-сдвиговые смеси нормальных законов	190
7.6	Обобщенные гиперболические распределения	192
7.7	Обобщенные дисперсионные гамма-распределения .	201
7.8	Двухэтапный сеточный метод разделения диспер- сионно-сдвиговых смесей нормальных законов . . .	207
7.9	Устойчи́вость представления вероятностных рас- пределений в виде дисперсионно-сдвиговых смесей нормальных законов	213
8	Вероятностные модели типа распределений времени жизни	217
8.1	Определение и примеры макс-обобщенных дважды стохастических пуассоновских процессов	218
8.2	Предельные распределения для макс-обобщенных дважды стохастических пуассоновских процессов . .	228
8.3	Предельные теоремы для мин-обобщенных дважды стохастических пуассоновских процессов	241
	Литература	247