

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Государственное образовательное учреждение
высшего профессионального образования

Учебное пособие

«Теория вероятностей»

А.А. Натан, О.Г. Горбачев, С.А. Гуз

Москва - 2008

УДК 519.8
ББК 22.18
М 51

Серия «Естественные науки. Математика. Информатика»

Редакционный совет серии:

Е. П. Велихов,
В. П. Иванников,
А. С. Кингсеп,
Е. И. Леванов,
А. И. Лобанов (ответственный секретарь серии),
Г. Ю. Ризниченко,
А. С. Холодов,
А. А. Шананин.

Рецензенты: д.ф.-м.н., профессор *Г.А. Агасандян*

Кафедра «Анализ данных и искусственный интеллект» отделения прикладной математики Государственного университета – Высшая школа экономики.

Зав. кафедрой д.ф.-м.н. *С.О. Кузнецов*.

А.А.Натан, О.Г.Горбачев, С.А.Гуз. Теория вероятностей — М., МЗ Пресс, 2007, 253 с.

Сжато излагаются лекции по семестровому курсу «Теория вероятностей», читаемого авторами студентам факультета управления и прикладной математики Московского физико-технического института перед курсами «Основы теории случайных процессов» и «Математическая статистика» (читаемых теми же авторами).

Пособие рассчитано на подготовку математиков – прикладников, но может служить основой и для более продвинутого изучения математических основ теории вероятностей.

Для студентов старших курсов и аспирантов.

ББК 22.18

ISBN 5-94073-099-X

© А.А.Натан, О.Г.Горбачев, С.А.Гуз, 2007 г.

Издание осуществлено при содействии НПО Наука

Оглавление

Введение.....	5
Исходные понятия и интуитивные предпосылки теории вероятностей.....	7
1.1. Классическое определение понятия «вероятность».....	10
1.2. «Геометрическая» вероятность.....	13
1.3. Статистическое определение понятия «вероятность»	15
1. Аксиоматическое определение вероятности.....	23
2. Условная вероятность. Стохастическая зависимость событий	34
1.4. Условная вероятность	35
1.5. Стохастическая (вероятностная) зависимость событий.....	36
1.6. О соотношении между вероятностными и причинно-следственными связями между событиями.	44
1.7. Формулы полной вероятности и Байеса	47
3. Случайные величины	50
1.8. Общая теория.....	50
1.9. Дискретные распределения.....	58
1.10. Непрерывные распределения	64
1.11. Функции случайных величин.....	71
1.12. Условное распределение случайной величины.....	72
4. Числовые характеристики случайных величин	78
1.13. Математическое ожидание случайной величины	78
1.14. Математическое ожидание функции случайной величины. Моменты	84
1.15. Дисперсия случайной величины	86
1.16. Условное математическое ожидание.....	89
1.17. Математическое ожидание в форме интегралов Лебега и Стильеса	91
1.18. Математическое ожидание и дисперсия для некоторых конкретных типов распределений	95
1.19. Параметры распределения, числовые характеристики, критерии.....	97
5. Случайный вектор.....	99
1.20. Общее определение и свойства случайного вектора.....	100
1.21. Числовые характеристики случайного вектора.....	109
1.22. Преобразование случайного вектора	117
1.23. Конкретные распределения случайного вектора.....	118
1.24. Свертка распределений. Устойчивость распределений относительно суммирования.....	126
6. Характеристические и производящие функции.....	129
1.25. Характеристические функции	129
1.26. Производящие функции	137
7. Предельные теоремы.....	140
1.27. Сходимость последовательностей случайных величин	140
1.28. Законы больших чисел	150
1.29. Предельные теоремы для распределений сумм случайных величин.	161
Список литературы	173
ТЕОРИЯ ВЕРОЯТНОСТЕЙ	174

Предисловие

Предлагаемое учебное пособие содержит курс лекций по одноименной дисциплине, читаемой авторами на протяжении многих лет на факультете управления и прикладной математики Московского физико-технического института, и представляющей собой первый курс из трех дисциплин, образующих вероятностно-статистический (стохастический) цикл учебного плана факультета.

Учебные пособия по двум последующим дисциплинам («Основы теории случайных процессов» и «Математическая статистика») изданы тем же авторским коллективом в 2003 году [5] и в 2004 году [6] соответственно.

Авторы выражают благодарность своим коллегам А.В.Гасникову и Е.В.Бурнаеву за сделанные ими по рукописи пособия замечания и пожелания.

Введение

Теория вероятностей представляет собой математическую дисциплину, формализующую случайные явления, т.е. призванную описывать строгими математическими соотношениями события, величины и процессы, имеющие недетерминированный (случайный) характер. Необходимость в таких операциях возникает во многих (практически – во всех) прикладных областях, где возникают задачи выводов и принятия решений по недетерминированным исходным данным.

Учебное пособие «Теория вероятностей» основано на лекциях, читаемых авторами по одноименному семестровому курсу на факультете управления и прикладной математики Московского физико-технического института. Этот курс является первым в цикле стохастических (вероятностно-статистических) курсов, предшествуя, согласно учебному плану, курсам по случайным процессам и по математической статистике.

Будучи достаточно традиционным по тематике и структуре, предлагаемое пособие обладает определенной спецификой по подбору материала и по его акцентированию. Основная преследуемая при этом цель состоит в выделении и обсуждении тех разделов теории, которые имеют важное практическое значение и определяют корректность стохастических моделей в приложениях, входящих в орбиту профессиональной подготовки выпускников факультета (моделирование технологических, физических и экономических реалий).

В первой главе пособия приведены исходные понятия и интуитивные предпосылки теории вероятностей. Здесь принимается в расчет, что интуитивные понятия вероятности, обладая содержательным смыслом, обычно лежат в основе описания исходных данных в прикладных стохастических моделях. Поэтому, несмотря на становление и развитие формальной теории вероятностей, корректное обращение с интуитивными определениями вероятности продолжает иметь крайне важное значение.

Во второй и третьей главах излагается аксиоматическое построение теории вероятностей и непосредственно следующие из нее выводы. Особое внимание здесь уделяется необходимости корректного обращения с понятиями вероятностной зависимости между событиями и ее соотношении с причинно-следственными связями между ними.

В четвертой и пятой главах приведены определения и свойства скалярных случайных величины, формы описания их распределений, числовые характеристики случайных величин.

В шестой главе пособия вводится понятие случайного вектора – многомерной случайной величины, описаны формы и свойства распределения и числовых характеристик случайного вектора. Приводится полное описание свойств нормального случайного вектора, играющего главную роль в многомерном статистическом анализе.

В седьмой главе сжато изложены основные факты из теории характеристических и производящих функций случайных величин, находящие широкое применение в стохастических моделях, содержащих предельные соотношения.

В восьмой главе излагаются и приведены доказательства основных предельных теорем теории вероятностей: законы больших чисел и предельные теоремы для распределения сумм случайных величин.

Принятые в пособии терминология и обозначения совпадают с используемыми в пособиях «Основы теории случайных процессов» [6] и «Математическая статистика» [5]

Исходные понятия и интуитивные предпосылки теории вероятностей

Будем понимать под *опытом* всякое действие, состоящее в создании или фиксации некоторого комплекса условий с целью получения (наблюдения) некоторого предполагаемого события или проверки какого-либо утверждения. Результат опыта назовем *событием*, полагая, что событие произошло, если результат опыта соответствует нашему ожиданию или сформулированное утверждение оказалось истинным.

Комплекс условий, при которых производится опыт, может контролироваться наблюдателем с той или иной полнотой, но, в силу неисчерпаемости природы, всегда не абсолютно полно. Иногда, однако, уже неполный контроль условий опыта достаточен, чтобы вполне определенно предсказать, что некоторое интересующее нас событие в каждом опыте непременно произойдет или, напротив, определенно не произойдет. Тогда событие будем называть *детерминированным*.

Чаще, однако, отсутствие полного контроля за условиями опыта не позволяет гарантировать появление нашего события: в результате каждого опыта оно может появиться или не появиться, т.е. является *недетерминированным*.

Приведем пример.

Пример 1.1. Пусть в урне находятся m белых и n черных шаров, и опыт состоит в последовательном поштучном их удалении из нее наугад (без возвращения) до момента извлечения первого черного шара. Обозначим через A событие, состоящее в том, что этот процесс завершится за конечное число шагов $k \leq m + n$. Ясно, что $k \leq m + 1$ и при любых конечных m и n это событие осуществляется непременно, т.е. является *детерминированным*. Пусть теперь событие B означает появление черного шара при первом извлечении. Это событие является *недетерминированным*, поскольку осуществляется необязательно. Следует при этом предполагать, что осуществимость этого события зависит от значений m и n (что не имеет места для события A).

Итак, в зависимости от условий опыта всякое интересующее нас событие может принадлежать к классу *детерминированных* или к классу *недетерминированных* событий, причем второй класс, видимо, более широк, поскольку *недетерминированные* события различаются по степени их осуществимости. С этой точки зрения, если ввести меру осуществимости события, то *детерминированным* событиям можно приписать максимальное значение осуществимости (равное, например, единице). При таком подходе все события можно включить в класс *недетерминированных*, считая *детерминированные* события их частным случаем.

Далее мы будем заниматься недетерминированными событиями, точнее – их подклассом, состоящим из случайных событий (это понятие будет вскоре определено). Абстрактное описание модели возникновения недетерминированного события удобно начать с представления всевозможных взаимоисключающих результатов опыта множеством $\Omega = \{\omega\}$. Каждый элемент ω этого множества, носит название элементарного исхода опыта, а множество Ω – множества исходов. Подчеркнем, что результатом каждого опыта является один и только один какой либо исход из Ω .

Важно также заметить, что наполнение множества Ω зависит от содержания описываемого (моделируемого) реального явления. Так, при моделировании карточной игры (при 52 карт в колоде) множество Ω может содержать 52 элемента (если результат игры зависит от извлечения конкретной карты), 13 элементов (если результат не зависит от масти карты) или 4 элемента (если важна только масть извлеченной карты).

В зависимости от решаемой задачи, множество Ω может быть конечным ($\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$), счетным ($\Omega = \{\omega_1, \omega_2, \dots\}$) или несчетным ($\Omega = \{\omega\}$).

Любое событие A , которое может быть наблюдаемо в результате проведения опыта, должно быть выражено через элементарные исходы. Другими словами, для анализа появления события A множество Ω должно быть сформировано таким образом, чтобы A выражалось через $\omega \in \Omega$. Это означает, что для каждого события A , наблюдаемого в данном опыте, множество исходов Ω оказывается разбитым на два подмножества A^* и $\Omega \setminus A^*$, из которых первое содержит все исходы, при которых осуществляется событие A (т.е. которые, как принято говорить, «благоприятствуют» событию A). Реализация события A идентична, следовательно, попаданию элементарного исхода ω в множество A^* , что можно выразить равенством

$$A = \{\omega \mid \omega \in A^*\} \quad (1.1)$$

Отсюда следует, что между событиями A и соответствующими им (согласно (1.1)) множествами A^* ($A^* \subseteq \Omega$) существует взаимно однозначное соответствие, причем теоретико-множественные операции над множествами соответствуют логическим (булевым) операциям над событиями: дополнению множества A^* соответствует отрицание события A

$$\Omega \setminus A^* \leftrightarrow \bar{A},$$

объединению множеств A_1^* и A_2^* соответствует дизъюнкция (логическая сумма) событий A_1 и A_2

$$A_1^* \cup A_2^* \leftrightarrow A_1 \vee A_2,$$

а пересечению множеств A_1^* и A_2^* – конъюнкция (логическое произведение) событий A_1 и A_2 а пересечению множеств A_1^* и A_2^* – конъюнкция (логическое произведение) событий A_1 и A_2

$$A_1^* \cap A_2^* \equiv A_1^* \cdot A_2^* \leftrightarrow A_1 \wedge A_2.$$

Это позволяет обычно одинаково обозначать множества и события, используя знак \cup – для операции дизъюнкции, знак \cap или \cdot – для операции конъюнкции, знак $\bar{}$ – для операции отрицания.

Детерминированное событие U , которое всегда происходит, соответствует полному множеству исходов Ω : $U = \{\omega \in \Omega\}$ и далее будет именоваться достоверным с обозначением $U = \Omega$. Для обратного детерминированного события \bar{U} , которое никогда не происходит, имеет место равенство $\bar{U} = \{\omega \in \emptyset\}$ и используется обозначение $\bar{U} = \emptyset$. Далее это событие именуется невозможным.

Полагая, что читатель знаком с основами булевой алгебры (алгебры логики) (см., например,[4]) , предлагаем ему вспомнить, как выводятся (с помощью таблиц истинности или диаграмм Венна) следующие часто используемые нами соотношения между событиями: пусть A и B – события, определенные на общем множестве исходов Ω , тогда справедливы соотношения:

$$A \cup B = A \cup \bar{A} \cdot B; \quad (1.2)$$

$$B = \bar{A} \cdot B \cup A \cdot B; \quad (1.3)$$

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B} \text{ – формулы Де Моргана.} \quad (1.4)$$

Известно, что на множестве всех булевых операций дизъюнкция и отрицание образуют полную систему операций: все остальные булевые операции могут быть выражены с помощью лишь этой пары операций. Заметим, что таким же свойством полноты обладает и пара «конъюнкция, отрицание». К этим и другим фактам из булевой алгебры мы будем обращаться, по мере необходимости, в следующих главах. Сейчас же перейдем к интуитивным определениям понятия вероятности события, как количественной характеристики его осуществимости.

Наблюдая появление различных событий при многократно повторяемых однородных опытах, мы склонны сравнивать степень их осуществимости по частоте их повторения. Наша задача состоит в построении математических моделей реальных явлений, в которых степень осуществимости события определялось бы числом, называемым его вероятностью. Такие модели мы будем в дальнейшем называть вероятностными или стохастическими (используя эти прилагательные в совпадающем смысле).

Ясно, что вводимые определения вероятностей событий должно отвечать постулатам, вытекающим из наших интуитивных представлений о вероятности.

Будем далее называть событие случайным, если его реальная или виртуальная осуществимость может быть количественно выражена его вероятностью (ниже мы увидим, что не всякое недетерминированное событие является случайным).

Применение в математической дисциплине термина «вероятность» требует, однако, пояснения.

В бытовом обиходе термин «вероятность» применяется, когда существует неопределенность в осуществлении того или иного события, и носит, обычно, качественный смысл, отражая его относительную осуществимость (как это имеет место, например, в фразах типа «высока вероятность, что через час пойдет дождь», «в предстоящей игре вероятность выигрыша у команды ЦСКА выше, чем у команды «Спартак» и т.п. Таким образом, вероятность нередко приписывается событию, которое принципиально не может быть результатом опыта, отвечающего требованиям возможности его многократного проведения в сходных условиях. Примером такого события является, например, извержение Ключевской сопки на Камчатке в течение следующего года.

Все это заставляет искать такое определение понятия «вероятность», которое приводило бы к ее корректному количественному выражению.

Решение этой проблемы «математизации» бытового термина «вероятность» потребовало более двух столетий усилий крупнейших ученых и привело к формированию теории вероятностей, как строгой математической дисциплины.

Вместе с тем, истоками и фундаментом современной теории вероятностей, остаются интуитивные понятия вероятности, не утратившие своего практического значения. С них мы и начнем.

Классическое определение понятия «вероятность»

На первых шагах становления теории вероятностей одним из стимулов ее развития служили задачи, возникающие в азартных играх и состоящие, обычно, в попытках вычисления вероятностей событий (выигрыш, проигрыш, ничья), реализующихся в результате случайного извлечения игровых карт или выпадения того или иного числа очков при бросании игральной кости. Такие задачи способствовали так называемому классическому определению вероятности, основанному на следующей схеме рассуждений.

Во-первых, отправляясь от игровых задач, множество исходов Ω принимается конечным (конечное число различных вариантов извлечения карт из колоды или числа

выпавших очков при бросании игральной кости). Во-вторых, множество Ω формируется из таких элементарных исходов, которые из интуитивных соображений считаются равновероятными, т.е. постулируется тот факт, что в Ω ни один из исходов не имеет приоритета в своей реализации по сравнению с каким либо другим исходом. В игровой задаче это означает, например, что все карты колоды имеют равную возможность быть извлеченными, т.е. игра ведется честно.

При всем этом заметим, что здесь, как обычно при построении стохастических моделей, предполагается принципиальная возможность многократного осуществления опыта в одних и тех же условиях.

При такой постановке предлагается следующее «классическое» определение вероятности.

Пусть $\Omega = (\omega_1, \omega_2, \dots, \omega_n)$ конечное множество n всех элементарных исходов, которые принимаются равновероятными, A и A^* – событие и множество ($A^* \subseteq \Omega$), связанные (1.1), n_{A^*} – число элементов множества A^* (число исходов, благоприятствующих событию A).

Тогда вероятность события A принимается равной

$$P(A) = \frac{n_{A^*}}{n}. \quad (1.5)$$

Это определение, естественно вытекающее из интуитивных представлений о вероятности в схеме равновероятных исходов опыта, позволяет установить ряд ее следующих свойств.

Прежде всего, отметим, что каждому событию, определенному (1.1), соответствует вероятность, вычисляемая согласно (1.5), т.е. все такие события являются случайными.

Из очевидного неравенства $n_{A^*} \leq n$ следует, что вероятность каждого случайного события отвечает условию

$$0 \leq P(A) \leq 1$$

События $A_1 = \{\omega \in A_1^*\}$ и $A_2 = \{\omega \in A_2^*\}$ называются несовместимыми, если не могут одновременно произойти в результате опыта. Ясно, что это имеет место, если A_1^* и A_2^* – непересекающиеся множества: $A_1^* \cap A_2^* = \emptyset$ (это обозначение сохраним и для событий: $A_1 \cap A_2 = \emptyset$).

Для дизъюнкции $A = A_1 \cup A_2$ двух несовместимых событий A_1 и A_2 из (1.5) следует

$$P(A) = P(A_1 \cup A_2) = \frac{n_{A^*}}{n} = \frac{n_{A_1^*} + n_{A_2^*}}{n} = P(A_1) + P(A_2), \quad (1.6)$$

т.е. вероятность дизъюнкции двух несовместимых событий равна сумме их вероятностей.

Этот результат читатель легко распространит на конечную группу k попарно несовместимых событий $\{A_j\}_{j=1}^k$ ($k \geq 2$), получив для вероятности их дизъюнкции A равенство

$$P(A) = P\left(\bigcup_{j=1}^k A_j\right) = \sum_{j=1}^k P(A_j)$$

Это свойство вероятности позволяет рассматривать ее как конечно-аддитивную меру множеств.

Пусть теперь события A_1 и A_2 совместимы, т.е. $A_1 \cap A_2 \neq \emptyset$. Тогда из (1.2) и (1.3), положив $A = A_1$ и $B = A_2$ легко получим

$$P(A) = P(A_1) + P(A_2) - P(A_1 \cdot A_2). \quad (1.7)$$

Пусть U – достоверное событие, которое, как было определено, осуществляется при любом исходе $\omega \in \Omega$. Из равенств $\Omega = \bigcup_{i=1}^n \omega_i$, $P(\omega_i) = \frac{1}{n}$ и несовместимости исходов опыта:

$\forall i, j: \omega_i \cap \omega_j = \emptyset$, следует $n_U = n$ и

$$P(U) = P\{\omega \in \Omega\} = \frac{n_U}{n} = \sum_{i=1}^n P(\omega_i) = 1 \quad (1.8)$$

т.е. достоверное событие принадлежит к числу событий с единичной вероятностью (забегая вперед заметим, что событие с единичной вероятностью не всегда является достоверным).

Пусть, далее, A и \bar{A} некоторое событие и его отрицание, т.е. несовместимые события, для которых $A \cup \bar{A} = \Omega$. Тогда из (1.6) и (1.8) следует

$$P(\bar{A}) = 1 - P(A) \quad (1.9)$$

Если положить в (1.9) $A = U$, то событие $W = \bar{U}$ представит собой невозможное событие, для которого множество благоприятствующих исходов пусто ($n_W = 0$) и $P(W) = \frac{n_W}{n} = 0$. Следовательно невозможное событие принадлежит к числу событий с нулевой вероятностью, которые мы будем называть нулевыми. Заметим, что и здесь, как выяснится ниже, нулевое событие не обязательно является невозможным.

До сих пор мы использовали различные обозначения (например – A и A^*) для событий и соответствующих им подмножеств множества исходов Ω . Поскольку эти понятия в нашем употреблении являются, по-существу, синонимами и определяются контекстом, в дальнейшем такое различие в обозначениях обычно отсутствует. В частности, как уже было принято, достоверное (U) и невозможное (W) события будут иметь обозначения, соответственно, Ω и \emptyset .

Описанный классический подход к определению и вычислению вероятностей в схеме равновозможных исходов опыта находит успешное применение во многих приложениях

(молекулярная и квантовая физика, теория надежности сложных систем, генетика, математическая микроэкономика и др.). Вместе с тем, как общая основа для теории вероятностей классическое определение вероятности имеет следующие ограничения.

Прежде всего, далеко не всегда реальные наблюдения, содержащие элемент случайности, могут быть сведены к опыту с конечным множеством равновозможных исходов. При этом понятие равновозможности элементарных исходов опыта совпадает, по существу, с понятием их равновероятности. Такая «зацикленность» определения снижает его математическую строгость, хотя и оправдана попыткой определения сложной сущности через более простую.

Следует также заметить, что предположение о равновозможности элементарных исходов в данном опыте целиком основывается на интуиции исследователя, которая иногда может не соответствовать реальности.

Поясним сказанное простым примером.

Пример 1.1. Пусть две частицы одна за другой случайным образом размещаются в двух ячейках. Требуется найти вероятность того, что в ячейках окажется по одной частице.

На первый взгляд, ответ очевиден: эта вероятность равна $1/4$. Однако он получен в предположении, что все четыре варианта размещения частиц по ячейкам равновозможны, что имеет место, если попадания каждой из них в первую или вторую ячейку равновозможны независимо от того, куда попала другая частица. Это условие в задаче не оговорено, что делает ответ в общем случае некорректным: нетрудно убедиться, что отказ от этого условия может приводить к другому решению. Так, например, если первая частица имеет равную возможность попасть в первую или во вторую ячейку, а вторая частица с двумя шансами против одного попадает в ту же ячейку, что и первая, то искомая вероятность оказывается равной $1/3$.

Заметим, кстати, что такой же ответ можно получить, полагая, что частицы неразличимы, т.е. неразличимы (точнее – тождественны) оба варианта расположения частиц по одной в каждой ячейке (напомним, что принцип неразличимости частиц является одним из постулатов квантовой статистики, отличающих ее от классической статистики Больцмана; к этому вопросу мы еще вернемся чуть ниже).

«Геометрическая» вероятность

В рассмотренном классическом определении вероятности предполагалось, что множество элементарных исходов Ω конечно.

Существуют, однако, задачи, в которых условие «равновозможности» исходов с некоторыми оговорками можно распространить и на случай, когда множество Ω несчетно.

Рассмотрим простой пример. Пусть стало известно, что газопровод длиной в L км, залегающий в однородной местности, имеет утечку в одной точке, расположение которой неизвестно. Требуется оценить вероятность p того, что утечка газа расположена на конкретном участке длиной ΔL км. Интуитивный ответ очевиден: $p = \frac{\Delta L}{L}$.

К такому ответу приводят два соображения: а) определение положения точки утечки газа можно уподобить выбору точки на прямой конечной длины L при предположении, что все ее положения на этой прямой равновозможны; б) вероятность расположения выбранной точки на заданном отрезке равна его относительной длине, т.е. его длине, нормированной единицей.

В приведенном простом примере задача имеет естественную геометрическую интерпретацию. Возможны, однако, случаи, когда геометрический подход к вычислению вероятностей оказывается успешным и при решении задач, не имеющих, на первый взгляд, геометрического содержания.

Покажем это на следующем примере.

Пример 1.2. Пусть каждое из двух наблюдаемых событий непременно осуществляется (независимо от другого) на отрезке времени $[0, \theta]$ в неопределенные равновозможные моменты времени. Требуется определить и вычислить вероятность того, что интервал между событиями будет не более τ .

Поясним, что под «равновозможностью» моментов появления события здесь понимается тот факт, что каждый из них выбирается «природой» наугад, вне зависимости от его положения на отрезке $[0, \theta]$ (в примере с газопроводом это предположение обосновано однородностью местности его залегания).

Итак, следует найти вероятность $P\{|t_1 - t_2| \leq \tau\}$ где t_1 и t_2 – моменты появления событий, $t_1 \in [0, \theta]$, $t_2 \in [0, \theta]$. Множество исходов Ω здесь представляет собой множество возможных значений вектора (t_1, t_2) $\Omega = [0, \theta] \times [0, \theta]$, изображенное квадратом на рис. 1.1. Исходы, благоприятствующие событию $A_\tau = \{|t_1 - t_2| \leq \tau\}$, образуют заштрихованную часть квадрата (G). Ввиду предположения о «равновозможности» исходов (в указанном смысле), возникает интуитивная предпосылка для выражения вероятности события A_τ отношением площади фигуры G к площади квадрата Ω :

$$P(A_\tau) = \frac{\text{mess} G}{\text{mess} \Omega} = \frac{2\tau\theta - \tau^2}{\theta^2}.$$

Приведенный пример можно распространить на большее число наблюдаемых событий, используя его многомерное геометрическое представление.

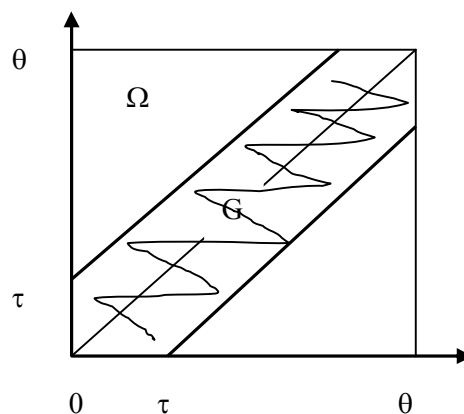


Рис. 0.1.

Хотя подобная геометрическая интерпретация вероятности в практических расчетах встречается сравнительно редко, в ней заложен глубокий смысл, состоящий в представлении вероятности как меры множеств (по аналогии с их геометрическими мерами – площадью и объемом). Заметим, что особенностью вероятности, как меры множества, является ее нормированность единицей. Ниже мы будем часто ссылаться на факты из теории меры, являющейся фундаментом современной теории вероятностей как строгой математической дисциплины.

Статистическое определение понятия «вероятность»

Иной («статистический») подход к определению понятия вероятности основан на том простом интуитивном соображении, что изучаемому недетерминированному событию следует приписать тем большую вероятность, чем чаще оно происходит при повторении опыта, т.е. чем выше частота его появления. Суть этого подхода состоит в следующем.

Пусть опыт, в котором интересующее нас событие A может осуществляться или не осуществляться, повторяется N раз (последовательность N опытов мы будем далее называть серией). Обозначим N_A число опытов из серии, в которых событие A произошло, и назовем отношение

$$W_N(A) = \frac{N_A}{N} \quad (1.10)$$

частотой события A . В общем случае появления события A , их число в серии N_A и, следовательно, частота $W_N(A)$ зависят от изменяющихся условий проведения опыта, некоторые из которых контролируются и управляются наблюдателем, а некоторые таковыми не являются, порождая (в силу своей изменчивости) неуправляемый разброс значений частоты $W_N(A)$ от серии к серии.

Так, к примеру, в ходе полигонных испытаний ракетной техники к контролируемым условиям при N -кратном обстреле цели могут относиться управляемые параметры ракетной установки и прицеливания (сохраняющиеся неизменными в сериях опытов), а к неконтролируемым – параметры атмосферы, которые могут флуктуировать, изменяясь от опыта к опыту (от выстрела к выстрелу).

Во многих приложениях, когда уровень неконтролируемых условий относительно невысок и они обладают некоторой стационарностью (не подвержены, к примеру, заметному тренду), частота $W_N(A)$ проявляет свойство *статистической устойчивости*: при больших значениях N ее значения мало отклоняются от некоторой константы.

Это свойство частоты служит основанием для широкого использования ее в качестве меры осуществимости наблюдаемого события, т.е. в роли оценки его вероятности. На этом пути возникает, однако, ряд вопросов методического (а, подчас, и концептуального) характера, из которых выделим следующие.

Прежде всего, заметим, что величина $W_N(A)$ зависит от общего числа проводимых опытов N , контролируемого исследователем, и от независимого от него числа N_A «благоприятных» из них (в которых произошло событие A). Последнее обстоятельство делают величину $W_N(A)$ зависящей от случая, что не позволяет непосредственно применить ее в качестве меры осуществимости события A (как это имеет место при классическом и геометрическом определениях вероятности).

Вместе с тем, практика показывает, что многие события обладают *асимптотической статистической устойчивостью*, состоящей в том, что с ростом N группировка значений частоты их появления около указанной константы делается все теснее. Обозначая эту константу p , представим этот факт равенством

$$\lim_{N \rightarrow \infty} W_N(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N} = p \quad (1.11)$$

Заимствованное из математического анализа обозначение предела имеет здесь, однако, не вполне ясный («стохастический») смысл, поскольку выражает сходимость недетерминированной (изменяющейся от опыта к опыту) величины $W_N(A)$ к константе p .

Характер этого типа сходимости, который в дальнейшем получит у нас строгое определение, иллюстрирует рис.1.2, на котором приведен результат статистического моделирования многократной реализации события A в сериях с различным числом опытов ($N = 2, 4, \dots, 200$) и вычисления по полученным данным зависимости от N значений частоты $W_N(A)$. Моделируемое событие A в этом опыте имело заданную классическую вероятность $P(A) = 0,5$ (что соответствует, например, выпадению герба при бросании «правильной»

монеты). Из полученного графика видно, что частота события A быстро сходится к константе p , которая в нашем случае оказывается равной классической вероятности этого события, т.е. $p = P(A) = 0,5$.

Сформулируем точнее свойство асимптотической статистической устойчивости частоты события.

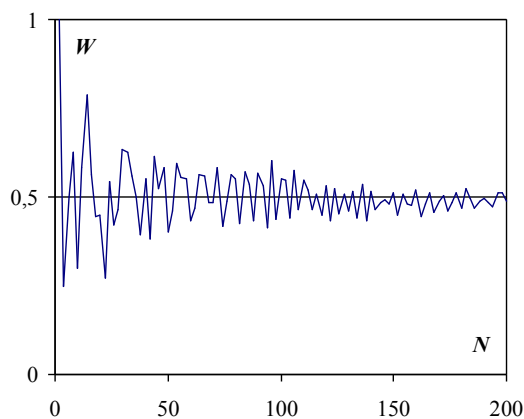


Рис. 1.2.

Пусть при неизменных контролируемых условиях для каждого N , принимающего значения $N_1 < N_2 < \dots$, проведено L серий по N опытов в каждой и $\{W_N^{(j)}(A)\}_{j=1}^L$ – полученные в этих сериях значения частоты события A (для различных серий, вообще говоря, разные ввиду флуктуации неконтролируемых факторов).

Обозначим $\Delta_\varepsilon = (p - \varepsilon, p + \varepsilon)$ ($0 < \varepsilon < p$), $L_\varepsilon(N)$ – число серий, в которых при данном N значение $\{W_N^{(j)}(A)\}_{j=1}^L$ оказалось лежащим в интервале Δ_ε . Тогда условие (1.11) можно записать в виде

$$\forall \varepsilon: V(N) \stackrel{\Delta}{=} \frac{L_\varepsilon(N)}{L} \xrightarrow{N \rightarrow \infty} 1. \quad (1.12)$$

На рис.1.3 приведен график, иллюстрирующий выполнение зависимости (1.12), полученный для тех же условий, что и график рисунка 1.2, при числе серий $L = 1000$ для каждого N и при $\varepsilon = 0,08$.

Приведенные факты, обладающие, как показывает практика, высокой общностью, приводят к выводу, что вероятностью события следует считать не частоту его появления, а ее предел в указанном выше смысле (полагая, что он существует, представляя собой детерминированную константу), т.е. дать *статистическое определение понятия вероятности* как предела (в указанном выше смысле)

$$P(A) = \lim_{N \rightarrow \infty} W_N(A) \quad (1.13)$$

(ниже определение (1.13) возникнет в модифицированном виде как теорема, в которой предел примет строгий формальный смысл).

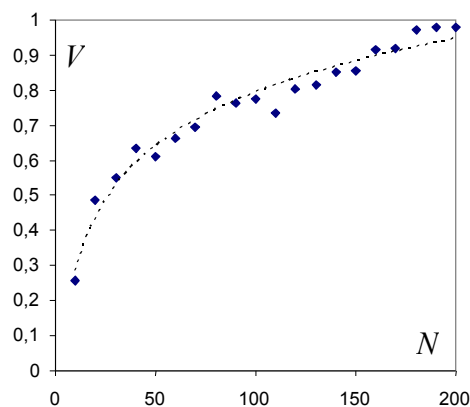


Рис. 0.2.

Читатель может самостоятельно убедиться, что свойства вероятности событий, вытекающие из ее классического определения, сохраняются и при ее статистическом определении. Это вытекает из того, что свойства «классической» вероятности событий совпадают с свойствами их частот, сохраняясь и при предельных переходах.

Остановимся кратко на концептуальных вопросах, связанных с интуитивным определением понятия вероятности. Прежде всего, подчеркнем, что в основе каждой стохастической модели лежат исходные события, вероятности которых задаются на основе интуитивных соображений: в схеме равновозможных исходов или путем их частотной интерпретации по результатам эксперимента. Мощный математический аппарат теории вероятности направлен на обработку этих исходных данных с целью корректного описания вероятностных свойств сложных явлений и связей. Отсюда можно сделать вывод о важной роли, которую играет при построении стохастической модели корректность интуитивных подходов к вероятностному описанию базовых данных.

В связи с существованием двух способов интуитивного определения вероятности – классического и статистического (геометрическое определение вероятности можно отнести к первому), возникает вопрос, какому из них следует отдать предпочтение, если оба реализуемы. Достаточно очевидные соображения приводят к выводу, что при построении схемы равновозможных исходов подразумевается, что частоты их появления в серии опытов при возрастании ее размера должны сходиться к одному и тому же числу, равному вероятности отдельного исхода. Отсюда можно ожидать, что в такой схеме значения оценок классической и статистической вероятности события должны совпадать, что и подтверждается практикой (в частности, это иллюстрируется описанным выше статистическим экспериментом).

Несовпадение этих значений вероятности исследуемого события может быть следствием как ошибочности построенной схемы равновозможных исходов при классическом определении вероятности, так и неповторяемости (тренда) или

недостаточности статистических данных при ее статистическом определении. Если последние условия соблюдены (что можно проверить, используя графики, аналогичные приведенным на рис.1.2 и 1.3), то «конфликт» в оценке вероятности следует решать в пользу статистической вероятности, внося коррективы в схему равновозможных исходов.

Приведем классический пример такой ситуации, именуемый парадоксом шевалье де Мере, и состоящий в сравнении вероятностей получения в сумме 11 и 12 очков при бросании трех игральных костей. Исходя из того, что обе суммы могут образоваться шестью способами каждая ($11 = 4 + 4 + 3 = 5 + 3 + 3 = 5 + 4 + 2 = 5 + 5 + 1 = 6 + 3 + 2 = 6 + 4 + 1$ и $12 = 4 + 4 + 4 = 5 + 4 + 3 = 5 + 5 + 2 = 6 + 3 + 3 = 6 + 4 + 2 = 6 + 5 + 1$) и полагая эти способы равновозможными, де Мере нашел, что обе суммы должны появляться с равными вероятностями. Этот вывод, однако, противоречил опытным данным, согласно которым двенадцать очков выпадает реже, чем одиннадцать. Ошибка де Мере (на которую ему указал знаменитый Паскаль) состояла в неправомерном представлении указанных исходов равновозможными. Читатель может самостоятельно найти, что число равновозможных исходов, приводящих к суммам очков 11 и 12 равно, соответственно, 27 и 25, а вероятности их появления, соответственно $27/216$ и $25/216$.

Подобные проблемы согласования вероятностных моделей с результатами эксперимента возникают и в гораздо более серьезных приложениях, к которым относится, например, построение вероятностных моделей современной физики.

Возникновение статистической физики сопровождалось широким использованием вероятностных моделей, описывающих поведение коллективов частиц – атомов и молекул, образующих ту или иную материальную субстанцию.

Исходными моделям при этом являются схемы распределения частиц по их состоянию, т.е. по ячейкам фазового пространства. Рассмотрим варианты расположения n частиц в k ячейках. Если все частицы различимы (т.е., к примеру, могут быть пронумерованы), то общее число различных вариантов их расположения равно k^n . Полагая, что попадание каждой частицы в ту или иную ячейку не зависит от расположения других частиц и что все варианты их размещения равновозможны (с вероятностью реализации, равной k^{-n}), мы приходим к статистике (модели) Максвелла – Больцмана, лежащей в основе классической статистической физики. Подчеркнем, что эта модель вполне соответствует нашим интуитивным представлениям о сути явления и классическому определению вероятности.

Развитие статистической физики показало, однако, что поведение далеко не всех микрочастиц можно описать, отправляясь от модели Максвелла – Больцмана. Это потребовало поиска иных моделей, позволяющих адекватно объяснить эмпирические

результаты квантовой физики. Первой из них является статистика Бозе – Эйнштейна, основанная на следующих соображениях. Будем полагать, что располагающиеся в k ячейках n частиц принципиально неразличимы (тождественны). Это означает, что каждый вариант размещения частиц по ячейкам (n_1, n_2, \dots, n_k) , $n_1 + n_2 + \dots + n_k = n$, считается неизменным при любой перестановке частиц с сохранением чисел (n_1, n_2, \dots, n_k) . Ясно, что при этом число различных вариантов, описывающих состояние системы n частиц существенно сокращается. Действительно, если при различимых частицах (в статистике Максвелла – Больцмана) для каждого набора (n_1, n_2, \dots, n_k) число различных вариантов размещения частиц равно

$$N(n_1, \dots, n_k) = C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot \dots \cdot 1 = \frac{n!}{n_1! \cdot \dots \cdot n_k!},$$

то в статистике Бозе – Эйнштейна это число полагается равным единице. Важно также заметить, что в этой модели, как и в модели Максвелла – Больцмана, все варианты размещения частиц принимаются равновероятными.

Для объяснения поведения некоторых микрочастиц статистику Бозе – Эйнштейна приходится дополнить ограничением («принципом Паули»), запрещающим размещение в каждой ячейке более одной частицы. Такая модель носит название статистики Ферми – Дирака.

Кратко описанные вероятностные модели, используемые в теоретической физике, иллюстрируют доминирующую роль опыта в формировании модельного описания исследуемого явления, включая случаи, когда это связано с необходимостью существенной коррекции интуитивных предпосылок при определении понятия вероятности.

В течение всего процесса становления теории вероятностей продолжались (и продолжают) дискуссии об интерпретации понятия вероятности, как философской категории.

Не останавливаясь подробно на этой обширной теме, укажем, что в центре этих обсуждений находится вопрос о мере объективности понятия вероятности. Наиболее распространенной позицией в этом вопросе является следующая.

Пусть появление или неappearance некоторого события A зависит от множества факторов, часть из которых контролируется пассивным наблюдателем (не оказывающим активного влияния на эти факторы). Ясно, что степень осуществимости (вероятность), которую наблюдатель приписывает событию A , в общем случае зависит от его информированности об условиях, в которых осуществляется это событие. Так человек, попавший в страну с неизвестным климатом и не имеющий информации о текущей погоде, вряд ли может прогнозировать дождливую погоду на следующий день иначе, как событие с вероятностью, равной 0,5. Напротив, наблюдатель, обладающий такой информацией и

соответствующей погодной статистикой, может предсказать завтрашнюю погоду более точно.

Аналогичная ситуация имеет место в следующем «урновом» примере. Пусть оператор извлекает шар из урны, содержащей m белых и n черных шаров, и, не глядя, кладет его в карман. Его коллега подсмотрел, что этот шар был белый. Легко убедиться, что вероятность события, состоящего в том, что второй извлеченный шар будет белым для оператора составляет $p_1 = \frac{m}{m+n}$, а для его (более «информированного») коллеги $p_2 = \frac{m-1}{m+n-1}$.

Подобные примеры могут привести к выводу, что вероятность события, будучи зависимой от информированности наблюдателя, имеет «интерсубъективный» характер, оставаясь неизменной для одинаково информированных наблюдателей. Такой вывод, однако, требует следующей коррективы.

Как мы видели (и еще более подробно убедимся ниже) базой всякой стохастической модели реального явления является тройка: множество исходов, система связанных между собой булевыми соотношениями событий и распределением вероятностей этих событий. Эту тройку мы назовем вероятностным пространством. Его структура зависит, естественно, от информированности исследователя, в то время как вероятности событий представляют собой константы, имеющие объективный смысл в рамках данного вероятностного пространства, т.е. бытующего представления наблюдателя об изучаемой реальности. Поэтому следует приписывать свойство интерсубъективности не вероятности события, а вероятностного пространства, в котором данное событие фигурирует. В таком ракурсе свойство интерсубъективности можно широко распространить на системы гипотез, используемые при моделировании реального мира.

Заметим, что в представлениях современной квантовой физики понятие вероятности приобрело более глубокий объективный смысл, выступая как внутреннее свойство элементарных частиц, не зависящее от внешних условий и информированности наблюдателя.

Примером может служить событие, состоящее в распаде свободного нейтрона. Известно, что нейтрон, находясь в свободном состоянии, в случайный момент времени, распадается (точнее – превращается в три различные частицы). Пусть τ – время, прошедшее с начала наблюдения, t – момент распада нейтрона, $A = \{t < \tau\}$ – событие, состоящее в распаде нейтрона в интервале времени $[0, \tau)$, $P(A)$ – его вероятность. Существующие опытные данные позволяют заключить, что вероятность $P(A)$ является объективной величиной, зависящей только от τ и остающейся неизменной при различных внешних

условий и от информирования наблюдателя о возрасте нейтрона к моменту начала наблюдения.

Подобное поведение самораспада большинства элементарных частиц, будучи основой явления радиоактивности, выдвигает вероятность на роль фундаментального свойства материи.

Введя понятия вероятностей случайных событий, естественно задать вопросы: 1) какой практический смысл имеет знание значения вероятности интересующего нас события и 2) какова цель предстоящего изучения нами математических основ теории вероятностей.

Отвечая на первый вопрос, следует указать, что вычисление вероятностей событий осуществляется, как правило, в процедурах принятия решений, имеющих альтернативный смысл. Это приводит к тому, что на практике оценка вероятностей событий преследует цели сравнения их осуществимости, т.е. носит сравнительный характер.

Такое положение имеет место, например, при медицинском диагнозе заболевания больного, когда вариантам болезни могут быть приписаны их вероятности (ясно, конечно, что выбор лечения должен определяться не только этими вероятностями, но и возможными последствиями от неправильного диагноза). Другой пример относится к выбору варианта ценовой политики торговой фирмы с учетом риска, выражаемого вероятностью ее разорения. В теории и практике исследования надежности сложных технических систем сравнительная оценка вероятности безотказной работы в течение срока службы ее элементов может использоваться при альтернативном выборе структуры системы.

Даже в тех случаях, когда значение вероятности имеет, казалось бы, самодовлеющий характер (например, декларируемая вероятность безотказной работы рекламируемого изделия в течение его срока службы), в контексте предполагается его сопоставление по этому параметру с конкурирующим изделием или рекламирование достигнутого прогресса в части его надежности.

При ответе на второй вопрос подчеркнем, что технология вероятностных расчетов состоит, обычно, в вычислении вероятностей сложных («выходных») событий по вероятностям простых («входных») событий, часто имеющих эмпирическое происхождение и получаемых в результате практического приложения описанных определений вероятности. Результатом таких вычислительных процедур нередко является уточнение и обнаружение взаимозависимости параметров изучаемого явления.

Математическая теория вероятностей, не претендуя на новое *содержательное* понятие вероятности, строит систему формальных правил, позволяющих корректно реализовать технологию вероятностных расчетов. При этом основные (приведенные выше) свойства интуитивной вероятности принимаются в математической теории вероятностей как аксиомы.

Сказанное свидетельствует о единстве практического применения в исчислении вероятностей интуитивного и аксиоматического подходов, к последнему из которых мы и переходим.

Аксиоматическое определение вероятности

Решающим этапом формирования теории вероятностей как математической дисциплины стало построение ее аксиоматики, завершенное к середине прошлого столетия А.Н.Колмогоровым. Чтобы сформулировать систему аксиом теории вероятностей, введем ряд новых понятий.

Пусть, как и прежде, Ω – множество всех возможных исходов опыта (наблюдения, испытания). Рассмотрим некоторую совокупность A' подмножеств множества Ω . Предположим, что эта совокупность подмножеств замкнута относительно операции дополнения и объединения любой пары подмножеств из A' , т.е. выполняются импликации

$$\begin{aligned} \forall A : A \in A' \Rightarrow \bar{A} \in A'; \\ \forall A_1, A_2 : A_1 \in A', A_2 \in A' \Rightarrow A_1 \cup A_2 \in A', \end{aligned} \quad (2.1)$$

означающие, что дополнение любого множества из A' принадлежит A' и объединение любой пары множеств из A' также принадлежит A' .

Из второй импликации по индукции получаем, что для любой конечной группы l подмножеств из A' ($\{A_i\}_{i=1}^l, l < \infty$) их объединение также принадлежит A' :

$$\bigcup_{i=1}^l A_i \in A'. \quad (2.2)$$

Совокупность подмножеств A' после включения в нее множества Ω образует систему множеств A , называемую алгеброй множеств.

Переходя к терминологии, применяемой для операций над событиями, получаем, что алгебра множеств A соответствует системе событий, содержащей достоверное событие Ω и замкнутой относительно операций отрицания и дизъюнкции или, ввиду полноты этой группы операций, относительно любых булевых операций над любым конечным набором событий из этой системы. Такую систему событий будем называть алгеброй событий, сохранив для нее обозначение A .

Итак, алгебра событий представляет собой систему событий, связанных между собой булевыми операциями, что, как обнаружится ниже, приводит к возможности исчисления их вероятностей. Пока, однако, это относилось лишь к тем случаям, когда исследуемая связь между событиями распространялась на конечное их число, что имеет место, например, когда множество Ω конечно. Нередко, однако, нас могут интересовать события, представляющие

собой булевы функции от счетного числа событий; это происходит, например, при рассмотрении предельных соотношений теории вероятностей.

Это обстоятельство заставляет в общем случае вместо алгебры событий рассматривать более широкую систему, для которой (наряду с (2.1)) всякое событие, представляющее собой булеву функцию не более, чем от счетного числа событий из этой системы, также является элементом этой системы. Назовем такую систему σ -алгеброй событий F , включив в нее достоверное событие Ω и события, отвечающие условиям

$$\begin{aligned} \forall A: A \in F \Rightarrow \bar{A} \in F; \\ \forall n \leq \infty, A_j \in F, j = \overline{1, n} \Rightarrow \bigcup_{j=1}^n A_j \in F \end{aligned} \quad (2.3)$$

(подчеркнем, что n может принимать здесь конечные и бесконечно большое значения).

Подведем итоги сказанного выше в виде двух определений.

Определение 2.1. Алгеброй событий на множестве исходов Ω называется система событий, содержащая достоверное событие Ω и замкнутая относительно булевых операций над не более, чем конечным числом любых событий из этой системы.

Определение 2.2. σ -алгеброй событий на множестве исходов Ω называется система событий, содержащая достоверное событие Ω и замкнутая относительно булевых операций над любой не более, чем счетной группой событий из этой системы.

Далее, описывая стохастические модели, мы будем говорить о существовании σ -алгебр событий, включая и те случаи, когда алгебры и σ -алгебры событий совпадают (что имеет место, например, при конечности множества исходов Ω).

Приведем поясняющие примеры.

Пример 2.1. Пусть результат некоторого наблюдения описывается конечным множеством исходов $\Omega = (\omega_0, \omega_1, \omega_2, \dots, \omega_n)$, где ω_i – элементарный исход, содержательный смысл которого состоит в появлении в заданном интервале времени ровно i интересующих нас событий из известного их числа n ($n < \infty$). На практике это может быть число проданных к определенному сроку изготовленных фирмой n изделий, число заболевших в течение контрольного срока из n наблюдаемых пациентов при медицинском мониторинге, число проданных билетов на футбольный матч (из поступивших в продажу n билетов) и т. д.

В силу конечности исходного множества Ω , каждая построенная здесь σ -алгебра событий представляет собой алгебру событий. При этом выбор подходящей алгебры неоднозначен и определяется содержанием и условиями решения практической задачи. Действительно, полнота построенной алгебры событий должна определяться требуемой и возможной детализацией получаемых выводов. Так, при наиболее полной алгебре событий,

содержащей в качестве ее элементов все исходы ω_i , в нее оказываются включенными все события, которые могут интересовать исследователя. Так, например, применительно к коммерческой операции, к ним могут относиться события B_k , означающие продажу к заданному сроку более k изделий ($k = \overline{1, n}$)

$$B_k = \bigcup_{i=k+1}^n \omega_i = \overline{\bigcup_{i=1}^k \omega_i}.$$

Полнота используемой в стохастической модели алгебры событий не всегда, однако, оправдана, поскольку приводит к ее усложнению и, главное, требует соответствующей полноты данных о значениях вероятности исходных событий (в нашем примере – вероятностей исходов ω_i), без чего расчет вероятностей более сложных событий невозможен и модель теряет практический смысл (оставаясь, быть может, полезной при теоретических исследованиях).

Так, если (по каким-то соображениям) фирму интересуют лишь такие события, как продажа хотя бы одного изделия (событие C_1) и продажа более 1 из них, $l = \text{const}$ (событие C_l), то достаточно, чтобы соответствующая задаче алгебра событий содержала события

$$A_1 = \omega_0 \text{ и } A_2 = \bigcup_{i=1}^l \omega_i;$$

тогда для интересующих нас событий получим

$$C_1 = \overline{A_1}, \quad C_l = \overline{A_1 \cup A_2}$$

и для вычисления вероятностей этих событий достаточно, как мы увидим, знать вероятности событий A_1 и A_2 (не исключается, конечно, и непосредственное получение вероятностей событий C_1 и C_l).

Пример 2.2. Изменим условия предыдущего примера, положив $n = \infty$, т.е. не ограничивая число интересующих нас событий, которые могут произойти к заданному сроку. Такое допущение является естественным, если, например, эти события представляют собой число заказов (спроса) на изделия фирмы, число отказов аппаратуры, число заболеваний в ходе эпидемии и т. д.

Теперь исходное множество счетно: $\Omega = (\omega_0, \omega_1, \omega_2, \dots)$, и существуют события, которые не принадлежат ни одной из алгебр событий, построенных на этом множестве, поскольку не могут быть выражены в виде булевых функций от конечного числа элементов этих алгебр. К таким событиям, в частности, относится событие

$$G = \bigcup_{i=1}^{\infty} \omega_k,$$

означающее, что число состоявшихся событий кратно заданному целому k ($k \geq 2$).

Это приводит к необходимости использования в этом примере σ -алгебры, наполнение которой, как и в случае алгебры событий, определяется условиями задачи.

В связи с разнообразием σ -алгебр, которые могут быть образованы на каждом исходном множестве Ω , рассмотрим обычный процесс их «порождения». Пусть при моделировании некоторого стохастического явления с исходным множеством Ω определена некоторая группа событий $\{A\}$, которые должны служить аргументами при булевом представлении других интересующих нас событий (обычно с целью вычисления вероятностей последних). Для этого все эти события должны принадлежать одной общей для них σ -алгебре. В общем случае выбор такой σ -алгебры неоднозначен и нередко возникает задача определения из всех подходящих из них ту, которая минимальна в смысле числа входящих в нее событий (точнее – мощности их множества). Минимальная σ -алгебра, содержащая группу событий $\{A\}$, называется σ -алгеброй, порожденной этими событиями.

Минимальная σ -алгебра для данной группы событий $\{A\}$ определяется следующим образом. Рассмотрим совокупность всех σ -алгебр, содержащих группу $\{A\}$; то, что эта совокупность не пуста, следует из того, что в нее непременно входит σ -алгебра, содержащая все события (подмножества) из Ω . Можно убедиться, что пересечение этих σ -алгебр снова представляет собой σ -алгебру, содержащую $\{A\}$ и являющуюся минимальной, поскольку всякое подмножество элементов этой σ -алгебры не является σ -алгеброй (к вопросу о существовании минимальной σ -алгебры с более общих позиций мы вернемся в четвертой главе).

Большую роль в дальнейшем изложении будут играть борелевские σ -алгебры, которые возникают в тех случаях, когда исходное множество совпадает с числовой осью ($\Omega = R^1$), с конечномерным евклидовым пространством ($\Omega = R^n$) или с евклидовым пространством счетной размерности ($\Omega = R^\infty$). Начнем со случая $\Omega = R^1$ (другие случаи будем рассматривать по мере необходимости).

Итак, пусть $\Omega = R^1 = (-\infty, \infty)$. Рассмотрим σ -алгебру B^1 , порожденную интервалами (a, b) , $-\infty \leq a < b \leq \infty$. Такая борелевская σ -алгебра весьма богата: помимо всевозможных комбинаций конечного и счетного набора интервалов, она содержит все отдельные точки $\{a\}$ ($a \in R^1$), замкнутые и полузамкнутые интервалы $[a, b]$, $[a, b)$, $(a, b]$ и их комбинации. Это следует из следующих очевидных соотношений:

$$\begin{aligned}\forall a \in \mathbb{R}^1 : \{a\} &= \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n} \right) \in \mathcal{B}; \\ [a, b] &= \{a\} \cup (a, b) \cup \{b\}; \\ [a, b) &= \{a\} \cup (a, b); \\ (a, b] &= (a, b) \cup \{b\}.\end{aligned}$$

Читатель легко проверит, что эта σ -алгебра содержит полубесконечные интервалы типа $(-\infty, a)$ и (b, ∞) , а также тот факт, что эта же σ -алгебра возникает и в случаях порождения ее промежутками другого типа (полуинтервалами типа $[a, b)$, $(a, b]$ или отрезками типа $[a, b]$).

Переходя к аксиоматике теории вероятностей, будем обозначать в общем случае построенную на исходном множестве Ω σ -алгебру буквой \mathcal{F} , а пару (Ω, \mathcal{F}) назовем измеримым пространством. Каждое множество A^* , являющееся элементом σ -алгебры \mathcal{F} , назовем \mathcal{F} – измеримым множеством.

Итак, пусть задано некоторое измеримое пространство (Ω, \mathcal{F}) , определенное конкретной задачей моделирования реального опыта (наблюдения, принятия решения, прогнозирования и т. д.). Исходя из описанного в главе 1 интуитивного представления о вероятности события, как о мере его неопределенности, припишем каждому событию A , $A \in \mathcal{F}$, вероятность его осуществления. Суть математической аксиоматики теории вероятности состоит в постулировании минимального набора свойств вероятности события, который позволял бы осуществлять всевозможные операции исчисления вероятностей. Адекватность предлагаемой аксиоматики проверяется, как обычно, путем сопоставления теории с интуитивными предпосылками и с практикой.

Рассматривая вероятность $P(A)$ события A , $A \in \mathcal{F}$, как числовую функцию множеств, определенную на элементах \mathcal{F} , припишем ей следующие свойства.

$$\textbf{Аксиома 1:} \text{ Для } \forall A \in \mathcal{F} \ P(A) \geq 0; \quad (2.4)'$$

$$\textbf{Аксиома 2:} \ P(\Omega) = 1; \quad (2.4)''$$

Аксиома 3: Пусть $\{A_i\}_{i=1}^n$ – конечная или счетная совокупность несовместимых событий из \mathcal{F} (т.е. $n \leq \infty$ и для $\forall i, j, i \neq j \ A_i \cap A_j = \emptyset$). Тогда

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (2.4)'''$$

т.е. вероятность дизъюнкции конечного или счетного числа несовместимых событий равна сумме вероятностей этих событий.

Вербальный смысл аксиом теории вероятностей заключается в определении вероятности события A как нормированной единицей счетно-аддитивной меры множества A^* (см.(1.1)). Каждое конкретное соответствие между событиями из \mathcal{F} и их вероятностями

мы будем называть распределением вероятностей (вероятностным распределением) на измеримом пространстве (Ω, F) и обозначать P , а тройку $H = \langle \Omega, F, P \rangle$ назовем вероятностным пространством.

Вид вероятностного пространства определяется конкретным содержанием решаемой задачи, состоящей, как правило, в моделировании какого-либо реального явления, обладающего стохастичностью. Преследуемая при этом цель состоит в описании связей между входными данными, представленными набором случайных событий с известными распределениями вероятностей и выходными данными, содержащими распределение вероятностей интересующих исследователя случайных событий, связанных с входными данными булевыми соотношениями. Входные данные содержат входные события, порождающие σ - алгебру F в данном вероятностном пространстве; выходные события являются элементами этой σ - алгебры и обычно существенно сложнее входных.

Так, типичными задачами стохастического моделирования являются оценка надежности сложной технической системы (вероятность ее безотказной работы в течение заданного срока службы) по надежности ее элементов, диагноз болезни пациента по многим факторам, оценка коммерческого риска инвестиции по параметрам проекта и т. д.

Распределения вероятностей для входных событий принимается заданными интуитивными или опытными (статистическими) оценками; для определения распределений вероятностей выходных событий служит теория вероятностей – ее аксиоматика и вытекающие из нее следствия, образующие формальный инструмент исчисления вероятностей случайных событий, подвергающихся булевым операциям.

Все это делает полезным начинать построение стохастических моделей с формирования соответствующих им вероятностных пространств, дабы избегать ошибок, иллюстрируемых следующим простым примером.

Пример 2.3. Студент ожидает вызова для сдачи зачета к одному из двух преподавателей А или В. Статистика показывает, что преподаватель А ставит зачет с вероятностью 0.7, а преподаватель В – с вероятностью 0.8. События C_1 и C_2 – успешная сдача зачета преподавателю А и В (соответственно) – несовместимы, и, используя аксиому 3, студент находит, что вероятность успешной сдачи зачета (событие C) составляет

$$P(C) = P(C_1 \cup C_2) = P(C_1) + P(C_2) = 1,5!$$

Эта ошибка вызвана тем, что события C_1 и C_2 принадлежат разным вероятностным пространствам: H_1 с алгеброй A_1 , порожденной событием C_1 (вероятностное пространство преподавателя А) и H_2 с алгеброй A_2 , порожденной событием C_2 (вероятностное пространство преподавателя В). Правильное решение будет нами ниже найдено в результате объединения пространств H_1 и H_2 (см. §3.1).

Введение вероятностного пространства позволяет уточнить термин «случайное событие», который теперь присваивается всякому событию A , являющемуся элементом F ; $A \in F$, и потому, согласно аксиоме 1, обладающему вероятностью $P(A)$. Иными словами, случайному событию A соответствует F – измеримое множество A^* (см. (1.1)).

Очевидно, что не всякое событие, определенное (1.1), является случайным, поскольку не всякое множество $A^* \in F$ – измеримо. Более того, одно и то же событие может быть случайным или не быть таковым в зависимости от конкретного вида σ -алгебры F .

Приведем примеры.

Пример 2.4. Пусть $\Omega = [0,1)$ и рассмотрим три построенные на этом исходном множестве вероятностные пространства:

а) $H_1 = \langle \Omega, F_1, P_1 \rangle$, где F_1 – σ -алгебра, порожденная всеми событиями вида $A = \{\omega \in [a, b)\}$, $0 \leq a < b < 1$, $P(A) = \text{mess}(b - a)$;

б) $H_2 = \langle \Omega, F_2, P_2 \rangle$, где F_2 – σ -алгебра, порожденная всеми событиями вида

$$A_k = \{\omega \in [a_k, b_k)\}, \quad a_k = \frac{2^{k-1} - 1}{2^{k-1}}, \quad b_k = \frac{2^k - 1}{2^k}, \quad k = 1, 2, \dots, \quad P(A_k) = \text{mess}(b_k - a_k);$$

в) $H_3 = \langle \Omega, F_3, P_3 \rangle$, где F_3 – алгебра, порожденная событиями

$$A_1 = [0, \frac{1}{4}), \quad A_2 = [\frac{1}{4}, \frac{1}{2}), \quad A_3 = [\frac{1}{2}, \frac{3}{4}), \quad A_4 = [\frac{3}{4}, 1), \\ P(A_1) = P(A_2) = P(A_3) = P(A_4) = \frac{1}{4}.$$

Читатель легко убедится, что эти вероятностные пространства отвечают аксиомам (2.4) (впрочем, в противном случае они не могли бы так называться). Наиболее «богатым» из них является пространство H_1 , содержащее случайные события, соответствующие всем интервалам, полуинтервалам, отрезкам и точкам из Ω (что это так, читатель проверит по аналогии с свойствами рассмотренной выше борелевской σ -алгебры).

Вероятностное пространство H_2 «беднее», чем H_1 ; в нем не содержится, например, случайное событие $B = \left\{ \omega \in [0, \frac{1}{4}) \right\}$. Вместе с тем это случайное событие присутствует в еще в более «бедном» вероятностном пространстве H_3 .

В дальнейшем изложении при рассмотрении той или иной стохастической модели по умолчанию будет постоянно предполагаться, что существует соответствующее вероятностное пространство и фигурирующие в модели события являются относительно этого пространства случайными (если не оговорено противное).

Сформулированная система аксиом теории вероятностей непротиворечива и неполна; первое означает, что ей удовлетворяет хотя бы одно распределение вероятностей, второе – что такое распределение не единственно. В этом мы будем ниже постоянно убеждаться.

Приведем ряд следствий, непосредственно вытекающих из сформулированных аксиом.

Следствие 2.1. Для каждого (случайного) события A вероятность его отрицания \bar{A} равна

$$P(\bar{A}) = 1 - P(A), \quad (2.5)$$

что следует из очевидного равенства $A \cup \bar{A} = \Omega$ и из аксиом (2.4)" и (2.4)'''.

Следствие 2.2. Пусть $B \subseteq A$ (т.е., иначе, $B \rightarrow A$). Тогда

$$P(B) \leq P(A). \quad (2.6)$$

Это следует из равенства $A = B \cup A \cdot \bar{B}$ и из (2.4)''': $P(A) = P(B) + P(A \cdot \bar{B}) \geq P(B)$ (где $P(A \cdot \bar{B}) \geq 0$ как вероятность F – измеримого события).

Следствие 2.3. (Теорема сложения вероятностей). Пусть A_1 и A_2 – совместимые события, определенные на общем вероятностном пространстве. Тогда

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cdot A_2). \quad (2.7)$$

Чтобы убедиться в этом, следует использовать равенства

$$A_1 \cup A_2 = A_1 \cup \bar{A}_1 \cdot A_2, \quad A_2 = A_1 \cdot A_2 \cup \bar{A}_1 \cdot A_2$$

(см. (1.2), (1.3)) и аксиому (2.4)'''.

Из (2.7) по индукции можно получить общее выражение для вероятности дизъюнкции n совместимых определенных на общем вероятностном пространстве событий ($n \geq 2$), которое (как читатель легко проверит) имеет вид

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cdot A_{i_2}) + \\ &+ \sum_{i_1 < i_2, i_2 < i_3} P(A_{i_1} \cdot A_{i_2} \cdot A_{i_3}) - \dots + (-1)^{n-1} P(A_1 \cdot \dots \cdot A_n) \end{aligned} \quad (2.8)$$

В силу громоздкости этой формулы на практике для вычисления искомой вероятности чаще используется соотношение

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcap_{i=1}^n \bar{A}_i\right), \quad (2.9)$$

которое очевидным способом выводится из формулы де Моргана (1.4)

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i, \quad n \leq \infty; \quad (2.10)$$

смысл этого равенства прост: слева стоит событие, состоящее в отрицании того, что из n событий $\{A_i\}$ произошло хотя бы одно событие, а справа – событие, означающее, что не произошло ни одного из n событий. Ясно, что эти события тождественны.

Соотношения (2.8) и (2.9) были получены нами в предположении, что $n < \infty$. Ниже будет рассмотрен случай $n = \infty$.

Как уже было сказано, аксиомы теории вероятностей рассматривают вероятность как нормированную единицей (см. (2.4)'') функцию (вероятностную меру) множеств из F , обладающую свойством конечной аддитивности (при $n < \infty$) или счетной аддитивности (при $n = \infty$) (см. (2.4)'''). Последнее свойство вероятности (счетная аддитивность) проявляется в тех случаях, когда F – σ -алгебра; если $F = A$ – алгебра, то для исчисления вероятностей событий достаточно свойства конечной аддитивности вероятности (последнее имеет место, например, в тех случаях, когда в вероятностной модели используется классическое определение вероятности (см. главу 1)).

Постулируемая счетная аддитивность вероятности влечет за собой свойство ее непрерывности, смысл которого раскрывается ниже.

Назовем последовательности случайных событий (т.е. измеримых множеств) $\{C_i\}_{i=1}^{\infty}$ и $\{D_i\}_{i=1}^{\infty}$ соответственно монотонно неубывающей и монотонно невозрастающей, если $C_1 \subseteq C_2 \subseteq \dots \subseteq C_n \subseteq \dots$ и $D_1 \supseteq D_2 \supseteq \dots \supseteq D_n \supseteq \dots$. Очевидны равенства:

$$\bigcup_{i=1}^n C_i = C_n, \quad \bigcap_{i=1}^n D_i = D_n.$$

Назовем пределами дизъюнкции $\bigcup_{i=1}^n C_i$ и конъюнкции $\bigcap_{i=1}^n D_i$ соответственно случайные события

$$C = \bigcup_{i=1}^{\infty} C_i \text{ и } D = \bigcap_{i=1}^{\infty} D_i.$$

Следствие 2.4 (теорема о непрерывности вероятности). Вероятностная мера событий обладает свойством непрерывности, которое при принятых обозначениях выражается равенствами

$$a) P(C) = \lim_{n \rightarrow \infty} P(C_n); \quad b) P(D) = \lim_{n \rightarrow \infty} P(D_n).$$

Докажем п. а). Используя очевидное тождество

$$C = \bigcup_{i=1}^{\infty} C_i = C_1 \cup C_2 \cup \dots = C_1 \cup \bar{C}_1 \cdot C_2 \cup \bar{C}_2 \cdot C_3 \cup \dots = C_n \cup \left(\bigcup_{i \geq n} \bar{C}_i \cdot C_{i+1} \right),$$

из аксиомы 3 получаем

$$\begin{aligned} P(C) &= P(C_1) + P(\bar{C}_1 \cdot C_2) + P(\bar{C}_2 \cdot C_3) + \dots + P(\bar{C}_i \cdot C_{i+1}) + \dots = P(C_n) + P\left(\bigcup_{i \geq n} \bar{C}_i \cdot C_{i+1}\right) = \\ &= P(C_n) + \sum_{i \geq n} P(\bar{C}_i \cdot C_{i+1}). \end{aligned}$$

Ввиду существования $P(C)$ ($P(C) \leq 1$) ряд справа сходится, т.е.

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i \geq n} \bar{C}_i \cdot C_{i+1}\right) = 0 \text{ и } \lim_{n \rightarrow \infty} P(C_n) = P(C);$$

п. а) теоремы доказан. Доказательство п. б) читатель легко получит, применяя формулу де Моргана, переходя к событиям $\{\bar{D}_i\}$ и используя результат п. а).

Иногда теорема о непрерывности вероятности принимается за аксиому, что позволяет снизить требование счетной аддитивности вероятностной меры до условия ее конечной аддитивности (т.е. заменить в аксиоме 3 условие $n \leq \infty$ на условие $n < \infty$).

В первой главе указывалось, что равенство единице или нулю вероятностей событий не означает их, соответственно, достоверность или невозможность (в принятом нами смысле). Убедимся в истинности этого утверждения, используя вероятностное пространство H_1 в примере 2.4 «а». Для этого рассмотрим одноточечное множество $A^* = \{\omega: \omega = c\}$, т.е. событие $A = \{\omega = c\}$ (c – внутренняя точка полуинтервала $[0,1)$) и убедимся, что в вероятностном пространстве H_1 $P(A) = 0$. Действительно, введем события $A_n = \{\omega \in \Delta_n\}$, где $\Delta_n = [c - 1/n, c + 1/n)$, и для простоты положим $n = n_0, n_0 + 1, n_0 + 2, \dots$, где n_0 – минимальное число, при котором $\Delta_n \subset [0,1)$. Тогда

$$P(A_n) = \frac{2}{n}, \quad A = \bigcap_{n=n_0}^{\infty} A_n \text{ и по теореме о непрерывности вероятности } P(A) = \lim_{n \rightarrow \infty} P(A_n) = 0,$$

т.е. событие A имеет нулевую вероятность (т.е. является «нулевым»), не будучи невозможным ($A \neq \emptyset$).

В вероятностном пространстве H_1 нулевыми (но не невозможными) оказываются и события, соответствующие множествам, образованным из конечного или счетного числа точек.

В нашем примере нетрудно построить и события, имеющие единичную вероятность, не являясь достоверным. К таким событиям относится, например, событие $B = \Omega \setminus A$, соответствующее множеству Ω с «выколотой» точкой ω .

Подобные факты могут иметь место, конечно, и в других вероятностных пространствах.

Вернемся к соотношениям (2.8) и (2.9), чтобы выяснить, как они преобразуются при $n = \infty$.

Используя обозначения, принятые в этих равенствах, введем события

$$\begin{aligned}
C_1 &= A_1; \\
C_2 &= C_1 \cup \bar{C}_1 \cdot A_2; \\
C_3 &= C_2 \cup \bar{C}_2 \cdot A_3; \\
&\dots\dots\dots \\
C_i &= C_{i-1} \cup \bar{C}_{i-1} \cdot A_i; \\
&\dots\dots\dots
\end{aligned}
\tag{2.11}$$

Нетрудно убедиться, что

$$C_n = \bigcup_{i=1}^n A_i, \quad \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} C_i \text{ и } C_1 \subseteq C_2 \subseteq \dots,$$

и согласно теореме о непрерывности вероятности (п. а)) получаем вместо (2.8) соотношение

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(C_n), \tag{2.12}$$

использование которого на практике (с учетом (2.11)) достаточно сложно. В то же время соотношение (2.10) при $n = \infty$ сохраняет относительно простой вид

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1 - P\left(\bigcap_{i=1}^{\infty} \bar{A}_i\right)$$

(отметим, что при переходе к $n = \infty$ фигурирующие здесь события остаются элементами σ -алгебры F и согласно аксиоме 1 для них определены вероятности).

В рассматриваемых ниже задачах у нас часто будут фигурировать полные группы событий, отвечающие равенству $\{A_i\}_{i=1}^n = \Omega$, $n \leq \infty$, означающему, что в результате опыта происходит хотя бы одно из событий, образующую группу. Нетрудно установить, что для полной группы несовместимых событий имеет место равенство

$$\sum_{i=1}^n P(A_i) = 1,$$

а для группы, в которой присутствуют совместимые события, справедливо неравенство

$$\sum_{i=1}^n P(A_i) > 1.$$

Все полученные нами выражения для вероятности дизъюнкции совместимых событий (начиная с (2.7)) содержали вероятности конечных и счетных конъюнкций случайных событий, т.е. булевых функций, вероятности которых определены (т.е. имеют смысл) для рассматриваемого вероятностного пространства. Пока, однако, остался в стороне вопрос о правилах исчисления вероятностей конъюнкций случайных событий. Этот вопрос рассматривается в следующей главе.

Условная вероятность. Стохастическая зависимость событий

Аксиомы (2.4), представляя собой основу математической теории вероятностей, формализуют понятие вероятности события, не претендуя на содержательный («физический») ее смысл, как на меру неопределенности. Решая ту или иную задачу, требующую вероятностной трактовки, исследователь пользуется исходной оценкой вероятности события, основанной на классической, статистической или геометрической ее интерпретации, применяя затем, по мере необходимости, аппарат математической теории вероятностей. Это заставляет нас нередко возвращаться к содержательному смыслу вероятности (заметим, что необходимость постоянно сочетать физический и математический подход к описанию реальных явлений является характерной чертой вероятностного моделирования).

Как уже говорилось, случайность события обусловлена неопределенностью неконтролируемых условий, от которых зависит его появление («успех») или непоявление («неудача»). Существенно, что вариация этих условий от опыта к опыту происходит так, что в серии опытов имеют место и «успехи» и «неудачи». В примере с бросанием игральной кости этими неконтролируемыми варьируемыми условиями являются начальные положение кости и приложенный к ней импульс; в примере (2.3) – способность и прилежность студента.

Аксиомы теории вероятности постулируют существование определенной стационарности неконтролируемых условий опыта, приводящей к стохастической устойчивости результатов серии опыта, что позволяет приписывать наблюдаемым событиям вероятности их появления.

В некоторых случаях можно полагать, что различные серии опытов идентичны по их условиям и вероятность исследуемого события для этих серий совпадает. Нередки, однако, случаи, когда на результаты опыта воздействуют факторы, остающиеся неизменными внутри серии опыта, но изменяющиеся от серии к серии (часто – случайным образом). В результате этого вероятность события также может меняться от серии к серии.

Так, к примеру, пусть проводятся серии бросания двух ассиметричных костей, различающихся по положению их центра тяжести. Тогда можно ожидать, что вероятность интересующего нас события (например – выпадения «шестерки») для этих костей будет различна. Такое положение имеет место и в примере (2.3): здесь новым условием, влияющим на результат зачета, является вызов студента тем или иным преподавателем.

Эти рассуждения и примеры приводят к понятию *условных вероятностей* событий, т.е. вероятностей их осуществления при фиксированных заданных условиях.

Условная вероятность

Начнем с простого случая, когда условие, влияющее на появление случайного события A , состоит в осуществлении некоторого другого события B (оба события, конечно, определены на общем вероятностном пространстве). Задача заключается в вычислении вероятности события A при условии, что осуществилось событие B , и в сравнении ее со значением вероятности события A , вычисленном без этого условия.

Введем определение.

Определение 3.1. Пусть A и B – случайные события, $P(B) \neq 0$. *Условной вероятностью события A относительно события B (т.е. вероятностью появления события A при условии, что осуществилось событие B), называется величина*

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}. \quad (3.1)$$

Это формальное определение условной вероятности является обобщением этого понятия, вытекающего из интуитивных представлений смысла вероятности. Действительно, для классического определения вероятности в схеме равновозможных исходов опыта условную вероятность $P(A|B)$ естественно положить равной отношению исходов, благоприятствующих совместному появлению событий A и B (n_{AB}) к числу исходов, благоприятствующих событию B (n_B):

$$P(A|B) = \frac{n_{AB}}{n_B};$$

Разделив числитель и знаменатель на общее число исходов n , получим (3.1). Подобный результат получается и при статистическом определении вероятностей.

Нетрудно установить, что условная вероятность отвечает аксиомам (2.4) при следующей модификации вероятностного пространства.

Множество исходов Ω заменяется множеством $\Omega_B = \{\omega: \omega \in B^*\}$ (B^* – подмножество множества Ω , соответствующее событию B , см. (1.1)), σ -алгебра в Ω заменяется σ_B -алгеброй в Ω_B , элементами которой являются события $A \cdot B$ (где A – элементы σ -алгебры в Ω), а вероятностные меры для этих событий определены (3.1).

Определение (3.1) можно трактовать как следующую теорему.

Теорема 3.1. Теорема умножения вероятностей (для двух событий). Вероятность конъюнкции двух случайных событий A и B , определенных на общем вероятностном пространстве, равна

$$P(A \cdot B) = P(A)P(B | A) = P(B)P(A | B). \quad (3.2)$$

Мы можем теперь найти правильное решение примера 2.3. Для этого введем единое вероятностное пространство, которое будет содержать все присутствующие в задаче события. Множества исходов Ω_1 и Ω_2 вероятностных пространств H_1 и H_2 содержат по два исхода – (C_1, \bar{C}_1) при событии A (преподаватель A) и (C_2, \bar{C}_2) при событии B (преподаватель B). Объединенное вероятностное пространство H (называемое произведением вероятностных пространств: $H = H_1 \times H_2$) имеет множество исходов $\Omega = \Omega_1 \times \Omega_2$ с четырьмя исходами: $A \cdot C_1, A \cdot \bar{C}_1, B \cdot C_2, B \cdot \bar{C}_2$. Для успешной сдачи зачета отсюда получаем $C = A \cdot C_1 \cup B \cdot C_2$ и (согласно (3.2))

$$P(C) = P(A \cdot C_1) + P(B \cdot C_2) = P(C_1 | A)P(A) + P(C_2 | B)P(B) = 0,7P(A) + 0,8P(B) \leq 1$$

(здесь $P(A)$ и $P(B)$ вероятности вызова студента к преподавателю A и B соответственно).

В определении 3.1 и далее мы полагаем, что $P(B) \neq 0$, т.е. событие B ненулевое. Ниже встретятся и будут обсуждены случаи, когда это предположение не выполняется.

Стохастическая (вероятностная) зависимость событий.

Введя понятие условной вероятности, мы можем приступить к изучению такого важного вопроса, как взаимозависимость случайных событий.

Содержательно зависимость случайного события A от случайного события B определяется зависимостью вероятности события A от того, произошло событие B , или нет. Используя условные вероятности, существование этой зависимости определяется неравенством

$$P(A | B) \neq P(A | \bar{B}); \quad (3.3)$$

независимость A от B выражается равенством

$$P(A | B) = P(A | \bar{B}). \quad (3.4)$$

В качестве упражнения читатель может легко показать, что (3.3) и (3.4) эквивалентны, соответственно, соотношениям

$$P(A | B) \neq P(A) \quad (3.5)$$

и

$$P(A | B) = P(A). \quad (3.6)$$

Важно отметить, что вероятностная связь случайных событий обладает, в отличие от причинно-следственной связи, свойством симметрии: если событие A зависит (не зависит) от

события B , то и событие B зависит (не зависит) от события A . В этом читатель легко убедится, если из (3.5) и (3.6) получит, соответственно, $P(B|A) \neq P(B)$ и $P(B|A) = P(B)$.

Свойство симметричности вероятностной связи означает, что для двух случайных событий, из которых одно является причиной, а другое – следствием, их вероятностная связь проявляется как в зависимости следствия от причины, так и причины от следствия (вероятность успешной сдачи студентом экзамена зависит от его усердной подготовки и, наоборот, успешная сдача им экзамена указывает на высокую вероятность его усердной подготовки).

Из (3.2) и (3.6) следует, что для независимых случайных событий вероятность их конъюнкции равна произведению их вероятностей

$$P(A \cdot B) = P(A)P(B). \quad (3.7)$$

Равенство (3.7) используется обычно как определение независимости двух событий. Действительно, нетрудно убедиться, что из (3.7) следуют равенства (3.4) и (3.6). Заметим, что в математическом отношении условие независимости событий в форме (3.7) предпочтительней, поскольку включает и случай, когда одно из событий имеет нулевую вероятность. Однако, с другой стороны, условия независимости в форме (3.4) или (3.6) имеют более содержательный смысл и на практике чаще используются для экспериментальной или интуитивной проверки независимости событий.

Обобщим все сказанное в следующем определении.

Определение 3.2. Случайные события A и B независимы, если выполняются эквивалентные равенства (3.4), (3.6) и (3.7).

Читателю предлагается проверить, что зависимость (независимость) событий A и B влечет за собой зависимость (независимость) событий A и \bar{B} , \bar{A} и B , \bar{A} и \bar{B} .

Выясним теперь, как результаты, полученные нами для двух событий, распространяются на случай n событий ($n > 2$).

Обозначим $K = \{A_i\}_{i=1}^n$ группу событий, определенных на общем вероятностном пространстве, и займемся изучением возможных форм их взаимозависимости.

Для каждой пары событий из K их связь, как было установлено, выражается условными вероятностями и теоремой произведения вероятностей. При $n > 2$, однако, возникает новая задача, состоящая в необходимости учета и описания взаимосвязей между подгруппами событий, входящих в K .

Пусть $K_1 = \{A_i^{(1)}\}_{i=1}^{n_1}$ и $K_2 = \{A_j^{(2)}\}_{j=1}^{n_2}$ – две непересекающиеся подгруппы событий из K : $K_1 \subset K$, $K_2 \subset K$, $K_1 \cdot K_2 = \emptyset$. Чтобы описать вероятностные связи между этими подгруппами, следовало бы, казалось, рассмотреть связи между различными булевыми функциями,

образуемыми событиями из K_1 и K_2 (по отдельности). Задача, однако, упрощается ввиду того, что все булевы функции могут быть выражены операциями конъюнкции и отрицания, образующими полную группу булевых операций. Поэтому для описания связей между группами событий достаточно использовать операции конъюнкции и отрицания.

Обозначим K_1 и K_2 конъюнкции событий, входящих, соответственно, в K_1 и K_2 . Ввиду (3.1) и (3.7)) для условной вероятности конъюнкции K_1 (относительно K_2) получим

$$P(K_1 | K_2) = \frac{P(K_1 \cdot K_2)}{P(K_2)} = \frac{P\left(\left(\bigcap_{i=1}^{n_1} A_i^{(1)}\right) \cap \left(\bigcap_{j=1}^{n_2} A_j^{(2)}\right)\right)}{P\left(\bigcap_{j=1}^{n_2} A_j^{(2)}\right)}, \quad (3.8)$$

а независимость конъюнкций K_1 и K_2 выразится равенством

$$P(K_1 \cdot K_2) = P(K_1)P(K_2) = P\left(\bigcap_{i=1}^{n_1} A_i^{(1)}\right)P\left(\bigcap_{j=1}^{n_2} A_j^{(2)}\right). \quad (3.9)$$

(заметим, что здесь и далее под конъюнкцией понимаются и отдельные события – «одноэлементные конъюнкции»).

Введем теперь новый тип независимости событий – *независимость в совокупности*.

Определение 3.3. Группа событий $K = \{A_i\}_{i=1}^n$ независимы в совокупности, если для любой пары непересекающихся подгрупп событий $K_1 = \{A_i^{(1)}\}_{i=1}^{n_1}$ и $K_2 = \{A_j^{(2)}\}_{j=1}^{n_2}$ ($K_1 \subset K$, $K_2 \subset K$, $K_1 \cdot K_2 = \emptyset$) конъюнкции их событий K_1 и K_2 независимы, т.е. выполняется (3.9).

Это определение распространяется и на случай $K = K^\infty = \{A_i\}_{i=1}^\infty$, если независимы в совокупности все конечные подгруппы группы событий K^∞ .

Иногда для стилистического удобства независимость событий в совокупности будет именоваться *групповой независимостью событий*.

Приведенное определение имеет достаточно ясный содержательный практический смысл: независимость случайных событий в совокупности означает, что вероятность любого события, представляющего собой булеву функцию событий из подгруппы K_1 , входящей в K , не зависит от того, как «ведут себя» события из K , не входящих в K_1 .

Более удобное в математическом отношении (и обычно используемое в учебниках) определение независимости в совокупности имеет следующий вид.

Определение 3.4. Группа событий $K = \{A_i\}_{i=1}^n$ независимы в совокупности, если для любых k , $k \leq n$, $(i_1, i_2, \dots, i_k) \subseteq (1, 2, \dots, n)$ имеет место равенство

$$P(A_{i_1} \cdot A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) \quad (3.10)$$

(здесь (i_1, i_2, \dots, i_k) – набор из $(1, 2, \dots, n)$).

Нетрудно убедиться, что определения 3.3 и 3.4 эквивалентны: из первого следует второе и наоборот (читателю полезно проверить это самостоятельно).

Элементами конъюнкции K_1 и K_2 в (3.8) и (3.9) являются события из K . Возникает вопрос, не изменится ли критерий групповой независимости событий, если в эти конъюнкции мы будем включать и их отрицания. Оказывается, что такое расширение задачи не вносит изменений в определения 3.3 и 3.4, т.к. из фигурирующих в них условий следует групповая независимость событий и их отрицаний. Действительно, пусть выполняется (3.10) и рассмотрим конъюнкцию $A_{i_1} \cdot A_{i_2} \cdots A_{i_k} \bar{A}_j$ ($A_j \in K$). Имеем

$$A_{i_1} \cdot A_{i_2} \cdots A_{i_k} \bar{A}_j \cup A_{i_1} \cdot A_{i_2} \cdots A_{i_k} A_j = A_{i_1} \cdot A_{i_2} \cdots A_{i_k},$$

$$P(A_{i_1} \cdot A_{i_2} \cdots A_{i_k} \bar{A}_j) = (1 - P(A_j)) \prod_{s=1}^k P(A_{i_s}) = P(\bar{A}_j) \prod_{s=1}^k P(A_{i_s}),$$

т.е. выполнение условия независимости в совокупности (3.10) автоматически распространяется на совокупность событий из K и их отрицаний.

При построении вероятностных моделей свойство независимости событий в совокупности нередко используется по следующей схеме. На основании эвристических соображений постулируется выполнение условий определения 3.3 и затем, как следствие, используются соотношения (3.10), существенно упрощающие модель. Отсюда лишнее можно заключить, что корректность исходных эвристических посылок требует строгой проверки.

Укажем, в связи с этим, на некоторые факты, игнорирование которых может приводить к серьезным ошибкам и упущениям.

Прежде всего, установим, что при $n > 2$ условие независимости группы событий в совокупности сильнее условия их попарной независимости. Действительно, второй тип независимости событий является следствием первого, в то время как из попарной независимости событий не следует их независимость в совокупности. Последнее утверждение не очевидно и требует проверки, которую проведем, используя следующий пример.

Пример 3.1. Пусть опыт состоит в наблюдении трех событий A_1 , A_2 и A_3 , каждое из которых появляется или не появляется в результате опыта. Каждый результат опыта (из восьми возможных) обозначим тройкой $(A_1^{\alpha_1}, A_2^{\alpha_2}, A_3^{\alpha_3})$, где α_j ($j = 1, 2, 3$) принимают значения 0 или 1 и $A_j^0 = \bar{A}_j$, $A_j^1 = A_j$. Допустим, далее, что для всех конъюнкций $A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3}$ известны их вероятности $P(A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3})$, по значениям которых следует найти характер стохастической зависимости между событиями A_1 , A_2 и A_3 .

Для наглядности сопроводим этот пример следующей иллюстрацией. Представим себе, что плоский восьмиугольник, сторонам которого взаимно однозначно соответствуют тройки $(A_1^{\alpha_1}, A_2^{\alpha_2}, A_3^{\alpha_3})$, случайным образом (путем вбрасывания) располагается на прямой, опираясь на нее одной из своих сторон. Результатом каждого такого опыта является реализация тройки событий $(A_1^{\alpha_1}, A_2^{\alpha_2}, A_3^{\alpha_3})$, соответствующей выпавшей (нижней) стороне многоугольника. Предположим, что вероятность выпадения каждой стороны пропорциональна ее длине, и построим таблицу 3.1, в которой каждому набору $(\alpha_1, \alpha_2, \alpha_3)$, т.е. тройке $(A_1^{\alpha_1}, A_2^{\alpha_2}, A_3^{\alpha_3})$, поставлена в соответствие вероятность их конъюнкции $P(A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3})$.

Таблица 3.1

α_1	α_2	α_3	$P(A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3})$
0	0	0	$p_1 = 0.125 + \varepsilon$
0	0	1	$p_2 = 0.125 - \varepsilon$
0	1	0	$p_3 = 0.125 - \varepsilon$
0	1	1	$p_4 = 0.125 + \varepsilon$
1	0	0	$p_5 = 0.125 - \varepsilon$
1	0	1	$p_6 = 0.125 + \varepsilon$
1	1	0	$p_7 = 0.125 + \varepsilon$
1	1	1	$p_8 = 0.125 - \varepsilon$

Здесь $0 \leq \varepsilon \leq 0.125$, $\sum_{i=1}^8 p_i = 1$ и значения p_i зависят от формы многоугольника; при симметричном многоугольнике ($\varepsilon = 0$) для $\forall i$ $p_i = 0.125$.

Для того, чтобы выяснить характер зависимости между событиями A_1 , A_2 и A_3 (т.е. определить, являются ли они зависимыми или независимыми попарно или в совокупности), найдем, используя данные, представленные в таблице, вероятности этих событий $\{P(A_i)\}$, их парных конъюнкций $\{P(A_i \cdot A_j)\}$ и общей конъюнкции $P(A_1 \cdot A_2 \cdot A_3)$.

Так, для вычисления вероятности $P(A_1)$ используем очевидное равенство

$$A_1 \cdot \bar{A}_2 \cdot \bar{A}_3 \cup A_1 \cdot \bar{A}_2 \cdot A_3 \cup A_1 \cdot A_2 \cdot \bar{A}_3 \cup A_1 \cdot A_2 \cdot A_3 = A_1 \cdot \Omega = A_1,$$

откуда (ввиду несовместимости событий, образующих дизъюнкцию слева)

$$P(A_1) = P(A_1 \cdot \bar{A}_2 \cdot \bar{A}_3) + P(A_1 \cdot \bar{A}_2 \cdot A_3) + P(A_1 \cdot A_2 \cdot \bar{A}_3) + P(A_1 \cdot A_2 \cdot A_3) = p_5 + p_6 + p_7 + p_8.$$

Для $P(A_2)$ и $P(A_3)$ аналогично получим

$$P(A_2) = p_3 + p_4 + p_7 + p_8; \quad P(A_3) = p_2 + p_4 + p_6 + p_8.$$

Далее, используя равенство

$$A_1 \cdot A_2 = A_1 \cdot A_2 \cdot \bar{A}_3 \cup A_1 \cdot A_2 \cdot A_3,$$

находим $P(A_1 \cdot A_2) = p_7 + p_8$, а затем, аналогично,

$$P(A_1 \cdot A_3) = p_6 + p_8 \text{ и } P(A_2 \cdot A_3) = p_4 + p_8.$$

Наконец, $P(A_1 \cdot A_2 \cdot A_3) = p_8$.

Если $\forall i \ p_i = 0.125$ (случай симметричного многоугольника) то для $\forall i, j$ получаем $P(A_i) = 0.5$, $P(A_i \cdot A_j) = P(A_i)P(A_j) = 0.25$ и $P(A_1 \cdot A_2 \cdot A_3) = P(A_1)P(A_2)P(A_3) = 0.125$, т.е. события A_1, A_2 и A_3 независимы и попарно и в совокупности.

Пусть теперь наш многоугольник несимметричен, т.е. $\varepsilon > 0$ и

$$p_1 = p_4 = p_6 = p_7 = 0.125 + \varepsilon,$$

$$p_2 = p_3 = p_5 = p_8 = 0.125 - \varepsilon.$$

Тогда

$$P(A_1) = P(A_2) = P(A_3) = 0.5$$

$$P(A_1 \cdot A_2) = P(A_1)P(A_2) = P(A_1 \cdot A_3) = P(A_1)P(A_3) =$$

$$= P(A_2 \cdot A_3) = P(A_2)P(A_3) = 0.25,$$

$$P(A_1 \cdot A_2 \cdot A_3) = 0.125 - \varepsilon \neq P(A_1)P(A_2)P(A_3) = 0.125,$$

т.е. события A_1, A_2 и A_3 обладают попарной независимостью, но не являются независимыми в совокупности. Поэтому вывод о групповой независимости событий, сделанный по результату проверки только их попарной независимости, может приводить к ошибке, которая в нашем примере выражается в завышении вероятности $P(A_1 \cdot A_2 \cdot A_3)$ – совместного осуществления событий A_1, A_2 и A_3 .

Этот пример подтверждает высказанное выше утверждение о соотношении попарной и групповой независимости событий.

Оперируя с группами случайных событий, исследователь может допускать (наряду с ошибкам описанного типа) ошибки, состоящие в упущении возможности получить положительный эффект от групповой зависимости между событиями. Опишем такую ситуацию несколько подробнее.

Во многих областях практического приложениях стохастических методов решается задача следующего типа. Пусть исследователь ведет мониторинг появления некоторого случайного события B , которое им непосредственно не наблюдается, но стохастически связано с группой наблюдаемых событий $\{A_i\}_{i=1}^n$, которые назовем *событиями – признаками*. Задача сводится к вычислению вероятности события B при различных реализациях событий из группы $\{A_i\}_{i=1}^n$; предполагается, что все необходимые для этого исходные данные известны. Оставляя до времени общее изучение задач такого типа, укажем, что

существенную роль в них играет оценка эффективности используемых событий-признаков, определяемой вероятностной связью между ними и исследуемым событием. Ясно, что независимость события B и отдельного признака A_i означает его нулевую индивидуальную эффективность, поскольку вероятность события B в этом случае не зависит от того, произошло событие A_i или нет.

Пусть в результате статистического эксперимента установлено, что событие B попарно независимо со всеми событиями-признаками из группы $\{A_i\}_{i=1}^n$. Можно ли отсюда заключить, что эта группа не может служить групповым признаком в задаче обнаружения события B ?

Следующий пример показывает, что такой вывод несостоятелен.

Пример 3.2. Пусть $n = 2$ и, сохраняя обозначения, принятые в предыдущем примере, обозначим B^β событие, равное \bar{B} при $\beta = 0$ и B при $\beta = 1$. Исходные данные содержатся в таблице 3.2, задающей, аналогично таблице 3.1, распределение вероятностей на конъюнкциях $(B^\beta \cdot A_1^{\alpha_1} \cdot A_2^{\alpha_2})$.

Таблица 3.2

α_1	α_2	β	$P(A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot B^\beta)$
0	0	0	$p_1 = 20/36$
0	0	1	$p_2 = 5/36$
0	1	0	$p_3 = 5/36$
0	1	1	$p_4 = 0$
1	0	0	$p_5 = 5/36$
1	0	1	$p_6 = 0$
1	1	0	$p_7 = 0$
1	1	1	$p_8 = 1/36$

Задача состоит в оценке эффективности событий A_1 и A_2 как событий-признаков при мониторинге события B . Из таблицы получаем

$$P(A_1) = p_5 + p_6 + p_7 + p_8 = P(A_2) = p_3 + p_4 + p_7 + p_8 = P(B) = p_2 + p_4 + p_6 + p_8 = \frac{1}{6};$$

$$P(A_1 \cdot B) = p_6 + p_8 = P(A_2 \cdot B) = p_4 + p_8 = \frac{1}{36},$$

т.е. событие B и признаки A_1 и A_2 попарно независимы и каждый из признаков в отдельности имеет нулевую эффективность. Однако, вместе с тем, получаем равенства

$$\begin{aligned}
P(B | A_1 \cdot A_2) &= \frac{P(A_1 \cdot A_2 \cdot B)}{P(A_1 \cdot A_2)} = \frac{p_8}{p_7 + p_8} = 1, \\
P(B | \bar{A}_1 \cdot A_2) &= \frac{P(\bar{A}_1 \cdot A_2 \cdot B)}{P(\bar{A}_1 \cdot A_2)} = \frac{p_4}{p_3 + p_4} = 0, \\
P(B | A_1 \cdot \bar{A}_2) &= \frac{P(A_1 \cdot \bar{A}_2 \cdot B)}{P(A_1 \cdot \bar{A}_2)} = \frac{p_6}{p_5 + p_6} = 0, \\
P(B | \bar{A}_1 \cdot \bar{A}_2) &= \frac{P(\bar{A}_1 \cdot \bar{A}_2 \cdot B)}{P(\bar{A}_1 \cdot \bar{A}_2)} = \frac{p_2}{p_1 + p_2} = \frac{1}{5},
\end{aligned}$$

свидетельствующие о высокой групповой эффективности признаков.

Использование эффекта «групповой полезности» признаков затруднено, конечно, необходимостью иметь объемную исходную информацию (представленную в нашем примере таблицей 3.2), причем задача быстро усложняется с ростом числа событий-признаков. Пути ее решения состоят в сочетании теоретического анализа и интуитивных посылок, а также и в применении методов стохастического моделирования.

Укажем еще на распространенную ошибку, состоящую в замене набора равенств (3.10) в определении 3.4 единственным равенством $P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n)$, которое, однако, не является достаточным условием независимости событий A_1, \dots, A_n в совокупности. Приведем пример.

Пример 3.3. Пусть вероятности конъюнкций $A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3}$ заданы таблицей 3.3 (обозначения имеют тот же смысл, что и в примере 3.1).

Таблица 3.3

Здесь имеем:

α_1	α_2	α_3	$P(A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot A_3^{\alpha_3})$
0	0	0	$0.125 + \varepsilon$
0	0	1	$0.125 - \varepsilon$
0	1	0	$0.125 - \varepsilon$
0	1	1	$0.125 + \varepsilon$
1	0	0	0.125
1	0	1	0.125
1	1	0	0.125
1	1	1	0.125

$$\begin{aligned}
P(A_1) &= P(A_2) = P(A_3) = \frac{1}{2}, \\
P(A_1 \cdot A_2 \cdot A_3) &= \frac{1}{8} = P(A_1)P(A_2)P(A_3),
\end{aligned}$$

но при $\varepsilon > 0$ $P(A_2 \cdot A_3) = \frac{1}{4} + \varepsilon \neq P(A_2)P(A_3) = \frac{1}{4}$, т.е. условие (3.10) независимости событий в совокупности не выполняется.

Вернемся теперь к (3.8), чтобы сформулировать теорему умножения вероятностей для группы n событий ($n \geq 2$) (в развитие теоремы 3.1).

Теорема 3.2. Теорема умножения вероятностей. Пусть $\{A_i\}_{i=1}^n$ группа конечного числа случайных событий, принадлежащих общему вероятностному пространству. Вероятность их конъюнкции равна

$$P(A_1 \cdot A_2 \cdots A_n) = P(A_1 | A_2 \cdots A_n)P(A_2 | A_3 \cdots A_n) \cdots P(A_{n-1} | A_n)P(A_n). \quad (3.11)$$

Легко видеть, что это утверждение теоремы есть простое следствие итерационного применения (3.1) и (3.8).

О соотношении между вероятностными и причинно-следственными связями между событиями.

Распространенной ошибкой, допускаемой исследователями-практиками при стохастическом анализе реальных явлений, является отождествление вероятностных и причинно-следственных (функциональных) связей между событиями.

Изучая феномен стохастической (вероятностной) связи между событиями, мы обнаруживаем существенное отличие ее свойств от свойств, присущих причинно-следственным отношениям между событиями. Как уже отмечалось, это проявляется в симметричности вероятностных связей, которой причинно-следственные связи не обладают. Поэтому из обнаружения между событиями стохастической зависимости нельзя безоговорочно сделать вывод о существовании между ними причинной связи.

Пусть, к примеру, в результате обработки большого массива данных обнаружена значимая вероятностная связь между двумя событиями A_1 и A_2 , выражающаяся в неравенстве $P(A_1 | A_2) > P(A_1)$, которое расценивается наблюдателем как свидетельство положительной причинной роли события A_2 по отношению к событию A_1 (следствию). Этот вывод, сделанный на основании результата статистической обработки данных, относящихся только к событиям A_1 и A_2 , может, однако, быть ошибочным, поскольку, во-первых, он с наименьшим основанием свидетельствует о причинной роли события A_1 по отношению к событию A_2 и, во-вторых (и это – главное), полученный результат вообще не доказывает существования между событиями A_1 и A_2 причинно-следственных отношений, поскольку обнаруженная вероятностная связь между ними может являться следствием их зависимости

от третьего события A_3 , оставшегося вне контроля. Рассмотрим возможность такой ситуации в контексте следующей реальной задачи.

Существует обширная статистика, свидетельствующая о существовании значимой вероятностной зависимости между здоровьем человека и его пищевым рационом. Интересно выяснить, в какой мере эта статистика может быть использована для рекомендаций по выбору полезной диеты.

Пусть, в частности, в результате широкого сбора данных установлено, что люди, в рационе которых преобладают рыбные продукты, реже других страдают от сердечных заболеваний (такой факт подтверждается рядом публикаций). Выясним, можно ли отсюда заключить, что увеличением доли рыбных продуктов в рационе населения можно сократить распространенность сердечных заболеваний. Подчеркнем, что используемая здесь статистика получена в результате «пассивного» эксперимента. В этих условиях полученная вероятностная связь между событиями A («рыбный рацион») и B («больное сердце») может быть результатом их зависимости от третьего (скрытого) события C – условия обитания (морской климат) и родом деятельности (рыболовство); причинно-следственная связь между событиями A и B может при этом вообще отсутствовать.

Покажем это на следующем числовом примере (в котором используется гипотетические статистические данные).

Пример 3.4. Пусть вероятностная связь между событиями A («рыбный рацион»), B («больное сердце») и C («условие обитания») выражена вероятностями их конъюнкций (аналогично таблицам 3.1 – 3.3). Требуется исследовать существование вероятностной и причинно-следственной связей между событиями A и B .

Таблица 3.4

α	β	γ	$P(A\alpha \cdot B\beta \cdot C\gamma)$
0	0	0	$p_1 = 2/16$
0	0	1	$p_2 = 4/16$
0	1	0	$p_3 = 2/16$
0	1	1	$p_4 = 1/16$
1	0	0	$p_5 = 1/16$
1	0	1	$p_6 = 4/16$
1	1	0	$p_7 = 1/16$
1	1	1	$p_8 = 1/16$

Исследуем сначала вероятностную зависимость события B от A без фиксации события C (условия обитания). Из таблицы легко получить следующие значения соответствующих условных вероятностей:

$$P(B|A) = \frac{P(A \cdot B)}{P(A)} = \frac{p_7 + p_8}{p_5 + p_6 + p_7 + p_8} \cong 0.29$$

$$P(B|\bar{A}) = \frac{P(\bar{A} \cdot B)}{P(\bar{A})} = \frac{p_3 + p_4}{p_1 + p_2 + p_3 + p_4} \cong 0.33,$$

из которых можно заключить, что соблюдение рыбного рациона будто бы снижает вероятность сердечного заболевания примерно на 13% (повторим, что приведенные здесь числовые данные условны). Выясним теперь, какова вероятностная зависимость события B от A при условии осуществления события C , т.е. когда все наблюдаемые пациенты находятся в благоприятных условиях жизни; для этого найдем условные вероятности

$$P(B|A \cdot C) = \frac{P(A \cdot B \cdot C)}{P(A \cdot C)} = \frac{p_8}{p_6 + p_8} = 0.2,$$

$$P(B|\bar{A} \cdot C) = \frac{P(\bar{A} \cdot B \cdot C)}{P(\bar{A} \cdot C)} = \frac{p_4}{p_2 + p_4} = 0.2,$$

$$P(B|A \cdot \bar{C}) = \frac{P(A \cdot B \cdot \bar{C})}{P(A \cdot \bar{C})} = \frac{p_7}{p_5 + p_7} = 0.5,$$

$$P(B|\bar{A} \cdot \bar{C}) = \frac{P(\bar{A} \cdot B \cdot \bar{C})}{P(\bar{A} \cdot \bar{C})} = \frac{p_3}{p_1 + p_3} = 0.5,$$

из которых видно, что в нашем примере при фиксированных событиях C или \bar{C} вероятностная связь между событиями A и B исчезает, т.е. предположение о положительном влиянии рыбного рациона на здоровье требует, якобы, пересмотра. Такой вывод, однако, не безупречен, поскольку можно составить пример, в котором отсутствие вероятностной связи между событиями не исключает существования между ними причинно-следственных отношений (в рассмотренном приложении это может иметь место, например, если положительный эффект от рыбного рациона компенсируется сопутствующим избыточным потреблением алкоголя).

Приведенный пример говорит о том, что убедительные выводы о причинно-следственных отношениях между событиями по их стохастическим связям требуют корректной обработки статистических данных и (по возможности) постановки активного статистического эксперимента, направленных на исключение влияния «третьих» искажающих факторов.

Формулы полной вероятности и Байеса

В заключение этой главы получим еще два важных соотношения, непосредственно следующих из аксиом теории вероятности и введенных определений.

Начнем со следующего простого примера.

Пример 3.5. Допустим, что на складе хранятся идентичные изделия, изготовленные различными фирмами и имеющие, в связи с этим, различное качество, выраженное вероятностью безотказной работы в течение некоторого испытательного (общего для фирм) срока службы. Требуется найти вероятность того, что выбранное наугад изделие проработает этот срок без отказа (число фирм-изготовителей, количество лежащих на складе их изделий и указанные вероятности безотказной работы известны).

Обобщим этот пример. Пусть A – случайное событие, вероятность осуществления которого зависит от несовместимых событий из совокупности $\{B_i\}_{i=1}^n$, причем выполняются условия

$$\forall i, j: B_i \cdot B_j = \emptyset, \quad \bigcup_{i=1}^n B_i \supseteq A.$$

Эта зависимость выражается условными вероятностями $\{P(A|B_i)\}$, значения которых известны. Известны также вероятности $\{P(B_i)\}_{i=1}^n$ событий $\{B_i\}_{i=1}^n$. Требуется найти безусловную (полную) вероятность $P(A)$ события A .

Составим для этого очевидное равенство

$$A = A \cdot \bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A \cdot B_i$$

и, применяя аксиому 3 (учитывая несовместимость событий, образующих правую дизъюнкцию) и теорему умножения вероятностей, получаем

$$P(A) = \sum_{i=1}^n P(A \cdot B_i) = \sum_{i=1}^n P(A|B_i)P(B_i). \quad (3.12)$$

Полученное выражение носит название *формулы полной вероятности*.

Используя ее, читатель без труда завершит рассмотрение примера 3.3, обозначая B_i факт изготовления изделия i -ой фирмой, A – его безотказную работу в течение установленного срока. Эта формула позволит ему найти правильное решение и примера 2.3.

Заметим, что события $\{B_i\}_{i=1}^n$ не обязательно образуют полную группу, как это имеет место в рассмотренном примере; требуется лишь, чтобы они образовывали покрытие события A . Следует также заметить, что формула (3.12) справедлива при $n \leq \infty$.

Изменим задачу, решаемую в примере 3.5. Пусть теперь она состоит в вычислении вероятности того, что изделие, проработавшее без отказа установленный срок, изготовлено j -ой фирмой.

В общей формулировке задача, следовательно, состоит в вычислении вероятности $P(B_j|A)$ при известных вероятностях $\{P(B_i)\}_{i=1}^n$ и $\{P(A|B_i)\}_{i=1}^n$ (при прежних обозначениях). Решение ее получается из следующих равенств (см.(3.2))

$$P(B_j A) = P(B_j | A)P(A) = P(A | B_j)P(B_j), \quad P(B_j | A) = \frac{P(A | B_j)P(B_j)}{P(A)}$$

и, ввиду (3.12),

$$P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^n P(A | B_i)P(B_i)}. \quad (3.13)$$

Равенство (3.13), называемое *формулой Байеса*, и рассматриваемые ниже его модификации играют фундаментальную роль в *теории статистических решений* (см., например, [5]), позволяя формулировать и решать задачи выбора из альтернативных гипотез (события B_j) по наблюдению события A , выполняющего теперь роль события-признака. Вероятности $\{P(B_i)\}_{i=1}^n$ и $\{P(B_i | A)\}_{i=1}^n$ носят, соответственно, названия априорных и апостериорных вероятностей гипотез B_j .

Пример 3.6 Наблюдаемый объект находится в одном из n классов состояний $\{H_i\}_{i=1}^n$ с известными наблюдателю априорными вероятностями $\{P(H_i)\}_{i=1}^n$. Для уточнения состояния объекта используется признак – случайное событие A , связанное с состоянием объекта известными условными вероятностями $\{P(A|H_i)\}_{i=1}^n$. Построим алгоритм классификации состояния объекта по наблюдению событий A и \bar{A} , минимизирующий вероятность ошибочной классификации, и вычислим эту вероятность.

Пусть осуществилось событие A . Тогда апостериорные вероятности состояний объекта $\{P(H_i | A)\}_{i=1}^n$ вычисляются согласно (3.13) с заменой $\{B_i\}_{i=1}^n$ на $\{H_i\}_{i=1}^n$, и принимается решение о том, что он находится в состоянии H_k , для которого апостериорная вероятность максимальна:

$$H_k = \arg \max_{H_i} P(H_i | A)$$

и, потому, вероятность ошибочной классификации $P_{\text{ош}}$, равная

$$P_{\text{ош}}(A) = 1 - P(H_k | A),$$

минимальна. При \bar{A} рассуждения повторяются с заменой A на \bar{A} и k на l (в общем случае $l \neq k$). Для полной вероятности ошибочной классификации получим

$$P_{\text{ош}} = P_{\text{ош}}(A)P(A) + P_{\text{ош}}(\bar{A})P(\bar{A}) .$$

Читателю рекомендуется завершить решение примера при следующих данных: $n = 2$, $P(H_1) = 0.20$, $P(H_2) = 0.80$, $P(A|H_1) = 0.90$, $P(A|H_2) = 0.05$ (ответ: $P_{\text{ош}} = 0.06$).

Область практического приложения байесовского подхода к принятию решений весьма широка. Сюда относится медицинская диагностика, выбор альтернативных бизнес-проектов, мониторинг конфликтных ситуаций в международных отношениях и т.д.

Случайные величины

Общая теория

В рассмотренных выше вероятностных моделях результат опыта выражался в появлении или в не появлении одного или группы случайных событий. В описываемых нами реальных явлениях, однако, наряду с недетерминированными событиями обычно присутствуют также недетерминированные величины, априорная неопределенность которых состоит в множественности их возможных значений (*реализаций*). К таким величинам относится, например, рост наугад выбранного человека, стоимость акции фирмы на предстоящих торгах, возможная оценка студента на экзамене по теории вероятностей.

При определенных условиях, которые будут сформулированы ниже, априорная неопределенность величины может быть выражена вероятностями ее возможных значений (точечных или интервальных); в этом случае она именуется *случайной величиной*. Приступим к описанию и изучению свойств этого нового случайного объекта.

В главах 4 и 5 будут изучаться скалярные случайные величины, в главе 6 – векторные случайные величины.

Предположим, что задано вероятностное пространство $\langle \Omega, \mathcal{F}, P \rangle$, и обозначим через X некоторую величину, имеющую определенный содержательный смысл и принимающую в результате опыта одно из множества X возможных значений, но какое конкретно априори не известно, а зависит в каждом опыте от его исхода ω , т.е. $X = x(\omega)$.

Чтобы величина X приобрела статус *случайной* величины необходимо, как следует из сказанного выше, выполнить условие, позволяющее каждому ее точечному значению или интервалу ее значений поставить в соответствие определенную вероятность.

Сформулируем это условие применительно (пока) к скалярной вещественной величине X : $X = x(\omega) \in X \subseteq R^1$.

Пусть \mathcal{B} – борелевская σ -алгебра в R^1 . Как было установлено в главе 2, \mathcal{B} содержит все интервалы и отдельные точки из R^1 . Существование вероятностной меры события $\{x(\omega) \in B\}$ для каждого элемента B σ -алгебры \mathcal{B} влечет за собой существование вероятностной меры для каждого (включая точечные) интервала значений функции $x(\omega)$ и, обратно, существование вероятностей для всех интервалов значений $x(\omega)$ означает существование вероятностей событий $\{x(\omega) \in B\}$ для всех $B \in \mathcal{B}$ (последнее следует из свойств борелевской σ -алгебры и ввиду установленных в главе 2 правил исчисления вероятностей событий, связанных булевыми функциями).

Поэтому необходимым и достаточным условием того чтобы величина X была случайной, является существование вероятности $P(B) = P\{x(\omega) \in B\}$ для каждого события $B \in \mathcal{B}$. Выясним, когда это условие выполняется.

Обозначим парой (X, \mathcal{B}) измеримое борелевское пространство в \mathbb{R}^1 . Для каждого множества (события) $B \in \mathcal{B}$ определим его прообраз $A(B)$ в пространстве Ω равенством

$$A(B) = \{\omega : x(\omega) \in B\}; \quad (4.1)$$

ясно, что между событиями A и B выполняется условие эквивалентности, т.е. $A \leftrightarrow B$. Тогда, если $A \in \mathcal{F}$ (т.е. A – случайное событие с вероятностной мерой $P(A)$), то $P(B) = P(A)$. Следовательно, величина $X = x(\omega)$ является случайной, если (4.1) выполняется для всех $B \in \mathcal{B}$.

Подведем итог в виде определения.

Определение 4.1. Случайной величиной X , определенной на вероятностном пространстве $\langle \Omega, \mathcal{F}, P \rangle$, называется функция исходов $X = x(\omega)$, $\omega \in \Omega$, отвечающая условию

$$\forall B \in \mathcal{B}: A(B) = \{\omega : x(\omega) \in B\} \in \mathcal{F}, \quad (4.2)$$

где B – элемент борелевской σ -алгебры \mathcal{B} в пространстве (X, \mathcal{B}) значений функции $x(\omega)$, A – его прообраз в Ω .

Функция $x(\omega)$, отвечающая (4.2), носит название \mathcal{F} -измеримой. Поэтому при вероятностном пространстве $\langle \Omega, \mathcal{F}, P \rangle$ случайную величину X можно определить как \mathcal{F} -измеримую функцию исходов $X = x(\omega)$.

Важно подчеркнуть, что величина X может в одном вероятностном пространстве быть случайной величиной, а в другом – нет. Приведем пример.

Пример 4.1. Вернемся к примеру 2.4а, т.е. пусть $\mathcal{H}_1 = \langle \Omega, \mathcal{F}_1, P_1 \rangle$, где $\Omega = [0, 1)$, \mathcal{F}_1 – σ -алгебра, порожденная всеми событиями вида $A = \{\omega \in [a, b)\}$, $0 \leq a < b < 1$, $P(A) = \text{mess}(b - a)$ и $X_1 = x_1(\omega) = |\omega|$ – расстояние исхода (точки ω отрезка $[0, 1)$ от его начала, см. рис.4.1). Ясно, что здесь $x_1(\omega)$ – \mathcal{F}_1 -измеримая функция, и X_1 – случайная величина.

Перейдем теперь к вероятностному пространству примера 2.4б: $\mathcal{H}_2 = \langle \Omega, \mathcal{F}_2, P_2 \rangle$, где \mathcal{F}_2 – σ -алгебра, порожденная всеми событиями вида

$$A_k = \{\omega \in [a_k, b_k)\}, \quad a_k = \frac{2^{k-1} - 1}{2^{k-1}}, \quad b_k = \frac{2^k - 1}{2^k}, \quad k = 1, 2, \dots,$$

$$P(A_k) = \text{mess}(b_k - a_k).$$

В этом вероятностном пространстве функция $x_1(\omega)$ не измерима; так, например, прообраз $A(B)$ интервала $B = [0.25; 0.50) \in \mathcal{B}$ не принадлежит σ -алгебре \mathcal{F}_2 (см. рис.4.2).

Иначе обстоит дело с функцией $x_2 = x_2(\omega)$:

$$x_2(\omega) = \begin{cases} 0, & \text{если } \omega \in A_1, \\ 1, & \text{если } \omega \in \bar{A}_1, \end{cases}$$

которая измерима в пространствах H_1 и H_2 , а также и в таком «бедном» пространстве, как $H_3 = \langle \Omega, F_3, P_3 \rangle$, где F_3 – алгебра, порожденная событием A_1 .

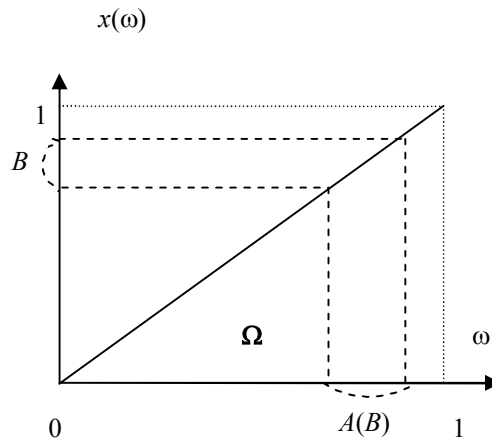


Рис. 4.2.

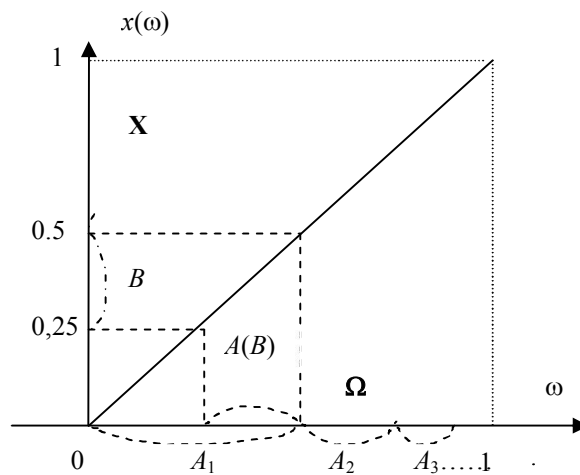


Рис. 4.3.

Из приведенного примера следует, что случайная величина задается функцией $x(\omega)$, которая в общем случае является измеримой для многих σ -алгебр различной «полноты». В связи с этим возникает задача определения минимальной σ -алгебры среди тех, для которых данная функция $x(\omega)$ измерима, т.е. величина $X = x(\omega)$ имеет статус случайной величины. Практический смысл определения и использования минимальной σ -алгебры заключается в ее экономичности, т.е. в минимизации вычислительных ресурсов, затрачиваемых для формирования статистических баз данных в стохастических моделях.

Во второй главе мы уже ввели понятие минимальной σ -алгебры, порождаемой случайными событиями. Здесь, повторяя, по существу, те же рассуждения, образуем минимальную σ -алгебру, порожденную случайной величиной.

Пусть $\{F_i\}$ – множество всех σ -алгебр, для которых функция $x(\omega)$ измерима, и F_{\min} – минимальная из них, т.е. такая, что

$$\forall i: F_{\min} \subseteq F_i \quad (4.3)$$

и для $\forall F' \subset F_{\min}$ имеем $F' \notin \{F_i\}$.

Обозначим F^* – пересечение всех σ -алгебр из $\{F_i\}$. Легко установить, что

$$F_{\min} = F^* = \bigcap_i F_i.$$

Это следует из того, что любая булева функция счетного числа элементов, входящих в пересечение σ -алгебр, принадлежит этому пересечению (в противном случае, хотя бы одно из множеств F_i не является σ -алгеброй); поэтому F^* – σ -алгебра. Далее, функция $x(\omega)$ F^* – измерима, поскольку содержит прообразы всех борелевских множеств пространства X (как и все σ -алгебры F_i). И, наконец, F^* – минимальная из σ -алгебр, для которых функция $x(\omega)$ измерима. Действительно, справедливы импликации

$$\forall i: F^* \subseteq F_i \Rightarrow F^* \subseteq F_{\min} \Rightarrow F^* = F_{\min}.$$

Поскольку каждой случайной величине X соответствует минимальная σ -алгебра, для которой функция $X = x(\omega)$ измерима, говорят, что случайная величина X порождает (в общем случае – свою) минимальную σ -алгебру $F_{\min} = F(X)$.

Полученные выводы можно сформулировать в виде теоремы.

Теорема 4.1. Пусть $H = \langle \Omega, F, P \rangle$ – вероятностное пространство и $x(\omega)$ – заданная на нем F – измеримая функция. Функция $x(\omega)$ порождает (индуцирует) в множестве исходов Ω минимальную σ -алгебру $F_{\min} = F(X)$, $F_{\min} \subseteq F$, для которой функция $x(\omega)$ измерима.

(заметим, что вероятностная мера P , определенная на H , распространяется на элементы F_{\min}).

Приведем некоторые важные свойства измеримых функций.

1. Пусть на вероятностном пространстве $\langle \Omega, F, P \rangle$ задана случайная величина $X = x(\omega)$ с значениями в X с борелевской σ -алгеброй B_X . Назовем борелевской функцию $y = y(x)$ с значениями в Y с борелевской σ -алгеброй B_Y , если для $\forall B_Y \in B_Y$ выполняется условие $B_X = \{x: y(x) \in B_Y\} \in B_X$. Для такой функции для $\forall B_Y \in B_Y$ имеем цепочку равенств

$$\{\omega: y(x(\omega)) \in B_Y\} = \{\omega: x(\omega) \in B_X\} = \{\omega \in B_X^{-1}\} \in F,$$

из которых следует, что борелевская функция от измеримой функции (иначе говоря – от случайной величины) измерима, т.е. также является случайной величиной.

Понятие и свойства борелевских функций очевидным образом распространяется и на случаи, когда пространство X векторное.

2. Для любого вероятностного пространства функция $x(\omega) \equiv c = \text{const}$ измерима (константа – частный случай случайной величины, порождающая тривиальную минимальную алгебру $F(c) = (\Omega, \emptyset)$).

3. Из измеримости функций $x(\omega)$ и $y(\omega)$ следует измеримость функций $u(\omega) = x(\omega) + y(\omega)$ и $v(\omega) = x(\omega) \cdot y(\omega)$ (если X и Y – случайные величины, то и $X + Y$ и $X \cdot Y$ – случайные величины).

4. Пусть $\{x_n(\omega)\}$ – упорядоченное по n множество измеримых функций. Тогда

а) функции $z(\omega) = \sup_n x_n(\omega)$, $w(\omega) = \inf_n x_n(\omega)$ тоже измеримы;

б) если для $\forall \omega \exists x(\omega): \lim_{n \rightarrow \infty} x_n(\omega) = x(\omega)$, то функция $x(\omega)$ измерима.

Доказательства п.п. 3 и 4 – см.[8].

Аксиоматическое определение вероятности и вытекающие из него математически строгие определения понятий случайного события и случайной величины, являясь основой для развития теории вероятностей и математической статистики, позволяют уже на этом этапе сделать ряд важных для практики методологических выводов.

Так, при построении стохастической модели функционирования системы любой сложности, ее основой служит единое вероятностное пространство, полнота которого должна быть достаточной для описания присутствующих в модели случайных объектов (случайных событий и величин) с требуемой детальностью. Это означает, что такое общее вероятностное пространство должно включать σ -алгебру, представляющую собой (как минимум) объединение минимальных σ -алгебр, порождаемых присутствующими в модели случайными событиями и величинами. Полнота такой σ -алгебры (и, следовательно, общего вероятностного пространства) определяет требуемые вычислительные ресурсы, необходимые для формирования статистической базы данных, представляющей собой, по существу, материальное воплощение вероятностного пространства.

В связи с этим заметим, что излишне высокая полнота используемой в модели общего вероятностного пространства, хотя и обеспечивает заведомо измеримость присутствующих в модели случайных объектов, приводит к завышенной емкости базы данных, т.е. к неоправданно высокому требованию к вычислительным ресурсам и к объему исходных данных.

Перейдем теперь к вопросу о способах описания вероятностных свойств случайных величин.

Определение случайной величины $X = x(\omega)$ как F –измеримой функции в вероятностном пространстве $H = \langle \Omega, F, P \rangle$ приводит, как мы видели, к тому, что каждому борелевскому множеству $B \in \mathcal{B}$ в пространстве X значений X ставится в соответствие вероятностная мера, т.е. вероятность $P\{X \in B\}$, равная вероятностной мере прообраза множества B в σ -алгебре F . В результате этого, как будет видно в дальнейшем, оказываются определенными все вероятностные свойства случайной величины X .

При этом, вместо того, чтобы для вычисления вероятностей типа $P\{X \in B\}$ каждый раз обращаться к вероятностному пространству H (которое на практике не всегда задается в явном виде), обычно удобнее (и естественнее) задавать вероятностную меру непосредственно на множествах $B \in \mathcal{B}$, т.е. в измеримом пространстве $\langle X, \mathcal{B} \rangle$, используя, таким образом, *вторичное вероятностное пространство* $\langle X, \mathcal{B}, P_X \rangle$ (здесь P_X – вероятностная мера, определенная на всех элементах борелевской σ -алгебры \mathcal{B}).

Однако, ввиду возможного многообразия всех элементов σ -алгебры \mathcal{B} , непосредственное задание их вероятностных мер часто неконструктивно. В связи с этим используем тот факт, что любой полуинтервал $[a, b) \in \mathcal{B}$ может быть выражен через полубесконечные интервалы типа $\Delta_x = (-\infty, x)$:

$$[a, b) = \Delta_b \cap \bar{\Delta}_a$$

и, следовательно, (ввиду свойств борелевской σ -алгебры) с помощью таких полуинтервалов можно представить любое событие $B \in \mathcal{B}$ и, используя аксиоматически введенные правила исчисления вероятностей событий, выразить вероятность $P(B)$ через вероятности вида $P(\Delta_x) = P\{-\infty < X < x\}$ (x – действительное число из R^1).

Заметим, что это обстоятельство позволяет в определении 4.1 условие (4.2) заменить на эквивалентное

$$\forall x \in R^1 : A(\Delta_x) = \{\omega : x(\omega) \in \Delta_x\} \in F.$$

Итак, полное описание свойств случайной величины X содержится в вероятности $P(\Delta_x)$, рассматриваемой как функция x .

Определение 4.2. *Функцией распределения скалярной случайной величины X называется функция*

$$F_X(x) = P(\Delta_x) = P\{X < x\}, x \in R^1. \quad (4.4)$$

В тех случаях, когда ясно, о какой случайной величине идет речь, ниже будет использоваться запись $F_X(x) = F(x)$.

Существование функции распределения $F(x)$ величины X определяет ее как случайную величину, полностью задавая распределение вероятностей ее значений, определяемых борелевскими множествами в пространстве X . Эквивалентность условий измеримости величины X (см.(4.2)) и существования ее функции распределения $F(x)$ позволяет определить случайную величину как недетерминированную величину, обладающую функцией распределения. Задание функции распределения $F(x)$ полностью определяет вероятностную меру P_X во вторичном вероятностном пространстве $\langle X, \mathcal{B}, P_X \rangle$.

Опишем общие свойства функции распределения.

1. Ясно, что, представляя собой вероятность, функция распределения удовлетворяет неравенствам

$$0 \leq F(x) \leq 1.$$

2. $F(x)$ – монотонно неубывающая функция. Действительно, пусть $x_1 < x_2$. Тогда $\Delta_1 = (-\infty, x_1) \subset \Delta_2 = (-\infty, x_2)$ и (поскольку соотношение между Δ_1 и Δ_2 сохраняется и для их прообразов в Ω) получаем(см. (2.7))

$$F(x_1) = P\{X \in (-\infty, x_1)\} \leq F(x_2) = P\{X \in (-\infty, x_2)\}.$$

Из равенства $(-\infty, x_2) = (-\infty, x_1) \cup [x_1, x_2)$ следует

$$P\{x_1 \leq X < x_2\} = F(x_2) - F(x_1);$$

(заметим, что если на интервале $[x_1, x_2)$ функция распределения остается постоянной, то вероятность попадания значения случайной величины X в этот интервал равна нулю).

3. Очевидны равенства

$$F(-\infty) = P\{X < -\infty\} = P(\emptyset) = 0;$$

$$F(\infty) = P\{X < \infty\} = P(\Omega) = 1.$$

Отсюда и ввиду монотонности функции распределения следуют предельные соотношения

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Покажем справедливость второго из них (первое доказывается аналогично). Свойство монотонности функции распределения позволяет описать ее поведение при $x \rightarrow \infty$, рассматривая ее значения при дискретных возрастающих значениях x :

$$x_1 < x_2 < \dots < x_n < \dots (x_n \rightarrow \infty \text{ при } n \rightarrow \infty). \quad (4.5)$$

Обозначим интервалы $\Delta_i = (-\infty, x_i)$; для них $\Delta_1 \subset \Delta_2 \subset \dots \subset \Delta_n \subset \dots$ и $\bigcup_{i=1}^{\infty} \Delta_i = (-\infty, \infty)$,

$P\left(\bigcup_{i=1}^{\infty} \Delta_i\right) = P\{X \in (-\infty, \infty)\} = 1$. Из теоремы о непрерывности вероятности (глава 2), получаем

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n \Delta_i\right) = \lim_{n \rightarrow \infty} P(\Delta_n) = \lim_{x_n \rightarrow \infty} P\{X < x_n\}, \text{ т.е. } \lim_{x \rightarrow \infty} F(x) = 1.$$

4. Из монотонности и ограниченности функции распределения следует, что она может иметь точки разрыва только первого рода, причем в этих точках функция распределения непрерывна слева. Чтобы убедиться в этом, повторим рассуждения предыдущего пункта, начиная с (4.5), при условии, что $xn \rightarrow x^*$ при $n \rightarrow \infty$, где x^* – точка разрыва. Снова используя непрерывность вероятности, легко получим

$$\lim_{x \uparrow x^*} F(x) = F(x^*).$$

Из равенства $P\{x^* \leq X < x^* + \varepsilon\} = F(x^* + \varepsilon) - F(x^*)$, $\varepsilon > 0$, при $\varepsilon \rightarrow 0$ получаем для точки разрыва x^* функции распределения равенство

$$P\{X = x^*\} = \Delta(x^*), \quad (4.6)$$

где $\Delta(x^*)$ – скачок функции распределения $F(x)$ в точке x^* .

Заметим, что для каждой точки непрерывности x функции распределения $F(x)$ вероятность $P\{X = x\}$ оказывается равной нулю. Поэтому, если функция $F(x)$ непрерывна всюду, то каждое значение случайной величины X имеет нулевую вероятность реализации, хотя в результате опыта какое либо его значение из X непременно реализуется. Это означает, что все события $\{X = x\}$, не являясь невозможными, являются нулевыми.

5. Можно показать, что число разрывов функции распределения не более чем счетно. Для этого обозначим L_n множество разрывов функции $F(x)$, размер которых превышает величину $\frac{1}{n}$; так как число таких разрывов не превышает n (ввиду установленных выше свойств функции распределения), то для каждого фиксированного n множество L_n конечно. Множество всех разрывов функции $F(x)$ равно

$$L = \bigcup_{n=1}^{\infty} L_n$$

т.е. счетно, так как представляет собой объединение счетного числа конечных множеств.

Ясно, конечно, что функция распределения, вполне определяя вероятностные свойства случайной величины, не задает саму величину, как измеримую функцию $x(\omega)$, определенную на множестве исходов Ω . Так, разные случайные величины X и Y :

$$X = \begin{cases} 1 \text{ с вероятностью } 0,5; \\ -1 \text{ с вероятностью } 0,5, \end{cases} \quad Y = -X,$$

имеют одну и ту же функцию распределения $F_X(x) \equiv F_Y(x)$.

Хотя функция распределения представляет собой универсальную форму описания распределения вероятностей случайной величины X , на практике обычно удобнее

использовать иные (эквивалентные) формы такого описания, выбор которых зависит от типа множества X возможных ее значений.

Рассмотрим возможные классы распределений случайных величин, введя следующие определения.

Определение 4.3. Если все значения случайной величины X образуют конечное или счетное изолированное множество: $X = (x_1, \dots, x_n)$ или $X = (x_1, \dots, x_n, \dots)$, то распределение X называют дискретным.

Напомним, что множество точек числовой прямой называют *изолированным* (или, иначе, *всюду неплотным*), когда для каждой из них существует ненулевая окрестность, не содержащая других точек из X (к таким множествам не относятся, например, счетное множество рациональных чисел).

Определение 4.4. Если множество X значений случайной величины X континуально и ее функция распределения $F(x)$ абсолютно непрерывна, т.е. существует неотрицательная функция $f(z)$ такая, что

$$\forall x: F(x) = \int_{-\infty}^x f(z) dz, \quad (4.7)$$

то распределение случайной величины X называется абсолютно непрерывным или, часто, просто непрерывным.

Определение 4.5. Распределение, не отвечающее ни определению 4.3, ни определению 4.4. будем называть сингулярным.

Особенностей сингулярного распределения мы коснемся ниже.

Рассмотрим важнейшие конкретные типы дискретных и непрерывных распределений случайной величины.

Дискретные распределения

В общем случае дискретное распределение скалярной случайной величины X с конечным или счетным множеством всех ее возможных значений $X = (x_1, \dots, x_n)$ ($n \leq \infty$) задается вероятностной мерой P_X , определенной на X конечным или счетным набором вероятностей событий $\{X = x_i\}$, $i = \overline{1, n}$, т.е.

$$P_X = \{P\{X = x_i\}\}_{i=1}^n \stackrel{\Delta}{=} \{p(x_i)\}_{i=1}^n; \quad (4.8)$$

для конкретных распределений (4.8) задается формулой или таблицей (при $n < \infty$).

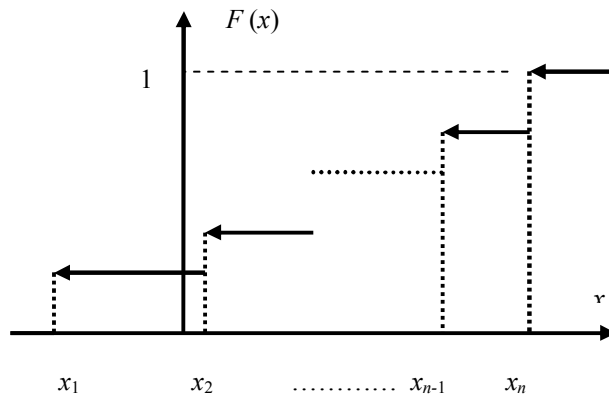


Рис. 4.3.

Поскольку события $\{X = x_i\}_{i=1}^n$ образуют полную группу несовместимых событий, имеет место равенство

$$\sum_{i=1}^n P\{X = x_i\} = \sum_{i=1}^n p(x_i) = 1.$$

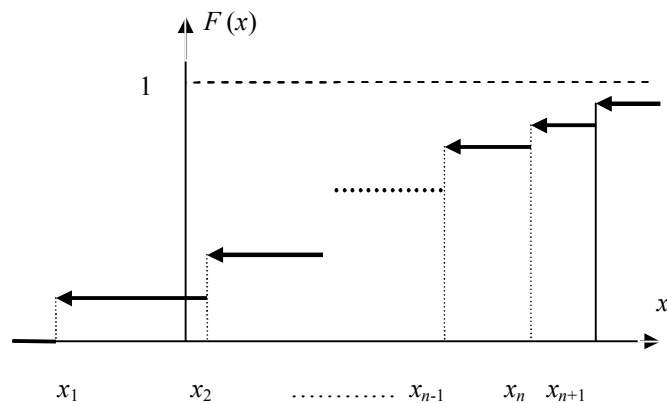


Рис. 4.4.

Нетрудно установить, что при дискретном распределении случайной величины X ее функция распределения $F(x)$ имеет ступенчатый вид (см. рис.4.3 для $n < \infty$ и рис.4.4 для $n = \infty$) с разрывами первого рода в точках x_i , ($i = 1, 2, \dots$); величина разрыва для каждого i равна $P\{X = x_i\} = p(x_i)$ (в справедливости всего сказанного читатель может убедиться самостоятельно, используя рассмотренные выше общие свойства функции распределения).

4.2.1. Распределение Бернулли

Рассмотрение конкретных типов распределения случайных величин мы начнем с самого простого и, одновременно, основополагающего – с *распределения Бернулли*.

Определение 4.6. Случайная величина X , принимающая значения 0 или 1 ($X = \{0, 1\}$) имеет распределение Бернулли с параметром p (обозначение $X \in \Pi(p)$), если выполняется условие

$$X = \begin{cases} 0 & \text{с вероятностью } 1-p, \\ 1 & \text{с вероятностью } p, \end{cases}$$

или, компактнее,

$$P\{X = \alpha\} = p^\alpha (1-p)^{1-\alpha}, \quad \alpha = 0, 1. \quad (4.9)$$

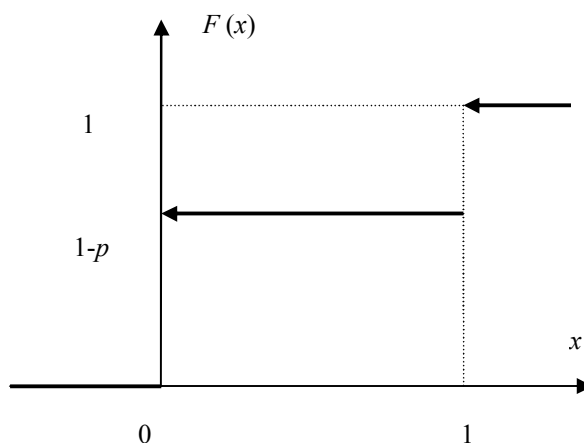


Рис. 4.5.

На рис.4.5 приведен вид функции распределения Бернулли.

Распределение Бернулли непосредственно связано с исчислением вероятностей случайных событий. Пусть A – случайное событие, т.е. событие, измеримое в некотором вероятностном пространстве $\langle \Omega, \mathcal{F}, P \rangle$, и $P(A) = p$. Тогда индикатор Z_A события A :

$$Z_A = \begin{cases} 0 & \text{при } \bar{A}, \\ 1 & \text{при } A \end{cases}$$

имеет распределение Бернулли (4.9). Этим объясняется широкое применение этого типа распределения в практических расчетах.

4.2.2. Биномиальное распределение

Рассмотрим серию n опытов, в каждом i -ом из которых с вероятностью p осуществляется и с вероятностью $1-p$ не осуществляется случайное событие A_i , причем события $\{A_i\} = (A_1, \dots, A_n)$ независимы в совокупности. Нередко такую серию опытов называют *схемой Бернулли*, а каждое появление события A_i – *успехом* в i -ом опыте. Обозначим X – число реализаций событий из $\{A_i\}$ в серии n опытов (т.е. число успехов в

схеме Бернулли). Во многих прикладных моделях, в которых величина X имеет конкретный смысл (число удачных коммерческих сделок, число попаданий в цель при бомбометании, число отказавших изделий при проверке их качества, число заболеваний при медицинском мониторинге и т.д.) возникает необходимость в определении распределения величины X .

При решении этой задачи убедимся, прежде всего, что X – случайная величина, т.е. $X = x(\omega)$ – измеримая функция на множестве исходов Ω_X вероятностного пространства $H_X = \langle \Omega_X, \mathcal{F}_X, P_X \rangle$. Построим это вероятностное пространство. Обозначим $Z_{A_i} = Z_i$ индикатор события A_i , который является случайной величиной ($P\{Z_i=1\} = P(A_i)$), порождающей вероятностное пространство $H_i = \langle \Omega_i, \mathcal{F}_i, P_i \rangle$, где

$$\Omega_i = (\omega_i^{(0)}, \omega_i^{(1)}), \quad z_i(\omega_i^{(0)}) = 0, \quad z_i(\omega_i^{(1)}) = 1, \\ \mathcal{F}_i = (\omega_i^{(0)}, \omega_i^{(1)}, \Omega_i, \emptyset), \quad P(\omega_i^{(0)}) = 1 - p, \quad P(\omega_i^{(1)}) = p.$$

Образуем теперь вероятностное пространство, на котором определена случайная величина X . Оно представляет собой *произведение вероятностных пространств* $\{H_i\}$:

$$H = H_1 \times \dots \times H_n$$

с множеством исходов Ω , содержащим 2^n элементов вида

$$\omega_k = (\omega_1^{(\alpha_1)} \cdot \omega_2^{(\alpha_2)} \dots \omega_n^{(\alpha_n)}), \quad k = \overline{1, 2^n},$$

где α_i принимают значения 0 или 1. Каждому исходу ω_k соответствует вероятность, равная

$$P(\omega_k) = p^{\sum_{i=1}^n \alpha_i} (1-p)^{n-\sum_{i=1}^n \alpha_i}, \quad k = \overline{1, 2^n}.$$

Объединим исходы ω_k , при которых реализуется событие $\{X = x\}$, т.е. для каждого из которых выполняются равенства $\sum_{i=1}^n \alpha_i = x$ и, следовательно, $P(\omega_k) = p^x (1-p)^{n-x}$; число таких исходов равно C_n^x . Проделав это для всех $x = \overline{0, n}$, получим вероятностное пространство с множеством исходов $\Omega_X = (\omega(0), \omega(1), \dots, \omega(n))$, где

$$\omega(x) = \bigcup_{\omega_r: \sum_{i=1}^n \alpha_i = x} \omega_r = \{X = x\}, \quad P(\omega(x)) = P\{X = x\} = \\ = C_n^x p^x (1-p)^{n-x}$$

Итак, получено вероятностное пространство $H_X = \langle \Omega_X, \mathcal{F}_X, P_X \rangle$, которое определяет X как случайную величину. Одновременно найдено распределение этой случайной величины, отвечающее следующему определению.

Определение 4.7. Случайная величина X , принимающая целые значения от 0 до n ($X = \{0, 1, \dots, n\}$) имеет биномиальное распределение с параметрами n и p (обозначение $X \in \text{Вб}(n, p)$), если выполняются равенства

$$P\{X = x\} = C_n^x p^x (1-p)^{n-x}, x = 0, 1, \dots, n. \quad (4.10)$$

При $n = 1$ биномиальное распределение совпадает с распределением Бернулли.

4.2.3. Распределение Пуассона

Прежде чем перейти к следующему типу дискретного распределения случайной величины, остановимся на простой модели применения биномиального распределения.

Пусть коммерческая фирма в течение некоторого заданного интервала времени T заключает сделки с фиксированным числом n клиентов. Каждая сделка с вероятностью p оказывается для фирмы успешной. Предполагая, что результаты сделок независимы в совокупности, находим, что число успешных сделок K_n за время T имеет биномиальное распределение с параметрами n и p : $K_n \in \text{Bi}(n, p)$. Допустим теперь, что фирма увеличивает число заключаемых сделок, обеспечивая постоянство вероятности p . Из интуитивных соображений следует, что в этом случае число успешных сделок также должно иметь тенденцию возрастать. Действительно, можно показать, что при этом для $\forall \bar{k}$ выполняется соотношение

$$\lim_{n \rightarrow \infty} P\{K_n > \bar{k}\} = 1.$$

Как в этих условиях описать предельное поведение случайной величины K_n мы узнаем в главе 7.

Иначе обстоит дело, если с ростом числа обслуживаемых клиентов n вероятность p успешности отдельной сделки падает (что на практике часто имеет место).

Рассмотрим случай, когда при неограниченном возрастании n выполняется равенство

$$\lim_{n \rightarrow \infty} np = a, \quad a = \text{const} < \infty, \quad (np = a + o(\frac{1}{n}))$$

и найдем предельное распределение величины K при $n \rightarrow \infty$.

Отправляясь от (4.10), имеем

$$\begin{aligned} P\{K = k\} &= \lim_{n \rightarrow \infty} C_n^k p^k (1-p)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} n^k \left(1 - \frac{a + o(\frac{1}{n})}{n}\right)^{n-k} = \frac{a^k e^{-a}}{k!}, \quad k = 0, 1, \dots \end{aligned}$$

Полученный результат представляет собой содержание **теоремы Пуассона**.

Итак, мы получили новое распределение, отвечающее определению:

Определение 4.8. Случайная величина X , принимающая целые значения от 0 до ∞ ($X \in \{0, 1, \dots\}$) имеет распределение Пуассона с параметром a (обозначение $X \in \text{Po}(a)$), если выполняются равенства

$$P\{X = x\} = \frac{a^x e^{-a}}{x!}, x = 0, 1, \dots \quad (4.11)$$

Распределение Пуассона имеет широкое применение в моделях, описывающих потоки случайных событий, реализующихся в случайные моменты времени (потоки отказов аппаратуры, потоки коммерческих сделок, потоки заболеваний при мониторинге эпидемий и пр.). Покажем, как возникает это распределение при моделировании потока случайных событий (*точечного случайного процесса*).

Пусть на оси времени $T = [0, \infty)$ в случайные моменты происходят случайные однородные события; их однородность означает, что они равноценны для наблюдателя, интересующегося лишь их числом. Назовем эти события элементарными, а их последовательность – потоком элементарных событий, и обозначим A_k событие, состоящее в том, что в интервале времени $\theta = [t, t + \tau)$ произошло ровно k элементарных событий: $A_k = \{K = k\}$, $k = 0, 1, 2, \dots, n, \dots$. Предположим, что поток элементарных событий удовлетворяет следующим трем требованиям.

1. *Стационарность*: вероятность $P_{t, t+\tau}(k) = P(A_k)$ появления ровно k событий в интервале времени θ зависит от τ , но не зависит от t , т.е. от его положения на оси времени:

$$\forall \tau \in [0, \infty) \quad P_{t, t+\tau}(k) = P_{\tau}(k), \quad k = 0, 1, 2, \dots$$

2. *Отсутствие последовательности*: для любого конечного набора непересекающихся интервалов времени

$$\theta_1 = [t_1, t_1 + \tau_1), \dots, \theta_n = [t_n, t_n + \tau_n)$$

события $(A_{k_1}, \dots, A_{k_n})$ независимы в совокупности.

3. *Ординарность*: вероятность появления более одного события в интервале времени $\theta = [t, t + \Delta t)$ — величина, малая относительно Δt , т.е.

$$\frac{P_{\Delta t}\{K > 1\}}{\Delta t} = \frac{\sum_{i=2}^{\infty} P(A_i)}{\Delta t} \rightarrow 0 \quad \text{при } \Delta t \rightarrow 0.$$

Оказывается, что при выполнении этих условий число K элементарных событий, происходящих в интервале времени $\theta = [t, t + \tau)$, имеет распределение, определяемое равенством

$$P\{K = k\} = P_{t, t+\tau}(k) = \frac{(\lambda \tau)^k e^{-\lambda \tau}}{k!}, k = 0, 1, \dots, \quad (4.12)$$

т.е. распределение Пуассона с параметром $a = \lambda \tau$ (содержательный смысл этого параметра будет раскрыт ниже).

Доказательство приведенного результата содержится, например, в [6]. Там же приведены важные для практики свойства рассмотренного потока событий, который носит название *простого пуассоновского случайного процесса* или *простейшего потока событий*.

Непрерывные распределения

Вернемся к определению 4.4. Функция $f(z)$ в (4.7) носит название *функции плотности распределения* (или просто *плотности распределения*) случайной величины X . Это название раскрывает ее смысл, как предела

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P\{x \leq X < x + \Delta x\}}{\Delta x} = \frac{dF(x)}{dx}.$$

(при условии ее непрерывности, которое обычно выполняется).

Иногда, когда это нужно, мы будем снабжать функцию плотности распределения нижним индексом, обозначающим случайную величину, к которой она относится, т.е. писать $f(x) = f_X(x)$.

Функция $f(x)$ имеет следующие очевидные свойства (вытекающие из свойств функции распределения $F(x)$).

$$1. f(x) \geq 0;$$

$$2. \int_{-\infty}^{\infty} f(x) dx = 1;$$

$$3. P\{a \leq X < b\} = F(b) - F(a) = \int_a^b f(x) dx.$$

Бесконечно малую величину $P\{x \leq X < x + dx\} = f(x)dx$ иногда называют *элементом вероятности*.

Каждая интегрируемая функция $f(x)$, удовлетворяющая условиям 1 и 2, может претендовать на роль плотности распределения некоторой случайной величины.

Переходя к конкретным типам непрерывного распределения случайных величин, заметим также, что в практических расчетах преимущественно используются функции плотности распределения, которые по полноте характеристики вероятностных свойств случайных величин эквивалентны функциям распределения, обладая при этом большей наглядностью.

4.3.1. Равномерное распределение

Самый простой, но, как мы убедимся ниже, весьма важный для практики тип непрерывного распределения случайной величины задается следующим определением.

Определение 4.9. *Случайная величина X , принимающая непрерывные значения на отрезке $X = [a, b]$ ($b > a$) имеет равномерное распределение с параметрами a и b , если ее плотность распределения равна*

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in X, \\ 0, & \text{если } x \notin X, \end{cases} \quad (4.13)$$

т.е. постоянна и отлична от нуля только на отрезке X (примем для этого распределения обозначение $X \in \text{Eq}(a, b)$).

Из (4.7) и (4.13) следует, что функция распределения случайной величины при равномерном распределении равна

$$F(x) = \begin{cases} 0, & \text{если } x < a, \\ \frac{x-a}{b-a}, & \text{если } a \leq x \leq b, \\ 1, & \text{если } x > b. \end{cases}$$

На рис. 4.6 приведен пример графиков плотности распределения $f(x)$ (сплошная кривая) и функции распределения $F(x)$ (пунктирная кривая) для $X \in \text{Eq}(a, b)$.

Хотя равномерное распределение для непосредственного моделирования поведения случайных величин используется сравнительно редко, оно играет главную роль в процедурах генерирования случайных событий и значений случайных величин с заданными распределениями (практически любого типа). Такая возможность, широко используемая в стохастическом компьютерном моделировании случайных явлений и процессов, основана на следующих фактах.

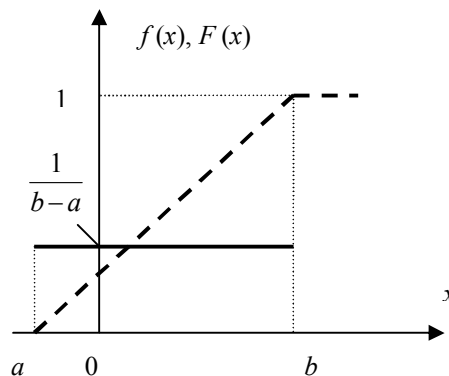


Рис. 4.6.

Программное обеспечение современных компьютеров содержит операцию генерирования случайных (точнее – «псевдослучайных») чисел, т.е. реализаций y случайной величины Y с равномерным распределением на единичном отрезке $[0, 1]$: $Y \in \text{Eq}(0, 1)$. Это

сразу позволяет моделировать реализацию случайного события с вероятностью p : событие каждый раз считается состоявшимся, если выполняется условие $y \leq p$.

Генерирование реализаций случайной величины с заданным дискретным распределением также несложно. Пусть случайная величина X имеет дискретное распределение, заданное вероятностями $\{p_i = P\{X = x_i\}\}_{i=1}^n$ (n – число значений X). Отрезок $[0,1]$ разбивается на n непересекающихся промежутков $\{\Delta_i\}_{i=1}^n$ длиной p_i каждый ($\sum_{i=1}^n p_i = 1$) и генерирование реализации случайной величины X сводится к выполнению импликации: $y \in \Delta_i \Rightarrow X = x_i$.

Одним из методов генерирования реализаций случайной величины с непрерывной функцией распределения основан на следующем интересном свойстве таких функций распределения.

Пусть случайная величина X имеет непрерывную функцию распределения $F_X(x)$. Используя ее как функцию преобразования случайной величины X в новую величину Y

$$Y = F_X(X),$$

получаем равенства

$$F_Y(y) = P\{Y < y\} = P\{F_X(X) < y\} = \begin{cases} 0, & \text{если } y < 0, \\ P\{X < F_X^{-1}(y)\} = F_X(F_X^{-1}(y)) = y, & \text{если } 0 \leq y \leq 1, \\ 1, & \text{если } y > 1, \end{cases}$$

из которых следует, что при любой (непрерывной) функции распределения $F_X(x)$ Y представляет собой случайную величину с равномерным распределением на единичном отрезке $[0,1]$: $Y \in \text{Eq}(0,1)$.

Итак, между реализациями случайных величин X и Y существует зависимость $y = F_X(x)$ или $x = F_X^{-1}(y)$, где y – реализация равномерно распределенной случайной величины Y , поставляемая генератором случайных чисел, а $F_X^{-1}(\cdot)$ функция, обратная функции распределения $F_X(\cdot)$. Последнее равенство позволяет получать реализации случайной величины X с известной непрерывной функцией распределения, т.е. решать поставленную задачу и в этом случае.

Соответствующие программные мероприятия позволяют генерировать реализации случайных величин, обладающих кусочно непрерывными и смешанными (см. (§4.5)) распределениями.

4.3.2. Гамма-распределение

Определим формально следующее непрерывное распределение, являющееся «прародителем» ряда распределений, имеющих большое практическое значение для построения вероятностных моделей и решения задач математической статистики.

Определение 4.10. *Случайная величина X имеет гамма-распределение с параметрами α и λ ($X \in (\alpha, \lambda)$), если ее плотность распределения равна*

$$f(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & \text{если } x > 0, \end{cases} \quad (4.14)$$

где $\alpha > 0$, $\lambda > 0$, $\Gamma(\alpha)$ – гамма-функция:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

(напомним: $\Gamma(\alpha) = \alpha\Gamma(\alpha-1)$, $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$; если $\alpha = n$ – целое, то

$$\Gamma(n+1) = n\Gamma(n) = n!).$$

Корректность определения 4.10 обеспечена тем, что функция (4.14) отвечает всем свойствам плотности распределения.

4.3.3. Показательное распределение

В п.4.2.3 был упомянут простейший поток событий, т.е. случайный процесс, в котором число событий, происходящих за интервал времени $[t, t+\tau)$, имеет распределение Пуассона с параметром $a = \lambda\tau$ (см.(4.12)). Значение параметра a вполне задает свойства простейшего потока. Оказывается, что его свойства столь же полно могут быть определены распределением другой случайной величины – значением интервала T между смежными (происходящим друг за другом) событиями. Найдем это распределение.

Пусть в момент t^* произошло очередное элементарное событие в череде событий, образующих простейший поток; следующее элементарное событие произойдет в момент $t^* + T$, где T – интересующая нас случайная величина. Для ее функции распределения справедливо равенство

$$F_T(t) = P\{T < t\} = 1 - P_{t^*, t^*+t}(0) = \begin{cases} 1 - e^{-\lambda t}, & \text{если } t \geq 0, \\ 0, & \text{если } t < 0 \end{cases} \quad (4.15)$$

(справа стоит вероятность того, что в интервале $[t^*, t^* + t)$ произойдет хотя бы одно элементарное событие, см.(4.12)).

Для плотности распределения случайной величины T из (4.15) получим

$$f_T(t) = \begin{cases} 0, & \text{если } t \leq 0, \\ \lambda e^{-\lambda t}, & \text{если } t > 0. \end{cases} \quad (4.16)$$

В результате получено непрерывное распределение, отвечающее следующему определению.

Определение 4.11. *Случайная величина X имеет показательное (экспоненциальное) распределение с параметром λ , $\lambda > 0$, если ее функция и плотность распределения определены (4.15) и (4.16) (с заменой T на X).*

Примем для этого распределения обозначение $X \in \text{Exp}(\lambda)$.

Из сравнения (4.14) и (4.16) следует, что показательное распределение есть частный случай гамма-распределения при $\alpha = 1$.

4.3.4. Распределение Эрланга

При построении вероятностных моделей, содержащих описание потоков событий (потоков заявок на обслуживание в системах массовой обработки, потоков сделок в бизнесе и т.д.) часто появляется необходимость в распределении случайного интервала времени, в течение которого произойдет ровно n событий. При определенных условиях (о которых речь пойдет ниже) такой интервал времени имеет распределение, представляющее собой частный случай гамма-распределения при целом $\alpha = n$ и называемое распределением Эрланга.

Определение 4.12. *Случайная величина X имеет распределение Эрланга, $X \in \text{Erl}(n, \lambda)$, с параметрами n (положительное целое) и $\lambda > 0$, если плотность распределения X равна*

$$f(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, & \text{если } x > 0. \end{cases} \quad (4.17)$$

4.3.5. Распределение χ^2

Этот тип распределения широко используется в операциях математической статистики, являясь частным случаем гамма-распределения при значениях параметров $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$.

Определение 4.13. *Случайная величина X имеет распределение χ^2 с параметром n (обозначение $X \in \chi^2(n)$), если плотность распределения X равна*

$$f(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}, & \text{если } x > 0. \end{cases} \quad (4.18)$$

(n – положительное целое).

Иногда обозначение χ^2 приписывают самой величине X . Функция распределения для $X = \chi^2$ равна

$$F(x) = F_{\chi^2}(x; n) = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{1}{2^{\frac{n}{2}} \Gamma(n/2)} \int_0^x z^{\frac{n}{2}-1} e^{-\frac{z}{2}} dz, & \text{если } x > 0. \end{cases}$$

Ввиду того, что распределение χ^2 является одним из основных распределений, используемых в математической статистике, оно широко представлено в табличном виде в литературе и программно в современных компьютерах. При этом часто табулируется не функция распределения $F_{\chi^2}(x)$, а так называемый интеграл вероятностей χ^2 :

$$I_{\chi^2}(x; n) = 1 - F_{\chi^2}(x; n).$$

Для практических расчетов полезна связь распределения χ^2 с распределением Пуассона. Действительно, интегрируя (4.18) последовательно по частям в пределах от 0 до $2a$ при $n = 2k$, получим равенство

$$I_{\chi^2}(2a, 2k) = 1 - F_{\chi^2}(2a; 2k) = \sum_{r=0}^{k-1} \frac{a^r}{r!} e^{-a},$$

где сумма справа представляет собой вероятность $P\{K < k\}$, когда $K \in \text{Po}(a)$ (см. (4.11)), т.е. значение функции распределения Пуассона с параметром a . Таким образом, в результате табулирования интеграла вероятностей χ^2 «обслуживается» как распределение χ^2 , так и распределение Пуассона.

4.3.6. Нормальное (гауссовское) распределение

Завершим рассмотрение наиболее часто применяемых типов распределений скалярной случайной величины распределением, занимающим, как выяснится в дальнейшем, центральное положение в теории вероятностей.

Определение 4.14. Случайная величина X имеет нормальное (гауссовское) распределение с параметрами m и σ (обозначение $X \sim N(m, \sigma^2)$), если плотность распределения X равна

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (4.19)$$

где $-\infty < m < \infty$, $\sigma > 0$.

Корректность этого определения (в смысле соответствия (4.19) свойствам плотности распределения) основана на равенстве

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sigma^2}} dz = \sqrt{2\pi}\sigma,$$

где слева стоит известный интеграл Пуассона.

Для функции нормального распределения получим

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(z-m)^2}{2\sigma^2}} dz \quad (4.20)$$

На рис.4.7 приведены графики функций $f(x)$ и $F(x)$ при $X \in N(1, 0.25)$ (сплошная и пунктирная кривые соответственно).

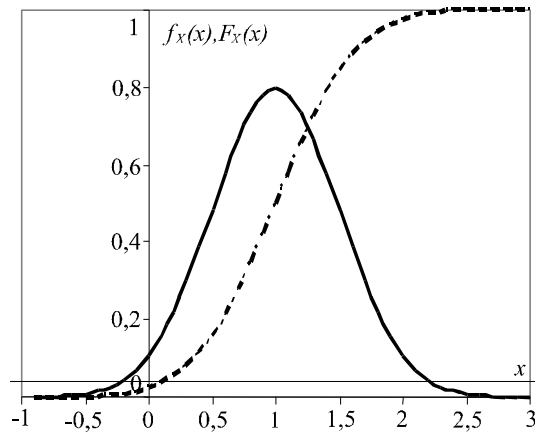


Рис. 4.7.

Нормальное распределение с параметрами $m = 0$, $\sigma^2 = 1$ называется *стандартным нормальным распределением* с функцией распределения, обозначаемой $\Phi^*(x)$:

$$\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz, \quad (4.21)$$

график которой приведен на рис. 4.8. Отметим справедливость равенства

$$\Phi^*(-x) = 1 - \Phi^*(x). \quad (4.22)$$

Как видно из (4.20) и (4.21), при нормальном распределении случайной величины X ее функция распределения представляет собой интеграл, не представимый элементарными функциями. Это неудобство преодолевается табулированием функции $\Phi^*(x)$ и возможностью, как будет показано в следующем параграфе, легко сводить произвольное нормальное распределение к стандартному.

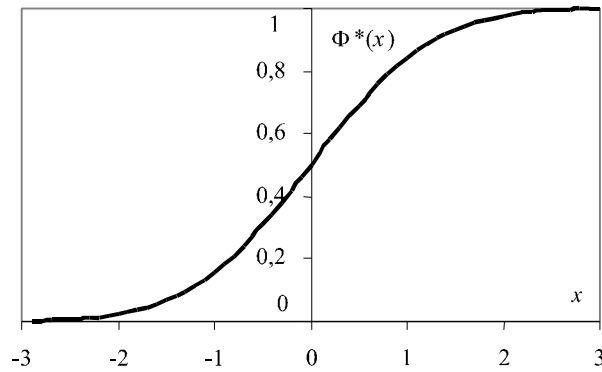


Рис. 4.8.

На этом мы завершим краткое рассмотрение наиболее представительных типов распределений скалярных случайных величин. В последующем изложении список типов распределений будет расширяться, а описание свойств уже рассмотренных распределений пополняться.

Функции случайных величин

В §4.1, говоря о свойствах измеримых функций, мы установили, что борелевские функции от измеримых функций также являются измеримыми, т.е. представляют собой случайные величины. Вместе с тем, борелевскими является большинство функций, встречающихся в практических расчетах (в частности – непрерывные функции и функции с конечным числом разрывов).

Приведем несколько примеров получения распределений простых борелевских функций $Y = \psi(X)$ при известном распределении аргумента – случайной величины X .

Пусть $Y = aX + b$, где $X \in F_X(x)$. Тогда

$$F_Y(y) = P\{aX + b < y\} = \begin{cases} F_X\left(\frac{y-b}{a}\right), & \text{если } a > 0, \\ P\left\{X \geq \frac{y-b}{a}\right\} = 1 - F_X\left(\frac{y-b}{a}\right), & \text{если } a < 0. \end{cases} \quad (4.23)$$

Обратимся к более общему случаю, когда функция $y = \psi(x)$ является монотонной (например – неубывающей) при условии существования обратной функции $x = \psi^{-1}(y)$. Тогда

$$F_Y(y) = P\{\psi(X) < y\} = P\{X < \psi^{-1}(y)\} = F_X(\psi^{-1}(y)) \text{ и}$$

если X имеет непрерывное распределение с плотностью $f_X(x)$, а функция $x = \psi^{-1}(y)$ дифференцируема, то для плотности распределения случайной величины Y получим

$$f_Y(y) = f_X(\psi^{-1}(y))(\psi^{-1}(y))'.$$

В том случае, если функция $y = \psi(x)$ является убывающей, аналогично приходим к равенству

$$f_Y(y) = f_X(\psi^{-1}(y)) \left| (\psi^{-1}(y))' \right|, \quad (4.24)$$

которое является общим для случаев монотонной функции $\psi(x)$.

Используем (4.24) для получения одного важного вывода о свойстве нормального распределения. Пусть $X \in N(m, \sigma^2)$ и $Y = aX + b$. Тогда из (4.22), (4.23) и (4.24) следует

$$f_Y(y) = \frac{1}{\sqrt{2\pi} |a| \sigma} e^{-\frac{(x-b-am)^2}{2a\sigma^2}},$$

т.е. $Y \in N(b+am, a^2\sigma^2)$. Это означает, что при линейном преобразовании нормально распределенной случайной величины тип ее распределения остается нормальным. Такая устойчивость нормального распределения будет ниже рассмотрена в более общем аспекте.

Если $X \in N(m, \sigma^2)$, то $Z = \frac{X-m}{\sigma} \in \square(0,1)$ и получаем формулу

$$F_X(x) = P\{X < x\} = P\{\sigma Z + m < x\} = P\{Z < \frac{x-m}{\sigma}\} = \Phi\left(\frac{x-m}{\sigma}\right),$$

которая позволяет находить значения функции нормального распределения при любых значениях параметров m и σ^2 по таблице значений функции стандартного нормального распределения $\Phi^*(z)$ (4.21).

Остановимся на частном случае квадратично зависимости $Y = X^2$, где X имеет стандартное нормальное распределение. Находим

$$F_Y(y) = P\{X^2 < y\} = P\{-\sqrt{y} \leq X < \sqrt{y}\} = \Phi^*(\sqrt{y}) - \Phi^*(-\sqrt{y}) = 2\Phi^*(\sqrt{y}) - 1$$

и для плотности распределения Y получим

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}. \quad (4.26)$$

Сравнивая (4.26) с (4.18) убеждаемся, что Y имеет распределение χ^2 при $n = 1$.

Условное распределение случайной величины

Предположим, что в нашей модели на вероятностном пространстве $H = \langle \Omega, F, P \rangle$ определены случайная величина $X = x(\omega)$ и случайное событие B с известными функцией распределения $F(x)$ и вероятностью $P(B)$. Очень часто возникает необходимость вычислить распределение случайной величины X при условии, что произошло событие B , или, точнее говоря, найти распределение той случайной величины, которая получается из X при осуществлении события B . Назовем это распределение *условным* (относительно события B) в

отличие от *безусловного (полного)* распределения с функцией $F(x)$, которое имеет место, когда никаких условий на осуществление события B не накладывается.

Приведем простые примеры.

Пример 4.2. Пусть Ω – множество взрослых людей разного пола, X – рост наугад выбранного человека, B – событие, состоящее в принадлежности выбранного человека к мужскому полу. Ясно, что распределение роста, когда выбор осуществляется вне всякой зависимости от пола (безусловное распределение), отличается от искомого условного распределения этой величины при отборе только мужчин.

Пример 4.3. Спрос на услуги, осуществляемые коммерческой фирмой, представляет собой поток событий, число которых в заданном интервале времени – случайная величина X с известным распределением (например, Пуассона). Следует найти распределение числа реальных сделок, которые заключит фирма при условии, что число принимаемых ею заказов лимитировано и не превышает числа x_m , т.е. $B = \{X \leq x_m\}$.

В общей постановке задача состоит, следовательно, в нахождении условного распределения, т.е. условной функции распределения, которую случайная величина X приобретает при осуществлении события B . Обозначим эту функцию $F(x|B)$. Полагая, что событие B ненулевое (т.е. имеет ненулевую вероятность), и обозначая $A = \{X < x\}$, ввиду (3.1) определим функцию $F(x|B)$ равенством

$$F(x|B) = P\{X < x|B\} = P(A|B) = \frac{P(A \cdot B)}{P(B)}, \quad (4.27)$$

т.е. как условную вероятность, отвечающую, как было замечено, аксиоматике теории вероятностей. Отсюда можно заключить, что условная функция распределения обладает теми же свойствами, что и безусловная функция распределения.

Если случайная величина X имеет дискретное распределение, то обычно в расчетах оперируют не с условной функцией распределения, а с условными вероятностями $p(x_i|B)$, определяемыми равенствами

$$p(x_i|B) = \frac{P(\{X = x_i\} \cap B)}{P(B)}, \quad i = 1, 2, \dots \quad (4.28)$$

Случайная величина X имеет абсолютно непрерывное условное распределение, если существует неотрицательная *условная плотность распределения* $f(x|B)$, удовлетворяющая равенству

$$\forall x: F(x|B) = \int_{-\infty}^x f(z|B) dz, \quad (4.29)$$

и обладающая, при этом, теми же свойствами, что и безусловная плотность распределения.

Нередко возникает задача вычисления безусловной функции распределения при известных условных функциях распределения. Ее решение сводится к применению формулы полной вероятности (3.12). Положим в (3.12) $A = \{X < x\}$ и пусть $\{B_j\}_{j=1}^n$ – полная конечная или счетная группа несовместимых событий ($n \leq \infty$). Тогда

$$F(x) = P(A) = \sum_{j=1}^n P(A | B_j) P(B_j) = \sum_{j=1}^n F(x | B_j) P(B_j) \quad (4.30)$$

При дискретном распределении случайной величины X положим в (3.12) $A = A_i = \{X = x_i\}$ и получим

$$P(A_i) = P\{X = x_i\} = p(x_i) = \sum_{j=1}^n p(x_i | B_j) P(B_j), \quad i = 1, 2, \dots \quad (4.31)$$

Если X имеет абсолютно непрерывные условные распределения с непрерывными плотностями $f(x|B_j)$, то, дифференцируя (4.30), получим

$$f(x) = \sum_{j=1}^n f(x | B_j) P(B_j). \quad (4.32)$$

Представление распределения случайной величины X в виде суммы (4.30) иногда называют *смесью распределений*, в которой условные распределения играют роль компонент смеси, а вероятности $P(B_j)$ – их весов.

Компоненты смеси распределений могут отличаться друг от друга как по типу, так и по классу. В тех случаях, когда в смеси $F(x)$ присутствуют как непрерывные, так и дискретные распределения, смесь представляет собой так называемое *смешанное распределение*.

Предположим, что смесь распределений содержит группу непрерывных и группу дискретных распределений. Тогда эти группы могут быть объединены и смесь представлена как взвешенная сумма распределений двух классов – непрерывного и дискретного.

Мы, однако, забыли о возможности реализации третьего – сингулярного – класса распределения, осуществимость которого связана с ограничениями, принятыми при определении дискретного и непрерывного распределений: изолированность значений случайной величины при ее дискретном распределении и существование функции $fX(x)$, отвечающей (4.7), при ее непрерывном распределении.

Оказывается, что отказ от этих ограничений приводит к появлению третьего (сингулярного) класса распределения с функцией $F(x)$, обладающей следующими весьма специфическим свойством: она почти всюду непрерывна, $F(\infty) - F(-\infty) = P\{-\infty < X < \infty\} = 1$, но почти всюду $\frac{dF(x)}{dx} = 0$. Это означает, что множество точек роста функции $F(x)$ имеют нулевую меру.

Показано, что функция распределения $F(x)$ всякой случайной величины однозначно представима взвешенной суммой, (т.е. в виде смеси)

$$F(x) = p_1 F^{\circ}(x) + p_2 F''(x) + p_3 F^c(x), \quad p_1 + p_2 + p_3 = 1;$$

(подробнее – см. [7]).

Заметим, что сингулярное распределение в практических приложениях встречается весьма редко; оно не появится и в нашем дальнейшем изложении.

Напротив, смесь дискретного и непрерывного распределений встречается в стохастических моделях достаточно часто. Приведем пример.

Пример 4.4. Фирма продает товар, имеющий непрерывную консистенцию (жидкую или сыпучую) в розницу («вразвес») и оптом. Каждая продажа с вероятностью p_1 является розничной и с вероятностью p_2 – оптовой ($p_1 + p_2 = 1$). X_n – вес одной розничной продажи (непрерывная случайная величина с функцией распределения $F''(x)$), x^* – вес одной оптовой единицы товара, K – число оптовых единиц в одной оптовой продаже (случайная величина, $K = 0, 1, 2, \dots$), $X_d = Kx^*$ – вес одной оптовой продажи (дискретная случайная величина с функцией распределения $F^{\circ}(x)$). Пусть в общей задаче торгового менеджмента требуется найти распределение случайной величины X – веса одной продажи. Имеем

$$X = \begin{cases} X_n & \text{с вероятностью } p_1; \\ X_d & \text{с вероятностью } p_2, \end{cases}$$

откуда

$$P\{X < x\} = p_1 P\{X_n < x\} + p_2 P\{X_d < x\},$$

т.е.

$$F(x) = p_1 F''(x) + p_2 F^{\circ}(x) - \quad (4.33)$$

смесь непрерывного и дискретного распределений.

Выясним теперь, как модифицируется формула Байеса (3.13), когда в моделях, использующих эту формулу, наряду со случайными событиями фигурируют случайные величины. Уточним типичную постановку задачи, в том или ином виде возникающей при построении разнообразных процедур статистической классификации, отбора и распознавания.

Пусть $\{B_j\}_{j=1}^n$ – полная группа несовместимых событий, определяющих сущность наблюдаемого объекта (например, его класс или состояние), которая, однако, наблюдателю доподлинно не известна и вероятность которой оценивается по косвенному признаку, представляющему собой случайную величину X (в отличие от случайного события A в (3.13)). Такая задача возникает, например, в медицине при диагностике типа заболевания

пациента (выбор из событий $\{B_j\}_{j=1}^n$) по значению некоторого доступного к измерению его физиологического параметра – случайной величины X .

Если признак X имеет дискретное распределение, то для каждого его значения $X = x_k$, полагая в (3.13) $A = \{X = x_k\}$, находим

$$P(B_i | X = x_k) = \frac{P(X = x_k | B_i)P(B_i)}{\sum_{j=1}^n P(X = x_k | B_j)P(B_j)},$$

т.е. для каждого полученного значения признака x_k получаем распределение апостериорных вероятностей событий (гипотез) из $\{B_j\}_{j=1}^n$ и возможность выбора из них наиболее вероятного.

В прикладных байесовских моделях чаще, однако, признак X представляет собой случайную величину с непрерывным распределением и поэтому событие $\{X = x_k\}$ имеет нулевую вероятность. Тогда принятое при выводе (3.13) условие $P(A) \neq 0$ не позволяет без оговорок положить $A = \{X = x\}$. Рассмотрим этот случай в предположении, что признак X обладает непрерывными условными плотностями распределения $\{f(x|B_j)\}$.

Обратимся к (4.32) и обозначим $X = \{x: f(x) \neq 0\}$ (т.е. для $\forall x \in X \exists j: f(x|B_j) \neq 0$). Пусть, далее, $\Delta_\varepsilon = [x, x+\varepsilon)$, где $x \in X$. Тогда

$$P\{X \in \Delta_\varepsilon\} = \int_x^{x+\varepsilon} f(x)dx = \varepsilon f(x^*) > 0, \quad (x^* \in \Delta_\varepsilon).$$

Полагая в (3.13) $A = \{X \in \Delta_\varepsilon\}$, получим

$$P(B_i | X \in \Delta_\varepsilon) = \frac{P(X \in \Delta_\varepsilon | B_i)P(B_i)}{P\{X \in \Delta_\varepsilon\}},$$

где

$$P\{X \in \Delta_\varepsilon | B_i\} = \int_x^{x+\varepsilon} f(x | B_i)dx = \varepsilon f(x^{**} | B_i), \quad (x^{**} \in \Delta_\varepsilon).$$

При $\varepsilon \rightarrow 0$ получаем $f(x^*) \rightarrow f(x)$, $f(x^{**}|B_i) \rightarrow f(x|B_i)$, $\{X \in \Delta_\varepsilon\} \rightarrow \{X=x\}$ и

$$P(B_i | x) = \frac{f(x | B_i)P(B_i)}{f(x)} \quad (4.34)$$

(заметим, что при несовпадении множеств X и $X_i = \{x: f(x|B_i) \neq 0\}$ ($X_i \subset X$) для $x \in X \setminus X_i$ получаем равенство $P(B_i | x) = 0$).

Полезной модификацией формулы полной вероятности (3.12) является очевидное равенство

$$P(B_i) = \int_{-\infty}^{\infty} P(B_i | x)f(x)dx. \quad (4.35)$$

Числовые характеристики случайных величин

Формализуя понятие скалярной случайной величины, мы установили, что ее свойства полностью описываются функцией распределения. В тех случаях, когда тип распределения задан, конкретный вид функции распределения определяется набором числовых параметров; в рассмотренных типах распределения этот набор обычно состоял из 1–2 величин (исключение составляют смеси распределений, см. (4.32), для которых в число параметров следует включить весовые коэффициенты $\{p_i = P(B_i)\}$).

Наряду с параметрами распределения, имевшими в наших представлениях пока формальный смысл, введем в рассмотрение величины, несущие содержательную информацию о свойствах случайной величины: об ее среднем значении, о величине разброса ее значений и пр. Эти величины в дальнейшем будем называть *числовыми характеристиками* случайной величины.

В общем случае числовые характеристики случайной величины по информативности уступают функции распределения. Однако при заданном типе распределения они могут конкретизировать его вид так же, как и параметры распределения (находясь с ними обычно в простых функциональных отношениях). В некоторых задачах знание этих характеристик оказывается достаточным и при неизвестных функциях распределения.

Центральным понятием, используемым при описании числовых характеристик случайных величин, является среднее значение случайной величины и средние значения ее функций. В теории вероятности синонимом «*среднее значение*» является термин «*математическое ожидание*».

Математическое ожидание случайной величины

Прежде чем дать формальное определение понятию «математическое ожидание», коснемся кратко его интуитивных истоков.

Начнем с простого случая, когда наблюдаемая случайная величина X имеет конечное множество возможных значений $X = \{x_j\}_{j=1}^n$. Пусть проведено N независимых измерений этой величины, что можно уподобить измерению N случайных величин X_1, X_2, \dots, X_N , каждая из которых имеет то же распределение, что и случайная величина X , вне зависимости от результатов измерения других величин из $\{X_i\}$.

Представим полученные реализации случайных величин X_1, X_2, \dots, X_N (*выборку значений случайной величины X*) набором

$$O_N = (x(1), x(2), \dots, x(N)),$$

каждый член которого является одним из элементов множества X значений X .

В повседневной практике при оценке той или иной недетерминированной величины широко используется величина

$$\bar{x} = \frac{\sum_{i=1}^N x(i)}{N}, \quad (5.1)$$

представляющая собой значение (реализацию) *среднего арифметического* случайной величины X

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} \quad (5.2)$$

(в математической статистике величина \bar{X} обычно именуется *средним выборочным*).

Пусть $N \gg n$. Группируя в (5.1) слагаемые, имеющие равные значения, приводим (5.1) к виду

$$\bar{x} = \frac{\sum_{j=1}^n x_j k_j}{N} = \sum_{j=1}^n x_j w_j, \quad (5.3)$$

где k_j – полученное значение (реализация) случайной величины K_j – числа членов выборки O_N , равных x_j , $w_j = \frac{k_j}{N}$ – реализация случайной величины $W_j = \frac{K_j}{N}$ – частоты события $\{X = x_j\}$.

В новых обозначениях (5.2) принимает вид

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} = \frac{\sum_{j=1}^n x_j K_j}{N} = \sum_{j=1}^n x_j W_j. \quad (5.4)$$

Исходя из статистического определения вероятности, приведенного в главе 1 (см.(1.10), (1.11)), можно заключить, что при определенных условиях с ростом N частоты W_j событий $\{X = x_j\}$ сходятся (в некотором, пока не сформулированном четко смысле) к вероятностям $P\{X_j = x_j\} = p_j$ и, следовательно, величина \bar{X} – к неслучайной величине MX

$$MX = \sum_{i=1}^n x_i p_i, \quad (5.5)$$

которую мы и определим в нашем случае как математическое ожидание (среднее значение) случайной величины X (буква M используется здесь и далее как знак оператора).

Итак, отправляясь от интуитивного (статистического) понятия вероятности, мы определили математическое ожидание случайной величины MX как предел ее среднего арифметического (среднего выборочного) \bar{X} . В наших рассуждениях присутствовала,

однако, некорректность, состоявшая в сопоставлении случайной допредельной величины \bar{X} с неслучайным пределом MX . Это недоразумение ниже мы устраним в рамках принятой аксиоматики теории вероятностей.

Из (5.5) следует, что для дискретной случайной величины с конечным множеством значений ($n < \infty$) математическое ожидание всегда существует и вполне определяется ее распределением.

Пусть теперь случайная величина X остается дискретной, но множество ее возможных значений неограниченно: $X = \{x_j\}_{j=1}^{\infty}$, т.е., иначе говоря, для ограничения n нет оснований (что имеет место, например, когда X – число регистрируемых в единицу времени космических частиц).

Повторяя рассуждения, приведшие нас к (5.5), примем в качестве определения математического ожидания случайной величины X ряд

$$MX = \sum_{j=1}^{\infty} x_j p_j. \quad (5.6)$$

В отличие от определения математического ожидания конечной суммой (5.5), теперь существование математического ожидания MX обусловлено сходимостью ряда (5.6). Эта сходимость должна быть, при этом, абсолютной, поскольку в противном случае значение ряда оказывается зависимым от перестановки его членов, т.е. от порядка поступления и группировки значений выборки $x(1), x(2), \dots$ при формировании суммы ряда, что делает значение MX неопределенным.

Предположим теперь, что случайная величина X имеет абсолютно непрерывное распределение с непрерывной плотностью $f(x)$, и проследим бегло, как, исходя из интуитивных представлений о вероятности, формируется определение математического ожидания в этом случае.

Предположим, что $X = [a, b]$, т.е. $-\infty < a \leq X \leq b < \infty$, и разобьем сегмент $[a, b]$ на n равных непересекающихся отрезков $\{\Delta_i\} : [a, b] = \bigcup_{i=1}^n \Delta_i$. Пусть снова $O_N = (x(1), x(2), \dots, x(N))$ – выборка значений случайной величины X , а k_i теперь число элементов выборки O_N , лежащих в Δ_i . Тогда значение среднего арифметического (5.1) можно приближенно выразить равенством

$$\bar{x} \cong \frac{\sum_{i=1}^n x_i^* k_i}{N} = \sum_{i=1}^n x_i^* w_i \cong \sum_{i=1}^n x_i^* P\{X \in \Delta_i\} = \sum_{i=1}^n x_i^* f(x_i^*) \Delta_i, \quad (5.7)$$

где $x_i^* \in \Delta_i$. Это равенство получено в предположении о непрерывности функции плотности распределения $f(x)$ и в результате применения теоремы о среднем, а также представления

отношения $w_i = \frac{k_i}{N}$ как частоты события $\{X \in \Delta_i\}$, которая снова рассматривается (в интуитивном представлении) как допредельное значение вероятности этого события. Когда $N \rightarrow \infty$, $n \rightarrow \infty$ (так, что для $\forall i \quad |\Delta_i| \rightarrow 0$) правая сумма в (5.7) превращается в интеграл, и мы можем в рассматриваемом случае в качестве математического ожидания случайной величины принять величину

$$MX = \int_a^b xf(x)dx. \quad (5.8)$$

Здесь, когда a и b конечны, MX всегда существует. Чаше, однако, приходится иметь дело с случайными величинами, значения которых не ограничены: $a = -\infty$ и (или) $b = \infty$. В таких случаях рассуждения, подобные приведенным выше, приводят к определению математического ожидания в форме несобственного интеграла (вида, например,

$$MX = \int_{-\infty}^{\infty} xf(x)dx, \quad (5.9)$$

который, как известно, существует и конечен, если сходится абсолютно, т.е. выполнено условия

$$M|X| = \int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$$

(в противном случае математическое ожидание равно $+\infty$ или $-\infty$, или не существует).

Подчеркнем, что изложенные выше соображения не имели доказательной строгости и служили только для содержательного обоснования приводимого ниже определения математического ожидания. При этом, поскольку форма представления математического ожидания случайной величины существенно зависит от класса ее распределения, целесообразно представить это определение в двух формах.

Определение 5.1. Математическим ожиданием MX дискретной скалярной случайной величины X с множеством значений $X = (x_1, x_2, \dots, x_n)$ называется сумма $MX = \sum_{i=1}^n x_i P\{X = x_i\} = \sum_{i=1}^n x_i p_i$, если $n < \infty$, или ряд $MX = \sum_{i=1}^{\infty} x_i P\{X = x_i\} = \sum_{i=1}^{\infty} x_i p_i$, если $n = \infty$ и ряд сходится абсолютно (иначе случайная величина X математического ожидания не имеет).

Определение 5.2. Математическим ожиданием MX скалярной случайной величины X , имеющей абсолютно непрерывное распределение с плотностью $f(x)$, называется величина

$$MX = \int_{-\infty}^{\infty} xf(x)dx,$$

если интеграл сходится абсолютно (иначе случайная величина X конечного математического ожидания не имеет).

Сформулированные определения достаточно полны для практических приложений, в которых, как правило, фигурируют случайные величины с «чистыми» (дискретными и непрерывными) классами распределения; как вскоре мы увидим, не трудно в рамках этих определений рассматривать и случаи смешанных распределений. Вместе с тем, приведенные определения математического ожидания не обладают универсальностью, присущей общей теории. В связи с этим в §5.5 будет дано более общее (универсальное) определение этой (и других) числовой характеристики случайной величины.

Укажем основные свойства математического ожидания, непосредственно вытекающие из определений 5.1 и 5.2.

Если $X = c = \text{const}$, то $MX = c$;

Если $X \geq 0$, то $MX \geq 0$;

$M(cX + b) = cMX + b$ ($c = \text{const}$, $b = \text{const}$);

$|MX| \leq M|X|$.

Заметим, что из определения математического ожидания следует, что перемена значений случайной величины $X = x(\omega)$ для множества нулевой меры не приводит к изменению ее математического ожидания. Поэтому, например, соотношения для X в п.1 и п.2 могут выполняться почти наверное, т.е. для всех исходов ω , за исключением, быть может, их множества нулевой меры.

По ходу изложения список свойств математического ожидания будет пополняться.

Остановимся на вопросе о практическом значении и применении в вероятностных моделях введенной числовой характеристики случайной величины – ее математического ожидания.

Прежде всего подчеркнем, что математическое ожидание случайной величины является характеристикой, приобретающей практический смысл лишь в вероятностных моделях, описывающих массовые явления – соударения большого числа частиц, формирование суммарной прибыли в интенсивном потоке коммерческих сделок, расчеты надежности многокомпонентных технических объектов, оценка ущерба от массовых катаклизмов и т.д. Применение этой характеристики для оценки результата единичной операции (например, от эксклюзивной коммерческой операции) бессмысленно.

Представление о математическом ожидании как о числе, позволяющем оценить саму случайную величину, нередко приводит к ошибочным заключениям. Пусть, например,

случайная величина X с равными вероятностями принимает одно из двух значений c или $-c$, $|c| > 0$. Ясно, что $MX = 0$ при любом сколь угодно большом c .

В связи с этим бытующая во многих приложениях практика оценки состояния конкретных реалий только лишь математическим ожиданием (средним значением) той или иной величины не только некорректна, но подчас и вредна, если одновременно не используются другие характеристики и параметры, позволяющие более полно анализировать моделируемое явление.

Все это не умаляет, однако, важности математического ожидания как центра рассеяния («характеристики расположения») значений случайной величины на числовой прямой и его ключевой роли, которую оно играет при формировании других числовых характеристик случайных величин, о которых речь пойдет ниже.

Уместно заметить, что альтернативой математическому ожиданию, как характеристике расположения значений случайной величины на числовой прямой, в некоторых задачах могут служить *мода* и *медиана* ее распределения, имеющие следующий смысл.

Пусть случайная величина X имеет абсолютно непрерывное распределение с плотностью $f(x)$, обладающей глобальным максимумом. Его модой x_{mod} называют значение X , при котором плотность распределения имеет максимум:

$$x_{mod} = \arg \max_x f(x).$$

Грубо говоря, мода распределения соответствует «наиболее вероятному» значению случайной величины.

Иногда понятие моды распространяют и на дискретное распределение, полагая $x_{mod} = x_{i^*}$, где x_{i^*} удовлетворяет условию $P\{X = x_{i^*}\} = \max_{x_i} P\{X = x_i\}$.

Медианой распределения называют число μ_e , удовлетворяющее условию

$$P\{X < \mu_e\} = P\{X > \mu_e\},$$

означающему равновероятность того, что реализация случайной величины X окажется меньше или больше μ_e . Другими словами, медиана разбивает упорядоченное по возрастанию множество X значений случайной величины X на два подмножества $X_1 = (x : x < \mu_e)$, $X_2 = (x : x > \mu_e)$ с равными вероятностными мерами:

$$P\{X \in X_1\} = P\{X \in X_2\}.$$

При непрерывном распределении X $P\{X = \mu_e\} = 0$ и $F_X(\mu_e) = 1 - F_X(\mu_e)$, $\mu_e = F_X^{-1}(1/2)$.

Математическое ожидание функции случайной величины. Моменты

Во многих приложениях, использующих вероятностные модели реальных явлений и процессов, возникает необходимость вычисления математического ожидания функции случайной величины при известном распределении последней. Простым примером такой задачи является вычисление средней мощности, выделяемой в электроприборе, при известном распределении входного напряжения.

Пусть $Y = \psi(X)$ – борелевская функция, X – случайная величина с известным распределением и задача заключается в вычислении математического ожидания MY случайной величины Y . Возможный путь ее решения состоит в нахождении распределения случайной величины Y (см. §4.5) и вычислении MY по одной из формул, приведенных в определениях 5.1 и 5.2 (в зависимости от класса распределения Y):

$$MY = \begin{cases} \sum_{i=1}^n y_i P\{Y = y_i\} \quad (n \leq \infty) \text{ при дискретном} \\ \text{распределении } Y; \\ \int_{-\infty}^{\infty} y f_Y(y) dy \text{ при непрерывном.} \\ \text{распределении } Y \end{cases} \quad (5.10)$$

Оказывается, однако, что этот путь неоправданно усложнен, поскольку для вычисления MY переход к распределению случайной величины Y вовсе не нужен, и для вычисления математического ожидания случайной величины Y (функции) можно непосредственно использовать распределение случайной величины X (аргумента) согласно формулам

$$MY = \begin{cases} \sum_{i=1}^n \psi(x_i) P\{X = x_i\} \quad (n \leq \infty) \text{ при дискретном} \\ \text{распределении } X; \\ \int_{-\infty}^{\infty} \psi(x) f_X(x) dx \text{ при непрерывном} \\ \text{распределении } X, \end{cases} \quad (5.11)$$

(условием существования MY в (5.11) является абсолютная сходимость фигурирующих здесь рядов и интегралов).

Чтобы убедиться в справедливости (5.11), следует проверить, что эти формулы дают те же значения MY , что и (5.10). Рассмотрим случай дискретного распределения случайных величин X и Y (непрерывный случай основан на аналогичных умозаключениях).

Представим сумму в (5.11) в виде

$$S_n = \sum_{i=1}^n \psi(x_i) P\{X = x_i\} = \sum_{j=1}^{n'} y_j P\{X \in \Delta_j\} = MY, \quad (5.12)$$

где n' – число возможных значений случайной величины $Y = \psi(X)$ ($n' \leq n$) и

$$\Delta_j = \{x : \psi(x) = y_j\}, \quad P\{X \in \Delta_j\} = P\{Y = y_j\}.$$

При $n < \infty$ равенство $S_n = MY$ очевидно. Оно имеет место и при $n = \infty$ ввиду абсолютной сходимости ряда S_n и допустимости, в связи с этим, перегруппировки его членов.

Из эквивалентных представлений (5.10) и (5.11) математического ожидания функции случайной величины $Y = \psi(X)$ второе используется более часто; поэтому его мы и используем в следующем определении.

Определение 5.3. Математическим ожиданием борелевской скалярной функции $Y = \psi(X)$ от скалярной случайной величины X называется величина MY , определенная равенствами (5.11). Конечное математическое ожидание MY существует тогда и только тогда, когда ряд или интеграл в (5.11) сходится абсолютно.

Ниже функция $Y = \psi(X)$ будет у нас принимать различный вид в зависимости от решаемой задачи.

Важную роль в качестве числовых характеристик случайной величины играют ее моменты, вводимые следующими определениями.

Определение 5.4. Начальным моментом r -го порядка α_r ($r = 1, 2, \dots$) случайной величины X называют математическое ожидание степени X^r , т.е. величину

$$\alpha_r = MX^r = \begin{cases} \sum_{i=1}^n x_i^r P\{X = x_i\} & (n \leq \infty) \text{ при} \\ \text{дискретном распределении } X; \\ \int_{-\infty}^{\infty} x^r f_X(x) dx & \text{при} \\ \text{непрерывном распределении } X. \end{cases} \quad (5.13)$$

Начальный момент α_r существует, если выражающий его ряд или интеграл сходится абсолютно. Ясно, что $\alpha_1 = MX$.

Полезно заметить, что из существования начального момента r -го порядка следует существование всех начальных моментов порядка ниже r . Покажем это для непрерывного распределения X (для дискретного распределения доказательство аналогично). Пусть $s < r$. Тогда справедливо неравенство

$$\begin{aligned} \int_{-\infty}^{\infty} |x^s| f_X(x) dx &= \int_{|x| < 1} |x^s| f_X(x) dx + \int_{|x| \geq 1} |x^s| f_X(x) dx \leq \\ &\leq \int_{|x| < 1} |x^s| f_X(x) dx + \int_{-\infty}^{\infty} |x^r| f_X(x) dx < \infty \end{aligned}$$

(ввиду очевидной ограниченности первого правого интеграла и второго – по условию существования момента α_r).

Определение 5.5. Центральным моментом r – го порядка μ_r ($r = 1, 2, \dots$) случайной

величины X называют математическое ожидание степени $\left(\overset{0}{X} \right)^r$, где $\overset{0}{X} = X - MX$ – центрированное значение случайной величины X , т.е. величину

$$\mu_r = M \overset{0}{X}^r = \begin{cases} \sum_{i=1}^n (x_i - MX)^r P\{X = x_i\} & (n \leq \infty) \\ \text{при дискретном распределении } X; \\ \int_{-\infty}^{\infty} (x - MX)^r f_X(x) dx & \\ \text{при непрерывном распределении } X. \end{cases} \quad (5.14)$$

Центральный момент μ_r существует, если выражающий его ряд или интеграл сходится абсолютно.

В некоторых случаях используются абсолютные начальные и абсолютные центральные моменты, представляющие собой соответственно математические ожидания случайных величин $|X|^r$ и $|X - MX|^r$. Чтобы не множить обозначения, будем обозначать их соответственно $|\alpha|_r$ и $|\mu|_r$. Из приведенных определений следует, что существование для данного r абсолютных моментов $|\alpha|_r$ и $|\mu|_r$ необходимо и достаточно для существования моментов α_r и μ_r .

Дисперсия случайной величины

Как было подчеркнуто, математическое ожидание случайной величины, характеризуя в определенной степени расположение ее значений на числовой прямой, далеко не исчерпывает необходимую для практики информацию об ее свойствах. В первую очередь это относится к разбросу ее значений.

Так, бессмысленно сравнивать социальное положение населения разных стран по среднему материальному достатку их представителей (игнорируя его разброс), опасно оценивать положение торговой фирмы по размеру ее средней прибыли без учета риска, т.е. вероятности получения недопустимо низкой прибыли вследствие разброса ее возможных значений и т.д.

Приступая к определению характеристики разброса значений случайной величины X , рассмотрим возможные меры ее отклонений от некоторого произвольно заданной числовой константы a . Использованию в этой роли непосредственно разности $\Delta a = X - a$ препятствует ее случайность, а переход к ее математическому ожиданию приводит к величине $M\Delta a = MX - a$, которая реагирует на знак отклонения и, поэтому, занижает средний размер отклонения

(так, например, при $a = MX$ он равен нулю, хотя отдельные значения отклонения могут быть сколь угодно велики).

Приемлемыми мерами разброса случайной величины X относительно точки a могут служить величины $M|\Delta_a| = M|X - a|$ (среднее отклонение от a) и $M(\Delta_a)^2 = M(X - a)^2$ (среднее квадратическое отклонение от a).

Выбор константы a в этих величинах основан на том естественном соображении, что она должна представлять собой центр рассеивания случайной величины X , т.е. минимизировать $M|\Delta_a|$ или $M(\Delta_a)^2$ (заметим, что такие значения константы a для этих характеристик разброса могут не совпадать).

Из двух рассмотренных характеристик разброса $M|\Delta_a|$ или $M(\Delta_a)^2$ на практике подавляющее применение имеет вторая, что объясняется, главным образом, удобством ее использования. Покажем, что для нее значение константы a , отвечающее приведенному условию, оказывается равным

$$\arg \min_a M(X - a)^2 = MX.$$

Действительно, пусть $a \neq MX$. Тогда, обозначая $MX = m$, получим

$$\begin{aligned} M(X - a)^2 &= M(X - m + m - a)^2 = M(X - m)^2 + 2(m - a)M(X - m) + (m - a)^2 = \\ &= M(X - m)^2 + (m - a)^2 \geq M(X - m)^2 = M(X - MX)^2. \end{aligned}$$

Полученный результат подтверждает смысл математического ожидания случайной величины как центра ее разброса, выраженного средним квадратическим отклонением $M(\Delta_a)^2$: при $a = MX$ эта величина достигает минимального значения, т.е. значения случайной величины X располагаются относительно MX наиболее компактно.

Все это позволяет применять среднее квадратическое отклонение случайной величины X от ее математического ожидания, т.е. величину $M(X - MX)^2$, как общепринятую меру ее разброса, называемую *дисперсией* и обозначаемую обычно DX (буква D играет роль оператора). Дадим развернутое определение этой величины.

Определение 5.6. Дисперсией скалярной случайной величины X называется величина

$$DX = M(X - MX)^2 = \begin{cases} \sum_{i=1}^n (x_i - MX)^2 P\{X = x_i\}, & n \leq \infty \\ \text{при дискретном распределении } X; \\ \int_{-\infty}^{\infty} (x - MX)^2 f_X(x) dx & \\ \text{при непрерывном распределении } X. \end{cases} \quad (5.15)$$

Дисперсия случайной величины представляет собой, следовательно, ее второй центральный момент: $DX = \mu_2$ (см. (5.14)).

Из (5.15) непосредственно следует удобная для расчетов формула

$$DX = MX^2 - (MX)^2 = \alpha_2 - \alpha_1^2, \quad (5.16)$$

т.е. дисперсия случайной величины существует, если существует ее второй начальный момент α_2 (см.(5.13)); как мы убедились, тогда существует и ее первый начальный момент α_1 .

Читатель легко найдет, что из (5.15) с учетом свойств математического ожидания вытекают следующие основные свойства дисперсии:

1. $DX \geq 0$.
2. Если $X = c = \text{const}$, то $DX = 0$.
3. Если $Y = aX + b$ ($a = \text{const}$, $b = \text{const}$), то $DY = a^2 DX$

(здесь, как и для математического ожидания, достаточно выполнение приведенных условий почти наверное).

Ниже для дисперсии иногда будет использоваться обозначение $DX = \sigma_X^2$, где $\sigma_X = \sqrt{DX}$ именуется *стандартным* или *средним квадратическим отклонением*.

Математическое ожидание и дисперсия случайной величины, не обладая информативностью функции распределения, находят, тем не менее, широкое практическое применение в стохастическом модельном анализе реальных процессов.

Пусть, например, исследуются варианты некоторого делового проекта, показателем которого является *размер успеха*, достигаемого к заданному сроку. Если размер успеха (прибыли, числа перевозок, размер урожая и пр.) – случайная величина X , то используемые в проекте ресурсы естественно, казалось бы, направить на достижение максимального значения математического ожидания MX . Следует, однако, иметь в виду, что разным вариантам проекта соответствуют обычно различные значения дисперсии DX , которая существенно влияет на свойства проекта: чем выше значение DX , тем (при прочих равных условиях) выше *риск* – вероятность получения размера успеха ниже заданного (например – нулевого). Поэтому выбор варианта проекта необходимо делать в рамках консенсуса «успех – риск» с привлечением на первых этапах оценок математического ожидания и дисперсии случайной величины X (подробнее об этом мы поговорим в §5.7).

Приближенным, но очень важным инструментом решения подобных задач является следующая теорема.

Теорема 5.1. (Неравенство Чебышева). Пусть X – случайная величина и существует ее дисперсия DX (следовательно, и ее математическое ожидание MX). Тогда справедливо соотношение

$$\forall \varepsilon > 0: P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}. \quad (5.17)$$

Доказательство проведем для случая абсолютно непрерывного распределения, случай дискретного распределения рассматривается аналогично.

Установим вначале, что для любой неотрицательной случайной величины Y при $\forall a > 0$ имеют место соотношения

$$MY = \int_0^{\infty} y f_Y(y) dy \geq \int_a^{\infty} y f_Y(y) dy \geq a \int_a^{\infty} f_Y(y) dy = aP\{Y \geq a\}, \quad P\{Y \geq a\} \leq \frac{MY}{a}. \quad (5.18)$$

Пусть теперь $Y = (X - MX)^2$, $a = \varepsilon^2$. Тогда

$$P\{(X - MX)^2 \geq \varepsilon^2\} \leq \frac{DX}{\varepsilon^2},$$

откуда следует (5.17).

Заметим, что иногда (5.18) и (5.17) называют, соответственно, первым и вторым неравенствами Чебышева.

Применим неравенство Чебышева для дополнения к свойству 2 дисперсии: если $DX = 0$, то имеет место равенство $P\{X = MX\} = 1$, т.е. случайная величина с вероятностью единица (иначе говоря, почти наверное) совпадает с константой, равной ее математическому ожиданию. Действительно, в этом случае из (5.17) получаем

$$\forall \varepsilon > 0: P\{|X - MX| > \varepsilon\} = 0,$$

откуда, ввиду аддитивности вероятности, следует

$$P\{|X - MX| > 0\} = P\{|X - MX| > 1\} + P\left\{\frac{1}{2} < |X - MX| \leq 1\right\} + \dots + P\left\{\frac{1}{2^k} < |X - MX| \leq \frac{1}{2^{k-1}}\right\} + \dots = 0$$

и $P\{|X - MX| = 0\} = P\{X = MX\} = 1$.

Условное математическое ожидание

В §4.6 было введено понятие условного распределения случайной величины X относительно события B . Возникает естественный вопрос о вычислении математического ожидания величины X (или, в общей постановке, ее функции) при условии осуществления события B , т.е. ее *условного математического ожидания* относительно события B . Когда никаких условий на осуществление события B не ставится, математическое ожидание X (или функции X) называется *безусловным* или *полным* (по отношению к событию B); именно такое математическое ожидание до сих пор нами и рассматривалось.

Переход от безусловного математического ожидания функции $Y = \psi(X)$, определяемого равенствами (5.11), к условному (относительно события B), сводится к замене

в этих соотношениях безусловных распределений X на условные согласно (4.28) и (4.29). Обозначая получаемое условное математическое ожидание $M(Y|B)$, вместо (5.11) находим

$$M(Y|B) = \begin{cases} \sum_{i=1}^n \psi(x_i) P\{X = x_i | B\} & \text{при дискретном} \\ \text{распределении } X, n \leq \infty; \\ \int_{-\infty}^{\infty} \psi(x) f(x|B) dx & \text{при непрерывном} \\ \text{распределении } X. \end{cases} \quad (5.19)$$

Для условного математического ожидания $M(X|B)$ и условной дисперсии $D(X|B)$ случайной величины X соответственно получаем

$$M(X|B) = \begin{cases} \sum_{i=1}^n x_i P\{X = x_i | B\} & \text{при дискретном} \\ \text{распределении } X, n \leq \infty; \\ \int_{-\infty}^{\infty} x f(x|B) dx & \text{при непрерывном} \\ \text{распределении } X \end{cases} \quad (5.20)$$

и

$$D(X|B) = \begin{cases} \sum_{i=1}^n (x_i - MX)^2 P\{X = x_i | B\} & \text{при дискретном распределении } X, n \leq \infty; \\ \int_{-\Gamma}^{\Gamma} (x - MX)^2 f(x|B) dx & \text{при непрерывном распределении } X \end{cases} \quad (5.21)$$

(предполагается, что входящие в (5.19) – (5.21) ряды и интегралы сходятся абсолютно).

Важно подчеркнуть, что условное математическое ожидание (5.19) (и, следовательно, (5.20) и (5.21)) принимает, в общем случае, различные значения для событий B и \bar{B} и потому может рассматриваться как реализация случайной величины Z с распределением: $Z = M(Y|B)$ с вероятностью $P(B)$, $Z = M(Y|\bar{B})$ с вероятностью $1 - P(B)$.

Весьма часто возникает практическая задача вычисления безусловного (полного) математического ожидания случайной величины при известных ее условных распределениях. Простым примером такой задачи является оценка среднего времени безотказной работы наугад выбранного на складе изделия, если складской запас состоит (в известной пропорции) из изделий различных фирм и имеющих, в связи с этим, разные (известные) значения среднего времени безотказной работы.

Пусть снова $Y = \psi(X)$ и $\{B_j\}_{j=1}^k$ – полная конечная или счетная группа несовместимых событий ($k \leq \infty$). Справедлива *формула полного математического ожидания*

$$MY = \sum_{j=1}^k M(Y|B_j)P(B_j), \quad (5.22)$$

где $M(Y|B_j)$ находятся из (5.19).

Убедимся в справедливости (5.22) при абсолютно непрерывном распределении случайной величины X . Отправляясь от (5.11) и используя (4.32), приходим к

$$MY = \int_{-\infty}^{\infty} \psi(x)f(x)dx = \int_{-\infty}^{\infty} \psi(x) \sum_{j=1}^k f(x|B_j)P(B_j)dx = \sum_{j=1}^k P(B_j) \int_{-\infty}^{\infty} \psi(x)f(x|B_j)dx,$$

т.е. к (5.22).

Дискретный случай рассматривается аналогично.

Полагая $\psi(X) = X$ или $\psi(X) = (X - MX)^2$, получаем, соответственно, условное математическое ожидание и условную дисперсию случайной величины X и формулы полного математического ожидания и полной дисперсии X .

Формула полного математического ожидания (5.22) может использоваться для вычисления математического ожидания случайной величины X и ее функций в случаях смешанного распределения X . Пусть, например, X имеет известное дискретное распределение при осуществлении события B_1 и известное непрерывное распределение при осуществлении события B_2 , $B_1 \cup B_2 = \Omega$; вероятности $P(B_1)$ и $P(B_2)$ полагаются известными (на практике события B_1 и B_2 могут быть связаны с значением X : например, $B_1 = \{X < c\}$, $B_2 = \bar{B}_1$). Тогда для полного математического ожидания случайной величины X согласно (5.22) получим

$$MX = M(X|B_1)P(B_1) + M(X|B_2)P(B_2),$$

где $M(X|B_1)$ и $M(X|B_2)$ находятся из (5.20).

Математическое ожидание в форме интегралов Лебега и Стильеса

Основные числовые характеристики случайной величины представляют собой, как мы видели, результат операции математического ожидания над самой случайной величиной или над ее функцией. Такая ведущая роль этой операции в вероятностных моделях предъявляет повышенные требования к ее определению, которое должно сочетать достаточную общность и прагматизм (состоящий в наглядности и в вычислительном удобстве).

С этой точки зрения, приведенные выше определения математического ожидания в большей мере удовлетворяют второму пожеланию: из этих определений относительно легко строятся вычислительные процедуры, но их общность нарушена зависимостью от класса распределения случайной величины и они не охватывают случая сингулярных распределений.

Этот недостаток принятых определений оправдан тем, что в приложениях практически не встречаются случаи, требующие их большей общности. В частности, как было показано, в рамках введенных определений могут рассматриваться и случаи смешанных распределений.

Рассмотрим, тем не менее, кратко возможность более общего определения математического ожидания в форме *интегралов Лебега* и *Стилтьеса*, которое позволяет (в более полной теории) обобщить полученные выше выводы о свойствах этой характеристики. При этом мы полагаем, что читатель знаком с интегралом Лебега, являющимся одним из основных инструментов теории меры – раздела функционального анализа.

Определения математического ожидания 5.1 и 5.2 сформулированы в терминах вторичного вероятностного пространства $\langle X, \mathcal{B}, P_X \rangle$, в котором представление вероятностной меры зависит от класса распределения (дискретного или непрерывного), что и вызывает разнообразие в определении математического ожидания. Для получения общего (универсального) определения этой характеристики, введем его на исходном уровне, т.е. с использованием первичного вероятностного пространства $H = \langle \Omega, \mathcal{F}, P \rangle$.

Итак, пусть заданы вероятностное пространство H и определенная в нем случайная величина (\mathcal{F} - измеримая функция) $X = x(\omega)$. Предположим, поначалу, что множество возможных значений функции $x(\omega)$ не более чем счетно: $X = (x_1, x_2, \dots)$ и для $\forall j$: $\{X = x_j\} = \{\omega \in A_j\}$, $A_j = \{\omega: x(\omega) = x_j\} \in \mathcal{F}$. Обозначим $\mu(A_j) = P\{\omega \in A_j\}$ (выразив таким образом функцией $\mu(\cdot)$ заданную в H вероятностную меру P) и составим сумму (ряд)

$$S_n = \sum_{j=1}^n x_j \mu(A_j), \quad n \leq \infty. \quad (5.23)$$

Ясно, что здесь X – дискретная случайная величина с математическим ожиданием (если оно существует), определяемым равенством (5.6). Очевидно также, что $S_n = MX$ и различие в определении математического ожидания MX равенствами (5.23) и (5.6) состоит лишь в том, что в первом случае мы не выходим за рамки исходного вероятностного пространства H с вероятностной мерой P , а во втором используем вторичное пространство $\langle X, \mathcal{B}, P_X \rangle$ с вероятностной мерой P_X .

Рассмотренная нами функция $X = x(\omega)$ в теории меры носит название *простой функции*, а сумма (ряд) (5.23) – *интеграла Лебега для простой функции*. Абсолютная сходимость ряда S_n (если она имеет место) означает интегрируемость (по Лебегу) функции $x(\omega)$ или, иначе говоря, существование математического ожидания случайной величины X .

Рассмотрим теперь общий случай, когда $X = x(\omega)$ – произвольная F - измеримая функция. Пусть существует последовательность простых интегрируемых функций $\{X_k = x_k(\omega)\}$, таких, что

$$\lim_{k \rightarrow \infty} x_k(\omega) = x(\omega) \text{ (равномерно по } \omega \text{)}.$$

Тогда, если существует предел

$$J = \lim_{k \rightarrow \infty} S_k = \lim_{k \rightarrow \infty} \sum_{j=1}^k x_j \mu(A_j), \quad (5.24)$$

то он представляет собой интеграл Лебега от функции $X = x(\omega)$ по мере μ и обозначается

$$J = \int x(\omega) d\mu(\omega).$$

Подробности, связанные с условиями существования интеграла Лебега и с его свойствами, мы опускаем, отсылая читателя к [8].

Применение интеграла Лебега позволяет обобщить понятие математического ожидания случайной величины следующим определением.

Определение 5.7. Математическим ожиданием случайной величины $X = x(\omega)$ называется значение интеграла Лебега

$$MX = \int x(\omega) d\mu(\omega) \quad (5.25)$$

(при условии существования интеграла $\int |x(\omega)| d\mu(\omega)$).

Это определение математического ожидания, обладая наибольшей общностью, широко используется в общей теории вероятностей.

Вместе с тем его непосредственное применение в прикладных задачах стохастического анализа встречается редко. В связи с этим обратимся еще к одному представлению математического ожидания, сочетающему достаточную универсальность с практическими достоинствами.

Для этого заметим, что функция распределения $F(x)$ для каждой случайной величины X с значениями в пространстве X порождает в исходном вероятностном пространстве вероятностную меру $\mu_F(\omega)$, определяя ее значение для каждого прообраза A борелевского множества B ($A \subseteq \Omega$, $B \subseteq X$). Поэтому математическое ожидание случайной величины $X = x(\omega)$ можно представить интегралом

$$MX = \int x(\omega) d\mu_F(\omega), \quad (5.26)$$

где $\mu_F(\omega)$ – вероятностная мера, порожденная функцией распределения $F(x)$ данной случайной величины X ; формула (5.26) представляет математическое ожидание случайной величины в форме *интеграла Лебега – Стильеса*. Для абсолютно непрерывного распределения с плотностью $f(x) = F'(x)$ равенство

$$\mu_F(\omega) = f(x)dx$$

переводит интеграл (5.26) в интеграл Римана (5.9).

Определенную общность определения математического ожидания, достаточную для формализации прикладных задач, можно, однако, получить, используя *интеграл Стильеса*, формируемый во вторичном вероятностном пространстве по следующей общей схеме.

Пусть $F(x)$ – функция с ограниченным изменением (пока – не обязательно функция распределения), $\psi(x)$ – некоторая произвольная функция.

Назовем интегралом Стильеса от функции $\psi(x)$ с интегрирующей функцией $F(x)$ на интервале $[a, b)$ предел

$$J_{ab}(S) = \lim_{\substack{n \rightarrow \infty \\ \sup \Delta_j \rightarrow 0}} \sum_{j=1}^n \psi(x_j^*) (F(x_j) - F(x_{j-1})),$$

$$(a = x_0 < x_1 < \dots < x_n = b, \quad x_j^* \in [x_{j-1}, x_j],$$

$$\Delta_j = x_j - x_{j-1}),$$
(5.27)

если этот предел существует и не зависит от разбиения интервала $[a, b)$.

При существовании предела

$$J(S) = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow \infty}} J_{ab}(S),$$
(5.28)

сохраняющего значение при произвольной сходимости $a \rightarrow -\infty$, $b \rightarrow \infty$, получаем несобственный интеграл Стильеса от функции $\psi(x)$ с интегрирующей функцией $F(x)$. Для интегралов Стильеса (5.27) и (5.28) приняты обозначения

$$J_{ab}(S) = \int_a^b \psi(\xi) dF(\xi)$$
(5.29)

и

$$J(S) = \int \psi(\xi) dF(\xi).$$
(5.30)

Применим теперь интеграл Стильеса в нашем приложении, конкретно – для универсального представления математического ожидания случайной величины при ее дискретном, непрерывном и смешанном распределениях.

Пусть X – случайная величина с функцией распределения $F_X(x)$. Полагая в (5.26) $F(x) = F_X(x)$, находим (ввиду (5.27), (5.28) и свойств функции распределения), что

$$J(S) = \int \psi(\xi) dF_X(\xi) = \begin{cases} \sum_j \psi(x_j) P\{X = x_j\} \\ \text{при дискретном } X; \\ \int_{-\infty}^{\infty} \psi(\xi) f_X(\xi) d\xi. \\ \text{при непрерывном } X \end{cases}$$

Сравнивая полученные равенства с (5.11), убеждаемся, что для обоих классов распределений случайной величины X значение интеграла Стильеса равно математическому ожиданию случайной величины $Y = \psi(X)$

$$J(S) = \int \psi(\xi) dF_X(\xi) = M(\psi(X)) = MY$$

(заметим, что необходимым и достаточным условием существования конечного математического ожидания является и здесь абсолютная сходимость интеграла Стильеса:

$$\int |\psi(\xi)| dF_X(\xi) < \infty).$$

Нетрудно убедиться, что интеграл Стильеса универсально выражает математическое ожидание случайной величины $Y = \psi(X)$ и при смешанном распределении случайной величины X , когда ее функция распределения представляет собой смесь непрерывного и дискретного распределений.

Вернемся для иллюстрации этого факта к примеру 4.4, дополнив его задачей вычисления математического ожидания MX веса X товара в одной продаже. Используя с этой целью (4.33), выразим MX интегралом Стильеса

$$F(x) = p_1 F''(x) + p_2 F^\circ(x)$$

$$MX = \int x dF(x) = p_1 \int x dF''(x) + p_2 \int x dF^\circ(x),$$

который легко приводится к виду

$$MX = p_1 \int x f_{X''}(x) dx + p_2 x^* \sum_k k P\{K = k\}.$$

В силу такой универсальности интеграла Стильеса далее мы часто будем использовать его в целях общности изложения.

Математическое ожидание и дисперсия для некоторых конкретных типов распределений

Выразим математическое ожидание и дисперсию некоторых распределений случайных величин в виде функций параметров этих распределений.

Распределение Бернулли ($X \in \square(p)$). Для этого распределения (см.(4.9)) находим

$$MX = 1 \cdot p = p, \quad (5.31)$$

$$DX = MX^2 - (MX)^2 = p - p^2 = p(1 - p). \quad (5.32)$$

Биномиальное распределение ($X \in \text{Bi}(n, p)$). Из (4.10) в результате простых преобразований получаем

$$MX = \sum_{x=0}^n x C_n^x p^x (1-p)^{n-x} = np, \quad (5.33)$$

$$DX = \sum_{x=0}^n x^2 C_n^x p^x (1-p)^{n-x} - (np)^2 = np(1-p) \quad (5.34)$$

(заметим, что здесь $MX = nMX_1$ и $DX = nDX_1$, где $X_1 \in B(p)$).

Распределение Пуассона ($X \in \text{Po}(a)$). Из (4.11) легко находим

$$MX = \sum_{x=0}^{\infty} \frac{x a^x}{x!} e^{-a} = a, \quad (5.35)$$

$$DX = \sum_{x=0}^{\infty} \frac{x^2 a^x}{x!} e^{-a} - a^2 = a, \quad (5.36)$$

т.е. для этого распределения $MX = DX = a$.

Рассматривая распределение Пуассона как модель потока событий, мы в главе 4 пришли к формуле (4.12), выражающей распределение числа K событий, происходящих в интервале времени $\theta = [t, t+\tau)$. Теперь из равенств $a = \lambda\tau = MK$, $\lambda = \frac{MK}{\tau}$ определяется смысл параметра λ как среднего числа событий, происходящих за единицу времени в простейшем (пуассоновском) потоке событий, т.е. как *интенсивность* такого потока событий.

Гамма-распределение ($X \in G(\alpha, \lambda)$). Для этого распределения (см. (4.14)) получаем

$$MX = \frac{\int_0^{\infty} \lambda^{\alpha} x^{\alpha} e^{-\lambda x} dx}{\Gamma(\alpha)} = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}, \quad (5.37)$$

$$DX = \frac{\int_0^{\infty} \lambda^{\alpha+1} x^{\alpha+1} e^{-\lambda x} dx}{\lambda \Gamma(\alpha)} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}. \quad (5.38)$$

Из (5.37) и (5.38) получаем MX и DX для распределений, являющихся частными случаями гамма-распределения (т.е. для показательного распределения (4.15), распределений Эрланга (4.17) и χ^2 (4.18)).

Показательное (экспоненциальное) распределение ($X \in \text{Exp}(\lambda)$). Полагая в (5.37) и (5.38) $\alpha = 1$, имеем

$$MX = \frac{1}{\lambda}, \quad (5.39)$$

$$DX = \frac{1}{\lambda^2}. \quad (5.40)$$

Распределение Эрланга ($X \in \text{Erl}(n, \lambda)$). Полагая в этом случае $\alpha = n$, получаем

$$MX = \frac{n}{\lambda}, \quad (5.41)$$

$$DX = \frac{n}{\lambda^2}. \quad (5.42)$$

Распределение χ^2 ($X \in \chi^2(n)$). Это распределение возникает при $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$; отсюда

$$MX = n, \quad (5.43)$$

$$DX = 2n. \quad (5.44)$$

Равномерное распределение ($X \in \text{Eq}(a, b)$, см.(4.13)). Для этого распределения имеем

$$MX = \int_a^b \frac{x dx}{b-a} = \frac{a+b}{2}, \quad (5.45)$$

$$DX = \int_a^b \frac{x^2 dx}{b-a} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}. \quad (5.46)$$

Нормальное (гауссовское) распределение ($X \in N(m, \sigma^2)$). Используя (4.19), найдем

$$MX = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + m) e^{-\frac{y^2}{2}} dy = \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = m \quad (5.47)$$

(здесь справа – интеграл Пуассона),

$$DX = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy = \sigma^2. \quad (5.48)$$

Параметры распределения, числовые характеристики, критерии

Все рассмотренные в §5.6 распределения задаются не более чем двумя параметрами и потому вполне определяются (при известном их типе) числовыми характеристиками – математическим ожиданием и дисперсией. Более того, при однопараметрических распределениях (Бернулли, Пуассона, χ^2 , показательного) для их конкретизации достаточно задать одну из числовых характеристик (математическое ожидание или дисперсию). На практике, однако, нередки стохастические модели, содержащие случайные величины с многопараметрическими распределениями. К ним относятся, например, случайные величины, распределения которых представляют собой смеси распределений различного типа (заметим различие в терминах «смесь распределений» и «смешанное распределение»: в первом случае имеется в виду любое различие распределений, представляющих собой компоненты смеси, во втором – различие классов распределений, входящих в смешанное распределение).

Пусть, например, распределение случайной величины X является смесью k нормальных распределений с весами и параметрами $p_j, m_j, \sigma_j^2, j = \overline{1, k}$, т.е. имеет плотность

$$f_X(x) = \sum_{j=1}^k \frac{p_j}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-m_j)^2}{2\sigma_j^2}}, \quad \sum_{j=1}^k p_j = 1, \quad (5.49)$$

задаваемую $3k - 1$ параметрами. Уже при $k = 2$ математическое ожидание MX и дисперсия DX случайной величины X

$$MX = \sum_{j=1}^k p_j m_j \text{ и } DX = \sum_{j=1}^k p_j \sigma_j^2$$

недостаточны для определения ее распределения. Это означает, что два распределения типа (5.49) с различающимися параметрами могут иметь совпадающие математические ожидания и дисперсии.

Во многих задачах анализа данных и принятия решений случайные величины выступают в роли критериальных переменных, по значениям которых осуществляется оценка и сравнение состояний изучаемых объектов и альтернативный выбор проектов. В таких задачах вводятся *критериальные числовые характеристики* распределений случайных величин, имеющие ясный содержательный смысл в контексте решаемой задачи и позволяющие, поэтому, осуществлять требуемый сравнительный анализ. Будем для краткости называть эти числовые характеристики *критериями*.

Хотя математическое ожидание и дисперсия случайной величины в общем случае имеют больший содержательный смысл, чем параметры распределения, они не всегда могут представлять собой критерии. Рассмотрим пример.

Пример 5.1. Выше мы отмечали недостаточность величины среднего материального достатка жителя страны как обобщенной оценки (критерия) благосостояния ее населения. Включение в число критериев дисперсии этой величины непосредственно не проясняет картину. Действительно, пусть для двух стран величины индивидуального достатка X_1 и X_2 имеют нормальные распределения $X_1 \in N(m_1, \sigma_1^2)$ и $X_2 \in N(m_2, \sigma_2^2)$. Предположим, что $m_1 < m_2$. Если при этом $\sigma_2^2 \leq \sigma_1^2$, то приоритет по благосостоянию населения следует отдать, несомненно, второй стране. При $\sigma_2^2 > \sigma_1^2$, однако, вопрос остается открытым, т.е. дисперсия непосредственно не может служить критерием и должна быть заменена другой более содержательной характеристикой. В нашем примере такой критериальной величиной может служить, например, вероятность того, что индивидуальный достаток жителя страны превышает прожиточный минимум x_0 , т.е. величина

$$\beta_j = P\{X_j > x_0\} = 1 - \Phi\left(\frac{x_0 - m_j}{\sigma_j}\right), \quad j = 1, 2.$$

Теперь приоритет по благосостоянию населения следует отдать второй стране, если $m_1 < m_2$ и $\beta_1 < \beta_2$ (одно из неравенств может быть нестрогим) или первой стране, если $m_1 > m_2$ и

$\beta_1 > \beta_2$ (одно из неравенств может быть нестрогим). При $m_1 = m_2$ и $\beta_1 = \beta_2$ страны по критериям m и β не различаются. Особым является случай, когда $m_1 < m_2$ и $\beta_1 > \beta_2$, т.е. по этим критериям страны формально несопоставимы. Однако, исходя из содержательного смысла используемых критериев, предпочтения заслуживает, по-видимому, социально-экономическое состояние первой страны, в которой доля населения с достатком, превышающим прожиточный минимум, больше, чем во второй стране. Имеющий при этом место более высокий средний достаток жителя второй страны свидетельствует о существовании в ней большего разброса значений индивидуального достатка.

Заметим, что принятое в этом примере предположение о нормальном распределении индивидуального достатка X является крайним упрощением. Действительно, ввиду высокого расслоения населения каждой страны по материальному достатку, распределение величины X в реальности представляет собой многопараметрическую смесь распределений, и формирование системы критериев, позволяющих проводить эффективную оценку и сравнение социально-экономического состояния стран, представляет собой весьма сложную задачу. Не имея возможности рассмотреть ее (и подобные ей задачи в других приложениях) более подробно, ограничимся следующими краткими выводами.

Число эффективно используемых в стохастической модели критериев l_k ограничено сверху числом параметров l_p распределения фигурирующей в ней критериальной случайной величины, поскольку при $l_k > l_p$ $l_k - l_p$ критериев оказываются зависимыми от остальных через параметры распределения и не несут дополнительной информации. С другой стороны, при $l_k < l_p$ «недобор» числа используемых критериев снижает различимость выводов при анализе (как это имеет место, например, в тривиальном случае использования в качестве критерия только математического ожидания при двухпараметрическом нормальном распределении).

Сложность задач описанного типа требует широкого привлечения для ее решения методов стохастического моделирования.

Случайный вектор

До сих пор нами рассматривались скалярные случайные величины; каждая такая величина X определялась как скалярная измеримая функция $X = x(\omega)$, заданная в вероятностном пространстве $\langle \Omega, F, P \rangle$ с множеством значений X в евклидовом пространстве: $X \subseteq R^1$. Большой (если – не главный) интерес представляет стохастическое моделирование реальных явлений, в которых фигурируют совокупности взаимосвязанных случайных величин. К описанию свойств таких совокупностей мы и переходим.

Общее определение и свойства случайного вектора

Пусть в общем вероятностном пространстве $H = \langle \Omega, F, P \rangle$ задано n ($2 \leq n < \infty$) скалярных случайных величин $\{X_j\}_{j=1}^n$, т.е. n F – измеримых функций $\{x_j(\omega)\}_{j=1}^n$ с множествами возможных значений $\{X_j\}_{j=1}^n$, $\forall j: X_j \subseteq R^1$.

В практическом приложении эти случайные величины могут иметь разнообразную сущность, обладая, вместе с тем, в совокупности важной информацией о состоянии или свойствах изучаемого явления (свойства объекта). Примером может служить совокупность физиологических параметров, описывающих состояние живого организма.

Это приводит к представлению набора случайных величин $\{X_j\}$ в виде вектора $X = (X_1, \dots, X_n)'$ (' – знак транспонирования) или, иначе говоря, вектор-функцией $x(\omega) = (x_1(\omega), \dots, x_n(\omega))'$.

Поскольку решаемые практические задачи сводятся, как правило, к вычислению распределения вероятностей такого вектора (вероятностей его точечных или интервальных значений), здесь, как и в скалярном случае, возникает вопрос об измеримости вектор-функции $x(\omega)$, который решается принципиально также, как и для скалярной случайной величины.

Представим пространство X значений вектора $X = x(\omega)$ как произведение $X = X_1 \times \dots \times X_n$, $X \subseteq R^n$, и введем в X борелевскую σ -алгебру B , т.е. минимальную σ -алгебру, содержащую все n -мерные полуинтервалы вида $[a_1, b_1) \times \dots \times [a_n, b_n)$ (как и в скалярном случае, при этом в B попадают и все промежутки других типов, включая и вырожденные случаи: точки, прямые и пр.). Функция $x(\omega)$ называется измеримой, а вектор X становится *случайным вектором*, если F -измеримо каждое множество B из B , т.е. выполняется условие

$$\forall B \in B: A(B) = \{\omega: x(\omega) \in B\} \in F, \quad (6.1)$$

где $A(B)$ – прообраз множества B в Ω (ср.(4.2)).

Читатель может самостоятельно проверить, что условие (6.1) измеримости вектор-функции $x(\omega)$ эквивалентно условию измеримости всех ее компонент $x_i(\omega)$, т.е. выполнению соотношений

$$\forall j, B_j \in B_j: A_j(B_j) = \{\omega: x_j(\omega) \in B_j\} \in F, \quad (6.2)$$

где $j = \overline{1, n}$, B_j – борелевская σ -алгебра в X_j , B_j – ее элемент, $A_j(B_j)$ – прообраз B_j в Ω .

Сформулируем в итоге определение:

Определение 6.1. Определенным на вероятностном пространстве $\langle \Omega, \mathcal{F}, P \rangle$ n -мерным случайным вектором (n -мерной случайной величиной, системой n случайных величин) называется n -мерная F – измеримая вектор-функция $X = x(\omega)$.

Как и в скалярном случае, в практических задачах для описания вероятностных свойств случайного вектора X используется обычно не первичное вероятностное пространство $\langle \Omega, \mathcal{F}, P \rangle$, а вторичное (выборочное) вероятностное пространство $\langle X, \mathcal{B}, P_x \rangle$, состоящее из пространства X значений вектора X , борелевской σ -алгебры \mathcal{B} в этом пространстве и распределения вероятностной меры P_X на элементах этой σ -алгебры, которое задает вероятности $P\{X \in B\}$ для всех $B \in \mathcal{B}$.

Оказывается, что и здесь распределение P_X исчерпывающе задается функцией распределения случайного вектора X , определяемой равенством

$$F_X(x) = F_X(x_1, \dots, x_n) = P\{X_1 < x_1, \dots, X_n < x_n\}. \quad (6.3)$$

Чтобы убедиться в этом, следует показать, что функция $F_X(x)$ определяет вероятностные меры, т.е. вероятности $P\{X \in \Delta\}$, для всех n – мерных интервалов (полуинтервалов) $\Delta = \Delta_1 \times \dots \times \Delta_n$ в X (поскольку каждое борелевское множество $B \in \mathcal{B}$ может быть представлено как результат булевых операций над не более чем счетным числом таких интервалов).

Пусть $n = 2$, $X = (X_1, X_2)'$ и $\Delta = \Delta_1 \times \Delta_2 = [x_1^{(0)}, x_1^{(1)}) \times [x_2^{(0)}, x_2^{(1)})$ – произвольный полуинтервал в $X \subseteq R^2$. В результате несложных преобразований, поясняемых рисунком (6.1), получим

$$\begin{aligned} P\{X \in \Delta\} &= P\{X_1 \in \Delta_1, X_2 \in \Delta_2\} = \\ &= P\{x_1^{(0)} \leq X_1 < x_1^{(1)}, x_2^{(0)} \leq X_2 < x_2^{(1)}\} = \\ &= P\{x_1^{(0)} \leq X_1 < x_1^{(1)}, X_2 < x_2^{(1)}\} - P\{x_1^{(0)} \leq X_1 < x_1^{(1)}, X_2 < x_2^{(0)}\} = \\ &= P\{X_1 < x_1^{(1)}, X_2 < x_2^{(1)}\} - P\{X_1 < x_1^{(0)}, X_2 < x_2^{(1)}\} - \\ &\quad - P\{X_1 < x_1^{(1)}, X_2 < x_2^{(0)}\} + P\{X_1 < x_1^{(0)}, X_2 < x_2^{(0)}\} = \\ &= F_X(x_1^{(1)}, x_2^{(1)}) - F_X(x_1^{(0)}, x_2^{(1)}) - F_X(x_1^{(1)}, x_2^{(0)}) + F_X(x_1^{(0)}, x_2^{(0)}) = \\ &= (-1)^{\sum_i \alpha_i} \sum_{(\alpha_1, \alpha_2)} F_X(x_1^{(\alpha_1)}, x_2^{(\alpha_2)}), \end{aligned} \quad (6.4)$$

где $\alpha_i = 0$ или 1 (суммирование выполняется по всем наборам (α_1, α_2)).

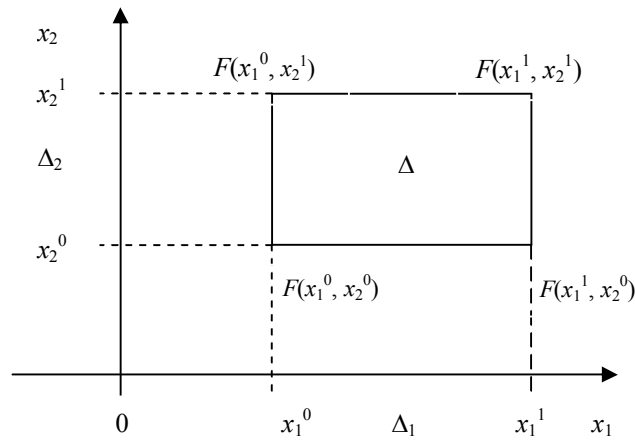


Рис. 0.1RABIC \s 1 1.

При $n > 2$ для $\Delta = [x_1^{(0)}, x_1^{(1)}] \times \dots \times [x_n^{(0)}, x_n^{(1)}]$ простые (хотя и более громоздкие) вычисления приводят к равенству

$$P\{X \in \Delta\} = \begin{cases} (-1)^{\sum_i \alpha_i} \sum_{(\alpha_1, \dots, \alpha_n)} F_X(x_1^{(\alpha_1)}, \dots, x_n^{(\alpha_n)}), & \text{если } n \text{ четно,} \\ (-1)^{\sum_i \alpha_i - 1} \sum_{(\alpha_1, \dots, \alpha_n)} F_X(x_1^{(\alpha_1)}, \dots, x_n^{(\alpha_n)}), & \text{если } n \text{ нечетно.} \end{cases} \quad (6.5)$$

Заметим, что (6.4) и (6.5) представляют вероятность $P\{X \in \Delta\}$ как разность n -гопорядка функции $F_X(\mathbf{x}) = F_X(x_1, \dots, x_n)$ по аргументам x_1, \dots, x_n с приращениями $h_j = x_j^{(1)} - x_j^{(0)}$, $j = 1, 2, \dots, n$. Отсюда следует, что одним из необходимых свойств функции распределения случайного вектора является неотрицательность этой ее разности для $\forall \Delta$.

Поскольку, таким образом, функция распределения $F_X(x_1, \dots, x_n)$ вполне определяет вероятностные свойства вектора X , случайный вектор можно определить (наряду с определением 6.1) как совокупность n заданных во вторичном пространстве случайных величин, обладающих совместной функцией распределения.

Рассмотрим свойства функции распределения случайного вектора.

Прежде всего заметим, что, как и в скалярном случае, характер функции распределения случайного вектора зависит от свойства множества его возможных значений X . Когда оно конечно или состоит из счетного числа изолированных точек, то случайный вектор имеет дискретное распределение и его функция распределения представляет собой разрывную ступенчатую функцию.

В тех случаях, когда существует неотрицательная функция $f_X(x_1, \dots, x_n)$, такая, что

$$\begin{aligned} \forall (x_1, \dots, x_n); F_X(x_1, \dots, x_n) = \\ = \int_{-\infty}^{x_1} \int_{-\infty}^{x_n} f_X(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n, \end{aligned} \quad (6.6)$$

распределение и сам случайный вектор X называют *абсолютно непрерывным* (или просто непрерывным); функция $f_X(x_1, \dots, x_n)$, равная

$$f_X(x_1, \dots, x_n) = \frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n},$$

называется *плотностью распределения случайного вектора X* . Для малой области Δ в пространстве X

$$\Delta = [x_1, x_1 + \Delta x_1) \times \dots \times [x_n, x_n + \Delta x_n)$$

при существовании непрерывной плотности распределения $f_X(\cdot)$ имеет место равенство

$$\begin{aligned} P\{X \in \Delta\} &= P\{x_1 \leq X_1 < x_1 + \Delta x_1, \dots, x_n \leq X_n < x_n + \Delta x_n\} = \\ &= \left[\frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} \right]_{x_1, \dots, x_n} \Delta x_1 \dots \Delta x_n + o(\Delta x_1 \dots \Delta x_n) = \\ &= f(x_1, \dots, x_n) \Delta x_1 \dots \Delta x_n + o(\Delta x_1 \dots \Delta x_n), \end{aligned}$$

и для произвольной области $G \in \mathcal{B}$ получим

$$P\{X \in G\} = \int_G f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Бесконечно малую $f_X(x_1, \dots, x_n) dx_1 \dots dx_n$ называют элементом вероятности.

Наряду с дискретными и непрерывными распределениями случайного вектора возможны, конечно, и случаи смешанного распределения, причем смешивание классов распределений здесь может иметь место как между компонентами случайного вектора, так и внутри компонент. Как и скалярная случайная величина, случайный вектор или его компоненты могут иметь сингулярное распределение (представляющее собой, впрочем, в основном теоретический интерес и далее в нашем изложении не фигурирующее).

Вспомнив вывод свойств функции распределения скалярной случайной величины, читатель без труда убедится в аналогичных свойствах функции $F_X(x_1, \dots, x_n)$:

1. Неотрицательность:

$$F_X(x_1, \dots, x_n) \geq 0 \text{ для } \forall (x_1, \dots, x_n) \in X.$$

2. Монотонность: для $\forall j$, если $x_j'' > x_j'$, то

$$F_X(x_1, \dots, x_j'', \dots, x_n) \geq F_X(x_1, \dots, x_j', \dots, x_n).$$

3. Для $\forall j$ справедливо равенство

$$\lim_{x_j \rightarrow -\infty} F_X(x_1, \dots, x_n) = F_X(x_1, \dots, x_{j-1}, -\infty, x_{j+1}, \dots, x_n) = 0.$$

4. $\lim_{\substack{x_1 \rightarrow \infty \\ x_n \rightarrow \infty}} F_X(x_1, \dots, x_n) = F_X(\infty, \dots, \infty) = 1.$

5. Непрерывность слева в точках разрыва:

$$\forall j \lim_{x_j' \uparrow x_j^*} F_X(x_1, \dots, x_j', \dots, x_n) = F_X(x_1, \dots, x_j^*, \dots, x_n),$$

где x_j^* – точка разрыва $F_X(x)$ по j -му аргументу (возможны разрывы первого рода, число разрывов не более, чем счетно).

Перечисленные свойства функции $F_X(x_1, \dots, x_n)$ дополняются следующими новыми, связанными с многомерностью ее аргумента:

6. Для $\forall \Delta = [x_1^{(0)}, x_1^{(1)}] \times \dots \times [x_n^{(0)}, x_n^{(1)}]$ разность n -го порядка функции $F_X(x_1, \dots, x_n)$ по аргументам x_1, \dots, x_n с приращениями $h_j = x_j^{(1)} - x_j^{(0)}$, $j = 1, 2, \dots, n$ неотрицательна (см.(6.5));

7. Согласованность:

$$\begin{aligned} \forall j : \lim_{x_j \rightarrow \infty} F_X(x_1, \dots, x_j, \dots, x_n) &= \\ &= F_{X'}(x_1, \dots, x_{j-1}, \infty, x_{j+1}, \dots, x_n) = \\ &= F_{X'}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \end{aligned} \quad (6.7)$$

(здесь $F_{X'}(\cdot)$ – функция распределения вектора X' , получающегося из вектора X исключением компоненты X_j).

Свойство 6 было объяснено выше; свойство 7, как и свойства 3, 4 и 5, легко доказывается с привлечением теоремы о непрерывности вероятности (в чем читателю предлагается убедиться самостоятельно).

На первый взгляд может показаться, что свойства 2 и 6 дублируют друг друга. Что это не так, читатель может проверить, придумав примеры, когда из свойства 2 не следует свойство 6 и наоборот.

Плотность распределения $f_X(x_1, \dots, x_n)$ имеет следующие очевидные свойства:

$$1. f_X(x_1, \dots, x_n) \geq 0;$$

$$2. \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(x_1, \dots, x_n) dx_1 \dots dx_n = 1;$$

$$3. \forall j : f_{X'}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \int_{-\infty}^{\infty} f_X(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) dx_j$$

(последнее равенство следует из свойства 7 функции распределения).

Приведенные свойства функций распределения и плотности распределения случайного вектора позволяют находить распределение любого его подвектора, в том числе – частные (*маргинальные*) распределения его компонент. Так, например, частная функция распределения компоненты X_1 выражается равенством:

$$F_{X_1}(x_1) = F_X(x_1, \infty, \dots, \infty);$$

а для плотности распределения (если она существует) получаем:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} (n-1) \int_{-\infty}^{\infty} f_X(x_1, \dots, x_n) dx_2 \dots dx_n.$$

Все это означает, что распределение случайного вектора полностью определяет распределения всех его подвекторов. Чтобы найти частную функцию распределения группы компонент случайного вектора, следует в его функции распределения положить бесконечности значения всех компонент, не входящих в данную группу. Частная плотность распределения этой группы (при непрерывном распределении X) получается в результате интегрирования плотности распределения вектора X по всем значениям его компонент, не вошедших в данную группу.

Аналогично скалярному случаю (см.(4.27)), определим условную функцию распределения случайного вектора X относительно случайного события B равенством

$$F_X(x_1, \dots, x_n | B) = F_X(\mathbf{x} | B) = P\{X_1 < x_1, \dots, X_n < x_n | B\} = P\{X < \mathbf{x} | B\}.$$

Тождество $F_X(\mathbf{x} | B) \equiv F_X(\mathbf{x})$ означает взаимную независимость случайного вектора X и случайного события B .

Пусть $\{B_j\}_{j=1}^k$ – полная группа несовместимых событий ($k \leq \infty$). Тогда, как и в скалярном случае (см.(4.30)), справедлива формула полной вероятности:

$$F_X(\mathbf{x}) = \sum_{j=1}^k F_X(\mathbf{x} | B_j) P(B_j). \quad (6.8)$$

Перейдем теперь к вопросу о вероятностной взаимозависимости компонент случайного вектора, во многом повторяя рассуждения, приведенные в главе 3 о зависимости случайных событий.

Начнем, однако, со следующего замечания. Говоря о зависимости между случайными величинами, мы постоянно имеем в виду ее вероятностный смысл, заключающийся в зависимости распределения одной из них от значения другой. Не следует, однако, исключать и возможность существования между ними функциональной зависимости в обычном смысле. В самом простом случае для двух случайных величин X_1 и X_2 такая зависимость, называемая *линейной собственной зависимостью*, имеет вид

$$\begin{aligned} c_1 X_1 + c_2 X_2 &= c, \\ c_1 &= \text{const} \neq 0, c_2 = \text{const} \neq 0, c = \text{const}. \end{aligned} \quad (6.9)$$

Появления такого рода зависимостей (которые очевидным образом определяются и при $n > 2$), будут оговариваться особо.

Возвращаясь к вероятностной зависимости, начнем с двумерного случайного вектора $X = (X_1, X_2)'$; X_1 и X_2 – пространства значений его компонент X_1 и X_2 , B_1 и B_2 – борелевские σ -алгебры в этих пространствах. Существование или отсутствие зависимости между

случайными величинами X_1 и X_2 определяется зависимостью или независимостью распределения одной из них от значений другой.

Так, если $\exists B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2$ такие, что

$$P\{X_1 \in B_1 | X_2 \in B_2\} \neq P\{X_1 \in B_1\},$$

то случайная величина X_1 зависит от X_2 и, в силу свойства симметричности вероятностной зависимости случайных событий, случайная величина X_2 зависит от X_1 , т.е.

$$P\{X_2 \in B_2 | X_1 \in B_1\} \neq P\{X_2 \in B_2\}.$$

Наоборот, если для $\forall B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2$

$$P\{X_1 \in B_1 | X_2 \in B_2\} = P\{X_1 \in B_1\},$$

то X_1 не зависит от X_2 и, наоборот, X_2 не зависит от X_1 .

Взаимную независимость компонент X_1 и X_2 можно, следовательно, определить условием:

$$\text{для } \forall B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2 \quad P\{X_1 \in B_1, X_2 \in B_2\} = P\{X_1 \in B_1\} \cdot P\{X_2 \in B_2\}, \quad (6.10)$$

которое в терминах функций распределения эквивалентно условию

$$F_X(x_1, x_2) \stackrel{(x_1, x_2)}{=} F_{X_1}(x_1) \cdot F_{X_2}(x_2). \quad (6.11)$$

Последнее утверждение требует пояснения. Действительно, (6.11) есть частный случай (6.10) при $B_1 = (-\infty, x_1)$, $B_2 = (-\infty, x_2)$. Обратно, используя (6.4), легко вывести (6.10) из (6.11) для полуинтервалов $B_1 = \Delta_1 = [x_1^{(0)}, x_1^{(1)})$ и $B_2 = \Delta_2 = [x_2^{(0)}, x_2^{(1)})$.

Распространить последний результат на случай произвольных борелевских множеств B_1 и B_2 можно на основании следующих соображений.

Используя правила булевой алгебры, множества B_1 и B_2 представим в виде объединений непересекающихся полуинтервалов:

$$B_1 = \bigcup_{i=1}^{n_1} \Delta_1^i, \quad n_1 \leq \infty, \quad B_2 = \bigcup_{j=1}^{n_2} \Delta_2^j, \quad n_2 \leq \infty,$$

где $\Delta_1^i = [x_1^{(1)}(i), x_1^{(2)}(i))$, $\Delta_2^j = [x_2^{(1)}(j), x_2^{(2)}(j))$. Обозначим их вероятностные меры $p_i = P\{X_1 \in \Delta_1^i\}$ и $q_j = P\{X_2 \in \Delta_2^j\}$. Имеем

$$P\{X_1 \in B_1\} = \sum_{i=1}^{n_1} p_i, \quad P\{X_2 \in B_2\} = \sum_{j=1}^{n_2} q_j, \quad B_1 \times B_2 = \bigcup_{i,j} (\Delta_1^i \times \Delta_2^j).$$

Как мы установили, из (6.4) и (6.11) следует

$$P\{X_1, X_2 \in (\Delta_1^i \times \Delta_2^j)\} = P\{X_1 \in \Delta_1^i\} P\{X_2 \in \Delta_2^j\} = p_i q_j, \quad \text{откуда}$$

$$P\{(X_1, X_2) \in B_1 \times B_2\} = \sum_{i,j} p_i q_j = \sum_{i=1}^{n_1} p_i \sum_{j=1}^{n_2} q_j = P\{X_1 \in B_1\} \cdot P\{X_2 \in B_2\},$$

в чем мы и хотели убедиться.

При абсолютно непрерывном распределении вектора X с плотностью $f_X(x_1, x_2)$ из (6.11) следует

$$f_X(x_1, x_2) \stackrel{(x_1, x_2)}{=} f_{X_1}(x_1) \cdot f_{X_2}(x_2) \quad (6.12)$$

Обратимся теперь к общему случаю n -мерного случайного вектора $X = (X_1, \dots, X_n)'$. Для каждой пары его компонент X_i, X_j можно определить условие их взаимной независимости равенствами (6.9) или (6.10) при замене в них индексов 1 и 2 на i и j соответственно. При выполнении этих условий для всех пар (i, j) случайный вектор X имеет *попарно независимые* компоненты.

Как и для случайных событий, попарная независимость компонент случайного вектора не означает, однако, их независимость в более общем смысле – независимость в совокупности, т.е. независимость не отдельных компонент случайного вектора X , а их групп – подвекторов вектора X . Убедиться в этом можно, вернувшись к примеру 3.1, заменив в нем случайные события A_1, A_2, A_3 их индикаторами – бернуллиевыми случайными величинами X_1, X_2, X_3 .

Сформулируем критерий независимости компонент случайного вектора в совокупности.

Составим из компонент вектора X два непересекающихся подвектора $X(1) = (X_{i_1}, \dots, X_{i_r})'$ и $X(2) = (X_{j_1}, \dots, X_{j_t})'$ ($i_l \neq j_m$ для $\forall l = \overline{1, r}, m = \overline{1, t}$), $r + t \leq n$.

Обозначим соответственно $X(1)$ и $X(2)$ пространства значений этих векторов, $B_{X(1)}$ и $B_{X(2)}$ – борелевские σ -алгебры в этих пространствах. Тогда условие независимости векторов $X(1)$ и $X(2)$ имеет вид:

$$\begin{aligned} &\text{для } \forall B_1 \in B_{X(1)}, B_2 \in B_{X(2)} \\ &P\{X(1) \in B_1, X(2) \in B_2\} = P\{X(1) \in B_1\} \cdot P\{X(2) \in B_2\}. \end{aligned} \quad (6.13)$$

Распространяя это условие на любую пару непересекающихся подвекторов $X(1)$ и $X(2)$ случайного вектора X , мы приходим к критерию независимости его компонент в совокупности, которое включает в себя, как частный случай, попарную независимость компонент. Нетрудно убедиться, что это условие может быть выражено по индукции из (6.13) в форме:

$$\begin{aligned} &\text{для } \forall B_j \in B_j, j = \overline{1, n}, \\ &P\left\{\bigcap_{j=1}^n X_j \in B_j\right\} = \prod_{j=1}^n P\{X_j \in B_j\}. \end{aligned} \quad (6.14)$$

(здесь B_j – борелевская σ -алгебра в пространстве значений компоненты X_j).

Как и в случае $n = 2$ (см. (6.10), (6.11)) последнее условие может быть представлено в терминах функций распределений в виде:

$$\text{для } \forall x_1, \dots, x_n; \quad F_X(x_1, \dots, x_n) = \prod_{j=1}^n F_{X_j}(x_j). \quad (6.15)$$

Действительно, (6.15) есть просто частный случай (6.14), а показать как (6.14) следует из (6.15) можно тем же путем, что и при $n = 2$.

Если X – непрерывный случайный вектор, то условие (6.15) можно выразить с помощью функций плотности распределения:

$$f_X(x_1, \dots, x_n) \equiv \prod_{i=1}^n f_{X_i}(x_i). \quad (6.16)$$

Для дискретного вектора X условие (6.14) можно представить в виде:

$$\forall j_1, \dots, j_n; P\{X_1 = x_1^{(j_1)}, \dots, X_n = x_n^{(j_n)}\} = \prod_{i=1}^n P\{X_i = x_i^{(j_i)}\} \quad (6.17)$$

(здесь $\{x_i^{(j)}\}_{j=1}^{m_i}$ – множество значений i -ой компоненты X_i , m_i – их число, $(m_i \leq \infty)$).

В итоге получаем определение:

Определение 6.2. Компоненты случайного вектора называются независимыми в совокупности, если выполняется условие (6.15).

Сравнивая определение 6.2 с определением 3.4 независимости в совокупности случайных событий, заметим, что условие (6.15) содержит одно равенство, в то время как условие (3.10) представляет собой набор равенств. Причина этого заключается в том, что, ввиду свойства (6.7) функции распределения случайного вектора, из (6.15) следует аналогичное равенство для распределения любого подвектора вектора X , если положить равными ∞ все компоненты, не входящие в этот подвектор.

Условия (6.15) - (6.17), играя роль определения независимости компонент случайного вектора в совокупности, на практике, однако, обычно выступают как следствия предположения о независимости компонент, вытекающего из содержательного анализа исследуемой реальности. Подчеркнем, что хотя такие предположения и способствуют существенному упрощению модели реального явления, в случае их недостаточной обоснованности могут приводить к серьезным ошибкам и к потере важной информации, заключенной в связях между компонентами. Иллюстрацией этого может служить пример 3.2, если фигурирующие там случайные события заменить их индикаторами.

Как будет показано ниже, существуют различные формы описания взаимозависимости компонент случайного вектора. Наиболее полно она выражается (в прежних обозначениях) системой условных распределений вида

$$P\{X(1) \in B_1 | X(2) \in B_2\} = \frac{P\{X(1) \in B_1, X(2) \in B_2\}}{P\{X(2) \in B_2\}}$$

или условных функций распределения

$$F(x(1) | x(2)) = P\{X(1) < x(1) | X(2) = x(2)\} = \frac{P\{X(1) < x(1), X(2) = x(2)\}}{P\{X(2) = x(2)\}}. \quad (6.18)$$

Для непрерывного случайного вектора, обладающего функцией плотности распределения $f_X(x_1, \dots, x_n) = f_X(x)$, зависимость между его компонентами вполне определяется системой условных плотностей распределения

$$f(x(1) | x(2)) = \frac{f(x(1), x(2))}{f(x(2))} \quad (6.19)$$

(здесь для простоты записи индексы у функций опущены; вид функций определяется обозначением их аргументов).

Фигурирующие в последнем выражении частные плотности распределения подвекторов $X(1)$ и $X(2)$ определяются, как было сказано выше, из плотности распределения вектора X . Так, если $X = \begin{pmatrix} X(1) \\ X(2) \end{pmatrix}$, $f(x(1), x(2)) = f_X(x)$, то $f(x(2)) = \int_{X(1)} f_X(x) dx(1)$.

Следует обратить внимание, что при непрерывном распределении вектора $X(2)$ событие $X(2) = x(2)$ имеет нулевую вероятность и правую часть в (6.18) следует рассматривать как существующий предел при $|\Delta| \rightarrow 0$ отношения вероятностей $\frac{P\{X(1) < x(1), X(2) \in \Delta\}}{P\{X(2) \in \Delta\}}$, где Δ – малая окрестность точки $x(2)$.

Числовые характеристики случайного вектора

Как и для скалярной случайной величины, свойства случайного вектора X с той или иной полнотой задаются его числовыми характеристиками.

Введем их, для лаконичности полагая, что X обладает непрерывным распределением (что обычно имеет место в многомерных стохастических моделях).

Пусть $Y = \psi(X) = \psi(X_1, \dots, X_n)$ – скалярная борелевская функция векторного аргумента X . Распространяя на многомерный случай определение 5.3, для математического ожидания MY примем определение

$$MY = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n \quad (6.20)$$

(как и в скалярном случае, здесь конечное математическое ожидание существует тогда и только тогда, когда этот интеграл сходится абсолютно).

В частности, при $\psi(X_1, \dots, X_n) = X_j$ получим математическое ожидание компоненты X_j :

$$MY = \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} x_j f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{-\infty}^{\infty} x_j f_{X_j}(x_j) dx_j = MX_j \quad (j=1, \dots, n)$$

и при $\psi(X_1, \dots, X_n) = (X_j - MX_j)^2 = \overset{0}{X_j}^2$ – дисперсию компоненты X_j :

$$MY = \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} (x_j - MX_j)^2 f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{-\infty}^{\infty} (x_j - MX_j)^2 f_{X_j}(x_j) dx_j \overset{\Delta}{=} DX_j \quad (j=1, \dots, n).$$

Эти числовые характеристики описывают свойства отдельных компонент вектора X ; теперь, однако, появляются числовые характеристики, содержащие информацию и об их групповых свойствах и связях. Рассмотрим некоторые из них.

Пусть $Y = \psi(X) = \sum_{j=1}^k \psi_j(X)$ где $\psi_j(X)$ – борелевские функции. Нетрудно проверить, что тогда справедливо полезное равенство

$$MY = M(\psi(X)) = M\left(\sum_{j=1}^k \psi_j(X)\right) \quad (6.21)$$

В частности, полагая для $\forall j \psi_j(X) = X_j$, т.е. $Y = \sum_{j=1}^n X_j$, найдем, что

$$MY = M\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n MX_j, \quad (6.22)$$

т.е. математическое ожидание суммы компонент случайного вектора равна сумме их математических ожиданий. Нетрудно убедиться, что такой же вывод можно сделать для любого подвектора вектора X .

Когда компоненты вектора X имеют равные математические ожидания (например, когда они одинаково распределены), (6.22) принимает вид

$$MY = nMX. \quad (6.23)$$

Важно заметить, что (6.21) – (6.23) получены без каких либо условий, налагаемых на зависимость компонент вектора X .

Положим, далее, $Y = \psi(X_1, \dots, X_n) = (X_i - MX_i)(X_j - MX_j) = \overset{0}{X_i} \overset{0}{X_j}$ и назовем *корреляционным моментом (ковариацией)* случайных величин (компонент вектора X) X_i и X_j число

$$\begin{aligned} R_{ij} &= M(\overset{0}{X_i} \overset{0}{X_j}) = \\ &= \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} (x_i - MX_i)(x_j - MX_j) f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - MX_i)(x_j - MX_j) f_X(x_i, x_j) dx_i dx_j \end{aligned} \quad (6.24)$$

(заметим сразу, что $R_{ii} = DX_i$ и $R_{ij} = R_{ji}$).

Из вида (6.24) можно заключить, что корреляционный момент R_{ij} выражает меру стохастической связи между компонентами X_i и X_j , состоящей в согласованности отклонений их значений от математических ожиданий (в одном и том же или в разных направлениях). Для независимых X_i и X_j ($i \neq j$) из (6.24) следует $R_{ij} = 0$, т.е. их *некоррелированность*.

Вместе с тем, из некоррелированности X_i и X_j не следует их независимость, Чтобы убедиться в этом, приведем пример.

Пример 6.1. Двумерный случайный вектор $X = (X_1, X_2)'$ равномерно распределен на круге единичного радиуса с плотностью

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi}, & \text{если } x_1^2 + x_2^2 \leq 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Для условной плотности распределения X_1 получаем

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \begin{cases} \frac{1}{2\sqrt{1-x_2^2}}, & \text{если } -\sqrt{1-x_2^2} \leq x_1 < \sqrt{1-x_2^2}, \\ 0 & \text{в прочих случаях,} \end{cases}$$

т.е. X_1 и X_2 взаимно зависимы. Вместе с тем

$$R_{x_1 x_2} = M(X_1 X_2) = \frac{1}{\pi} \int_{-1}^1 x_1 \left(\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} x_2 dx_2 \right) dx_1 = 0,$$

т.е. X_1 и X_2 некоррелированы.

Отсюда можно заключить, что, анализируя связь между компонентами случайного вектора, следует различать (в порядке доминирования) их независимость в совокупности, попарную независимость и некоррелированность.

Часто вместо корреляционного момента R_{ij} случайных величин X_i и X_j используется их *коэффициент корреляции* r_{ij} , равный

$$r_{ij} = \frac{R_{ij}}{\sqrt{DX_i \cdot DX_j}} = \frac{R_{ij}}{\sqrt{R_{ii} \cdot R_{jj}}}. \quad (6.25)$$

Удобство коэффициента корреляции состоит в том, что сохраняя информацию о связи между случайными величинами, он безразмерен и нормирован единицей:

$$|r_{ij}| = \left| \frac{R_{ij}}{\sqrt{DX_i \cdot DX_j}} \right| \leq 1 \quad (6.26)$$

Последнее соотношение представляет собой *неравенство Коши – Буняковского*, которое читатель легко докажет, раскрыв очевидное неравенство

$$M \left(\frac{\overset{0}{X}_i}{\sqrt{M \overset{0}{X}_i^2}} - \frac{\overset{0}{X}_j}{\sqrt{M \overset{0}{X}_j^2}} \right)^2 \geq 0.$$

Заметим также, что коэффициент корреляции r_{ij} компонент X_i и X_j представляет собой корреляционный момент случайных величин, получаемых в результате их нормировки дисперсиями:

$$r_{ij} = M \left(\frac{\overset{0}{X}_i}{\sqrt{DX_i}} \cdot \frac{\overset{0}{X}_j}{\sqrt{DX_j}} \right),$$

т.е. является характеристикой связи между компонентами, свободной от значений их дисперсий (в отличие от корреляционного момента).

Найдем теперь формулу для вычисления дисперсии суммы компонент случайного вектора \mathbf{X} . Снова полагая

$$Y = \psi(\mathbf{X}) = \sum_{j=1}^n X_j, \text{ получим}$$

$$\begin{aligned} DY &= M(Y - MY)^2 = \\ &= \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} ((x_1 - MX_1) + \dots + (x_n - MX_n))^2 f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \sum_{j=1}^n DX_j + 2 \sum_{\substack{i,j=1 \\ (i < j)}}^n R_{ij}, \end{aligned} \quad (6.27)$$

т.е. дисперсия суммы компонент случайного вектора в общем случае не равна сумме их дисперсий, отличаясь от нее на удвоенную сумму их попарных ковариаций. При некоррелированных (и, тем более, при независимых) компонентах (6.27) принимает вид

$$DY = D \left(\sum_{j=1}^n X_j \right) = \sum_{j=1}^n DX_j \quad (6.28)$$

т.е. дисперсия суммы компонент случайного вектора оказывается равной сумме их дисперсий. Если, к тому же, компоненты вектора \mathbf{X} одинаково распределены (или, хотя бы имеют равные дисперсии), (6.28) принимает вид

$$DY = nDX. \quad (6.29)$$

Смысл и свойства введенного в главе 5 условного математического ожидания скалярной случайной величины просто переносятся на векторный случай, если (по аналогии с (5.19)) определить условное относительно события B математическое ожидание функции $Y = \psi(\mathbf{X})$ равенством

$$M(Y | B) = \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} \psi(x_1, \dots, x_n) f_X(x_1, \dots, x_n | B) dx_1 \dots dx_n \quad (6.30)$$

(напомним, что с целью упрощения записи мы рассматриваем случай непрерывного распределения вектора X ; при дискретном и смешанном распределениях принципиально новых ситуаций не возникает).

В частности, в векторном случае сохраняется формула полного математического ожидания (5.22), принимая при непрерывном распределении вектора X вид

$$\begin{aligned} MY &= \sum_{j=1}^k M(Y | B_j) P(B_j) = \\ &= \sum_{j=1}^k P(B_j) \int_{-\infty}^{\infty} (n) \int_{-\infty}^{\infty} \psi(x_1, \dots, x_n) f_X(x_1, \dots, x_n | B_j) dx_1 \dots dx_n, \quad k \leq \infty, \end{aligned} \quad (6.13)$$

где $\{B_j\}_{j=1}^k$ – полная уруппа несовместимых событий.

Применим (6.30) для вычисления математического ожидания суммы одинаково распределенных компонент случайного вектора X , когда его размерность N сама является целочисленной случайной величиной с распределением вероятностей

$$P\{N = n\} = p(n), \quad n = 1, 2, \dots$$

Полагая в (6.31) $j = n$, $B_n = \{N = n\}$, $Y = \psi(X_1, \dots, X_n) = X_1 + \dots + X_n$, из (6.23) и (6.31) получим

$$MY = \sum_{n=1}^{\infty} P\{N = n\} nMX = MX \cdot MN. \quad (6.32)$$

Обратим внимание, что это равенство получено в предположении отсутствия зависимости между компонентами вектора X и его размерностью N ; в свое время (в курсе по математической статистике) это равенство, именуемое *тождеством Вальда*, будет получено и при существовании такой зависимости.

Соотношение, подобное (6.32), легко выводится и для дисперсии суммы независимых одинаково распределенных компонент вектора X . Для этого используем формулу

$$DY = MY^2 - (MY)^2,$$

где

$$\begin{aligned} MY^2 &= \sum_{n=1}^{\infty} M(Y^2 | n) p(n) = \sum_{n=1}^{\infty} M(X_1 + \dots + X_n)^2 p(n) = \\ &= \sum_{n=1}^{\infty} (nMX^2 + n(n-1)(MX)^2) p(n) = MN \cdot MX^2 + MN^2 \cdot (MX)^2 - MN \cdot (MX)^2, \\ &MY = MN \cdot MX, \end{aligned}$$

т.е.

$$\begin{aligned} DY &= MN \cdot MX^2 + MN^2 \cdot (MX)^2 - MN \cdot (MX)^2 - \\ &- (MN)^2 \cdot (MX)^2 = MN \cdot DX + (MX)^2 \cdot DN. \end{aligned} \quad (6.33)$$

До сих пор, оперируя с условными математическими ожиданиями, мы полагали, что условие задается случайным ненулевым событием B . В некоторых (весьма частых) случаях стохастическая модель содержит условные математические ожидания, в которых условия задаются нулевыми случайными событиями.

Рассмотрим, к примеру, двумерный случайный вектор $X = (X_1, X_2)'$, обладающий непрерывным распределением с плотностью распределения $f(x_1, x_2)$, и пусть задача состоит в вычислении среднего значения его компоненты X_1 при фиксированном $X_2 = x_2$, т.е. условного математического ожидания $M(X_1|B)$ при нулевом событии $B = \{X_2 = x_2\}$. Обозначим и определим $M(X_1|B)$ равенством

$$M(X_1 | B) = \mu_1(x_2) = M(X_1 | x_2) = \int_{-\infty}^{\infty} x_1 dF(x_1 | x_2),$$

которое имеет смысл в предположении существования предела

$$F(x_1 | x_2) = \lim_{|\Delta| \rightarrow 0} \frac{P\{X(1) < x(1), X(2) \in \Delta\}}{P\{X(2) \in \Delta\}},$$

где Δ – малая окрестность точки x_2 (см. §6.1). Заметим, что это условие обычно выполняется.

Условное математическое ожидание, используемое как инструмент выражения зависимости среднего значения одной компоненты случайного вектора от значений других, лежит в основе *регрессионного анализа*, представляющего собой один из основных разделов математической статистики.

Не раскрывая подробно сущность этого метода стохастического анализа (см., например [5]), коснемся его простых исходных положений.

Пусть снова $X = (X_1, X_2)'$. Используя терминологию регрессионного анализа, назовем функцию $\mu_1(x_2) = M(X_1 | x_2)$ *функцией регрессии* X_1 на X_2 . Можно показать (см.[5]), что среди всех функций $v(x_2)$, претендующих на выражении зависимости от значений x_2 значений компоненты X_1 , функция $\mu_1(x_2)$ обладает тем свойством, что среднее квадратическое отклонение от нее значений X_1 минимально. Это и объясняет широкое применение ее (при определенном обобщении) в регрессионном анализе.

Значение функции регрессии $\mu_1(x_2)$ может, однако, применяться и как еще одна характеристика взаимозависимости компонент X_1 и X_2 : при $\mu_1(x_2) \neq MX_1$ компоненты следует признать зависимыми. В противном случае назовем их *регрессионно независимыми*.

Для всех рассмотренных нами типов зависимости между компонентами случайного вектора можно в итоге установить следующую импликационную цепочку:

независимость в совокупности \Rightarrow попарная независимость \Rightarrow регрессионная независимость \Rightarrow некоррелированность (обратные импликации везде отсутствуют).

Свойства случайного вектора удобно описывать свойствами его числовых характеристик, представленных в векторно-матричной форме. С этой целью введем ряд определений.

Определение 6.3. *Определенной на вероятностном пространстве $\langle \Omega, F, P \rangle$ случайной матрицей $Z = z(\omega)$ называется матрица $Z = (Z_{ij})$, элементами которой являются определенные на этом пространстве скалярные случайные величины $Z_{ij} = z_{ij}(\omega)$.*

Определение 6.4. *Математическим ожиданием MZ случайной матрицы Z называется матрица, элементами которой являются математические ожидания соответствующих элементов случайной матрицы Z*

$$MZ = (MZ_{ij}).$$

Пусть

$$Z = AYB + C, \quad (6.33)$$

где Y – случайная матрица с математическим ожиданием MY , A , B и C – неслучайные матрицы. Используя свойства математического ожидания, непосредственной проверкой легко установить справедливость равенства

$$MZ = A(MY)B + C. \quad (6.34)$$

Введем понятие вектора математического ожидания случайного вектора:

Определение 6.5. Вектором математического ожидания MX случайного вектора X называется вектор, компоненты которого суть математические ожидания компонент вектора X :

$$MX = m_X = \begin{pmatrix} MX_1 \\ \cdot \\ \cdot \\ \cdot \\ MX_n \end{pmatrix}.$$

Обозначим $\overset{0}{X} = X - MX$ центрированное значение случайного вектора X и образуем случайную матрицу $Z = \overset{0}{X} \overset{0}{X}'$. Нетрудно проверить, что математическое ожидание этой матрицы MZ есть матрица, элементами которой являются попарные корреляционные моменты компонент вектора X :

$$\begin{aligned}
MZ &= \begin{pmatrix} 0 & 0 \\ X & X' \end{pmatrix} = \begin{pmatrix} M \begin{pmatrix} 0 & 0 \\ X_i & X_j \end{pmatrix} \end{pmatrix} = \\
&= (R_{ij}), \quad R_{ij} = M((X_i - MX_i)(X_j - MX_j))
\end{aligned}$$

Определение 6.6. Матрица R_X , элементы которой представляют собой корреляционные моменты компонент случайного вектора X , называется корреляционной (ковариационной) матрицей этого вектора.

Корреляционная матрица обладает приведенными ниже свойствами.

1. R_X – симметричная матрица, так как $R_{ij} = R_{ji}$.

2. Для $\forall_i, R_{ii} \geq 0$, т.к. $R_{ii} = M \overset{0}{X_i^2} = DX_i$, где DX_i – дисперсия компоненты X_i .

3. Элементы корреляционной матрицы удовлетворяют неравенству $|R_{ij}| \leq \sqrt{R_{ii}R_{jj}}$ (см. (6.25) и (6.26)).

4. Корреляционная матрица неотрицательно определена, т.е. для любого неслучайного вещественного вектора a справедливо неравенство $a'R_X a \geq 0$. Этот факт вытекает из следующих соотношений:

$$a'R_X a = a'M \begin{pmatrix} 0 & 0 \\ X & X' \end{pmatrix} a = M(a' \overset{0}{X} \overset{0}{X}' a) = M(a' \overset{0}{X})^2 \geq 0.$$

Можно показать, что в силу симметричности корреляционной матрицы последнее её свойство связано с знаком её определителя $\det R_X$: корреляционная матрица положительно определена, т.е. для $\forall a \neq 0 \quad a'R_X a > 0$, тогда и только тогда, когда $\det R_X > 0$, и лишь неотрицательно определена (т.е. для $\forall a \neq 0 \quad a'R_X a \geq 0$ и $\exists a \neq 0: a'R_X a = 0$), если $\det R_X = 0$.

Если корреляционная матрица только неотрицательно определена (не являясь положительно определенной), из равенства $a'R_X a = 0$ (для некоторого $a \neq 0$) следуют импликации

$$\begin{aligned}
a'R_X a = 0 &\Rightarrow a'M \begin{pmatrix} 0 & 0 \\ X & X' \end{pmatrix} a = 0 \Rightarrow M(a' \overset{0}{X})^2 = D(a' X) = 0 \Rightarrow \\
&\Rightarrow P\{a'X = a'm_X\} = 1 \quad (\text{см. §5.3}),
\end{aligned}$$

что означает существование с вероятностью 1 линейной собственной зависимости между компонентами вектора X (см. (6.9)), свидетельствующие о равенстве нулю дисперсии случайной величины $a'X$. Из полной теории известно, что равенство нулю дисперсии любой случайной величины Y следует ее равенство с вероятностью 1 константе MY : $DY = 0 \Rightarrow P\{Y = MY\} = 1$ (в этом свойстве дисперсии читатель может убедиться для частного случая дискретного распределения случайной величины Y при конечном множестве ее значений). В нашем случае имеем $P\{a'X = a'm_X\} = 1$, что означает существование с вероятностью 1 линейной собственной зависимости между компонентами вектора X (см. (6.9)).

Наряду с корреляционной матрицей часто используется *нормированная корреляционная матрица* $\mathbf{r}=(r_{ij})$, элементами которой являются коэффициенты корреляции компонент вектора \mathbf{X} (см.(6.25)).

Корреляционная матрица содержит информацию о корреляционных связях между компонентами случайного вектора. Когда такая связь для каждой пары компонент отсутствует, корреляционная матрица имеет диагональный вид:

$$R_X = \begin{pmatrix} DX_1 & 0 & \dots & 0 \\ 0 & DX_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & DX_n \end{pmatrix}$$

Напомним, что из независимости двух случайных величин следует их некоррелированность, но их некоррелированность не влечет за собой их независимости. Поэтому в общем случае диагональный вид матрицы R_X не гарантирует попарной независимости компонент вектора (и, тем более, их независимости в совокупности).

Преобразование случайного вектора

Пусть заданы случайный вектор $\mathbf{X} = (X_1, \dots, X_n)'$ с известной плотностью распределения $f_X(x_1, \dots, x_n)$ и вектор $\mathbf{Y} = (Y_1, \dots, Y_n)'$, значения компонент которого связаны с компонентами вектора \mathbf{X} взаимно однозначным непрерывно дифференцируемым соответствием

$$y_i = v_i(x_1, \dots, x_n) \quad (i = \overline{1, n})$$

При неравном нулю якобиане $J\left(\frac{\mathbf{X}}{\mathbf{Y}}\right) = \left(\frac{\partial x_i}{\partial y_j}\right)$. Задача состоит в нахождении распределения вектора \mathbf{Y} .

Обозначим B_X и B_Y борелевские σ -алгебры в пространствах X и Y значений векторов \mathbf{X} и \mathbf{Y} .

Из математического анализа известно (см. теорему о замене переменных в кратном интеграле), что при указанном соответствии между векторами для каждой положительной интегрируемой функции $f(x_1, \dots, x_n)$ справедливо равенство

$$\forall B_x \in B_x : \int_{B_x} f(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{B_y} f(v_1^{-1}(y_1, \dots, y_n), \dots, v_n^{-1}(y_1, \dots, y_n)) \left| J\left(\frac{\mathbf{X}}{\mathbf{Y}}\right) \right| dy_1 \dots dy_n,$$

где $v_i^{-1}(\cdot)$ – функция, обратная $v_i(\cdot)$, B_x – прообраз B_y , $B_y \in B_Y$.

Полагая $f(x_1, \dots, x_n) = f_X(x_1, \dots, x_n)$, получаем равенство

$$\begin{aligned}
P\{X \in B_x \in \mathbf{B}_X\} &= \int_{B_x} f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \\
&= \int_{B_y} f_X(v_1^{-1}(y_1, \dots, y_n), \dots, v_n^{-1}(y_1, \dots, y_n)) \left| \mathbf{J} \left(\frac{X}{Y} \right) \right| dy_1 \dots dy_n = P\{Y \in B_y \in \mathbf{B}_Y\},
\end{aligned}$$

справедливое для $\forall B_y \in \mathbf{B}_Y$ и поэтому определяющее подынтегральную функцию в правом интеграле как плотность распределения случайного вектора Y . Поставленная задача решается, следовательно, равенством

$$\begin{aligned}
f_Y(y_1, \dots, y_n) &= \\
&= f_X(v_1^{-1}(y_1, \dots, y_n), \dots, v_n^{-1}(y_1, \dots, y_n)) \left| \mathbf{J} \left(\frac{X}{Y} \right) \right|.
\end{aligned} \tag{6.35}$$

Ясно, что задача имеет решение и в том случае, когда размерность вектора Y ниже размерности вектора X : при этом размерность вектора Y доводится до размерности вектора X путем включением в него «лишних» компонент (приравнивая их, например, соответствующим компонентам вектора X) с последующим интегрированием полученной плотности распределения по этим компонентам.

При вычислении числовых характеристик случайного вектора $Y = Y(X)$ можно использовать полученное распределение вектора Y или, как и в скалярном случае, распределение вектора X (см. (5.11)).

Конкретные распределения случайного вектора

В этом параграфе мы рассмотрим два типа распределений случайного вектора – дискретный вектор с полиномиальным распределением и непрерывный вектор с нормальным распределением, связанные между собой предельными соотношениями.

6.4.1. Полиномиальное распределение случайного вектора

Рассмотрим следующую вероятностную модель. Пусть производится n опытов, в каждом из которых происходит одно из k несовместимых событий, образующих полную группу $\{A_j\}_{j=1}^k$, причем появление каждого из них имеет известную вероятность

$$P(A_j) = p_j, \quad \sum_{j=1}^k p_j = 1. \text{ Предполагается, что результаты опытов независимы в совокупности.}$$

Обозначим $X_j, \quad j = \overline{1, k}$, случайную величину, равную числу опытов, в которых произошло событие A_j , и найдем распределение случайного вектора $X = (X_1, \dots, X_k)'$ при

$$\sum_{j=1}^k X_j = n \text{ (последнее равенство означает вырожденность вектора } X, \text{ см. (6.9)).}$$

При указанных предположениях каждая реализация этого вектора $\mathbf{x} = (x_1, \dots, x_k)'$ осуществляется с вероятностью $p_1^{x_1} \dots p_k^{x_k}$, не зависящей варианта последовательности (очередности) появления событий. Число этих равновероятных вариантов при фиксированной реализации равно

$$N(x_1, \dots, x_k) = C_n^{x_1} C_{n-x_1}^{x_2} \dots 1 = \frac{n!}{x_1! \dots x_k!},$$

откуда получаем

$$P\{\mathbf{X} = (x_1, \dots, x_k)'\} = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad (6.36)$$

или, ввиду вырожденности вектора \mathbf{X} ,

$$P\{\mathbf{X} = (x_1, \dots, x_k)'\} = P\{\mathbf{X} = (x_1, \dots, x_{k-1})'\} = \frac{n!}{x_1! \dots (n - \sum_{j=1}^{k-1} x_j)!} p_1^{x_1} \dots p_{k-1}^{x_{k-1}} (1 - \sum_{j=1}^{k-1} p_j)^{(n - \sum_{j=1}^{k-1} x_j)}.$$

которое будем называть *полиномиальным* и обозначать $P(n, \mathbf{p})$, где $\mathbf{p} = (p_1, \dots, p_k)'$. Название этого распределения следует из того очевидного факта, что (6.36) представляет собой элемент разложения степени полинома $(p_1 + \dots + p_k)^n$.

Легко видеть, что при $k = 2$ полиномиальное распределение превращается в биномиальное.

Читатель легко найдет, что вектор математического ожидания и корреляционная матрица для полиномиального распределения равны

$$M\mathbf{X} = (np_1, \dots, np_k)', \quad \mathbf{R}_X = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_1p_2 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_1p_k & -np_2p_k & \dots & np_k(1-p_k) \end{pmatrix}.$$

Полиномиальное распределение играет, в частности, важную роль допредельного распределения для многомерного нормального распределения и распределения χ^2 в задачах математической статистики (см. [5]).

6.4.2. Нормальный случайный вектор

В этом параграфе рассматривается частный, но широко используемый в приложениях тип распределения случайного вектора — нормальное (гауссовское) распределение, являющееся естественным обобщением на векторный случай нормального распределения скалярной случайной величины (см. п.4.3.6).

Определение 6.7. *Невырожденным нормальным случайным вектором $\mathbf{X} = (X_1, \dots, X_n)'$ называется случайный вектор, обладающий плотностью распределения вероятностей вида:*

$$f_X(x_1, \dots, x_n) = f_X(\mathbf{x}) = k e^{-\frac{(\mathbf{x}-\mathbf{b})' \mathbf{A} (\mathbf{x}-\mathbf{b})}{2}},$$

где \mathbf{A} – симметричная положительно определенная матрица, k – коэффициент, определяемый условием нормировки, \mathbf{b} – неслучайный вектор.

Раскроем смысл параметров \mathbf{A} , k и \mathbf{b} . Известно, что для каждой симметричной матрицы \mathbf{A} существует ортогональная матрица \mathbf{C} , такая, что

$$\mathbf{C}' \mathbf{A} \mathbf{C} = \mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & d_n \end{pmatrix}, \quad (6.37)$$

где, в силу положительной определенности матрицы \mathbf{A} , для $\forall j \ d_j > 0$.

Образую матрицу $\mathbf{B} = \mathbf{C} \mathbf{D}^{-\frac{1}{2}}$, где

$$\mathbf{D}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{d_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{d_2}} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{d_n}} \end{pmatrix},$$

и введем вектор \mathbf{Z} , удовлетворяющий преобразованию $\mathbf{X} - \mathbf{b} = \mathbf{B} \mathbf{Z}$, при котором

$$(\mathbf{x} - \mathbf{b})' \mathbf{A} (\mathbf{x} - \mathbf{b}) = \mathbf{Z}' \mathbf{B}' \mathbf{A} \mathbf{B} \mathbf{Z} = \mathbf{Z}' \mathbf{D}^{-\frac{1}{2}} \mathbf{C}' \mathbf{A} \mathbf{C} \mathbf{D}^{-\frac{1}{2}} \mathbf{Z} = \mathbf{Z}' \mathbf{Z},$$

якобиан этого преобразования равен

$$\mathbf{J} \left(\frac{\mathbf{X}}{\mathbf{Z}} \right) = \left| \frac{dx_i}{dz_j} \right| = \det \mathbf{B}.$$

Отсюда, в виду (6.37),

$$f_Z(\mathbf{z}) = k \det \mathbf{B} e^{-\frac{\mathbf{z}' \mathbf{z}}{2}} = k \det \mathbf{B} e^{-\frac{\sum_{i=1}^n z_i^2}{2}} \quad (6.38)$$

и, используя интеграл Пуассона: $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$, найдем (из условия нормировки функции

плотности распределения $f_Z(\mathbf{z})$ вектора \mathbf{Z})

$$\int_{-\infty}^{\infty} f_Z(\mathbf{z}) d\mathbf{z} = (2\pi)^{\frac{n}{2}} k \det \mathbf{B} = 1, \text{ т.е. } k = \frac{1}{(2\pi)^{\frac{n}{2}} \det \mathbf{B}}.$$

Из (6.38) видно, что вектор \mathbf{Z} образован n независимыми компонентами, каждая из которых имеет стандартное нормальное распределение $Z_i \in N(0, 1)$, т.е.

$$f_Z(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{j=1}^n z_j^2}{2}}$$

($MZ = \mathbf{m}_Z = \mathbf{0}$, $\mathbf{R}_Z = \mathbf{I}$ – единичная матрица).

Далее получим

$$M(\mathbf{X} - \mathbf{b}) = M(\mathbf{BZ}) = \mathbf{0}, \text{ т.е.}$$

$$\mathbf{b} = M\mathbf{X} = \mathbf{m}_X, \quad \mathbf{R}_X = M(\overset{0}{\mathbf{X}} \overset{0}{\mathbf{X}}') = M\mathbf{BZZ}'\mathbf{B}' = \mathbf{B}\mathbf{B}'.$$

С другой стороны справедливо равенство:

$$\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{D}^{-\frac{1}{2}}\mathbf{C}'\mathbf{A}\mathbf{C}\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I}, \text{ откуда}$$

$$\mathbf{A} = (\mathbf{B}\mathbf{B}')^{-1} = \mathbf{R}_X^{-1}.$$

В результате получаем $\det \mathbf{B} = \sqrt{\det \mathbf{R}_X}$ и

$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{R}_X}} e^{-\frac{(\mathbf{x} - \mathbf{m}_X)' \mathbf{R}_X^{-1} (\mathbf{x} - \mathbf{m}_X)}{2}} \quad (6.39)$$

Из (6.39) следует, что нормальное распределение случайного вектора полностью определяется вектором математического ожидания \mathbf{m}_X и корреляционной матрицей \mathbf{R}_X . Для $\forall c = \text{const}$ уравнение $f_X(\mathbf{x}) = c$ описывает в пространстве X эллипсоид с центром в \mathbf{m}_X .

Тот факт, что вектор \mathbf{X} имеет нормальное распределение с параметрами \mathbf{m}_X и \mathbf{R}_X , мы будем выражать записью

$$\mathbf{X} \in \square(\mathbf{m}_X, \mathbf{R}_X).$$

Осуществляя преобразование случайного вектора \mathbf{X} с помощью ортогональной матрицы \mathbf{C} :

$$\mathbf{X} - \mathbf{m}_X = \mathbf{C}\mathbf{Y},$$

при котором

$$\mathbf{C}'\mathbf{R}_X^{-1}\mathbf{C} = \mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & d_n \end{pmatrix}$$

(ср. (6.37)), мы получаем каноническую форму многомерного нормального распределения:

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{j=1}^n \sigma_j} e^{-\frac{\mathbf{y}'\mathbf{D}\mathbf{y}}{2}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi y_j}} e^{-\frac{y_j^2}{2y_j^2}} = \prod_{j=1}^n f_j(y_j), \quad (6.40)$$

где $f_j(y_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{y_j^2}{2\sigma_j^2}}$ – частная плотность распределения компоненты Y_j вектора Y ,
 $\sigma_j = 1/\sqrt{d_j}$.

Из (6.39) следует, что некоррелированность компонент нормального случайного вектора влечет за собой их независимость в совокупности (и, следовательно, попарную независимость). Переход от вектора X к вектору Y сопровождается, следовательно, исчезновением зависимости между компонентами. Это, однако, не следует понимать как потерю зависимости между реальными случайными величинами, описывающими стохастическую модель: хотя в новой «декоррелированной» форме зависимость между компонентами вектора Y отсутствует, она сохраняется между их линейными комбинациями, выражающими компоненты исходного вектора X .

Следует заметить, что процедура декорреляции компонент случайного вектора в указанном смысле возможна не только при его нормальном распределении. Её значение для вектора с нормальным распределением объясняется именно тем, что в этом случае она сопровождается исчезновением вероятностной зависимости между компонентами.

Пусть $X(1)$ и $X(2)$ непересекающиеся подвекторы нормального случайного вектора X , причем компоненты, входящие в эти подвекторы, взаимно не коррелированы. Тогда векторы $X(1)$ и $X(2)$ независимы.

Действительно, представим X в виде

$$X = \begin{pmatrix} X(1) \\ X(2) \end{pmatrix}, \quad (6.41)$$

где $X(1)$ и $X(2)$ имеют, соответственно, размерности r и t (для простоты, не умаляя общности, рассмотрим случай $r + t = n$).

Тогда корреляционная матрица R_X вектора X представляется в виде блочной матрицы

$$R_X = \begin{pmatrix} R(1) & \vdots & R(12) \\ \dots\dots\dots & \ddots & \dots\dots\dots \\ R(21) & \vdots & R(2) \end{pmatrix} = \begin{pmatrix} R(1) & \vdots & 0 \\ \dots\dots\dots & \ddots & \dots\dots\dots \\ 0 & \vdots & R(2) \end{pmatrix},$$

где $R(1)$ и $R(2)$ – корреляционные матрицы векторов $X(1)$ и $X(2)$, а $R(12)$ и $R(21)$ – прямоугольные матрицы, содержащие корреляционные моменты компонент, относящихся к разным подвекторам, равные (по условию) нулю. Используя правила обращения с блочными матрицами, нетрудно получить:

$$\begin{aligned} f_X(\vec{x}) &= \frac{1}{(2\pi)^{\frac{r}{2}} (2\pi)^{\frac{s}{2}} (\det R(1) \det R(2))^{\frac{1}{2}}} e^{-\frac{1}{2}[(x(1)-m(1))' R(1)^{-1} (x(1)-m(1)) + (x(2)-m(2))' R(2)^{-1} ((x(2)-m(2)))]} = \\ &= f_{X(1)}(x(1)) f_{X(2)}(x(2)), \quad m(1) = MX(1), \quad m(2) = MX(2), \end{aligned}$$

что свидетельствует о независимости векторов $X(1)$ и $X(2)$.

Ниже будет показано, что любое линейное преобразование нормального случайного вектора приводит к нормальному случайному вектору. Покажем это пока для случая неособенной матрицы преобразования.

Пусть $X \in N(m_X, R_X)$ и $U = FX$, где F – квадратная неособенная матрица. Имеют место очевидные равенства:

$$(x - m_X)' R_X^{-1} (x - m_X) = (u - Fm_X)' (F^{-1})' R_X^{-1} F^{-1} (u - Fm_X) = (u - Fm_X)' (FR_X F')^{-1} (u - Fm_X),$$

$$J\left(\frac{X}{U}\right) = \left\| \frac{\partial x_i}{\partial u_j} \right\| = F^{-1},$$

$$f_U(u) = \frac{1}{(2\pi)^2 (\det FR_X F')^2} e^{-\frac{(u - Fm_X)' (FR_X F')^{-1} (u - Fm_X)}{2}},$$

т.е. $U \in N(Fm_X, FR_X F')$.

Перейдем теперь к вопросу о частном распределении группы компонент (подвектора) нормального случайного вектора. Пусть $X \in N(m_X, R_X)$, $X = \begin{pmatrix} X(1) \\ X(2) \end{pmatrix}$ (см. (6.41)) и мы ищем частное распределение группы компонент, образующих вектор $X(2)$. Преобразуем X с помощью неособенной матрицы S :

$$S = \begin{pmatrix} I_1 & \vdots & H \\ \cdots & \ddots & \cdots \\ 0 & \vdots & I_2 \end{pmatrix},$$

где I_1 и I_2 – единичные матрицы размерности $r \times r$ и $t \times t$ соответственно, H – пока произвольная матрица. Получим

$$Y = SX = \begin{pmatrix} Y(1) \\ Y(2) \end{pmatrix}, \quad Y(1) = X(1) + HX(2), \quad Y(2) = X(2) \text{ и}$$

$$Y \in N(S\bar{m}_X, SR_X S'), \quad R_Y = SR_X S' = S \begin{pmatrix} R(1) & \vdots & R(12) \\ \cdots & \ddots & \cdots \\ R(21) & \vdots & R(2) \end{pmatrix} S' =$$

$$= \begin{pmatrix} R(1) + HR(21) + R(12)H' + HR(2)H' & \vdots & R(12) + HR(2) \\ \cdots & \ddots & \cdots \\ R(21) + R(2)H' & \vdots & R(2) \end{pmatrix}.$$

Конкретизируем теперь вид матрицы H , положив $R(12) + HR(2) = 0$ (при этом $R(21) + R(2)H' = 0$, так как $R(21) = R'(12)$).

Тогда $H = -R(12)R^{-1}(2)$ и

$$\begin{aligned} \mathbf{R}_Y &= \begin{pmatrix} \mathbf{R}(1) - \mathbf{R}(12)\mathbf{R}^{-1}(2)\mathbf{R}(21) & \vdots & \mathbf{0} \\ \cdots\cdots\cdots & \ddots & \cdots\cdots\cdots \\ \mathbf{0} & \vdots & \mathbf{R}(2) \end{pmatrix}; \\ m_Y &= \begin{pmatrix} \mathbf{I}_1 & \vdots & -\mathbf{R}(12)\mathbf{R}^{-1}(2) \\ \cdots\cdots\cdots & \ddots & \cdots\cdots\cdots \\ \mathbf{0} & \vdots & \mathbf{I}_2 \end{pmatrix} m_X = \begin{pmatrix} m_X(1) - \mathbf{R}(12)\mathbf{R}^{-1}(2)m_X(2) \\ m(2) \end{pmatrix}. \end{aligned}$$

Отсюда $Y(2) = X(2) \in N(\mathbf{m}_X(2), \mathbf{R}(2))$ т.е. частное распределение группы компонент, образующих подвектор $X(2)$ имеет нормальное распределение, вектор математического ожидания и корреляционная матрица которого образуется простым отбором элементов вектора \mathbf{m}_X и корреляционной матрицы \mathbf{R}_X , относящихся к компонентам данной группы. Ясно, что такое правило образования частного распределения распространяется на любую совокупность компонент нормального вектора.

Найдем условную плотность распределения компонент нормального случайного вектора, т.е., например, плотность $f(x(1)|x(2))$. Для этого воспользуемся тем же преобразованием $Y = SX$.

Из независимости $Y(1)$ и $Y(2)$ (при выбранной матрице \mathbf{H}), имеем:

$$f(\mathbf{y}(1) | \mathbf{y}(2)) = f(\mathbf{y}(1) | \mathbf{x}(2)) = f(\mathbf{y}(1)) .$$

Ho

$$\begin{aligned} Y(1) &\in \mathbf{N}(m_x(1) - \mathbf{R}(12)\mathbf{R}^{-1}(2)m_x(2), \\ &\quad \mathbf{R}(1) - \mathbf{R}(12)\mathbf{R}^{-1}(2)\mathbf{R}(21)), \\ X(1) &= Y(1) - \mathbf{H}X(2) \end{aligned}$$

и при $X(2) = \mathbf{x}(2)$ получаем $X(1) = Y(1) - \mathbf{H}\mathbf{x}(2)$, т.е. вектор $X(1)$ имеет нормальное условное распределение с параметрами

$$\begin{aligned} \mathbf{m}_{1|2}^{\Delta} &= M[X(1) | \mathbf{x}(2)] = \mathbf{m}_Y(1) - \mathbf{H}\mathbf{x}(2) = m_X(1) + \mathbf{R}(12)\mathbf{R}^{-1}(2) (\mathbf{x}(2) - \mathbf{m}_X(2)), \\ \mathbf{R}_{1|2}^{\Delta} &= \mathbf{R}(1) - \mathbf{R}(12)\mathbf{R}^{-1}(2)\mathbf{R}(21). \end{aligned}$$

Заметим, что функция

$$\mathbf{m}_{1|2} = \mathbf{m}_{1|2}(\mathbf{x}(2)) = M[X(1) | \mathbf{x}(2)] = \mathbf{m}_X(1) + \mathbf{R}(12)\mathbf{R}^{-1}(2)(\mathbf{x}(2) - \mathbf{m}_X(2)),$$

называемая *функцией регрессии вектора X_1 на вектор X_2* , представляет собой обобщение на случай $n > 2$ введенной выше функции регрессии $\mu_1(x_2)$ для $n = 2$.

Выше было показано, что линейное невырожденное преобразование нормального вектора при сохранении его размерности приводит снова к нормально распределенному вектору. Покажем, что при линейном преобразовании нормального вектора с снижением его размерности также сохраняется нормальное распределение.

Пусть $X \in N(m_X, R_X)$ и $Y = SX$, где S – матрица, имеющая m строк, n столбцов ($m < n$) и ранг m . Всегда можно найти матрицу A , дополняющую матрицу S до квадратной неособенной матрицы B размера $n \times n$ с рангом n :

$$B = \begin{pmatrix} S \\ A \end{pmatrix}.$$

Тогда $Z = BX \in N(m_Z, R_Z)$, где $m_Z = Bm_X$, $R_Z = BR_X B'$. Вектор Y является подвектором вектора Z и, как было выше установлено, также имеет нормальное распределение, причем, как нетрудно видеть, $m_Y = Sm_X$, $R_Y = SR_X S'$.

Пусть, в частности, $S = (1, \dots, 1)$; тогда

$$Y = SX = \sum_{i=1}^n X_i \in N\left(\sum_{i=1}^n m_i, \sum_{i,j=1}^n R_{ij}\right) \left(m_i = MX_i, R_{ij} = M\begin{pmatrix} 0 & 0 \\ X_i & X_j \end{pmatrix}\right),$$

т.е. сумма компонент нормального случайного вектора имеет нормальное распределение.

Следует подчеркнуть, что это утверждение получено в предположении нормальности распределения вектора, т.е. совместного распределения его компонент. В общем случае нормальность частных распределений компонент случайного вектора не означает нормальности их совместного распределения.

Так, например, двумерный случайный вектор с плотностью распределения

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi\sigma^2} e^{-\frac{(x_1^2 + x_2^2)}{2\sigma^2}}, & \text{если } x_1 \cdot x_2 \geq 0; \\ 0 & , \text{ если } x_1 \cdot x_2 < 0, \end{cases}$$

имеет нормально распределенные компоненты: $X_1 \in N(0, \sigma^2)$, $X_2 \in N(0, \sigma^2)$, но не является нормальным и распределение суммы его компонент не нормально. Ситуация упрощается при независимости компонент, поскольку вектор, образованный независимыми нормально распределенными компонентами, всегда нормален, и потому сумма независимых нормально распределенных случайных величин имеет нормальное распределение.

До сих пор мы рассматривали нормальный случайный вектор с неособенной корреляционной матрицей, что давало нам возможность определять этот вид распределения видом функции плотности распределения (см. определение 6.7). В следующей главе, используя аппарат характеристических функций, мы дадим более широкое определение класса нормальных распределений, что позволит расширить и класс допустимых преобразований, сохраняющих этот тип распределения случайного вектора. Забегая вперед, сформулируем этот результат в виде следующего утверждения.

Утверждение 6.1. *Любое линейное преобразование с матрицей S нормального случайного вектора $X \in N(m_X, R_X)$ приводит к нормальному случайному вектору*

$$Y = SX \in N(Sm_X, SR_X S'). \quad (6.42)$$

Свертка распределений. Устойчивость распределений относительно суммирования

При построении стохастических моделей нередко возникает необходимость в получении распределения суммы независимых случайных величин, определенных на общем вероятностном пространстве.

В предыдущем параграфе мы установили, что сумма независимых нормально распределенных случайных величин представляет собой снова нормальную случайную величину с параметрами, вполне определенными параметрами распределения слагаемых величин. Такая «устойчивость» нормального распределения проявляется и при сложении независимых случайных величин, если они имеют совместное нормальное распределение, т.е. образуют нормальный случайный вектор.

Далее, однако, решая общую задачу вычисления распределения суммы случайных величин, мы будем постоянно полагать их независимыми. Основное внимание в этом параграфе уделяется сумме двух случайных величин; размерность задачи может повышаться методом индукции.

Итак, пусть (X, Y) — две независимые случайные величины с известными распределениями. Задача состоит в нахождении распределения их суммы $Z = X + Y$, которое носит название *свертки распределений* случайных величин X и Y .

Для дискретных распределений, задаваемых вероятностями

$$p_X(x) = P\{X = x\}, x \in X, \quad p_Y(y) = P\{Y = y\}, y \in Y, \quad (6.43)$$

(X и Y — конечные или счетные множества значений X и Y), используя формулу полной вероятности и независимость X и Y , получаем

$$\begin{aligned} p_Z(z) &= P\{Z = z\} = P\{X + Y = z\} = \\ &= \sum_Y P\{X = z - Y \mid Y = y\} P\{Y = y\} = \\ &= \sum_Y P\{X = z - y\} P\{Y = y\} = \sum_Y p_X(z - y) p_Y(y) \end{aligned} \quad (6.44)$$

(или «симметричное» равенство

$$p_Z(z) = \sum_X p_Y(z - x) p_X(x),$$

если X и Y поменять местами).

Пример 6.2. Пусть $X \in \text{Po}(a_1)$ и $Y \in \text{Po}(a_2)$ (см.(4.11)). Для $Z = X + Y$ из (6.44) получаем

$$p_Z(z) = \sum_Y p_X(z - y) p_Y(y) = \sum_{y=0}^z \frac{a_1^{z-y} a_2^y}{(z-y)! y!} e^{-(a_1+a_2)} = \frac{(a_1 + a_2)^z}{z!} e^{-(a_1+a_2)}, \quad z = 0, 1, \dots,$$

т.е. $Z \in \text{Po}(a_1+a_2)$. Полученный результат означает, что распределение Пуассона обладает устойчивостью в том смысле, что сумма двух (а по индукции, следовательно, и более двух) независимых случайных величин, имеющих распределение Пуассона, сохраняет это распределение, причем при суммировании их параметры складываются. Такое свойство пуассоновского распределения, которое будем называть *устойчивостью относительно суммирования*, имеет важное практическое значение при построении стохастических моделей, в которых фигурируют случайные точечные процессы. Так, например, если поток спроса на некоторый товар представляет собой сумму независимых пуассоновских потоков с различной интенсивностью, то суммарный поток спроса сохраняет пуассоновское распределение с интенсивностью, равной сумме интенсивностей отдельных потоков.

Пусть теперь независимые случайные величины X и Y имеют абсолютно непрерывное распределение с плотностями $f_X(x)$ и $f_Y(y)$. Тогда, используя формулу полной вероятности в форме (4.35), получим

$$F_Z(z) = P\{Z < z\} = \int_Y P\{X + Y < z \mid y\} f_Y(y) dy = \int_Y F_X(z - y) f_Y(y) dy$$

и, переходя к плотностям распределений,

$$f_Z(z) = \int_Y f_X(z - y) f_Y(y) dy \quad (6.45)$$

(или, меняя местами X и Y , эквивалентное равенство

$$f_Z(z) = \int_X f_Y(z - x) f_X(x) dx).$$

Пример 6.3. Пусть случайные величины X и Y имеют показательное распределение (4.15) с плотностями $f_X(x) = \lambda_1 e^{-\lambda_1 x}$, $x \geq 0$ и $f_Y(y) = \lambda_2 e^{-\lambda_2 y}$, $y \geq 0$. Для $Z = X + Y$ получим

$$f_Z(z) = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (e^{-\lambda_2 z} - e^{-\lambda_1 z}), \quad z \geq 0,$$

если $\lambda_1 \neq \lambda_2$, или

$$f_Z(z) = \lambda^2 z e^{-\lambda z}, \quad z \geq 0, \quad (6.46)$$

если $\lambda_1 = \lambda_2 = \lambda$.

Полученные распределения не являются показательными, т.е. показательное распределение не обладает устойчивостью относительно суммирования. Однако при $\lambda_1 = \lambda_2 = \lambda$ это распределение является частным случаем гамма-распределения (конкретнее – распределения Эрланга (4.17) при $n = 2$), которое, как мы убедимся в следующем примере, этим свойством частично обладает.

Пример 6.4. Найдем свертку распределений независимых случайных величин X и Y , имеющих распределения Эрланга (4.17) с одинаковым значением параметра λ и с

несовпадающими (в общем случае) значениями параметра n : $X \in \text{Er}(n_1, \lambda)$, $Y \in \text{Er}(n_2, \lambda)$. Из (6.45) и (4.17) для плотности распределения суммы $Z = X + Y$ получаем

$$f_Z(z) = \frac{\lambda^{n_1+n_2} e^{-\lambda z}}{(n_1-1)!(n_2-1)!} \int_0^z (z-y)^{n_1-1} y^{n_2-1} dy = \frac{\lambda^{n_1+n_2} z^{n_1+n_2-1} e^{-\lambda z}}{(n_1+n_2-1)!} \quad (6.47)$$

и $Z \in \text{Er}(n_1+n_2, \lambda)$, т.е. распределение Эрланга обладает устойчивостью относительно суммирования при варьируемом n и фиксированном λ . Отсюда, ввиду (6.47) и используя метод индукции, легко заключить, что распределение Эрланга $\text{Er}(n, \lambda)$ представляет собой свертку n показательных распределений $\text{Exp}(\lambda)$ или (в практическом смысле) распределение интервала времени между событиями в точечном процессе, возникающим из простого процесса Пуассона в результате удаления из него подряд каждой группы из $n-1$ событий.

В заключение укажем, что свертку распределений часто выражают символом \oplus , представляя распределение F , получаемое в результате свертки распределений F_1 и F_2 , равенством

$$F = F_1 \oplus F_2.$$

Характеристические и производящие функции

Многие теоретические и практические задачи теории вероятностей успешно решаются с помощью аппарата характеристических и производящих функций, возникающего в результате применения к процедурам исчисления вероятностей методов Фурье – анализа. Полагая, что читатель достаточно подробно знаком с этим разделом математического анализа (см., например, [9]), ниже основное внимание мы обратим на прикладные результаты Фурье – анализа в их теоретико-вероятностной интерпретации. Более полные теоретические основы метода характеристических и производящих функций, включая доказательства приводимых ниже теорем, читатель может найти, например, в [2].

Характеристические функции

До сих пор полное описание свойств случайной величины (вектора) возлагалось на ее функцию распределения. Оказывается, однако, что не менее полно ее свойства определяются ее *характеристической функцией*, находящейся с функцией распределения в соотношении прямого и обратного преобразования Фурье. Для формального определения этого нового математического инструмента нам потребуется ввести в рассмотрение понятие *комплексной случайной величины*: $Y = X_1 + iX_2$, где X_1 и X_2 – действительные определенные на общем вероятностном пространстве случайные величины, i – мнимая единица.

По определению математическое ожидание и дисперсия комплексной случайной величины Y равны (если существуют)

$$MY = MX_1 + iMX_2, \quad DY = M \overset{0}{Y} \overset{0}{Y}^* = M \left((\overset{0}{X_1} + i \overset{0}{X_2})(\overset{0}{X_1} - i \overset{0}{X_2}) \right) = DX_1 + DX_2$$

(* – знак сопряженности комплексной величины).

Легко проверить, что основные свойства математического ожидания, приведенные в главе 5, сохраняются и для математического ожидания комплексной величины.

Пусть $Y = e^{iXt} = \cos Xt + i \sin Xt$, где X – действительная случайная величина, t – действительное неслучайное число. Назовем *характеристической функцией случайной величины X* математическое ожидание комплексной величины Y , рассматриваемое как функция t

$$\varphi_X(t) = MY = Me^{iXt} = M(\cos Xt) + iM(\sin Xt). \quad (7.1)$$

Используя правило вычисления математического ожидания функции случайной величины (5.11) (которое сохраняется при комплекснозначности функции), получим

а) для дискретной случайной величины X :

$$\varphi_X(t) = \sum_{k=1}^n e^{ix_k t} p(x_k), \quad (n \leq \infty, p(x_k) = P\{X = x_k\}, -\infty < t < \infty); \quad (7.2)$$

б) для абсолютно непрерывной случайной величины X с плотностью $f_X(x)$:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{ixt} f_X(x) dx. \quad (7.3)$$

Формулы (7.2) и (7.3) представляют характеристическую функцию как Фурье-преобразование распределений $\{p(x_i)\}$ или $f_X(x)$ случайной величины X . (для простоты мы опускаем случай смешенного распределения, для рассмотрения которого можно использовать представление математического ожидания в виде интеграла Стильеса, см. §5.5).

Из приведенных определений характеристической функции сразу следует ее существование при $\forall t$ ввиду неравенства $|e^{ixt}| \leq 1$. Очевидны также следующие простые свойства характеристической функции:

а) $\varphi_X(0) = 1$;

б) для $\forall t$ $|\varphi_X(t)| \leq 1$ (ввиду неравенств $|\varphi_X(t)| \leq \sum_{x_k} |e^{ix_k t}| p(x_k)$ и $|\varphi_X(t)| \leq \int_{-\infty}^{\infty} |e^{ixt}| f_X(x) dx$;

в) для $\forall a, b$ ($a = \text{const}, b = \text{const}$) и $Y = bX + a$ имеет место равенство $\varphi_Y(t) = e^{iat} \varphi_X(bt)$; в частности при $b = 0$

$$\varphi_a(t) = e^{iat}. \quad (7.4)$$

Известные в математическом анализе формулы обращения для преобразований Фурье устанавливают взаимно однозначное соответствие между прямым и обратным преобразованиями, т.е. — в теоретико-вероятностных приложениях — между распределением случайной величины и ее характеристической функцией. Сформулируем этот факт в виде теоремы.

Теорема 7.1. *Соответствие между распределением случайной величины и ее характеристической функцией, определенной формулой (7.1), является взаимно однозначным.*

Это означает, что характеристическая функция случайной величины задает ее свойства так же полно, как и ее распределение. Вместе с тем, использование характеристических функций существенно облегчает решение ряда теоретических и практических задач. Этому способствует ряд их свойств, устанавливаемых следующими теоремами.

Теорема 7.2. Пусть $\{X_k\}_{k=1}^n$ – независимые в совокупности случайные величины с характеристическими функциями $\{\varphi_{X_k}\}_{k=1}^n$. Тогда характеристическая функция суммы

$$S = \sum_{k=1}^n X_k \text{ равна}$$

$$\varphi_S(t) = \prod_{k=1}^n \varphi_{X_k}(t). \quad (7.5)$$

Это утверждение следует из равенства $\varphi_S(t) = Me^{iSt} = \prod_{k=1}^n Me^{iX_k t}$, справедливость которого вытекает из определения математического ожидания комплексной случайной величины (подробно проверить этот результат читатель может, рассмотрев случай $n = 2$ и обобщив его затем по индукции).

Теорема 7.3. Если случайная величина X имеет ограниченный начальный момент r – го порядка $\alpha_r = MX^r$ (см. §5.2), то существует производная r -го порядка ее характеристической функции $\varphi_X^{(r)}(t)$, причем

$$\varphi_X^{(r)}(0) = i^r MX^r. \quad (7.6)$$

Покажем справедливость этой теоремы для X с абсолютно непрерывным распределением. Как нам известно из главы 5, условием существования момента α_r является существование абсолютного начального момента $|\alpha|_r = M|X|^r < \infty$, а также что из существования момента α_r следует существование всех моментов α_k при $k < r$. Исходя из этого, начнем с $k = 1$, формально продифференцировав (7.3) по t под знаком интеграла:

$$\varphi_X^{(1)}(t) = \frac{d\varphi_X(t)}{dt} = \int_{-\infty}^{\infty} ixe^{ixt} f_X(x) dx. \quad (7.7)$$

Как известно из математического анализа, допустимость дифференцирования под знаком интеграла и, следовательно, корректность полученного равенства имеет место при равномерной сходимости интеграла в (7.7) по t , которое, в свою очередь, выполняется ввиду неравенств

$$\left| \int_{-\infty}^{\infty} ixe^{ixt} f_X(x) dx \right| \leq \int_{-\infty}^{\infty} |ixe^{ixt}| f_X(x) dx = \int_{-\infty}^{\infty} |x| f_X(x) dx = |\alpha|_1 < \infty.$$

Из (7.7) получаем

$$\varphi_X^{(1)}(0) = i \int_{-\infty}^{\infty} xf_X(x) dx = i\alpha_1.$$

Далее поступаем по индукции. Допустив существование производной

$$\varphi_X^{(k)}(t) = \int_{-\infty}^{\infty} i^k x^k e^{ixt} f_X(x) dx$$

для $k < r$, убеждаемся в правомерности перехода к производной $k + 1$ -го порядка путем дифференцирования под знаком интеграла по t ввиду неравенств

$$\left| \int_{-\infty}^{\infty} (ix)^{k+1} e^{ixt} f_X(x) dx \right| \leq \int_{-\infty}^{\infty} |(ix)^{k+1} e^{ixt}| f_X(x) dx = \int_{-\infty}^{\infty} |x|^{k+1} f_X(x) dx = |\alpha|_{k+1} < \infty.$$

В результате итераций от $k=2$ до $k=r-1$ получаем

$$\varphi_X^{(r)}(t) = \int_{-\infty}^{\infty} (ix)^r e^{ixt} f_X(x) dx$$

и

$$\varphi_X^{(r)}(0) = i^r M X^r = i^r \alpha_r.$$

Полученная связь производных характеристической функции случайной величины с начальными моментами позволяет удобно представить характеристическую функцию формулой Тейлора, что нам потребуется делать при изложении предельных теорем.

Так, если существует $\alpha_1 = MX$ и, следовательно, производная $\varphi_X^{(1)}(t)$, отправляясь от формулы Тейлора в форме Пеано для функции $\varphi_X(t)$ в точке $t = 0$, получим

$$\varphi_X(t) = \varphi_X(0) + t\varphi_X^{(1)}(0) + to(t) = 1 + it\alpha_1 + to(t), \quad \left(\frac{o(t)}{t} \xrightarrow{t \rightarrow 0} const \right). \quad (7.8)$$

При существовании $\alpha_2 = MX^2$ (и, следовательно, и $\alpha_1 = MX$) формула Тейлора может быть записана в виде

$$\varphi_X(t) = \varphi_X(0) + t\varphi_X^{(1)}(0) + \frac{t^2}{2}\varphi_X^{(2)}(0) + t^2O(t) = 1 + it\alpha_1 + \frac{(it)^2}{2}\alpha_2 + t^2O(t). \quad (7.9)$$

Найдем характеристические функции некоторых часто встречающихся распределений случайных величин.

Распределение Бернулли $X \in \mathcal{B}(p)$ (см.(4.9)). Имеем $P\{X = x\} = p^x(1-p)^{1-x}$ ($x = 0, 1$), отсюда

$$\varphi_X(t) = \sum_{x=0}^1 P\{X = x\} e^{ixt} = 1 - p + pe^{it} \quad (7.10)$$

Биномиальное распределение $X \in \mathcal{Bi}(n, p)$. Используя (4.10), получаем

$$\varphi_X(t) = \sum_{x=0}^n P\{X = x\} e^{ixt} = \sum_{x=0}^n C_n^x (pe^{it})^x (1-p)^{n-x} = (1-p + pe^{it})^n. \quad (7.11)$$

Из сравнения (7.10) и (7.11) и ввиду теорем 7.1 и 7.2 следует, что случайная величина X имеет распределение, совпадающее с распределением суммы n независимых случайных величин, каждая из которых имеет распределение Бернулли с параметром p . Иными словами, биномиальное распределение с параметрами n и p есть свертка n распределений Бернулли с параметром p .

Распределение Пуассона $X \in \text{Po}(a)$. Из (4.11) получаем

$$\varphi_X(t) = \sum_{x=0}^{\infty} P\{X=x\} e^{ixt} = e^{-a} \sum_{x=0}^{\infty} \frac{(ae^{it})^x}{x!} = e^{a(e^{it}-1)}. \quad (7.12)$$

Заметим, что из (7.12) и свойств характеристических функций нетрудно повторить вывод об устойчивости распределения Пуассона относительно суммирования, полученный в §6.5.

Гамма-распределение (см. (4.14)). Имеем

$$\varphi_X(t) = \frac{\int_0^{\infty} \lambda^\alpha x^{\alpha-1} e^{(it-\lambda)x} dx}{\Gamma(\alpha)}.$$

Отсюда

$$\varphi'_X(t) = \frac{i \int_0^{\infty} \lambda^\alpha x^\alpha e^{(it-\lambda)x} dx}{\Gamma(\alpha)} = \frac{i\alpha \varphi_X(t)}{\lambda - it}$$

и, решая это уравнение при $\varphi_X(0) = 1$, получаем

$$\varphi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}. \quad (7.13)$$

Полученная характеристическая функция гамма-распределения позволяет сразу найти характеристические функции еще для трех распределений, представляющих собой его частные случаи, а именно:

Показательное распределение (см. п.4.33); здесь $\alpha = 1$ и

$$\varphi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

Распределение Эрланга (см. п.4.34); здесь $\alpha = n$ (целое) и

$$\varphi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-n}.$$

Распределение χ^2 (см. п.4.35); здесь $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$ и

$$\varphi_X(t) = (1 - 2it)^{-\frac{n}{2}}.$$

Нормальное (гауссовское) распределение $X \in \mathcal{N}(m, \sigma^2)$ (см.4.19)). Начнем со случая $Y \in \mathcal{N}(0, 1)$:

$$\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cos xtdx + i \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \sin xtdx \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cos xtdx$$

(ввиду равенства нулю второго справа интеграла от нечетной функции). В результате дифференцирования полученного равенства по t (что допустимо, поскольку полученный интеграл равномерно сходится по t) получим

$$\varphi_Y^{(1)}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \sin xt e^{-\frac{x^2}{2}} dx$$

или, в результате интегрирования по частям, уравнение

$$\varphi_Y^{(1)}(t) = -t\varphi_Y(t),$$

решение которого при начальном условии $\varphi_Y(0) = 1$ приводит к результату

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}. \quad (7.14)$$

Пусть теперь $X \in N(m, \sigma^2)$. Это распределение случайная величина X имеет, будучи связанной со случайной величиной Y равенством $X = \sigma Y + m$. Отсюда для характеристической функции X получаем

$$\varphi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}. \quad (7.15)$$

По виду характеристических функций легко проверяется устойчивость относительно суммирования (воспроизводимость) распределений по их параметрам. Так, из приведенных здесь распределений полной воспроизводимостью (по всем параметрам) обладают распределения Пуассона, Гаусса и χ^2 , частичной воспроизводимостью (по параметру α) – гамма-распределение и (по параметру n) – биномиальное и распределение Эрланга.

Отсюда, в частности, обнаруживается важная связь между нормальным распределением и распределением χ^2 . Найдем распределение случайной величины $Y = X^2$, где $X \in N(0, 1)$:

$$F_Y(y) = P\{Y < y\} = P\{-\sqrt{y} \leq X < \sqrt{y}\} = 2\Phi^*(\sqrt{y}) - 1, \quad f_Y(y) = \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}, \quad (y > 0),$$

которое, как легко видеть (см.(4.18)), имеет распределение χ^2 при $n = 1$, т.е. с одной степенью свободы. Ввиду воспроизводимости распределения χ^2 это означает, что сумма квадратов n независимых случайных величин $\{X_j\}_{j=1}^n$ с стандартным нормальным распределением ($X_j \in N(0, 1)$) имеет распределение χ^2 с n степенями свободы:

$$\sum_{j=1}^n X_j^2 \in \chi^2(n).$$

Наряду с взаимно однозначным соответствием между распределением случайной величины и ее характеристической функцией (теорема 7.1) оказывается, что это соответствие обладает свойством непрерывности, смысл которого разъясняет приводимая ниже без доказательства теорема 7.4.

Прежде, однако, забегаая вперед, применим к функциям распределения известное из математического анализа понятие слабой сходимости функций. Пусть заданы функции

распределения $\{F_n(x)\}$ последовательности случайных величин $\{X_n\}$ и функция распределения $F(x)$ случайной величины X . Последовательность $\{F_n(x)\}$ слабо сходится к функции $F(x)$, если выполняется условие

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$$

в каждой точке непрерывности функции $F(x)$. Для такой последовательности функций распределения принята запись $F_n(x) \xrightarrow{сл} F(x)$, а для последовательности случайных величин $\{X_n\} - X_n \xrightarrow{d} X$ (сходимость по распределению).

Теорема 7.4. Пусть $\{X_n\}$, $\{F_n(x)\}$ и $\{\varphi_n(t)\}$ – последовательности случайных величин, их функций распределения и характеристических функций. Тогда

1) если для $\forall t \varphi_n(t) \xrightarrow{n \rightarrow \infty} \varphi(t)$ и $\varphi(t)$ непрерывна в $t = 0$, то $\varphi(t)$ – характеристическая функция некоторой случайной величины X с функцией распределения $F(x)$, к которой слабо сходится последовательность $\{F_n(x)\}$: $F_n(x) \xrightarrow{сл} F(x)$;

2) если $F_n(x) \xrightarrow{сл} F(x)$, где $F(x)$ функция распределение некоторой случайной величины X , то для $\forall t \varphi_n(t) \xrightarrow{n \rightarrow \infty} \varphi(t)$, где $\varphi(t)$ – характеристическая функция распределения случайной величины X .

Приведем следствие теоремы 7.4, важность которого придает ему статус теоремы.

Теорема 7.5 (интегральная теорема Муавра – Лапласа). Пусть $\{X_j\}$ последовательность независимых в совокупности случайных величин, имеющих одинаковое распределение Бернулли $B(p)$. Тогда последовательность случайных величин $\{S_n\}$, где S_n – нормированная сумма

$$S_n = \frac{\sum_{j=1}^n X_j - np}{\sqrt{np(1-p)}}, \quad (7.16)$$

имеют функции распределения $F_{S_n}(s)$, сходящиеся при $n \rightarrow \infty$ к стандартной нормальной функции распределения:

$$F_{S_n}(s) \rightarrow \Phi^*(s).$$

Поскольку из условия этой теоремы сумма $\sum_{j=1}^n X_j = W_n$ имеет биномиальное распределение ($W_n \in Be(n, p)$), смысл теоремы следует трактовать, как утверждение о сходимости этого распределения к нормальному при $n \rightarrow \infty$.

Доказательство теоремы 7.5 основано на вычислении характеристической функции случайной величины $S_n = \frac{W_n - np}{\sqrt{np(1-p)}}$, где $W_n \in \text{Be}(n, p)$ и, соответственно, $\varphi_{W_n}(t) = (q + pe^{it})^n$. Из свойства «в» характеристических функций следует

$$\varphi_{S_n}(t) = e^{-it\sqrt{\frac{np}{1-p}}} \varphi_{W_n}\left(\frac{t}{\sqrt{np(1-p)}}\right) = \left((1-p)e^{-it\sqrt{\frac{p}{n(1-p)}}} + pe^{-it\sqrt{\frac{1-p}{np}}} \right)^n.$$

Далее, используя разложение в ряд Маклорена, можно получить

$$(1-p)e^{-it\sqrt{\frac{p}{n(1-p)}}} + pe^{-it\sqrt{\frac{1-p}{np}}} = 1 - \frac{t^2}{2n}(1 + R_n),$$

где $R_n \rightarrow 0$ при $n \rightarrow \infty$, откуда следует

$$\varphi_{S_n}(t) = \left(1 - \frac{t^2}{2n}(1 + R_n) \right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}},$$

что, согласно теореме 7.4, означает сходимость $F_{S_n}(s) \rightarrow \Phi^*(s)$, причем ввиду непрерывности функции распределения $\Phi^*(s)$, эта сходимость имеет место при всех s .

Метод характеристических функций без принципиальных изменений применим и в многомерном стохастическом анализе при исследовании свойств случайных векторов. Практически все результаты, полученные для скалярных случайных величин, распространяются и на векторные величины. Поэтому ограничимся здесь общим определением и двумя примерами характеристических функций случайного вектора.

Характеристической функцией случайного вектора $X = (X_1, \dots, X_n)'$ называют функцию $\varphi_X(\mathbf{t})$ вещественного векторного неслучайного аргумента $\mathbf{t} = (t_1, \dots, t_n)'$

$$\varphi_X(\mathbf{t}) = M e^{i\mathbf{t}'X}$$

(принимаящую действительное или комплексное значение).

Найдем, например, характеристическую функцию случайного вектора, имеющего полиномиальное распределение: $X \in P(n, \mathbf{p})$ (см. п.6.4.1).

Из (6.36) в результате несложных преобразований (с учетом вырожденности вектора X) получаем

$$\begin{aligned} \varphi_X(\mathbf{t}) &= \sum_{(x_1 + \dots + x_{k-1})} \frac{n!}{x_1! \dots x_{k-1}! (n - \sum_{j=1}^{k-1} x_j)!} p_1^{x_1} \dots p_{k-1}^{x_{k-1}} (1 - \sum_{j=1}^{k-1} p_j)^{(n - \sum_{j=1}^{k-1} x_j)} e^{i(t_1 x_1 + \dots + t_{k-1} x_{k-1})} = \\ &= (p_1 e^{it_1} + \dots + p_{k-1} e^{it_{k-1}} + 1 - \sum_{j=1}^{k-1} p_j)^n. \end{aligned}$$

Для вектора X с нормальным распределением: $X \in N(\mathbf{m}_X, \mathbf{R}_X)$ (см. (6.39)) характеристическая функция равна (см., например, [1], [7]).

$$\varphi_X(t) = Me^{it'X} = e^{it'm_X - \frac{t'R_X t}{2}}. \quad (7.17)$$

Используя приведенные нами характеристические функции для случайных векторов, имеющих полиномиальное и нормальное распределения, можно получить многомерный аналог теоремы Муавра – Лапласа, имеющий следующее содержание.

Теорема 7.6. Пусть случайный вектор X_n имеет полиномиальное распределение $X_n \in P(n, p)$, $p = (p_1, \dots, p_k)'$, $\forall p_j > 0$. Тогда функция распределения $F_{Y_n}(y)$ нормированного вектора $Y_n = \frac{X_n - MX_n}{\sqrt{n}}$ при $n \rightarrow \infty$ сходится к нормальному (вырожденному) k -мерному распределению с нулевым вектором математического ожидания и с корреляционной матрицей

$$R_Y = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_k & -p_2p_k & \cdots & p_k(1-p_k) \end{pmatrix}.$$

Не приводя доказательства этой теоремы, укажем, что принципиально оно аналогично доказательству теоремы 7.5.

Тот факт, что характеристическая функция нормального случайного вектора существует и при вырожденной корреляционной матрице, позволяет расширить класс нормальных случайных векторов, определяя нормальное распределение не видом плотности распределения (см. определение 6.7), что не имеет смысла в вырожденном случае, а видом характеристической функции (7.17). Заметим, что при таком определении нормального случайного вектора нетрудно установить справедливость сформулированного в главе 6 утверждения 6.1.

Производящие функции

В некоторых случаях при построении и исследовании стохастических моделей оказывается удобным применение аппарата *производящих функций*, имеющего то же прикладное предназначение, что и метод характеристических функций. Существенно уступая последнему по универсальности, метод производящих функций прост и не связан с теорией Фурье-анализа. Здесь мы приведем основные сведения об этом методе.

Пусть X – случайная величина, принимающая только целочисленные значения, т.е. значения из множества $X = (0, 1, 2, \dots, n, \dots)$, с вероятностями $P\{X = j\} = p_j \geq 0$, $j = 0, 1, 2, \dots, n$, $n \leq \infty$, $\sum_{j=0}^n p_j = 1$. Образует случайную величину $V = z^X$, где z –

нелучайная вещественная или комплексная величина с модулем $|z| \leq 1$. Производящая функция $\phi_X(z)$ случайной величины X определяется как математическое ожидание величины V :

$$\phi_X(z) = MV = \sum_{j=0}^n z^j p_j, \quad (n \leq \infty). \quad (7.18)$$

Ввиду очевидной сходимости ряда в (7.18), производящая функция существует при любом распределении целочисленной случайной величины X .

Нетрудно убедиться, что производящая функция при представлении $z = e^{it}$ превращается в характеристическую функцию для частного случая величины X с целочисленными значениями, и свойства характеристических функций, описанные в §7.1 присущи (в определенной модификации) и производящим функциям. Перечислим основные (достаточно очевидные) из них.

1. $\phi_X(1) = 1$;

2. Пусть целочисленные случайные величины $\{X_j\}_{j=1}^k$ независимы в совокупности и $Y = \sum_{j=1}^k X_j$. Тогда $\phi_Y(z) = \prod_{j=1}^k \phi_{X_j}(z)$;

3. Между распределениями целочисленных случайных величин и их производящими функциями существует взаимно однозначное соответствие, выраженное (7.18) и обратным преобразованием

$$p_j = \frac{1}{j!} \phi_X^{(j)}(0), \quad j = 0, 1, 2, \dots \quad (7.19)$$

(здесь $\phi_X^{(j)}(0)$ – j -ая производная $\phi_X(z)$ в точке $z = 0$);

4. Между производными производящей функции целочисленной случайной величины и ее моментами существует связь, позволяющая сравнительно просто их вычислять. Так, например, в общем случае

$$MX = \alpha_1 = \phi_X^{(1)}(1), \quad DX = \alpha_2 - \alpha_1^2 = \phi_X^{(2)}(1) + \phi_X^{(1)}(1) - \left(\phi_X^{(1)}(1)\right)^2.$$

Читателю предлагается проверить, что для конкретных распределений производящие функции имеют следующий вид:

для $X \in B(p)$ (распределение Бернулли) $\phi_X(z) = 1 - p + pz$;

для $X \in Bi(n, p)$ (биномиальное распределение) $\phi_X(z) = (1 - p + pz)^n$;

для $X \in Po(a)$ (распределение Пуассона) $\phi_X(z) = e^{-a(z-1)}$.

Приведем пример совместного применения характеристической и производящей функций.

Пример 7.1. Рассмотрим поток событий, отличающийся от простого пуассоновского процесса, рассмотренного в главе 4, тем, что каждое событие обладает «весом», представляющим собой реализацию v случайной величины V с известным распределением. Такой процесс, носящий название сложного пуассоновского процесса (см. [6]), описывает, например, поток коммерческих сделок (проектов), приносящих случайный доход. Задача состоит в вычислении параметров суммарного веса, доставляемого таким потоком за некоторое заданное время, т.е. случайной величины Q , представляющей собой сумму случайного числа случайных величин

$$Q = \sum_{j=1}^L V_j,$$

где $L \in \text{Po}(\lambda\tau)$ – число событий в простейшем (пуассоновском) потоке с интенсивностью λ в заданном интервале времени τ , V_j – вес j -го события, имеющий известное распределение, которое в этом примере примем одинаковым для всех событий.

Получить в общем случае распределение случайной величины Q обычно непросто, что же касается моментов (в том числе математического ожидания и дисперсии этого распределения), то для их вычисления удобно использовать метод характеристических функций.

Найдем характеристическую функцию случайной величины Q , используя формулу полного математического ожидания (6.31) при $Y = e^{iQ^t}$, $B_j = \{L = j\}$:

$$\begin{aligned} \varphi_Q(t) &= M e^{iQ^t} = \sum_{j=0}^{\infty} M(e^{iQ^t} | L = j) P\{L = j\} = \sum_{j=0}^{\infty} (M e^{iV^t})^j p(j) = \sum_{j=0}^{\infty} (\varphi_V(t))^j p(j), \\ (p(j) &= P\{L = j\}), \end{aligned}$$

т.е. характеристическая функция случайной величины Q равна производящей функции случайной величины L с распределением Пуассона ($L \in \text{Po}(\lambda\tau)$), при $z = \varphi_V(t)$. Следовательно

$$\varphi_Q(t) = e^{\lambda\tau(\varphi_V(t)-1)}.$$

Используя связь между характеристической функцией и моментами случайной величины, находим

$$MQ = \lambda\tau MV \text{ и } DQ = \lambda\tau MV^2$$

(подробнее см. [6]).

Предельные теоремы

До сих пор нами рассматривались, как правило, задачи, в которых фигурировали конечные совокупности случайных событий и величин. Однако уже в первой главе, рассматривая статистическое определение вероятности события, мы столкнулись с необходимостью говорить о пределе последовательности случайных величин $W_n(A)$ – частот события A при неограниченном возрастании n , оставив, до времени, обсуждение смысла понятия предела в стохастических моделях. В настоящей главе излагаются основные предельные теоремы, описывающие предельное поведение последовательностей случайных событий и величин.

Эти последовательности представляют собой счетные совокупности случайных событий $\{A_n\}_{n=1}^{\infty}$ или случайных величин $\{X_n\}_{n=1}^{\infty}$, определенные на общем вероятностном пространстве и упорядоченные индексом (аргументом) n .

Вводя понятие предела таких последовательностей, приходится убедиться в многообразии понятия их сходимости в приложениях к стохастическим моделям. Ниже рассматриваются основные типы сходимости последовательностей случайных величин (к которым относятся и последовательности случайных событий, представленных их индикаторами).

Сходимость последовательностей случайных величин

При изучении асимптотического поведения последовательностей случайных величин возникает вопрос о смысле, вкладываемом в понятие их сходимости к пределу. В отличие от последовательности детерминированных величин, под *стохастической сходимостью* членов последовательности случайных величин к некоторому пределу следует понимать их приближение к нему в некотором вероятностном смысле, который может варьироваться в зависимости от содержания задачи. Опишем возникающие при этом вопросы в теоретико-вероятностных терминах.

Пусть параллельно наблюдается динамика состояний большого (теоретически – неограниченного) множества однотипных объектов (частиц, живых организмов, экономических субъектов, технических устройств и пр.), состояние каждого из которых выражается некоторым скалярным числом, принимающим в дискретные моменты $n = 1, 2, \dots, j, \dots$ различные (в общем) значения. В схеме вероятностного пространства множество всех объектов образует исходное множество Ω , каждому элементу которого ω соответствует,

следовательно, выбор конкретного объекта. Предположим, далее, что на множестве Ω определен случайный N -мерный вектор

$X^N = (X_1, X_2, \dots, X_N)' = x^N(\omega) = (x_1(\omega), x_2(\omega), \dots, x_N(\omega))'$, описывающий состояние объектов в моменты $n = 1, 2, \dots, N$. Здесь последовательность $x_1(\omega), x_2(\omega), \dots, x_N(\omega)$ для каждого фиксированного $\omega \in \Omega$ представляет собой реализацию траектории состояний конкретного объекта в моменты времени $n = 1, 2, \dots, N$, а X_n – случайную величину, множество возможных значений которой определено множеством возможных состояний объектов в фиксированный момент времени n .

Допустим, далее, что в рамках конкретной стохастической модели нас интересуют асимптотические свойства этого вектора, выражающиеся в сходимости последовательности X_1, X_2, \dots, X_N к некоторому случайному или неслучайному пределу X при $N \rightarrow \infty$. Исследования такого (и подобного) рода требуют, однако, определения понятия сходимости для случайных величин, которое в наших приложениях оказывается неоднозначным, будучи зависимым, как уже говорилось, от конкретной постановки задачи. Для иллюстрации рассмотрим два (из возможных) подхода к определению понятия стохастической сходимости (ниже они будут изучены подробнее вместе с другими определениями этого понятия).

В первом подходе понятию стохастической сходимости придается следующий смысл. Зададим любую малую окрестность величины X и будем оценивать стохастическую близость X_n к X вероятностной мерой («долей») множества исходов (объектов) из Ω , для которых при данном n случайная величина X_n лежит в заданной окрестности величины X . Когда для каждой фиксированной окрестности X эта вероятность при $n \rightarrow \infty$ стремится к единице, то говорят о сходимости X_n к X «по вероятности». Другими словами это означает, что чем больше n , тем больше вероятность того, что выбранный наугад объект имеет значение X_n , лежащее в заданной окрестности величины X , выполняющей для последовательности $(X_1, X_2, \dots, X_n, \dots)' = \{X_n\}_{n=1}^{\infty}$ роль предела по вероятности.

Оказывается, однако, что существование сходимости X_n к X «по вероятности» вовсе не означает, что состояния всех или хотя бы почти всех объектов образуют траектории $x_1(\omega), x_2(\omega), \dots, x_n(\omega), \dots$, сходящиеся к какому либо пределу. Более того, при этом все траектории могут не иметь предела.

В связи с этим возможен другой подход к определению сходимости случайных величин. Будем полагать, что последовательность случайных величин $\{X_n\}$ сходится к X , если множество исходов, для которых реализации вектора X^∞ , т.е. траектории $x_1(\omega), x_2(\omega), \dots, x_n(\omega), \dots$, сходятся (в обычном смысле) к $x(\omega)$, имеет вероятностную меру, равную единице.

Такой тип сходимости называют сходимостью «с вероятностью единица» или «сходимостью почти наверное».

Полезно подчеркнуть, что применение того или иного критерия стохастической сходимости в конкретной модели определяется ее содержательным смыслом. Так, в моделях налогообложения коммерческих фирм налоговая власть может принимать решения, основанные на описании состояния налогооблагаемой базы в терминах сходимости по вероятности, в то время, как для действий руководства отдельных фирм более естественным является описание стохастических процессов с использованием критерия сходимости почти наверное.

8.1.1. Сходимость по вероятности

Выраженный выше словесно смысл сходимости последовательности случайных величин к некоторому пределу по вероятности формально имеет следующий вид.

Пусть на общем вероятностном пространстве $\langle \Omega, F, P \rangle$ заданы последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ и случайная (возможно – вырожденная, т.е. равная константе) случайная величина X . Тогда имеет место

Определение 8.1. *Последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ сходится по вероятности к пределу X , если выполняется условие для*

$$\forall \varepsilon > 0, 0 < \delta < 1, \exists n_0(\varepsilon, \delta) : \forall n > n_0 \quad P\{|X_n - X| \leq \varepsilon\} > 1 - \delta \quad (8.1)$$

или, короче,

$$\text{для } \forall \varepsilon > 0, P\{|X_n - X| \leq \varepsilon\} \xrightarrow{n \rightarrow \infty} 1 \quad (8.2)$$

(используемое далее обозначение: $X_n \xrightarrow{P} X$).

Полагая $\varepsilon = \frac{1}{k}$ и обозначая $A_n^k = \left\{ |X_n - X| \leq \frac{1}{k} \right\}$, условие (8.1) можно записать в виде

$$\text{для } \forall k, \delta, 0 < \delta < 1, \exists n_0 : P\left(\bigcap_{n > n_0} A_n^k\right) > 1 - \delta$$

или, проще, для $\forall k \quad P(A_n^k) \xrightarrow{n \rightarrow \infty} 1$.

Приведем пример сходимости по вероятности.

Пример 8.1. Пусть в вероятностном пространстве $\langle \Omega, F, P \rangle$ $\Omega = [0, 1)$, F – σ -алгебра с полуинтервалами вида $\Omega_{mj} = \left[\frac{(j-1)}{m}, \frac{j}{m} \right)$, $m = 1, 2, \dots; j = 1, \dots, m$ и P – вероятностная мера, заданная для $\forall m, j$ равенством $P\{\omega \in \Omega_{mj}\} = \frac{1}{m}$. Зададим последовательность случайных величин $\{X_{mj}\}$, каждая из которых представляет собой функцию вида

$$X_{mj} = x_{mj}(\omega) = \begin{cases} m, & \text{если } \omega \in \Omega_{mj}, \\ 0, & \text{если } \omega \notin \Omega_{mj}. \end{cases}$$

Перенумеровав их (положив $n = \frac{m(m-1)}{2} + j$) получим последовательность случайных величин $\{X_n\}$ с распределением вероятностей

$$P\{X_n = x\} = \begin{cases} \frac{1}{m} & \text{при } x = m, \\ \frac{m-1}{m} & \text{при } x = 0. \end{cases}$$

Отсюда следует, что эта последовательность сходится по вероятности к нулю:

$$\forall \varepsilon > 0, P\{|X_n - 0| = 0 < \varepsilon\} = \frac{m-1}{m} \xrightarrow[n \rightarrow \infty]{(m \rightarrow \infty)} 1,$$

что соответствует (8.2).

8.1.2. Сходимость с вероятностью единица (почти наверное)

Как уже говорилось, каждая реализация последовательности случайных величин может рассматриваться как траектория процесса с целочисленным аргументом. При исследовании поведения множества этих реализаций в совокупности естественно использовать понятие сходимости по вероятности. При изучении поведения отдельных траекторий больший интерес, однако, может представлять описываемая ниже сходимость с вероятностью единица (или, иначе говоря, сходимость почти наверное).

Сохраняя смысл исходных данных и обозначений, принятых в начале предыдущего пункта, и представляя реализацию случайных величин X_n как функции исходов $x_n(\omega)$ и $x(\omega)$ (см. определение 4.1), введем следующее определение.

Определение 8.2. Последовательность случайных величин $\{X_n = x_n(\omega)\}_{n=1}^{\infty}$ сходится с вероятностью единица (почти наверное) к пределу $X = x(\omega)$, если выполняется условие

$$P\{\omega : \lim_{n \rightarrow \infty} x_n(\omega) = x(\omega)\} = 1; \quad (8.3)$$

(используемое далее обозначение: $X_n \xrightarrow{П.Н.} X$).

Иногда вместо (8.3) удобно использовать эквивалентное условие:

$$\text{для } \forall \varepsilon > 0 \quad \lim_{n_0 \rightarrow \infty} P\left\{\omega : \sup_{n \geq n_0} |x_n(\omega) - x(\omega)| > \varepsilon\right\} = 0. \quad (8.4)$$

Проверим эквивалентность условий (8.3) и (8.4). Действительно, снова обозначая $A_n^k = \left\{|X_n - X| \leq \frac{1}{k}\right\} = \left\{|x_n(\omega) - x(\omega)| \leq \frac{1}{k}\right\}$, представим событие $B = \{\lim_{n \rightarrow \infty} x_n(\omega) = x(\omega)\}$ в виде

$$B = \bigcap_{k=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n \geq n_0} A_n^k,$$

и условие (8.3) примет вид $P(B) = 1$ или $P(\bar{B}) = 0$, где (согласно правилу де Моргана)

$$\bar{B} = \bigcup_{k=1}^{\infty} \bigcap_{n_0=1}^{\infty} \bigcup_{n \geq n_0} \bar{A}_n^k. \text{ Из последнего представления очевидно, что равенство } P(\bar{B}) = 0$$

осуществляется тогда и только тогда, когда при $\forall k$ выполняется равенство $P(\bigcap_{n_0=1}^{\infty} \bigcup_{n \geq n_0} \bar{A}_n^k) = 0$

или, ввиду теоремы о непрерывности вероятности, равенство $\lim_{n_0 \rightarrow \infty} P(\bigcup_{n \geq n_0} \bar{A}_n^k) = 0$. Отсюда, ввиду равенства

$$\bigcup_{n \geq n_0} \bar{A}_n^k = \left\{ \omega : \sup_{n \geq n_0} |x_n(\omega) - x(\omega)| > \frac{1}{k} \right\}, \text{ следует } \lim_{n_0 \rightarrow \infty} P(\bigcup_{n \geq n_0} \bar{A}_n^k) = \lim_{n_0 \rightarrow \infty} P\left\{ \omega : \sup_{n \geq n_0} |x_n(\omega) - x(\omega)| > \frac{1}{k} \right\}$$

и вместо (8.3) в качестве критерия сходимости с вероятностью единица может использоваться условие (8.4).

Приведенное определение сходимости последовательности случайных величин, характеризую индивидуальное асимптотическое поведение ее траекторий, требует выполнения условия (8.3) на множестве исходов, имеющих единичную вероятностную меру, допуская его невыполнение на множестве нулевой меры (отсюда термин «сходимость почти наверное»).

Сходимость последовательности случайных величин с вероятностью единица влечет за собой ее сходимость по вероятности. Чтобы убедиться в этом, перепишем (8.2) в виде

$$\text{для } \forall \varepsilon > 0, P\left\{ |x_{n_0}(\omega) - x(\omega)| > \varepsilon \right\} \xrightarrow{n_0 \rightarrow \infty} 0 \quad (8.5)$$

и, используя очевидное соотношение

$$\left\{ |x_{n_0}(\omega) - x(\omega)| > \varepsilon \right\} \subseteq \left\{ \sup_{n \geq n_0} |x_n(\omega) - x(\omega)| > \varepsilon \right\},$$

ввиду (8.4) получаем (8.5), т.е. имеет место импликация

$$X_n \xrightarrow{П.Н.} X \Rightarrow X_n \xrightarrow{P} X. \quad (8.6)$$

Обратное утверждение неверно: сходимость по вероятности не влечет за собой сходимости с вероятностью единица. Чтобы убедиться в этом, достаточно привести контрпример, роль которого выполняет пример 8.1, при условиях которого имеет место сходимость по вероятности (8.2), но не выполняется условие сходимости с вероятностью единица (8.3). Действительно, для этого примера имеем

$$\text{для } \forall k, k > 0, P\left\{ \bigcap_{n_0=1}^{\infty} \bigcup_{n \geq n_0} \bar{A}_n^k \right\} \neq 0,$$

что противоречит (8.4), означая, что последовательность $\{X_n\}$ не сходится к $X = 0$ с вероятностью единица.

Более точно соотношение между этими типами сходимости можно сформулировать в виде утверждения:

Утверждение 8.1. *Если существует сходимость последовательности $\{X_n\}$ к X по вероятности, то или имеет место сходимость $\{X_n\}$ с вероятностью единица к тому же пределу X , или $\{X_n\}$ не сходится с вероятностью единица ни к какому пределу (возможность существования различных пределов для этих двух типов сходимости исключается ввиду очевидной единственности предела по вероятности).*

Мы убедились, следовательно, что сходимость с вероятностью единица «сильнее» сходимости по вероятности. Для практики, однако, важным является не только (иногда – и не столько) такое «иерархическое» различие между этими типами сходимости, но, как уже отмечалось, и их различный практический смысл.

8.1.3. Сходимость в среднем квадратическом

Рассмотренные типы сходимости последовательности случайных величин не содержат прямой информации о разбросе значений ее членов относительно предельной величины. Даже если установлено, что последовательность $\{X_n\}$ сходится к некоторому пределу по вероятности или с вероятностью единица, остается открытым вопрос о том, как ведет себя разброс отклонений значений X_n от этого предела при возрастании n . Вместе с тем, такой вопрос возникает во многих приложениях (стохастическое моделирование систем автоматического управления, моделирование экономических процессов с учетом риска и др.).

В связи с этим, в теории случайных процессов (см., например, [6]) широко применяется тип сходимости последовательности случайных величин, именуемый *сходимостью в среднем квадратическом*, суть которого раскрывается следующим определением.

Определение 8.3. *Последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ сходится в среднем квадратическом к пределу X , если выполняется условие*

$$\lim_{n \rightarrow \infty} M(X_n - X)^2 = 0 \quad (8.7)$$

(используемое далее обозначение: $X_n \xrightarrow{C.K.} X$).

Используя неравенство Чебышева, легко установить, что сходимость в среднем квадратическом влечет за собой сходимость по вероятности. Действительно, положим в (5.18) $Y = (X_n - X)^2$. Тогда для $\forall \varepsilon^2 = a$ получим

$$P\{|X_n - X| \geq \varepsilon\} = P\{(X_n - X)^2 \geq \varepsilon^2\} \leq \frac{M(X_n - X)^2}{\varepsilon^2}$$

и, сравнивая (8.2) и (8.7), приходим к импликации

$$X_n \xrightarrow{C.K.} X \Rightarrow X_n \xrightarrow{P} X.$$

Таким образом, сходимость последовательности случайных величин $\{X_n\}$ к пределу X в среднем квадратическом означает, что эта последовательность имеет предел по вероятности и, одновременно, среднее значение квадрата ее отклонения от этого предела стремится к нулю с ростом n .

Как и в отношении к сходимости с вероятностью единица, сходимость по вероятности «слабее» сходимости в среднем квадратическом: из существования первого типа сходимости не следует существование второго. Проверить это можно, снова используя в качестве контрпримера пример 8.1. Действительно, для условий этого примера получаем

$$M(X_n - X)^2 = MX_n^2 = m \xrightarrow{n \rightarrow \infty} \infty.$$

Соотношение между сходимостью по вероятности и в среднем квадратическом можно выразить утверждением, аналогичным утверждению 8.1:

Утверждение 8.2. Если существует сходимость последовательности $\{X_n\}$ к X по вероятности, то или имеет место сходимость $\{X_n\}$ в среднем квадратическом к тому же пределу X , или $\{X_n\}$ не сходится в среднем квадратическом ни к какому пределу (возможность существования различных пределов для этих двух типов сходимости исключается ввиду очевидной единственности предела по вероятности).

Доминируя над сходимостью по вероятности, сходимости с вероятностью единица и в среднем квадратическом не находятся в отношении доминирования между собой: существование одной из них не влечет за собой существование другой. Для проверки истинности такого положения приведем следующие примеры.

Пример 8.2. Сохраняя вероятностное пространство, описанное в примере 8.1, зададим на нем последовательность случайных величин $\{X_{mj}\}$ равенствами

$$X_{mj} = x_{mj}(\omega) = \begin{cases} \frac{1}{m}, & \text{если } \omega \in \Omega_{mj}, \\ 1 - \frac{1}{m}, & \text{если } \omega \notin \Omega_{mj}. \end{cases}$$

Перенумеровав их (так же, как в примере 8.1), получим последовательность случайных величин $\{X_n\}$ с распределением вероятностей

$$P\{X_n = x\} = \begin{cases} \frac{1}{m} & \text{при } x = \frac{1}{m}, \\ \frac{m-1}{m} & \text{при } x = 1 - \frac{1}{m}. \end{cases}$$

Читатель легко убедится, что в этом примере последовательность $\{X_n\}$ сходится к $X = 1$ в среднем квадратическом (действительно: $M(X_n - 1)^2 = \frac{n(n-1)}{n^3} \xrightarrow{n \rightarrow \infty} 0$) и, следовательно, и по вероятности, но не имеет предела с вероятностью единица (для $\forall \omega$ $x_n(\omega)$ с ростом n не сходится к пределу, принимая бесконечно много раз значения, приближающиеся к 0 и 1).

Таким образом, сходимость в среднем квадратическом не влечет за собой сходимость с вероятностью единица.

Симметричное утверждение вытекает из следующего примера.

Пример 8.3. Пусть в вероятностном пространстве $\langle \Omega, F, P \rangle$ $\Omega = [0, 1)$, F – σ -алгебра с полуинтервалами вида $\Omega_n = [0, \frac{1}{n})$, $n = 1, 2, \dots$ и P – вероятностная мера, задаваемая для $\forall n$ равенством $P\{\omega \in \Omega_n\} = \frac{1}{n}$. Зададим последовательность случайных величин $\{X_n\}$ равенством

$$X_n = x_n(\omega) = \begin{cases} n, & \text{если } \omega \in \Omega_n, \\ 0, & \text{если } \omega \notin \Omega_n \end{cases}$$

с распределением вероятностей

$$P\{X_n = x\} = \begin{cases} \frac{1}{n} & \text{при } x = n, \\ \frac{n-1}{n} & \text{при } x = 0 \end{cases}$$

и установим, какими типами сходимости эта последовательность обладает. Очевидно, что $X_n \xrightarrow{П.Н.} 0$, поскольку для $\forall \varepsilon$ существует n_0 такое, что для $\forall n > n_0$ $X_n = 0$. Отсюда следует, что $X_n \xrightarrow{P} 0$. Далее находим, что $M(X_n - 0)^2 = n \rightarrow \infty$, т.е. рассматриваемая последовательность не сходится к нулю в среднем квадратическом и, следовательно, сходимость с вероятностью единица не влечет за собой существование сходимости в среднем квадратическом.

Обратим внимание на тот очевидный факт, что при одновременной сходимости последовательности $\{X_n\}$ с вероятностью единица и в среднеквадратическом, пределы этих сходимостей совпадают, поскольку их различие означало бы, что последовательность $\{X_n\}$ одновременно сходится по вероятности к двум различным пределам, что исключено.

8.1.4. Сходимость по распределению

До сих пор, описывая типы сходимости последовательности случайных величин, мы интересовались существованием и величиной предела ее элементов. Как будет показано ниже, типичной прикладной задачей, в которой это оказывается необходимым, является

анализ предельного поведения среднего арифметического, как оценки математического ожидания случайной величины.

Часто, однако, задача состоит не в поиске предела последовательности самих случайных величин, а в нахождении их предельного распределения. Как мы вскоре выясним, необходимость в этом возникает при построении стохастических моделей, содержащих суммы большого числа случайных величин.

Пусть, как всегда, $\{X_n\}$ и X – последовательность случайных величин и случайная величина, претендующая на роль предела этой последовательности. Обозначим $F_n(x)$ и $F(x)$ функции распределения случайных величин X_n и X соответственно ($n = 1, 2, \dots$).

Теперь, говоря о сходимости X_n к X в смысле сходимости их распределений, мы имеем в виду выполнение условия $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$, которое, однако, требует уточнения в связи с возможной разрывностью функции $F(x)$.

Оказывается, что достаточно потребовать, чтобы эта сходимость осуществлялась в каждой точке непрерывности функции $F(x)$, т.е. являлась бы слабой сходимостью: $F_n(x) \xrightarrow{cl} F(x)$. Действительно, при таком типе сходимости предельна функция распределения $F(x)$ аппроксимируется функцией, отличающейся от нее («выколотой») не более чем на счетном множестве значений X . Не трудно установить, что такая аппроксимация эквивалентна $F(x)$, поскольку задает то же распределение в борелевской σ -алгебре B_X (при использовании, в случае необходимости, теоремы о непрерывности вероятности). С другой стороны, требование сходимости $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$ всюду приводило бы к случаям неоправданного отрицания сходимости X_n к X по распределению. Следует также напомнить, что в теореме (7.4) о непрерывном соответствии характеристических функций и функций распределения последовательности случайных величин фигурирует слабая сходимость функций распределения.

В итоге сформулируем определение:

Определение 8.4. Последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ с функциями распределения $\{F_n(x)\}_{n=1}^{\infty}$ сходится по распределению к пределу X с функцией распределения $F(x)$, если выполняется условие

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$$

для $\forall x$, для которых $F(x)$ непрерывна, т.е. при $F_n(x) \xrightarrow{cl} F(x)$;

(используемое далее обозначение: $X_n \xrightarrow{d} X$).

Выясним, какое место занимает этот новый тип сходимости в иерархической системе доминирования типов сходимости, рассмотренных выше. Естественно начать сравнение сходимости по распределению со сходимости по вероятности (которая пока является самой «слабой» из рассмотренных). Прежде всего, установим, что из сходимости по вероятности всегда следует сходимость по распределению:

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X.$$

Обозначим событие $A_n^\varepsilon = \{|X_n - X| < \varepsilon\}$ и пусть для $\forall \varepsilon$ $P(A_n^\varepsilon) \xrightarrow{n \rightarrow \infty} 1$, т.е. $X_n \xrightarrow{P} X$.

Представим событие A_n^ε в виде

$$A_n^\varepsilon = \{X - \varepsilon < X_n < X + \varepsilon\},$$

из которого следует, что для $\forall x$ при осуществлении события A_n^ε имеют место импликации:

$$\begin{aligned} \{X_n < x\} &\Rightarrow \{X - \varepsilon < x\}, \\ \{X + \varepsilon < x\} &\Rightarrow \{X_n < x\}, \end{aligned}$$

из которых следует

$$F_X(x - \varepsilon) \leq F_{X_n}(x) \leq F_X(x + \varepsilon).$$

Из полученного соотношения очевидно, что если x – любая точка непрерывности функции распределения $F_X(x)$, то $F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_X(x)$, т.е. $X_n \xrightarrow{d} X$.

Обратное утверждение в общем случае неверно: из сходимости последовательности случайных величин по распределению делать вывод о существовании ее сходимости по вероятности нельзя. Чтобы убедиться в этом, достаточно вспомнить, что различные случайные величины могут иметь совпадающие распределения.

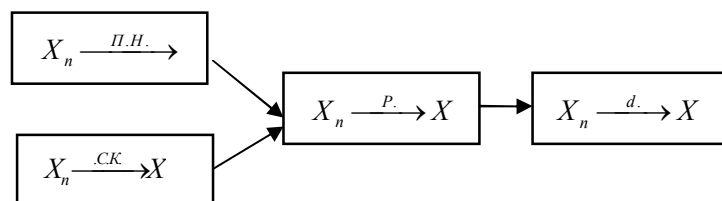
Из этого факта есть, однако, следующее важное исключение. Пусть $X_n \xrightarrow{d} c = \text{const}$, т.е., другими словами, X имеет вырожденное распределение. Тогда имеет место и сходимость по вероятности: $X_n \xrightarrow{P} c$. Действительно, в этом случае для $\forall \varepsilon > 0$ получаем

$$F_{X_n}(c - \varepsilon) \xrightarrow{n \rightarrow \infty} 0, F_{X_n}(c + \varepsilon) \xrightarrow{n \rightarrow \infty} 1,$$

$$\text{т.е. } P\{c - \varepsilon \leq X_n < c + \varepsilon\} \xrightarrow{n \rightarrow \infty} 1,$$

что и означает сходимость по вероятности $X_n \xrightarrow{P} c$.

В итоге можно представить связь между различными типами сходимости последовательности случайных величин схемой, приведенной на рис.8.1.



Законы больших чисел

Одной из центральных задач теории вероятностей является изучение предельного поведения приведенных сумм членов последовательности случайных величин $\{X_k\}_{k=1}^{\infty}$, т.е. последовательности случайных величин

$$\{S_n\}_{n=1}^{\infty}, \text{ где } S_n = \frac{\sum_{k=1}^n X_k}{n}.$$

Потребность в этом возникает уже при статистическом определении вероятности (см. §1.3), когда вероятность события A определяется (согласно (1.13) при замене обозначения N на n) как предел

$$P(A) = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{n},$$

где X_k – индикатор события A в k -ом наблюдении. Напомним, что мы оставили пока открытым вопрос о смысле предела в этом равенстве; пришло время к нему вернуться.

Аналогично обстоит дело с представлением математического ожидания MX случайной величины X как предела ее среднего арифметического \bar{X} (5.2) (см. §5.1).

Эти и подобные им задачи решаются рядом теорем, носящих название законов больших чисел (ЗБЧ).

Общая постановка задачи в этих теоремах звучит так.

Пусть задана последовательность случайных величин $\{X_k\}_{k=1}^{\infty}$, которая преобразуется в

последовательность случайных величин $\{S_n\}_{n=1}^{\infty}$, где $S_n = \frac{\sum_{k=1}^n X_k}{n}$. Необходимо определить тип и условия сходимости последовательности $\{S_n\}_{n=1}^{\infty}$ к пределу S и найти этот предел, если он существует.

Разнообразие конкретных постановок задач и условий их решения влечет за собой множественность теорем, представляющих ЗБЧ в той или иной форме. Остановимся на их главных различиях.

Форма ЗБЧ зависит от предположения о глубине связи между членами последовательности $\{X_k\}_{k=1}^{\infty}$: они могут полагаться независимыми в совокупности, попарно

независимыми, некоррелированными или зависимыми. Наиболее проработаны и практически применимы ЗБЧ, основанные на первых трех предположениях.

Существенное влияние на форму ЗБЧ оказывает различие предположений о распределении X_k : содержание теорем о ЗБЧ зависит от того, полагаются ли члены последовательности одинаково распределенными или нет, а также от предположений о существовании и свойствах их моментов (математического ожидания и дисперсии).

Принципиальным, наконец, является устанавливаемый теоремой тип сходимости в предельном соотношении

$$S_n = \frac{\sum_{k=1}^n X_k}{n} \xrightarrow{n \rightarrow \infty} S. \quad (8.8)$$

В существующих вариантах ЗБЧ фигурируют два типа сходимости: по вероятности и с вероятностью единица. Поскольку второй тип сходимости сильнее первого, ЗБЧ, в которых действует сходимость с вероятностью единица, именуются *усиленными законами больших чисел* (УЗБЧ).

Отметим некоторую условность такого определения. Предположим, что существует «слабый» ЗБЧ, теорема о котором утверждает существование сходимости (8.8) по вероятности, и в результате дополнительных исследований установлено, что при сохранении условий теоремы имеет место и сходимость (8.8) с вероятностью единица (т.е. при сохранении условий задачи имеет место теорема о существовании УЗБЧ). В этом случае необходимость в теореме об «обычном» ЗБЧ исчезает и термин «усиленный» теряет свой обычный смысл (сохраняя, возможно, историческое значение).

Иначе дело обстоит в том случае, когда замена в (8.8) сходимости по вероятности на сходимость с вероятностью единица оказывается возможной в результате изменения (например – ужесточения) некоторых условий теоремы о «слабом» ЗБЧ. Следует, однако, полагать, что и в этом случае прилагательное «усиленный» имеет условный характер, поскольку теоремы о слабом и усиленном ЗБЧ сохраняют свою самостоятельность.

Начнем рассмотрение теорем о законах больших чисел с вариантов, в которых используется сходимость по вероятности.

Теорема 8.1 (ЗБЧ в форме Чебышева). Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность попарно некоррелированных случайных величин с конечными математическими ожиданиями $MX_k = a_k$ и с дисперсиями DX_k , удовлетворяющими ограничению: $DX_k \leq ck^{\alpha}$, где $c < \infty$, $k = 1, 2, \dots$, $0 \leq \alpha < 1$. Тогда

$$\frac{\sum_{k=1}^n X_k}{n} \xrightarrow[n \rightarrow \infty]{P} \frac{\sum_{k=1}^n a_k}{n}. \quad (8.9)$$

В обозначениях, принятых в (8.8), утверждение (8.9) означает

$$S_n \xrightarrow[n \rightarrow \infty]{P} MS_n. \quad (8.10)$$

Для доказательства теоремы найдем, что, ввиду попарной некоррелированности членов последовательности $\{X_k\}_{k=1}^{\infty}$ и согласно (8.8),

$$DS_n = \frac{\sum_{k=1}^n DX_k}{n^2} \leq \frac{c \sum_{k=1}^n k^{\alpha}}{n^2} < \frac{c \int_0^{n+1} x^{\alpha} dx}{n^2} = \frac{c(n+1)^{1+\alpha}}{(1+\alpha)n^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

т.е. применимо неравенство Чебышева (5.17), которое здесь принимает вид

$$\forall \varepsilon > 0: P\{|S_n - MS_n| \geq \varepsilon\} \leq \frac{DS_n}{\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

что и означает выполнение соотношений (8.9) и (8.10) (заметим, что обычно в формулировке доказанной теоремы принимается $\alpha = 0$).

Из теоремы 8.1 сразу вытекает ряд следствий, имеющих значение частных случаев.

Так, если члены последовательности $\{X_k\}_{k=1}^{\infty}$ попарно некоррелированы и имеют одинаковое распределение с конечной дисперсией $DX_k = \sigma^2$ и, следовательно, с конечным математическим ожиданием $MX_k = a$, то (8.9) примет вид

$$\frac{\sum_{k=1}^n X_k}{n} \xrightarrow[n \rightarrow \infty]{P} a = MX,$$

где левая дробь представляет собой среднее арифметическое (выборочное среднее) значение

для n членов последовательности $\{X_k\}_{k=1}^{\infty}$: $\bar{X}(n) = \frac{\sum_{k=1}^n X_k}{n}$. Следовательно,

$$\bar{X}(n) \xrightarrow[n \rightarrow \infty]{P} MX$$

и теорема 8.1 разъясняет смысл сходимости среднего арифметического к математическому ожиданию, о которой шла речь в §5.1.

Пусть теперь члены последовательности $\{X_k\}_{k=1}^{\infty}$ попарно некоррелированы и имеют одинаковое распределение Бернулли: $X_k \in B(p)$. Если рассматривать случайную величину X_k как индикатор появления события A в k -ом опыте, то величина S_n принимает смысл частоты появления события A в n опытах: $S_n = W_n(A)$ (см. (1.10) при замене обозначений N на n). В результате (8.10) превращается в соотношение

$$W_n(A) \xrightarrow[n \rightarrow \infty]{P} P(A) = p,$$

определяющее смысл предела в (1.11) и, таким образом, post factum устраняющее в определенной мере нестрогость в статистическом определении понятия вероятности (см. §1.3).

Заметим, что приведенные следствия из теоремы 8.1, ввиду их важности и исторического приоритета, нередко представляются в литературе по теории вероятностей как самостоятельные теоремы.

Перейдем к изложению вариантов теорем о ЗБЧ, в которых используется сходимость с вероятностью единица, т.е. об усиленных законах больших чисел. Для доказательства основных из этих теорем нам потребуются факты, устанавливаемые следующими утверждениями.

Лемма 8.1. (Бореля – Кантелли).

Пусть $\{A_k\}_{k=1}^{\infty}$ – последовательность случайных событий, принадлежащих общему вероятностному пространству и имеющих вероятности $\{P(A_k)\}_{k=1}^{\infty}$. Тогда,

1) если выполнено условие

$$\sum_{k=1}^{\infty} P(A_k) < \infty, \quad (8.11)$$

то в последовательности $\{A_k\}_{k=1}^{\infty}$ с вероятностью единица происходит лишь конечное число событий;

2) если события из $\{A_k\}_{k=1}^{\infty}$ независимы в совокупности и

$$\sum_{k=1}^{\infty} P(A_k) = \infty, \quad (8.12)$$

то в последовательности $\{A_k\}_{k=1}^{\infty}$ с вероятностью единица происходит бесконечное число событий.

Для доказательства леммы обозначим B событие, состоящее в том, что в последовательности $\{A_k\}_{k=1}^{\infty}$ происходит бесконечное число событий, т.е.

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k. \quad (8.13)$$

Тогда

$$\bar{B} = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_k \quad (8.14)$$

является событием, состоящим в том, что в последовательности $\{A_i\}_{i=1}^{\infty}$ происходит лишь конечное число событий. Пусть выполнено (8.11). Тогда для $\forall n$ имеет место (ввиду теоремы о непрерывности вероятности и свойства сходящегося ряда) соотношение

$$P(B) \leq P\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k) \xrightarrow{n \rightarrow \infty} 0,$$

из которого следует, что $P(B) = 0$ и $P(\bar{B}) = 1$, т.е. выполняется первое утверждение леммы.

Для доказательства ее второго утверждения используем независимость событий $\{A_k\}$ (напомним, что под независимостью в совокупности событий, образующих бесконечную последовательность $\{A_k\}_{k=1}^{\infty}$, понимается независимость в совокупности событий из любого конечного набора событий из $\{A_k\}_{k=1}^{\infty}$). Из независимости событий $\{A_k\}$ следует, как известно, и независимость событий $\{\bar{A}_k\}$. Используя неравенство $1 - a \leq e^{-a}$, $0 \leq a \leq 1$, получим для $\forall n$, $m \geq n$ (с учетом свойства расходящегося ряда (8.12))

$$P\left(\bigcap_{k=n}^m \bar{A}_k\right) = P(\bar{A}_n) \dots P(\bar{A}_m) = (1 - P(A_n)) \dots (1 - P(A_m)) \leq e^{-\sum_{k=n}^m P(A_k)} \xrightarrow{m \rightarrow \infty} 0.$$

Снова используя теорему о непрерывности вероятности, находим, что

$$P\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) = \lim_{m \rightarrow \infty} P\left(\bigcap_{k=n}^m \bar{A}_k\right) = 0,$$

т.е. все события, входящие в объединение в (8.14), имеют нулевую вероятность, откуда получаем $P(\bar{B}) = 0$ и $P(B) = 1$, что соответствует второму утверждению леммы.

Лемма 8.2. Необходимым и достаточным условием существования конечного математического ожидания случайной величины X является сходимость ряда

$$\sum_{n=1}^{\infty} P\{|X| \geq n\} < \infty.$$

Напомним, что необходимым и достаточным условием существования конечного математического ожидания MX случайной величины X является существование конечного математического ожидания ее модуля $M|X|$. Очевидны соотношения

$$\begin{aligned} \sum_{n=1}^{\infty} (n-1)P\{n-1 \leq |X| < n\} &\leq M|X| \leq \sum_{n=1}^{\infty} nP\{n-1 \leq |X| < n\}, \\ \sum_{n=1}^{\infty} nP\{n-1 \leq |X| < n\} &= \sum_{n=0}^{\infty} P\{|X| \geq n\} \leq 1 + \sum_{n=1}^{\infty} P\{|X| \geq n\}, \\ \sum_{n=1}^{\infty} (n-1)P\{n-1 \leq |X| < n\} &= \sum_{n=1}^{\infty} nP\{n-1 \leq |X| < n\} - P\{|X| \geq 0\} = \sum_{n=1}^{\infty} P\{|X| \geq n\}, \end{aligned}$$

из которых следует соотношение

$$\sum_{n=1}^{\infty} P\{|X| \geq n\} \leq M|X| \leq 1 + \sum_{n=1}^{\infty} P\{|X| \geq n\},$$

доказывающее справедливость леммы.

Теорема 8.2. (Неравенство Колмогорова).

Пусть $\{X_j\}_{j=1}^{\infty}$ – независимые случайные величины с конечными математическими ожиданиями MX_j и дисперсиями DX_j , $Y_k = \sum_{j=1}^k X_j$, $k = 1, 2, \dots, n$. Тогда для $\forall u = \text{const} > 0$

$$P\left\{\max_{1 \leq k \leq n} |Y_k - MY_k| \geq u\right\} \leq \frac{DY_n}{u^2}. \quad (8.15)$$

Эта теорема, будучи в определенном смысле обобщением неравенства Чебышева, ограничивает отклонение сумм $Y_k = \sum_{j=1}^k X_j$ от их математического ожидания величиной, определяемой дисперсией суммы Y_n (обладающей, как нетрудно видеть, наибольшей дисперсией для величин из набора $\{Y_k\}_{k=1}^n$).

Положим без ограничения общности для $\forall j$ $MX_j = 0$ (в противном случае перейдем к случайным величинам $X'_j = X_j - MX_j$); неравенство (8.15) принимает вид

$$P\left\{\max_{1 \leq k \leq n} |Y_k| \geq u\right\} \leq \frac{DY_n}{u^2}. \quad (8.16)$$

Введем случайную величину $V = \min\{k : |Y_k| \geq u\}$, равную номеру суммы в их последовательности $\{Y_k\}_{k=1}^n$, которая первой достигла по модулю заданный порог u . Обозначим событие $A_k = \{V = k\}$ и его индикатор I_k . Ясно, что $\sum_{k=1}^n I_k \leq 1$, так как события A_k образуют неполную группу несовместимых событий (все I_k равны нулю, когда для всех $k = 1, 2, \dots, n$ выполняется событие $\{|Y_k| < u\}$). Поэтому справедливы неравенства

$$\begin{aligned} Y_n^2 &\geq Y_n^2 \sum_{k=1}^n I_k, \\ MY_n^2 &\geq \sum_{k=1}^n MY_n^2 I_k = \sum_{k=1}^n M((X_1 + \dots + X_k + X_{k+1} + \dots + X_n)^2 I_k) \geq \\ &\geq \sum_{k=1}^n M((X_1 + \dots + X_k)^2 I_k) + 2 \sum_{k=1}^n M((X_1 + \dots + X_k)(X_{k+1} + \dots + X_n) I_k). \end{aligned}$$

Каждый индикатор I_k является случайной величиной, зависящей от величин X_1, \dots, X_k и не зависящей от величин X_{k+1}, \dots, X_n . Поэтому для последней суммы получаем

$$\sum_{k=1}^n M((X_1 + \dots + X_k)(X_{k+1} + \dots + X_n) I_k) = \sum_{k=1}^n M((X_1 + \dots + X_k) I_k) M(X_{k+1} + \dots + X_n) = 0$$

и

$$MY_n^2 \geq \sum_{k=1}^n M((X_1 + \dots + X_k)^2 I_k) = \sum_{k=1}^n M(Y_k^2 I_k).$$

В последней сумме для каждого отличного от нуля индикатора I_k имеем $Y_k^2 \geq u^2$.

Поэтому $MY_n^2 \geq u^2 \sum_{k=1}^n MI_k = u^2 P\{V \leq n\} = u^2 P\left\{\max_{1 \leq k \leq n} |Y_k| \geq u\right\}$, т.е. получаем (8.15).

Изложение теорем об усиленных законах больших чисел начнем с теоремы, которая (как и теорема 8.1 среди теорем о «слабых» ЗБЧ) может рассматриваться как «базовая».

Теорема 8.3.

Пусть $\{X_j\}_{j=1}^{\infty}$ — последовательность независимых в совокупности случайных величин с конечными дисперсиями DX_j и выполняется условие

$$\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty. \quad (8.17)$$

Тогда

$$\frac{\sum_{j=1}^n X_j}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \frac{\sum_{j=1}^n a_j}{n}, \quad (a_j = MX_j). \quad (8.18)$$

Сравнивая теоремы 8.1 и 8.3 заметим, что, различаясь типами сходимости в (8.9) и (8.18), они содержат вместе с тем разные требования к независимости случайных величин из последовательности $\{X_j\}_{j=1}^{\infty}$ и к свойствам их дисперсий.

Не ограничивая общности, мы можем положить для $\forall k$ $a_k = 0$ (в противном случае можно перейти к случайным величинам $X'_j = X_j - a_j$, что не повлияет на результат доказательства). Тогда, снова обозначая $Y_n = \sum_{j=1}^n X_j$, получим (8.18) в виде $\frac{Y_n}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} 0$, т.е. (см. (8.4)),

$$P\left\{\sup_{n \geq n_0} \left|\frac{Y_n}{n}\right| > \varepsilon\right\} \xrightarrow[n_0 \rightarrow \infty]{} 0 \text{ при } \forall \varepsilon > 0. \quad (8.19)$$

Разобьем множество всех целых чисел N на непересекающиеся отрезки $N = \bigcup_{n=1}^{\infty} [2^{n-1}, 2^n)$

и, обозначив

$$A_n = \left\{ \max_{2^{n-1} \leq k < 2^n} \left| \frac{Y_k}{k} \right| > \varepsilon \right\},$$

убеждаемся, что (8.19) эквивалентно соотношению

$$P\left(\bigcup_{n=n_0}^{\infty} A_n\right) \xrightarrow{n_0 \rightarrow \infty} 0 \quad (8.20)$$

(оба соотношения означают, что для $\forall \varepsilon > 0$ вероятность выхода случайной величины $\frac{Y_n}{n}$ из ε -окрестности нуля при каком либо $n > n_0$ может быть сколь угодно малой, если n_0 достаточно велико).

Очевидна цепочка неравенств (последнее из них – неравенство Колмогорова (8.19))

$$P(A_n) \leq P\left\{\max_{2^{n-1} \leq k < 2^n} |Y_k| > 2^{n-1} \varepsilon\right\} \leq P\left\{\max_{1 \leq k \leq 2^n} |Y_k| > 2^{n-1} \varepsilon\right\} \leq \frac{4DY_{2^n}}{2^{2n} \varepsilon^2}, \text{ где } DY_{2^n} = \sum_{j=1}^{2^n} DX_j,$$

откуда, ввиду (6.28), получим

$$\sum_{k=1}^{\infty} P(A_k) \leq 4\varepsilon^{-2} \sum_{k=1}^{\infty} 2^{-2k} \sum_{n=1}^{2^k} DX_n \leq 4\varepsilon^{-2} \sum_{n=1}^{\infty} DX_n \sum_{\left(k: 2^{-2k} \geq n^{-2}\right)} 2^{-2k} \leq 8\varepsilon^{-2} \sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty.$$

Из сходимости ряда $\sum_{k=1}^{\infty} P(A_k)$ следует

$$P\left(\bigcup_{n=n_0}^{\infty} A_n\right) \leq \sum_{k=n_0}^{\infty} P(A_k) \xrightarrow{n_0 \rightarrow \infty} 0,$$

т.е. выполнение (8.20) и доказательство теоремы.

Условия теоремы 8.3 не содержат требования об одинаковом распределении случайных величин, входящих в последовательность $\{X_j\}_{j=1}^{\infty}$, хотя на практике это условие нередко выполняется. Поэтому интересно выяснить, какое «послабление» в других условиях закона больших чисел можно получить при включении в них требования одинакового распределения случайных величин $\{X_j\}$. В классе слабых ЗБЧ, основанных на сходимости по вероятности, ответ на этот вопрос дает следующая теорема.

Теорема 8.4 (ЗБЧ в форме Хинчина). Пусть $\{X_j\}_{j=1}^{\infty}$ – последовательность независимых в совокупности одинаково распределенных случайных величин с конечным математическим ожиданием $MX_j = a < \infty$. Тогда

$$\frac{\sum_{j=1}^n X_j}{n} \xrightarrow[n \rightarrow \infty]{P} a. \quad (8.21)$$

Сравнивая теоремы 8.3 и 8.4, мы видим, что при одинаковом распределении независимых случайных величин $\{X_j\}$ снимаются какие либо требования к их дисперсии (включая требование об ее существовании). Оказывается, что утверждение теоремы 8.4 (при сохранении ее условий) может быть усилено заменой сходимости по вероятности в (8.21) на

сходимость с вероятностью единица (почти наверное). В результате оказывается справедливой следующая теорема.

Теорема 8.5 (УЗБЧ Колмогорова). Пусть $\{X_j\}_{j=1}^{\infty}$ — последовательность независимых в совокупности одинаково распределенных случайных величин. Тогда существование конечного математического ожидания $a = MX_j$ является необходимым и достаточным условием сходимости

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} a. \quad (8.22)$$

Поскольку теорема 8.4 следует из 8.5, приведем доказательство второй из них.

Начнем с доказательства достаточности условия теоремы.

Итак, допустим, что все случайные величины из $\{X_j\}_{j=1}^{\infty}$ имеют одинаковые функции распределения $F(x)$ и конечные математические ожидания: $MX_j = MX = a < \infty$, и, следовательно, равные конечные математические ожидания их модулей $M|X| = \int |x| dF(x)$.

В лемме 8.2 было получено неравенство

$$\sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \sum_{n=1}^{\infty} P\{|X| > n\} \leq M|X| = \int |x| dF(x) < \infty \quad (8.23)$$

Введем в рассмотрение случайные величины

$$\tilde{X}_n = \begin{cases} X_n, & \text{если } |X_n| < n, \\ 0, & \text{если } |X_n| \geq n; \end{cases}$$

$$\tilde{Y}_n = \sum_{j=1}^n \tilde{X}_j.$$

Ясно, что случайные величины $\{\tilde{X}_j\}_{j=1}^{\infty}$, являясь функциями независимых величин, также независимы.

Обозначая, по-прежнему, $Y_n = \sum_{j=1}^n X_j$, представим (8.22) в виде

$$\frac{Y_n - na}{n} \equiv \frac{Y_n - \tilde{Y}_n}{n} + \frac{\tilde{Y}_n - M\tilde{Y}_n}{n} + \frac{M\tilde{Y}_n}{n} - a \xrightarrow[n \rightarrow \infty]{\text{п.н.}} 0. \quad (8.24)$$

Теорема будет доказана, если выяснится, что все три слагаемые в (8.24) сходятся с вероятностью единица к нулю.

Начиная с первого слагаемого, обозначим $B_n = \{X_n \neq \tilde{X}_n\}$. Имеем

$$\sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} P\{|X_n| \geq n\} < \infty \quad (8.25)$$

(ввиду существования MX_n (см.(8.23)). Сходимость ряда в (8.26) согласно лемме Бореля-Кантелли означает, что событие $B_n = \{X_n \neq \tilde{X}_n\}$ осуществляется лишь для конечного числа значений n , что приводит к сходимости $\frac{Y_n - \tilde{Y}_n}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} 0$.

Переходя к второму слагаемому в (8.24), покажем, что последовательность $\{\tilde{X}_j\}_{j=1}^{\infty}$ удовлетворяет условию (8.17) в теореме 8.3 при замене $\{X_j\}_{j=1}^{\infty}$ на $\{\tilde{X}_j\}_{j=1}^{\infty}$, т.е. выполняется условие

$$\sum_{n=1}^{\infty} \frac{D\tilde{X}_n}{n^2} < \infty.$$

Следует при этом заметить, что появление здесь дисперсий $D\tilde{X}_n$ не налагает условий на существование дисперсий DX_n ; как мы убедимся, для этого оказывается достаточным предположение о конечности математических ожиданий. (8.24).

Третье слагаемое в (8.24) не содержит случайных элементов и ее сходимость с вероятностью единица к нулю означает существование обычного предела

$$\lim_{n \rightarrow \infty} \frac{M\tilde{Y}_n - an}{n} = 0,$$

что имеет место, поскольку дробь $\frac{M\tilde{Y}_n - an}{n}$ представляет собой среднее арифметическое, члены которого стремятся к нулю при $n \rightarrow \infty$ (проверку этого читатель может выполнить самостоятельно).

Тем самым доказывается достаточность условия теоремы.

Необходимость условия теоремы доказывается проще. Из сходимости $\frac{Y_n}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} a$

получаем $\frac{X_n}{n} = \frac{Y_n}{n} - \frac{(n-1)}{n} \frac{Y_{n-1}}{(n-1)} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} 0$, т.е. событие $\left\{\left|\frac{X_n}{n}\right| \geq 1\right\} = \{|X_n| \geq n\}$ с вероятностью единица осуществляется лишь конечное число раз. Отсюда, в силу независимости и одинакового распределения случайных величин X_n и согласно лемме 8.1 (п.2) получаем

$$\sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \sum_{n=1}^{\infty} P\{|X| \geq n\} < \infty,$$

что, ввиду леммы 8.2, означает существование конечного математического ожидания MX . Теорема доказана.

В рассмотренных нами теоремах о законах больших чисел присутствовали требования к той или иной форме независимости случайных величин, образующих последовательность $\{X_k\}_{k=1}^{\infty}$, — их попарная некоррелированность или независимость в совокупности. Поскольку в

реальных задачах эти требования могут не выполняться, рассмотрим теорему о ЗБЧ, в которой никаких ограничений на зависимость случайных величин не налагается.

Теорема 8.6 (теорема Маркова). Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность случайных величин с математическими ожиданиями $MX_j = a_j < \infty$ и удовлетворяющие условию

$$\frac{1}{n^2} D\left(\sum_{k=1}^n X_k\right) \xrightarrow{n \rightarrow \infty} 0, \quad (8.26)$$

тогда выполняется ЗБЧ в смысле сходимости по вероятности, т.е.

$$\frac{\sum_{k=1}^n X_k}{n} \xrightarrow[n \rightarrow \infty]{P} \frac{\sum_{k=1}^n a_k}{n}. \quad (8.27)$$

(подчеркнем, что здесь не налагается условий на независимость случайных величин).

Доказательство этой теоремы основано на неравенстве Чебышева. Обозначая снова

$$S_n = \frac{\sum_{k=1}^n X_k}{n}, \text{ получаем}$$

$$P\{|S_n - MS_n| \geq \varepsilon\} \leq \frac{DS_n}{\varepsilon^2} = \frac{D\left(\sum_{k=1}^n X_k\right)}{n^2 \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0, \text{ т.е. (8.27).}$$

Условие (8.26) является достаточным, но не необходимым для выполнения (8.27). Можно, однако, найти не только достаточное, но и необходимое условие выполнения (8.27); в отличие от (8.26) оно имеет вид

$$M\left(\frac{\sum_{k=1}^n (X_k - MX_k)^2}{n^2 + \sum_{k=1}^n (X_k - MX_k)^2}\right) \xrightarrow{n \rightarrow \infty} 0 \quad (\text{см. [9]}).$$

Отказ от требования независимости случайных величин, фигурирующих в теореме, приводит лишь к кажущемуся упрощению, поскольку сопряжен с вычислением дисперсий их сумм с учетом их зависимости.

Можно показать, что при независимости случайных величин $\{X_k\}_{k=1}^{\infty}$ необходимое и достаточное условие выполнения (8.27) существенно упрощается, принимая вид

$$\sum_{k=1}^n M\left(\frac{(X_k - MX_k)^2}{n^2 + (X_k - MX_k)^2}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Подробнее о приведенных результатах см. [10].

Пределные теоремы для распределений сумм случайных величин.

В предыдущем параграфе были рассмотрены теоремы, описывающие предельное поведение нормированных сумм случайных величин, конкретно – сходимость этих сумм к их математическим ожиданиям.

Позволяя корректно ставить и решать задачи, связанные с оценкой среднего значения суммарного эффекта от действия многочисленных случайных факторов (величин), законы больших чисел не приносят, однако, информации о вероятностном распределении этого эффекта. Так, например, с помощью ЗБЧ можно оценить математическое ожидание суммарной прибыли от большого числа сделок, осуществляемых коммерческой фирмой, но, не зная распределение суммарной прибыли, нельзя оценить размер риска, выраженного, например, вероятностью получения фирмой отрицательной прибыли.

Поэтому большой интерес представляет собой вопрос о сходимости распределений сумм случайных величин к некоторым предельным распределениям.

8.3.1. Пределная теорема для суммы одинаково распределенных независимых случайных величин

Будем полагать, что содержанием некоторого реального явления (процесса) является суммирование большого числа одинаково распределенных независимых в совокупности случайных величин из последовательности $\{X_k\}_{k=1}^{\infty}$, и ставится задача получения предельного распределения суммы $Y_n = \sum_{k=1}^n X_k$ при $n \rightarrow \infty$. Нетрудно, однако, убедиться, что такая постановка задачи применительно непосредственно к случайной величине Y_n некорректна в связи с неограниченностью параметров ее распределения (MY_n, DY_n). Поэтому далее мы будем изучать предельное распределение нормированной суммы

$$S_n = \frac{Y_n - MY_n}{\sqrt{DY_n}} = \frac{\sum_{k=1}^n X_k - nm}{\sigma\sqrt{n}}$$

в предположении существования конечных $m = MX$ и $\sigma^2 = DX$.

В этих условиях справедлива теорема, именуемая *центральной* или *классической предельной теоремой*.

Теорема 8.7. Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность одинаково распределенных независимых в совокупности случайных величин с конечными $m = MX$ и $\sigma^2 = DX$, $\sigma^2 > 0$. Тогда

$$\forall s : F_{S_n}(s) = P\left\{S_n = \frac{\sum_{k=1}^n X_k - nm}{\sigma\sqrt{n}} < s\right\} \xrightarrow[n \rightarrow \infty]{cl} \Phi^*(s) \quad (8.28)$$

т.е.

$$S_n \xrightarrow[n \rightarrow \infty]{d} S \in N(0, 1)$$

(напомним, что $\Phi^*(\cdot)$ – функция распределения стандартного нормального распределения, см.(4.21)). Для доказательства этой теоремы применим аппарат характеристических функций. Переходя к центрированным случайным величинам $\overset{0}{X}_i = X_i - MX$, получим

$$S_n = \frac{\sum_{i=1}^n \overset{0}{X}_i}{\sigma\sqrt{n}}.$$

Для характеристических функций случайных величин $\overset{0}{X}_i$ согласно (7.9) имеем $\varphi_X(t) = 1 - \frac{t^2\sigma^2}{2} + t^2O(t)$, откуда для характеристической функции случайной величины S_n для $\forall t$ получаем

$$\varphi_{S_n}(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2 O(\frac{t}{\sigma\sqrt{n}})}{\sigma^2 n} \right)^n \xrightarrow[n \rightarrow \infty]{} e^{-\frac{t^2}{2}},$$

что ввиду (7.14) означает выполнение (8.28), т.е. доказательство теоремы.

Читатель легко проверит, что в частном случае, когда X_k имеет распределение Бернулли с параметром p , т.е. $\sum_{k=1}^n X_k \in \text{Bi}(n, p)$, доказанная теорема превращается в теорему 7.5, т.е. в интегральную теорему Муавра – Лапласа (доказанную, заметим, задолго до теоремы 8.7 методами комбинаторного анализа и представляющую собой первообраз появившихся позднее предельных теорем).

Поскольку практическое применение предельной теоремы осуществляется при достаточно больших, но конечных n (обычно при $n > 10$), возникает вопрос об оценке величины расхождения функций распределения $F_{S_n}(s)$ и $\Phi^*(s)$ при фиксированном n . Не останавливаясь на доказательствах, приведем результаты исследования этого вопроса.

Ввиду того, что сходимость функции $F_{S_n}(s)$ к функции $\Phi^*(s)$ при $n \rightarrow \infty$ оказывается равномерной по s , примем за меру расхождения между $F_{S_n}(s)$ и $\Phi^*(s)$ величину $\sup_s |F_{S_n}(s) - \Phi^*(s)|$. Допустим, что случайные величины $\{X_i\}_{i=1}^\infty$ в теореме 8.7 обладают конечным третьим абсолютным центральным моментом $M|X-MX|^3$. Тогда имеет место *неравенство Берри – Эссеена*:

$$\sup_s |F_{S_n}(s) - \Phi^*(s)| \leq \frac{CM|X - MX|^3}{\sigma^3 \sqrt{n}}, \quad (8.29)$$

где C – абсолютная константа, $C \cong 0.8$; (доказательство (8.29) см. в [2], [8]).

Приведем пример применения этого соотношения.

Пример 8.4. Рассмотрим последовательность событий A_1, A_2, \dots , результатом каждого из которых является реализация случайной величины X_k с одинаковым для всех A_k распределением, обладающим конечными $m = MX_k$, $\sigma^2 = DX_k$, $|\mu|^3 = M|X_k - MX_k|^3$; случайные величины $\{X_k\}$ полагаются независимыми в совокупности. Задача состоит в определении минимального числа событий \bar{n} , в результате которых сумма $Y_{\bar{n}} = \sum_{k=1}^{\bar{n}} X_k$ с вероятностью, не меньшей α , окажется не ниже заданной величины v . Содержательно такая модель может соответствовать, например, определению минимального числа коммерческих сделок, приносящих фирме с заданной надежностью заданную прибыль.

Итак, задача состоит в определении \bar{n} из условия

$$\bar{n} = \min \{n : P\{Y_n \geq v\} < \alpha\} = \min \{n : F_{Y_n}(v) < 1 - \alpha\},$$

где $F_{Y_n}(v)$ – функция распределения суммы Y_n .

Исходя из теоремы 8.7, решим сначала задачу с точностью до аппроксимации распределения $F_{Y_n}(v)$, т.е. аргумент полагая \bar{n} достаточно большим для возможности использования как точного приближенного равенства

$$F_{Y_n}(v) = P\{Y_{\bar{n}} < v\} = P\left\{\frac{Y_{\bar{n}} - \bar{n}m}{\sigma\sqrt{\bar{n}}} < \frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right\} = P\left\{S_{\bar{n}} < \frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right\} = F_{S_{\bar{n}}}\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right) \cong \Phi^*\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right) < 1 - \alpha,$$

из которого при известных m и σ нетрудно найти приближенное значение \bar{n} . Если оно оказывается достаточно большим, часто делается заключение, что это и есть решение задачи.

Такой вывод, однако, не всегда правомерен. Действительно, из (8.29) следует, что точность аппроксимации $F_{S_{\bar{n}}}\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right) \cong \Phi^*\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right)$ зависит от отношения моментов

$\frac{M|X - MX|^3}{\sigma^3}$: с его ростом расхождение между $F_{S_{\bar{n}}}\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right)$ и $\Phi^*\left(\frac{v - \bar{n}m}{\sigma\sqrt{\bar{n}}}\right)$ возрастает и может

оказаться при найденном \bar{n} слишком большим.

Учет этого обстоятельства приводит к определению \bar{n} из следующих очевидных соотношений

$$\begin{aligned}
P\{Y_n < v\} &= F_{S_n}\left(\frac{v-nm}{\sigma\sqrt{n}}\right) = \Phi^*\left(\frac{v-nm}{\sigma\sqrt{n}}\right) + F_{S_n}\left(\frac{v-nm}{\sigma\sqrt{n}}\right) - \Phi^*\left(\frac{v-nm}{\sigma\sqrt{n}}\right) \leq \\
&\leq \Phi^*\left(\frac{v-n}{\sigma\sqrt{n}}\right) + \sup_s |F_{S_n}(s) - \Phi^*(s)| \leq \Phi^*\left(\frac{v-nm}{\sigma\sqrt{n}}\right) + \frac{CM|X-MX|^3}{\sigma^3\sqrt{n}},
\end{aligned}$$

т.е. к выбору \bar{n} из условия:

$$\bar{n} = \min\{n : \Phi^*\left(\frac{v-nm}{\sigma\sqrt{n}}\right) + \frac{CM|X-MX|^3}{\sigma^3\sqrt{n}} < 1-\alpha\}.$$

8.3.2. Предельные теоремы для суммы разно распределенных независимых случайных величин

Замечательный результат теоремы 8.7, состоящий в возможности аппроксимировать нормальным распределением распределение суммы достаточно большого числа случайных величин с одинаковым произвольным распределением (с конечными двумя первыми моментами), приводит к вопросу о возможности подобного результата и для различно распределенных случайных величин, образующих сумму. Конечно, и в этом случае необходимо потребовать существования конечных математических ожиданий и дисперсий этих случайных величин, однако выполнение этих требований здесь оказывается недостаточным ввиду следующих обстоятельств.

В теореме 8.7 все случайные величины, образующие сумму, будучи одинаково распределенными, оказывают равное влияние на формирование ее распределения. Теперь такое «равноправие» случайных величин-слагаемых может исчезнуть: случайные величины, обладающие большей дисперсией, могут оказывать большее влияние на суммарное распределение, препятствуя возникновению его в стандартной форме (в силу центрированности суммы, различие математических ожиданий такое влияние на суммарное распределение не оказывает). Интуитивные соображения подсказывают, что указанный нежелательный эффект можно устранить, если выдвинуть требование, ограничивающее внос в общую (суммарную) дисперсию со стороны случайных величин-слагаемых на «хвостах» их распределений.

Введем обозначения (используемые и далее):

$$a_k = MX_k, b_k^2 = DX_k, B_n^2 = \sum_{k=1}^n b_k^2 = D\left(\sum_{k=1}^n X_k\right),$$

и выразим сформулированное требование *условием Линдберга*, имеющим вид

$$\forall \tau, \tau > 0, \frac{\sum_{k=1}^n \int_{|x-a_k| > \tau B_n} (x-a_k)^2 dF_k(x)}{B_n^2} \xrightarrow{n \rightarrow \infty} 0, \quad (8.30)$$

где $F_k(x)$ – функция распределения случайной величины X_k .

Справедлива следующая теорема.

Теорема 8.8 (Линдеберга). Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность независимых в совокупности случайных величин с конечными математическими ожиданиями и дисперсиями

$$a_k = MX_k, \quad b_k^2 = DX_k$$

и выполняется условие (8.30). Тогда имеет место равномерная по x сходимост

$$P\left\{\frac{1}{B_n} \sum_{k=1}^n (X_k - a_k) < x\right\} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz. \quad (8.31)$$

Приведем доказательство этой теоремы (придерживаясь его изложения в [3]).

Обозначая

$$X_{nk} = \frac{X_k - a_k}{B_n}, \quad F_{nk}(x) = P\{X_{nk} < x\}, \quad S_n^* = \frac{1}{B_n} \sum_{k=1}^n (X_k - a_k) = \sum_{k=1}^n X_{nk}, \text{ имеем}$$

$$MX_{nk} = 0, \quad DX_{nk} = \frac{DX_k}{B_n^2}, \quad \sum_{k=1}^n DX_{nk} = 1.$$

Тогда (8.30) примет вид

$$\forall \tau, \tau > 0, \quad \sum_{k=1}^n \int_{|x| > \tau} x^2 dF_{nk}(x) \xrightarrow{n \rightarrow \infty} 0. \quad (8.32)$$

Обозначим, далее, характеристические функции случайных величин X_{nk} и суммы S_n^* соответственно $\varphi_{nk}(t)$ и $\varphi_{S_n^*}(t)$. Ввиду независимости случайных величин $\{X_k\}$, а следовательно и случайных величин $\{X_{nk}\}$, имеет место равенство

$$\varphi_{S_n^*}(t) = \prod_{k=1}^n \varphi_{nk}(t). \quad (8.33)$$

Наша задача состоит в доказательстве справедливости предельного соотношения

$$\varphi_{S_n^*}(t) \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}}. \quad (8.34)$$

Установим, что все характеристические функции $\varphi_{nk}(t)$ в (8.33) для каждого фиксированного t при $n \rightarrow \infty$ равномерно относительно k ($1 \leq k \leq n$) стремятся к 1. Для этого выпишем очевидное равенство

$$\varphi_{nk}(t) - 1 = \int (e^{itx} - 1 - itx) dF_{nk}(x)$$

(интеграл для третьего слагаемого в правой части равен нулю ввиду равенства $MX_{nk} = 0$).

Используя известное неравенство, справедливое для любого вещественного α :

$$|e^{i\alpha} - 1 - i\alpha| \leq \frac{\alpha^2}{2}, \quad (8.35)$$

получим

$$|\varphi_{nk}(t) - 1| \leq \frac{t^2}{2} \int x^2 dF_{nk}(x). \quad (8.36)$$

Для $\forall \varepsilon, \varepsilon > 0$, очевидно соотношение

$$\int x^2 dF_{nk}(x) = \int_{|x| \leq \varepsilon} x^2 dF_{nk}(x) + \int_{|x| > \varepsilon} x^2 dF_{nk}(x) \leq \varepsilon^2 + \int_{|x| > \varepsilon} x^2 dF_{nk}(x),$$

в котором последнее слагаемое согласно (8.32) при достаточно большом n может быть сделано меньше ε^2 . Поэтому для всех достаточно больших n равномерно относительно k ($1 \leq k \leq n$) при $|t| \leq T$ (T – любой конечный интервал) получим

$$|\varphi_{nk}(t) - 1| \leq \varepsilon^2 T^2,$$

т.е. $\varphi_{nk}(t) \xrightarrow{n \rightarrow \infty} 1$ равномерно по k .

В частности, для $|t| \leq T$ и при достаточно больших n выполняется неравенство

$$|\varphi_{nk}(t) - 1| < \frac{1}{2} \quad (8.37)$$

(ниже мы будем постоянно полагать это неравенство выполненным).

Из (8.33) для главного значения логарифма $\ln \varphi_{S_n}(t)$ получим

$$\ln \varphi_{S_n}(t) = \sum_{k=1}^n \ln \varphi_{nk}(t) \equiv \sum_{k=1}^n \ln(1 + (\varphi_{nk}(t) - 1))$$

или, представляя слагаемые правой суммы рядами,

$$\ln \varphi_{S_n}(t) = \sum_{k=1}^n (\varphi_{nk}(t) - 1) + R_n \quad (8.38)$$

где

$$R_n = \sum_{k=1}^n \sum_{j=2}^{\infty} \frac{(-1)^{j-1}}{j} (\varphi_{nk}(t) - 1)^j.$$

Нетрудно проверить соотношения (с учетом (8.37)):

$$|R_n| \leq \sum_{k=1}^n \sum_{j=2}^{\infty} \frac{1}{2} |\varphi_{nk}(t) - 1|^j \leq \frac{1}{2} \sum_{k=1}^n \frac{|\varphi_{nk}(t) - 1|^2}{1 - |\varphi_{nk}(t) - 1|} \leq \sum_{k=1}^n |\varphi_{nk}(t) - 1|^2.$$

Далее, ввиду (8.36), получаем

$$|R_n| \leq \max_k |\varphi_{nk}(t) - 1| \sum_{k=1}^n |\varphi_{nk}(t) - 1| \leq \frac{t^2}{2} \max_k |\varphi_{nk}(t) - 1| \sum_{k=1}^n \int x^2 dF_{nk}(x) = \frac{t^2}{2} \max_k |\varphi_{nk}(t) - 1| \xrightarrow{n \rightarrow \infty} 0$$

и из (8.38) следует

$$\lim_{n \rightarrow \infty} \ln \varphi_{S_n}(t) = \lim_{n \rightarrow \infty} \sum_{k=1}^n (\varphi_{nk}(t) - 1).$$

Выпишем формально следующее очевидное равенство

$$\sum_{k=1}^n (\varphi_{nk}(t) - 1) = -\frac{t^2}{2} + \frac{t^2}{2} + \sum_{k=1}^n \int (e^{itx} - 1 - itx) dF_{nk}(x) = -\frac{t^2}{2} + \rho_n$$

$$\rho_n = \frac{t^2}{2} + \sum_{k=1}^n \int (e^{itx} - 1 - itx) dF_{nk}(x) \xrightarrow{n \rightarrow \infty} 0.$$

Для этого, наряду с (8.35), используем подобное соотношение, справедливое для любого вещественного α :

$$\left| e^{i\alpha} - 1 - i\alpha + \frac{\alpha^2}{2} \right| \leq \left| \frac{\alpha^3}{6} \right|,$$

и выполним цепочку преобразований (для произвольного ε)

$$\begin{aligned} |\rho_n| &= \left| \frac{t^2}{2} + \sum_{k=1}^n \int (e^{itx} - 1 - itx) dF_{nk}(x) \right| = \\ &= \left| \sum_{k=1}^n \int_{|x| \leq \varepsilon} (e^{itx} - 1 - itx - \frac{(itx)^2}{2}) dF_{nk}(x) + \sum_{k=1}^n \int_{|x| > \varepsilon} (e^{itx} - 1 - itx - \frac{(itx)^2}{2}) dF_{nk}(x) \right| \leq \\ &\leq \frac{|t^3|}{6} \sum_{k=1}^n \int_{|x| \leq \varepsilon} |x^3| dF_{nk}(x) + t^2 \sum_{k=1}^n \int_{|x| > \varepsilon} x^2 dF_{nk}(x) \leq \\ &\leq \frac{|t^3|}{6} \varepsilon \sum_{k=1}^n \int_{|x| \leq \varepsilon} x^2 dF_{nk}(x) + t^2 \sum_{k=1}^n \int_{|x| > \varepsilon} x^2 dF_{nk}(x) = \frac{|t^3|}{6} \varepsilon + t^2 \left(1 - \frac{|t| \varepsilon}{6} \right) \sum_{k=1}^n \int_{|x| > \varepsilon} x^2 dF_{nk}(x). \end{aligned}$$

Согласно условию Линдеберга (в форме (8.32)) второе слагаемое в полученной сумме при любом ε может быть сделано сколь угодно малым при достаточно большом n , т.е. $|\rho_n| \xrightarrow{n \rightarrow \infty} 0$ и

$$\lim_{n \rightarrow \infty} \ln \varphi_{S_n^*}(t) = \lim_{n \rightarrow \infty} \sum_{k=1}^n (\varphi_{nk}(t) - 1) = -\frac{t^2}{2}.$$

Теорема доказана.

Нетрудно установить, что теорема 8.7 представляет собой частный случай теоремы 8.8, так как при одинаковом распределении случайных величин из последовательности $\{X_k\}_{k=1}^{\infty}$ и конечности их дисперсий выполняется условие Линдеберга. Это, однако, не лишает теорему 8.7 самостоятельного значения, поскольку проверка условий ее выполнения проще, чем проверка условия Линдеберга.

В смысле простоты проверки выполнения условий теоремы с теоремой Линдеберга может конкурировать приведенная ниже *теорема Ляпунова*.

Сохраняя прежние обозначения, заменим условие Линдеберга (8.30) *условием Ляпунова*, состоящим в возможности подобрать число $\delta > 0$, при котором имеет место сходимость

$$\frac{\sum_{k=1}^n M |X_k - a_k|^{2+\delta}}{B_n^{2+\delta}} \xrightarrow{n \rightarrow \infty} 0. \quad (8.39)$$

Справедлива теорема.

Теорема 8.9 (Ляпунова). Пусть $\{X_k\}_{k=1}^{\infty}$ — последовательность независимых в совокупности случайных величин с конечными математическими ожиданиями и дисперсиями $a_k = MX_k, b_k^2 = DX_k$ и выполняется условие (8.39). Тогда имеет место равномерная по x сходимость

$$P\left\{\frac{1}{B_n} \sum_{k=1}^n (X_k - a_k) < x\right\} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

Доказательство этой теоремы сводится к проверке очевидных соотношений, показывающих, что выполнение условия (8.39) влечет за собой выполнение условия (8.30):

$$\begin{aligned} \frac{\sum_{k=1}^n \int_{|x-a_k| > \tau B_n} (x-a_k)^2 dF_k(x)}{B_n^2} &= \frac{\sum_{k=1}^n \int_{|x-a_k| > \tau B_n} \tau^\delta B_n^\delta (x-a_k)^2 dF_k(x)}{B_n^2 \tau^\delta B_n^\delta} \leq \frac{\sum_{k=1}^n \int_{|x-a_k| > \tau B_n} |x-a_k|^{2+\delta} dF_k(x)}{\tau^\delta B_n^{2+\delta}} \leq \\ &\leq \frac{\sum_{k=1}^n \int |x-a_k|^{2+\delta} dF_k(x)}{\tau^\delta B_n^{2+\delta}} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Пусть, например, для всех X_i существуют абсолютные центральные моменты третьего порядка $c_k^3 = M|X_k - a_k|^3$; обозначим $C_n^3 = \sum_{k=1}^n c_k^3$. Тогда условие Ляпунова выполняется, если

$\frac{C_n}{B_n} \xrightarrow{n \rightarrow \infty} 0$ (что соответствует $\delta = 1$). Отметим относительную простоту проверки этого (достаточного) условия выполнения предельной теоремы.

8.3.3. Предельные теоремы с управляющим параметром

Рассмотренные нами предельные теоремы обычно именуются *классическими*, играя фундаментальную роль при формировании прикладных стохастических моделей. Вместе с тем, реальные задачи могут подчас приводить к необходимости той или иной модификации этих теорем, отражающей условия практики.

Так, в приведенных теоремах число слагаемых случайных величин задается неслучайным целым числом, принимающим детерминированные значения $n = 1, 2, \dots$. На практике, однако, число слагаемых может оказаться случайной величиной с распределением, зависящим от значения некоторого параметра, выбираемого исследователем. Самым простым примером такой ситуации является зависимость числа покупок, сделанных посетителями магазина за день, от цены товара.

Этот и подобные ему примеры приводят к некоторому классу предельных теорем, в которых фигурирует *управляющий параметр*, опосредованно определяющий те условия теоремы, которые ранее мы полагали заданными. Приведем пример такой теоремы.

Теорема 8.10. Пусть $\{X_k\}_{k=1}^N$ — последовательность случайного числа N одинаково распределенных независимых в совокупности случайных величин с конечными математическими ожиданиями $m = MX_k$ и дисперсиями $\sigma^2 = DX_k$; функция распределения случайной величины N зависит от управляющего параметра r : $F_N(n) = F_N(n; r)$, причем для $\forall n$ выполняется условие

$$\lim_{r \rightarrow \infty} F_N(n; r) = 0. \quad (8.40)$$

Тогда для функции распределения $F_{S_N}(s; r)$ суммы

$$S_N = \frac{\sum_{k=1}^N X_k - Nm}{\sigma \sqrt{N}} \quad (8.41)$$

имеет место сходимост

$$\forall s: F_{S_N}(s; r) \xrightarrow{r \rightarrow \infty} \Phi^*(s).$$

Введем обозначение

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

и выпишем следующие очевидные равенства (пока для произвольных фиксированных r и \bar{n}):

$$\begin{aligned} F_{S_N}(s) &= P\{S_N < s\} = \sum_{n=1}^{\infty} P\{S_N < s | N = n\} P\{N = n; r\} = \\ &= \sum_{n < \bar{n}} P\{S_N < s | N = n\} P\{N = n; r\} + \sum_{n \geq \bar{n}} P\{S_N < s | N = n\} P\{N = n; r\} = \\ &= A_{\bar{n}} + \sum_{n \geq \bar{n}} F_{S_n}(s) P\{N = n; r\} = A_{\bar{n}} + \sum_{n \geq \bar{n}} [\Phi^*(s) + \varepsilon(n)] P\{N = n; r\} = \\ &= A_{\bar{n}} + \Phi^*(s)(1 - P\{N < \bar{n}; r\}) + B_{\bar{n}}, \end{aligned}$$

где

$$\begin{aligned} A_{\bar{n}} &= \sum_{n < \bar{n}} P\{S_N < s | N = n\} P\{N = n\}; \\ B_{\bar{n}} &= \sum_{n \geq \bar{n}} \varepsilon(n) P\{N = n\}, \end{aligned}$$

$\varepsilon(n) = F_{S_n}(s) - \Phi^*(s)$ — отклонение $F_{S_n}(s)$ от стандартной нормальной функции распределения $\Phi^*(s)$ (там где это полезно, подчеркнута зависимость вероятности от параметра r). Согласно предельной теореме (в её классической форме) $|\varepsilon(n)| \rightarrow 0$ при $n \rightarrow \infty$ и ввиду очевидных оценок

$$\begin{aligned} |A_{\bar{n}}| &\leq P\{N < \bar{n}\} = F_N(\bar{n}); \\ |B_{\bar{n}}| &\leq \sum_{n \geq \bar{n}} |\varepsilon(n)| P\{N = n\} < |\varepsilon(\bar{n})|, \end{aligned}$$

получаем

$$|F_{S_N}(s) - \Phi^*(s)| < F_N(\bar{n}; r) + |\varepsilon(\bar{n})|.$$

Для $\forall \varepsilon > 0$ можно выбрать \bar{n} таким, что бы выполнялось неравенство $|\varepsilon(n)| < \varepsilon/2$, а затем r , при котором $F_N(n; r) < \varepsilon/2$, что означает выполнение (8.41).

К полученному результату следует сделать несколько добавлений.

В теореме рассматривался управляющий параметр r , увеличение которого приводило к возрастанию числа слагаемых N , точнее говоря, к смещению функции распределения числа N вправо. Не составляет, однако, трудности рассмотрение и параметров, оказывающих на N обратное действие (так, в приведенном примере торговой операции управляющим параметром служит цена товара, повышение которой снижает спрос, сдвигая распределение N влево).

На практике условие (8.40) обычно не выполняется (что, впрочем, свойственно всем предельным теоремам). Причиной этого может быть как ограничение, налагаемое на сам управляющий параметр r , так и естественный эффект насыщения в зависимости распределения N от r (снова обращаясь к приведенному примеру, заметим, что снижение цены на товар не может быть безграничным и не влечет за собой неограниченный рост спроса). В связи с этим возникает задача оценки ошибки аппроксимации распределения суммы S_N нормальным распределением (т.е. задача, аналогичная рассмотренной в примере 8.4).

Для этого снова используем неравенство Берри – Эссеена в предположении существования конечного третьего абсолютного центрального момента $M|X - MX|^3$. Пусть \bar{r} – заданное значение управляющего параметра (выбранное из соображений указанных ограничений) и $F_N(n; \bar{r})$ – соответствующее ему распределение числа слагаемых N . Обозначив в (8.29) правую часть

$$\delta = \frac{CM|X - MX|^3}{\sigma^3 \sqrt{n}} = \frac{C\mu_3}{\sigma^3 \sqrt{n}}, \quad (\mu_3 \stackrel{\Delta}{=} M|X - MX|^3),$$

получим

$$n = \bar{n} = \left(\frac{C\mu_3}{\sigma^3 \delta} \right)^2,$$

откуда следует, что для того, чтобы максимальная ошибка аппроксимации не превышала заданную величину δ , достаточно выполнение условия

$$n \geq \bar{n} = \left(\frac{0,8\mu_3}{\sigma^3\delta} \right)^2, \quad (8.42)$$

вероятность которого зависит от параметра r и равна

$$P \left\{ N \geq \left(\frac{0,8\mu_3}{\sigma^3\delta} \right)^2 \right\} = 1 - F_N \left[\left(\frac{0,8\mu_3}{\sigma^3\delta} \right)^2; r \right] = \alpha(r). \quad (8.43)$$

При известных оценках моментов распределения случайных величин X_k и функции распределения $F_N(n; r)$ полученное соотношение позволяет находить значения управляющего параметра, для которых модуль отклонения распределения $F_{S_N}(s; r)$ от предельного стандартного нормального распределения $\Phi^*(s)$ не превышает значения δ с вероятностью не ниже заданного значения α .

Приведем пример такого расчета.

Пример 8.5. Предположим, что число слагаемых N в сумме (8.41) – случайная величина, имеющая распределение Пуассона $N \in \text{Po}(a)$, равная числу событий в интервале времени $(0, \tau)$, т.е. где $a = \lambda\tau$ и $\lambda = r$ – управляющий параметр, значение которого определяет интенсивность потока событий в интервале $(0, \tau)$.

Используя равенство (см. п.4.3.5)

$$1 - F_{\chi^2}(2a; 2k) = \sum_{r=0}^{k-1} \frac{a^r}{r!} e^{-a},$$

получим

$$F_N(n; \lambda) = \sum_{j=0}^{n-1} \frac{(\lambda\tau)^j}{j!} e^{-\lambda\tau} = 1 - F_{\chi^2}(2\lambda\tau; 2n) \xrightarrow{\lambda \rightarrow \infty} 0, \quad (8.44)$$

т.е. (8.40) выполняется. Пусть, далее, случайные величины X_k имеют равномерное распределение на отрезке $[0, 1]$ и, следовательно, моменты $m = \frac{1}{2}$, $\sigma^2 = \frac{1}{12}$, $\mu_3 = \frac{1}{32}$. Отсюда

(см.(8.42)) получаем $\bar{n} \cong \frac{1}{\delta^2}$ и (8.43) принимает вид (с учетом (8.44))

$$P \left\{ N \geq \frac{1}{\delta^2} \right\} = 1 - F_N \left[\frac{1}{\delta^2}; r \right] = F_{\chi^2} \left(2\lambda\tau; \frac{2}{\delta^2} \right) = \alpha.$$

Пусть, к примеру, $\tau = 10$ и заданы значения $\delta = 0.1$, $\alpha = 0.9$. Тогда из последнего равенства следует, что минимальное значение управляющего параметра, начиная с которого заданная точность аппроксимации δ выполняется с заданной вероятностью α , составляет $\lambda = 12$.

В условиях теоремы 8.10 предполагалось, что параметры распределения случайных величин X_k не зависят от управляющего параметра r , что, однако, не всегда соответствует действительности. Так, например, в торговой операции цена на товар не только влияет на его

спрос, но и определяет среднее значение и дисперсию получаемой от его продажи прибыли, т.е. случайной величины X_k . Не останавливаясь подробно на описании модели, учитывающей такую зависимость, и на соответствующей ей параметрической предельной теореме, отметим лишь возникающие здесь особенности.

В связи с зависимостью $\sigma = \sigma(r)$, $\mu_3 = \mu_3(r)$, соотношение (8.29) принимает теперь вид

$$n \geq \bar{n}(r) = \left(\frac{0,8\mu_3(r)}{\sigma^3(r)\delta} \right)^2,$$

из которого следует

$$P\{\sup_s |F_{S_n}(s) - \Phi^*(s)| \leq \delta\} = P\{N \geq \bar{n}(r); r\} = 1 - F_N(\bar{n}(r); r).$$

Сравнивая полученное равенство с (8.43), можно заключить, что в новой постановке выполнение предельной теоремы связано с зависимостью от r не только вида функции распределения $F_N(\cdot)$, но и значений ее аргумента.

Завершая рассмотрение предельных теорем для распределений сумм независимых случайных величин, заметим, что все приведенные выше теоремы относятся к так называемым *интегральным* предельным теоремам, поскольку в них предельные распределения представлены функцией нормального распределения. Наряду с широко применяемыми на практике предельными теоремами этого класса, в некоторых задачах теории вероятностей используются *локальные* предельные теоремы, в которых предельное распределение представлено плотностью нормального распределения. С предельными теоремами этого класса читатель может ознакомиться в [2], [3], [7].

Список литературы

- Андерсон Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963. – 500 с.
- Боровков А.А. Теория вероятностей. – М.: УРСС, 2003. – 472 с.
- Гнеденко Б.В. Курс теории вероятностей. – М.: УРСС, 2005. – 448 с.
- Журавлев Ю.И., Флеров Ю.А. Дискретный анализ, ч.1: Учеб. пособие. – М.: МФТИ, 1999. – 136 с.
- Натан А.А., Горбачев О.Г., Гуз С.А. Математическая статистика: Учеб. пособие. – М.: МЗ Пресс - МФТИ, 2004. – 158 с.
- Натан А.А., Горбачев О.Г., Гуз С.А. Основы теории случайных процессов: Учеб. пособие. – М.: МЗ Пресс – МФТИ, 2003. – 168 с.
- Севастьянов Б.А. Курс теории вероятностей и математической статистики. – М.: РХД, 2004. – 272с.
- Ширяев А.Н. Вероятность-1, Вероятность-2. – М.: МЦНМО, 2007. – 552 + 416 с.
- Яковлев Г.Н. Лекции по математическому анализу. ч.2. – М.: Изд-во МФТИ, 1994. - 480 с.
- Петров В.В. Пределные теоремы для сумм независимых случайных величин. - М.: Физматгиз, 1987. – 317 с.
- Стоянов Й. Контрпримеры в теории вероятностей. – М.: Факториал, 1999. – 288 с.
- Секей Г. Парадоксы в теории вероятностей и математической статистике. – М.: РХД, 2003. – 272 с.

Издательство «МЗ Пресс»
приглашает к сотрудничеству авторов
и книготорговые организации
Тел. (495)156-84-82
Почтовый адрес: 125183, г. Москва, пр. Черепановых, д. 56
ООО «У СЫтина», для МЗ Пресс
e-mail: mzpress@mail.ru

А.А.Натан, О.Г.Горбачев, С.А.Гуз

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Учебное пособие
Издание в авторской редакции

Москва, МЗ Пресс, 2007

ИЗДАТЕЛЬ З.А. Отарашвили.
КОМПЬЮТЕРНАЯ ВЕРСТКА Б.И. Оводов

«МЗ Пресс», 125299, г. Москва, ул. Приорова, д. 2А
Подписано в печать 07.09.2007 Формат 60×84/16
Печать офсетная. Бумага офсетная. Усл.- печ. 16 л.
Тираж 1000 экз. Заказ №

Отпечатано в ОАО «Калужская типография стандартов»
248021, г. Калуга, ул. Московская, д. 256