

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МЕХАНИКО–МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

кафедра теории вероятностей и математической статистики

И. С. БОРИСОВ

ЛЕКЦИИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Новосибирск – 2010

А Н Н О Т А Ц И Я

Учебное пособие подготовлено по материалам 24 лекций курса «Теория вероятностей», который на протяжении ряда лет читался автором на механико-математическом факультете НГУ (отделение «Математика», 5-й семестр). В него вошли классические разделы как комбинаторной, так и абстрактной теории вероятностей, которые традиционно освещаются при чтении подобных курсов. Некоторые утверждения приводятся в редакции, допускающей более простые и короткие доказательства по сравнению с известными. По ходу доказательств часть возникающих вопросов в виде упражнений отнесена читателю. Эти упражнения могут быть использованы для практических занятий по данному курсу.

Учебное пособие подготовлено в рамках реализации Программы развития НИУ—НГУ.

ВВЕДЕНИЕ

Теория вероятностей — раздел математики, посвященный исследованию закономерностей, возникающих в длинных сериях однородных стохастических экспериментов. Под *стохастическим экспериментом (испытанием)* мы будем понимать некое действие, результатом которого могут быть по меньшей мере два исхода, при этом ни один из них предсказать наверняка заранее невозможно. Каждый стохастический эксперимент определяется комплексом условий или, другими словами, *стохастическим алгоритмом*. Так что говоря об «однородных» испытаниях, мы имеем в виду многократное воспроизведение одного и того же стохастического алгоритма.

Показательным примером стохастического испытания является подбрасывание *симметричной* монеты, т. е. априори у вас нет никакой информации о приоритете той или иной ее стороны. Если бросать монету с интенсивным вращением на высоту двух метров с отскоком от упругого пола (тем самым, мы описали «комплекс условий» или «стохастический алгоритм» проводимого эксперимента), то предсказать заранее наверняка исход такого действия, очевидно, невозможно.

Рассмотрим эксперимент с подбрасыванием монеты подробнее. Проведем серию из n испытаний, каждое из которых заключается в бросании монеты в одних и тех же описанных выше условиях. Будем считать испытание «успешным», если выпал «орёл». Число «успехов» в серии из n испытаний обозначим через S_n и определим *частоту* ν_n появления «успеха» в серии из n испытаний как $\nu_n = S_n/n$. Поразительным является тот факт, что при возрастании числа испытаний в серии частота ν_n начинает стабилизироваться в окрестности точки $1/2$. Иначе говоря, при $n \rightarrow \infty$ величина ν_n сходится (в известном смысле, который далее будет пояснен) к величине $1/2$. Этот факт является иллюстрацией ***всемирного эмпирического закона стабилизации частот***: для любого стохастического эксперимента с двумя исходами существует $\lim_{n \rightarrow \infty} \nu_n$ (не обязательно равный $1/2$). Или в более общей постановке: для широкого класса стохастических экспериментов среднее арифметическое суммарного результата стабилизируется с ростом числа однородных испытаний. Слово «всемирный» подчеркивает всеобщность описанного феномена, т. е. отсутствие «контр-примеров» стохастических экспериментов, исходы которых, скажем, априори ограничены некоторым числом, для которых среднее арифметическое наблюдений не стабилизировалось бы с ростом числа воспроизведений данного эксперимента.

Наглядного физического объяснения отмеченного феномена стабилизации не существует. Если сказать коротко, то теорию вероятностей как раз и интересуют указанные точки стабилизации, точнее, способы их вычисления при выполнении тех или иных аксиом касательно проводимого стохастического эксперимента. Адекватность реальным стохастическим экспериментам тех или иных вероятностных моделей, определяемых соответствующей системой аксиом, может быть установлена только опытным путем.

Рассмотрим следующий

Пример (игра «Спортлото — 5 из 36»). Имеется карточка с 36 натуральными числами. Наудачу зачёркиваем 5 из них. После этого происходит розыгрыш пяти «счастливых» чисел. Денежный приз тем существеннее, чем больше совпало зачёркнутых и «счастливых» чисел. Ясно, что угадать все 5 чисел весьма затруднительно. Поэтому посмотрим, насколько возможно угадать ровно 3 номера. Приведём конкретные циф-

ры, заимствованные с сайта «Спортлото». Возьмём результаты наудачу выбранных 10 розыгрышей. В каждом из них участвовало n карточек (для каждого розыгрыша приблизительно $n \approx 24$ тысячи). Три из пяти чисел угадали S_n человек. Значения частот $\nu_n = S_n/n$ оказались такими:

0,0121; 0,0178; 0,0110; 0,0091; 0,0120; 0,0112; 0,0130; 0,0132; 0,0122; 0,0127.

Теперь объединим результаты всех 10 розыгрышей. Тогда имеется всего $\tilde{n} = 241172$ карточек, причем три номера угадали $S_{\tilde{n}} = 3026$ человек, откуда $\nu_{\tilde{n}} = 0,0125$. Скоро мы научимся считать точку стабилизации в этой задаче (по формуле гипергеометрического распределения). Пока лишь приведем ответ:

$$p = \frac{C_5^3 \cdot C_{31}^2}{C_{36}^5} \approx 0,0123.$$

Заметим, что величины p и $\nu_{\tilde{n}}$ отличаются всего лишь на 0,0002, в то время как частоты в отдельных тиражах, приведенные выше (где количество карточек, т. е. число проведенных стохастических экспериментов, примерно в десять раз меньше), более значимо отличаются от «истинной» вероятности p . Это и есть иллюстрации отмеченной выше стабилизации частоты при увеличении числа наблюдений. Таким образом, при достаточно большом числе карточек (или, что то же самое, однородных стохастических экспериментов) $\nu_n = S_n/n \approx p$, откуда $S_n \approx np$. Поэтому ещё задолго до проведения розыгрыша, принимая во внимание всемирный закон стабилизации частот, можно посчитать прибыль лотереи при той или иной стоимости карточек.

Глава 1. КОМБИНАТОРНАЯ ВЕРОЯТНОСТЬ

Модель классической вероятности.

Для расчёта точек стабилизации нам придётся строить математические модели стохастических испытаний. Эти модели будут описываться набором аксиом, соответствующих экспериментальным данным. Введем аксиоматику простейшей модели — *модели классической вероятности*. Результаты каждого испытания описывается элементом так называемого *пространства элементарных исходов*, природа которого может быть самой разнообразной.

Аксиомы классической вероятностной модели

1. Пространство элементарных исходов Ω *конечно*.
2. Все элементы (исходы) $\omega_i \in \Omega$ *равновозможны* (симметричны), то есть нет никаких оснований предпочесть один исход другому, они все «бесприоритетны».
3. Под *событием* A понимаем любое подмножество пространства элементарных исходов: $A \subseteq \Omega$. *Вероятностью* события A (обозначаем $\mathbf{P}(A)$) называется отношение числа благоприятных исходов к числу всех исходов:

$$\mathbf{P}(A) = \frac{\#(A)}{\#(\Omega)},$$

где $\#(\cdot)$ — число элементов множества (*считающая мера*).

Отметим очевидные свойства введенной вероятности. Во-первых, $\mathbf{P}(\Omega) = 1$, а во-вторых, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ при условии, что A и B — *несовместные* (т. е. непересекающиеся) события. Событие $\bar{A} = \Omega \setminus A$ будем называть *дополнительным* к событию A . Так как $A \cup \bar{A} = \Omega$ и $A \cap \bar{A} = \emptyset$, то в силу аддитивности вероятности $\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A})$. Эта формула полезна в том случае, когда вероятность дополнительного события вычисляется проще, нежели исходного.

Мы можем *априори* (до проведения эксперимента) рассчитать $\mathbf{P}(A)$, если стохастический эксперимент удовлетворяет первым двум аксиомам. Обычно для проверки второй аксиомы используют так называемые «соображения симметрии». Частоту ν_n можно рассчитать только после проведения испытания. По анонсированному выше эмпирическому закону стабилизации частот при большом числе испытаний должно выполняться соотношение $\nu_n \approx \mathbf{P}(A)$. Если этого не происходит, то, скорее всего, мы неверно построили математическую модель (здесь центральное место занимает обоснование правомерности использования аксиомы 2).

Пример. Подбрасываем с вращением на высоту двух метров от пола горсть из 5 монет. С каждой монетой мы связываем «успех» (скажем, число «1») или «неудачу» (пусть это «0»). Опишем пространство элементарных исходов. Для начала мы заменим наш эксперимент (т. е. проведём *редукцию* задачи) на эквивалентный исходному, производя пятикратное подбрасывание одной монеты.

Если положить $\Omega = \{0, 1, 2, 3, 4, 5\}$ (по количеству выпавших «единиц»), то мы не получим классическую схему, ибо элементарные исходы (элементы Ω) в этом случае, очевидно, несимметричны: пять единиц может выпасть только одним способом расположения монет, в то время как одна единица — целыми пятью! Ясно, что в силу симметричности монет и отсутствия между ними каких-либо видимых связей у нас нет никаких оснований предпочесть ту или иную комбинацию «успехов» и «неудач» лежащих на полу монет при условии, что все они различимы (мы можем на них написать номера!). Поэтому для нас все такие комбинации равновозможны. Итак, в данном случае не выполняется аксиома 2 классической модели. Для приведенного выше конкретного примера один «успех» при пяти бросаниях монеты (не важно, на каком шаге произошел этот «успех»!) будет наблюдаться в пять раз чаще, нежели комбинация из пяти успехов.

Рассмотрим другой вариант. Именно, пусть $\Omega = \{(x_1, x_2, x_3, x_4, x_5) : x_i \in \{0, 1\}\}$. Нетрудно понять, что здесь все исходы (векторы длины 5, каждая компонента которых 0 или 1) равновозможны. Их количество есть $\#(\Omega) = 2^5 = 32$.

Таким образом, мы приходим к выводу, что в модели классической вероятности очень важно правильно задать пространство элементарных исходов.

Теперь посчитаем вероятность того, что будет хотя бы одна «1», то есть вероятность события $A = \{\omega \in \Omega : \text{вектор } \omega \text{ содержит хотя бы одну «1»}\}$. Воспользуемся уже упомянутой формулой перехода к дополнительному событию.

У нас $\bar{A} = \{\omega \in \Omega : \text{число «единиц» в векторе } \omega \text{ равно } 0\}$. Тогда $\#(\bar{A}) = 1$, значит, $\mathbf{P}(\bar{A}) = 1/32$, откуда $\mathbf{P}(A) = 31/32$.

Упражнение. Найти вероятность того, что при подбрасывании пяти симметричных монет число успехов превзойдет число неудач.

Модель дискретной вероятности.

Классическая вероятностная модель имеет весьма ограниченную область применимости. Чтобы понять это, рассмотрим

Пример. Два человека играют в такую игру: они по очереди подкидывают монету, а выигрывает тот, у кого выпадет «орёл». Задача: описать пространство элементарных исходов и сравнить шансы на победу у каждого игрока. Ясно, что классической моделью здесь пользоваться нельзя, так как пространство исходов перестает быть конечным. Оказывается, эта задача решается в рамках модели со следующей аксиоматикой.

Аксиомы дискретной вероятностной модели.

1. Пространство Ω элементарных исходов *не более чем счётно*.
2. *Дискретная вероятность* — это набор положительных чисел $\{p_i\}$, где p_i называется вероятностью исхода ω_i , при выполнении условия нормировки $\sum_{\omega_i \in \Omega} p_i = 1$.

Вероятность события A определяется как сумма вероятностей благоприятных исходов:

$$\mathbf{P}(A) = \sum_{\omega_i \in A} p_i.$$

Отметим, что и в этой модели вероятность, очевидно, обладает свойствами, отмеченными в классической модели — аддитивностью и нормировкой. Ясно, что классическая модель включается в дискретную при $p_i = \frac{1}{\#(\Omega)}$.

Теперь вернёмся к нашему примеру. Если «орёл» — это 1, «решка» — 0, то пространство элементарных исходов имеет вид $\Omega = \{(1), (01), (001), (0001), \dots\}$. Таким образом, Ω счётно. Посчитаем вероятность p_i i -го исхода. Берём вектор из i компонент: $(0 \dots 01)$. Ему соответствует i -кратное подбрасывание монеты. Согласно классической модели, в которой 0 или 1 могут при каждом подбрасывании появиться с вероятностью $1/2$, вероятность рассматриваемого исхода $(\underbrace{0 \dots 0}_i 1)$ равна $p_i = 1/2^i$. Элементарно прове-

ряется, что условие нормировки выполнено: $\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$. Итак, мы находимся в пределах применимости дискретной модели. Определим шансы на победу каждого игрока. Пусть событие A_1 и A_2 — выигрыши первого и второго игроков соответственно. Тогда $A_1 = \{(1), (001), (00001), \dots\}$. Так как $A_2 = \{(01), (0001), \dots\}$ и $A_2 = \bar{A}_1$, то в силу аддитивности меры достаточно посчитать вероятность только одного из этих событий: $\mathbf{P}(A_1) + \mathbf{P}(\bar{A}_1) = \mathbf{P}(A_1) + \mathbf{P}(A_2) = 1$. Имеем

$$\mathbf{P}(A_1) = \sum_{i \geq 1} \frac{1}{2^{2i-1}} = \frac{2}{3}.$$

Следовательно, преимущество первого хода у игрока под номером 1 даёт ему вдвое больше шансов победить.

Упражнение. В предыдущей задаче сравнить вероятности победить для каждого игрока в случае, когда играют трое.

Модель геометрической вероятности.

Бывают случаи, когда введенными выше двумя моделями обойтись нельзя. Рассмотрим стохастический эксперимент «стрельба по мишени». По условиям этого эксперимента стреляющий попадает в квадрат размером, скажем, 10 на 10 см. Если считать, что внутри квадрата нет приоритетных зон, то по сути происходит бросание точки *наудачу* внутрь квадрата. В нашей математической модели мы приходим к необходимости считать все точки этого квадрата элементарными исходами, т. е. пространством Ω . Но это множество Ω *несчетно* (имеет мощность континуума). С другой стороны, заметим, что при отсутствии приоритетных зон внутри Ω из соображений симметрии естественно считать, что вероятность попасть в фиксированную половину квадрата есть $1/2$, в фиксированную четверть — $1/4$ и т. д. Значит, если мы рассматриваем разбиение Ω на квадраты со стороной $1/2^n$, то вероятность попасть в любой из них есть $1/2^n$. Отсюда естественным образом можно перейти и к любым измеримым по Жордану множествам, т. е. для которых площадь теоретико-множественной разности объединения попарно непересекающихся квадратиков, содержащих данное множество, и аналогичного объединения, содержащегося в данном множестве, может быть сделана сколь угодно малой. Более того, мы можем перейти к более широкому классу подмножеств — измеримых в смысле Лебега. Тогда на указанном пространстве элементарных исходов для любого измеримого по Лебегу подмножества A его *нормированную меру Лебега* естественно взять в качестве вероятности попадания в него при бросании точки наудачу в Ω :

$$\mathbf{P}(A) = \frac{\Lambda_2(A)}{\Lambda_2(\Omega)},$$

где Λ_2 — двумерная мера Лебега.

Теперь мы переходим к описанию аксиоматики третьей вероятностной модели.

Аксиомы геометрической вероятностной модели.

1. Ω — измеримое по Лебегу множество в R^n с конечной мерой Лебега: $\Lambda_n(\Omega) < \infty$.
2. $\mathcal{F} = \{A\}$ — все измеримые по Лебегу подмножества Ω .
3. Отсутствие приоритетных зон (измеримых подмножеств ненулевой меры) элементарных исходов.
4. Вероятность события A определяется по формуле

$$\mathbf{P}(A) = \frac{\Lambda_n(A)}{\Lambda_n(\Omega)}.$$

Эта модель — континуальный аналог классической вероятности.

Пример («задача о встрече»). Два лица договорились встретиться между 23 и 24 часами. При этом моменты их прихода априори выбираются наудачу, т. е. не имеют приоритета в указанном часовом интервале, и никак не связаны друг с другом. Имеется договоренность этих лиц: первый пришедший ждет второго не более 15 минут, иначе встреча не состоится. Найти вероятность того, что встреча произойдет.

Решение: Редукция задачи: на отрезок $[0, 1]$ наудачу бросаются две точки; пусть x и y - координаты первой и второй точек (соответственно время прихода первого и второго лица). Тогда, полагая $\omega = (x, y)$, приходим к двумерной модели геометрической вероятности с единичным квадратом в качестве Ω , поскольку из условия задачи можно заключить, что приоритетных зон в Ω нет. Множество благоприятных исходов есть $A = \left\{ (x, y) : |x - y| \leq \frac{1}{4} \right\}$. Тогда $\Lambda_2(\Omega) = 1$ и $\Lambda_2(A) = 1 - \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot 2 = \frac{7}{16}$. Следовательно, $\mathbf{P}(A) = \frac{\Lambda_2(A)}{\Lambda_2(\Omega)} = \frac{7}{16}$.

Упражнение. Решить эту задачу для трёх участников встречи со временем ожидания в 15 минут для прибывшего первым двух других участников.

Упражнение. Имеются 3 спички единичной длины. Наудачу рассекаем каждую из спичек на 2 части. Из шести полученных частей наудачу выбираем три. Какова вероятность, что из этих частей можно сложить треугольник? Остроугольный треугольник?

Основная проблема при выборе той или иной модели геометрической вероятности при решении конкретных задач состоит в обосновании отсутствия приоритетных зон внутри пространства элементарных исходов. В качестве иллюстрации сказанного рассмотрим следующую задачу стохастической геометрии:

Окружность единичного радиуса наудачу рассекается прямой на две дуги (обозначим через l меньшую дугу). Найти вероятность события A , состоящего в том, что длина $|l|$ меньшей дуги не превосходит $1/3$ длины окружности, то есть найти $\mathbf{P}(|l| \leq 2\pi/3)$.

Решение: В качестве элементарного исхода описанного стохастического эксперимента рассмотрим величину центрального угла α , опирающегося на меньшую дугу. Ясно, что $\Omega = [0, \pi]$, а множество благоприятных исходов есть $A = [0, 2\pi/3]$. Тогда согласно одномерной модели геометрической вероятности $\mathbf{P}(|l| \leq 2\pi/3) = 2/3$.

Также в качестве элементарного исхода можно взять расстояние h от центра круга до секущей хорды. Тогда $h \in [0, 1] \Rightarrow \Omega = [0, 1]$, $A = [1/2, 1] \Rightarrow \mathbf{P}(A) = 1/2$.

Если в эксперименте отслеживать середину секущей хорды, то в этом случае мы имеем двумерную модель, где Ω - весь круг, A - кольцо с внутренним радиусом $1/2$. Тогда

$$\mathbf{P}(A) = \frac{\Lambda_2(A)}{\Lambda_2(\Omega)} = \frac{\pi - \pi/4}{\pi} = \frac{3}{4}.$$

На самом деле, в этой задаче можно получить континуум ответов (моделей):

Упражнение. Для любого наперед заданного числа $p \in (0, 1)$ построить такое пространство элементарных исходов, что в рамках геометрической модели вероятности будет иметь место равенство $\mathbf{P}(A) = p$.

Кажущееся противоречие объясняется просто - мы не определили понятие «*рассечения наудачу окружности прямой*». Иными словами, задача была поставлена некорректно. Каждый раз при выборе той или иной модели мы неявно доопределяли задачу, но каждый раз по-своему. Например, сравним две модели: величину центрального угла α и расстояние h . В первой модели величина α наудачу распределена на отрезке $[0, \pi]$ и $\mathbf{P}(A) = 2/3$. При этом $\alpha \in [0, 2\pi/3]$ тогда и только тогда, когда $h \in [1/2, 1]$. Стало

быть, вероятности указанных событий должны совпадать, т. е. $\mathbf{P}(h \in [1/2, 1]) = 2/3$. Таким образом, если величина α распределена наудачу, то расстояние h будет иметь приоритетные зоны в $\Omega = [0, 1]$ — эту точку как бы притягивает «магнитом» к правой границе отрезка $[0, 1]$. К такому же выводу относительно распределения α мы приходим, если объявим величину h распределенной наудачу на отрезке $[0, 1]$.

Если в Ω есть приоритетные зоны, то они нередко задаются функцией *плотности распределения* (о которой мы еще будем говорить), т. е. для любого события $A \subseteq \Omega$

$$\mathbf{P}(A) = \int_A p(t) dt.$$

Скажем, в двумерной модели бросания точки наудачу в круг единичного радиуса расстояние от этой точки до центра круга имеет плотность распределения $p(t) = 2t$ на единичном отрезке. Если $p \equiv \text{const}$, то точка в Ω распределена *равномерно* (это и означает, что у исходов такого стохастического эксперимента нет приоритетных зон), и мы приходим к модели геометрической вероятности.

Упражнение. *Вычислить плотность распределения расстояния h от центра круга до секущей хорды в описанном выше стохастическом эксперименте, когда центральный угол α , опирающийся на эту хорду, распределен равномерно на отрезке $[0, \pi]$.*

Элементы комбинаторики. Урновая модель.

Предположим, что у нас имеется урна, содержащая конечный набор различных элементов. Имеем два алгоритма выбора этих элементов.

1. *Выборка с возвращением.* Пусть в урне имеется N шаров, называемых *генеральной совокупностью*. Наудачу извлекаем по одному шару, фиксируем его метку и возвращаем обратно в генеральную совокупность. Тем самым, мы получим вектор (a_1, a_2, \dots, a_n) (a_i — метка i -го вытасченного шара) — «слово», если совокупность шаров интерпретировать как «алфавит». Это есть *выборка объема n с возвращением из генеральной совокупности N* .
2. *Выборка без возвращения* (например, как в игре «Спортлото»). Рассмотрим вектор (a_1, a_2, \dots, a_n) (здесь уже с необходимостью $n \leq N$) — *выборка без возвращения объема n из генеральной совокупности, содержащей N элементов*, где a_i — это и есть сам i -ый извлеченный шар.

Принцип умножения (основной прием решения задач комбинаторной теории вероятностей). Пусть имеется конечный набор конечных множеств A_1, A_2, \dots, A_m . Строится *слово* (вектор) (a_1, a_2, \dots, a_m) по алгоритму: $a_1 \in A_1, a_2 \in A_2, \dots, a_m \in A_m$. Число различных набранных слов в таком случае равно числу элементов декартова произведения

$$\#(A_1 \times A_2 \times \dots \times A_m) = \prod_{i=1}^m \#(A_i),$$

где $\#$ — считающая мера. Доказывается индукцией по m .

1. Тогда для выборки объёма n из N -элементного множества с возвращением и учетом порядка извлечения существует N^n возможностей.
2. Аналогично для выборки без возвращения с учетом порядка существует

$$N(N-1)\dots(N-n+1) = \frac{N!}{(N-n)!} = A_N^n$$

различных вариантов. Эта величина называется *числом размещений из N элементов по n* .

3. Для выборки без возвращения и без учета порядка извлечения элементов имеем

$$C_N^n = \frac{A_N^n}{n!}$$

различных вариантов, отличающихся составом. Величина C_N^n — это *число сочетаний из N элементов по n* (так называемый *биномиальный коэффициент*).

Гипергеометрическое распределение.

Рассмотрим урну, содержащую N_1 элементов первого типа и N_2 второго. Произведем выборку без возвращения объёма n . Требуется найти вероятность того, что в выборке содержатся $k \leq N_1$ элементов первого типа. Пусть для определенности $n - k \leq N_2$.

Так как мы различаем выборки только по составу, то $\#(\Omega) = C_{N_1+N_2}^n$.

При подсчете числа благоприятных исходов в этой задаче мы продемонстрируем еще один приём комбинаторики — *принцип независимого выбора*. Прежде всего, сделаем *редукцию* (замену исходной постановки на эквивалентную) задачи, а именно — сепарацию элементов двух типов, размещая их отдельно в двух “подурнах” исходной урны.

Ясно, что перебрать все благоприятные наборы элементов можно следующим образом. Сначала выбираем k элементов первого типа из соответствующей подурны, потом добираем на оставшиеся места элементы второго типа из своей подурны. Тогда по принципу умножения

$$\#(A) = C_{N_1}^k C_{N_2}^{n-k}.$$

Согласно модели классической вероятности

$$\mathbf{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_{N_1+N_2}^n}, \quad k = 0, 1, \dots, \min\{n, N_1\}, \quad k \geq n - N_2.$$

Набор этих вероятностей называется *гипергеометрическим распределением*.

Пример. Игра «Спортлото 5 из 36». В карточке «Спортлото» наудачу зачеркнем 5 чисел из 36. Найти вероятность того, что среди этих пяти зачеркнутых присутствуют три числа из пяти, выделенных тиражной комиссией. Редукция этой задачи состоит в том, что выбранные тиражной комиссией 5 занумерованных шаров мы объявим элементами первого типа (“мечеными” шарами), а оставшийся 31 шар — элементами второго типа. Мы можем считать, что в момент зачеркивания пяти номеров в карточке (это и есть выборка объёма 5) меченые шары уже содержатся в урне (т. е. что заседание тиражной

комиссии уже состоялось!) Тогда наша задача, очевидно, сводится к вышеприведенной схеме. В этом случае $N_1 = 5$, $N_2 = 31$, $n = 5$, $k = 3$. Следовательно,

$$\mathbf{P}(A) = \frac{C_5^3 C_{31}^2}{C_{36}^5} \approx 0,0123.$$

Упражнение. Вывести формулу обобщенного гипергеометрического распределения для более чем двух типов элементов генеральной совокупности.

Абстрактное вероятностное пространство.

Рассмотрим тройку $(\Omega, \mathcal{F}, \mathbf{P})$, где Ω - пространство элементарных исходов, \mathcal{F} - класс событий, \mathbf{P} - вероятностная мера. Тогда говорят, что $(\Omega, \mathcal{F}, \mathbf{P})$ – *вероятностное пространство*. Предполагается, что \mathcal{F} удовлетворяет следующим аксиомам:

Аксиомы σ -алгебры \mathcal{F} .

F1. $\Omega \in \mathcal{F}$ – достоверное событие.

F2. $\cup A_i, \cap A_i \in \mathcal{F}$ для любого не более чем счетного набора событий $\{A_i\}$ – замкнутость относительно счетного числа операций объединения и пересечения.

F3. Если $A \in \mathcal{F}$, то $\bar{A} \in \mathcal{F}$ – замкнутость относительно операции дополнения.

В силу принципа двойственности $\overline{\cup A_i} = \cap \bar{A}_i$ в пункте F2 достаточно требовать выполнения только одного из двух условий.

Аксиомы вероятностной меры \mathbf{P} .

P1. $\mathbf{P}(\cdot) \geq 0$, $\mathbf{P}(\Omega) = 1$.

P2. Счетная аддитивность: $\forall \{A_i\} \in \mathcal{F}$, удовлетворяющих условию $A_i \cap A_j = \emptyset$ при всех $i \neq j$, имеет место равенство

$$\mathbf{P}\left(\bigcup A_i\right) = \sum \mathbf{P}(A_i).$$

Иными словами, \mathbf{P} – нормированная мера.

Свойства вероятности.

1) $\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A})$, так как $A \cap \bar{A} = \emptyset$, и в силу аддитивности меры получаем

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \bar{A}) = \mathbf{P}(A) + \mathbf{P}(\bar{A}).$$

2) **Монотонность.** Если $A \subseteq B$, то $\mathbf{P}(A) \leq \mathbf{P}(B)$. В самом деле, $A \cap (B \setminus A) = \emptyset$, т. е. $\mathbf{P}(A) + \mathbf{P}(B \setminus A) = \mathbf{P}(A \cup (B \setminus A)) = \mathbf{P}(B)$, при этом $\mathbf{P}(B \setminus A) \geq 0$.

3) *Вероятность объединения пересекающихся событий.* Пусть $A, B \in \mathcal{F}$. Тогда $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

4) *Полуаддитивность вероятности:* $\forall \{A_i; i \leq n\} \in \mathcal{F}$

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

Эта свойство доказывается методом математической индукции, где база индукции — это пункт 3.

5) *Формула включения-исключения:*

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = & \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i_1 < i_2 \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3 \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots \\ & + (-1)^{n+1} \mathbf{P}\left(\bigcap_{i=1}^n A_i\right). \end{aligned}$$

Эта формула также доказывается по индукции, где база индукции — пункт 3.

В качестве дополнения к свойствам 4 — 5 сформулируем следующее упражнение:

Упражнение. Доказать двойственный аналог формулы включения-исключения:

$$\begin{aligned} \mathbf{P}\left(\bigcap_{i=1}^n A_i\right) = & \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i_1 < i_2 \leq n} \mathbf{P}(A_{i_1} \cup A_{i_2}) + \sum_{i_1 < i_2 < i_3 \leq n} \mathbf{P}(A_{i_1} \cup A_{i_2} \cup A_{i_3}) - \dots \\ & + (-1)^{n+1} \mathbf{P}\left(\bigcup_{i=1}^n A_i\right). \end{aligned}$$

Упражнение. Доказать следующие неравенства:

a) $\forall \{A_i; i \leq n\} \in \mathcal{F}$

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i < j \leq n} \mathbf{P}(A_i \cap A_j).$$

b) $\forall \{A_i; i \leq n\} \in \mathcal{F}$

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i < j \leq n} \mathbf{P}(A_i \cap A_j) + \sum_{i_1 < i_2 < i_3 \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}).$$

Условная вероятность. Независимость событий.

Пусть в урне имеется 100 пронумерованных шаров. Какова вероятность при выборе наудачу вытянуть шар с номером, кратным четырём? По классической модели — это $1/4$. Добавим частичную информацию о проведенном эксперименте: известно, что вытянули шарик с четным номером. Тогда понятно, что после проведения эксперимента и указанной частичной информации *апостериорная* вероятность интересующего нас события увеличится вдвое. Этот элементарный пример приводит нас к понятию условной вероятности.

Определение. Имеется два события: $A, B \in \mathcal{F}, \mathbf{P}(B) \neq 0$. Условная вероятность события A при условии, что произошло событие B , определяется по формуле

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

В этом определении пространство элементарных исходов сужено до события B . При этом вероятность A при произошедшем событии B есть доля $A \cap B$ по отношению ко всему B (доля в смысле меры \mathbf{P}). Легко видеть, что при любом фиксированном событии B условная вероятность будет также вероятностной мерой на исходном вероятностном пространстве.

Формула полной вероятности.

Теорема. Пусть имеется не более чем счетное разбиение пространства элементарных исходов $\{H_k\}_{k \geq 1}$ (попарно непересекающиеся множества и $\bigcup_{i \geq 1} H_i = \Omega$ — это так называемая полная группа событий.) Тогда $\forall A \in \mathcal{F}$ при условии, что $\mathbf{P}(H_k) \neq 0 \quad \forall k$, вероятность события A может быть представлена в виде ряда

$$\mathbf{P}(A) = \sum_{k \geq 1} \mathbf{P}(A|H_k) \mathbf{P}(H_k).$$

События H_k иногда называют *гипотезами* касательно проводимого стохастического эксперимента.

ДОКАЗАТЕЛЬСТВО. Рассмотрим попарно несовместные события $B_k = A \cap H_k$. Тогда в силу счетной аддитивности \mathbf{P} и свойства дистрибутивности соответствующих операций с множествами получаем

$$\mathbf{P}(A) = \mathbf{P}\left(\bigcup_k B_k\right) = \sum_{k \geq 1} \mathbf{P}(B_k) = \sum_{k \geq 1} \frac{\mathbf{P}(B_k)}{\mathbf{P}(H_k)} \mathbf{P}(H_k) = \sum_{k \geq 1} \mathbf{P}(A|H_k) \mathbf{P}(H_k). \quad \square$$

Пример. Рассмотрим шахматную доску размером 8 на 8 клеток. Наудачу на доску ставятся две ладьи разного цвета. Найти вероятность того, что они «пробивают» друг друга. Имеем две неопределенности — положение двух фигур (исключаем вариант постановки фигур в одну клетку). Редукция задачи: сначала ставим наудачу первую фигуру в одну из 64 клеток, затем вторую в одну из оставшихся 63-х. Определим полную группу гипотез $H_k = \{\text{первая фигура находится в } k\text{-ой клетке}\}, k = 1, \dots, 64$. При этом

$\forall i \neq j H_i \cap H_j = \emptyset$ и $\cup H_k = \Omega$ (так как одно из H_k наверняка произойдет). Очевидно, что по классической модели $\mathbf{P}(H_k) = 1/64$ и $\mathbf{P}(A|H_k) = 14/63 \forall k$. Значит, по формуле полной вероятности

$$\mathbf{P}(A) = \sum \mathbf{P}(A|H_k)\mathbf{P}(H_k) = \frac{14}{63} \sum_{k=1}^{64} \mathbf{P}(H_k) = \frac{14}{63}.$$

Упражнение. Найти вероятность взаимного пробития друг друга трёх фигур, если на доске два ферзя и ладья.

Формула Байеса.

Имеется событие A и полная группа событий (гипотез) $\{H_k\}$. Каждое из этих событий имеет известную *априорную* вероятность. Наша задача — вычислить вероятность $\mathbf{P}(H_k|A)$ при $\mathbf{P}(A) \neq 0$, т. е. произвести *апостериорный* пересчет априорных вероятностей имеющихся гипотез.

Теорема (формула Байеса). При вышеприведенных условиях

$$\mathbf{P}(H_k|A) = \frac{\mathbf{P}(H_k \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \cap H_k) \frac{\mathbf{P}(H_k)}{\mathbf{P}(H_k)}}{\sum_{i>0} \mathbf{P}(A|H_i) \mathbf{P}(H_i)} = \frac{\mathbf{P}(A|H_k) \mathbf{P}(H_k)}{\sum_{i>0} \mathbf{P}(A|H_i) \mathbf{P}(H_i)}.$$

Пример. Рассмотрим урновую модель: 3 белых и 3 черных шарика. Из урны «пропал» шар неизвестного цвета. После чего при выборе наудачу двух шаров происходит событие $A = \{\text{оба шарика оказались белыми}\}$. По цвету потерянного шарика определим две гипотезы $H_б$ и $H_ч$. Тогда $\mathbf{P}(H_б) = \mathbf{P}(H_ч) = 1/2$, а условные вероятности события A при условии наступления одного из событий $H_б$ или $H_ч$ будут $\mathbf{P}(A|H_б) = 1/C_5^2$, $\mathbf{P}(A|H_ч) = C_3^2/C_5^2$. По формуле Байеса получаем $\mathbf{P}(H_б|A) = 1/4$ и $\mathbf{P}(H_ч|A) = 3/4$.

Заметим, что условная вероятность обладает свойствами обычной вероятности: $\mathbf{P}(\Omega|B) = 1$, $\mathbf{P}(\cup A_i|B) = \sum \mathbf{P}(A_i|B)$ для любых попарно несовместных событий $\{A_i\}$.

Понятие независимости событий.

Одно из наиболее важных понятий теории вероятности — это понятие независимости событий. Следующее два определения вводят принципиальное понятие, выделяющее этот раздел математики из функционального анализа, в частности, из теории меры.

Определение. События A и B (для определенности пусть $\mathbf{P}(B) \neq 0$) называются *независимыми*, если $\mathbf{P}(A|B) = \mathbf{P}(A)$, то есть априорная и апостериорная относительно B вероятности события A совпадают. Другими словами, информация об эксперименте, содержащаяся в событии B , не дает никакой существенной информации о событии A и не меняет его априорную вероятность.

Замечание. Ясно, что если $\mathbf{P}(A) \neq 0$, то приведенное определение будет инвариантным относительно перестановки событий A и B . При этом мы получаем, что $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. Полученное соотношение нередко используют в качестве определения независимости. При этом здесь уже не требуется никаких ограничений на вероятности рассматриваемых событий.

Пример. Рассмотрим единичный квадрат. Пусть внутри этого квадрата выделены два измеримых подмножества: A_x — некоторая «вертикальная» (точнее, параллельная

одной из сторон квадрата) полоса, а B_y — «горизонтальная», перпендикулярная A_x . Бросаем наудачу точку в этот квадрат. Событие A — попадание точки в полосу A_x , событие B — попадание точки в полосу B_y . Покажем, что события A и B независимы.

Пусть Δ_1 — ширина полосы A_x , Δ_2 — ширина полосы B_y . Очевидно, что $\mathbf{P}(A) = \Delta_1$, $\mathbf{P}(B) = \Delta_2$. В то же время, $\mathbf{P}(A \cap B) = \Delta_1 \cdot \Delta_2$. Таким образом, в нашем примере выполнено равенство $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, откуда и следует независимость событий A и B .

Зависимые события в рассматриваемом квадрате построить гораздо проще, чем независимые. Например, зависимыми будут два непересекающихся множества ненулевой меры.

Упражнение. Проверить, что если в предыдущем примере в качестве пространства элементарных событий взять круг, то события A и B , вообще говоря, будут зависимыми.

Определение. Конечный набор событий A_1, \dots, A_n называется набором *независимых в совокупности* событий, если для любого конечного поднабора этих событий вероятность их совместного наступления факторизуется, т. е.

$$\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_m}) = \prod_{j=1}^m \mathbf{P}(A_{i_j}) \text{ для любого } \{i_1, \dots, i_m\} \subset \{1, \dots, n\}.$$

Понятно, что из независимости в совокупности следует попарная независимость. Однако обратное утверждение, как показывает следующий пример, не всегда имеет место.

Пример Бернштейна. Дан правильный тетраэдр. Одна его грань окрашена в красный цвет, другая — в зелёный, третья — в синий, а на четвертую грань нанесли все три цвета. Бросаем этот тетраэдр («игральную кость») с интенсивным вращением на плоскость. После его остановки фиксируем цвета его основания. Рассмотрим события: R, G, B — появление красного, зелёного и синего цвета на нижней грани. Легко проверить, что эти события *попарно независимы* (к примеру, $\mathbf{P}(R \cap B) = 1/4 = \mathbf{P}(R) \cdot \mathbf{P}(B) = 1/2 \cdot 1/2$). Однако независимыми в совокупности эти события не являются, так как $\mathbf{P}(R \cap B \cap G) = 1/4 \neq 1/2 \cdot 1/2 \cdot 1/2 = \mathbf{P}(R) \cdot \mathbf{P}(B) \cdot \mathbf{P}(G)$.

Для того чтобы мы могли рассматривать содержательные задачи, необходимо постулировать одну из основных аксиом теории вероятностей:

Аксиома независимости. Причинно несвязанные (в бытовом смысле) события являются независимыми и в математическом (вероятностном) смысле.

Схема Бернулли.

Схема Бернулли описывает следующий стохастический эксперимент. В одинаковых условиях проводятся последовательно n испытаний, каждое из которых имеет два исхода («успех» и «неудача»). Если вероятности успеха и неудачи одинаковы, то говорят о *симметричной схеме Бернулли*. В схеме Бернулли основным является предположение о том, что любой набор событий в этой конечной цепочке испытаний объявляется независимым в совокупности.

Обозначим через p вероятность успеха (разумеется, в общем случае $p \neq 1/2$). Сформулируем теорему о распределении числа успехов в схеме Бернулли.

Теорема (формула Бернулли). Пусть S_n — число успехов в серии из n независимых однородных испытаний в схеме Бернулли. Тогда распределение этой величины имеет вид $\mathbf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k}$.

ДОКАЗАТЕЛЬСТВО. Вероятностное пространство — множество n -мерных векторов, компоненты которых 0 (неудача) или 1 (успех). Посчитаем вероятность появления вектора $\omega = (10100 \dots 01)$, в котором k единиц стоят на фиксированных местах. Введём соответствующие этому вектору события: A_1 — появление 1 на первом месте, A_2 — появление 0 на втором месте, A_3 — появление 1 на третьем месте, \dots , A_n — появление 1 на n -м месте. Тогда

$$\mathbf{P}(\omega) = \mathbf{P}\left(\bigcap_{j=1}^n A_j\right) = \prod_{j=1}^n \mathbf{P}(A_j) = p(1-p)(1-p)p \dots (1-p)p = p^k (1-p)^{n-k},$$

где второе равенство написано на основе того, что события A_1, \dots, A_n независимы в совокупности.

Для подсчёта $\mathbf{P}(S_n = k)$ надо учесть все векторы, содержащие k единиц. Таких векторов будет C_n^k , и вероятности появления каждого из них одинаковы. Поэтому можно воспользоваться дискретной моделью:

$$\mathbf{P}(S_n = k) = C_n^k \mathbf{P}(\omega) = C_n^k p^k (1-p)^{n-k}. \quad \square$$

Легко понять, что в случае симметричной схемы Бернулли ($p = 1/2$) мы находимся в рамках классической вероятностной модели. Ясно также, что существует много экспериментов в схеме Бернулли с вероятностью успеха, отличной от $1/2$. Допустим, в отдельном тираже игры «Спортлото 6 из 49», в котором приняло участие n игроков, заполнивших свои карточки *наудачу или по одному и тому же стохастическому алгоритму* без каких-либо контактов друг с другом (иначе говоря, независимо друг от друга!), мы фиксируем только те карточки, в которых угаданы все 6 счастливых чисел. Вероятность угадать 6 номеров из 6, помеченных тиражной комиссией, есть $p = 1/C_{49}^6 \approx 7,2 \cdot 10^{-8}$. Это и есть вероятность успеха в рассматриваемой схеме Бернулли.

Упражнение. В рассмотренной выше игре каковы шансы, что ровно 10 человек из 1000 игравших угадали 3 числа из 6, помеченных тиражной комиссией?

Формула Бернулли задаёт дискретное распределение, называемое *биномиальным*: $\tilde{p}_k = C_n^k p^k (1-p)^{n-k}$, $k = 0, \dots, n$.

Связь гипергеометрического и биномиального распределений.

Напомним, что к гипергеометрическому распределению нас привела урновая модель: в урне N_1 белых и N_2 чёрных шаров, производится выборка без возвращения объёма n . Тогда вероятность того, что она содержит k чёрных шаров (конечно, предполагается, что $0 \leq k \leq \min\{n, N_2\}$ и $n - k \leq N_1$) равна

$$\tilde{p}_k = \frac{C_{N_2}^k C_{N_1}^{n-k}}{C_{N_1+N_2}^n}.$$

Посмотрим, как ведёт себя гипергеометрическое распределение, когда $N_1 \rightarrow \infty$, $N_2 \rightarrow \infty$, причем так, что $\frac{N_2}{N_1 + N_2} \rightarrow p$, где $p \in (0, 1)$. В этом случае можно считать, что N_1

и N_2 настолько велики, что выбор конечного числа шаров не изменяет генеральную совокупность. Но это значит, что вероятность вынуть чёрный шар на любом шаге равна $p_0 = N_2/(N_1 + N_2)$. Тогда осуществление выборки объёма n соответствует последовательности из n испытаний в схеме Бернулли с вероятностью успеха p_0 . Следовательно, справедливо приближённое равенство $\tilde{p}_k \approx C_n^k p_0^k (1 - p_0)^{n-k}$.

Теперь докажем отмеченное выше предельное соотношение. Именно, покажем, что в оговоренных выше условиях

$$\tilde{p}_k \rightarrow C_n^k p_0^k (1 - p_0)^{n-k}.$$

Действительно,

$$\begin{aligned} \tilde{p}_k &= \frac{C_{N_2}^k C_{N_1}^{n-k}}{C_{N_1+N_2}^n} = \frac{\frac{N_2(N_2-1)\dots(N_2-k+1)}{k!} \cdot \frac{N_1(N_1-1)\dots(N_1-n+k+1)}{(n-k)!}}{\frac{(N_1+N_2)(N_1+N_2-1)\dots(N_1+N_2-n+1)}{n!}} = \\ &= C_n^k \cdot \frac{\frac{N_2}{N_1+N_2} \cdot \frac{N_2-1}{N_1+N_2} \cdot \dots \cdot \frac{N_2-k+1}{N_1+N_2} \cdot \frac{N_1}{N_1+N_2} \cdot \dots \cdot \frac{N_1-n+k+1}{N_1+N_2}}{\frac{N_1+N_2}{N_1+N_2} \cdot \frac{N_1+N_2-1}{N_1+N_2} \cdot \dots \cdot \frac{N_1+N_2-n+1}{N_1+N_2}} \\ &\xrightarrow[\frac{N_2}{N_1+N_2} \rightarrow p]{N_1 \rightarrow \infty, N_2 \rightarrow \infty} C_n^k \frac{p \cdot \dots \cdot p \cdot (1-p) \cdot \dots \cdot (1-p)}{1 \cdot 1 \cdot \dots \cdot 1} = C_n^k p_0^k (1 - p_0)^{n-k}. \end{aligned}$$

Полиномиальное распределение.

В одних и тех же условиях проводятся n стохастических испытаний, каждое из которых имеет конечное число исходов (не менее двух!). Занумеруем эти исходы числами $1, \dots, m$. Вероятности этих исходов обозначим p_1, \dots, p_m соответственно. Какова вероятность того, что в результате проведенных n испытаний мы будем иметь ровно k_1 исходов 1-го типа, \dots , k_m исходов m -го типа (запяты в нашей записи играют роль пересечения)? Таким образом, нас интересует значение $\mathbf{P}(S_n^{(1)} = k_1, \dots, S_n^{(m)} = k_m)$. Очевидно, что $\sum_{i=1}^m k_i = n$ и $\sum_{i=1}^m p_i = 1$. Согласно дискретной вероятностной модели, вероятность любой фиксированной последовательности испытаний, содержащей соответственно k_1, \dots, k_m исходов типов $1, \dots, m$, равна $p_1^{k_1} \cdot \dots \cdot p_m^{k_m}$. Число различных таких последовательностей, как нетрудно убедиться с помощью принципа умножения, есть так называемый *полиномиальный коэффициент*

$$C_n^{k_1} \cdot C_{n-k_1}^{k_2} \cdot \dots \cdot C_{n-\sum_{i=1}^{m-1} k_i}^{k_m} = \frac{n!}{k_1! \cdot \dots \cdot k_m!}.$$

Итак, мы получили формулу для *полиномиального распределения*

$$\mathbf{P}(S_n^{(1)} = k_1, \dots, S_n^{(m)} = k_m) = \frac{n!}{k_1! \cdot \dots \cdot k_m!} p_1^{k_1} \cdot \dots \cdot p_m^{k_m}.$$

Задача о размещении частиц по ячейкам.

Вкратце стохастический эксперимент состоит в следующем. Происходит «бросание» n точек (дробинок) по r ячейкам согласно некоторому алгоритму. Нас интересует вероятность того или иного варианта их совместного размещения.

1. Полиномиальный алгоритм. В одинаковых условиях бросаем дробинки по одной. Каждая дробинка с фиксированной вероятностью p_i может попасть в i -ю ячейку. Тогда вероятность того, что в 1-й ячейке оказалось k_1 дробинок, . . . , в r -й — k_r дробинок даётся формулой для полиномиального распределения, выведенной в предыдущем пункте.

В частности, если вероятность попадания в любую ячейку равна $1/r$, то мы имеем дело с *распределением Максвелла*.

2. Схема Бозе—Эйнштейна. Дробинки разбрасываются горстью по всем r ячейкам. При этом предполагается, что все варианты размещения дробинок по ячейкам равно-возможны. Посчитаем число всевозможных размещений дробинок в ячейках. Заметим, что каждое фиксированное размещение можно описать вектором длины $n + r + 1$, состоящим из нулей и единиц, в котором $r + 1$ единиц играют роль перегородок между ячейками, а та или иная серия нулей (возможно, пустая) между двумя единицами — количество дробинок в соответствующей ячейке. Ясно, что при такой интерпретации две крайние единицы можно из рассмотрения исключить (они присутствуют в любом варианте размещения). Так что задача свелась к подсчёту числа всевозможных размещений неразличимых дробинок (нулей) на $n+r-1$ местах. Значит, их число равно C_{n+r-1}^n . Таким образом, вероятность каждого такого размещения вычисляется по формуле $1/C_{n+r-1}^n$.

Упражнение. В обеих схемах размещения найти вероятность того, что в результате эксперимента останется l пустых ячеек.

Асимптотический анализ биномиального распределения. Пуассоновская аппроксимация.

При изучении схемы Бернулли мы получили формулу для биномиального распределения $\mathbf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k}$. Обозначим левую часть этой формулы через $B_{n,p}(k)$, где $k = 0, \dots, n$. В этом разделе мы обсуждаем вычислительные аспекты, связанные с биномиальным распределением. В реальных задачах обычно n велико, при этом p может быть близким, скажем, к значениям 0 и 1. Так что при вычислении вероятностей $B_{n,p}(k)$ могут возникать неопределённости типа умножения бесконечности на ноль. Рассмотрим ситуацию, когда p близко к нулю (в силу симметрии задачи к этому же условию сводится и случай близости p к 1). Это условие описывается следующим двойным предельным переходом по направлению:

$$n \rightarrow \infty, \quad p \rightarrow 0, \quad np \rightarrow \lambda > 0. \quad (*)$$

Теорема (локальная предельная теорема Пуассона). Пусть выполнено условие (*). Тогда при любом фиксированном k

$$B_{n,p}(k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} =: \pi_\lambda(k).$$

Набор $\{\pi_\lambda(k) : k = 0, \dots, n\}$ задаёт так называемое *распределение Пуассона с параметром λ* (элементарно проверяется, что $\sum \pi_\lambda(k) = 1$).

ДОКАЗАТЕЛЬСТВО. Достаточно выписать определение биномиального коэффициента и воспользоваться вариантом «второго замечательного предела»:

$$\begin{aligned} B_{n,p}(k) &= \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{k!} p^k (1-p)^{n-k} \frac{n^k}{n^k} \sim \frac{(pn)^k}{k!} \left(1 - \frac{pn}{n}\right)^{n-k} \\ &\sim \frac{\lambda^k \left(1 - \frac{\lambda}{n}\right)^n}{k! (1-p)^k} \sim \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

что и завершает доказательство.

Следствие 1. Если $A \subset \mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ и одно из множеств A или \bar{A} конечно, то при выполнении условия (*)

$$\left| \mathbf{P}(S_n \in A) - \Pi_\lambda(A) \right| \rightarrow 0,$$

где $\Pi_\lambda(A) = \sum_{k \in A} \pi_\lambda(k)$.

ДОКАЗАТЕЛЬСТВО. Поскольку $\mathbf{P}(S_n \in A) = \sum_{k \in A} B_{n,p}(k)$, то для конечного A можно воспользоваться локальной предельной теоремой Пуассона. Если же \bar{A} конечно, то используя свойство $\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A})$, сводим всё к предыдущему случаю. \square

Следствие 2 (классическая интегральная теорема Пуассона). При выполнении условия (*)

$$\sup_{A \subset \mathbb{Z}_+} \left| \mathbf{P}(S_n \in A) - \Pi_\lambda(A) \right| \rightarrow 0.$$

ДОКАЗАТЕЛЬСТВО. Для любого натурального N справедливо неравенство

$$\begin{aligned} \left| \mathbf{P}(S_n \in A) - \Pi_\lambda(A) \right| &\leq \sum_{k \leq N} |B_{n,p}(k) - \pi_\lambda(k)| \\ &\quad + \mathbf{P}(S_n > N) + \Pi_\lambda(k : k > N). \end{aligned}$$

Отметим, что правая часть в этой оценке не зависит от множества A . Далее, при фиксированном N первая сумма в правой части этого неравенства при $n \rightarrow \infty$ стремится к нулю в силу локальной предельной теоремы Пуассона. Кроме того, следствие 1 утверждает, что при $n \rightarrow \infty$

$$\mathbf{P}(S_n > N) \rightarrow \Pi_\lambda(k : k > N),$$

причем вероятность $\Pi_\lambda(k : k > N)$ может быть сделана сколь угодно малой выбором достаточно большого N , что и требовалось доказать. \square

Приведенные теоремы часто называют *теоремами о редких событиях*.

Сформулируем без доказательства результат, позволяющий оценивать точность пуассоновской аппроксимации.

Теорема. Для любых n и p имеет место оценка

$$\sup_{A \subset \mathbb{Z}_+} \left| \mathbf{P}(S_n \in A) - \Pi_{np}(A) \right| \leq \min\{p, np^2\}.$$

Пример. Пусть в игре «Спортлото» «6 из 49» участвует $n = 10^7$ человек. Какова вероятность, что хотя бы один угадает все 6 номеров?

Здесь $p = 7,2 \cdot 10^{-8}$. Множество $A = \{k : k \geq 1\}$, $\min\{p, np^2\} = 0,72 \cdot 7,2 \cdot 10^{-8}$. Очень малое значение последней величины позволяет нам утверждать, что в данном случае значение биномиального распределение на любом подмножестве практически совпадает с соответствующей пуассоновской вероятностью. В рассматриваемом случае имеем: $\lambda = np \approx 0,7$ и $\Pi_\lambda(A) \approx 0,5$. Если же $n = 10^6$, то $\lambda \approx 0,07$ и $\Pi_\lambda(A) \approx 0,09$.

Пример. В аудитории 100 человек. Какова вероятность, что хотя бы у одного из присутствующих день рождения 1 октября?

В этом примере мы имеем дело со схемой Бернулли с вероятностью успеха $p = 1/365$ и параметром пуассоновского распределения $\lambda = np \approx 0,3$. Снова рассматриваем множество $A = \{k : k \geq 1\}$. Искомая вероятность довольно точно определяется по теореме Пуассона: $\Pi_\lambda(A) \approx 0,26$.

Нормальное (гауссовское) приближение биномиального распределения.

Теперь займемся изучением асимптотического поведения биномиального распределения при $n \rightarrow \infty$, когда p фиксировано (например, мы изучаем распределения числа «орлов» при бросании симметричной монеты достаточно много раз). Всюду в дальнейшем символом \log обозначается натуральный логарифм.

Теорема (обобщенная локальная предельная теорема (ЛПТ)). Пусть $k \rightarrow \infty$ и $n - k \rightarrow \infty$. Тогда

$$B_{n,p}(k) \sim \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\{-nH(p^*)\},$$

где $p^* = k/n$, $H(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$, $x \in (0,1)$.

ДОКАЗАТЕЛЬСТВО. Сначала мы представим факториалы в выражении для $B_{n,p}(k)$ с помощью формулы Стирлинга

$$n! = \sqrt{2\pi n} n^n e^{-n+\theta(n)}, \text{ где } \frac{1}{12n+1} \leq \theta(n) \leq \frac{1}{12n}.$$

Имеем

$$\begin{aligned} B_{n,p}(k) &= \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n+\theta(n)} \cdot p^k \cdot (1-p)^{n-k}}{\sqrt{2\pi k} \cdot k^k \cdot e^{-k+\theta(k)} \cdot \sqrt{2\pi(n-k)} \cdot (n-k)^{n-k} \cdot e^{-n+k+\theta(n-k)}} \sim [\text{т. к. } \theta(\cdot) \rightarrow 0] \sim \\ &\sim \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \left(\frac{p}{p^*}\right)^k \left(\frac{1-p}{1-p^*}\right)^{n-k} = \end{aligned}$$

$$\begin{aligned}
&= [\text{обозначим } C_n(p^*) = \frac{1}{\sqrt{2\pi np^*(1-p^*)}}] = \\
&= C_n(p^*) \exp \left\{ k \log \frac{p}{p^*} + (n-k) \log \frac{1-p}{1-p^*} \right\} \\
&= C_n(p^*) \exp \left\{ -n \left[p^* \log \frac{p^*}{p} + (1-p^*) \log \frac{1-p^*}{1-p} \right] \right\} \\
&= C_n(p^*) \exp \{-nH(p^*)\}. \quad \square
\end{aligned}$$

В приведенном доказательстве можно было бы вместо знака « \sim » подставить « $=$ », используя O -символику. В такой форме суммарная погрешность будет равна

$$\theta(n) - \theta(k) - \theta(n-k) = O(1/k) + O(1/(n-k)),$$

а так как $e^{O(1/m)} = 1 + O(1/m)$, то окончательное асимптотическое представление биномиальной точечной массы принимает вид

$$B_{n,p}(k) = C_n(p^*) \exp\{-nH(p^*)\} \left[1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right) \right].$$

Пример. Бросаем симметричную монету 10000 раз. Какова вероятность, что гербов и решек выпадет одинаковое количество?

Итак, в нашем случае $p = 1/2$, $n = 10000$, $k = 5000$, $p^* = 1/2$ и $H(p^*) = 0$. Тогда

$$B_{n,p}(k) \sim \frac{\sqrt{2}}{\sqrt{\pi n}} = \frac{\sqrt{2}}{100\sqrt{\pi}} \approx 8 \cdot 10^{-3}.$$

Для того чтобы сформулировать следствие из только что доказанной теоремы, введем следующее определение.

Определение. Плотностью нормального распределения (распределения Гаусса) с параметрами α , σ называется функция

$$\varphi_{(\alpha,\sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2}.$$

Стандартной нормальной (гауссовской) плотностью называется функция

$$\varphi_{(0,1)}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Заметим, что $\varphi_{(\alpha,\sigma)}(x) = \frac{1}{\sigma} \varphi_{(0,1)}\left(\frac{x-\alpha}{\sigma}\right)$.

Следствие (классическая ЛПТ). Пусть $n \rightarrow \infty$ и $n|p^* - p|^3 \rightarrow 0$. Тогда

$$B_{n,p}(k) \sim \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)}\left(\frac{k-np}{\sigma\sqrt{n}}\right), \quad \text{где } \sigma^2 = p(1-p),$$

или в терминах $p^* = k/n$

$$B_{n,p}(k) \sim \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)}\left(\frac{\sqrt{n}(p^* - p)}{\sigma}\right).$$

ДОКАЗАТЕЛЬСТВО. Сначала отметим, что $H(p) = 0$. Далее, вычислим первую производную функции $H(x)$:

$$\begin{aligned} H'(x) &= \log \frac{x}{p} + x \cdot \frac{p}{x} \cdot \frac{1}{p} + \log \frac{1-x}{1-p} \cdot (-1) - 1 \\ &= \log \frac{x}{p} - \log \frac{1-x}{1-p}. \end{aligned}$$

Значит, $H'(p) = 0$. Вторая производная имеет вид

$$H''(x) = \frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)}.$$

Тогда $H''(p) = 1/\sigma^2$. Тогда по формуле Тейлора $H(p^*) = \frac{1}{2\sigma^2}(p^* - p)^2 + O(|p^* - p|^3)$. Поэтому в условиях теоремы

$$\begin{aligned} nH(p^*) &= \frac{n}{2\sigma^2}(p^* - p)^2 + O(n|p^* - p|^3) = \frac{n}{2\sigma^2}(p^* - p)^2 + o(1) \\ &= \frac{1}{2} \left(\frac{k - np}{\sigma\sqrt{n}} \right)^2 + o(1). \end{aligned}$$

Таким образом,

$$\exp\{-nH(p^*)\} \sim \exp\left\{-\frac{1}{2} \left(\frac{k - np}{\sigma\sqrt{n}} \right)^2\right\}.$$

Так как $n|p^* - p|^3 \rightarrow 0$, то $p^* \sim p$, т. е.

$$\frac{1}{\sqrt{2\pi np^*(1-p^*)}} \sim \frac{1}{\sigma\sqrt{2\pi n}}. \quad \square$$

Интегральная предельная теорема Муавра–Лапласа.

Перейдём теперь к вычислению интегральных вероятностей (т. е. сумм локальных точечных вероятностей). Именно, нас интересует приближенное значение для вероятности

$$\mathbf{P}(S_n \in A) = \sum_{k \in A} B_{n,p}(k).$$

Сформулируем сначала

Определение. Функцией Лапласа $\Phi(x)$, $x \in \mathbb{R}$, называется функция распределения стандартного нормального закона

$$\Phi(x) = \int_{-\infty}^x \varphi_{(0,1)}(t) dt.$$

Теорема (Муавр, Лаплас). *Рассматривается схема Бернулли с вероятностью успеха p . Пусть S_n – число успехов в серии из n независимых испытаний. Тогда для любых $a < b$*

$$\mathbf{P} \left(a \leq \frac{S_n - np}{\sigma\sqrt{n}} \leq b \right) \xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a),$$

где $\sigma^2 = p(1 - p)$.

Эта теорема даёт возможность оценить вероятность попадания числа «успехов» S_n в отрезок $[a\sigma\sqrt{n} + np, b\sigma\sqrt{n} + np]$ при большом числе испытаний. Скажем, если величина $\Phi(b) - \Phi(a)$ близка к единице, то при значительном числе испытаний мы вправе рассчитывать на то, что число успехов S_n со значимой вероятностью будет находиться в окрестности числа np , причем размер этой окрестности имеет существенно меньший порядок \sqrt{n} .

Заметим, что график функции Лапласа обладает свойством центральной симметрии относительно точки $(0, 1/2)$, т. е. $\Phi(x) = 1 - \Phi(-x)$ (всё это легко следует из чётности функции $\varphi_{(0,1)}(x)$). Если интервал в теореме симметричный, т. е. $-a = b > 0$, то

$$\mathbf{P} \left(-b \leq \frac{S_n - np}{\sigma\sqrt{n}} \leq b \right) \rightarrow \Phi(b) - \Phi(-b) = 1 - 2\Phi(-b) = 1 - \varepsilon.$$

Задав малое ε и решив уравнение $\Phi(-b) = \varepsilon/2$, можно восстановить симметричный отрезок, в который попадет величина S_n с вероятностью, близкой к $1 - \varepsilon$.

Пример. Симметричная монета бросается 10000 раз. Требуется указать интервал, в который число S_n «навверняка» попадает.

Очевидно, из «соображений симметрии», за центр этого интервала нужно принять 5000. Под «навверняка» будем понимать $\varepsilon = 0,01$ в вышеприведенной формуле. Из таблиц найдем, что $b = 2,6$. Отсюда получаем $b\sigma\sqrt{n} = 2,6 \cdot \frac{1}{2} \cdot 100 = 130$, что дает на удивление достаточно узкий «доверительный» интервал $4870 \leq S_n \leq 5130$, в котором величина S_n окажется почти навверняка. Отметим, что для частоты появления «успеха» $\nu_n = \frac{S_n}{n}$ справедливо двойное включение

$$-\frac{b\sigma}{\sqrt{n}} + p \leq \nu_n \leq \frac{b\sigma}{\sqrt{n}} + p.$$

В рассматриваемом случае $0,487 \leq \nu_n \leq 0,513$ – подтверждение закона стабилизации частот (порядок разброса $1/\sqrt{n} = 0,01$).

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ. Согласно модели дискретной вероятности ($S_n = 0, 1, 2, \dots$) имеем

$$\mathbf{P} \left(a \leq \frac{S_n - np}{\sigma\sqrt{n}} \leq b \right) = \sum_{k \in Z_+ : a \leq \frac{k - np}{\sigma\sqrt{n}} \leq b} B_{n,p}(k).$$

Так как $np + a\sigma\sqrt{n} \leq k \leq np + b\sigma\sqrt{n}$, то число слагаемых суммы справа имеет порядок \sqrt{n} . Мы хотим воспользоваться локальной предельной теоремой, но так как число слагаемых в указанной сумме растет как \sqrt{n} , то напрямую доказанные выше асимптотические формулы использовать нельзя, поскольку здесь уже нужно учитывать погрешность приближения.

Справедлива следующая цепочка импликаций:

$$\begin{aligned} \frac{a\sigma}{\sqrt{n}} \leq \frac{k}{n} - p \leq \frac{b\sigma}{\sqrt{n}} &\Rightarrow |p^* - p| \leq \max \left\{ \frac{|b\sigma|}{\sqrt{n}}, \frac{|a\sigma|}{\sqrt{n}} \right\} \Rightarrow n|p^* - p|^3 \leq \max \left\{ \frac{|b^3\sigma^3|}{\sqrt{n}}, \frac{|a^3\sigma^3|}{\sqrt{n}} \right\} = \\ &= \frac{\sigma^3}{\sqrt{n}} \max \{ |b|^3, |a|^3 \}. \end{aligned}$$

Стало быть, условие классической ЛПТ выполнено, причем

$$n|p^* - p|^3 = O\left(\frac{1}{\sqrt{n}}\right).$$

Таким образом, мы получили представление

$$nH(p^*) = \frac{1}{2} \left(\frac{k - np}{\sigma\sqrt{n}} \right)^2 + O\left(\frac{1}{\sqrt{n}}\right).$$

Далее, из арифметики O -символов мы имеем

$$e^{-(\dots)+O(1/\sqrt{n})} = e^{-(\dots)} \left(1 + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

Тогда

$$B_{n,p}(k) = \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp \left\{ -\frac{1}{2} \left(\frac{k - np}{\sigma\sqrt{n}} \right)^2 \right\} \left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right) + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

Кроме того, из ограничений на k следует, что обе величины k и $n - k$ имеют порядок n . Точнее, при фиксированном $p \in (0, 1)$ мы имеем $k \sim np$ и $n - k \sim n(1 - p)$. Тогда

$$B_{n,k}(k) = \frac{1}{\sqrt{\dots}} \exp \{ \dots \} \left(1 + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

Также мы отметим, что

$$p + \frac{a\sigma}{\sqrt{n}} \leq p^* \leq p + \frac{b\sigma}{\sqrt{n}}.$$

Следовательно, если

$$f(x) = \frac{1}{\sqrt{x(1-x)}},$$

то

$$f(p^*) = f(p) + O(|p^* - p|) = f(p) + O\left(\frac{1}{\sqrt{n}}\right).$$

Последнее равенство — это результат применения формулы конечных приращений Лагранжа. В итоге мы получим

$$B_{n,p}(k) = \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)} \left(\frac{k - np}{\sigma\sqrt{n}} \right) \left(1 + O\left(\frac{1}{\sqrt{n}}\right) \right) = \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)} \left(\frac{k - np}{\sigma\sqrt{n}} \right) + O\left(\frac{1}{n}\right) \quad (1)$$

Заметим, что если p , а стало быть, и σ малы, то погрешность может оказаться значительной. Здесь эту погрешность надо бы сравнить с погрешностью в теореме Пуассона, чтобы выбор нормальной аппроксимации был обоснован. Кроме того, данная формула верна только для множеств вида

$$A = \left\{ k : a \leq \frac{k - np}{\sigma\sqrt{n}} \leq b \right\}.$$

Из (1) получаем, что

$$\begin{aligned} P\left(\frac{S_n - np}{\sigma\sqrt{n}} \in [a, b]\right) &= P(S_n \in A) = \sum_{k \in A} B_{n,p}(k) \\ &= \sum_{k \in A} \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)}\left(\frac{k - np}{\sigma\sqrt{n}}\right) + \sum_{k \in A} O\left(\frac{1}{n}\right). \end{aligned}$$

В этой области O -символ зависит только от a, b и p и равномерен по всем k таким, что $np + a\sigma\sqrt{n} \leq k \leq np + b\sigma\sqrt{n} \forall a < b$. Значит,

$$\sum_{k \in A} O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Обозначим

$$x_k := \frac{k - np}{\sigma\sqrt{n}}.$$

Тогда

$$x_{k+1} - x_k = \frac{1}{\sigma\sqrt{n}}.$$

Точки x_k образуют измельчающееся разбиение отрезка $[a, b]$. Для простоты считаем, что концы отрезка $[a, b]$ попадают в решетку $\{x_k\}$. В итоге мы получили риманову интегральную сумму с шагом разбиения $1/(\sigma\sqrt{n})$:

$$\begin{aligned} \sum_{k \in A} \frac{1}{\sigma\sqrt{n}} \varphi_{(0,1)}\left(\frac{k - np}{\sigma\sqrt{n}}\right) &= \sum_{x_k \in [a, b]} (x_{k+1} - x_k) \varphi_{(0,1)}(x_k) \\ &\rightarrow \int_a^b \varphi_{(0,1)}(x) dx = \Phi(b) - \Phi(a) \end{aligned} \quad (2)$$

при $n \rightarrow \infty$. Таким образом, теорема доказана. \square

На самом деле, мы можем уточнить соотношение (2), оценив погрешность при переходе от суммы к интегралу для гладкой функции f . Имеем

$$\sum := \sum_{a \leq x_k \leq b} f(x_k) \Delta x_k = \int_a^b f(x) dx + O(?).$$

Вспомним определения нижней и верхней сумм Римана. Если \hat{x}_k — точка минимума на $[x_k, x_{k+1}]$, а \tilde{x}_k — точка максимума непрерывной функции f , то для любой интегральной суммы Римана, соответствующей разбиению $\{x_k\}$, имеем

$$\sum f(\hat{x}_k)\Delta x_k \leq \sum f(\bar{x}_k)\Delta x_k \leq \sum f(\tilde{x}_k)\Delta x_k.$$

С другой стороны, легко видеть, что

$$\sum f(\hat{x}_k)\Delta x_k \leq \int_a^b f(x) dx \leq \sum f(\tilde{x}_k)\Delta x_k.$$

Из этих двух соотношений получаем

$$\left| \sum_{a \leq x_k \leq b} f(x_k)\Delta x_k - \int_a^b f(x) dx \right| \leq \sum (f(\tilde{x}_k) - f(\hat{x}_k))\Delta x_k.$$

Теперь если дополнительно потребовать *липшицевость* f , то есть выполнение неравенства $|f(x) - f(y)| \leq K|x - y|$ при всех x и y , где K — постоянная, то

$$\sum (f(\tilde{x}_k) - f(\hat{x}_k))\Delta x_k \leq K \sum (\Delta x_k)^2 = K\Delta x_1(b - a).$$

Для непрерывно дифференцируемой функции f из формулы конечных приращений $f(x + \Delta) = f(x) + f'(x + \theta\Delta)$ заключаем, что $K = \sup |f'|$.

В нашем случае мы имеем дело с функцией f вида

$$\varphi_{(0,1)}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Модуль производной

$$|\varphi'_{(0,1)}(x)| = \frac{1}{\sqrt{2\pi}} |x| e^{-x^2/2}$$

достигает своего максимума при $|x| = 1$, т. е. функция плотности стандартного нормального распределения липшицева. Стало быть, для нее будут иметь место приведенные выше оценки. Так как для функции $\varphi_{(0,1)}$ выполнено

$$\Delta x_k = \frac{1}{\sigma\sqrt{n}} \Rightarrow \left| \sum_{a \leq x_k \leq b} f(x_k)\Delta x_k - \int_a^b f(x) dx \right| = O\left(\frac{1}{\sigma\sqrt{n}}\right),$$

то

$$P(S_n \in A) = \sum_{k \in A} B_{n,p}(k) = \int_a^b \varphi_{(0,1)}(x) dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Последняя формула — это уточненная теорема Муавра—Лапласа. Из приведенных рассуждений следует, что погрешность $O\left(\frac{1}{\sqrt{n}}\right)$ зависит от a и b , но ее можно сделать и не зависящей от этих параметров. Приведем без доказательства следующее утверждение.

Теорема (уточнение теоремы Муавра - Лапласа).

$$\sup_{a < b} \left| P \left(\frac{S_n - np}{\sigma \sqrt{n}} \in [a, b] \right) - \int_a^b \varphi_{(0,1)}(x) dx \right| < \frac{1}{\sigma \sqrt{n}}.$$

Эта форма записи позволяет оценивать погрешность при замене биномиального распределения нормальным вне зависимости от отрезка $[a, b]$. При этом мы можем получать равномерные оценки и в случаях $a = -\infty$ или $b = \infty$. Заметим, что в последних двух случаях точность приближения будет оцениваться сверху величиной $\frac{1}{2\sigma\sqrt{n}}$.

Итак, для приближенного вычисления биномиального распределения мы доказали две теоремы: Пуассона и Муавра–Лапласа. Когда какую применять? Для обоснования выбора нужно сравнить погрешности соответствующих приближений. В теореме Пуассона оценка точности приближения равна $p \min(1, np)$, а в теореме Муавра–Лапласа при малых p

$$\frac{1}{\sigma \sqrt{n}} \sim \frac{1}{\sqrt{np}}.$$

По тому, какая из двух погрешностей меньше, выбираем и соответствующую аппроксимацию. Если n велико, а np «сравнимо с единицей», применяем теорему Пуассона. Если же при этом и величина np велика, — теорему Муавра–Лапласа.

Глава 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

До этого мы изучали комбинаторную теорию вероятностей. Теперь приступаем к изучению аналитической части курса.

Определение. Измеримое отображение $\xi: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ — случайная величина, где \mathcal{B} — борелевская σ -алгебра (минимальная σ -алгебра, порожденная интервалами). Измеримое отображение

$$\bar{\xi} = (\xi_1, \dots, \xi_k): (\Omega, \mathcal{F}) \rightarrow (R^k, \mathcal{B}^k),$$

где ξ_i — случайные величины, называется k -мерным случайным вектором.

Напомним, что такое измеримость. Полный прообраз множества B для отображения $\xi(\omega)$ есть множество $\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}$. Тогда ξ измеримо, если $\xi^{-1}(B) \in \mathcal{F}$ для каждого $B \in \mathcal{B}$. Мы требуем измеримость для корректности записи

$$\mathbf{P}(\xi \in B) = \mathbf{P}(\omega : \xi(\omega) \in B) = \mathbf{P}(\xi^{-1}(B)).$$

Определение. Распределением случайной величины ξ называется функция множества $\mathcal{P}_\xi(B) = \mathbf{P}(\xi \in B)$.

Упражнение. Проверить, что $\mathcal{P}_\xi(\cdot)$ — σ -аддитивная мера на вещественной прямой.

Определение. Функция $F_\xi = \mathcal{P}_\xi((-\infty, t))$ называется функцией распределения случайной величины ξ .

Отметим, что любая функция распределения однозначно восстанавливает само распределение. В самом деле, в силу аддитивности вероятностной меры имеем $F_\xi(b) - F_\xi(a) = \mathcal{P}_\xi([a, b]) \forall a < b$, так как $\mathcal{P}_\xi((-\infty, b)) - \mathcal{P}_\xi((-\infty, a)) = \mathcal{P}_\xi([a, b])$. Причем на полукольце открытых справа интервалов введенная функция множества обладает свойством σ -аддитивности. А по теореме о продолжении меры этого достаточно, чтобы на σ -алгебре всех борелевских множеств \mathcal{B} (минимальной σ -алгебре, содержащей указанные полуоткрытые интервалы) однозначно определить меру $\mathcal{P}_\xi(\cdot)$.

Классификация распределений (случайных величин).

- 1. Дискретные распределения.** Пусть случайная величина ξ принимает не более чем счетный набор значений $\{a_i\}$. В этом случае распределение \mathcal{P}_ξ называют *дискретным*. Точки a_i называются *атомами* \mathcal{P}_ξ , а $p_i = \mathbf{P}(\xi = a_i) = \mathcal{P}_\xi(\{a_i\}) > 0$ — *массами* соответствующих атомов.

Типичный вид дискретной функции распределения в случае возрастающего набора атомов следующий: $F_\xi(t) = \mathbf{P}(\xi < t) = 0$ при $t \leq a_1$; если $t \in (a_1, a_2]$, то $\mathbf{P}(\xi < t) = \mathbf{P}(\xi = a_1) = p_1$; при $t \in (a_2, a_3]$ имеем $\mathbf{P}(\xi < t) = \mathbf{P}(\xi = a_2) + \mathbf{P}(\xi = a_1) = p_1 + p_2$ и так далее.

Заметим, что для этой разрывной функции имеет место непрерывность слева

$$\mathbf{P}(\xi < t) = \mathbf{P}(\xi = a_1) + \mathbf{P}(\xi = a_2), \quad t \in (a_2, a_3].$$

В результате мы имеем монотонно неубывающую ступенчатую функцию со скачками в атомах этого распределения, величины скачков — массы соответствующих атомов.

- 2. Непрерывное распределение** — это распределение, у которого нет ни одного атома. Этот класс расщепляется на два.

- 2.1 Абсолютно непрерывные распределения.** Если распределение представимо в виде интеграла Лебега как функции множества от некоторой неотрицательной функции $p(t)$, т.е. $\mathcal{P}_\xi(A) = \int_A p(t) dx \forall A \in \mathcal{B}$, то распределение называется абсолютно непрерывным с плотностью распределения $p(t)$.

Например, в схеме геометрической вероятности $p(t) = const$, а в теореме Муавра—Лапласа мы рассматривали стандартное нормальное распределение

$$\mathcal{P}_\xi(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dx.$$

Плотность также называют обобщенной производной или производной Радона—Никодима абсолютно непрерывной меры \mathcal{P} относительно меры Лебега:

$$p(t) = \frac{d\mathcal{P}_\xi}{d\Lambda}(t).$$

Для нас важно, что плотность и функция распределения находятся во взаимно-однозначном соответствии. Если $A = (\infty, t)$, то

$$F_{\xi}(t) = \int_{-\infty}^t p(x) dx.$$

Если p непрерывна, т.е. $p \in C(\mathbb{R})$, то $F_{\xi}(t)$ — непрерывно дифференцируемая функция и $F'_{\xi}(t) \equiv p(t)$. В общем случае можно утверждать, что $F'(t) = p(t)$ для почти всех t , так как любая монотонная функция почти всюду дифференцируема, и что ее производная совпадает с плотностью, *но только для абсолютно непрерывных распределений*. Заметим, что под плотностью мы понимаем функциональный класс эквивалентности (так как она определяется с точностью до множества меры ноль). Значит, для вычисления плотности достаточно продифференцировать функцию распределения там, где есть производная и произвольным образом доопределить ее в остальных точках. При этом, если у нас нет информации о характере рассматриваемого распределения, мы должны проверить условие нормировки полученной производной — интеграл по всей прямой от этой функции с необходимостью должен равняться 1.

2.2 Сингулярные распределения. (*Непрерывные, но не абсолютно непрерывные*). Возникает вопрос существования таких распределений. Приведем пример, который в вероятностной литературе называется «Канторова лестница». Рассматривается случай, когда $\text{supp}F = [0, 1]$, $F(0) = 0$ и $F(1) = 1$. Делим отрезок на три равные части, причем на интервале $(1/3, 2/3)$ полагаем $F(t) = 1/2$. Оставшиеся два отрезка мы вновь делим на три равные части и на средних интервалах задаем функцию как полусумму двух ближайших справа и слева уже заданных значений этой функции. Тем самым, мы определили рекуррентную процедуру задания функции распределения Кантора на отрезках «троичного» деления. Построенная функция распределения, очевидно, непрерывна, поскольку по построению (так называемая, «диадическая схема» деление отрезка $[0, 1]$) все двоично-рациональные точки $k/2^n \in [0, 1]$ принадлежат области значений этой монотонной функции. Напомним, что двоично-рациональные точки вида $k/2^n$ образуют всюду плотное множество на прямой. Следовательно, в силу монотонности $F(t)$ разрывов у нее быть не может.

Далее, по построению функция кусочно-постоянна, поскольку для любой двоично-рациональной точки из $(0, 1)$ существует отрезок в области определения, где функция распределения Кантора тождественно равна этому двоично-рациональному числу. Суммируя меры отрезков постоянства, получаем

$$\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \dots = \frac{1/3}{1 - 2/3} = 1.$$

Значит, дополнительное множество (*канторово множество*) имеет нулевую меру Лебега. Заметим, что на всех построенных «троичных» интервалах производная F' тождественно равна нулю.

Предположим, что F — абсолютно непрерывная функция распределения. Тогда для любых точек t_1 и t_2 , одновременно принадлежащих любому из построенных выше троичных интервалов, имеем

$$0 \equiv F_\xi(t_1) - F_\xi(t_2) = \int_{t_1}^{t_2} p(x) dx,$$

т.е. $p(t) = 0$ почти всюду (по мере Лебега). Следовательно $\forall A$

$$\mathcal{P}_\xi(A) = \int_A 0 dt = 0.$$

Получили противоречие, так как $\mathcal{P}_\xi(R) = 1$. Таким образом, F не является абсолютно непрерывной.

3. Смеси распределений. Этот класс получается как производный от первых двух. Прежде дадим следующее

Определение. *Смесью* двух распределений \mathcal{P}_1 и \mathcal{P}_2 с *весом* α называется распределение $\alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2$, которое еще называют *выпуклой линейной комбинацией* двух распределений.

Понятно, что мы можем определить выпуклые линейные комбинации любого конечного числа распределений — это такие же смеси.

Класс смесей распределений определяется как всевозможные распределения, в которых первая компонента смеси \mathcal{P}_1 — непрерывное распределение, а вторая \mathcal{P}_2 — дискретное.

Примеры наиболее распространенных распределений.

1. Дискретные распределения (случайные величины).

- *Бернуллиевская (простейшая) случайная величина.*

Она принимает только два значения: 1 с вероятностью p , и 0 с вероятностью $1 - p$. Запишем это так:

$$\xi = \begin{cases} 1, & p, \\ 0, & 1 - p. \end{cases}$$

- *Биномиальное распределение.*

Это распределение имеет атомы $0, \dots, n$ с массами p_0, \dots, p_n :

$$\xi = \begin{cases} 0, \\ \vdots \\ k, & p_k = C_n^k p^k (1-p)^{n-k}, \\ \vdots \\ n. \end{cases}$$

- *Гипергеометрическое распределение.*

Это распределение имеет атомы $0, 1, \dots, n$ (при этом $n \leq \min\{N_1, N_2\}$):

$$\xi = \begin{cases} 0, \\ \vdots \\ k, \\ \vdots \\ n. \end{cases} \quad p_k = \frac{C_{N_1}^k C_{N_2}^{m-k}}{C_{N_1+N_2}^m},$$

- *Пуассоновское распределение.*

Для него множество атомов совпадает с \mathbb{Z}_+ :

$$\xi = \begin{cases} 0, \\ \vdots \\ k, \\ \vdots \end{cases} \quad p_k = \frac{\lambda^k}{k!} e^{-\lambda}.$$

2. Абсолютно непрерывные распределения.

- *Гауссовское распределение с параметрами α, σ .*

Плотность гауссовского распределения

$$p(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\alpha)^2/2\sigma^2}, \quad \alpha \in R, \quad \sigma > 0.$$

Это пример двупараметрического семейства плотностей.

- *Равномерное распределение.*

Случайная величина ξ распределена на $[a, b]$ с плотностью

$$p(t) = \begin{cases} \frac{1}{b-a}, & t \in [a, b], \\ 0, & \text{иначе.} \end{cases}$$

- *Показательное (экспоненциальное) распределение с параметром $\alpha > 0$.*

Плотность этого распределения задается формулой

$$p(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0, \\ 0, & \text{иначе.} \end{cases}$$

Лемма непрерывности. Свойства функций распределения.

Лемма (свойство непрерывности меры). Пусть $\{A_i\}_{i=1}^{\infty}$ – монотонно возрастающий поток событий, то есть $A_n \subseteq A_{n+1}$ при всех n . Тогда

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n).$$

ДОКАЗАТЕЛЬСТВО. Введём события $B_0 = A_1, B_n = A_{n+1} \setminus A_n, n \geq 1$. Легко заметить, что $B_i \cap B_j = \emptyset$, если $i \neq j$. Используя свойство σ -аддитивности меры, придём к равенству

$$\mathbf{P}\left(\bigcup_{i=0}^{\infty} B_i\right) = \sum_{i=0}^{\infty} \mathbf{P}(B_i).$$

Очевидно, что $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=0}^{\infty} B_i$. Тогда следующая цепочка равенств завершает доказательство:

$$\sum_{i=0}^{\infty} \mathbf{P}(B_i) = \lim_{m \rightarrow \infty} \sum_{i=0}^m \mathbf{P}(B_i) = \lim_{m \rightarrow \infty} \mathbf{P}\left(\bigcup_{i=0}^m B_i\right) = \lim_{m \rightarrow \infty} \mathbf{P}(A_{m+1}). \quad \square$$

Упражнение. Доказать, что если некая функция множества обладает свойством конечной аддитивности и свойством непрерывности для любого возрастающего потока событий, то эта функция будет σ -аддитивной, т. е. мерой.

Теперь сформулируем двойственный аналог доказанного утверждения.

Лемма. Пусть $\{A_i\}_{i=1}^{\infty}$ – монотонно убывающий поток событий, т.е. $A_n \supseteq A_{n+1}$ при всех n . Тогда

$$\mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{m \rightarrow \infty} \mathbf{P}(A_m).$$

ДОКАЗАТЕЛЬСТВО. Согласно принципу двойственности $\mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{\infty} \bar{A}_i\right)$, причём \bar{A}_i удовлетворяют условию предыдущей теоремы. Тогда

$$1 - \lim_{m \rightarrow \infty} \mathbf{P}(\bar{A}_m) = 1 - \lim_{m \rightarrow \infty} (1 - \mathbf{P}(A_m)) = \lim_{m \rightarrow \infty} \mathbf{P}(A_m). \quad \square$$

В дальнейшем на любой из этих двух вариантов мы будем ссылаться, как на *лемму непрерывности*.

Упражнение. Доказать следующие свойства функции распределения:

1. Функция распределения монотонно не убывает: $F_{\xi}(t) \leq F_{\xi}(t + \Delta) \forall \Delta > 0 \forall t$.
2. Поведение на бесконечности: $\lim_{t \rightarrow \infty} F_{\xi}(t) = 1, \lim_{t \rightarrow -\infty} F_{\xi}(t) = 0$.
3. Функция распределения непрерывна слева: $\lim_{\Delta \rightarrow 0+0} F_{\xi}(t - \Delta) = F_{\xi}(t)$.

Отметим, что класс всех функций распределения задаётся этими тремя свойствами.

Перейдём к рассмотрению распределений случайных векторов.

Пусть $\bar{\xi} = (\xi_1, \dots, \xi_n): (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ — случайный вектор. Под распределением этого вектора понимается $\mathcal{P}_{\bar{\xi}}(A) = \mathbf{P}(\bar{\xi} \in A)$, а под функцией распределения

$$F_{\bar{\xi}}(t_1, \dots, t_n) = \mathbf{P}(\{\omega : \xi_1(\omega) < t_1\} \cap \dots \cap \{\omega : \xi_n(\omega) < t_n\}) = \mathbf{P}(\xi_1 < t_1, \dots, \xi_n < t_n).$$

Говорят также, что распределение вектора $\bar{\xi} = (\xi_1, \dots, \xi_n)$ — это *совместное* распределение набора ξ_1, \dots, ξ_n .

Распределение случайного вектора назовём *дискретным*, если оно сосредоточено в не более чем счетном числе атомов. Распределение называется *непрерывным*, если оно не имеет атомов, т.е. $\mathbf{P}(\bar{\xi} = \bar{a}) = 0$ для любого вектора \bar{a} .

Абсолютно непрерывным распределением назовем такое, которое допускает представление

$$\mathcal{P}_{\bar{\xi}}(A) = \int_A p(t) dt = \int \dots \int_{(t_1, \dots, t_n) \in A} p(t_1, \dots, t_n) dt_1 \dots dt_n,$$

где $p(t)$ — *плотность* совместного распределения ξ_1, \dots, ξ_n .

Непрерывное распределение называется *сингулярным*, если оно не допускает такого представления.

Пример. Бросаем точку наудачу в единичный квадрат. Рассматривая координаты этой точки, получим случайный вектор $\bar{\xi} = (\xi_1, \xi_2)$, который равномерно распределён с единичной плотностью:

$$\mathcal{P}(A) = \Lambda_2(A) = \iint_A 1 dt_1 dt_2.$$

Продолжая пример, рассмотрим случайный вектор $\bar{\xi}' = (\xi_1, \xi_1)$. Очевидно, носителем распределения этого вектора является диагональ квадрата, то есть одномерное многообразие. Таким образом, в виде двойного интеграла вероятность попадания в множество A для такого вектора представить нельзя. Такого типа распределения также называются *вырожденными* (с носителями на многообразиях меньшей размерности, чем исходное).

Независимость случайных величин. Свертка распределений.

Прежде всего, нам понадобится важное понятие независимости случайных величин.

Определение. Совокупность (ξ_1, \dots, ξ_n) случайных величин, заданных на одном пространстве называется семейством *независимых случайных величин*, если для любых борелевских множеств $A_i \in \mathcal{B}$ совместное распределение этих случайных величин факторизуется на маргинальные (одномерные) распределения

$$\mathcal{P}_{\bar{\xi}}(A_1 \times \dots \times A_n) = \mathbf{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \prod_{i=1}^n \mathbf{P}(\xi_i \in A_i) = \prod_{i=1}^n \mathcal{P}_{\xi_i}(A_i).$$

Напомним, что знак « \cdot » обозначает операцию « \cap ».

Заметим, что если $A_m = \mathbb{R}$, то $\{\xi_{im} \in \mathbb{R}\} = \Omega$, и это событие исчезнет в пересечении под знаком вероятности в вышеприведенном определении. Значит, если (ξ_1, \dots, ξ_n) — независимый набор, то любое его подсемейство также состоит из независимых случайных величин.

Упражнение. Доказать, что если ξ_1, \dots, ξ_n — дискретные случайные величины с соответствующими наборами атомов $\{a_i^{(1)}\}, \dots, \{a_i^{(n)}\}$, то независимость этих величин эквивалентна факторизации совместной точечной массы: $\forall a_i^{(k)} \in \{a_j^{(k)}\}, k = 1, \dots, n,$

$$\mathbf{P}(\xi_1 = a_1, \dots, \xi_n = a_n) = \prod_{i=1}^n \mathbf{P}(\xi_i = a_i).$$

Критерий независимости для случайных величин с абсолютно непрерывным совместным распределением.

Пусть случайный вектор $\bar{\xi} = (\xi_1, \xi_2)$ имеет абсолютно непрерывное распределение в \mathbb{R}^2

$$\mathcal{P}_{\bar{\xi}}(A) = \iint_A p(t_1, t_2) dt_1 dt_2,$$

где $p(t_1, t_2) \geq 0$ — плотность распределения вектора $\bar{\xi}$.

Лемма. Если распределение случайного вектора допускает вышеприведенное представление, то координаты ξ_1 и ξ_2 также имеют абсолютно непрерывные распределения в \mathbb{R} , т. е. у них существуют плотности (маргинальные) распределения p_{ξ_1} и p_{ξ_2} .

ДОКАЗАТЕЛЬСТВО. Мы хотим получить интегральное представление для $\mathcal{P}_{\xi_1}(B)$

$$\mathcal{P}_{\xi_1}(B) = \mathcal{P}_{\bar{\xi}}(B \times \mathbb{R}) = \mathbf{P}(\xi_1 \in B, \xi_2 \in \mathbb{R}) = \iint_{B \times \mathbb{R}} p(t_1, t_2) dt_1 dt_2 =$$

[используем теорему Фубини]

$$= \int_B \left(\int_{\mathbb{R}} p(t_1, t_2) dt_2 \right) dt_1, \text{ значит, } p_{\xi_1}(t_1) = \int_{\mathbb{R}} p(t_1, t_2) dt_2.$$

Аналогично проводятся рассуждения для ξ_2 и для случайных векторов. \square

Теорема (критерий независимости). Пусть вектор $\bar{\xi} = (\xi_1, \dots, \xi_n)$ имеет абсолютно непрерывное распределение. Тогда $\{\xi_i\}_{i=1}^n$ — независимы тогда и только тогда, когда $\forall t_i$

$$p_{\bar{\xi}}(t_1, \dots, t_n) = \prod_{i=1}^n p_{\xi_i}(t_i),$$

где p_{ξ_i} — маргинальные плотности.

ДОКАЗАТЕЛЬСТВО. Достаточно провести рассуждения для двумерного случая.

(\rightarrow) Чтобы доказать равенство для всех борелевских множеств, по теореме о продолжении меры его достаточно доказать только для класса канонических прямоугольников. Имеем

$$\mathcal{P}_{\bar{\xi}}(A) = \mathcal{P}_{\bar{\xi}}([a_1, b_1] \times [a_2, b_2]) = \mathbf{P}(\xi_1 \in [a_1, b_1], \xi_2 \in [a_2, b_2]) =$$

[так как ξ_1 и ξ_2 независимы]

$$= \mathbf{P}(\xi_1 \in [a_1, b_1])\mathbf{P}(\xi_2 \in [a_2, b_2]) =$$

[так как ξ_1, ξ_2 — абсолютно непрерывны]

$$= \int_{a_1}^{b_1} p_{\xi_1}(t_1) dt_1 \int_{a_2}^{b_2} p_{\xi_2}(t_2) dt_2 =$$

[по теореме Фубини]

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} p_{\xi_1}(t_1)p_{\xi_2}(t_2) dt_1 dt_2. \quad \square$$

Упражнение. Показать достаточность в критерии независимости.

Аддитивное преобразование. Рассмотрим некоторое измеримое отображение

$$f: (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^k, \mathcal{B}^k), n, k = 1, 2, \dots$$

Тогда k -мерная величина $\bar{\eta} := f(\bar{\xi})$ — суперпозиция отображения f и вектора $\bar{\xi}$, будет случайной величиной (так как суперпозиция измеримых отображений измерима). Мы хотим понять, как считать распределения таких преобразований.

Рассмотрим *аддитивное преобразование*

$$f(\bar{\xi}) = \sum_{i=1}^n \xi_i.$$

Этот тип функций для нас очень важен. Скажем, если ξ_i имеют структуру индикаторных случайных величин (т. е. принимающих только значения 1 или 0), то, например, число «успехов» (или число «1») в соответствующей схеме Бернулли как раз и представляется в виде

$$S_n = \sum_{i=1}^n \xi_i.$$

Распределение суммы независимых случайных величин называют *свертками распределений слагаемых*.

Сначала получим формулу свертки для одного частного случая. Пусть ξ_1 имеет произвольное распределение, ξ_2 — произвольное дискретное, т. е. $\xi_2 \in \{a_2\}$, $a_i < a_{i+1}$. Задача: найти распределение суммы этих случайных величин $\mathbf{P}(\xi_1 + \xi_2 \in A)$, которое и называется *сверткой* распределений \mathcal{P}_{ξ_1} и \mathcal{P}_{ξ_2} .

Применим формулу полной вероятности. Напомним, что если $\{H_i\}$ — разбиение пространства элементарных исходов (полная группа событий), то для любого события A

$$\mathbf{P}(A) = \sum_i \mathbf{P}(A|H_i)\mathbf{P}(H_i) = \sum_i \mathbf{P}(A \cap H_i), \quad \text{где } \mathbf{P}(H_i) > 0.$$

В качестве H_i возьмем $\{\xi_2 = a_i\}$ — это полная группа событий, так как $\forall i \neq j$ $H_i \cap H_j = \emptyset$ и $\cup H_i = \Omega$. Тогда

$$\begin{aligned} \mathbf{P}(\xi_1 + \xi_2 \in A) &= \sum_i \mathbf{P}(\xi_1 + \xi_2 \in A, \xi_2 = a_i) = \sum_i \mathbf{P}(\xi_1 + a_i \in A, \xi_2 = a_i) = \\ &= \sum_i \mathbf{P}(\xi_1 \in A - a_i, \xi_2 = a_i) = \end{aligned}$$

[(где запись $A - a_i$ обозначает снос влево множества A на число a_i), и с учетом независимости событий получаем]

$$= \sum_i \mathbf{P}(\xi_1 \in A - a_i) \mathbf{P}(\xi_2 = a_i) =$$

[так как атомы упорядочены, то $\mathbf{P}(\xi_2 = a_i) = F_{\xi_2}(a_{i+1}) - F_{\xi_2}(a_i) = \Delta F_{\xi_2}(a_i)$]

$$= \sum_i \mathbf{P}(\xi_1 \in A - a_i) \Delta F_{\xi_2}(a_i) =$$

[Последняя сумма называется *интегральной суммой Римана—Стилтьеса*, которая для дискретной случайной величины ξ_2 может быть представлена как соответствующий интеграл Римана—Стилтьеса]

$$= \int \mathcal{P}_{\xi_1}(A - a) dF_{\xi_2}(a).$$

Такой интеграл называется *интегралом типа свертки*; здесь необходимо отметить, что в данной конструкции интеграла Римана—Стилтьеса F_{ξ_2} — функция ограниченной вариации (точнее, ее полная вариация равна 1) и что для дискретной случайной величины все приращения функции $F_{\xi_2}(x)$ вне окрестностей атомов равны нулю.

Лемма. Пусть случайные величины ξ_1 и ξ_2 независимы и имеют абсолютно непрерывные распределения. Тогда

$$p_{\xi_1 + \xi_2}(u) = \int_{\mathbb{R}} p_{\xi_1}(u - v) p_{\xi_2}(v) dv.$$

ДОКАЗАТЕЛЬСТВО. Имеем $\mathbf{P}(\xi_1 + \xi_2 \in A) = \mathbf{P}((\xi_1, \xi_2) \in B)$, где $B = \{(x, y) : x + y \in A\}$. Тогда по критерию независимости

$$\mathbf{P}(\xi_1 + \xi_2 \in A) = \iint_B p_{\xi_1}(t) p_{\xi_2}(s) dt ds =$$

[сделаем замену переменных $t + s = u, s = v$, якобиан преобразования равен 1, тогда]

$$= \iint_{\tilde{B} = \{u \in A, v \in \mathbb{R}\}} p_{\xi_1}(u - v) p_{\xi_2}(v) du dv =$$

[по теореме Фубини]

$$= \int_A \left[\int_{\mathbb{R}} p_{\xi_1}(u-v)p_{\xi_2}(v) dv \right] du.$$

Следовательно,

$$p_{\xi_1+\xi_2}(u) = \int_{\mathbb{R}} p_{\xi_1}(u-v)p_{\xi_2}(v) dv.$$

Последний интеграл — это *свертка* двух плотностей. \square

Упражнение. Рассмотреть случай, когда ξ_1 и ξ_2 равномерно распределены на $[0, 1]$ и независимы. Построить свертку их распределений. Посчитать также плотность свертки трех и четырех равномерно распределенных плотностей. Заметим, что с увеличением числа независимых случайных величин ξ_i график плотности распределения их суммы всё более будет напоминать график плотности распределения Гаусса.

Верхние и нижние пределы последовательности множеств. Теорема Бореля–Кантелли.

Определение. Пусть имеется счетный набор событий $\{A_i\}$. Введем два события

$$A^+ = \bigcap_n \bigcup_{k \geq n} A_k, \quad A^- = \bigcup_n \bigcap_{k \geq n} A_k.$$

События A^+ и A^- называются соответственно верхним и нижним пределом последовательности $\{A_i\}$.

Лемма. Справедливо вложение $A^+ \supseteq A^-$.

ДОКАЗАТЕЛЬСТВО. Действительно, пусть элементарный исход $\omega \in A^-$. Тогда он принадлежит хотя бы одному из пересечений:

$$\exists n_0 : \omega \in \bigcap_{k \geq n_0} A_k \rightarrow \omega \in A_k \quad \forall k \geq n_0.$$

Значит,

$$\omega \in \bigcup_{k \geq n} A_k \quad \forall n,$$

т. е. $\omega \in A^+$. Следовательно, $A^- \subseteq A^+$. \square

Упражнение. Показать, что в рассматриваемой общности доказанное включение может быть строгим. Кроме того, привести пример, когда $A^- = A^+$.

Теорема (Борель–Кантелли). Пусть имеется счетная последовательность событий $\{A_i\}$. Тогда

1. Если $\sum_{i \geq 1} \mathbf{P}(A_i) < \infty$, то $\mathbf{P}(A^+) = 0$ (а так как $A^- \subseteq A^+$, то и $\mathbf{P}(A^-) = 0$).

2. Если $\{A_i\}$ независимы в совокупности (то есть любые конечные поднаборы независимы в совокупности) и $\sum_{i \geq 1} \mathbf{P}(A_i) = \infty$, то $\mathbf{P}(A^+) = 1$.

Эту теорему называют *законом нуля и единицы*. Докажем сначала две вспомогательные леммы.

Лемма 1. Для любого набора событий $\{A_i\}$ справедливо свойство *полуаддитивности меры* $\mathbf{P}(\cup A_i) \leq \sum \mathbf{P}(A_i)$.

ДОКАЗАТЕЛЬСТВО. Для конечного набора событий по уже доказанному свойству полуаддитивности вероятности имеем

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

Обозначим

$$B_N = \bigcup_{i=1}^N A_i.$$

Легко видеть, что B_N — расширяющийся поток событий и $\bigcup B_N = \bigcup A_i$. Тогда по лемме непрерывности меры

$$\mathbf{P}\left(\bigcup B_N\right) = \mathbf{P}\left(\bigcup A_i\right) = \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{i=1}^N A_i\right) \leq \lim_{N \rightarrow \infty} \sum_{i=1}^N P(A_i) = \sum_{i \geq 1} P(A_i). \quad \square$$

Лемма 2. Пусть события $\{A_i\}_{i=1}^M$ независимы. Тогда события $\{\bar{A}_i\}_{i=1}^M$ также независимы.

ДОКАЗАТЕЛЬСТВО. Достаточно доказать, что $\forall M \mathbf{P}\left(\bigcap_{i=1}^M \bar{A}_i\right) = \prod_{i=1}^M \mathbf{P}(\bar{A}_i)$. Имеем

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

[в силу независимости и формулы включения-исключения]

$$\begin{aligned} &= 1 - \sum_{k=1}^M \mathbf{P}(A_k) + \sum_{k_1 < k_2 \leq M} \mathbf{P}(A_{k_1})\mathbf{P}(A_{k_2}) - \sum_{k_1 < k_2 < k_3 \leq M} \mathbf{P}(A_{k_1})\mathbf{P}(A_{k_2})\mathbf{P}(A_{k_3}) + \dots = \\ &= \prod_{k=1}^M (1 - \mathbf{P}(A_k)) = \prod_{k=1}^M \mathbf{P}(\bar{A}_k). \quad \square \end{aligned}$$

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ.

(1) Пусть $B_n = \bigcup_{k \geq n} A_k$. Тогда B_n — убывающий поток, и по лемме непрерывности и свойству полуаддитивности меры получаем

$$\begin{aligned} \mathbf{P}(A^+) &= \lim_{n \rightarrow \infty} \mathbf{P} \left(\bigcup_{k \geq n} A_k \right) \leq \\ &\leq \lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbf{P}(A_k) = 0. \end{aligned}$$

Последнее равенство справедливо в силу того, что стоящая под знаком предела сумма является хвостом сходящегося ряда.

(2) Докажем, что в условиях второго пункта теоремы $\mathbf{P}(\bar{A}^+) = 0$. Имеем

$$\mathbf{P}(\bar{A}^+) = \mathbf{P} \left(\bigcup_n \bigcap_{k \geq n} \bar{A}_k \right) =$$

$[B_n = \bigcap_{k \geq n} \bar{A}_k$ — возрастающий поток, тогда используем лемму непрерывности]

$$= \lim_{n \rightarrow \infty} \mathbf{P} \left(\bigcap_{k \geq n} \bar{A}_k \right) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbf{P} \left(\bigcap_{k=n}^{n+m} \bar{A}_k \right) =$$

[по лемме 2]

$$= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \prod_{k=n}^{m+n} (1 - \mathbf{P}(A_k)) \leq$$

[Так как $1 - x \leq e^{-x}$]

$$\leq \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \exp \left\{ - \sum_{k=n}^{m+n} \mathbf{P}(A_k) \right\} = \lim_{n \rightarrow \infty} \exp \left\{ - \sum_{k=n}^{\infty} \mathbf{P}(A_k) \right\} = \lim_{n \rightarrow \infty} 0 = 0. \quad \square$$

В качестве одного из приложений теоремы Бореля–Кантелли рассмотрим вопрос сходимости рядов случайных величин

$$\sum_{i=1}^{\infty} \xi_i \equiv \sum_{i=1}^{\infty} \xi_i(\omega).$$

С точки зрения анализа это обычный функциональный ряд.

Следствие (Достаточные условия абсолютной почти наверное сходимости случайных рядов). Пусть ε_k — суммируемая последовательность положительных чисел. Введем $B_k = \{|\xi_k| > \varepsilon_k\}$. Тогда если $\sum \mathbf{P}(B_k) < \infty$, то ряд $\sum_{i=1}^{\infty} \xi_i(\omega)$ сходится абсолютно и равномерно по всем элементарным исходам.

В самом деле, в силу теоремы Бореля – Кантелли выполнены неравенства $|\xi_k| \leq \varepsilon_k$ $\forall k \geq n_0(\omega)$. Следовательно выполнен известный критерий абсолютной и равномерной сходимости функциональных рядов.

Упражнение. Пусть $\{\xi_k\}$ – последовательность независимых одинаково распределенных бернуллиевских случайных величин с вероятностью успеха $1/2$. Доказать, что функция распределения случайной величины $\eta = 2 \sum_{k=1}^{\infty} \frac{\xi_k}{3^k}$ есть “канторова лестница”, т.е. η имеет сингулярное распределение.

Сходимость последовательностей случайных величин.

Пусть на одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ задана бесконечная последовательность случайных величин $\{\xi_i\}$.

Определение 1. Говорят, что $\xi_n \xrightarrow{\text{п.н.}} \xi$ почти наверное (или с вероятностью 1), если $\xi_n(\omega) \rightarrow \xi(\omega) \forall \omega \in A_0$ и множество $A_0 = \{\omega : \xi_n(\omega) \rightarrow \xi(\omega)\}$ имеет полную меру, т.е. $\mathbf{P}(A_0) = 1$.

Определение 2. Говорят, что $\xi_n \xrightarrow{p} \xi$ по вероятности, если $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P}(|\xi_n - \xi| > \varepsilon) = 0$$

или в эквивалентной форме

$$\lim_{n \rightarrow \infty} \mathbf{P}(|\xi_n - \xi| \leq \varepsilon) = 1.$$

Определение 3. Говорят, что последовательность случайных величин ξ_n (или их распределений) слабо сходится (или сходится в основном, или сходится по распределению) к предельной случайной величине ξ (в наших обозначениях $\xi_n \Rightarrow \xi$), если $F_{\xi_n}(x) \rightarrow F_{\xi}(x)$ поточечно для всех точек непрерывности предельной функции $F_{\xi}(x)$.

Мы покажем, что из п.н. сходимости следует сходимость по вероятности, которая, в свою очередь, влечет за собой слабую сходимость. Обратные включения в этой цепочке неверны.

Сначала покажем, что из сходимости с вероятностью 1 следует сходимость по вероятности. Опишем множество элементарных исходов, в которых имеет место поточечная сходимость нашей последовательности. Для любого фиксированного ε , начиная с некоторого n , для благоприятных ω должно выполняться неравенство $|\xi_n(\omega) - \xi(\omega)| \leq \varepsilon$. При этом можно положить $\varepsilon = 1/m$ Тогда нижний предел этих событий (m фиксировано)

$$\bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ |\xi_k - \xi| \leq \frac{1}{m} \right\}$$

как раз и описывает приведенную выше логическую цепочку. Поскольку упомянутая логическая цепочка справедлива для каждого m , то мы получаем

$$A_0 = \{\omega : \xi_n(\omega) \rightarrow \xi\} = \bigcap_{m \geq 1} \bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ |\xi_k(\omega) - \xi(\omega)| \leq \frac{1}{m} \right\}.$$

Теорема (критерий сходимости почти наверное). *Последовательность случайных величин ξ_n почти наверное сходится к случайной величине ξ тогда и только тогда, когда*

$$\eta_n := \sup_{k \geq n} |\xi_k - \xi| \xrightarrow{p} 0.$$

Заметим, что $\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \leq \mathbf{P}(\sup_{k \geq n} |\xi_k - \xi| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$. Значит, из п.н. сходимости следует сходимость по вероятности.

ДОКАЗАТЕЛЬСТВО. Необходимость (\rightarrow). Пусть $A_0 = \{\omega : \xi_n(\omega) \rightarrow \xi\}$. Тогда для любого фиксированного m имеем

$$1 = \mathbf{P}(A_0) = [\text{можем убрать внешнее пересечение в силу монотонности меры}]$$

$$= \mathbf{P} \left(\bigcup_n \bigcap_{k \geq n} \left\{ |\xi_k - \xi| \leq \frac{1}{m} \right\} \right) =$$

[в силу леммы непрерывности]

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \mathbf{P} \left(\bigcap_{k \geq n} \left\{ |\xi_k - \xi| \leq \frac{1}{m} \right\} \right) = \lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{k \geq n} |\xi_k - \xi| \leq \frac{1}{m} \right) = \\ &= 1 - \lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{k \geq n} |\xi_k - \xi| > \frac{1}{m} \right). \end{aligned}$$

Таким образом,

$$\sup_{k \geq n} |\xi_k - \xi| \xrightarrow{p} 0. \quad \square$$

Упражнение. Доказать достаточность (\leftarrow) условий приведенной теоремы.

Упражнение. Пусть имеется счетный набор $\{B_n\}$ множеств полной меры, что означает $\mathbf{P}(B_n) = 1 \forall n$. Доказать, что $\mathbf{P}(\cap B_n) = 1$.

Покажем, что из сходимости по вероятности не следует п.н. сходимость. Рассмотрим в качестве вероятностного пространства единичную окружность с индуцированной мерой Лебега Λ на ней.

В качестве искомой последовательности случайных величин возьмем характеристические функции замкнутых дуг $\{\Delta_n\}$ единичной окружности:

$$\xi_1(\omega) = I_{\Delta_1}(\omega), \quad \xi_2 = I_{\Delta_2}(\omega), \dots, \xi_n = I_{\Delta_n}(\omega) \dots,$$

при условии, что дуги откладываются «впритык» (с одной общей точкой для двух соседних дуг), например, по часовой стрелке от некоторой начальной точки, причем $\Lambda(\Delta_n) = \frac{1}{n+1}$. При этом «цепочка» построенных дуг «опоясывает» окружность бесконечно много раз (так как $\sum \frac{1}{n} = \infty$).

Пусть ω_0 — произвольный элементарный исход, т. е. *наудачу* брошенная точка на рассматриваемую единичную окружность. Тогда

1. Существует счетная подпоследовательность дуг $\{\Delta_{n'}\}$ такая, что если $\omega_0 \in \Delta_{n'} \forall n'$, то $\xi_{n'}(\omega_0) = 1$.

2. Ясно, что существует другая счетная подпоследовательность $\{\Delta_{n''}\}$ такая, что если $\omega_0 \notin \Delta_{n''} \forall n''$, то $\xi_{n''}(\omega_0) = 0$, и значит, $\xi_n(\omega)$ не имеет предела ни в одной точке ω . Поэтому нет сходимости почти наверное. Но при этом

$$\mathbf{P}(\Delta_n) = \mathbf{P}(|\xi_n| > \varepsilon) = \frac{1}{n+1} \xrightarrow[n \rightarrow \infty]{} 0.$$

Следовательно, имеет место сходимость по вероятности: $\xi_n \xrightarrow[p]{} 0$. \square

Покажем, что из сходимости по вероятности следует слабая сходимость. По формуле полной вероятности имеем

$$\begin{aligned} \mathbf{P}(\xi_n < x) &= \mathbf{P}(\xi_n < x, |\xi_n - \xi| \leq \varepsilon) + \mathbf{P}(\xi_n < x, |\xi_n - \xi| > \varepsilon) \\ &\leq \mathbf{P}(\xi_n < x, |\xi_n - \xi| \leq \varepsilon) + \mathbf{P}(|\xi_n - \xi| > \varepsilon). \end{aligned}$$

Далее,

$$\begin{aligned} \mathbf{P}(|\xi_n - \xi| > \varepsilon) &\xrightarrow[n \rightarrow \infty]{} 0, \\ \mathbf{P}(\xi_n < x, |\xi_n - \xi| \leq \varepsilon) &\leq \mathbf{P}(\xi < x + \varepsilon, |\xi_n - \xi| \leq \varepsilon) \leq \mathbf{P}(\xi < x + \varepsilon), \end{aligned}$$

откуда $F_{\xi_n}(x) \leq F_\xi(x + \varepsilon) + o(1)$.

Теперь получим нижнюю оценку для $F_{\xi_n}(x)$. Имеем

$$\begin{aligned} \mathbf{P}(\xi_n < x) &\geq \mathbf{P}(\xi_n < x, |\xi_n - \xi| \leq \varepsilon) \geq \mathbf{P}(\xi < x - \varepsilon, |\xi_n - \xi| \leq \varepsilon) = \\ &= F_\xi(x - \varepsilon) - \mathbf{P}(\xi, x - \varepsilon, |\xi_n - \xi| > \varepsilon) \geq F_\xi(x - \varepsilon) - \mathbf{P}(|\xi_n - \xi| > \varepsilon). \end{aligned}$$

Таким образом,

$$F_\xi(x - \varepsilon) + o(1) \leq F_{\xi_n}(x) \leq F_\xi(x + \varepsilon) + o(1),$$

поэтому

$$F_\xi(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{\xi_n}(x) \leq \limsup_{n \rightarrow \infty} F_{\xi_n}(x) \leq F_\xi(x + \varepsilon),$$

где x — точка непрерывности функции $F_\xi(x)$. Значит, в силу произвольности ε предел существует, и $\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_\xi(x)$. \square

Покажем, что из слабой сходимости не следует сходимость по вероятности. Заметим, что в случае слабой сходимости даже не требуется задание последовательности случайных величин на одном вероятностном пространстве. Рассмотрим радемахеровскую случайную величину $\xi_1 = \begin{cases} 1, & 1/2, \\ -1, & 1/2. \end{cases}$ Эта величина имеет *симметричное* распределение,

так как $\mathcal{P}_{\xi_1} = \mathcal{P}_{-\xi_1}$. Положим $\xi_2 = -\xi_1, \xi_3 = -\xi_2 = \xi_1, \dots$

У всех введенных величин функции распределения совпадают, и значит, есть слабая сходимость. Но $|\xi_k - \xi_{k-1}| = 2 \forall k$, и поэтому сходимости по вероятности нет. \square

Упражнение. Доказать, что если последовательность $\{\xi_n\}$ задана на одном вероятностном пространстве и слабо сходится к вырожденной случайной величине (т. е. постоянной с вероятностью 1), то имеет место и сходимость по вероятности.

Упражнение. Пусть $\{\xi_i\}$ — последовательность одинаково распределенных независимых случайных величин. Доказать, что условие $\sum_{k \geq 1} \mathbf{P}(|\xi_1| > k) < \infty$ необходимо и достаточно для того, чтобы при $n \rightarrow \infty$

$$n^{-1} \max_{i \leq n} |\xi_i| \xrightarrow{п.н.} 0.$$

Существование бесконечных последовательностей независимых событий и случайных величин.

Напомним, что в лемме Бореля — Кантелли впервые рассматривался счетный набор независимых в совокупности событий $\{A_i\}$. Обсудим вопрос о существовании такого набора.

С конечными наборами проблемы не возникает. Рассмотрим модель двумерной геометрической вероятности: бросаем наудачу точку в единичный квадрат. В этом квадрате любые две перпендикулярные полосы с границами, параллельными смежным сторонам квадрата, как мы уже ранее выяснили, образуют независимые события. Аналогичным образом, в \mathbb{R}^3 можно построить независимые события A, B, C , реализовав их как слои с гранями параллельными соответственно трем смежным граням единичного куба с общей вершиной. Эта же процедура переносится на евклидово пространство \mathbb{R}^n , и, тем самым, мы можем привести пример n независимых в совокупности событий в n -мерном единичном кубе.

Значит, для построения бесконечного набора независимых событий мы должны задать меру в бесконечномерном пространстве (здесь нужно использовать теорему о продолжении меры), а потом рассматривать бесконечномерный куб и слои в нем. Однако мы сделаем по-другому.

Рассмотрим так называемый *пример Радемахера*. В качестве вероятностного пространства возьмем отрезок $[0, 1]$ с лебеговой мерой. Начнем процесс «диадического» деления: сначала делим пополам отрезок $[0, 1]$, затем образовавшиеся половинки снова делим пополам и т. д. Получившиеся точки (концы отрезков) будут двоично-рациональными точками вида $\{k/2^n\}$, $k \geq 0, n \geq 1$. Это множество всюду плотно на $[0, 1]$. Начнем построение счетного набора $\{\xi_i\}$ бернуллиевских случайных величин, независимых в совокупности. На первом шаге диадической схемы отрезок $[0, 1]$ делится на две половины: на промежутки $\Delta_1^{(1)} = [0, 1/2)$ и $\Delta_2^{(1)} = [1/2, 1]$. Введем случайную величину

$$\xi_1(\omega) = I_{\Delta_1^{(1)}}(\omega) = \begin{cases} 1, & \omega \in [0, 1/2), \\ 0, & \omega \in [1/2, 1]. \end{cases}$$

На втором шаге полуинтервал $\Delta_1^{(1)}$ делим на $\Delta_1^{(2)} = [0, 1/4)$ и $\Delta_2^{(2)} = [1/4, 1/2)$, а промежуток $\Delta_2^{(1)}$ делим на $\Delta_3^{(2)} = [1/2, 3/4)$ и $\Delta_4^{(2)} = [3/4, 1]$. Вводим случайную величину

$$\xi_2(\omega) = I_{\Delta_1^{(2)} \cup \Delta_3^{(2)}}(\omega) = I_{\Delta_1^{(2)}}(\omega) + I_{\Delta_3^{(2)}}(\omega).$$

Продолжая далее, построим бесконечный набор случайных величин (так называемые *функции Радемахера*)

$$\xi_n(\omega) = \sum_{k=1}^{2^n-1} I_{\Delta_{2^k-1}^{(n)}}(\omega), \quad n = 1, 2, \dots$$

Очевидно, мы получили симметричные бернуллиевские случайные величины, так как носитель любой из построенных индикаторных функций имеет меру $1/2$.

Обозначение через $\{\text{supp}\xi_n\}$ объединение нечетных интервалов на n -ом шаге описанной диадической процедуры (т. е. носитель функции (случайной величины) ξ_n).

Отметим важные свойства введенных носителей: $\forall n \geq 1$ имеем

$$\Lambda(\{\text{supp}\xi_n\} \cap \{\text{supp}\xi_{n+1}\}) =$$

[в силу симметричности схемы]

$$= \Lambda(\{\dots\} \cap \overline{\{\dots\}}) = \Lambda(\overline{\{\dots\}} \cap \{\dots\}) = \Lambda(\overline{\{\dots\}} \cap \overline{\{\dots\}}) = \frac{1}{4}.$$

Аналогично $\Lambda(\{\text{supp}\xi_n\} \cap \{\text{supp}\xi_k\}) = \dots = \frac{1}{4}$ при $n \neq k$.

Упражнение. Доказать, что $\Lambda(\{\text{supp}\xi_n\} \cap \{\text{supp}\xi_{n+k_1}\} \cap \dots \cap \{\text{supp}\xi_{n+k_{m-1}}\}) = \frac{1}{2^m}$. Также проверить это равенство для любой комбинации дополнений рассматриваемых событий.

Мы хотим доказать, что $\forall a_i \in \{0, 1\}$

$$\mathbf{P}(\xi_1 = a_1, \dots, \xi_n = a_n) = \prod_{i=1}^n \mathbf{P}(\xi_i = a_i).$$

Если $a_k = 1$, то событие $\{\xi_k = a_k\}$ означает попадание элементарного исхода в $\{\text{supp}\xi_k\}$, если $a_k = 0$, то $\{\xi_k = a_k\}$ — попадание в $\overline{\{\text{supp}\xi_k\}}$. Следовательно, $\forall a_1, \dots, a_n$ и $\forall n$

$$\mathbf{P}(\xi_1 = a_1, \dots, \xi_n = a_n) = \frac{1}{2^n} = \prod_{i=1}^n \mathbf{P}(\xi_i = a_i)$$

[как мера пересечения носителей и их дополнений]. Тогда $\{\xi_i\}_{i=1}^{\infty}$ — это счетный набор бернуллиевских независимых в совокупности случайных величин, заданных на одном вероятностном пространстве.

Таким образом, последовательность подмножеств $A_i = \{\text{supp}\xi_i\}$ (или $\overline{\{\text{supp}\xi_i\}}$) образует счетное семейство независимых событий (тем самым доказано их существование).

Далее попробуем расширить класс введенных случайных величин.

Лемма 1. Пусть $\{\xi_i\}$ — счетный набор независимых бернуллиевских случайных величин с вероятностью успеха $p = 1/2$. Рассмотрим случайную величину

$$\eta = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k}.$$

Тогда η равномерно распределена на $[0, 1]$.

ДОКАЗАТЕЛЬСТВО. Заметим, что если $\eta = \frac{k}{2^n} \in [0, 1]$, то, начиная с $(n+1)$ -го номера, все $\xi_k = 0$. Имеем

$$\mathbf{P}(\eta \in [0, 1/2)) = \mathbf{P}(\xi_1 = 0) = \frac{1}{2},$$

$$\mathbf{P}(\eta \in [0, 1/4]) = \mathbf{P}(\xi_1 = 0, \xi_2 = 0) = \frac{1}{4},$$

$$\mathbf{P}(\eta \in (1/2, 1]) = \mathbf{P}(\xi_1 = 1) = \frac{1}{2},$$

$$\mathbf{P}(\eta \in (1/4, 1/2]) = \mathbf{P}(\xi_1 = 0, \xi_2 = 1) = \frac{1}{4}.$$

По индукции $\mathbf{P}(\eta \in (k/2^n, (k+m)/2^n)) = m/2^n$ [последнее равенство написано в силу аддитивности], так как

$$\mathbf{P}\left(\eta \in \left(\frac{k}{2^n}, \frac{k+m}{2^n}\right)\right) = \mathbf{P}\{\xi_1 = a_1, \dots, \xi_n = a_n\} = \frac{1}{2^n} \left(\text{т. к. } \sum_{k>n} \frac{\xi_k}{2^k} \leq \frac{1}{2^n} \right).$$

Таким образом, для любых двоично-рациональных точек $a, b \in [0, 1]$ мы доказали, что $\mathbf{P}(\eta \in [a, b]) = b - a$.

Далее, $\forall a, b \in [0, 1] \exists \{a_n\} = \left\{ \frac{k_n}{2^n} \right\} \uparrow a$ и $\exists \{b_n\} = \left\{ \frac{k_n}{2^n} \right\} \downarrow b$. Тогда по лемме непрерывности получаем

$$A_n = \{\eta \in [a_n, b_n]\} \supseteq A_{n+1} \rightarrow \mathbf{P}(\eta \in [a, b]) = \lim_{n \rightarrow \infty} \mathbf{P}(\eta \in [a_n, b_n]) = \lim_{n \rightarrow \infty} \mathbf{P}(b_n - a_n) = b - a. \quad \square$$

Заметим, что $\eta = \sum_{k=1}^N \frac{\xi_k}{2^k}$ — это дискретное равномерное распределение, а $\eta = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k}$ — абсолютно непрерывное.

Лемма 2. Пусть $\{\xi_i\}$ — последовательность независимых симметричных бернуллиевских величин, заданных на одном вероятностном пространстве. Тогда на этом же пространстве можно задать счетный набор независимых случайных величин $\{\eta_i\}$, равномерно распределенных на $[0, 1]$.

ДОКАЗАТЕЛЬСТВО. Воспользуемся диагональным методом и построим счетный набор подпоследовательностей бернуллиевских симметричных случайных величин, независимых в совокупности:

$$(\xi_1, \xi_3, \xi_6, \xi_{10}, \xi_{15} \dots)$$

$$(\xi_2, \xi_4, \xi_7, \xi_{11}, \xi_{16} \dots)$$

$$(\xi_5, \xi_8, \xi_{12}, \xi_{17} \dots)$$

...

Для каждой из построенных последовательностей зададим равномерно распределенную случайную величину

$$\eta_n = \sum_{k \geq 1} \frac{\xi_k^{(n)}}{2^k}.$$

Интуитивно понятно, что это независимый счетный набор. Но это нужно обосновать, поскольку мы имеем дело с функциями от бесконечных наборов случайных величин. Рассмотрим частичные суммы рассматриваемых рядов:

$$\eta_n^N = \sum_{k=1}^N \frac{\xi_k^{(n)}}{2^k}.$$

Лемма. Пусть векторы $\bar{\xi}_1 \in \mathbb{R}^{N_1}, \bar{\xi}_2 \in \mathbb{R}^{N_2}$ заданы на одном вероятностном пространстве и независимы, а $f: \mathbb{R}^{N_1} \rightarrow \mathbb{R}^{m_1}; g: \mathbb{R}^{N_2} \rightarrow \mathbb{R}^{m_2}$ — произвольные борелевские функции (измеримые относительно борелевской σ -алгебры), $\mathbb{R}^{m_1}, \mathbb{R}^{m_2}$ — конечномерные евклидовы пространства. Тогда новые векторы $f(\bar{\xi}_1), g(\bar{\xi}_2)$ также независимы.

ДОКАЗАТЕЛЬСТВО. Проверим факторизацию $\mathbf{P}(f(\bar{\xi}_1) \in A_1, g(\bar{\xi}_2) \in A_2) \forall A_1, A_2$. Так как функции f и g измеримы, то

$$\mathbf{P}(f(\bar{\xi}_1) \in A_1, g(\bar{\xi}_2) \in A_2) = \mathbf{P}(\bar{\xi}_1 \in f^{-1}(A_1), \bar{\xi}_2 \in g^{-1}(A_2)) =$$

[положим $B_1 = f^{-1}(A_1) \in \mathcal{B}^{N_1}; B_2 = g^{-1}(A_2) \in \mathcal{B}^{N_2}$. Так как $\bar{\xi}_1, \bar{\xi}_2$ независимы, то]

$$= \mathbf{P}(\bar{\xi}_1 \in B_1) \mathbf{P}(\bar{\xi}_2 \in B_2) = \mathbf{P}(f(\bar{\xi}_1) \in A_1) \mathbf{P}(g(\bar{\xi}_2) \in A_2).$$

Аналогичное утверждение имеет место для любого конечного набора векторов, т. е. борелевские преобразование независимых элементов независимы. \square

Заметим, что η_n^N есть линейная комбинация координат вектора — это борелевская функция. В силу леммы

$$\mathbf{P}(\eta_{n_1}^{N_1} < t_1, \eta_{n_2}^{N_2} < t_2) \stackrel{n_1 \neq n_2}{=} \mathbf{P}(\eta_{n_1}^{N_1} < t_1) \mathbf{P}(\eta_{n_2}^{N_2} < t_2).$$

Итак, $\forall t_1, t_2$ доказали факторизацию, то есть независимость $\eta_{n_1}^{N_1}$ и $\eta_{n_2}^{N_2}$ для $n_1 \neq n_2$. Но мы работаем с рядами. Положим $N_1 = N_2 = N$ без уменьшения общности и перейдем к пределу по N .

Обозначим $A_N = \{\eta_{n_1}^N < t_1, \eta_{n_2}^N < t_2\}$ при фиксированных t_1, t_2 . Тогда последовательность $\{A_N\}$ — это сужающийся поток. Тогда по лемме непрерывности имеем:

$$\begin{aligned} \mathbf{P}(\eta_{n_1} < t_1, \eta_{n_2} < t_2) &= \lim_{N \rightarrow \infty} \mathbf{P}(\eta_{n_1}^N < t_1, \eta_{n_2}^N < t_2) \\ &= \lim_{N \rightarrow \infty} \mathbf{P}(\eta_{n_1}^N < t_1) \lim_{N \rightarrow \infty} \mathbf{P}(\eta_{n_2}^N < t_2) = \mathbf{P}(\eta_{n_1} < t_1) \mathbf{P}(\eta_{n_2} < t_2) \end{aligned}$$

(так как $\forall i$ последовательности событий $\{\eta_{n_i}^N < t_i\}, N \geq 1\}$ тоже образует сужающиеся потоки). Мы построили бесконечно много независимых случайных величин, равномерно распределенных на $[0, 1]$. \square

Определение. Пусть F — произвольная функция распределения. Тогда $F^{-1}(t) = \inf\{x : F(x) \geq t\} = \sup\{x : F(x) < t\}$ — обобщенная обратная функция (квантильное преобразование функции распределения F).

Упражнение. Показать, что если F — строго монотонна, то F^{-1} — это классическая обратная функция.

Заметим, что для некоторых функций распределения (например, гауссовских) квантильное преобразование нужно доопределять в 1 или 0 (так как иначе имеем пустые неравенства в соответствующих определениях). Например, $F^{-1}(1) = \infty$ (для определения через \inf) и $F^{-1}(0) = -\infty$ (для определения через \sup).

Таким образом, типичный вид обобщенной обратной функции — это кусочно-непрерывная функция, состоящая из не более чем счетного набора нестрого монотонных участков, где участки постоянства соответствуют разрывам исходной функции распределения, и наоборот — участки постоянства квантильного преобразования соответствует скачкам исходной функции распределения.

Упражнение. Доказать, что квантильное преобразование — непрерывная слева функция.

Упражнение. Доказать, что для любой функции распределения F справедливо тождество $\Lambda(\{t : F^{-1}(t) < x\}) = F(x)$, где $\Lambda(\cdot)$ — мера Лебега.

Отметим, что доказательство последнего упражнения очевидно, если F — строго монотонная и непрерывная функция. В этом случае F^{-1} — классическая обратная функция. Поэтому $F^{-1}(t) < x \leftrightarrow t < F(x)$. Поэтому

$$\Lambda(\{t : F^{-1}(t) < x\}) = \Lambda(\{t : t < F(x)\}) = F(x).$$

□

Переходим к последнему этапу построения бесконечной последовательности независимых случайных величин с любыми наперед заданными распределениями. Возьмем произвольный счетный набор функций распределения F_1, F_2, \dots . Построим соответствующие квантильные преобразования $F_1^{-1}, F_2^{-1}, \dots$. Тогда в силу упражнения 2 последовательность $\xi_k = F_k^{-1}(\eta_k)$ и есть искомая, где $\{\eta_k\}$ — ранее построенная последовательность независимых равномерно распределенных случайных величин.

В самом деле, $\mathbf{P}(\xi_k < x) = \mathbf{P}(F_k^{-1}(\eta_k) < x) = \Lambda(\{t : F_k^{-1}(t) < x\}) = F_k(x)$. Таким образом, мы получили счетный набор независимых случайных величин с произвольными маргинальными распределениями.

Глава 3. МОМЕНТЫ СЛУЧАЙНЫХ ВЕЛИЧИН (РАСПРЕДЕЛЕНИЙ)

Понятие математического ожидания.

Определение. Пусть ξ — дискретная случайная величина с конечным набором атомов $\{a_k; k \leq n\}$ и соответствующими массами $\{p_k\}$. Математическим ожиданием (или средним значением) величины ξ называется её первый момент:

$$\mathbf{E}\xi = \sum_{k \leq n} a_k p_k.$$

Заметим, что для вырожденных случайных величин $\xi \equiv c$, очевидно, $\mathbf{E}\xi = c$. Если вещественную прямую представить себе как невесомый бесконечный стержень, на котором

помещены атомы с данными массами, то первый момент — это координата центра масс.

Свойства оператора математического ожидания в классе дискретных величин.

1. *Линейность* $\mathbf{E}(c_1\xi_1 + c_2\xi_2) = c_1\mathbf{E}\xi_1 + c_2\mathbf{E}\xi_2$.

Однородности $\mathbf{E}c\xi = c\mathbf{E}\xi$ очевидна. Докажем аддитивность. Пусть ξ_1 имеет атомы $\{a_k\}_{k=1}^n$ с весами $\{p_k\}$, а величина ξ_2 — атомы $\{b_j\}_{j=1}^m$ с массами $\{q_j\}$. Величина $\xi_1 + \xi_2$ имеет атомы $\{a_k + b_j : k \leq n, j \leq m\}$ с массами $\tilde{p}_{k,j} = \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j)$. Тогда

$$\begin{aligned} \mathbf{E}(\xi_1 + \xi_2) &= \sum_{k \leq n, j \leq m} (a_k + b_j) \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) = \sum_{k \leq n, j \leq m} a_k \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) + \\ &+ \sum_{k \leq n, j \leq m} b_j \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) = \sum_{k=1}^n a_k \sum_{j=1}^m \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) + \\ &+ \sum_{j=1}^m b_j \sum_{k=1}^n \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) = \sum_{k=1}^n a_k p_k + \sum_{j=1}^m b_j q_j = \mathbf{E}\xi_1 + \mathbf{E}\xi_2. \end{aligned}$$

При переходе к предпоследнему равенству использована формула полной вероятности, так как наборы событий $H_{1k} = \{\xi_1 = a_k\}$, $k \leq n$, а также $H_{2j} = \{\xi_2 = b_j\}$, $j \leq m$, образуют полные группы. \square

2. *Если $\xi_1 \leq \xi_2$ почти наверное, то $\mathbf{E}\xi_1 \leq \mathbf{E}\xi_2$.*

Очевидно, это свойство можно сформулировать эквивалентным образом: если $\eta \geq 0$ почти наверное, то $\mathbf{E}\eta \geq 0$ (если положим $\eta = \xi_2 - \xi_1$, то из аддитивности получим требуемое).

2'. *Из свойства монотонности следует неравенство треугольника для математического ожидания: $|\mathbf{E}\xi| \leq \mathbf{E}|\xi|$, так как $\pm\xi \leq |\xi|$, откуда $\mathbf{E}(\pm\xi) = \pm\mathbf{E}\xi \leq \mathbf{E}|\xi|$.*

3. *Мультипликативность.*

Если ξ_1, ξ_2 — независимые дискретные случайные величины с конечным числом атомов, то $\mathbf{E}\xi_1\xi_2 = \mathbf{E}\xi_1\mathbf{E}\xi_2$.

ДОКАЗАТЕЛЬСТВО. Рассмотрим новую дискретную случайную величину

$$\xi_1\xi_2 \in \{a_k b_j : k \leq n, j \leq m\}$$

с атомами $p_{k,j} = \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j)$. Отметим, что в определении математического ожидания мы не подчеркивали попарное различие атомов: если имеются атомы неединичной кратности, то при желании мы можем рассматривать один атом, сложив массы соответствующих совпадающих с ним атомов. Среднее от такого преобразования, как легко видеть, не претерпит никаких изменений.

Считаем математическое ожидание (во втором равенстве использована независи-

мость):

$$\begin{aligned} \mathbf{E}\xi_1\xi_2 &= \sum_{k \leq n, j \leq m} a_k b_j \mathbf{P}(\xi_1 = a_k, \xi_2 = b_j) = \sum_{k \leq n, j \leq m} a_k b_j \mathbf{P}(\xi_1 = a_k) \mathbf{P}(\xi_2 = b_j) = \\ &= \sum_{k=1}^n a_k p_k \cdot \sum_{j=1}^m b_j q_j = \mathbf{E}\xi_1 \mathbf{E}\xi_2. \quad \square \end{aligned}$$

З'. Если ξ_1, \dots, ξ_n — независимые случайные дискретные величины с конечным числом атомов, то

$$\mathbf{E} \prod_{j=1}^n \xi_j = \prod_{j=1}^n \mathbf{E}\xi_j.$$

Доказывается индукцией по n (база установлена в пункте 3).

Понятие математического ожидания можно ввести и для счетного числа атомов. Необходимо только, чтобы соответствующий ряд $\mathbf{E}\xi = \sum_{k=1}^{\infty} a_k p_k$ сходилась абсолютно.

Пример. Пусть дискретная случайная величина ξ имеет атомы $a_i = i$ с весами $p_i = \frac{1}{i(i+1)}$, $i = 1, 2, \dots$. Отметим, что $\{p_i\}$ — это действительно распределение, так как

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{i} - \frac{1}{i+1} \right) = 1.$$

Величина ξ не имеет математического ожидания, поскольку ряд

$$\sum_{i=1}^{\infty} |a_i| p_i = \sum_{i=1}^{\infty} \frac{1}{i+1}$$

расходится. Под существованием математического ожидания мы в дальнейшем будем понимать конечность величины $\mathbf{E}|\xi|$.

Математическое ожидание дискретной величины можно представить и как интеграл Лебега. В самом деле, введем множества $A_i = \{\omega : \xi = a_i\}$. Пусть I_A — это индикатор множества A (это значит, что $I_A(\omega) = 1$, если $\omega \in A$, и 0 в противном случае). Тогда ξ — ступенчатая функция:

$$\xi(\omega) = \sum a_i I_{A_i}(\omega).$$

Поэтому математическое ожидание такой случайной величины — это интеграл Лебега для ступенчатой величины (или интегральная сумма):

$$\mathbf{E}\xi = \sum_{i=1}^n a_i \mathbf{P}(A_i) = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega).$$

Перейдем к определению математического ожидания для произвольной случайной величины. Но сначала рассмотрим случай ограниченной случайной величины: $-M \leq$

$\xi \leq M$ почти наверное для некоторого $M > 0$. Измеримую ограниченную (с вероятностью 1) функцию ξ приблизим ступенчатыми функциями. Именно, покажем, что существует последовательность дискретных случайных величин $\{\xi_n^{(d)}\}$ такая, что $\xi_n^{(d)} \rightarrow \xi$ равномерно по $\omega \in \Omega$. Для этого зададим атомы $a_i = i/n$ и множества

$$A_i = \left\{ \omega : \frac{i}{n} \leq \xi(\omega) < \frac{i+1}{n} \right\}.$$

Случайную величину $\xi_n^{(d)}$ определим по формуле

$$\xi_n^{(d)}(\omega) = \sum_i \frac{i}{n} I_{A_i}(\omega).$$

В написанной сумме всегда лишь одно слагаемое отлично от нуля, учитывая попарную несовместность A_i . Тогда легко видеть, что $|\xi - \xi_n^{(d)}| \leq 1/n \forall \omega$. Поэтому при $n \rightarrow \infty$

$$\sup_{\omega \in \Omega} |\xi - \xi_n^{(d)}| \leq \frac{1}{n} \rightarrow 0,$$

а это и означает равномерную по ω сходимость.

После сказанного естественно считать, что $\mathbf{E}\xi = \lim_{n \rightarrow \infty} \mathbf{E}\xi_n$. Но для этого надо доказать, что написанный справа предел существует. Обозначим $b_n = \mathbf{E}\xi_n^{(d)}$. Согласно критерию Коши $\lim_{n \rightarrow \infty} b_n$ существует тогда и только тогда, когда $\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} |b_n - b_m| = 0$. Имеем

$$\begin{aligned} |\mathbf{E}\xi_n^{(d)} - \mathbf{E}\xi_m^{(d)}| &= |\mathbf{E}(\xi_n^{(d)} - \xi_m^{(d)})| \leq \mathbf{E}|\xi_n^{(d)} - \xi_m^{(d)}| = \mathbf{E}|\xi_n^{(d)} - \xi_m^{(d)} - \xi + \xi| \leq \\ &\leq \mathbf{E}|\xi_n^{(d)} - \xi| + \mathbf{E}|\xi_m^{(d)} - \xi| \leq \frac{1}{n} + \frac{1}{m} \rightarrow 0. \end{aligned}$$

Первое равенство в этой цепочке написано на основе линейности математического ожидания, следующее за ним неравенство — это неравенство треугольника для математического ожидания, наконец, второе неравенство вытекает из обычного неравенства треугольника для модуля и монотонности оператора \mathbf{E} . Итак, существование предела для любой равномерно сходящейся к исходной ξ последовательности дискретных случайных величин доказано (осталось доказать еще корректность определения — см. упражнение ниже).

Пусть ξ — произвольная случайная величина (не обязательно ограниченная). С помощью операции *срезки* построим ограниченную случайную величину с двумя параметрами:

$$\xi_{M,N} = \xi I(-N \leq \xi \leq M),$$

где для удобства мы несколько изменили определение индикаторной функции. Математическим ожиданием случайной величины ξ назовем значение

$$\mathbf{E}\xi = \lim_{\substack{M \rightarrow \infty \\ N \rightarrow \infty}} \mathbf{E}\xi_{M,N},$$

если двойной предел в правой части существует.

Упражнение. Проверить корректность задания математического ожидания, т. е. что для любой последовательности дискретных случайных величин, равномерно сходящихся к предельной случайной величине, предел соответствующих средних останется неизменным.

Запишем теперь математическое ожидание в виде интеграла. Начнем опять со случайной величины ξ такой, что $-M \leq \xi \leq N$ почти наверное. Возьмем дискретные случайные величины, которые мы построили ранее. Так как

$$\mathbf{E}I\left(\frac{i}{n} \leq \xi < \frac{i+1}{n}\right) = \mathbf{P}\left(\frac{i}{n} \leq \xi < \frac{i+1}{n}\right) = F_\xi\left(\frac{i+1}{n}\right) - F_\xi\left(\frac{i}{n}\right),$$

то

$$\mathbf{E}\xi_n^{(n)} = \sum_i \frac{i}{n} \mathbf{P}\left(\frac{i}{n} \leq \xi < \frac{i+1}{n}\right) = \sum_i \frac{i}{n} \Delta F_\xi\left(\frac{i}{n}\right).$$

У нас получилась интегральная сумма Римана–Стилтьеса, которая при $n \rightarrow \infty$ переходит в интеграл Римана–Стилтьеса:

$$\mathbf{E}\xi = \int_{-M}^N x dF_\xi(x).$$

Интеграл Римана–Стилтьеса определяется для непрерывных интегрантов (в данном случае – линейная функция) и интегрирующих функций F_ξ ограниченной вариации, то есть для которых

$$\sup_{\{t_i\}} \sum |F_\xi(t_{i+1}) - F_\xi(t_i)| < \infty,$$

где супремум берется по всевозможным конечным разбиениям $\{t_i\}$ отрезка $[-M, N]$. В нашем случае F_ξ – это функция распределения. Ее полная вариация, очевидно, равна 1.

Если F_ξ – гладкая функция, то по формуле конечных приращений найдется точка $\tilde{x}_i \in [x_i, x_{i+1})$ такая, что $\Delta F_\xi(x_i) = F'_\xi(\tilde{x}_i) \Delta x_i$. Тогда (последнее приближенное равенство написано в предположении непрерывности плотности p_ξ)

$$\sum x_i \Delta F_\xi(x_i) = \sum x_i F'_\xi(\tilde{x}_i) \Delta x_i = \sum x_i p_\xi(\tilde{x}_i) \Delta x_i \sim \sum x_i p_\xi(x_i) \Delta x_i.$$

Последняя сумма стремится к интегралу Римана

$$\int_{-M}^N x p(x) dx,$$

где $p \geq 0$ и $\int p(x) dx = 1$, а $F_\xi(t) = \int_{-\infty}^t p(x) dx$.

В случае же произвольной случайной величины ξ , согласно общему определению, мы должны положить

$$\mathbf{E}\xi = \lim_{\substack{M \rightarrow \infty \\ N \rightarrow \infty}} \int_{-M}^N x dF_\xi(x) = \int x dF_\xi(x),$$

если двойной предел существует.

Математическое ожидания преобразований случайных величин.

Пусть $\eta = f(\xi)$, где f — непрерывная (почти всюду на прямой) функция, а ξ — случайная величина. Мы хотим посчитать $\mathbf{E}f(\xi)$. Нам известно, что

$$\mathbf{E}f(\xi) = \mathbf{E}\eta = \int x dF_\eta(x).$$

Наша задача состоит в том, чтобы выразить $\mathbf{E}f(\xi)$ в терминах F_ξ . Пусть для простоты рассуждений f непрерывна всюду, а $|\xi| \leq N$ с вероятностью 1. Снова используем построенные ранее дискретные случайные величины $\xi_n^{(1)}$, с помощью которых определим дискретные случайные величины $f(\xi_n^{(1)})$. Тогда

$$\mathbf{E}f(\xi_n^{(1)}) = \sum_k f\left(\frac{k}{n}\right) \mathbf{P}\left(\frac{k}{n} \leq \xi < \frac{k+1}{n}\right) = \sum_k f(x_k) \Delta F_\xi(x_k),$$

где точки $x_k \equiv k/n$ образуют измельчающееся разбиение отрезка $[-N, N]$ — области значений случайной величины ξ . Последняя сумма стремится к $\int_{-N}^N f(x) dF_\xi(x)$. Используя срезы исходных случайных величин на уровне N и переходя к пределу при $N \rightarrow \infty$, можно легко распространить этот результат на случай произвольных случайных величин. В результате получим *формулу замены переменных* в интеграле Римана—Стилтьеса:

$$\mathbf{E}f(\xi) = \int x dF_{f(\xi)}(x) = \int f(x) dF_\xi(x).$$

В частности, если ξ имеет абсолютно непрерывное распределение с плотностью p_ξ , то

$$\mathbf{E}f(\xi) = \int f(x) p_\xi(x) dx.$$

Приведенная конструкция легко переносится на случай многомерных ξ . Ограничимся рассмотрением двумерных $\xi = (\xi_1, \xi_2)$. Тогда

$$\begin{aligned} \mathbf{E}f(\xi_{1,n}^{(1)}, \xi_{2,n}^{(1)}) &= \sum_k \sum_l f\left(\frac{k}{n}, \frac{l}{n}\right) \mathbf{P}\left(\frac{k}{n} \leq \xi_1 < \frac{k+1}{n}, \frac{l}{n} \leq \xi_2 < \frac{l+1}{n}\right) \sim \\ &\sim \int \int f(x, y) dF_{\xi_1, \xi_2}(x, y) = \mathbf{E}f(\xi_1, \xi_2), \end{aligned}$$

где $dF_{\xi_1, \xi_2}(x, y) = F_{\xi_1, \xi_2}(x + dx, y + dy) - F_{\xi_1, \xi_2}(x + dx, y) - F_{\xi_1, \xi_2}(x, y + dy) + F_{\xi_1, \xi_2}(x, y)$ — так называемая двойная разность функции $F_{\xi_1, \xi_2}(x, y)$ на инфинитезимальном прямоугольнике $[x, x + dx] \times [y, y + dy]$, что совпадает с вероятностью попадания случайного вектора (ξ_1, ξ_2) в этот прямоугольник. Легко видеть, что для независимых компонент рассматриваемого случайного вектора вышеприведенная формула принимает вид

$$\mathbf{E}f(\xi_1, \xi_2) = \int \int f(x, y) dF_{\xi_1}(x) dF_{\xi_2}(y),$$

где кратный интеграл удобно понимать как повторный, при этом вычисленный в любой последовательности переменных интегрирования x и y . Это, на самом деле, и есть содержание классической теоремы Фубини. Тем самым, мы сформулировали правило вычисления математического ожидания скалярных преобразований от конечного набора независимых случайных величин.

Также важно отметить, что для любой ограниченной измеримой функции f имеет место более общая конструкция математического ожидания случайной величины $f(\xi)$ в виде интеграла Лебега, который, как известно, определяется несколько иной последовательностью ступенчатых функций $f_n^{(1)}(x) = \sum x_k I(x_k \leq f(x) < x_{k+1})$:

$$\mathbf{E}f^{(1)}(\xi) = \sum_k x_k \mathbf{P}(x_k \leq f(\xi) < x_{k+1}) \sim \int f(x) \mathcal{P}_\xi(dx),$$

если только упорядоченное по возрастанию семейство действительных чисел $\{x_k\}$ образует измельчающееся разбиение области значений функции $f(x)$ (ограниченной для простоты рассуждений), а не ее аргумента, т.е. выполнено условие $\max_k(x_{k+1} - x_k) \rightarrow 0$.

Упражнение. Доказать, что если непрерывная функция $f(x)$ интегрируема в смысле Римана—Стилтьеса (по некоторой функции распределения), то она интегрируема в смысле Лебега.

Предельным переходом с использованием вышеупомянутых ступенчатых функций $f_n^{(1)}(x)$ можно легко проверить выполнение всех уже отмеченных выше свойств для случайных величин $f(\xi)$, ограниченных с вероятностью 1. Например, проверим свойство мультипликативности. В самом деле, мы уже доказали равенство

$$\mathbf{E}f^{(1)}(\xi_1)g^{(1)}(\xi_2) = \mathbf{E}f^{(1)}(\xi_1)\mathbf{E}g^{(1)}(\xi_2)$$

для любых независимых случайных величин ξ_1 и ξ_2 (напомним, что любые борелевские преобразования независимых случайных величин снова независимы). Тогда, с одной стороны, при измельчении разбиения $\{x_k\}$ имеем

$$\mathbf{E}f_n^{(1)}(\xi_1) \rightarrow \mathbf{E}f(\xi_1), \quad \mathbf{E}g_n^{(1)}(\xi_2) \rightarrow \mathbf{E}g(\xi_2),$$

а с другой,

$$\begin{aligned} |\mathbf{E}f_n^{(1)}(\xi_1)g_n^{(1)}(\xi_2) - \mathbf{E}f(\xi_1)\mathbf{E}g(\xi_2)| &\leq \sup |g|\mathbf{E}|f_n^{(1)}(\xi_1) - f(\xi_1)| + \sup |f|\mathbf{E}|g_n^{(1)}(\xi_2) - g(\xi_2)| \\ &\leq (\sup |g| + \sup |f|) \max_k(x_{k+1} - x_k) \rightarrow 0. \end{aligned} \quad \square$$

Главное преимущество лебеговой конструкции математического ожидания перед конструкцией Римана—Стилтьеса состоит в том, что она не предполагает как непрерывности функций $f(x)$, так и одномерности случайных величин ξ .

Упражнение. Пусть ξ_1 и ξ_2 — независимые случайные величины. Доказать, что

$$F_{\xi_1+\xi_2}(x) = \int_R F_{\xi_1}(x-t)dF_{\xi_2}(x).$$

Упражнение. Пусть ξ_1 и ξ_2 — независимые случайные величины, при этом ξ_1 имеет абсолютно непрерывное распределение. Доказать, что сумма этих случайных величин также имеет абсолютно непрерывное распределение с плотностью

$$p_{\xi_1+\xi_2}(x) = \int_R p_{\xi_1}(x-t)dF_{\xi_2}(x).$$

Моменты распределений.

1. Если $f(x) = x^k$, то $\mathbf{E}\xi^k$ называют k -ым моментом случайной величины ξ .
2. Если $f(x) = |x|^k$, то $\mathbf{E}|\xi|^k$ называют k -ым абсолютным моментом случайной величины ξ .
3. Если $f(x) = (x - \mathbf{E}\xi)^k$, то говорят о k -ом центральном моменте $\mathbf{E}(\xi - \mathbf{E}\xi)^k$ случайной величины ξ .
4. Если $f(x) = |x - \mathbf{E}\xi|^k$, то говорят о k -ом абсолютном центральном моменте случайной величины ξ , т. е. о величине $\mathbf{E}|\xi - \mathbf{E}\xi|^k$.

Определение. Дисперсией случайной величины ξ называется её второй центральный момент (среднеквадратичный разброс случайной величины в окрестности её математического ожидания):

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2.$$

Свойства дисперсии.

1. $\mathbf{D}\xi \geq 0$; $\mathbf{D}\xi = 0 \leftrightarrow \xi = \text{const}$ с вероятностью 1.
2. $\mathbf{D}(c + \xi) = \mathbf{D}\xi$, где $c = \text{const}$.
3. Однородность второго порядка: $\mathbf{D}(c\xi) = c^2\mathbf{D}\xi$.
4. Аддитивность. Если случайные величины ξ_1, \dots, ξ_n независимы и имеют конечные дисперсии, то

$$\mathbf{D} \sum_{i=1}^n \xi_i = \sum_{i=1}^n \mathbf{D}\xi_i.$$

Оформим в виде леммы более общее утверждение.

Лемма. Если ξ_1, \dots, ξ_n — произвольные случайные величины, имеющие конечные дисперсии, то

$$\mathbf{D} \sum_{i=1}^n \xi_i = \sum_{i=1}^n \mathbf{D}\xi_i + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}(\xi_i, \xi_j),$$

где $\mathbf{Cov}(\xi_i, \xi_j) = \mathbf{E}(\xi_i - \mathbf{E}\xi_i)(\xi_j - \mathbf{E}\xi_j)$ — ковариация между величинами ξ_i и ξ_j (смешанный второй центральный момент).

ДОКАЗАТЕЛЬСТВО ЛЕММЫ. Обозначим $\xi'_i = \xi_i - \mathbf{E}\xi_i$. По определению дисперсии и свойствам математического ожидания имеем:

$$\begin{aligned} \mathbf{D} \sum_{i=1}^n \xi_i &= \mathbf{E} \left(\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbf{E}\xi_i \right)^2 = \mathbf{E} \left(\sum_{i=1}^n \xi'_i \right)^2 = \mathbf{E} \left(\sum_{i=1}^n (\xi'_i)^2 + 2 \sum_{1 \leq i < j \leq n} \xi'_i \xi'_j \right) = \\ &= \sum_{i=1}^n \mathbf{E}(\xi'_i)^2 + 2 \sum_{1 \leq i < j \leq n} \mathbf{E}(\xi'_i \xi'_j) = \sum_{i=1}^n \mathbf{D}\xi_i + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}(\xi_i, \xi_j). \quad \square \end{aligned}$$

С помощью этой леммы докажем свойство 4, сформулированное для независимых величин. Так как сдвиг на константу независимости не меняет, то величины $\xi_1 - \mathbf{E}\xi_1$ и $\xi_2 - \mathbf{E}\xi_2$ тоже независимы. Значит, при $i \neq j$ имеет место равенство

$$\mathbf{Cov}(\xi_i, \xi_j) = \mathbf{E}(\xi_i - \mathbf{E}\xi_i)(\xi_j - \mathbf{E}\xi_j) = \mathbf{E}(\xi_i - \mathbf{E}\xi_i) \cdot \mathbf{E}(\xi_j - \mathbf{E}\xi_j) = (\mathbf{E}\xi_i - \mathbf{E}\xi_i) \cdot (\mathbf{E}\xi_j - \mathbf{E}\xi_j) = 0,$$

откуда и следует требуемое. Тем самым, свойство 4 — это следствие мультипликативности математического ожидания.

Отметим полезное во многих случаях представление для дисперсии:

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{E}\xi^2 - 2\mathbf{E}(\xi \cdot \mathbf{E}\xi) + \mathbf{E}(\mathbf{E}\xi)^2 = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2,$$

откуда следует, что существование дисперсии случайной величины эквивалентно существованию ее второго момента и что $\mathbf{E}\xi^2 \geq (\mathbf{E}\xi)^2$. Это соотношение — частный случай неравенства Коши—Буняковского, о котором речь пойдет ниже.

Примеры.

1. Бернуллиевская случайная величина $\xi = \begin{cases} 1, & \text{с вероятностью } p, \\ 0, & \text{с вероятностью } 1 - p. \end{cases}$

Очевидно, что $\mathbf{E}\xi^k = p \forall k \in \mathbb{N}$ и $\mathbf{D}\xi = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2 = p(1 - p)$.

2. Случайная величина ξ с биномиальным распределением принимает целые значения $\{0, \dots, n\}$ с вероятностями $p_k = \mathbf{P}(\xi = k) = C_n^k p^k (1 - p)^{n-k}$. Так как $\xi = \sum_{i=1}^n \xi_i$, где ξ_i — независимые бернуллиевские величины, то $\mathbf{E}\xi = np$ и $\mathbf{D}\xi = np(1 - p)$.

Упражнение. Посчитать математическое ожидания и дисперсию биномиального распределения по определению.

3. Пуассоновская случайная величина ξ принимает целые неотрицательные значения с вероятностями $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$. Вычислим ее математическое ожидание:

$$\mathbf{E}\xi = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

Вычислим дисперсию. Для этого посчитаем второй момент:

$$\begin{aligned} \mathbf{E}\xi^2 &= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{k=1}^{\infty} (k+1-1) \frac{\lambda^k}{(k-1)!} = \\ &= e^{-\lambda} \sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} + \lambda = \lambda + e^{-\lambda} \sum_{k=2}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} = \lambda + \lambda^2. \end{aligned}$$

Итак, $\mathbf{D}\xi = \lambda + \lambda^2 - \lambda^2 = \lambda$.

Упражнение. Доказать, что у пуассоновского распределения третий центральный момент тоже равен λ .

Упражнение. Доказать, что у пуассоновского распределения любой центральный момент нечетного порядка неотрицателен.

4. *Равномерное распределение* на отрезке $[a, b]$ с плотностью

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{иначе.} \end{cases}$$

Математическое ожидание:

$$\mathbf{E}\xi = \int_a^b x \frac{1}{b-a} dx = \frac{b+a}{2}.$$

Упражнение. Вычислить дисперсию равномерного распределения.

5. *Распределение Гаусса* с плотностью $p_{\xi}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2}$, $\sigma > 0$, $\alpha \in \mathbb{R}$.

Математическое ожидание:

$$\mathbf{E}\xi = \int_{\mathbb{R}} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2} dx = \int_{\mathbb{R}} (x-\alpha) \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2} dx + \alpha \cdot 1 = \alpha.$$

Заметим, что у распределения Гаусса существуют все моменты.

Упражнение. Доказать, что $\mathbf{D}\xi = \sigma^2$ и $\mathbf{E}(\xi - \alpha)^{2n} = (2n - 1)!!\sigma^{2n}$ для любого натурального n .

Величину $\sigma = \sqrt{\mathbf{D}\xi}$ называют *стандартным отклонением*.

6. *Распределение Коши* с плотностью $p_{\xi}(x) = \frac{1}{\pi(1+x^2)}$.

У этого распределения математического ожидания не существует, так как

$$\int_{\mathbb{R}} |x| p_{\xi}(x) dx = \infty.$$

7. *Показательное распределение* с плотностью $p_{\xi}(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0, & \text{иначе.} \end{cases}$

Математическое ожидание:

$$\mathbf{E}\xi = \int_0^{\infty} \alpha x e^{-\alpha x} dx = \frac{1}{\alpha}.$$

Упражнение. Для любого натурального n вычислить n -й центральный момент показательного распределения.

Упражнение. Пусть $\{\xi_i\}$ – последовательность одинаково распределенных неза-

всисмых случайных величин. Доказать, что условие $\mathbf{E}|\xi_1|^r < \infty$ для некоторого $r > 0$ необходимо и достаточно для того, чтобы при $n \rightarrow \infty$

$$n^{-1/r} \max_{i \leq n} |\xi_i| \rightarrow 0$$

с вероятностью 1.

Моментные неравенства.

Сначала приведем одно неравенство, которое непосредственно вытекает из определения дисперсии. Так как $0 \leq \mathbf{D}|\xi| = \mathbf{E}\xi^2 - (\mathbf{E}|\xi|)^2$, то $|\mathbf{E}\xi| \leq \mathbf{E}|\xi| \leq \sqrt{\mathbf{E}\xi^2}$ (поэтому из конечности второго момента следует существование математического ожидания). Рассмотрим некоторые частные случаи приведенного неравенства, иногда называемого неравенством Шварца или Коши–Буняковского (частный случай). Например, для дискретной случайной величины ξ , принимающей значения $\{a_i\}_{i=1}^n$ с вероятностями $\{p_i\}_{i=1}^n$, $\sum p_i = 1$, неравенство Шварца принимает вид

$$\left| \sum_{i=1}^n a_i p_i \right| \leq \sqrt{\sum_{i=1}^n a_i^2 p_i}.$$

В частности, если $p_i = 1/n$, $i = 1, \dots, n$, то получим известное арифметическое неравенство

$$\left| \frac{1}{n} \sum_{i=1}^n a_i \right| \leq \sqrt{\frac{1}{n} \sum_{i=1}^n a_i^2}.$$

Аналогично для любой абсолютно непрерывно распределенной случайной величины ξ с плотностью p_ξ имеет место неравенство

$$\left| \int_{\mathbb{R}} t p_\xi(t) dt \right| \leq \sqrt{\int_{\mathbb{R}} t^2 p_\xi(t) dt}.$$

Теперь сформулируем и докажем классическое *неравенство Коши–Буняковского*.

Теорема Пусть случайные величины ξ и η имеют конечные вторые моменты. Тогда для второго смешанного момента справедливо неравенство

$$|\mathbf{E}\xi\eta| \leq \sqrt{\mathbf{E}\xi^2 \mathbf{E}\eta^2}.$$

Причем равенство достигается тогда и только тогда, когда $\eta = c\xi$, где $c = \text{const}$.

ДОКАЗАТЕЛЬСТВО. Рассмотрим квадратный трехчлен

$$P_2(x) = \mathbf{E}(x\xi + \eta)^2 = x^2 \mathbf{E}\xi^2 + 2x \mathbf{E}\xi\eta + \mathbf{E}\eta^2.$$

Очевидно, что $P_2(x) \geq 0$. Значит, дискриминант неположителен: $(\mathbf{E}\xi\eta)^2 - \mathbf{E}\xi^2 \cdot \mathbf{E}\eta^2 \leq 0$.

Разберем случай равенства. Легко проверить, что если $\eta = c\xi$ с некоторой константой c , то равенство имеет место. Обратное, из равенства нулю упомянутого выше дискриминанта следует существование такого x_0 , что $0 = P_2(x_0) = \mathbf{E}(x_0\xi + \eta)^2$. Тогда для почти всех ω верно равенство $\eta(\omega) = -x_0\xi(\omega)$. \square

З а м е ч а н и е. Поскольку в приведенной теореме можно вместо случайных величин ξ и η рассмотреть $|\xi|$ и $|\eta|$, то на самом деле неравенство Коши—Буняковского можно записать в несколько более сильной форме:

$$|\mathbf{E}\xi\eta| \leq \mathbf{E}|\xi\eta| \leq \sqrt{\mathbf{E}\xi^2\mathbf{E}\eta^2}.$$

Теорема (неравенство Гёльдера). Пусть ξ и η — произвольные случайные величины, а p и q — сопряженные положительные числа: $1/p + 1/q = 1$. Тогда справедливо неравенство

$$\mathbf{E}|\xi\eta| \leq (\mathbf{E}|\xi|^p)^{1/p} \cdot (\mathbf{E}|\eta|^q)^{1/q},$$

если моменты в правой части существуют.

Заметим, что при $p = q = 2$ мы получаем неравенство Коши—Буняковского.

ДОКАЗАТЕЛЬСТВО. Прежде всего покажем, что для любых $x \geq 0$, $y \geq 0$ и любых сопряженных p и q верно неравенство

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q.$$

В самом деле, без ограничения общности можно считать, что $x > 0$ и $y > 0$. Тогда в силу вогнутости функции $\log x$ получаем

$$\log(xy) = \frac{1}{p}\log(x^p) + \frac{1}{q}\log(y^q) \leq \log\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right),$$

откуда в силу монотонности функции $\log x$ и следует нужное неравенство.

Далее, обозначим $\|\xi\|_p = (\mathbf{E}|\xi|^p)^{1/p}$, $\|\eta\|_q = (\mathbf{E}|\eta|^q)^{1/q}$. Положим $x = \frac{|\xi|}{\|\xi\|_p}$, $y = \frac{|\eta|}{\|\eta\|_q}$ и применим к этим величинам только что доказанное арифметическое неравенство. К обеим частям полученного неравенства применим оператор математического ожидания. Учитывая его монотонность, окончательно получаем

$$\frac{\mathbf{E}|\xi\eta|}{\|\xi\|_p\|\eta\|_q} \leq 1. \quad \square$$

В частности, если $\eta = 1$, то для любого $p \geq 1$ имеет место неравенство $\mathbf{E}|\xi| \leq (\mathbf{E}|\xi|^p)^{1/p}$. Применяя последнее неравенство для случайной величины $|\xi|^m$, $m > 0$, где $p = m/(m - \varepsilon)$, $m - \varepsilon > 0$, получим, что если $\mathbf{E}|\xi|^m < \infty$, то и $\mathbf{E}|\xi|^{m-\varepsilon} < \infty$.

Теорема (неравенство Йенсена). Пусть f — выпуклая (вниз) функция. Тогда справедливо неравенство

$$\mathbf{E}f(\xi) \geq f(\mathbf{E}\xi),$$

если выписанные моменты существуют.

ДОКАЗАТЕЛЬСТВО. Проведем через точку $(x_0, f(x_0))$ опорную прямую («выше» этой прямой расположен график функции; если функция гладкая, то опорная прямая — это

обычная касательная). Пусть A — угловой коэффициент проведенной прямой. Тогда в силу выпуклости функции f имеем $f(x) - f(x_0) \geq A(x - x_0)$.

Положим $x = \xi$ и $x_0 = \mathbf{E}\xi$. Тогда, применив оператор математического ожидания к обеим частям неравенства, получим:

$$\mathbf{E}f(\xi) - f(\mathbf{E}\xi) \geq A(\mathbf{E}\xi - \mathbf{E}\xi) = 0. \quad \square$$

Упражнение. Доказать, что если функция f вогнута (выпукла вверх), то

$$\mathbf{E}f(\xi) \leq f(\mathbf{E}\xi).$$

Следующее неравенство известно в функциональном анализе как неравенство треугольника для L_p -нормы.

Теорема (неравенство Минковского). Для любого $p \geq 1$ имеет место неравенство

$$\|\xi_1 + \xi_2\|_p \leq \|\xi_1\|_p + \|\xi_2\|_p,$$

если выписанные моменты существуют.

Доказательство. Ясно, что нам достаточно рассмотреть случай $p > 1$. В силу неравенства треугольника для модуля имеем

$$\|\xi_1 + \xi_2\|_p^p \leq \mathbf{E}|\xi_1 + \xi_2|^{p-1}|\xi_1| + \mathbf{E}|\xi_1 + \xi_2|^{p-1}|\xi_2|.$$

Теперь применяя к каждому из двух слагаемых правой части этого неравенства неравенство Гёльдера для сопряженных чисел $p/(p-1)$ и p , получаем

$$\|\xi_1 + \xi_2\|_p^p \leq \|\xi_1 + \xi_2\|_p^{p-1} (\|\xi_1\|_p + \|\xi_2\|_p). \quad \square$$

Теорема (неравенство Кокса–Кемпермана). Пусть функция $f \in \mathbf{C}^2(\mathbb{R})$ имеет выпуклую вторую производную и $f(0) \leq 0$. Пусть ξ_1 и ξ_2 — независимые центрированные случайные величины. Тогда

$$\mathbf{E}f(\xi_1 + \xi_2) \geq \mathbf{E}f(\xi_1) + \mathbf{E}f(\xi_2)$$

при условии, что момент в левой части этого неравенства существует.

Доказательство. Для любых вещественных x и y справедливо тождество (вариант формулы Тейлора)

$$f(x + y) = f(x) + f'(x)y + y^2 \int_0^1 (1 - \theta)f''(x + \theta y)d\theta.$$

Положим в этом тождестве $y = \xi_2$ и возьмем математическое ожидание от обеих частей. В результате получим

$$\mathbf{E}f(x + \xi_2) = f(x) + \mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)f''(x + \theta\xi_2)d\theta. \quad (1)$$

Отметим, что момент слева (а стало быть, и справа!) существует в силу теоремы Фубини. Теперь усредним обе части этого тождества по распределению случайной величины ξ_1

(по переменной x). Вновь использую теорему Фубини и вспоминая правило вычисления моментов функций от конечного набора независимых случайных величин, получаем

$$\mathbf{E}f(\xi_1 + \xi_2) = \mathbf{E}f(\xi_1) + \mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)f''(\xi_1 + \theta\xi_2)d\theta. \quad (2)$$

К моменту в правой части последнего равенства вновь применим теорему Фубини, а для внутреннего интеграла – неравенство Йенсена:

$$\begin{aligned} \mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)f''(\xi_1 + \theta\xi_2)d\theta &= \mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)\mathbf{E}_{\xi_1}f''(\xi_1 + \theta\xi_2)d\theta \\ &\geq \mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)f''(\theta\xi_2)d\theta, \end{aligned}$$

где символом \mathbf{E}_{ξ_1} мы обозначили усреднение по распределению ξ_1 при фиксированном ξ_2 (внутреннее интегрирование вдоль сечения). Для правой части этого неравенства воспользуемся формулой (1) при $x = 0$:

$$\mathbf{E}\xi_2^2 \int_0^1 (1 - \theta)f''(\theta\xi_2)d\theta = \mathbf{E}f(\xi_2) - f(0).$$

Остается подставить полученную оценку в (2). □

Следствие. Пусть $\xi_1, \xi_2, \dots, \xi_n$ – независимые центрированные случайные величины с конечными моментами порядка $p \geq 3$. Тогда

$$\mathbf{E} \left| \sum_{i=1}^n \xi_i \right|^p \geq \sum_{i=1}^n \mathbf{E}|\xi_i|^p$$

или в терминах ранее введенных обозначений

$$\left\| \sum_{i=1}^n \xi_i \right\|_p^p \geq \sum_{i=1}^n \|\xi_i\|_p^p.$$

Переход к пределу под знаком математического ожидания. Эквивалентное определение слабой сходимости.

Теорема (критерий слабой сходимости). Пусть $\{\xi_n\}$ – последовательность случайных величин. Тогда $\xi_n \Rightarrow \xi$ в том и только в том случае, когда для любой непрерывной ограниченной функции f ($f \in C_B$) имеет место сходимость

$$\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi).$$

ДОКАЗАТЕЛЬСТВО. Напомним, что $\xi_n \Rightarrow \xi \leftrightarrow F_{\xi_n}(t) \rightarrow F_{\xi}(t)$ для всех точек непрерывности функции F .

(\rightarrow) (Теорема Хелли). Пусть $f \in CB$. Приближим эту функцию ступенчатыми. Зафиксируем число m и определим измельчающееся разбиение $\{x_k\}$ отрезка $[-M, M]$ (например, можно положить $x_k = 2Mk/m$). Обозначим

$$f^{(n)}(x) = \sum_k f(x_k)I(x_k \leq x < x_{k+1}).$$

Предположим, что $|\xi_n| \leq M$. Тогда сумма в только что приведенном представлении $f^{(n)}(x)$ конечная. В силу равномерной непрерывности функции f на компакте $[-M, M]$ мы имеем

$$|f(x) - f^{(n)}(x)| \leq \delta_m \quad \forall x \in [-M, M], \quad \text{где } \delta_m = \sup_{t,s,|t-s| \leq 1/m} |f(t) - f(s)| \xrightarrow{m \rightarrow \infty} 0.$$

Сначала докажем наше утверждение для ступенчатых функций. Имеем

$$\begin{aligned} \mathbf{E}f^{(n)}(\xi_n) &= \sum_k f(x_k)\mathbf{P}(x_k \leq \xi_n < x_{k+1}) = \sum_k f(x_k)(F_{\xi_n}(x_{k+1}) - F_{\xi_n}(x_k)), \\ \mathbf{E}f^{(n)}(\xi) &= \sum_k f(x_k)\mathbf{P}(x_k \leq \xi < x_{k+1}) = \sum_k f(x_k)(F_{\xi}(x_{k+1}) - F_{\xi}(x_k)). \end{aligned}$$

Так как любая монотонная функция имеет не более чем счетное число точек разрыв (а функция распределения монотонна), то без ограничения общности можно считать, что x_k — точки непрерывности функции F . Тогда для всех k

$$F_{\xi_n}(x_k) \xrightarrow{n \rightarrow \infty} F_{\xi}(x_k),$$

откуда $\mathbf{E}f^{(n)}(\xi_n) \rightarrow \mathbf{E}f^{(n)}(\xi)$, и для ступенчатых функций теорема доказана.

Для произвольной функции $f \in CB$ имеем

$$\begin{aligned} \mathbf{E}f(\xi_n) &= \mathbf{E}f^{(n)}(\xi_n) + \mathbf{E}(f(\xi_n) - f^{(n)}(\xi_n)) = \mathbf{E}f^{(n)}(\xi_n) + \Delta_n, \\ \mathbf{E}f(\xi) &= \mathbf{E}f^{(n)}(\xi) + \tilde{\Delta}_n. \end{aligned}$$

В силу неравенства треугольника и монотонности среднего получаем

$$|\Delta_n| \leq \mathbf{E}|f(\xi_n) - f^{(n)}(\xi_n)| \leq \delta_m, \quad |\tilde{\Delta}_n| \leq \mathbf{E}|f(\xi) - f^{(n)}(\xi)| \leq \delta_m.$$

Так что выбором m указанные погрешности могут быть сколь угодно малыми. А так как для дискретных функций мы уже доказали наше утверждение, то и $\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi)$.

Избавимся от ограничения $|\xi_n| \leq M$. Заметим, что если $\{H_i\}$ — полная группа событий, то справедливо разложение единицы в виде

$$1 = \sum_i I(H_i),$$

тогда $\mathbf{E}f(\xi_n) = \mathbf{E}f(\xi_n)I(|\xi_n| \leq M) + \mathbf{E}f(\xi_n)I(|\xi_n| > M)$. Обозначим $\Delta_n = \mathbf{E}f(\xi_n)I(|\xi_n| > M)$. В силу ограниченности f для любого наперед заданного $\varepsilon > 0$ можно выбрать M так, что

$$\begin{aligned} |\Delta_n| &\leq \sup_t |f(t)|\mathbf{P}(|\xi_n| > M) = \sup_t |f(t)|(\mathbf{P}(\xi_n > M) + \mathbf{P}(\xi_n < -M)) \\ &\leq \sup_t |f(t)|(1 - F_{\xi_n}(M) + F_{\xi_n}(-M)) \leq 2\varepsilon \sup_t |f| + o(1) \end{aligned}$$

при $n \rightarrow \infty$, так как $1 - F_{\xi_n}(M) \rightarrow 1 - F_{\xi}(M) \leq \varepsilon$, $F_{\xi_n}(-M) \rightarrow F_{\xi}(-M) \leq \varepsilon$.

Поэтому мы можем записать

$$\mathbf{E}f(\xi_n) = \mathbf{E}f(\xi_n I(|\xi_n| \leq M)) + \Delta_n(M),$$

где выбором ε величина верхнего предела $\limsup_{n \rightarrow \infty} \Delta_n(M)$ может быть сделана сколь угодно малой.

Далее, обозначив $\xi_n^{(M)} = \xi_n I(|\xi_n| \leq M)$, имеем

$$\begin{aligned} \mathbf{E}f(\xi_n I(|\xi_n| \leq M)) &= \mathbf{E}f(\xi_n^{(M)}) - \mathbf{E}f(\xi_n I(|\xi_n| \leq M))I(|\xi_n| > M), \\ \mathbf{E}f(\xi_n I(|\xi_n| \leq M))I(|\xi_n| > M) &\leq \sup_t |f(t)| \mathbf{P}(|\xi_n| > M). \end{aligned}$$

Аналогичное неравенство справедливо и для $\mathbf{E}f(\xi)$.

Упражнение. Доказать, что если $\xi_n \Rightarrow \xi$, то и $\xi_n^{(M)} \Rightarrow \xi^{(M)}$.

Так как $\mathbf{E}f(\xi_n^{(M)}) \rightarrow \mathbf{E}f(\xi^{(M)})$, то из вышеприведенных оценок следует, что $\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi)$.

(\leftarrow) Пусть $\forall f \in CB$ выполнено $\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi)$. Заметим, что $F_{\xi_n}(x) = \mathbf{E}I(\xi_n < x)$. Введем две ограниченные и непрерывные функции g_{ε}^+ и g_{ε}^- — «спрямленные» индикаторы множества $\{t : t < x\}$ с помощью линейных функции на отрезках $[x, x + \varepsilon]$ и $[x - \varepsilon, x]$ соответственно. В силу монотонности математического ожидания

$$\begin{aligned} \mathbf{E}g_{\varepsilon}^-(\xi_n) &\leq \mathbf{E}I(\xi_n < x) \leq \mathbf{E}g_{\varepsilon}^+(\xi_n), \\ \mathbf{E}g_{\varepsilon}^-(\xi) &\leq \mathbf{E}I(\xi < x) \leq \mathbf{E}g_{\varepsilon}^+(\xi). \end{aligned}$$

Пусть x — точка непрерывности функции F_{ξ} . Тогда

$$\mathbf{E}(g_{\varepsilon}^+(\xi) - g_{\varepsilon}^-(\xi)) \leq \mathbf{P}(x - \varepsilon \leq \xi < x + \varepsilon) = F_{\xi}(x + \varepsilon) - F_{\xi}(x - \varepsilon) = \delta(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0,$$

откуда, в силу принципа «двух милиционеров» и произвольности ε , имеет место сходимость $F_{\xi_n}(x) \rightarrow F_{\xi}(x)$ в точках непрерывности F_{ξ} . \square

Следствие. Если $\xi_n \Rightarrow \xi$, то для любой непрерывной функции g выполнено $g(\xi_n) \Rightarrow g(\xi)$.

Теорема. Пусть $\xi_n \Rightarrow \xi$ и существуют математические ожидания всех рассматриваемых случайных величин. Кроме того, пусть выполнено условие равномерной интегрируемости:

$$\sup_n \mathbf{E}|\xi_n| I(|\xi_n| \geq M) \xrightarrow{M \rightarrow \infty} 0.$$

Тогда $\mathbf{E}\xi_n \rightarrow \mathbf{E}\xi$ при $n \rightarrow \infty$.

ДОКАЗАТЕЛЬСТВО. Рассмотрим сглаженную срезку, которая является непрерывной ограниченной функцией

$$f_M(x) = \begin{cases} -M, & x < -M, \\ x, & -M \leq x \leq M, \\ M, & x > M. \end{cases}$$

Тогда $\mathbf{E}f_M(\xi_n) \rightarrow \mathbf{E}f_M(\xi)$. В то же время,

$$|\mathbf{E}f_M(\xi_n) - \mathbf{E}\xi_n| \leq \mathbf{E}|f_M(\xi_n) - \xi_n| \leq \sup_n \mathbf{E}|\xi_n|I(|\xi_n| \geq M) \xrightarrow{M \rightarrow \infty} 0.$$

Аналогично, $|\mathbf{E}f_M(\xi) - \mathbf{E}\xi| \xrightarrow{M \rightarrow \infty} 0$. \square

Следствие. Пусть $\sup_n |\xi_n| \leq \eta$ и $\mathbf{E}\eta < \infty$. Тогда выполнено условие равномерной интегрируемости. Следовательно, имеет место и теорема Лебега:

$$\mathbf{E}\xi_n = \int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega) \rightarrow \mathbf{E}\xi.$$

ДОКАЗАТЕЛЬСТВО. $\mathbf{E}|\xi_n|I(|\xi_n| > M) \leq \mathbf{E}\eta I(\eta > N) \rightarrow 0$ как хвост сходящегося интеграла. \square

Упражнение. 1) Доказать, что условие равномерной интегрируемости является необходимым для сходимости математических ожиданий неотрицательных случайных величин.

2) Привести пример, когда не выполнено условие равномерной интегрируемости и сходимости математических ожиданий нет.

Математическое ожидание суммы случайного числа случайных величин.

Пусть на некотором вероятностном пространстве задана последовательность $\{\xi_n\}$ произвольно связанных случайных величин с одинаковыми математическими ожиданиями: $\mathbf{E}\xi_i = a \quad \forall i$ (это, конечно, не означает, что случайные величины распределены одинаково). На этом же вероятностном пространстве задана целочисленная случайная величина $\nu \in \mathbb{Z}_+$, не зависящая от последовательности $\{\xi_n\}$ (т. е. случайная величина ν не зависит от каждого конечного набора из $\{\xi_n\}$). Нас интересует математическое ожидание величины

$$S_\nu = \sum_{i=1}^{\nu} \xi_i,$$

где по определению полагаем $S_\nu = 0$ при $\nu = 0$.

Теорема (тождество Вальда). Пусть $\mathbf{E}\nu < \infty$. Тогда при сделанных предположениях относительно последовательности $\{\xi_i\}$ справедливо равенство

$$\mathbf{E}S_\nu = a\mathbf{E}\nu.$$

ДОКАЗАТЕЛЬСТВО. Предположим, что $\nu \leq N$ и $\xi_i \geq 0$. Ограничение $\xi_i \geq 0$ несущественно в силу разложения $\xi_i = \xi_i^+ - \xi_i^-$, где $\xi_i^+ = \max\{0, \xi_i\}$, $\xi_i^- = \max\{0, -\xi_i\}$. Прежде всего, получим представление для распределения

$$\mathcal{P}_{S_\nu}(A) = \mathbf{P}(S_\nu \in A).$$

Положим $H_i = \{\nu = i\}$, $i = 0, 1, \dots$. Тогда $\{H_i\}$ — полная группа событий (а так как $\nu \leq N$, то в этой группе не более чем $N + 1$ событий). По формуле полной вероятности

$$\mathbf{P}(\tilde{A}) = \sum_{i \geq 0} \mathbf{P}(\tilde{A} \cap H_i),$$

где $\tilde{A} = \{\omega : S_{\nu(\omega)}(\omega) \in A\}$. Тогда

$$\mathbf{P}(\tilde{A}) = \sum_{i \geq 0} \mathbf{P}(S_\nu \in A, \nu = i) = \sum_{i \geq 0} \mathbf{P}(S_i \in A, \nu = i) = \sum_{i \geq 0} \mathbf{P}(S_i \in A)p_i,$$

где $p_i = \mathbf{P}(\nu = i)$ и последнее равенство справедливо вследствие независимости соответствующих событий под знаком вероятности.

Имеем

$$\begin{aligned} \mathbf{E}S_\nu &= \int x \mathcal{P}_{S_\nu}(dx) = \int x \sum_{i \geq 0} \mathbf{P}(S_i \in dx)p_i = \sum_{i \geq 0} p_i \int x \mathbf{P}(S_i \in dx) = \\ &= \sum_{i \geq 0} p_i \int x \mathcal{P}_{S_i}(dx) = \sum_{i \geq 0} p_i \mathbf{E}S_i = \sum_{i \geq 0} p_i \cdot i \cdot a = a \sum_{i \geq 0} ip_i = a\mathbf{E}\nu. \end{aligned}$$

В третьем равенстве знаки интеграла и суммы можно поменять местами, так как сумма конечна.

Если же ν не является ограниченной, то введем срезку $\nu^{(N)} = \nu I(\nu \leq N)$. Тогда $\nu^{(N)} \leq N$, и значит, $\mathbf{E}S_{\nu^{(N)}} = a\mathbf{E}\nu^{(N)}$. Если N растет, то $S_{\nu^{(N)}}$ — монотонная последовательность, так как $\xi_i \geq 0$ и $\nu^{(N)} \xrightarrow{\text{п.н.}} \nu$ также монотонно. Поэтому $\mathbf{E}S_{\nu^{(N)}} \rightarrow \mathbf{E}S_{\nu}$ и $\mathbf{E}\nu^{(N)} \rightarrow \mathbf{E}\nu$. \square

Ветвящиеся процессы Гальтона—Ватсона.

Введем набор целочисленных неотрицательных случайных величин $\{\xi_k^{(i)}, k \geq 1\}$, $i = 0, 1, \dots$. Величина $\xi_k^{(i)}$ интерпретируется как число потомков, которое производит k -я частица из i -го поколения за единицу времени (т. е. в течение жизни одного «поколения»). В момент времени $i = 0$ имеется одна частица-«родитель», которая производит за время своей жизни $\xi_1^{(0)}$ потомков.

Будем предполагать, что набор $\{\xi_k^{(i)}, k \geq 1\}$ при фиксированном i не зависит от двумерного массива $\{\xi_k^{(j)}, j \leq i-1, k \geq 1\}$; каждый потомок развивается независимо от родителя, но стохастический алгоритм деления остается тем же самым, т. е. $\xi_k^{(i)}$ одинаково распределены $\forall i, k$. При этом частицы одного поколения (т. е. $\xi_k^{(i)}$ при фиксированном i) могут быть зависимыми.

Получившееся во времени «дерево» (или плоский граф) — это ветвящийся процесс деления Гальтона—Ватсона (заметим, что частица может никого не производить, а просто исчезнуть в следующем поколении, не оставив потомства).

Основной интересующей нас характеристикой является общая численность потомков N -го поколения, которая задается следующей рекуррентной формулой:

$$S_N = \sum_{k=1}^{S_{N-1}} \xi_k^{N-1}.$$

Нас интересует поведение S_N в среднем, когда N неограниченно возрастает. Отметим, что в приведенном представлении мы имеем дело с суммой случайного числа случайных величин, где по условию задачи $\nu = S_{N-1}$ не зависит от всех величин ξ_k^{N-1} в поколении $N-1$, которые одинаково распределены и независимы. Следовательно, по тождеству Вальда, если $a = \mathbf{E}\xi_1^{(0)}$

$$\mathbf{E}S_N = \mathbf{E} \sum_{k=1}^{S_{N-1}} \xi_k^{N-1} = \mathbf{E}S_{N-1}a = a^N$$

(заметим, что зависимость слагаемых предусмотрена в тождестве Вальда). Если $a > 1$, то происходит эффект взрыва — число потомков растет экспоненциально. При $a \leq 1$ — критический процесс (происходит вырождение популяции); с вероятностью 1 рано или поздно произойдет событие $S_N = 0$.

Если при $t = 0$ было M частиц, то в силу линейности среднего мы имеем $\mathbf{E}S_N = Ma^N$. Так что множитель N на асимптотическое поведение среднего числа потомком никак не повлияет.

Глава 4. ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ.

Теорема (закон больших чисел (ЗБЧ) в форме Чебышева). Пусть ξ_1, \dots, ξ_n — независимые одинаково распределенные случайные величины с конечными дисперсиями. Тогда

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{p} \mathbf{E}\xi_1.$$

Прежде всего докажем вспомогательные утверждения.

Лемма (неравенство Маркова). Для любой неотрицательной случайной величины ξ при любом $x > 0$ справедливо неравенство

$$\mathbf{P}(\xi \geq x) \leq \frac{\mathbf{E}\xi I(\xi \geq x)}{x} \leq \frac{\mathbf{E}\xi}{x}.$$

ДОКАЗАТЕЛЬСТВО. Ясно, что $\xi I(\xi \geq \varepsilon) \geq \varepsilon I(\xi \geq \varepsilon)$ (действительно, если $\omega \in \{\xi \geq \varepsilon\}$, то имеем $\xi(\omega) \geq \varepsilon$, если же $\omega \notin \{\xi \geq \varepsilon\}$, то $0 = 0$). В силу монотонности среднего $\mathbf{E}\xi I(\xi \geq \varepsilon) \geq \varepsilon \mathbf{P}(\xi \geq \varepsilon)$, откуда и следует требуемое. \square

Отметим, что если $\mathbf{E}\xi < \infty$, то $\mathbf{P}(|\xi| \geq x) = o(1/x)$, так как

$$\mathbf{E}\xi I(\xi \geq x) = \int_{-\infty}^{\infty} t I(t \geq x) dF_{\xi}(t) = \int_x^{\infty} t dF_{\xi}(t)$$

является хвостом сходящегося интеграла.

Для любой случайной величины ξ обозначим через $\text{supp}\xi$ носитель ее распределения, т.е. минимальное замкнутое множество на вещественной прямой, в которое ξ попадает с вероятностью 1.

Следствие 1 (обобщенное неравенство Маркова) Пусть f — положительная неубывающая функция на множестве $\text{supp}\xi$ (носителе распределения ξ). Тогда для любого $x \in \text{supp}\xi$

$$\mathbf{P}(\xi \geq x) \leq \frac{\mathbf{E}f(\xi)}{f(x)}.$$

ДОКАЗАТЕЛЬСТВО. Ясно, что в силу свойств функции f

$$\mathbf{P}(\xi \geq x) \leq \mathbf{P}(f(\xi) \geq f(x)).$$

Осталось применить неравенство Маркова для величины $\xi' = f(\xi) \geq 0$. \square

Следствие 2 (неравенство Чебышева). Для любой случайной величины ξ с конечной дисперсией

$$\mathbf{P}(|\xi - \mathbf{E}\xi| \geq \varepsilon) \leq \frac{\mathbf{D}\xi}{\varepsilon^2}.$$

ДОКАЗАТЕЛЬСТВО. Введём случайную величину $\xi' = |\xi - \mathbf{E}\xi|$ и воспользуемся следствием 1 для $f(x) = x^2$. \square

ДОКАЗАТЕЛЬСТВО ЗБЧ. Обозначим $\eta = \frac{1}{n} \sum_{i=1}^n \xi_i$. Имеем

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \mathbf{E}\xi_1 \right| \geq \varepsilon \right) &= \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \mathbf{E} \frac{1}{n} \sum_{i=1}^n \xi_i \right| \geq \varepsilon \right) = \mathbf{P}(|\eta - \mathbf{E}\eta| \geq \varepsilon) \\ &\leq \frac{\mathbf{D}\eta}{\varepsilon^2} = \frac{\mathbf{D} \sum_{i=1}^n \xi_i}{n^2 \varepsilon^2} = \frac{\mathbf{D}\xi_1}{n \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Здесь сначала применено неравенство Чебышева, а два равенства после него следуют из однородности второго порядка дисперсии и её аддитивности на классе независимых случайных величин. \square

Надо сказать, что условия теоремы несколько завышены: для выполнения ЗБЧ достаточно требовать лишь существования математического ожидания.

Следующую теорему называют также законом больших чисел в форме Хинчина.

Теорема. ЗБЧ имеет место, если в условиях предыдущей теоремы вместо конечности дисперсии слагаемых предполагается лишь, что $\mathbf{E}|\xi_1| < \infty$.

ДОКАЗАТЕЛЬСТВО. Введем следующие обозначения:

$$\xi_i^{(N)} = \xi_i I\{|\xi_i| \leq N\}, \quad \tilde{\xi}_i^{(N)} = \xi_i - \xi_i^{(N)}.$$

Первая случайная величина — уже использованная ранее срезка случайной величины ξ_i на уровне N , а вторая — *хвост срезки*. Очевидно, что $\xi_i = \xi_i^{(N)} + \tilde{\xi}_i^{(N)}$. Стало быть,

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \mathbf{E}\xi_1 \right| \geq \varepsilon \right) &\leq \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i^{(N)} - \mathbf{E}\xi_1^{(N)} \right| \geq \varepsilon/2 \right) \\ &\quad + \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \tilde{\xi}_i^{(N)} - \mathbf{E}\tilde{\xi}_1^{(N)} \right| \geq \varepsilon/2 \right). \end{aligned}$$

Поскольку случайная величина $\xi_i^{(N)}$ ограничена, то в силу ЗБЧ в форме Чебышева первая вероятность в правой части этого неравенства при $n \rightarrow \infty$ стремится к нулю при любом фиксированном уровне срезки N . Вторая же вероятность оценивается с помощью обобщенного неравенства Маркова для функции $f(x) = |x|$ и неравенством треугольника для модуля и математического ожидания:

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \tilde{\xi}_i^{(N)} - \mathbf{E}\tilde{\xi}_1^{(N)} \right| \geq \varepsilon/2 \right) \leq \frac{4}{\varepsilon} \mathbf{E}|\tilde{\xi}_1^{(N)}|,$$

и выбором достаточно большого N может быть сделана сколь угодно малой, что и требовалось доказать. \square

Пример. Рассмотрим схему Бернулли с вероятностью успеха p . Как обычно, $S_n = \sum_{i=1}^n \xi_i$, где $\{\xi_i\}$ — независимые бернуллиевские случайные величины. Согласно закону больших чисел, имеет место сходимость частоты успехов: $\nu_n = S_n/n \xrightarrow{p}$.

Усиленный закон больших чисел.

В этом параграфе мы покажем, что в условиях ЗБЧ в форме Чебышева справедливо более сильное утверждение, которое носит название *усиленного закона больших чисел* (УЗБЧ).

Теорема (УЗБЧ). Пусть $\{\xi_i\}_{i=1}^{\infty}$ — последовательность независимых одинаково распределенных случайных величин, удовлетворяющих условию $\mathbf{E}\xi_1^2 < \infty$. Тогда

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{\text{п. н.}} \mathbf{E}\xi_1.$$

ДОКАЗАТЕЛЬСТВО. Сразу отметим, что в формулировке теоремы вместо конечности второго момента достаточно требовать лишь существование математического ожидания. Однако в этом случае доказательство значительно усложняется.

Обозначим $S_n = \sum_{i=1}^n \xi_i$ и введем последовательность событий

$$A_k = \left\{ \left| \frac{S_{k^2}}{k^2} - \mathbf{E}\xi_1 \right| \geq \varepsilon \right\},$$

где ε — произвольное положительное фиксированное число.

В силу теоремы Бореля—Кантелли и неравенства Чебышева ряд $\sum_{k \geq 1} \mathbf{P}(A_k)$ сходится, а стало быть, в этом случае с вероятностью 1 будет наблюдаться лишь конечное число событий из последовательности $\{A_k\}$. Поэтому, начиная с некоторого $k_0(\omega)$, будут наблюдаться соответствующие обратные события. Это означает, что подпоследовательность $\left\{ \frac{S_{k^2}}{k^2}; k = 1, 2, \dots \right\}$ с вероятностью 1 сходится к величине $\mathbf{E}\xi_1$.

Далее, без ограничения общности можно считать, что $\xi_i \geq 0$ почти наверное. В противном случае мы можем доказывать УЗБЧ отдельно для последовательностей $\{\xi_i^+\}$ и $\{\xi_i^-\}$, заметив при этом, что $\xi_i = \xi_i^+ - \xi_i^-$, откуда и будет следовать наше утверждение. Теперь положим $k \equiv k(n) = [\sqrt{n}]$. С учетом неотрицательности слагаемых имеем очевидное двойное неравенство

$$\frac{k(n)^2}{n} \cdot \frac{S_{k(n)^2}}{k(n)^2} \leq \frac{S_n}{n} \leq \frac{(k(n)+1)^2}{n} \cdot \frac{S_{(k(n)+1)^2}}{(k(n)+1)^2}.$$

Остается только заметить, что $\frac{k(n)^2}{n} \rightarrow 1$ и $\frac{(k(n)+1)^2}{n} \rightarrow 1$ при $n \rightarrow \infty$, и воспользоваться принципом “двух милиционеров”. \square

З а м е ч а н и е. Наиболее короткое доказательство УЗБЧ имеет место при более жестких ограничениях по сравнению с вышеприведенными, а именно при условии конечности четвертого момента слагаемых. В самом деле, введем в рассмотрение случайные величины $\xi'_i = \xi_i - \mathbf{E}\xi_i$ и $S'_n = \sum_{i=1}^n \xi'_i$, а также последовательность событий $A_n = \{|n^{-1}S'_n| \geq \varepsilon\}$. Далее, воспользовавшись неравенством Маркова с $f(x) = x^4$, получим

$$\mathbf{P}(A_n) \leq \frac{\mathbf{E}(S'_n)^4}{n^4 \varepsilon^4}.$$

Кроме того,

$$\mathbf{E}(S'_n)^4 = \sum_{i_1, i_2, i_3, i_4 \leq n} \mathbf{E}\xi'_{i_1} \xi'_{i_2} \xi'_{i_3} \xi'_{i_4} = n\mathbf{E}\xi_1'^4 + 3n(n-1)(\mathbf{E}\xi_1'^2)^2 \leq Cn^2,$$

где C зависит только от первого, второго и четвертого моментов (например, в случае, когда $i_1 \neq i_2, i_1 \neq i_3, i_1 \neq i_4$, величины ξ'_{i_1} и $\xi'_{i_2} \xi'_{i_3} \xi'_{i_4}$ независимы, поэтому $\mathbf{E}\xi'_{i_1} (\xi'_{i_2} \xi'_{i_3} \xi'_{i_4}) = \mathbf{E}\xi'_{i_1} \cdot \mathbf{E}\xi'_{i_2} \xi'_{i_3} \xi'_{i_4} = 0$, так как $\mathbf{E}\xi'_{i_1} = \mathbf{E}(\xi_i - \mathbf{E}\xi_i) = \mathbf{E}\xi_i - \mathbf{E}\xi_i = 0$).

Итак,

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \sum_{n=1}^{\infty} \mathbf{P}\left(\left|\frac{S_n}{n} - \mathbf{E}\xi_1\right| \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{\mathbf{E}(S'_n)^4}{n^4 \varepsilon^2} \leq \frac{C}{\varepsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Последний ряд, как известно, сходится. Остается воспользоваться теоремой Бореля–Кантелли. \square

Глава 5. УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ.

Напомним, что ранее у нас было введено понятие условной вероятности

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}, \text{ где } \mathbf{P}(B) > 0.$$

Пусть $A = \{\xi \in C\}$, где $C \in \mathcal{B}$. В этом случае мы приходим к определению условного распределения $\mathcal{P}_\xi(C|B)$ случайной величины ξ при условии, что произошло событие B . Тогда естественно назвать *условным математическим ожиданием ξ относительно события B* величину

$$\mathbf{E}(\xi|B) = \int x \mathcal{P}_\xi(dx|B).$$

Далее мы существенно расширим это понятие.

Элементы теории гильбертовых пространств.

Гильбертово пространство H – линейное пространство, в котором задан бинарный функционал (x, y) – *скалярное произведение* со следующими свойствами:

1. $(x, y) = (y, x)$;
2. $(c_1 x_1 + c_2 x_2, y) = c_1(x_1, y) + c_2(x_2, y)$;
3. $(x, x) \geq 0$, $(x, x) = 0 \Leftrightarrow x = 0$;

при этом H – *полное* нормированное пространство с евклидовой нормой $\|x\| = \sqrt{(x, x)}$.

Определим операцию проецирования.

Определение. Пусть L – замкнутое линейное подпространство H (тогда L – тоже полное в индуцированной топологии). Элемент $\hat{x} \in L$ – *ортпроекция x на подпространство L* , если $(x - \hat{x}, y) = 0 \forall y \in L$.

Известно, что при сделанных предположениях ортопроекция \hat{x} существует и единственна.

Упражнение. Доказать, что если ортопроекция \hat{x} существует, то она единственна (доказывается от противного).

Приведем эквивалентное определение ортопроекции. Теперь рассмотрим функционал $\|x - y\|$, где элемент $x \in H$ фиксирован, а y принимает всевозможные значения из замкнутого линейного подпространства L .

Теорема. Элемент $\hat{x} \in L$ будет ортопроекцией x на L тогда и только тогда, когда

$$\hat{x} = \arg \min_{y \in L} \|x - y\|$$

или, что то же,

$$\min_{y \in L} \|x - y\| = \|x - \hat{x}\|. \quad (*)$$

Этот результат называется в функциональном анализе теоремой о перпендикуляре, смысл которой прозрачен: на ортопроекции реализуется расстояние Хаусдорфа, т. е. минимальное расстояние от элемента x до замкнутого линейного подпространства L .

Доказательство. Необходимость. Для любого $y \in L$ имеем

$$\|x - y\|^2 = \|x - \hat{x}\|^2 + \|\hat{x} - y\|^2 + 2(x - \hat{x}, \hat{x} - y).$$

Поскольку $\hat{x} - y \in L$, то в силу тождества ортопроекции скалярное произведение в правой части этого равенства обращается в ноль. Отсюда немедленно следует соотношение (*), поскольку возведение в квадрат неотрицательной функции — монотонное преобразование на положительной полуоси, не меняющее множество ее экстремальных точек.

Достаточность. Покажем, что точка минимума в (*) единственная, откуда и будет следовать наше утверждение. Допустим противное: существует точка $x_0 \in L$, причем $x_0 \neq \hat{x}$, в которой достигается минимум в (*). Но поскольку x_0 не является ортопроекцией, то в линейном пространстве L найдется такой элемент Δ , что $(x - x_0, \Delta) > 0$. Исследуем поведение функции $f(y) = \|x - y\|^2$ на одномерном многообразии $\{y \in L : y = x_0 + t\Delta, t \in R\}$. Имеем

$$f(x_0 + t\Delta) = f(x_0) + t^2\|\Delta\|^2 - 2t(x - x_0, \Delta).$$

Очевидно, для достаточно малых положительных t в наших условиях будет выполнено неравенство $t^2\|\Delta\|^2 - 2t(x - x_0, \Delta) < 0$, что противоречит допущению, что $x_0 \neq \hat{x}$ — точка минимума функции $f(y)$. Теорема доказана.

Введем пространство $L_2(\Omega, \mathbf{P})$ — множество всех случайных величин, заданных на пространстве элементарных исходов и имеющих конечный второй момент (мы не различаем в этом пространстве случайные величины ξ и ξ' , если $\mathbf{P}(\xi = \xi') = 1$. Все такие случайные величины «стягиваются» в одну точку, т. е. мы тем самым вводим классы эквивалентности или, как говорят, строим фактор-пространство

$$L_2(\Omega, \mathbf{P}) = \{\xi : \mathbf{E}\xi^2 < \infty\}.$$

В этом пространстве введем скалярное произведение (второй смешанный момент): $(\xi, \eta) = \mathbf{E}\xi\eta$. Очевидно, что скалярное произведение задано корректно. Например, для проверки третьего свойства нужно отметить $\mathbf{E}\xi^2 = 0$ тогда и только тогда, когда $\xi = 0$ с вероятностью 1. Тогда получаем гильбертово пространство с нормой

$$\|\xi\| = \sqrt{\mathbf{E}\xi^2} = \left(\int_{\Omega} \xi^2(\omega) \mathbf{P}(d\omega) \right)^{1/2}.$$

Рассмотрим пару случайных величин $\xi \in \mathbb{R}$ и $\eta \in \mathbb{R}^n$. Введем замкнутое линейное подпространство в $L_2(\Omega, \mathbf{P})$:

$$L(\eta) = \{g(\eta) : \mathbf{E}g^2(\eta) < \infty\} = \left\{ g(x) : \int g^2(x) \mathcal{P}_{\eta}(dx) < \infty \right\} = L_2(\mathbb{R}^n, \mathcal{P}_{\eta}),$$

т. е. $L(\eta)$ — класс всех детерминированных (неслучайных) борелевских преобразований вектора η с конечным вторым моментом. Этот класс еще называют пространством, порожденным вектором *сопутствующих наблюдений* η .

Определение. Условное математическое ожидание (УМО) случайной величины ξ при фиксации вектора сопутствующих наблюдений η есть ортопроекция в $L_2 \equiv L_2(\Omega, \mathbf{P})$ случайной величины ξ на замкнутое линейное подпространство $L(\eta)$:

$$\mathbf{E}(\xi|\eta) = \widehat{\xi}|_{L(\eta)} = \widehat{g}(\eta).$$

Эту случайную величину еще называют УМО ξ относительно η .

З а м е ч а н и е. В классической вероятностной литературе аналогичное условное математическое ожидание вводится для более широкого класса случайных величин, имеющих лишь первый момент. На наш взгляд, это же понятие, определенное для более узкого класса случайных величин с конечными вторыми моментами (что вполне достаточно для многих приложений, таких как задачи прогноза или построения эффективных статистических оценок), выгодно отличается от общепринятой схемы своей наглядной геометрической интерпретацией, что позволяет лучше понять этот довольно непростой математический объект.

Свойства УМО

1. **Линейность.** $\mathbf{E}(c_1\xi_1 + c_2\xi_2|\eta) = c_1\mathbf{E}(\xi_1|\eta) + c_2\mathbf{E}(\xi_2|\eta)$. Напомним, что здесь равенство понимается как равенство в пространстве L_2 , т. е. мы не различаем случайные величины, отличные на множестве вероятности ноль. В дальнейшем мы будем понимать все равенства и неравенства именно так, не снабжая знаки равенство аббревиатурой «п.н.».

В силу определения ортопроекции $\forall g \in L(\eta)$

$$(\xi - \widehat{\xi}, g(\eta)) = 0 = \mathbf{E}(\xi - \widehat{\xi})g(\eta).$$

Отсюда мы получаем *тождество ортопроекции*

$$\mathbf{E}\widehat{\xi g}(\eta) = \mathbf{E}\xi g(\eta), \quad \forall g \in L(\eta).$$

Докажем линейность УМО. Мы хотим доказать, что

$$c_1\widehat{\xi_1} + c_2\widehat{\xi_2} = c_1\widehat{\xi_1} + c_2\widehat{\xi_2}.$$

Подставляя в тождество ортопроекции правую часть этого (предполагаемого!) равенства, получаем

$$\mathbf{E}(c_1\mathbf{E}(\xi_1|\eta) + c_2\mathbf{E}(\xi_2|\eta))g(\eta) = c_1\mathbf{E}(\mathbf{E}(\xi_1|\eta)g(\eta)) + c_2\mathbf{E}(\mathbf{E}(\xi_2|\eta)g(\eta)) =$$

[в силу тождества ортопроекции для ξ_1 и ξ_2]

$$= c_1\mathbf{E}\xi_1g(\eta) + c_2\mathbf{E}\xi_2g(\eta) = \mathbf{E}(c_1\xi_1 + c_2\xi_2)g(\eta)$$

Следовательно, в силу единственности ортопроекции,

$$c_1\widehat{\xi_1} + c_2\widehat{\xi_2} = c_1\widehat{\xi_1} + c_2\widehat{\xi_2}.$$

2. Монотонность. Если $\xi \geq 0$, то $\widehat{g}(\eta) := \mathbf{E}(\xi|\eta) \geq 0 \quad \forall \eta$.

Предположим противное: $\exists \Omega_0 \subset \Omega : \mathbf{P}(\Omega_0) \neq 0$ и $\widehat{g}(\eta) < 0$ для всех элементарных исходов из Ω_0 . Рассмотрим измеримую ограниченную функцию $g_0(\eta) = \mathbf{I}(\widehat{g}(\eta) < 0)$. Тогда $g_0(\eta) \in L_2$. Далее используем тождество ортопроекции:

$$\mathbf{E}(\mathbf{E}(\xi|\eta))g_0(\eta) = \mathbf{E}\xi g_0(\eta).$$

Тогда если $\mathbf{P}(\Omega_0) \neq 0$, то одновременно выполнены следующие два неравенства: $\mathbf{E}(\mathbf{E}(\xi|\eta))g_0(\eta) < 0$ и $\mathbf{E}\xi g_0(\eta) \geq 0$, что невозможно. Мы получили противоречие. Значит, $\mathbf{P}(\Omega_0) = 0$.

Следствие. Если $\xi_1 \geq \xi_2$, то $\mathbf{E}(\xi_1|\eta) \geq \mathbf{E}(\xi_2|\eta) \quad \forall \eta$. Доказательство напрямую следует из линейности и монотонности. В частности, отсюда немедленно (сравни с доказательством аналогичного утверждения для обычных средних) следует неравенство треугольника

$$|\mathbf{E}(\xi|\eta)| \leq \mathbf{E}(|\xi||\eta).$$

3. Пусть f – ограниченная функция. Тогда для любого вектора сопутствующих наблюдений η

$$\mathbf{E}(\xi f(\eta)|\eta) = f(\eta)\mathbf{E}(\xi|\eta).$$

Упражнение. Доказать, что это утверждение можно распространить на неограниченные функции f при условии, что $\xi f(\eta) \in L_2$ (использовать срезы случайной величины $f(\eta)$).

ДОКАЗАТЕЛЬСТВО. Нам нужно доказать, что

$$\widehat{\xi f(\eta)} = f(\eta)\mathbf{E}(\xi|\eta).$$

По тождеству ортопроекции

$$\mathbf{E}f(\eta)\mathbf{E}(\xi|\eta)g(\eta) = \mathbf{E}\mathbf{E}(\xi|\eta)\tilde{g}(\eta) = \mathbf{E}\xi\tilde{g}(\eta) = \mathbf{E}(\xi f(\eta))\tilde{g}(\eta).$$

Первое равенство справедливо в силу того, что $\tilde{g}(\eta) = f(\eta)g(\eta) \in L_2$ (так как f ограничена, а $g(\eta)$ имеет второй момент, то и $f(\eta)g(\eta)$ имеет второй момент).

4. *Формула полной вероятности для УМО:*

$$\mathbf{E}\mathbf{E}(\xi|\eta) = \mathbf{E}\xi.$$

Эта формула прямо получается из тождества ортопроекции при $g(\eta) \equiv 1$.

5. *Если случайные величины ξ и η независимы, то $\mathbf{E}(\xi|\eta) = \mathbf{E}\xi$.*

В самом деле,

$$\mathbf{E}(\mathbf{E}\xi)g(\eta) = \mathbf{E}\xi\mathbf{E}g(\eta) = \mathbf{E}\xi g(\eta).$$

Последнее равенство имеет место в силу независимости. Значит, тождество ортопроекции для правой части предполагаемого выше равенства выполнено.

Следующее свойство обобщает предыдущее.

6. *Пусть случайные величины ξ и η – независимы, $f(x, y)$ – измеримая функция, для которой $\mathbf{E}f^2(\xi, \eta) < \infty$. Тогда*

$$\mathbf{E}(f(\xi, \eta)|\eta) = \mathbf{E}f(\xi, y)|_{y=\eta}.$$

Приведенное равенство указывает нам правило вычисления $\mathbf{E}(f(\xi, \eta)|\eta)$ в случае независимых компонент.

ДОКАЗАТЕЛЬСТВО. Пусть $\varphi(y) = \mathbf{E}f(\xi, y)$. Тогда

$$\begin{aligned} \mathbf{E}\varphi(\eta)g(\eta) &= \int \varphi(y)g(y)\mathcal{P}_\eta(dy) = \int \left[\int f(x, y)\mathcal{P}_\xi(dx) \right] \bar{g}(y)\mathcal{P}_\eta(dy) = \\ &= \int \int f(x, y)g(y)\mathcal{P}_\xi(dx)\mathcal{P}_\eta(dy) = \mathbf{E}f(\xi, \eta)g(\eta). \end{aligned}$$

Так как $\iint |f(x, y)|^2\mathcal{P}_\xi(dx)\mathcal{P}_\eta(dy) < \infty$, то $\iint |f(x, y)|\mathcal{P}_\xi(dx)\mathcal{P}_\eta(dy) < \infty$. Следовательно, можно применить теорему Фубини и переставить интегралы (третье равенство).

Упражнение. Доказать формулу свертки для произвольных распределений:

$$F_{\xi_1+\xi_2}(x) = \int_R F_{\xi_1}(x-t)dF_{\xi_2}(t),$$

где случайные величины ξ_1 и ξ_2 независимы.

7. Определим условную функцию распределения случайной величины ξ при фиксации сопутствующего наблюдения η по формуле $F_\xi(t|\eta) = \mathbf{E}(I(\xi < t)|\eta)$.

Введенная функция по t обладает всеми свойствами обычных функций распределения (но все соотношения выполняются с вероятностью 1). Скажем, монотонность сразу следует из свойства (2) УМО. Покажем, что и пределы этих функций на бесконечности положительного или отрицательного знаков будут соответственно 1 либо 0. В самом деле, в силу монотонности существует измеримый предел $g^+(\eta) = \lim_{t \rightarrow \infty} F_\xi(t|\eta) \leq 1$. Если допустить, что событие $\Omega_\varepsilon = \{g^+(\eta) < 1 - \varepsilon\}$ будет иметь ненулевую вероятность для некоторого положительного ε , то в силу тождества ортопроекции и теоремы Лебега о мажорируемой сходимости мы получаем равенство $\mathbf{E}F_\xi(t|\eta)I(\Omega_\varepsilon) = \mathbf{E}I(\xi < t)I(\Omega_\varepsilon) \rightarrow \mathbf{P}(\Omega_\varepsilon)$ при $t \rightarrow \infty$. В то же время, левая часть этого тождества не превосходит величину $(1 - \varepsilon)\mathbf{P}(\Omega_\varepsilon)$. Получили противоречие, которое явилось следствием предположения, что $\mathbf{P}(\Omega_\varepsilon) > 0$.

Совершенно аналогично доказываются равенство нулю предела условной функции распределения при $t \rightarrow -\infty$.

Упражнение. Доказать, что любая условная функция распределения с вероятностью 1 непрерывна слева.

Упражнение. Доказать, что

$$\mathbf{E}(\xi|\eta) = \int_{\mathbb{R}} x dF_\xi(x|\eta),$$

где интеграл понимается в смысле Римана–Стилтьеса.

8. Пусть (ξ, η) – случайный вектор с плотностью совместного распределения координат $p_{\xi, \eta}(x, y)$ (где ξ – скаляр). Тогда верна формула Байеса

$$\mathbf{E}(\xi|\eta) = \int_{\mathbb{R}} x \frac{p_{\xi, \eta}(x, \eta)}{p_\eta(\eta)} dx,$$

где $\frac{p_{\xi, \eta}(x, \eta)}{p_\eta(\eta)}$ называется апостериорной плотностью ξ при фиксации η .

ДОКАЗАТЕЛЬСТВО. Обозначим

$$\varphi(\eta) = \int_{\mathbb{R}} x \frac{p_{\xi, \eta}(x, \eta)}{p_\eta(\eta)} dx.$$

Докажем, что $\varphi(\eta) = \widehat{g}(\eta)$. Вновь проверим тождество ортопроекции:

$$\mathbf{E}\varphi(\eta)g(\eta) = \int \varphi(y)g(y)\mathcal{P}_\eta(dy) = \int \left[\int x \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)} dx \right] g(y)\mathcal{P}_\eta(dy) =$$

[так как интегрируем на носителе η , то $p_\eta(y) = \int_{\mathbb{R}} p_{\xi, \eta}(x, y) dx \neq 0$]

$$= \int \left[\int x \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)} dx \right] g(y)p_\eta(y) dy =$$

[по теореме Фубини]

$$= \iint xg(y)p_{\xi,\eta}(x,y)dxdy = \mathbf{E}\xi g(\eta).$$

Из теоремы единственности вновь следует, что $\varphi(\eta) = \widehat{\xi}$.

9. Пусть даны произвольная случайная величина $\xi \in L_2$ и дискретная случайная величина η с атомами $\{a_i\}$ и соответствующими массами $\{p_i\}$. Тогда

$$\mathbf{E}(\xi|\eta) = \begin{cases} \mathbf{E}(\xi|\eta = a_i) & \text{с вероятностью } p_i, \\ i = 1, 2, \dots, & (p_i \neq 0). \end{cases}$$

ДОКАЗАТЕЛЬСТВО. Напомним, что классическое условное среднее определяется по формуле

$$\mathbf{E}(\xi|B) = \frac{\mathbf{E}\xi I(B)}{\mathbf{P}(B)},$$

где B — событие ненулевой вероятности. Нам надо доказать, что

$$\mathbf{E}(\xi|\eta) = \sum_{i \geq 1} \mathbf{E}(\xi|\eta = a_i)I(\eta = a_i).$$

По тождеству ортопроекции имеем

$$\begin{aligned} \mathbf{E}g(\eta) \sum_{i \geq 1} \mathbf{E}(\xi|\eta = a_i)I(\eta = a_i) &= \sum_{i \geq 1} \mathbf{E}(\xi|\eta = a_i)\mathbf{E}g(\eta)I(\eta = a_i) = \\ &= \sum_{i \geq 1} \frac{\mathbf{E}(\xi I(\eta = a_i))}{p_i} \mathbf{E}g(\eta)I(\eta = a_i) = \sum_{i \geq 1} \frac{\mathbf{E}(\xi I(\eta = a_i))}{p_i} g(a_i)p_i \\ &= \sum_{i \geq 1} \mathbf{E}(\xi I(\eta = a_i))g(a_i) = \mathbf{E}\xi \sum_{i \geq 1} g(a_i)I(\eta = a_i) = \mathbf{E}\xi g(\eta), \end{aligned}$$

так как дискретная случайная величина $g(\eta)$, очевидно, задается по формуле $g(\eta) = \sum_{i \geq 1} g(a_i)I(\eta = a_i)$.

Задачи прогноза случайных последовательностей.

Пусть имеется последовательность случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, которые могут быть зависимыми. Индекс интерпретируем как момент времени, считая « n » за настоящее время, то есть последовательность $\xi_1, \xi_2, \dots, \xi_n$ описывает всю предысторию вплоть до настоящего времени. Задача состоит в том, чтобы наиболее точно предсказать «завтрашнее» наблюдение, то есть ξ_{n+1} . Будем строить оценку наблюдения ξ_{n+1} в виде некоторой детерминированной функции от всей предыстории наблюдений: $\xi_{n+1}^* = g(\xi_1, \xi_2, \dots, \xi_n)$. Оценка должна быть оптимальной в следующем смысле: она должна минимизировать среднеквадратическое отклонение от будущего наблюдения ξ_{n+1} :

$$\mathbf{E}(\xi_{n+1} - \xi_{n+1}^*)^2 \rightarrow \min_{g \in L},$$

где L – то или иное пространство функций. Если найдем «точку» минимума $\hat{g}(\cdot)$, то будем говорить, что ξ_{n+1}^* – это *оптимальный прогноз в пространстве L* для случайной последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$. Введем в рассмотрение вектор сопутствующих (прогнозу) наблюдений – всю предысторию наблюдений:

$$\eta = (\xi_1, \xi_2, \dots, \xi_n).$$

Рассмотрим семейство случайных величин $L(\eta) = \{g(\eta) : \mathbf{E}g^2(\eta) < \infty\}$ – замкнутое линейное подпространство пространства $L_2(\Omega, \mathbf{P})$ (здесь g – всевозможные детерминированное преобразование с указанным конечным вторым моментом), порожденное вектором сопутствующих наблюдений.

Уже было отмечено, что имеется два эквивалентных определения ортопроекции:

$$(\xi - \hat{\xi}, y) \equiv 0 \quad \forall y \in L$$

или

$$\hat{\xi} = \arg \min_{y \in L} \|\xi - y\|^2 = \arg \min_{g \in L(\eta)} \mathbf{E}(\xi - g(\eta))^2.$$

Ясно, что мы ищем оптимальный прогноз в классе L_2 , т. е. при условии $\mathbf{E}g^2(\eta) < \infty$. Нам требуется минимизировать $\mathbf{E}(\xi_{n+1} - g(\xi_1, \xi_2, \dots, \xi_n))^2$ по всевозможным $g \in L_2(\mathbb{R}^n, \mathcal{P}_\eta)$. Иными словами, мы свели задачу к вычислению соответствующей ортопроекции:

$$\min_{g \in L(\eta)} \mathbf{E}(\xi_{n+1} - g(\eta))^2 = \mathbf{E}(\xi_{n+1} - \hat{\xi}_{n+1})^2,$$

где $\hat{\xi}_{n+1} = \hat{g}(\xi_1, \xi_2, \dots, \xi_n)$.

Важно отметить, что если мы не располагаем никакой информацией о совместном распределении компонент $\xi_1, \dots, \xi_n, \xi_{n+1}$, то оптимальный прогноз построить невозможно. То есть для того, чтобы заниматься прогнозом, нужно знать вероятностный закон, управляющий всей последовательностью. Только тогда задача будет корректна. Скажем, если указанное совместное распределение имеет плотность то, как мы уже знаем,

$$\xi_{n+1}^* = \int_{\mathbb{R}} x \frac{p_{\xi_1, \xi_2, \dots, \xi_n, \xi_{n+1}}(\xi_1, \xi_2, \dots, \xi_n, x)}{p_{\xi_1, \xi_2, \dots, \xi_n}(\xi_1, \xi_2, \dots, \xi_n)} dx.$$

Оптимальный линейный прогноз случайных последовательностей.

Пусть $\xi_1, \dots, \xi_n, \xi_{n+1}, \dots$ – последовательность случайных величин из $L_2(\Omega, \mathbf{P})$. Введем подпространство, состоящее из линейных комбинаций координат вектора сопутствующих наблюдений:

$$L_0(\xi_1, \dots, \xi_n) = \left\{ \sum_{i=1}^n c_i \xi_i : c_i \in \mathbb{R} \right\}.$$

Конечномерное линейное пространство $L_0(\eta)$ замкнуто относительно евклидовой нормы, причем $L_0(\eta) \subset L(\eta)$.

Построим *оптимальный линейный* прогноз, т. е. функцию $\widehat{g}(\eta) = \sum_{i=1}^n \widehat{c}_i \xi_i$, которая минимизирует на пространстве L_0 среднее квадратичное отклонение

$$\min_{g \in L_0} \mathbf{E}(\xi_{n+1} - g(\xi_1, \dots, \xi_n))^2.$$

Из предыдущих рассмотрений следует, что случайная величина $\widehat{g}(\xi_1, \dots, \xi_n)$ представляет собой ортопроекцию элемента ξ_{n+1} на замкнутое линейное подпространство $L_0(\eta)$, которое существенно уже пространства $L(\eta)$. Ясно, что в этом случае ортопроекции случайной величины ξ_{n+1} на линейные подпространства $L_0(\eta)$ и $L(\eta)$ не обязаны совпадать.

Далее, по тождеству ортопроекции $\mathbf{E}(\xi_{n+1} - \widehat{g}(\xi_1, \dots, \xi_n))\zeta = 0, \forall \zeta \in L_0(\eta)$. В качестве ζ возьмем n различных порождающих элементов из $L_0(\eta)$: ξ_1, \dots, ξ_n . Тогда имеем

$$(\xi_{n+1} - \widehat{\xi}_{n+1}, \xi_k) = 0, \text{ т. е. } \sum_{i=1}^n c_i \mathbf{E} \xi_i \xi_k = \mathbf{E} \xi_{n+1} \xi_k, \quad k = 1, \dots, n.$$

Получили систему n линейных уравнений относительно $\{c_i; i = 1, \dots, n\}$. Матрица этой системы — это классическая ковариационная матрица

$$C = \|\mathbf{E}(\xi_i, \xi_k)\|_{n \times n}.$$

Это симметричная неотрицательно определенная матрица, так как

$$\sum_{j,i=1}^n x_i x_j \mathbf{E} \xi_i \xi_j = \mathbf{E} \sum_{j,i=1}^n x_i x_j \xi_i \xi_j = \left(\sum_{i=1}^n x_i \xi_i \right)^2 \geq 0$$

для любых вещественных $\{x_i\}$. Отметим, что если эта ковариационная матрица строго положительно определена, то она будет обратимой, что немедленно приводит к *единственности* линейного прогноза.

Когда же условие строгой положительной определенности указанной матрицы может нарушаться? Допустим, что для некоторого набора чисел $\{\tilde{x}_i\}$, не все из которых равны нулю, выполнено

$$\mathbf{E} \left(\sum_{i=1}^n \tilde{x}_i \xi_i \right)^2 = 0.$$

Тогда $\sum_{i=1}^n \tilde{x}_i \xi_i = 0$ с вероятностью 1. Следовательно, $\{\xi_i\}$ *линейно связаны*, что превращает задачу прогноза в малосодержательную: мы можем представить наблюдение ξ_{n+1} как линейную комбинацию предшествующих. Такой «детерминированный» прогноз мы в дальнейшем рассматривать не будем. Тогда $C > 0$ (т. е. матрица C положительно определена), и система имеет единственное решение.

Конечно, возникает проблема вычислений смешанных моментов $\mathbf{E}(\xi_i, \xi_k)$ при всевозможных индексах i и j . Пусть имеются N серий наблюдений, которые представляют

собой независимые «копии» (т. е. совпадающие по распределению) исходной последовательности:

$$\begin{aligned} &\xi_1^{(1)}, \dots, \xi_n^{(1)}, \xi_{n+1}^{(1)}, \\ &\xi_1^{(2)}, \dots, \xi_n^{(2)}, \xi_{n+1}^{(2)}, \\ &\dots \end{aligned}$$

Если речь идет, например, о прогнозе погоды за год, то отмеченные последовательности-копии представляют собой соответствующие серии наблюдений за многие предыдущие годы (скажем, 100 лет). Построим среднюю сумму $\frac{1}{N} \sum_{j=1}^N \xi_i^{(j)} \xi_k^{(j)}$, где $\xi_i^{(j)}, \xi_k^{(j)}$ — независимые по верхнему индексу ограниченные случайные величины. Когда N велико, в силу ЗБЧ получаем

$$\frac{1}{N} \sum_{j=1}^N \xi_i^{(j)} \xi_k^{(j)} \sim \mathbf{E} \xi_i^{(1)} \xi_k^{(1)}.$$

с вероятностью 1, т. е. массив смешанных моментов $\{\mathbf{E} \xi_i \xi_k\}$ может быть легко оценен с той точностью, которая определяется объемом предшествующей информации. После этого остается только решить приведенную выше систему линейных уравнений.

Многомерные нормальные (гауссовские) распределения.

Сначала напомним определение скалярной случайной величина ξ с нормальным распределением: $\xi \in N(\alpha, \sigma)$, если

$$p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}}.$$

Определение. Вектор $\bar{\xi} = (\xi_1, \dots, \xi_n)$ имеет *многомерное нормальное распределение*, если

$$p_{\bar{\xi}}(\bar{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{|C|}} \exp\left\{-\frac{1}{2}((\bar{x} - \bar{\alpha})C^{-1}, \bar{x} - \bar{\alpha})\right\},$$

где $\bar{x} = (x_1, \dots, x_n)$, $|C|$ и C^{-1} — соответственно детерминант и обратная матрица для ковариационной матрицы $C = \|c_{ij}\|$, где $c_{ij} = \mathbf{Cov}(\xi_i, \xi_j) = \mathbf{E}(\xi_i - \mathbf{E}\xi_i)(\xi_j - \mathbf{E}\xi_j)$.

В дальнейшем условимся, что координаты ξ_i являются линейно независимыми (см. рассуждения предыдущего пункта), т. е. мы рассматриваем *невыврожденное* гауссовское распределение, для которого ковариационная матрица положительно определена (значит, $|C| > 0$). Без ограничения общности можно ограничиться рассмотрением только центрированных случайных величин (т. е. $\bar{\alpha} = \bar{0}$).

Напомним, что если ξ и η независимы, то $\mathbf{Cov}(\xi, \eta) = 0$. Обратное неверно. Например, если ξ *симметрично* распределена, а $\eta = \xi^2$ — невырождена, то эти две случайные величины будут зависимыми (они функционально связаны), но $\mathbf{Cov}(\xi, \eta) = 0$ (если ковариация определена).

Упражнение. Показать, что в последнем примере нет факторизации совместного распределения.

Теорема. Пусть $\bar{\xi}$ — гауссовский вектор (центрированный). Рассмотрим два подвектора $\bar{\xi}_1 = (\xi_1, \dots, \xi_m)$ и $\bar{\xi}_2 = (\xi_{m+1}, \dots, \xi_n)$. Тогда $\bar{\xi}_1$ и $\bar{\xi}_2$ независимы тогда и только тогда, когда $\mathbf{E}\xi_i\xi_j = 0, \forall i \leq m, \forall j \geq m+1$.

ДОКАЗАТЕЛЬСТВО. (\rightarrow) Доказано ранее.

(\leftarrow) Введем обозначения $\bar{x}_1 = (x_1, \dots, x_m)$ и $\bar{x}_2 = (x_{m+1}, \dots, x_n)$. Тогда

$$(\bar{x}C^{-1}, \bar{x}) = ((\bar{x}_1C_1^{-1}, \bar{x}_2C_2^{-1}), (\bar{x}_1, \bar{x}_2)) = (\bar{x}_1C_1^{-1}, \bar{x}_1) + (\bar{x}_2C_2^{-1}, \bar{x}_2),$$

$$|C| = |C_1||C_2|,$$

где C_1 и C_2 — ковариационные матрицы случайных векторов $\bar{\xi}_1$ и $\bar{\xi}_2$ соответственно. Следовательно, исходная плотность факторизуется на две и выполнен критерий независимости. \square

Заметим, что гауссовское распределение — единственное из абсолютно непрерывных, для которого верно утверждение приведенной теоремы.

Лемма. Пусть $\bar{\xi} = (\xi_1, \dots, \xi_n)$ — гауссовский вектор с плотностью

$$p_{\bar{\xi}}(\bar{\xi}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{|C|}} \exp\left\{-\frac{1}{2}((\bar{x} - \bar{\alpha})C^{-1}, \bar{x} - \bar{\alpha})\right\},$$

где $C = \|\mathbf{Cov}(\xi_i, \xi_j)\| > 0$. Тогда для любой невырожденной матрицы A размера $n \times n$ (т. е. $|A| = |\det A| > 0$), $\bar{\eta} = \bar{\xi}A$ — снова гауссовский вектор.

ДОКАЗАТЕЛЬСТВО. В одномерном случае это утверждение, очевидно, следует из формулы для плотности монотонного преобразования случайных величин с абсолютно непрерывным распределением. Докажем утверждение для n -мерного случая. Воспользуемся определением плотности через интегральное представление. Тогда для любого борелевского $B \subset \mathbb{R}^n$ должно выполняться

$$\mathcal{P}_{\bar{\eta}}(B) = \mathbf{P}(\bar{\eta} \in B) = \int_B p(\bar{t})d\bar{t}.$$

Пусть ξ — случайный вектор с абсолютно непрерывным распределением (не обязательно гауссовский). Тогда

$$\mathcal{P}_{\bar{\eta}}(B) = \mathbf{P}(\bar{\eta} \in B) = \mathbf{P}(\bar{\xi}A \in B) = \int_{\bar{x}A \in B} p_{\bar{\xi}}(\bar{x})d\bar{x} = [\bar{y} = \bar{x}A] = \int_B \frac{1}{|A|} p_{\bar{\xi}}(\bar{y}A^{-1})d\bar{y}.$$

Тогда

$$p_{\bar{\eta}}(\bar{y}) = \frac{1}{|A|} p_{\bar{\xi}}(\bar{y}A^{-1}) =$$

[теперь докажем, что если $p_{\bar{\xi}}$ — гауссовская плотность, то $p_{\bar{\eta}}$ — также гауссовская. Для простоты рассмотрим центрированные величины, т. е. положим $\bar{\alpha} = \bar{0}$]

$$= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{|C||A|}} \exp\left\{-\frac{1}{2}((\bar{y}A^{-1})C^{-1}, \bar{y}A^{-1})\right\} =$$

[известно, что $(\bar{z}, \bar{z}\tilde{A}) = (\bar{z}\tilde{A}^T, \bar{z})$]

$$= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{|C||A|}} \exp\left\{-\frac{1}{2}((\bar{y}A^{-1}C^{-1}(A^T)^{-1}, \bar{y}))\right\} =$$

[напомним, что $(AB)^{-1} = B^{-1}A^{-1}$, и используем эту формулу дважды. Также заметим, что $|A| = |A^T| \Rightarrow |A| = \sqrt{|A|}\sqrt{|A^T|}$]

$$= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{|A^TCA|}} \exp\left\{-\frac{1}{2}((\bar{y}(A^TCA)^{-1}, \bar{y}))\right\}. \quad \square$$

Покажем, что A^TCA – ковариационная матрица для вектора

$$\bar{\eta} := \bar{\xi}A = \left(\sum_{i=1}^n \xi_i a_{i,1}, \dots, \sum_{i=1}^n \xi_i a_{i,n}\right).$$

В случае центрированных величин ковариационная матрица совпадает с матрицей вторых смешанных моментов, совпадающей с матрицей A^TCA :

$$\tilde{c}_{k,j} := \mathbf{E} \sum_{i_1=1}^n \xi_{i_1} a_{i_1,k} \sum_{i_2=1}^n \xi_{i_2} a_{i_2,j} = \sum_{i_1, i_2=1}^n \mathbf{E} \xi_{i_1} \xi_{i_2} a_{i_1,k} a_{i_2,j}.$$

Прогноз гауссовских последовательностей.

Теорема. Пусть $\xi_1, \dots, \xi_n, \xi_{n+1}, \dots$ – гауссовская последовательность центрированных случайных величин, т. е. каждые n первых элементов этой последовательности имеют многомерное гауссовское распределение. Тогда оптимальный прогноз совпадает с линейным:

$$\hat{\xi}_{n+1} = \mathbf{E}(\xi_{n+1} | \xi_1, \dots, \xi_n) = \sum_{i=1}^n \hat{c}_i \xi_i,$$

где $\{\hat{c}_i; i = 1, \dots, n\}$ – единственное решение системы линейных уравнений

$$\mathbf{E} \xi_{n+1} \xi_k = \sum_{i=1}^n \hat{c}_i \mathbf{E} \xi_i \xi_k, \quad k = 1, \dots, n.$$

ДОКАЗАТЕЛЬСТВО. Рассмотрим вспомогательный вектор размерности $n + 1$

$$\bar{\zeta} = \left(\xi_1, \xi_2, \dots, \xi_n, \xi_{n+1} - \sum_{i=1}^n \hat{c}_i \xi_i\right).$$

Этот вектор представляет собой результат невырожденного линейного преобразования вектора $(\xi_1, \dots, \xi_{n+1})$: $\bar{\zeta} = (\xi_1, \xi_2, \dots, \xi_n, \xi_{n+1})A$, где

$$A = \begin{pmatrix} 1 & \dots & 0 & \dots & -\widehat{c}_1 \\ \vdots & & \ddots & & -\widehat{c}_n \\ 0 & \dots & & & 1 \end{pmatrix}$$

и $|A| = 1$. Вектор $(\xi_1, \dots, \xi_{n+1})$ гауссовский по условию теоремы. Тогда, как было установлено в предыдущем параграфе, вектор $\bar{\zeta}$ — тоже гауссовский. А так как $\mathbf{E}\xi_{n+1}\xi_k - \sum_{i=1}^n \widehat{c}_i \mathbf{E}\xi_i \xi_k = 0$ при всех $k = 1, \dots, n$, то $(n+1)$ -я координата вектора $\bar{\zeta}$ некоррелирована с предшествующими. Следовательно, случайная величина $\xi_{n+1} - \sum_{i=1}^n \widehat{c}_i \xi_i$ и n -мерный вектор (ξ_1, \dots, ξ_n) независимы между собой. Тогда

$$\widehat{\xi}_{n+1}|_{L(\bar{\eta})} = \mathbf{E}(\xi_{n+1} | \xi_1, \dots, \xi_n) =$$

[по свойствам 1, 5 и 6 УМО]

$$= \mathbf{E} \left(\xi_{n+1} - \sum_{i=1}^n \widehat{c}_i \xi_i | \xi_1, \dots, \xi_n \right) + E \left(\sum_{i=1}^n \widehat{c}_i \xi_i | \xi_1, \dots, \xi_n \right) = \sum_{i=1}^n \widehat{c}_i \xi_i.$$

Здесь мы воспользовались также центрированностью наблюдений:

$$\mathbf{E} \left(\xi_{n+1} - \sum_{i=1}^n \widehat{c}_i \xi_i | \xi_1, \dots, \xi_n \right) = \mathbf{E} \left(\xi_{n+1} - \sum_{i=1}^n \widehat{c}_i \xi_i \right) = 0. \quad \square$$

Оптимальный линейный прогноз часто используется в различных приложениях, например, при прогнозе погоды.

Глава 6. ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ (ПРЕОБРАЗОВАНИЕ ФУРЬЕ РАСПРЕДЕЛЕНИЙ).

Определение и свойства характеристических функций.

Определение. Под комплекснозначной случайной величиной ζ понимается $\zeta = \xi_1 + i\xi_2$, где (ξ_1, ξ_2) — двумерный случайный вектор. Математическим ожиданием этой величины мы назовём $\mathbf{E}\zeta = \mathbf{E}\xi_1 + i\mathbf{E}\xi_2$.

Понятно, что в этом случае не все свойства обычного математического ожидания будут выполняться (например, линейность имеет место, а монотонность — только «по-координатная»).

Определение. Комплекснозначные случайные величины ζ_1 и ζ_2 независимы, если соответствующие векторы их вещественных и мнимых частей $(\xi_1^{(1)}, \xi_2^{(1)})$ и $(\xi_1^{(2)}, \xi_2^{(2)})$ независимы.

Теорема (правило умножения для математических ожиданий). *Если случайные комплекснозначные величины ζ_1 и ζ_2 независимы, то при условии существования $\mathbf{E}\zeta_1$ и $\mathbf{E}\zeta_2$ имеет место равенство $\mathbf{E}\zeta_1\zeta_2 = \mathbf{E}\zeta_1\mathbf{E}\zeta_2$.*

ДОКАЗАТЕЛЬСТВО. Выделим в произведении величин действительную и мнимую части:

$$\zeta_1 \cdot \zeta_2 = \operatorname{Re} \zeta_1 \cdot \operatorname{Re} \zeta_2 - \operatorname{Im} \zeta_1 \cdot \operatorname{Im} \zeta_2 + i(\operatorname{Re} \zeta_1 \cdot \operatorname{Im} \zeta_2 + \operatorname{Re} \zeta_2 \cdot \operatorname{Im} \zeta_1).$$

Возьмём от обеих частей последнего равенства математическое ожидание, используя при этом теорему умножения для вещественных случайных величин:

$$\mathbf{E}\zeta_1\zeta_2 = \mathbf{E}\operatorname{Re} \zeta_1 \cdot \mathbf{E}\operatorname{Re} \zeta_2 - \mathbf{E}\operatorname{Im} \zeta_1 \cdot \mathbf{E}\operatorname{Im} \zeta_2 + i(\mathbf{E}\operatorname{Re} \zeta_1 \cdot \mathbf{E}\operatorname{Im} \zeta_2 + \mathbf{E}\operatorname{Re} \zeta_2 \cdot \mathbf{E}\operatorname{Im} \zeta_1) = \mathbf{E}\zeta_1 \cdot \mathbf{E}\zeta_2 \quad \square$$

Заметим, что с помощью математической индукции приведенная теорема распространяется на любое конечное число независимых множителей.

Далее, мы можем записать математическое ожидание комплекснозначных случайных величин так:

$$\mathbf{E}\zeta = \mathbf{E}\xi_1 + i\mathbf{E}\xi_2 = \int_{\Omega} \zeta(\omega) \mathbf{P}(d\omega).$$

Интеграл Лебега справа определяется для комплекснозначных функций точно так же, как и для вещественных. Без ограничения общности мы можем полагать дискретные приближения для ξ_1 и ξ_2 (построенные ранее) заданными на одном конечном разбиении вероятностного пространства. Тогда интегральная сумма для комплексных случайных величин будет выглядеть так же, как и для вещественных:

$$\sum_{k \leq N} \zeta_k \mathbf{P}(A_k);$$

причем дискретное приближение $\sum_{k \leq N} \zeta_k I(A_k)$ с ростом N сходится равномерно по всем ω к случайной величине ζ . Отсюда, в частности, легко получить неравенство треугольника и для интеграла (с помощью соответствующего неравенства для сумм комплекснозначных величин и переходом к пределу):

$$|\mathbf{E}\zeta| \leq \mathbf{E}|\zeta|.$$

Введём теперь основное понятие. Прежде напомним определение экспоненты Эйлера $e^{ix} = \cos x + i \sin x$, а также ее основные свойства: $|e^{ix}| = 1 \quad \forall x \in \mathbb{R}$, и свойство мультипликативности $e^{i(x+y)} = e^{ix} e^{iy} \quad \forall x, y \in \mathbb{R}$.

Определение. *Характеристическая функция* распределения случайной величины — это функция вещественного аргумента, определённая на всей числовой прямой по формуле

$$\varphi_\xi(t) = \mathbf{E} e^{it\xi} = \int e^{itx} \mathcal{P}_\xi(dx) = \mathbf{E} \cos(t\xi) + i \mathbf{E} \sin(t\xi).$$

Свойства характеристической функции.

1. $|\varphi_\xi(t)| \leq 1, \varphi_\xi(0) = 1$.

Следует из неравенства треугольника для математического ожидания.

2. Характеристическая функция линейно преобразованной случайной величины со сдвигом имеет вид $\varphi_{c\xi+a}(t) = e^{ita} \varphi_\xi(ct)$.

Следует из мультипликативности экспоненты Эйлера.

3. Свойство мультипликативности при суммировании независимых наблюдений: *если случайные величины ξ_1 и ξ_2 независимы, то $\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t)$.*

Это свойство является следствием теоремы умножения: $\varphi_{\xi_1+\xi_2}(t) = \mathbf{E} e^{it\xi_1} e^{it\xi_2} = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t)$. Последнее равенство написано на основе того, что борелевские преобразования независимых случайных векторов будут независимыми. Легко видеть, что это свойство обобщается на любое конечное число слагаемых: *для любого конечного набора независимых случайных величин $\{\xi_i\}_{i=1}^n$ имеем*

$$\varphi_{\sum_{i=1}^n \xi_i}(t) = \prod_{i=1}^n \varphi_{\xi_i}(t).$$

4. *Если ξ — случайная величина с абсолютно непрерывным распределением, то $\lim_{t \rightarrow \infty} \varphi_\xi(t) = 0$.*

Это утверждение о поведении интегралов быстро осциллирующих функций составляет содержание классической теоремы Римана. По теореме Римана оба слагаемых в написанной ниже сумме стремятся к нулю при $t \rightarrow \infty$:

$$\varphi_\xi(t) = \int \cos(tx) p_\xi(x) dx + i \int \sin(tx) p_\xi(x) dx \rightarrow 0.$$

5. Рассмотрим частный случай дискретного распределения. Если для всех элементарных исходов $\xi \in \{a + hk; k \in \mathbb{Z}\}$, то говорят, что случайная величина ξ имеет *решетчатое распределение с шагом $h > 0$* (обычно предполагается, что h — максимальное возможное с указанным свойством) и *сдвигом a* , где, разумеется, $|a| < h$. Например, решетчатыми с параметрами $a = 0$ и $h = 1$ будут бернуллевская и пуассоновская случайные величины, а также случайная величина с биномиальным распределением. Вообще, если $a = 0$ и $h = m$, где m — натуральное, то решетчатое распределение называется *арифметическим*.

Радемахеровская случайная величина — арифметическая с шагом 2.

Теорема (критерий решетчатости). *Случайная величина ξ распределена решетчато в том и только в том случае, если $|\varphi_\xi(t)|$ – периодическая функция.*

ДОКАЗАТЕЛЬСТВО. (\rightarrow) Пусть $p_k = \mathbf{P}(\xi = a + hk)$. Тогда

$$|\varphi_\xi(t)| = \left| \sum_k e^{it(a+hk)} p_k \right| = \left| \sum_k e^{ita} e^{ithk} p_k \right| = |e^{ita}| \cdot \left| \sum_k e^{ithk} p_k \right| = \left| \sum_k e^{ithk} p_k \right|.$$

Последняя функция периодическая с периодом $2\pi/h$.

(\leftarrow) Так как $\varphi_\xi(0) = 1$, то найдется $T > 0$ такое, что $|\varphi_\xi(T)| = 1$, т. е. $\varphi_\xi(T) = e^{ia}$. Так как $\varphi_{\xi+c}(t) = e^{itc} \varphi_\xi(t)$, то при $c = -a$ получим $\varphi_{\xi-a}(T) = 1$, или, иными словами, $\mathbf{E} \cos((\xi - a)T) = 1$, откуда $\mathbf{E}(\cos((\xi - a)T) - 1) = 0$. И так как $\cos((\xi - a)T) - 1 \leq 0$, то равенство нулю среднего неотрицательной функции влечет за собой очевидное тождество $\cos((\xi - a)T) \stackrel{\text{п.п.}}{=} 1$. Поэтому $(\xi - a)T \in \{2\pi k; k \in \mathbb{Z}\}$ с вероятностью 1, откуда $\xi \in \{a + (2\pi/T)k; k \in \mathbb{Z}\}$. Тем самым мы установили связь периода с шагом решетки: $T = 2\pi/h$. \square

Заметим, что так как $\varphi_\xi(0) = 1$ и для решетчатых распределений модуль характеристической функции периодичен, то значение 1 в указанном случае принимается бесконечно много раз, и поэтому свойство 4 не выполняется (при $t \rightarrow \infty$ предела вообще может не быть).

Упражнение. Доказать, что для любого решетчатого распределения с нулевым сдвигом характеристическая функция периодическая.

Упражнение. Привести пример решетчатого распределения с ненулевым сдвигом, для которого характеристическая функция является периодической.

Упражнение. Привести пример решетчатого распределения, для которого характеристическая функция не является периодической.

6. Характеристическая функция $\varphi_\xi(\cdot)$ вещественнозначна тогда и только тогда, когда ξ симметрично распределена: $\xi \stackrel{d}{=} -\xi$, т. е. функции распределения случайных величин ξ и $-\xi$ совпадают.

ДОКАЗАТЕЛЬСТВО. Имеем:

$$\varphi_\xi(\cdot) \in \mathbb{R} \leftrightarrow \forall t \varphi_\xi(t) = \overline{\varphi_\xi(t)} = \mathbf{E} e^{-it\xi} = \mathbf{E} e^{it(-\xi)} = \varphi_{-\xi}(t).$$

Второе равенство написано на основе того, что операции сопряжения и интегрирования коммутируют и синус – нечетная функция.

Тогда если $\xi \stackrel{d}{=} -\xi$, то $\varphi_\xi(t) = \varphi_{-\xi}(t) = \overline{\varphi_\xi(t)}$, откуда $\varphi_\xi(\cdot) \in \mathbb{R}$.

Доказательство в обратную сторону будет следовать из теоремы о взаимно-однозначном соответствии между распределениями и их Фурье-образами, которая будет доказана чуть позже. \square

Проиллюстрировать это свойство можно на примере радемахеровской случайной величины

$$\xi = \begin{cases} 1, & 1/2, \\ -1, & 1/2, \end{cases}$$

или стандартного нормального распределения $p(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ (проверьте его симметричность). Вообще любое абсолютно непрерывное распределение с четной плотностью будет симметричным.

Упражнение. Доказать, что характеристическая функция любого симметричного распределения является чётной функцией.

Упражнение. Может ли характеристическая функция иметь нечетное число нулей?

7. Любая характеристическая функция равномерно непрерывна.

ДОКАЗАТЕЛЬСТВО. Нужно показать, что $\sup_{t \in \mathbb{R}} |\varphi_\xi(t + \Delta) - \varphi_\xi(t)| \xrightarrow{\Delta \rightarrow 0} 0$. Действительно,

$$\begin{aligned} \sup_{t \in \mathbb{R}} |\varphi_\xi(t + \Delta) - \varphi_\xi(t)| &= \sup_{t \in \mathbb{R}} \left| \int (e^{i(t+\Delta)x} - e^{itx}) dF_\xi(x) \right| = \sup_{t \in \mathbb{R}} \int |e^{itx}| \cdot |e^{ix\Delta} - 1| dF_\xi(x) = \\ &= \sup_{t \in \mathbb{R}} \int |e^{ix\Delta} - 1| dF_\xi(x) = \int |e^{ix\Delta} - 1| dF_\xi(x) = \int g_\Delta(x) dF_\xi(x) \xrightarrow{\Delta \rightarrow 0} 0 \end{aligned}$$

по теореме Лебега о мажорируемой сходимости (функция g_Δ ограничена числом 2, которое, разумеется, интегрируемо по распределению). \square

8. Связь моментов случайных величин с характеристическими функциями.

Если существует момент порядка k , то есть $\mathbf{E}|\xi|^k < \infty$ ($k \in \mathbb{N}$), то характеристическая функция k раз непрерывно дифференцируема: $\varphi_\xi(\cdot) \in C^k(\mathbb{R})$. Кроме того, $\varphi_\xi^{(m)}(0) = i^m \mathbf{E}\xi^m$, $m = 1, \dots, k$.

ДОКАЗАТЕЛЬСТВО. Внесем производную под знак интеграла

$$\varphi'_\xi(t) = \int ixe^{itx} dF_\xi(x).$$

Чтобы обосновать законность такого действия, воспользуемся теоремой Лебега о мажорируемой сходимости. Действительно, для модуля интегранта $|ixe^{itx}|$ существует интегрируемая мажоранта: $|ixe^{itx}| = |x|$ — интегрируемая по распределению функция, т. к. $\int |x| dF_\xi(x)$ — это первый абсолютный момент. Аналогично для $\varphi_\xi^{(m)}(t)$, $m = 1, \dots, k$ имеем: $|i^m x^m e^{itx}| = |x|^m$ и $\int |x|^m dF_\xi(x) < \infty$, т. к. существует $\mathbf{E}|\xi|^k$ (вспомним, что из существования старшего момента следует существование младших). По свойству 7 последняя k -ая производная будет непрерывна. \square

Следствие. В окрестности нуля для $\varphi_\xi(t)$ справедлива формула Тейлора:

$$\varphi_\xi(t) = 1 + it\mathbf{E}\xi - \frac{t^2}{2}\mathbf{E}\xi^2 + \dots + \frac{(it)^k \mathbf{E}\xi^k}{k!} + o(|t|^k).$$

Остаточный член можно записать в таком виде в силу непрерывности $\varphi_\xi^{(k)}(t)$.

Примеры характеристических функций.

1. Бернуллиевская случайная величина с вероятностью успеха p (дискретное решетчатое распределение). Характеристическая функция $\varphi_\xi(t) = pe^{it \cdot 1} + (1-p)e^{it \cdot 0} = 1 + p(e^{it} - 1)$.

2. Биномиальное распределение, как мы уже знаем, есть распределение случайной величины $S_n = \sum_{i=1}^n \xi_i$, где все ξ_i независимы и распределены по закону Бернулли с вероятностью успеха p . По свойству мультипликативности $\varphi_{S_n}(t) = \varphi_{\xi_1}^n(t) = (1 + p(e^{it} - 1))^n$.

3. Пуассоновское распределение с параметром λ . Характеристическая функция вычисляется следующим образом:

$$\varphi_{\pi_\lambda}(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} e^{-\lambda} = e^{\lambda e^{it}} \cdot e^{-\lambda} = e^{\lambda(e^{it}-1)}.$$

4. Стандартное нормальное распределение $N_{0,1}$. Характеристическая функция этого распределения по определению записывается как

$$\varphi_\xi(t) = \int e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Для вычисления этого интеграла составим дифференциальное уравнение. По теореме Лебега о мажорируемой сходимости (производная — это предел, причем $|e^{itx}| = 1$, а функция $|x|e^{-x^2/2}$ интегрируема на прямой) данный интеграл можно дифференцировать по параметру t . Имеем

$$\begin{aligned} \varphi'_\xi(t) &= \int ixe^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -\frac{i}{\sqrt{2\pi}} \int e^{itx} d(e^{-x^2/2}) = \\ &= [\text{далее интегрируем по частям}] = \\ &= -\frac{t}{\sqrt{2\pi}} \int e^{-x^2/2} e^{itx} dx = -t\varphi_\xi(t). \end{aligned}$$

Итак, мы пришли к задаче Коши: $\varphi'_\xi(t) = -t\varphi_\xi(t)$, $\varphi_\xi(0) = 1$, решая которую находим искомую характеристическую функцию: $\varphi_\xi(t) = e^{-t^2/2}$.

Рассмотрим произвольное гауссовское распределение. Мы уже знаем, что если ξ имеет распределение $N_{0,1}$, то линейное преобразование $\tilde{\xi} = \sigma\xi + \alpha$ имеет распределение $N_{\alpha,\sigma}$. По свойству 2 имеем $\varphi_{\tilde{\xi}}(t) = e^{it\alpha - \frac{\sigma^2 t^2}{2}}$. Для любого центрированного нормального распределения характеристическая функция имеет вид $e^{-ct^2/2}$, где $c > 0$.

Формулы обращения.

Теорема (основная формула обращения). Пусть $F_\xi(\cdot)$ — функция распределения некоторой случайной величины, и $\varphi_\xi(\cdot)$ — соответствующая характеристическая функция. Тогда для любых точек $x < z$, в которых функция F_ξ непрерывна, имеет место формула обращения:

$$F_\xi(z) - F_\xi(x) = \frac{1}{2\pi} \lim_{\sigma \rightarrow 0} \int \frac{e^{-itx} - e^{-itz}}{it} \varphi_\xi(t) e^{-\sigma^2 t^2/2} dt.$$

Отметим, что множитель $\varphi_\xi(t) e^{-\sigma^2 t^2/2}$ под интегралом — это характеристическая функция свертки распределения ξ с распределением $N_{0,\sigma}$.

Важнейшее следствие приведенной формулы — это *однозначность восстановления распределения по его характеристической функции*, поскольку любая монотонная функция на прямой с известным конечным пределом на ∞ (или $-\infty$) однозначно

восстанавливается по своим приращениям. Иными словами, отсюда следует *теорема о взаимно-однозначном соответствии между классами распределений и их характеристическими функциями*.

Следствие. Если характеристическая функция φ_ξ абсолютно интегрируема на всей числовой прямой ($|\varphi_\xi(\cdot)| \in L_1(\mathbb{R})$), то распределение ξ является абсолютно непрерывным с плотностью

$$p_\xi(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_\xi(t) dt.$$

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ. Прежде всего докажем вспомогательное утверждение, которое называется *тождеством Парсевалья*.

Лемма. Для любых случайных величин ξ и η справедливо тождество

$$\int dF_\eta(t) e^{-ity} \varphi_\xi(t) = \int \varphi_\eta(x - y) dF_\xi(x) \quad \forall y \in \mathbb{R}.$$

ДОКАЗАТЕЛЬСТВО. Домножим обе части равенства $\varphi_\xi(t) = \int e^{itx} dF_\xi(x)$ на e^{-ity} и проинтегрируем по переменной t по распределению случайной величины η :

$$\begin{aligned} \int dF_\eta(t) e^{-ity} \varphi_\xi(t) &= \int \left[\int e^{it(x-y)} dF_\xi(x) \right] dF_\eta(t) = \int dF_\xi(x) \left[\int e^{it(x-y)} dF_\eta(t) \right] = \\ &= \int \varphi_\eta(x - y) dF_\xi(x). \end{aligned}$$

Второе равенство здесь написано на основании теореме Фубини: интегралы можно переставить местами, так как все участвующие здесь меры конечны, а интегранты — ограниченные функции. \square

Далее, пусть ξ — произвольная случайная величина, а η — центрированная нормальная случайная величина с параметром $1/\sigma$. Тогда подставляя в тождество Парсевалья $\varphi_\eta(x - y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$, получаем

$$\sigma \int dt \frac{e^{-\sigma^2 t^2/2}}{\sqrt{2\pi}} e^{-ity} \varphi_\xi(t) = \int e^{-\frac{(x-y)^2}{2\sigma^2}} dF_\xi(x).$$

Отметим важную для дальнейшего деталь — характеристическая функция распределения величины η с точностью до постоянного множителя совпадает с плотностью другого нормального распределения $N(0, \sigma)$.

Разделив полученное выражение на $\sigma\sqrt{2\pi}$, получим:

$$\frac{1}{2\pi} \int e^{-\sigma^2 t^2/2} e^{-ity} \varphi_\xi(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int e^{-\frac{(x-y)^2}{2\sigma^2}} dF_\xi(x) = p_{\xi+\tilde{\eta}}(y). \quad (1)$$

Плотность $p_{\xi+\tilde{\eta}}$ появилась из формулы свертки. Напомним, что если ξ — произвольная случайная величина, а $\tilde{\eta}$ — случайная величина с абсолютно непрерывным распределением, и эти величины независимы, то и их сумма имеет плотность, которую можно найти по формуле свертки:

$$p_{\xi+\tilde{\eta}}(y) = \int p_{\tilde{\eta}}(y - x) dF_\xi(x).$$

В нашем случае роль случайной величины с абсолютно непрерывным распределением играет «сглаживающая» нормальная случайная величина $\tilde{\eta}$ с параметрами 0 и σ .

От обеих частей тождества (1) возьмем интеграл по конечному отрезку $[x, z]$, при этом снова воспользуемся теоремой Фубини, а также примем во внимание соотношение

$$\int_x^z e^{-ity} dy = \frac{e^{-itx} - e^{-itz}}{it}.$$

Имеем

$$\frac{1}{2\pi} \int dt e^{-\sigma^2 t^2/2} \int_x^z e^{-ity} \varphi_\xi(t) dy = \int_x^z p_{\xi+\tilde{\eta}}(y) dy = F_{\xi+\tilde{\eta}}(z) - F_{\xi+\tilde{\eta}}(x). \quad (2)$$

Перейдем к пределу при $\sigma \rightarrow 0$. В левой части равенства (2) получим выражение, присутствующее в формулировке теоремы:

$$\frac{1}{2\pi} \lim_{\sigma \rightarrow 0} \int \frac{e^{-itx} - e^{-itz}}{it} \varphi_\xi(t) e^{-\sigma^2 t^2/2} dt.$$

Осталось показать, что правая часть (2) будет сходиться к соответствующей разности значений функции распределения F_ξ при условии, что x и z — ее точки непрерывности.

Согласно неравенству Чебышева $\mathbf{P}(|\zeta - \mathbf{E}\zeta| > \varepsilon) \leq \varepsilon^{-2} \mathbf{D}\zeta$. Значит,

$$\mathbf{P}(|\tilde{\eta} - \mathbf{E}\tilde{\eta}| > \varepsilon) = \mathbf{P}(|\tilde{\eta}| > \varepsilon) \leq \frac{\mathbf{D}\tilde{\eta}}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

Но ε фиксировано и $\sigma \rightarrow 0$, поэтому $\mathbf{P}(|\tilde{\eta}| > \varepsilon) \rightarrow 0$. Таким образом, $\xi + \tilde{\eta} \xrightarrow{p} \xi$. Следовательно, имеет место и слабая сходимост распределений указанных случайных величин, т. е. $F_{\xi+\tilde{\eta}}(t) \rightarrow F_\xi(t)$, где t — любая точка непрерывности функции F_ξ . Следовательно, $F_{\xi+\tilde{\eta}}(z) - F_{\xi+\tilde{\eta}}(x) \rightarrow F_\xi(z) - F_\xi(x)$ и теорема доказана. \square

ДОКАЗАТЕЛЬСТВО СЛЕДСТВИЯ. Проинтегрируем тождество (1) по y на отрезке $[x, z]$. Тогда если $|\varphi_\xi| \in L_1(\mathbb{R})$, то по теореме Лебега

$$\frac{1}{2\pi} \lim_{\sigma \rightarrow 0} \int_x^z dy \int_{\mathbb{R}} e^{-ity} \varphi_\xi(t) e^{-\sigma^2 t^2/2} dt =$$

[для интегранта существует интегрируемая мажоранта $|\varphi_\xi(t) e^{-\sigma^2 t^2/2}| \leq |\varphi_\xi(t)|$, не зависящая от параметра σ . Поэтому предел можно внести под знак двойного интеграла]

$$= \frac{1}{2\pi} \int_x^z dy \int_{\mathbb{R}} e^{-ity} \varphi_\xi(t) dt.$$

Иными словами, для любых точек непрерывности x и z функции распределения F_ξ мы получили представление для разности:

$$F_\xi(z) - F_\xi(x) = \int_x^z p_\xi(y) dy,$$

где

$$p_{\xi}(y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ity} \varphi_{\xi}(t) dt.$$

Это равенство, очевидно, можно распространить и на все точки вещественной прямой, так как множество точек непрерывности любой монотонной функции всюду плотно в \mathbb{R} , а любая функция распределения непрерывна слева. Это и будет означать, что $p_{\xi}(y)$ — плотность. Следствие доказано. \square

Упражнение. Доказать, что для интегрируемой характеристической функции соответствующая плотность распределения будет равномерно непрерывной.

Упражнение. Доказать, что если некоторая характеристическая функция неотрицательна и интегрируема на всей прямой, то для соответствующей плотности распределения $p_{\xi}(t)$ выполнено $p_{\xi}(0) > 0$ и отношение $p_{\xi}(t)/p_{\xi}(0)$ также является характеристической функцией.

Упражнение. Привести пример характеристической функции с ограниченным носителем.

Формула обращения для решетчатых распределений.

Пусть $\xi \in \{a + hk; k \in \mathbb{Z}\}$. Ясно, что $\xi = \xi_0 h + a$, где ξ_0 имеет арифметическое распределение ($a = 0, h = 1$). Связь характеристических функций при линейном преобразовании случайных величин известна (свойство 2), так что восстанавливая распределение ξ_0 , мы восстанавливаем распределение ξ . Итак,

$$\varphi_{\xi_0}(t) = \sum_{k \in \mathbb{Z}} e^{itk} p_k.$$

Умножим обе части равенства на e^{-itm} ($m \in \mathbb{Z}$) и проинтегрируем по t в промежутке от $-\pi$ до π :

$$\int_{-\pi}^{\pi} e^{-itm} \varphi_{\xi_0}(t) dt = \sum_{k \in \mathbb{Z}} \int_{-\pi}^{\pi} e^{it(k-m)} p_k dt = 2\pi p_m$$

(поскольку при $k \neq m$ мы имеем $\int_{-\pi}^{\pi} e^{-it(k-m)} dt = 0$). Следовательно,

$$p_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itm} \varphi_{\xi_0}(t) dt. \quad \square$$

Теорема непрерывности.

Основной результат главы — теорема о непрерывном взаимно-однозначном соответствии (гомеоморфизме), называемая часто для краткости *теоремой непрерывности*.

Теорема. Пространство всех функций распределений с топологией слабой сходимости гомеоморфно пространству всех характеристических функций с топологией поточечной сходимости:

$$\{F, \Rightarrow\} \leftrightarrow \{\varphi, \rightarrow\}.$$

ДОКАЗАТЕЛЬСТВО. Вопрос о взаимно-однозначном соответствии рассматриваемых двух классов уже был решен с помощью основной формулы обращения. Нам остается только показать, что указанное отображение будет непрерывным в соответствующих топологиях.

(\rightarrow) По критерию слабой сходимости $\xi_n \Rightarrow \xi$ тогда и только тогда, когда $\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi)$ для любой непрерывной ограниченной функции f . Нам же требуется доказать, что $\int e^{itx} dF_{\xi_n}(x) \rightarrow \int e^{itx} dF_{\xi}(x)$. Отметим, что для каждого фиксированного t мы здесь как раз и имеем дело с интегралами от ограниченных непрерывных функций: $\cos(tx)$ и $\sin(tx)$. Следовательно, необходимость доказана.

(\leftarrow) Имеем: $\varphi_{\xi_n}(t) \rightarrow \varphi_{\xi}(t) \forall t$. Требуется установить слабую сходимость соответствующих распределений. В процессе доказательства основной формулы обращения было доказано, что если η — нормальная случайная величина с параметрами $(0, \sigma)$, а ξ_n — произвольная, причем ξ_n и η независимы, то

$$F_{\xi_n+\eta}(x) - F_{\xi_n+\eta}(y) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ity} - e^{-itx}}{it} \varphi_{\xi_n}(t) e^{-\frac{\sigma^2 t^2}{2}} dt. \quad (1)$$

Аналогичное представление имеет место и при замене ξ_n на ξ .

Заметим, что $|\varphi_{\xi_n}(t) e^{-\frac{\sigma^2 t^2}{2}}| \leq e^{-\frac{\sigma^2 t^2}{2}}$ и, кроме того,

$$\left| \frac{e^{-itx} - e^{-ity}}{it} \right| \leq \begin{cases} |x - y|, & |t| < 1, \\ 2/|t|, & |t| \geq 1, \end{cases}$$

так как $|e^{i\varphi} - e^{i\phi}| \leq |\varphi - \phi|$. По теореме Лебега о мажорируемой сходимости предел по n можно внести под знак интеграла в (1). Тогда для всех $x, y \in \mathbb{R}$

$$F_{\xi_n+\eta}(x) - F_{\xi_n+\eta}(y) \rightarrow F_{\xi+\eta}(x) - F_{\xi+\eta}(y).$$

Поскольку мы можем взять в этом тождестве y сколь угодно близким к $-\infty$, то тем самым нами доказано, что для любого $x \in \mathbb{R}$

$$F_{\xi_n+\eta}(x) \rightarrow F_{\xi+\eta}(x).$$

Теперь отсюда надо извлечь требуемую слабую сходимость. Имеем

$$F_{\xi_n+\eta}(x) = \mathbf{P}(\xi_n + \eta < x) \leq$$

[воспользуемся формулой полной вероятности: $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap \bar{B})$, где $A = \{\xi_n + \eta < x\}$, $B = \{|\eta| \leq \varepsilon\}$, $\bar{B} = \{|\eta| > \varepsilon\} \Rightarrow \mathbf{P}(A \cap B) \leq \mathbf{P}(A) \leq \mathbf{P}(A \cap B) + \mathbf{P}(\bar{B})$]

$$\leq F_{\xi_n}(x + \varepsilon) + \mathbf{P}(|\eta| > \varepsilon) \leq$$

[по неравенству Чебышева]

$$\leq F_{\xi_n}(x + \varepsilon) + \frac{\sigma^2}{\varepsilon^2}.$$

Так как σ, ε — свободные независимые параметры, то σ можно выбрать таким, что σ^2/ε^2 будет сколь угодно малой величиной. С другой стороны,

$$\begin{aligned}
F_{\xi_n+\eta}(x) &\geq P(\xi_n + \eta < x, |\eta| \leq \varepsilon) \geq P(\xi_n < x - \varepsilon, |\eta| \leq \varepsilon) \\
&\geq F_{\xi_n}(x - \varepsilon) - P(|\eta| > \varepsilon) \geq F_{\xi_n}(x - \varepsilon) - \frac{\sigma^2}{\varepsilon^2}.
\end{aligned}$$

Таким образом,

$$F_{\xi_n}(x - \varepsilon) - \frac{\sigma^2}{\varepsilon^2} \leq F_{\xi_n+\eta}(x) \leq F_{\xi_n}(x + \varepsilon) + \frac{\sigma^2}{\varepsilon^2}.$$

Сделав преобразование $x - \varepsilon = \tilde{x}$, $x = \tilde{x} + \varepsilon$ (x — произвольная точка на прямой), мы «обратим» полученное двойное неравенство:

$$F_{\xi_n+\eta}(x - \varepsilon) - \frac{\sigma^2}{\varepsilon^2} \leq F_{\xi_n}(x) \leq F_{\xi_n+\eta}(x + \varepsilon) + \frac{\sigma^2}{\varepsilon^2}.$$

Так как $F_{\xi_n+\eta}(x - \varepsilon) \xrightarrow{n \rightarrow \infty} F_{\xi+\eta}(x - \varepsilon)$ и $F_{\xi_n+\eta}(x + \varepsilon) \xrightarrow{n \rightarrow \infty} F_{\xi+\eta}(x + \varepsilon)$, и для ξ имеем совершенно аналогичное двустороннее неравенство (заменой в предыдущем неравенстве ξ_n на ξ):

$$F_{\xi+\eta}(x - \varepsilon) - \frac{\sigma^2}{\varepsilon^2} \leq F_{\xi}(x) \leq F_{\xi+\eta}(x + \varepsilon) + \frac{\sigma^2}{\varepsilon^2},$$

то, выбирая в качестве x произвольную точку непрерывности функции $F_{\xi}(\cdot)$, избавляемся от сглаживающей компоненты, используя только что приведенные аргументы:

$$F_{\xi}(x - 2\varepsilon) - \frac{2\sigma^2}{\varepsilon^2} \leq F_{\xi+\eta}(x - \varepsilon) - \frac{\sigma^2}{\varepsilon^2} \leq F_{\xi}(x) \leq F_{\xi+\eta}(x + \varepsilon) + \frac{\sigma^2}{\varepsilon^2} \leq F_{\xi}(x + 2\varepsilon) + \frac{2\sigma^2}{\varepsilon^2}.$$

Полагая $\sigma^2 = \varepsilon^3$ и устремляя ε к 0, по «принципу двух милиционеров» получаем из вышеприведенных неравенств сходимость $F_{\xi_n}(x) \rightarrow F_{\xi}(x)$ в точках непрерывности функции F_{ξ} . \square

З а м е ч а н и е. Чаще всего в приложениях используется импликация \leftarrow приведенной теоремы непрерывности, что в полной мере проиллюстрировано в следующей главе. Исследование свойств распределений (в частности, слабой сходимости) с помощью соответствующих характеристических функций представляет собой так называемый *метод характеристических функций*.

Упражнение. Доказать замкнутость классов нормальных и пуассоновских распределений относительно операции «свёртка», используя метод характеристических функций.

Это так называемое свойство *устойчивости* нормальных и пуассоновских распределений.

Глава 7. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

В этой главе будет продемонстрирован *метод характеристических функций* при доказательстве различных предельных теорем. Наиболее впечатляющее в своей общности утверждение — это *центральная предельная теорема*, частным случаем которой является теорема Муавра — Лапласа. Достаточно сравнить объемы доказательств этих двух утверждений и их общность, чтобы по достоинству оценить предлагаемый новый метод доказательства, основанный на теореме непрерывности.

Центральная предельная теорема (ЦПТ).

Теорема. Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины с конечной дисперсией. Тогда

$$\frac{S_n - n\mathbf{E}\xi_1}{\sqrt{n\mathbf{D}\xi_1}} \Rightarrow \eta \in N(0, 1), \quad (1)$$

где $S_n = \sum_{i=1}^n \xi_i$.

В первую очередь, нас привлекает «собираемость» этой теоремы. Заметим, что теорема Муавра—Лапласа — это частный случай ЦПТ для одного и того же бернуллиевского распределения слагаемых:

$$\xi_i = \begin{cases} 1, & p, \\ 0, & 1 - p. \end{cases}$$

ДОКАЗАТЕЛЬСТВО. Введем новые центрированные и нормированные случайные величины $\tilde{\xi}_i = (\xi_i - \mathbf{E}\xi_1)/\sqrt{\mathbf{D}\xi_1}$. Тогда

$$\frac{S_n - n\mathbf{E}\xi_1}{\sqrt{n\mathbf{D}\xi_1}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\xi}_i$$

[так как $\mathbf{E}\tilde{\xi}_i = 0$, $\mathbf{D}\tilde{\xi}_i = 1 = \mathbf{E}\tilde{\xi}_i^2$]. Мы хотим доказать, что

$$\varphi_{\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\xi}_i}(t)$$

в пределе имеет характеристическую функцию гауссовской случайной величины, а затем воспользоваться теоремой непрерывности. Имеем

$$\varphi_{\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\xi}_i}(t) = \varphi_{\sum_{i=1}^n \tilde{\xi}_i}(t/\sqrt{n}) =$$

[по свойству 3)]

$$= \left(\varphi_{\tilde{\xi}_1} \left(\frac{t}{\sqrt{n}} \right) \right)^n =$$

[по свойству 8)]

$$= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{-\frac{t^2}{2}}$$

[t фиксировано, $n \rightarrow \infty$]. Предел есть характеристическая функция стандартного нормального закона. \square

Пример. Игральная кость бросается 10^4 раз (здесь n велико!). Здесь ξ_i имеют равномерное распределение на первых шести натуральных числах. Тогда для функции распределения нормированной суммы S_n вместо предела в (1) можно писать приближенное равенство с параметрами $n = 10^4$, $\mathbf{E}\xi_i = 3,5$ и $\mathbf{D}\xi_i = \mathbf{E}\xi_i^2 - (\mathbf{E}\xi_i)^2 = 91/6 - (7/2)^2 = 35/12$.

Упражнение. Каково отклонение суммарного количества S_n набранных очков от 35000 с вероятностью 0,99? (полагаем $S_n = 35000 \pm x$ и оцениваем минимальное x .)

Следствие («Закон трех сигм»). При любом фиксированном $x > 0$ и достаточно больших n

$$\mathbf{P}\left(-x < \frac{S_n - n\mathbf{E}\xi_i}{\sqrt{n\mathbf{D}\xi_i}} < x\right) \approx \Phi(x) - \Phi(-x) = 1 - 2\Phi(-x).$$

При $x = 3$ имеем $1 - 2\Phi(-x) \approx 0,997$, то есть при больших n почти наверняка будет выполнено двустороннее неравенство

$$-3 < \frac{S_n - n\mathbf{E}\xi_1}{\sqrt{n\mathbf{D}\xi_1}} < 3.$$

Обозначив $\sigma^2 = \mathbf{D}\xi_1$, получим, что $n\mathbf{E}\xi_1 - 3\sigma\sqrt{n} \leq S_n \leq n\mathbf{E}\xi_1 + 3\sigma\sqrt{n}$ с вероятностью почти 1. Однако надо иметь в виду погрешность приближения в ЦПТ. Грубо говоря, эта погрешность имеет такой же порядок, как и в теореме Муавра – Лапласа, т. е. $O(n^{-1/2})$.

Закон больших чисел и обобщенная теорема Пуассона.

Сначала с помощью теоремы непрерывности мы приведем новое доказательство ЗБЧ в форме Хинчина. Напомним этот результат.

Теорема (ЗБЧ в форме Хинчина). Пусть ξ_1, ξ_2, \dots – независимые одинаково распределенные и $\mathbf{E}|\xi_1| < \infty$. Тогда

$$\frac{1}{n}S_n \xrightarrow{p} \mathbf{E}\xi_1.$$

Мы приведем новое не менее простое доказательство этой формы ЗБЧ, применяя метод характеристических функций.

ДОКАЗАТЕЛЬСТВО. Имеем

$$\begin{aligned} \varphi_{\frac{1}{n}S_n}(t) &= \varphi_{S_n}\left(\frac{t}{n}\right) = \\ &= \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{it}{n}\mathbf{E}\xi_1 + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{it\mathbf{E}\xi_1} = \varphi_{\mathbf{E}\xi_1}(t). \end{aligned}$$

Тогда по теореме непрерывности

$$\frac{1}{n}S_n \Rightarrow \mathbf{E}\xi_1.$$

Но если предельная величина вырождена, то слабая сходимость эквивалентна сходимости по вероятности (доказано ранее). Следовательно,

$$\frac{1}{n}S_n \xrightarrow{p} \mathbf{E}\xi_1. \quad \square$$

Теперь мы докажем более общий вариант теоремы Пуассона. Напомним, что классическая теорема Пуассона была посвящена предельному поведению распределения числа успехов в серии из n независимых испытаний в схеме Бернулли с вероятностью успеха p , причем предполагалось, что $np \rightarrow \lambda$ (двойной предельный переход по направлению). Рассмотрим несколько более общую схему стохастических испытаний. Пусть имеются независимые одинаково распределенные случайные величины $\{\xi_i\}$ с арифметическим распределением

$$\xi_i = \begin{cases} 0, & p_0 \\ 1, & p_1 \\ 2, & p_2, \\ \cdot \\ \cdot \\ \cdot \end{cases} \quad \text{при этом предполагается, что } np_k \rightarrow \lambda_k \text{ для всех } k \geq 1 \\ \text{и что } \sum_{k \geq 1} \lambda_k < \infty.$$

Иначе говоря, здесь мы рассматриваем предел относительно бесконечномерного параметра (n, p_1, p_2, \dots) по направлению, задаваемому счетным набором соотношений $np_k \rightarrow \lambda_k, k = 1, 2, \dots$

Обозначим

$$S_n = \sum_{i=1}^n \xi_i, \quad \zeta = \sum_{k=1}^{\infty} k\pi_{\lambda_k},$$

где $\{\pi_{\lambda_k}\}$ — независимые пуассоновские случайные величины с параметрами $\{\lambda_k\}$ соответственно. Заметим, что $\zeta < \infty$ с вероятностью 1. Это следует из теоремы Бореля—Кантелли, поскольку

$$\sum_{k=1}^{\infty} \mathbf{P}(\pi_{\lambda_k} > 0) = \sum_{k=1}^{\infty} (1 - e^{-\lambda_k}) < \sum_{k=1}^{\infty} \lambda_k < \infty.$$

Это означает, что с вероятностью 1 в последовательности $\{\pi_{\lambda_k}\}$ лишь конечный набор случайных величин отличен от нуля. Распределение случайной величины ζ принадлежит важному классу так называемых *обобщенных пуассоновских распределений*. Ясно, что если $p_k = 0 = \lambda_k$ при всех $k \geq 2$, то $\zeta = \pi_{\lambda_1}$ — обычная пуассоновская случайная величина.

Теорема (обобщенная интегральная теорема Пуассона). *Для описанной выше схемы имеет место предельное соотношение*

$$\sup_{A \subset \mathbb{Z}_+} |\mathbf{P}(S_n \in A) - \mathbf{P}(\zeta \in A)| \rightarrow 0.$$

ДОКАЗАТЕЛЬСТВО. Имеем

$$\varphi_{S_n}(t) = (\varphi_\xi(t))^n =$$

[характеристическая функция решетчатой случайной величины с нулевым сдвигом]

$$\begin{aligned} &= \left(\sum_{k \geq 0} e^{itk} p_k \right)^n = \left(p_0 + \sum_{k \geq 1} e^{itk} p_k \right)^n \\ &= \left(1 + \sum_{k \geq 1} p_k (e^{itk} - 1) \right)^n \end{aligned}$$

$$\begin{aligned} &[\text{воспользуемся соотношением } (1+z)^n = \exp\{n \log(1+z)\} = \exp\{nz + O(nz^2)\}] \\ &= \exp \left\{ \sum_{k \geq 1} np_k (e^{itk} - 1) + O \left(n^{-1} \left(n \sum_{k \geq 1} p_k \right)^2 \right) \right\}. \end{aligned}$$

В силу леммы Фату $\limsup n \sum_{k \geq 1} p_k \leq \sum_{k \geq 1} \lim np_k = \sum_{k \geq 1} \lambda_k < \infty$. Поэтому O -символ в правой части приведенного асимптотического равенства можно записать как $O(1/n)$.

Далее, для любого натурального N имеем

$$\sum_{k \leq N} np_k (e^{itk} - 1) \rightarrow \sum_{k \leq N} \lambda_k (e^{itk} - 1).$$

Хвост ряда опять же с помощью леммы Фату легко оценим как

$$\limsup \left| \sum_{k > N} np_k (e^{itk} - 1) \right| \leq 2 \sum_{k > N} \lambda_k,$$

что может быть сделано сколь угодно малым выбором достаточно большого N . Таким образом, в условиях теоремы

$$\varphi_{S_n}(t) \rightarrow \exp \left\{ \sum_{k \geq 1} \lambda_k (e^{itk} - 1) \right\}.$$

Упражнение. Доказать, что правая часть последнего соотношения представляет собой характеристическую функцию случайной величины ζ .

Это означает, что любая точечная масса $\mathbf{P}(S_n = m)$ сходится к соответствующей обобщенной пуассоновской массе, т. е. имеет место локальная предельная теорема, которая, как мы уже знаем, влечет за собой интегральную. \square

ЛИТЕРАТУРА

1. Боровков А. А. *Теория вероятностей*.— М: Эдиториал УРСС, 1999.
4. Лозв М. *Теория вероятностей*.— М.: Иностранная литература, 1962.
2. Розанов Ю. А. *Введение в теорию случайных процессов*.— М.: Наука, 1982.
3. Феллер В. *Введение в теорию вероятностей и ее приложения*, т. I, II. — М.: Мир, 1984.