Что было, а также чего не было, но что вполне могло бы быть

прочитано в курсе лекций под названием

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1 курс ЭФ, отделение
«математические методы и исследование операций в экономике»

весенний семестр 1998-99 уч. года

Чернова Н.И.

— Знаете что, милый Арамис? — сказал д'Артаньян, ненавидевший стихи почти так же сильно, как латынь. — Добавьте к достоинству трудности достоинство краткости, и вы сможете быть уверены в том, что ваша поэма будет иметь никак не менее двух достоинств.

Раздел 1. Классическая вероятностная схема

1.1 Основные формулы комбинаторики

В данном разделе мы займемся подсчетом числа «шансов». О числе шансов говорят, когда возможно несколько различных результатов какого-либо действия (извлечение карты из колоды, подбрасывание кубика или монетки, двух кубиков и т.д.). Число шансов — это число таких возможных результатов, или, иначе говоря, число способов проделать это действие.

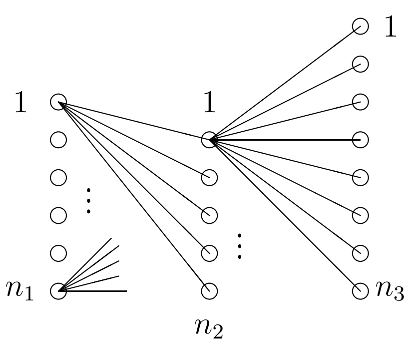
Теорема о перемножении шансов

Теорема 1. *Пусть имеется k, $k \in \mathbb{N}$, групп элементов, причем i-я группа содержит n_i элементов, $1 \leqslant i \leqslant k$. Выберем из каждой группы по одному элементу. Тогда общее число N способов, которыми можно произвести такой выбор, равняется*

$$N = n_1 \cdot n_2 \cdot \ldots \cdot n_k.$$

Замечание 1. В теореме **1** считается, что даже если все элементы в i-й группе неразличимы, выбрать один из них можно n_i способами. Представим *результат* выбора, описанного в теореме **1**, в виде набора $(a_1, \ldots, a_k)$, в котором a_i — выбранный из i-й группы элемент. Тогда общее число различных наборов $(a_1, \ldots, a_k)$ также равняется $N = n_1 \cdot n_2 \cdot \ldots \cdot n_k$.

Доказательство теоремы 1.



Занумеруем элементы i-й группы числами от 1 до n_i. Элемент из первой группы можно выбрать n_1 способами. Если мы выбрали элемент j, $1 \leqslant j \leqslant n_1$, то выбрать элемент из второй группы мы можем n_2 способами. Получаем, что с первым элементом j возможно составить n_2 пар (j, l), где $1 \leqslant l \leqslant n_2$.
Но столько же пар можно составить и с любым другим элементом первой группы. Тогда всего пар, в которых первый элемент выбран из первой группы, а второй — из второй, существует ровно $n_1 \cdot n_2$. Иначе говоря, есть $n_1 \cdot n_2$ способов выбрать по одному элементу из первых двух групп.

Возьмем одну такую пару (j, l). Заметим, что элемент из третьей группы можно выбрать n_3 способами, то есть возможно составить ровно n_3 троек (j, l, m), добавляя к данной паре (j, l) любой из n_3 элементов третьей группы.

Но столько же троек можно составить и с любой другой парой (j, l). Тогда всего троек, в которых первый элемент выбран из первой группы, второй — из второй, а третий — из третьей, существует ровно $n_1 \cdot n_2 \cdot n_3$.

Продолжая рассуждения, методом математической индукции заключаем справедливость утверждения теоремы. □

Упражнение 1. Сформулировать предположение индукции и доказать индукционный переход от $k-1$ к k.

Урны и шарики

Есть урна (то есть ящик), содержащая n занумерованных объектов, которые мы без ограничения общности будем считать шариками. Мы выбираем из этой урны k шариков. Нас интересует, сколькими способами можно выбрать k шариков из n, или *сколько различных результатов* (то есть наборов, состоящих из k шариков) получится.

На этот вопрос нельзя дать однозначный ответ, пока мы не определимся
а) с тем, как организован выбор (скажем, можно ли шарики возвращать в урну), и
б) с тем, что понимается под *различными* результатами выбора.

Рассмотрим следующие возможные схемы выбора:

1. Выбор с возвращением: каждый выбранный шарик возвращается в урну, то есть каждый из k шариков выбирается из полной урны. В полученном наборе, состоящем из k номеров шариков, могут встречаться одни и те же номера (*выборка с повторениями*).

2. Выбор без возвращения: выбранные шарики в урну *не* возвращаются, и в полученном наборе *не* могут встречаться одни и те же номера (*выборка без повторений*).

И в том, и в другом случае результатом выбора является набор из k номеров шариков. Удобно считать, что шарики всегда выбираются последовательно, по одному (с возвращением или без). Условимся, какие результаты мы будем считать *различными*. Есть ровно две возможности.

1. Выбор с учетом порядка: два набора номеров шариков считаются различными, если они отличаются составом или порядком номеров. Так, при выборе трех шариков из урны, содержащей 5 шариков, наборы $(1, 5, 2)$, $(2, 5, 1)$ и $(4, 4, 5)$ различны, если производится *выбор с учетом порядка*.

2. Выбор без учета порядка: два набора номеров шариков считаются различными, если они отличаются составом. Наборы, отличающиеся лишь порядком следования номеров, считаются одинаковыми. Так, в примере выше первые два набора $(1, 5, 2)$ и $(2, 5, 1)$ есть один и тот же результат выбора, а набор $(4, 4, 5)$ — другой результат выбора.

Подсчитаем теперь, сколько же возможно различных результатов при каждой из четырех схем (выбор с возвращением и без, и в каждом из этих случаев учитываем ли мы порядок или нет).

Урновая схема: выбор без возвращения, с учетом порядка

Теорема 2. *Общее количество выборок в схеме выбора k элементов из n без возвращения и с учетом порядка определяется формулой*

$$A_n^k = \underbrace{n(n-1)\cdot\ldots\cdot(n-k+1)}_{k} = \frac{n!}{(n-k)!}$$

*и называется **числом размещений из n элементов по k элементов**.*

Доказательство. Первый шарик можно выбрать n способами. При каждом из этих способов второй шарик можно выбрать $n-1$ способом, и т.д. Последний k-й шарик можно выбрать $n-k+1$ способом. По теореме **1**, общее число способов выбора равно $n\cdot(n-1)\cdot\ldots\cdot(n-k+1)$, что и требовалось доказать. □

Следствие 1. *Число возможных перестановок множества из n элементов есть $n!$*

Доказательство очевидно, если заметить, что перестановка есть не что иное, как результат выбора без возвращения и с учетом порядка всех n элементов из n. Так что общее число перестановок равно $A_n^n = n!$ □

Урновая схема: выбор без возвращения и без учета порядка

Теорема 3. *Общее количество выборок в схеме выбора k элементов из n без возвращения и без учета порядка определяется формулой*

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

*и называется **числом сочетаний из n элементов по k элементов**.*

Доказательство. Заметим, что, согласно следствию **1**, из каждой выборки данного состава (состоящей из k элементов) можно образовать $k!$ выборок, отличающихся друг от друга только порядком элементов.

То есть число выборок, различающихся еще и порядком, в $k!$ раз больше, чем число выборок, различающихся только составом. Поделив A_n^k на $k!$, получим утверждение теоремы.

□

Урновая схема: выбор с возвращением и с учетом порядка

Теорема 4. *Общее количество выборок в схеме выбора k элементов из n с возвращением и с учетом порядка определяется формулой*

$$n^k = \underbrace{n \cdot n \cdot \ldots \cdot n}_{k}.$$

Доказательство. Первый шарик можно выбрать n способами. При каждом из этих способов второй шарик можно выбрать также n способами, и так k раз. □

Урновая схема: выбор с возвращением и без учета порядка

Рассмотрим урну с двумя шариками и перечислим результаты выбора двух шариков из этой урны при выборе с возвращением:

<table>
<tr><th>с учетом порядка</th><th>без учета порядка</th></tr>
<tr><td>(1,1)</td><td>(1,1)</td></tr>
<tr><td>(2,2)</td><td>(2,2)</td></tr>
<tr><td>(1,2)</td><td rowspan="2">(1,2)</td></tr>
<tr><td>(2,1)</td></tr>
</table>

Заметим, что в схеме «без учета порядка» получилось 3 различных результата в отличие от четырех в схеме «с учетом порядка» (число 4 возникает и согласно теореме **4**); и что никаким делением на «число каких-нибудь перестановок» число 3 из 4 получить не удастся.

Теорема 5. *Общее количество выборок в схеме выбора k элементов из n с возвращением и без учета порядка определяется формулой*

$$C^k_{n+k-1} = C^{n-1}_{n+k-1}.$$

Упражнение 2. Проверить, что при $n = 2$ и $k = 2$ получается ровно 3.

Доказательство. Рассмотрим подробно, чем отличаются друг от друга два разных результата такой схемы выбора.

Нам не важен порядок номеров, то есть мы учитываем только, сколько раз в нашем наборе из k номеров шариков появился шарик номер 1, шарик номер 2, … , шарик номер n. То есть результат выбора можно представить набором чисел $k_1, k_2, \ldots, k_n$, в котором k_i — число появлений шарика номер i в выборке, и $k_1 + \ldots + k_n = k$. Числа k_i принимают значения из $\mathbb{N} \cup \{0\}$. При этом два результата эксперимента различны, если соответствующие им наборы $k_1, k_2, \ldots, k_n$ не совпадают (при этом учитывается и порядок элементов).

Представим себе другой эксперимент, имеющий точно такие же результаты (и, следовательно, их столько же). Есть n ящиков, в которых размещается k шариков. Нас интересует только *количество шариков в каждом ящике*.

То есть результатом эксперимента снова является набор чисел $k_1, k_2, \ldots, k_n$, в котором k_i — число шариков в ящике с номером i, и $k_1 + \ldots + k_n = k$. Числа k_i по-прежнему принимают натуральные значения или равны 0.

А теперь изобразим результат такого размещения в виде схемы, в которой вертикальные линии обозначают перегородки между ящиками, а кружки — находящиеся в ящиках шарики:

$$\Big|\ \bullet\ \bullet\ \bullet\ \Big|\quad\Big|\ \bullet\ \Big|\ \bullet\ \bullet\ \Big|\ \bullet\ \bullet\ \Big|\quad\Big|\ \bullet\ \Big|$$

Мы видим результат размещения 9 шариков по 7 ящикам. Здесь 1-й ящик содержит 3 шарика, 2-й и 6-й ящики пусты, 3-й ящик содержит 1 шарик, и в 4-м и 5-м ящиках есть по 2 шарика. Переложим один шарик из первого ящика во второй и изобразим таким же образом еще один результат размещения:

$$\Big|\ \bullet\ \bullet\ \Big|\ \bullet\ \Big|\ \bullet\ \Big|\ \bullet\ \bullet\ \Big|\ \bullet\ \bullet\ \Big|\quad\Big|\ \bullet\ \Big|\,. \qquad \text{И еще один:} \qquad \Big|\,\Big|\,\Big|\,\Big|\,\Big|\,\Big|\,\Big|\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \Big|$$

Видим, что все размещения можно получить, меняя между собой шарики и перегородки, или расставляя k шариков на $n-1+k$ месте. Число $n-1+k$ получается так: у n ящиков есть ровно $n+1$ перегородка, считая крайние, или $n-1$ перегородка, если не считать крайние, которые двигать нельзя. И есть k шариков. Перебрав все возможные способы расставить k шариков на этих $n-1+k$ местах (и ставя на оставшиеся места перегородки), переберем все нужные размещения.

Но способов расставить k шариков на $n-1+k$ местах ровно C^k_{n-1+k} — это в точности число способов выбрать из $n-1+k$ номеров мест k номеров мест (без учета порядка и без возвращения), на которые нужно поместить шарики. Заметим, что равенство $C^k_{n+k-1} = C^{n-1}_{n+k-1}$ верно как по определению биномиальных коэффициентов или свойствам треугольника Паскаля, так и в силу того, что можно *вместо* выбора k мест для шариков выбирать $n-1$ место для перегородок ящиков, заполняя шариками оставшиеся места.

□

1.2 Основные понятия элементарной теории вероятностей

Предмет теории вероятностей. Статистическая устойчивость.

Теория вероятностей изучает закономерности, возникающие в случайных экспериментах (явлениях). Случайным называют эксперимент, результат которого нельзя предсказать заранее. Невозможность предсказать заранее — основное, что отличает *случайное* явление от *детерминированного*.

Не все случайные явления (эксперименты) можно изучать методами теории вероятностей, а лишь те, которые могут быть воспроизведены в одних и тех же условиях и обладают (непонятно как проверяемым заранее) свойством «статистической устойчивости»: если A — некоторое событие, могущее произойти или не произойти в результате эксперимента, то доля $n(A)/n$ числа экспериментов, в которых данное событие произошло, имеет тенденцию стабилизироваться с ростом общего числа экспериментов n, приближаясь к некоторому числу $\mathsf{P}(A)$. Это число служит объективной характеристикой «степени возможности» событию A произойти.

В дальнейшем мы будем говорить лишь о случайных экспериментах, обладающих данными свойствами, а свойство статистической устойчивости докажем в утверждении, известном как закон больших чисел Я. Бернулли.

Пространство элементарных исходов. Операции над событиями

Определение 1. *Пространством элементарных исходов* Ω («óмега») называется множество, содержащее все возможные результаты данного случайного эксперимента, из которых в эксперименте происходит ровно один. Элементы этого множества называют *элементарными исходами* и обозначают буквой ω («óмега») с индексами или без.

Определение 2. *Событиями* мы будем называть подмножества множества Ω. Говорят, что в результате эксперимента *произошло событие* $A \subseteq \Omega$, если в эксперименте произошел один из элементарных исходов, входящих в множество A.

Замечание 2. Вообще говоря, можно назвать событиями не обязательно все подмножества множества Ω, а лишь множества из некоторого набора подмножеств. О смысле такого ограничения мы поговорим позднее.

Пример 1. Один раз подбрасывается одна *игральная кость* (кубик). Самый разумный способ задать пространство элементарных исходов таков: $\Omega = \{1, 2, 3, 4, 5, 6\}$, элементарные исходы здесь соответствуют числу выпавших очков.

Примеры событий: $A = \{1, 2\}$ — выпало одно или два очка; $A = \{1, 3, 5\}$ — выпало нечетное число очков.

Пример 2. Два раза подбрасывается одна *игральная кость* (кубик). Или, что то же самое, один раз подбрасываются две игральные кости. Как мы увидим в дальнейшем, здесь самый разумный способ задать пространство элементарных исходов — считать результатом эксперимента упорядоченную пару чисел (i, j), в которой $1 \leqslant i, j \leqslant 6$

и i (j) есть число очков, выпавших при первом (втором) подбрасывании: $\Omega = \{(i, j),\ \text{где } 1 \leqslant i, j \leqslant 6\}$.

Примеры событий:
$A = \{(1,2),(1,3),(1,4),(1,5),(1,6)\}$ — при первом подбрасывании выпало одно очко;
$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$ — при двух подбрасываниях выпало одинаковое число очков.

Пример 3. На поверхность стола бросается монета. Результатом эксперимента можно считать координату центра монеты (а если нам не безразличен угол поворота монеты, то можно добавить и величину этого угла). Пространство элементарных исходов — множество точек стола (в втором случае — множество пар $\{(x, \varphi)\}$, где $x \in \mathbb{R}^2$ — точка стола и $\varphi \in [0, 2\pi)$ — угол поворота). Число элементарных исходов такого эксперимента несчетно.

Пример 4. Монета подбрасывается до тех пор, пока не выпадет вверх гербом. Пространство элементарных исходов состоит из бесконечного, но счетного числа исходов: $\Omega = \{$*г, рг, ррг, рррг, ррррг, рррррг,* $\ldots, \}$, где *р* и *г* обозначают выпадение решки и герба при одном подбрасывании, соответственно.

Определение 3.

1. *Достоверным* называется событие, которое обязательно происходит в результате эксперимента, то есть единственное событие, включающее все без исключения элементарные исходы — событие Ω.

2. *Невозможным* называется событие, которое не может произойти в результате эксперимента, то есть событие, не содержащее ни одного элементарного исхода («пустое множество» $\varnothing$). Заметим, что всегда $\varnothing \subset \Omega$.

Определение 4. Пусть A и B — события.

1. *Объединением* $A \cup B$ событий A и B называется событие, состоящее в том, что произошло либо A, либо B, либо оба события одновременно. На языке теории множеств $A \cup B$ есть множество, содержащее как элементарные исходы, входящие в A, так и элементарные исходы, входящие в B.

2. *Пересечением* $A \cap B$ событий A и B называется событие, состоящее в том, что произошли оба события A и B одновременно. То есть $A \cap B$ есть множество, содержащее элементарные исходы, входящие одновременно в A и в B.

3. *Дополнением* $A \backslash B$ события B до A называется событие, состоящее в том, что произошло событие A, но не произошло B. То есть $A \backslash B$ есть множество, содержащее элементарные исходы, входящие в A, но не входящие в B.

4. *Противоположным* (или *дополнительным*) к событию A называется событие $\overline{A} = \Omega \backslash A$, состоящее в том, что событие A в результате эксперимента не произошло. Иначе говоря, $\overline{A}$ есть множество, содержащее элементарные исходы, не входящие в A.

Определение 5.

1. События A и B называются ***несовместными***, если $A \cap B = \varnothing$.

2. События $A_1, \ldots, A_n$ называются ***попарно несовместными***, если для любых $i \neq j$, $1 \leqslant i, j \leqslant n$, события A_i и A_j несовместны.

3. Говорят, что событие A ***влечет*** событие B, и пишут $A \subseteq B$, если всегда, как только происходит событие A, происходит и событие B. На языке теории множеств это означает, что любой элементарный исход, входящий в A, одновременно входит и в событие B.

Вероятность на дискретном пространстве элементарных исходов

Предположим, что мы имеем дело с ***дискретным*** пространством элементарных исходов, то есть пространством, состоящим из конечного или счетного числа элементов: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$.

Определение 6. Поставим каждому элементарному исходу $\omega_i \in \Omega$ в соответствие число $p(\omega_i) \in [0, 1]$ так, что

$$\sum_{\omega_i \in \Omega} p(\omega_i) = 1.$$

Назовем число $p(\omega_i)$ ***вероятностью*** элементарного исхода ω_i. ***Вероятностью*** события $A \subseteq \Omega$ называется число

$$\mathsf{P}(A) = \sum_{\omega_i \in A} p(\omega_i),$$

равное сумме вероятностей элементарных исходов, входящих в множество A.

Замечание 3. Позднее, познакомившись с аксиоматикой теории вероятностей, мы зададим вероятности событий непосредственно, а не через вероятности элементарных исходов. Тем более, что сложением вероятностей элементарных исходов можно получить лишь вероятность события, состоящего не более чем из счетного числа элементарных исходов (иначе само понятие суммирования не определено). Но на дискретном пространстве элементарных исходов определить вероятности событий так, как это сделано в определении **6**, всегда возможно.

Перечислим очевидные в случае дискретного пространства элементарных исходов свойства вероятности, которые мы скоро докажем сразу в общем случае.

1. $0 \leqslant \mathsf{P}(A) \leqslant 1$; **2**. $\mathsf{P}(\Omega) = 1$; **3**. $\mathsf{P}(\varnothing) = 0$; **4**. $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$;

5. если A и B несовместны, то $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$;

6. в общем же случае $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$;

7. если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Упражнение 3. Доказать перечисленные выше свойства, пользуясь определением **6**.

Как видно, вероятностью может быть названа совершенно абстрактная функция, удовлетворяющая нескольким необременительным требованиям. Однако о необходимости «соответствия теории практике» тоже надо подумать.

Классическое определение вероятности

Предположим, что мы имеем дело с пространством элементарных исходов, состоящим из конечного числа N элементов: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. Более того, предположим, что из каких-либо соображений мы можем считать элементарные исходы *равновозможными*. Тогда вероятность любого из них принимается равной $1/N$.

Эти соображения чаще всего не имеют отношения к математической модели и основаны на какой-либо симметрии в эксперименте (симметричная монета, хорошо перемешанная колода карт, правильная кость). Либо мы можем заранее считать исходы эксперимента равновозможными, но тогда рано или поздно все равно возникнет вопрос о соответствии такой математической модели реальному эксперименту.

Если событие $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$ состоит из k элементарных исходов, то вероятность этого события равняется отношению k/N:

$$\mathsf{P}(A) = p(\omega_{i_1}) + \ldots + p(\omega_{i_k}) = k \cdot \frac{1}{N} = \frac{|A|}{|\Omega|},$$

где символом $|A|$ обозначено число элементов конечного множества A.

Определение 7.

Говорят, что эксперимент удовлетворяет *классическому определению вероятности* (или классической вероятностной схеме), если пространство элементарных исходов состоит из конечного числа $|\Omega| = N$ равновозможных исходов.

В этом случае вероятность любого события A вычисляется по формуле

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|},$$

называемой *классическим определением вероятности*. Эта формула читается так: «вероятность события A равна отношению числа исходов, *благоприятствующих* событию A, к общему числу исходов».

Замечание 4. Полезно помнить классическую формулировку Якоба Бернулли: ***«Вероятность есть степень достоверности и отличается от нее как часть от целого»***

(*Ars Conjectandi*, 1713 г.)

Замечание 5. Мы видим теперь, что подсчет вероятности в классической схеме сводится к подсчету числа «шансов» (элементарных исходов), благоприятствующих какому-либо событию, и общего числа шансов. Как правило, это делается с помощью формул комбинаторики.

Рассмотрим описанные в параграфе **1.1** урновые схемы. Напомним, что речь идет об извлечении k шариков из урны, содержащей n шариков. При этом три схемы: с возвращением и с учетом порядка, без возвращения и с учетом порядка, а также без возвращения и без учета порядка удовлетворяют классическому определению вероятности. Общее число элементарных исходов в этих схемах подсчитано в теоремах **4**, **2**, **3** и равно, соответственно, n^k, A_n^k, C_n^k.

Четвертая же схема — схема выбора с возвращением и без учета порядка — имеет заведомо ***неравновозможные*** исходы.

Пример 5. Рассмотрим, скажем, выбор двух шариков из двух или, что то же самое, дважды подбросим монету. Если учитывать порядок, то исходов получится 4, и все они равновозможны, то есть имеют вероятность по 1/4:

(герб,герб), (решка,решка), (решка,герб), (герб,решка).

Если порядок не учитывать, то следует объявить два последних исхода одним и тем же результатом эксперимента, и получить три исхода вместо четырех: выпало

два герба, либо две решки, либо один герб и одна решка.

При этом первые два исхода имеют вероятность 1/4, а последний — вероятность $1/4+1/4=1/2$.

Упражнение 4. Посчитать число элементарных исходов в примере **2** (при подбрасывании двух игральных костей). Каким станет пространство элементарных исходов, если порядок костей не учитывать? Посчитать число элементарных исходов в таком пространстве (пользуясь теоремой **5** или прямым подсчетом). Убедиться, что их ровно $C_7^2 = 21$. Равновозможны ли эти исходы? Посчитать вероятность каждого исхода.

Гипергеометрическое распределение

Пример 6.

Из урны, в которой n_1 белых и $n-n_1$ черных шаров, наудачу, без возвращения вынимают k шаров, $k \leqslant n$. Термин «наудачу» означает, что появление любого набора из k шаров равновозможно. Найти вероятность того, что будет выбрано ровно k_1 белых и $k-k_1$ черных шаров.

$$\boxed{n_1 \;\;\big|\;\; n-n_1}$$
$$\Downarrow$$
$$\boxed{k_1 \;\;\big|\;\; k-k_1}$$

Решение. Заметим, что при $k_1 > n_1$ или $k - k_1 > n - n_1$ искомая вероятность равна 0, так как соответствующее событие невозможно. Пусть $k_1 \leqslant n_1$ и $k - k_1 \leqslant n - n_1$.

Результатом эксперимента является набор из k шаров. При этом можно не учитывать или учитывать порядок следования шаров.

1. Выбор без учета порядка. Общее число элементарных исходов есть число k-элементных подмножеств множества, состоящего из n элементов, то есть $|\Omega| = C_n^k$ (по теореме **3**).

Обозначим через A событие, вероятность которого требуется найти. Событию A благоприятствует появление любого набора, содержащего k_1 белых шаров и $k - k_1$ черных. Число благоприятных исходов равно произведению (по теореме **1**) числа способов выбрать k_1 белых шаров из n_1 и числа способов выбрать $k - k_1$ черных шаров из $n - n_1$: $|A| = C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}$. Вероятность события A равна

$$\mathsf{P}(A) = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}. \tag{1}$$

2. Выбор с учетом порядка. Общее число элементарных исходов есть число способов разместить n элементов на k местах: $|\Omega| = A_n^k = n(n-1)\ldots(n-k+1)$ (по теореме **2**).

При подсчете числа благоприятных исходов нужно учесть как число способов выбрать нужное число шаров, так и число способов расположить эти шары среди k. Можно, скажем, посчитать число способов выбрать k_1 мест среди k (равное $C_k^{k_1}$), затем число способов разместить на этих k_1 местах n_1 белых шаров (равное $A_{n_1}^{k_1}$ — не забывайте про учет порядка!), и затем число способов разместить на оставшихся $k - k_1$ местах $n - n_1$ черных шаров (равное $A_{n-n_1}^{k-k_1}$). Перемножив (почему?) эти числа, получим

$$|A| = C_k^{k_1} \cdot A_{n_1}^{k_1} \cdot A_{n-n_1}^{k-k_1}, \qquad \mathsf{P}(A) = \frac{C_k^{k_1} \cdot A_{n_1}^{k_1} \cdot A_{n-n_1}^{k-k_1}}{A_n^k} = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}.$$

В рассмотренной задаче мы сопоставили каждому набору из k_1 белых и $k - k_1$ черных шаров вероятность получить этот набор при выборе k шаров из урны, содержащей n_1 белых и $n - n_1$ черных шаров:

$$\mathsf{P}(A) = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}.$$

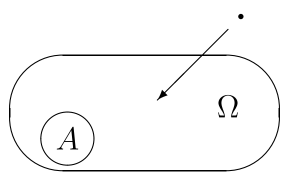
Определение 8. Соответствие $k_1 \mapsto \mathsf{P}(A) = \dfrac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}$, или следующий набор вероятностей

$$\left\{ \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}, \quad \text{где} \quad 0 \leqslant k_1 \leqslant \min(k, n_1), \quad k - k_1 \leqslant n - n_1 \right\}$$

называется *гипергеометрическим распределением*.

Раздел 2. Геометрическая вероятность

2.1 Что это такое



Рассмотрим какую-нибудь область Ω в $\mathbb{R}^m$ (на прямой, на плоскости, в пространстве). Предположим, что «мера» Ω (длина, площадь, объем, соответственно) конечна. Пусть случайный эксперимент состоит в том, что мы наудачу бросаем в эту область точку. Термин «наудачу» здесь означает, что вероятность попадания точки в любую часть $A \subset \Omega$ не зависит от формы или расположения A внутри Ω, а зависит лишь от «меры» области A (если A измеримо, см. замечание **6**).

Определение 9. Эксперимент удовлетворяет условиям «геометрического определения вероятности», если его исходы *можно* изобразить точками некоторой области Ω в $\mathbb{R}^m$ так, что вероятность попадания точки в любую часть $A \subset \Omega$ не зависит от формы или расположения A внутри Ω, а зависит лишь от меры области A (и, следовательно, пропорциональна этой мере):

$$\mathsf{P}(\cdot \in A) = \frac{\mu(A)}{\mu(\Omega)}, \quad \text{где} \quad \mu(A) \quad \text{обозначает меру области} \quad A.$$

«Мерой» мы пока будем называть длину, площадь, объем и т.д.

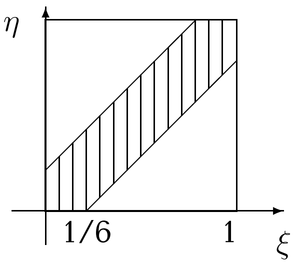
Если для точки, брошенной в область Ω, выполнены условия геометрического определения вероятности, то говорят, что точка *равномерно распределена в области* Ω.

Пример 7. Точка наудачу бросается на отрезок [0,1]. Вероятность точке попасть в точку $\{0.5\}$ равна нулю, так как мера множества, состоящего из одной точки («длина точки»), есть 0. Вместе с тем попадание в точку $\{0.5\}$ не является невозможным событием — это один из элементарных исходов эксперимента.

2.2 Задача о встрече

Пример 8. Два лица X и Y условились встретиться в определенном месте между двумя и тремя часами дня. Пришедший первым ждет другого в течении 10 минут, после чего уходит. Чему равна вероятность встречи этих лиц, если каждый из них может прийти в любое время в течение указанного часа независимо от другого?

Решение. Будем считать интервал с 14 до 15 часов дня отрезком [0,1] длиной 1 час. Пусть ξ («кси») и η («э́та») — моменты прихода X и Y (точки отрезка [0,1]). Все возможные результаты эксперимента — множество точек квадрата со стороной 1: $\Omega = \{(\xi, \eta) : \ 0 \leqslant \xi \leqslant 1, 0 \leqslant \eta \leqslant 1\} = [0,1] \times [0,1]$.



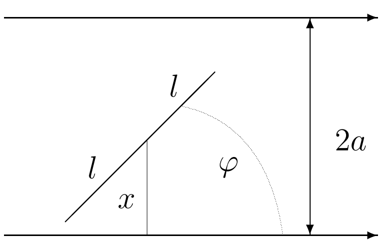
Можно считать, что эксперимент сводится к бросанию точки наудачу в квадрат. При этом благоприятными исходами являются точки множества $A = \{(\xi, \eta) : \ |\xi - \eta| \leqslant 1/6\}$ (10 минут = 1/6 часа). То есть попадание в множество A наудачу брошенной в квадрат точки означает, что X и Y встретятся. Тогда вероятность встречи равна

$$\mathsf{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1 - \left(\frac{5}{6}\right)^2}{1} = \frac{11}{36}.$$

2.3 Задача Бюффона

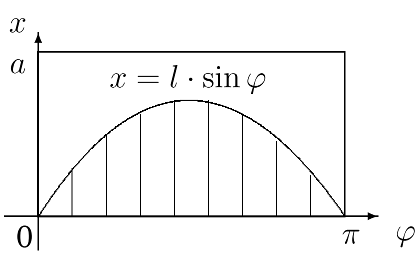
Пример 9. На плоскости начерчены параллельные прямые, находящиеся друг от друга на расстоянии $2a$. На плоскость наудачу брошена игла длины $2l < 2a$. Какова вероятность того, что игла пересечет одну из прямых?

Решение. Поймем, что означает здесь «наудачу брошена игла». Возможные положения иглы (отрезка) на плоскости полностью определяются положением середины иглы и углом поворота иглы относительно какого-либо направления. Причем две эти переменные (положение центра и угол поворота) меняются независимо друг от друга.



Обозначим через $x \in [0, a]$ расстояние от середины иглы до ближайшей прямой, а через $\varphi \in [0, \pi]$ — угол между каким-то направлением прямых и иглой. Множество возможных положений иглы целиком определяется выбором наудачу точки из прямоугольника $\Omega = [0, \pi] \times [0, a]$.

Игла пересекает ближайшую прямую, если координаты выбранной наудачу точки удовлетворяют неравенству: $x \leqslant l \cdot \sin \varphi$.



Площадь области $A \subset \Omega$, точки которой удовлетворяют такому неравенству, равна

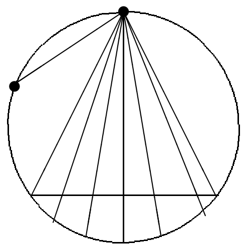
$$\mu(A) = \int_0^\pi l \cdot \sin \varphi \; d\varphi = -l \cdot \cos \varphi \Big|_0^\pi = 2l.$$

И так как $\mu(\Omega) = a\pi$, то искомая вероятность равна $\mathsf{P}(A) = \dfrac{2l}{a\pi}$.

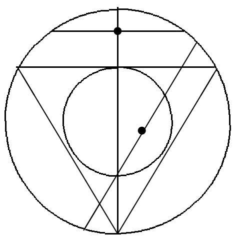
2.4 Парадокс Бертрана

Пример 10 (Josef Bertrand, “Calcul des Probabilites", 1888). В круге единичного радиуса наудачу выбирается хорда. Какова вероятность того, что ее длина будет больше, чем длина стороны вписанного в круг правильного треугольника?

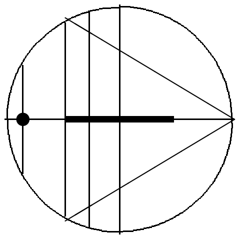
«Решение». Есть по крайней мере три способа «выбрать наудачу хорду в круге».



1. Зафиксируем одну точку (конец хорды) на окружности и выберем наудачу на окружности другую точку (второй конец хорды). Здесь $\Omega = [0, 2\pi]$, а благоприятными являются положения второй точки на интервале $[2\pi/3, 4\pi/3]$ (хорды, помеченные на рисунке красным цветом). Вероятность получить «длинную» хорду равна $\dfrac{1}{3}$.



2. Существует ровно одна хорда, для которой данная точка в круге является серединой[1]. Можно поэтому выбирать наудачу хорду, бросая наудачу точку (середину хорды) в круг. Здесь Ω — круг радиуса 1, $\mu(\Omega) = \pi$, а благоприятными являются положения середины хорды внутри вписанного в треугольник круга (радиусом 1/2). Вероятность получить «длинную» хорду равна отношению площадей кругов, то есть $\dfrac{1}{4}$.



3. Наконец, можно ограничиться рассмотрением только хорд, перпендикулярных какому-либо диаметру (остальные могут быть получены поворотом). То есть эксперимент может состоять в выборе середины хорды наудачу на диаметре круга — отрезке длиной 2. Благоприятными являются положения середины хорды на отрезке длиной 1. Искомая вероятность для такого эксперимента равна $\dfrac{1}{2}$.

В чем причина разницы в ответах на, казалось бы, один и тот же вопрос? На самом деле формулировка задачи не корректна с математической точки зрения. «Выбор наудачу хорды в круге» может быть по-разному описан с помощью геометрического определения вероятности (что мы и сделали). То есть этот «эксперимент» можно по-разному описать с помощью выбора наудачу точки в некоторой области.

[1]Кроме того случая, когда брошенная наудачу точка попадет в центр круга. Но поскольку вероятность этого события равна нулю, то учет или неучет такого события не влияет на итоговую вероятность

Слово «эксперимент» взято в кавычки не напрасно: сказав «в круге наудачу выбирается хорда», мы еще не описали физического эксперимента. Действительно, каждому из трех предложенных способов выбора хорд можно сопоставить конкретный физический эксперимент (всякий раз другой).

Так что парадокс исчезает сразу, как только получен ответ на вопрос: что значит «в круге наудачу выбирается хорда»?

Заканчивая обсуждение понятия геометрической вероятности, сделаем очень важное для дальнейшего замечание.

Замечание 6. Если даже эксперимент удовлетворяет геометрическому определению вероятности, далеко не для всех множеств $A \subset \Omega$ вероятность может быть вычислена как отношение меры A к мере Ω. Причиной этого является существование так называемых «неизмеримых» множеств, то есть множеств, мера которых не существует.

А если не для всех подмножеств Ω мы можем определить их вероятности, следует сузить класс множеств, называемых «событиями», оставив в этом классе только те множества, для которых мы можем определить вероятность.

В следующей главе мы займемся построением (вслед за Андреем Николаевичем Колмогоровым) аксиоматики теории вероятностей: познакомимся с понятиями σ-алгебры (или поля) событий, вероятностной меры, вероятностного пространства, а также докажем сформулированные в параграфе **1.2** свойства вероятности.

Раздел 3. Аксиоматика теории вероятностей

3.1 σ-алгебра событий

Пусть Ω — пространство элементарных исходов некоторого случайного эксперимента (то есть, вообще говоря, множество произвольной природы). Мы собираемся определить набор подмножеств Ω, которые будут называться событиями, и затем задать вероятность как функцию, определенную *только* на множестве событий.

То есть событиями мы будем называть не *любые* подмножества Ω, а лишь подмножества из некоторого «множества подмножеств» $\mathcal{F}$. При этом необходимо позаботиться, чтобы это множество $\mathcal{F}$ подмножеств Ω было «замкнуто» относительно введенных в параграфе **1.2** операций над событиями, то есть чтобы объединение, пересечение, дополнение событий (то есть элементов $\mathcal{F}$) снова давало событие (то есть элемент $\mathcal{F}$).

Определение 10. Множество $\mathcal{F}$, состоящее из подмножеств множества Ω (не обязательно всех!) называется ***σ-алгеброй событий***, или ***σ-алгеброй подмножеств*** Ω, если выполнены следующие условия:

(A1) $\Omega \in \mathcal{F}$ (σ-алгебра событий содержит достоверное событие);

(A2) если $A \in \mathcal{F}$, то $\overline{A} \in \mathcal{F}$ (вместе с любым событием σ-алгебра содержит противоположное событие);

(A3) если $A_1, A_2, \ldots \in \mathcal{F}$, то $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$ (вместе с любым конечным или счетным набором событий σ-алгебра содержит их объединение).

Условия (A1) — (A3) часто называют «аксиомами σ-алгебры».

Проверим, что этого набора аксиом достаточно (даже более чем достаточно - ***убедитесь, что первая аксиома следует из двух оставшихся***) для замкнутости множества $\mathcal{F}$ относительно других операций над событиями.

Свойство 1. $\varnothing \in \mathcal{F}$ *(σ-алгебра событий содержит невозможное событие).*

Доказательство. По (A1), $\Omega \in \mathcal{F}$, но $\varnothing = \Omega \backslash \Omega = \overline{\Omega} \in \mathcal{F}$ в силу (A2). □

Свойство 2. *При выполнении* (A1),(A2) *свойство* (A3) *эквивалентно свойству* (A4)

(A4) *если $A_1, A_2, \ldots \in \mathcal{F}$, то $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{F}$ (вместе с любым конечным или счетным набором событий σ-алгебра содержит их пересечение).*

Доказательство. Докажем, что при выполнении (A1),(A2) из (A3) следует (A4).

Если $A_1, A_2, \ldots \in \mathcal{F}$, то при всех $i = 1, 2, \ldots$ по свойству (A2) выполнено $\overline{A}_i \in \mathcal{F}$. Тогда из (A3) следует, что $\bigcup\limits_{i=1}^{\infty} \overline{A}_i \in \mathcal{F}$, и, по (A2), дополнение к этому множеству также принадлежит $\mathcal{F}$, то есть $\overline{\bigcup\limits_{i=1}^{\infty} \overline{A}_i} \in \mathcal{F}$. Но, в силу формул двойственности, $\overline{\bigcup\limits_{i=1}^{\infty} \overline{A}_i} = \bigcap\limits_{i=1}^{\infty} A_i$, что и т.д.

Доказательство в обратную сторону выглядит совершенно аналогично.

$\square$

Свойство 3. *Если $A, B \in \mathcal{F}$, то $A \backslash B \in \mathcal{F}$.*

Доказательство. $A \backslash B = A \cap \overline{B} \in \mathcal{F}$, так как $A \in \mathcal{F}$, $\overline{B} \in \mathcal{F}$, и по (A4) их пересечение тоже принадлежит $\mathcal{F}$. $\square$

Пример 11. Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$ — пространство элементарных исходов (например, при бросании игрального кубика). Следующие наборы подмножеств Ω являются σ-алгебрами (***доказать!***):

1. $\mathcal{F} = \{\Omega, \varnothing\} = \{\{1, 2, 3, 4, 5, 6\}, \varnothing\}$ — *тривиальная σ-алгебра.*

2. $\mathcal{F} = \{\Omega, \varnothing, \{1\}, \overline{\{1\}}\} = \{\{1, 2, 3, 4, 5, 6\}, \varnothing, \{1\}, \{2, 3, 4, 5, 6\}\}$.

3. $\mathcal{F} = \{\Omega, \varnothing, A, \overline{A}\} = \big\{\{1, 2, 3, 4, 5, 6\}, \varnothing, A, \overline{A}\big\}$, где A — произвольное подмножество Ω (в предыдущем примере $A = \{1\}$).

4. $\mathcal{F}$ — множество всех подмножеств Ω. ***Доказать, что если Ω состоит из n элементов, то в множестве всех его подмножеств ровно 2^n элементов.***

Итак, мы определили специальный класс $\mathcal{F}$ подмножеств пространства элементарных исходов Ω, названный σ-алгеброй событий, причем применение счетного числа любых операций (таких, как объединение, пересечение, дополнение) к множествам из $\mathcal{F}$ снова дает множество из $\mathcal{F}$ (не выводит за рамки этого класса). Множества $A \in \mathcal{F}$ мы и назвали «событиями».

Определим теперь понятие «вероятности» как функции, определенной на множестве событий (то есть функции, которая каждому *событию* ставит в соответствие число). А чтобы читателю сразу стало понятно, о чем пойдет речь, добавим: вероятность мы определим как *неотрицательную нормированную меру*, заданную на σ-алгебре $\mathcal{F}$ подмножеств Ω.

3.2 Вероятность как нормированная мера

Определение 11.

Пусть Ω — некоторое множество и $\mathcal{F}$ — σ-алгебра его подмножеств. Функция $\mu : \mathcal{F} \to \mathbb{R} \cup \{\infty\}$ называется ***мерой*** на $(\Omega, \mathcal{F})$, если она удовлетворяет условиям:

(M1) Для любого множества $A \in \mathcal{F}$ его мера неотрицательна: $\mu(A) \geqslant 0$.

(M2) Для любого счетного набора попарно непересекающихся множеств $A_1, A_2, A_3, \ldots \in \mathcal{F}$ (то есть такого, что $A_i \cap A_j = \varnothing$ при всех $i \neq j$) мера их объединения равна сумме их мер: $\mu\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ («счетная аддитивность» или «σ-аддитивность»)

Иначе говоря, мера есть неотрицательная, счетно-аддитивная функция множеств.

Определение 12. Пусть Ω — некоторое множество и $\mathcal{F}$ — σ-алгебра его подмножеств. Мера $\mu : \mathcal{F} \to \mathbb{R}$ называется ***нормированной***, если $\mu(\Omega) = 1$. Другое название нормированной меры — ***«вероятность»*** или ***«вероятностная мера»***.

То же самое еще раз и подробно:

Определение 13.

Пусть Ω — пространство элементарных исходов и $\mathcal{F}$ — σ-алгебра его подмножеств (событий). ***Вероятностью*** или ***вероятностной мерой*** на $(\Omega, \mathcal{F})$ называется функция $\mathsf{P} : \mathcal{F} \to \mathbb{R}$, обладающая свойствами:

(P1) Для любого события $A \in \mathcal{F}$ выполняется неравенство $\mathsf{P}(A) \geqslant 0$;

(P2) Для любого счетного набора попарно несовместных событий $A_1, A_2, A_3, \ldots \in \mathcal{F}$ имеет место равенство

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

(P3) Вероятность достоверного события равна единице: $\mathsf{P}(\Omega) = 1$.

Свойства (P1) – (P3) часто называют «аксиомами вероятности».

Определение 14. Тройка $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle$, в которой Ω — пространство элементарных исходов, $\mathcal{F}$ — σ-алгебра его подмножеств и P — вероятностная мера на $\mathcal{F}$, называется *вероятностным пространством*.

Докажем свойства вероятности, вытекающие из аксиом. Здесь и в дальнейшем под знаком вероятности появляются только события!

0. $\mathsf{P}(\varnothing) = 0$.

Доказательство. События $A_i = \varnothing$, $i = 1, 2, \ldots$, попарно несовместны, и их объединение есть также пустое множество. По аксиоме (P2)

$$\mathsf{P}(\varnothing) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \sum_{i=1}^{\infty} \mathsf{P}(\varnothing).$$

Это возможно только в случае $\mathsf{P}(\varnothing) = 0$.

1. Для любого *конечного* набора попарно несовместных событий $A_1, \ldots, A_n \in \mathcal{F}$ имеет место равенство

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

Доказательство. Пусть $A_i = \varnothing$ при любом $i > n$. Вероятности этих событий, по предыдущему свойству, равны нулю. События $A_1, \ldots, A_n, \varnothing, \varnothing, \varnothing, \varnothing, \ldots$ попарно несовместны, и, по аксиоме (P2),

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

2. $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$.

Доказательство. $A \cup \overline{A} = \Omega$, и события A, $\overline{A}$ несовместны. По аксиоме (P3) и предыдущему свойству, $1 = \mathsf{P}(\Omega) = \mathsf{P}(A) + \mathsf{P}(\overline{A})$.

3. Если $A \subseteq B$, то $\mathsf{P}(B \backslash A) = \mathsf{P}(B) - \mathsf{P}(A)$.

Доказательство. $B = A \cup (B \backslash A)$, и события A, $B \backslash A$ несовместны. По аксиоме (P2), $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \backslash A)$.

4. Если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Доказательство. По предыдущему свойству, $\mathsf{P}(A) = \mathsf{P}(B) - \mathsf{P}(B \backslash A) \leqslant \mathsf{P}(B)$. Последнее неравенство следует из (P1), т.к. $\mathsf{P}(B \backslash A) \geqslant 0$.

5. $0 \leqslant \mathsf{P}(A) \leqslant 1$.

Доказательство. $\mathsf{P}(A) \geqslant 0$ по (P1), и т.к. $A \subseteq \Omega$, то по предыдущему свойству $\mathsf{P}(A) \leqslant \mathsf{P}(\Omega) = 1$.

6. $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$.

Доказательство. $A \cap B \subseteq B$, поэтому $\mathsf{P}(B \backslash (A \cap B)) = \mathsf{P}(B) - \mathsf{P}(A \cap B)$. Но события A и $B \backslash (A \cap B)$ несовместны, поэтому

$$\mathsf{P}(A \cup B) = \mathsf{P}(A \ \cup \ B \backslash (A \cap B)) = \mathsf{P}(A) + \mathsf{P}(B \backslash (A \cap B)) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

7. $\mathsf{P}(A \cup B) \leqslant \mathsf{P}(A) + \mathsf{P}(B)$.

Доказательство. Сразу следует из предыдущего свойства и аксиомы (P1).

8. $\mathsf{P}(A_1 \cup \ldots \cup A_n) \leqslant \sum\limits_{i=1}^{n} \mathsf{P}(A_i)$. ***Доказать методом математической индукции.***

9.

$$\mathsf{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{i<j} \mathsf{P}(A_i \cap A_j) + \sum_{i<j<m} \mathsf{P}(A_i \cap A_j \cap A_m) - \ldots + (-1)^{n-1} \mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_n). \quad (2)$$

Доказательство. Воспользуемся методом математической индукции. Базис индукции при $n = 2$ — свойство **6** выше. Пусть свойство **9** верно при $n = k - 1$. Докажем, что тогда оно верно при $n = k$.

$$\mathsf{P}\left(\bigcup_{i=1}^{k} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i \ \cup \ A_k\right) = \mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i\right) + \mathsf{P}\left(A_k\right) - \mathsf{P}\left(A_k \cap \bigcup_{i=1}^{k-1} A_i\right) \quad (3)$$

По предположению индукции, первое слагаемое в правой части (3) равно

$$\mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i\right) = \sum_{i=1}^{k-1} \mathsf{P}(A_i) - \sum_{1 \leqslant i<j \leqslant k-1} \mathsf{P}(A_i \cap A_j) + \sum_{1 \leqslant i<j<m \leqslant k-1} \mathsf{P}(A_i \cap A_j \cap A_m) - \ldots + (-1)^{k-2} \mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_{k-1}). \quad (4)$$

Вычитаемое в правой части (3) равно

$$\mathsf{P}\left(A_k\cap\bigcup_{i=1}^{k-1}A_i\right)=\mathsf{P}\left(\bigcup_{i=1}^{k-1}(A_i\cap A_k)\right)=\sum_{i=1}^{k-1}\mathsf{P}(A_i\cap A_k)-\sum_{1\leqslant i<j\leqslant k-1}\mathsf{P}(A_i\cap A_j\cap A_k)+$$
$$+\sum_{1\leqslant i<j<m\leqslant k-1}\mathsf{P}(A_i\cap A_j\cap A_m\cap A_k)-\ldots+(-1)^{k-2}\mathsf{P}(A_1\cap A_2\cap\ldots\cap A_{k-1}\cap A_k). \quad (5)$$

Подставить (4),(5) *в* (3) *и довести до конца шаг индукции.* □

Приведем пример задачи, в которой использование свойства **9** — самый простой путь решения.

Пример 12. Есть n писем и n подписанных конвертов. Письма раскладываются в конверты наудачу. Найти вероятность того, что хотя бы одно письмо попадет в предназначенный ему конверт и предел этой вероятности при $n\to\infty$.

Р е ш е н и е. Пусть событие A_i, $i=1,\ldots,n$ означает, что i-е письмо попало в свой конверт. Тогда $A=\{$*хотя бы одно письмо попадет в свой конверт*$\}=A_1\cup\ldots\cup A_n$. И так как события $A_1,\ldots,A_n$ совместны, придется использовать формулу (2). Нетрудно убедиться, что

$$\mathsf{P}(A_i)=\frac{1}{n}\ \text{для всех}\quad i,\quad \mathsf{P}(A_i\cap A_j)=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)}\quad\text{для всех}\quad i\neq j,$$

$$\mathsf{P}(A_i\cap A_j\cap A_m)=\frac{(n-3)!}{n!}=\frac{1}{n(n-1)(n-2)}\ \text{для всех}\ i\neq j\neq m,\quad\ldots,\quad\mathsf{P}(A_1\cap\ldots\cap A_n)=\frac{1}{n!}$$

Вычислим количество слагаемых в каждой сумме в формуле (2). Например, в сумме $\sum\limits_{1\leqslant i<j<m\leqslant n}$ ровно C_n^3 слагаемых — ровно столько трех-элементных множеств можно образовать из n элементов, и каждое такое множество $\{i,j,m\}$ встречается в индексах данной суммы единажды.

Подставляя все вероятности в формулу (2), получим:

$$\mathsf{P}(A)=n\cdot\frac{1}{n}-C_n^2\cdot\frac{1}{n(n-1)}+C_n^3\cdot\frac{1}{n(n-1)(n-2)}-\ldots+(-1)^{n-1}\frac{1}{n!}\ =\ 1-\frac{1}{2!}+\frac{1}{3!}-\ldots+(-1)^{n-1}\frac{1}{n!}$$

Выписать разложение e^{-1} в ряд Тейлора и убедиться, что $\mathsf{P}(A)\longrightarrow 1-e^{-1}$ при $n\to\infty$.

3.3 О борелевской σ-алгебре и мере Лебега

Следующий параграф предназначен только для тех, кто не испугался всего сказанного выше и хочет познакомиться с понятиями ***«σ-алгебра борелевских множеств»*** и ***«мера Лебега»***.

Борелевская σ-алгебра на прямой

Пример 13. Пусть $\Omega = \mathbb{R}$ — вещественная прямая. Рассмотрим некоторые наборы множеств, не являющиеся σ-алгебрами, и увидим, как их можно дополнить до σ-алгебр.

1. Множество $\mathcal{A} = \{\Omega,\ \varnothing,\ [0,1],\ \{0\}\} = \{\mathbb{R},\ \varnothing,\ [0,1],\ \{0\}\}$ не является σ-алгеброй, так как, например, $\overline{[0,1]} = \mathbb{R}\backslash[0,1] = (-\infty,0)\cup(1,\infty) \notin \mathcal{A}$. Минимальный набор множеств, содержащий $\mathcal{A}$ и являющийся σ-алгеброй (*минимальная σ-алгебра*), получится, если включить в него всевозможные объединения, пересечения и дополнения множеств из $\mathcal{A}$:

   $$\mathcal{F} = \{\ \mathbb{R},\ \varnothing,\ [0,1],\ \{0\},\ (-\infty,0)\cup(1,\infty),\ (0,1],\ (-\infty,0]\cup(1,\infty),\ (-\infty,0)\cup(0,\infty)\ \}$$

   Более точно, *минимальной σ-алгеброй*, содержащей набор множеств $\mathcal{A}$, называется пересечение всех σ-алгебр, содержащих $\mathcal{A}$.

2. Найти минимальную σ-алгебру, содержащую $\mathcal{A} = \{\mathbb{R},\ \varnothing,\ [0,1],\ \{3\}\}$

3. Пусть множество $\mathcal{A}$ подмножеств вещественной прямой $\mathbb{R}$ состоит из всевозможных открытых интервалов (a,b), где $a < b$: $\quad \mathcal{A} = \{(a,b) : -\infty < a < b < \infty\}$.

   (a) Проверить, что $\mathcal{A}$ ни в коем случае не является σ-алгеброй!

   Указание: привести примеры двадцати множеств из $\mathcal{A}$, дополнения к которым не принадлежат $\mathcal{A}$; привести примеры пяти множеств из $\mathcal{A}$, любые объединения которых не принадлежат $\mathcal{A}$.

   (b) Минимальная σ-алгебра, содержащая множество $\mathcal{A}$ всех интервалов на вещественной прямой, называется ***борелевской σ-алгеброй*** в $\mathbb{R}$ (Felix Edouard Justin Emile Borel) и обозначается $\mathcal{B}$ или $\mathcal{B}(\mathbb{R})$.

   (c) Перечислим некоторые множества на прямой, содержащиеся в $\mathcal{B}$. Таковы все привычные нам множества. Чтобы получить множество, не содержащееся в $\mathcal{B}$, требуются специальные построения.

   Итак, мы знаем, что все интервалы на прямой принадлежат $\mathcal{B}$, и $\mathcal{B}$ — σ-алгебра.

   - $\mathbb{R}$ ***принадлежит*** $\mathcal{B}$. Это сразу следует из свойства (A1) σ-алгебры, но может быть доказано исходя из свойств (A2), (A3).

Действительно, $\mathbb{R} = \bigcup\limits_{n=1}^{\infty}(-n,n)$. Так как все эти интервалы лежат в $\mathcal{A}$, а $\mathcal{A} \subset \mathcal{B}$, то все эти интервалы принадлежат $\mathcal{B}$. Но $\mathcal{B}$ — σ-алгебра, поэтому она содержит счетное объединение любых своих элементов. Поэтому $\mathbb{R} = \bigcup\limits_{n=1}^{\infty}(-n,n) \in \mathcal{B}$.

- *Любой интервал вида $(a,b]$ (или $[a,b)$, или $[a,b]$), где $a < b$, принадлежит $\mathcal{B}$.*

  Действительно, $(a,b] = \bigcap\limits_{n=1}^{\infty}\left(a, b+\frac{1}{n}\right)$, и так как все эти интервалы лежат в $\mathcal{B}$, то их счетное пересечение должно по свойству (A4) принадлежать $\mathcal{B}$.

- *Любое одноточечное подмножество $\{b\} \subset \mathbb{R}$ принадлежит $\mathcal{B}$.*

  Действительно, $\{b\} = (a,b] \setminus (a,b)$, а разность двух множеств из σ-алгебры снова принадлежит σ-алгебре.

- Докажите, что, например, *любые множества вида $(a_1,b_1) \cup (a_2,b_2)$ принадлежит $\mathcal{B}$, множество натуральных чисел $\mathbb{N}$ принадлежит $\mathcal{B}$, множество рациональных чисел $\mathbb{Q}$ принадлежит $\mathcal{B}$.*

4. Борелевская σ-алгебра в $\mathbb{R}^n$ строится совершенно так же, как в $\mathbb{R}$. Это должна быть минимальная σ-алгебра, содержащая все множества вида $(a_1,b_1) \times \ldots \times (a_n,b_n)$ (уже не интервалы, как в $\mathbb{R}$, а «прямоугольники» в $\mathbb{R}^2$, «параллелепипеды» в $\mathbb{R}^3$ и т.д.).

Мера Лебега

Когда мы говорили о геометрической вероятности, мы использовали термин «мера области A в $\mathbb{R}^m$», имея ввиду «длину» на прямой, «площадь» на плоскости, «объем» в трехмерном пространстве. Являются ли все эти «длины-площади-объемы» настоящими мерами в смысле определения **11**? Мы решим этот вопрос для прямой, оставляя плоскость и пространство большей размерности читателю.

Если вам уже расхотелось читать дальше, сообщаем: мерой Лебега в задачниках и учебниках называют как раз «длину-площадь-объем», так что все в порядке.

Рассмотрим вещественную прямую с σ-алгеброй борелевских множеств. Эта σ-алгебра, по определению, есть наименьшая σ-алгебра, содержащая любые интервалы. Для каждого интервала $(a,b) \subset \mathbb{R}$ число $b-a$ назовем «длиной интервала (a,b)». Мы не станем доказывать следующее утверждение:

Лемма 1. *Существует единственная* ***мера*** *(то есть неотрицательная и σ-аддитивная функция) λ на $(\mathbb{R}, \mathcal{B})$, значение которой на любом интервале равно его длине: $\lambda\big((a,b)\big) = b-a$. Эта мера называется* ***мерой Лебега***.

Это утверждение является следствием теоремы Каратеодори о продолжении меры с алгебры на σ-алгебру, применительно к $(\mathbb{R}, \mathcal{B})$. См. А.Н.Колмогоров, С.В.Фомин, *Функциональный анализ* или А.А.Боровков, *Теория вероятностей.*

Итак, мы ограничили набор событий только множествами из какой-нибудь σ-алгебры событий. Мы потребовали, чтобы вероятность была функцией *только* на множестве событий. Покажем, что это необходимо: построим пример множества на отрезке, мера Лебега которого («длина») просто не существует (множество Витали).

То есть: если рассмотреть бросание точки наудачу на отрезок, то вычислить вероятность попадания точки в указанное множество в соответствии с геометрической вероятностью нельзя. Значит, это множество нельзя считать событием — мы не умеем вычислить его вероятность!

Пример 14. Рассмотрим окружность единичного радиуса (реально это тот же отрезок $[0, 2\pi]$). Возьмем любое иррациональное число α. Поскольку оно иррационально, число $n\alpha$ не является целым ни при каком целом $n \neq 0$ (то есть число $2\pi n\alpha$ равно $2\pi k\alpha$ лишь при $n = k$).

Поэтому если взять *произвольную* точку $x \in [0, 2\pi]$, то есть точку на окружности, и перечислить все точки, которые получаются поворотом точки x на угол $2\pi n\alpha$, $n = \pm 1, \pm 2, \ldots$, то мы ни разу не вернемся в точку x. Точек, получившихся из точки x такими поворотами, счетное число. Объединим их в один класс. С любой другой точкой окружности можно тоже связать класс точек, получающихся из нее поворотом на угол $2\pi n\alpha$ при каком-то $n \in \mathbb{Z}$.

То есть вся окружность разбивается на классы точек. В каждом классе счетное число точек, и все точки в одном классе получаются друг из друга такими поворотами. Причем эти классы не пересекаются.

Множество A_0 определим так: возьмем из каждого такого класса ровно по одной точке. Пусть множество A_n получается поворотом всех точек множества A_0 на угол $2\pi n\alpha$, $n = \pm 1, \pm 2, \ldots$.

Так как все точки одного класса можно получить, поворачивая любую из них на угол $2\pi n\alpha$, $n = \pm 1, \pm 2, \ldots$, а в множестве A_0 собрано по одной точке из каждого класса, то поворачивая это множество, получим все точки окружности.

Очевидно, что $\bigcup\limits_{n=-\infty}^{\infty} A_n = [0, 2\pi]$. Предположим, что лебегова мера («длина») множества A_0 существует. Заметим, что тогда все множества A_n имеют ту же лебегову меру, так как получены из A_0 поворотом. И так как все эти множества не пересекаются, то мера их объединения равна сумме их мер:

$$2\pi = \lambda\left(\bigcup_{n=-\infty}^{\infty} A_n\right) = \sum_{n=-\infty}^{\infty} \lambda(A_n) = \sum_{n=-\infty}^{\infty} \lambda(A_0) = \infty.$$

Полученное противоречие означает, что лебегова мера, или длина множества A_0 ***не существует***.

Упражнение: какими свойствами «длины» (или меры Лебега) мы воспользовались в этом примере?

Раздел 4. Условная вероятность, независимость

4.1 Условная вероятность

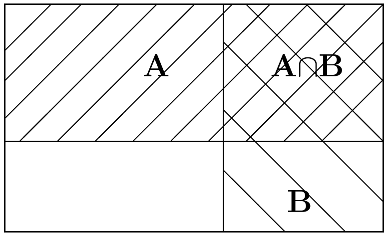
Пример 15. Кубик подбрасывается один раз. Известно, что выпало более трех очков. Какова при этом вероятность того, что выпало четное число очков?

В данном случае пространство элементарных исходов состоит из трех равновозможных элементарных исходов: $\Omega = \{4, 5, 6\}$, и событию $A = \{$выпало четное число очков$\}$ благоприятствуют 2 из них: $A = \{4, 6\}$. Поэтому $\mathsf{P}(A) = 2/3$.

Посмотрим на этот вопрос с точки зрения первоначального эксперимента. Пространство элементарных исходов при одном подбрасывании кубика состоит из шести точек: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Слова «известно, что выпало более трех очков» означают, что в эксперименте произошло событие $B = \{4, 5, 6\}$. Слова «какова при этом вероятность того, что выпало четное число очков?» означают, что нас интересует, в какой доле случаев при осуществлении B происходит и A. Вероятность события A, вычисленную в предположении, что нечто о результате эксперимента уже известно (событие B произошло), мы будем обозначать через $\mathsf{P}(A \big| B)$.

Мы хотим вычислить отношение числа исходов, благоприятствующих A внутри B (то есть благоприятствующих одновременно A и B), к числу исходов, благоприятствующих B.

$$\mathsf{P}(A \big| B) = \frac{2}{3} = \frac{2/6}{3/6} = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$



Какое отношение требуется вычислить, если элементарные исходы не являются равновозможными?

Определение 15. Условной вероятностью события A, при условии, что произошло событие B, называется число

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

Будем считать, что условная вероятность определена только в случае, когда $\mathsf{P}(B) > 0$.

Следующее свойство называется "теоремой умножения":

Теорема 6. $\mathsf{P}(A \cap B) = \mathsf{P}(B)\mathsf{P}(A \big| B) = \mathsf{P}(A)\mathsf{P}(B \big| A)$, *если соответствующие условные вероятности определены (то есть если* $\mathsf{P}(B) > 0$, $\mathsf{P}(A) > 0$*).*

Теорема умножения для большего числа событий:

Теорема 7. $\mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathsf{P}(A_1)\mathsf{P}(A_2 \big| A_1)\mathsf{P}(A_3 \big| A_1 \cap A_2) \cdot \ldots \cdot \mathsf{P}(A_n \big| A_1 \cap \ldots \cap A_{n-1})$, *если соответствующие условные вероятности определены.*

Доказать теорему 7 методом математической индукции.

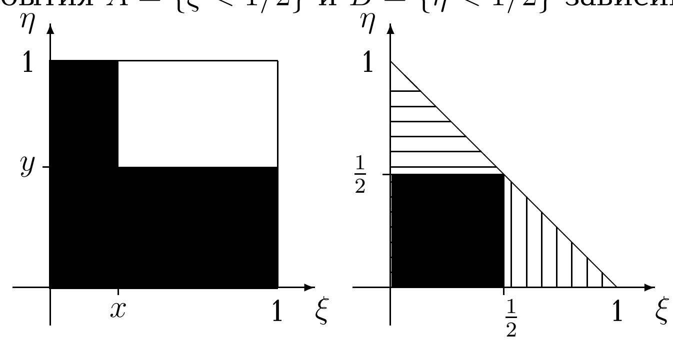
4.2 Независимость

Определение 16. События A и B называются ***независимыми***, если $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$.

Пример 16.

1. Точка с координатами ξ, η бросается наудачу в квадрат со стороной 1. Доказать, что для любых $x, y \in \mathbb{R}$ события $A = \{\xi < x\}$ и $B = \{\eta < y\}$ независимы.

2. Точка с координатами ξ, η бросается наудачу в треугольник с вершинами (1,0), (0,0) и (0,1). Доказать, что события $A = \{\xi < 1/2\}$ и $B = \{\eta < 1/2\}$ зависимы.



1. Рассмотрим $x, y \in [0,1]$ (***разобрать остальные случаи***). Видим, что $\mathsf{P}(A) = x$, $\mathsf{P}(B) = y$, $\mathsf{P}(A \cap B) = x \cdot y$, так что события $A = \{\xi < x\}$ и $B = \{\eta < y\}$ независимы.

2. На рисунке событие A заштриховано зеленым, событие B — синим. Видим, что $\mathsf{P}(A) = 3/4$, $\mathsf{P}(B) = 3/4$, $\mathsf{P}(A \cap B) = 1/2 \neq (3/4)^2$, так что события $A = \{\xi < 1/2\}$ и $B = \{\eta < 1/2\}$ зависимы. ***Доказать, что при $x \notin [0,1]$ или $y \notin [0,1]$ события $A = \{\xi < x\}$ и $B = \{\eta < y\}$ независимы.***

Замечание 7. Если события A и B несовместны, то они независимы, если и только если $\mathsf{P}(A) = 0$ или $\mathsf{P}(B) = 0$.

Доказать!!

Следствие 2. *Если $\mathsf{P}(B) > 0$, то события A и B независимы $\iff$ $\mathsf{P}(A \big| B) = \mathsf{P}(A)$.*
Если $\mathsf{P}(A) > 0$, то события A и B независимы $\iff$ $\mathsf{P}(B \big| A) = \mathsf{P}(B)$.

Доказать следствие, пользуясь определением условной вероятности.

Лемма 2. *Если события A и B независимы, то независимы и события A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Так как $A = A \cap B \bigcup A \cap \overline{B}$, и события $A \cap B$ и $A \cap \overline{B}$ несовместны, то $\mathsf{P}(A) = \mathsf{P}(A \cap B) + \mathsf{P}(A \cap \overline{B})$. Поэтому $\mathsf{P}(A \cap \overline{B}) = \mathsf{P}(A) - \mathsf{P}(A \cap B) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(\overline{B})$.
Вывести отсюда все остальные утверждения. $\square$

Определение 17. События $A_1, \dots, A_n$ называются *независимыми в совокупности*, если для любого набора $1 \leqslant i_1, \dots, i_k \leqslant n$

$$\mathsf{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathsf{P}(A_{i_1}) \dots \mathsf{P}(A_{i_k}). \tag{6}$$

Замечание 8. Если события $A_1, \dots, A_n$ независимы в совокупности, то они попарно независимы, то есть любые два события A_i, A_j независимы. Достаточно в равенстве (6) взять $k = 2$. Обратное, как показывает следующий пример, неверно.

Пример 17 (Пример С. Н. Бернштейна).

Рассмотрим правильный тетраэдр, 3 грани которого окрашены, соответственно, в красный, синий, зеленый цвета, а четвертая грань содержит все три цвета. Событие A (B, C) означает, что выпала грань, содержащая красный (синий, зеленый) цвета.

Вероятность каждого из этих событий равна 1/2, так как каждый цвет есть на двух гранях из четырех. Вероятность пересечения любых двух из них равна 1/4, так как только одна грань содержит два цвета. А так как $1/4 = 1/2 \cdot 1/2$, то все события попарно независимы.

Но вероятность пересечения всех трех тоже равна 1/4, а не 1/8, то есть события не являются независимыми в совокупности.

Заметьте, что равенство (6) выполнено для $k = 2$, но не выполнено для $k = 3$.

4.3 Формула полной вероятности

Пример 18. Есть 3 завода, производящих одну и ту же продукцию. При этом 1-й завод производит 25%, 2-й завод — 35% и 3-й завод — 40% всей производимой продукции. Брак составляет 5% от продукции 1-го завода, 3% от продукции 2-го и 4% от продукции 3-го завода.

Вся продукция смешивается и поступает в продажу. Найти а) вероятность купить бракованное изделие; б) условную вероятность того, что купленное изделие изготовлено 1-м заводом, если это изделие бракованное.

Первая вероятность равна доле бракованных изделий в объеме всей продукции, то есть $0.05{\cdot}0.25 + 0.03{\cdot}0.35 + 0.04{\cdot}0.4$. Вторая вероятность равна доле брака 1-го завода среди всего брака, то есть

$$\frac{0.05{\cdot}0.25}{0.05{\cdot}0.25 + 0.03{\cdot}0.35 + 0.04{\cdot}0.4}.$$

Определение 18. Набор попарно несовместных событий $H_1, H_2, \dots$ таких, что $\mathsf{P}(H_i) > 0$ для всех i и $\bigcup\limits_{i=1}^{\infty} H_i = \Omega$, называется *полной группой событий* или *разбиением пространства* Ω.

События $H_1, H_2, \ldots$, образующие полную группу событий, часто называют ***гипотезами***. При подходящем выборе гипотез для произвольного события A могут быть сравнительно просто вычислены $\mathsf{P}(A \big| H_i)$ (вероятность событию A произойти при выполнении «гипотезы» H_i) и собственно $\mathsf{P}(H_i)$ (вероятность выполнения «гипотезы» H_i). Как, используя эти данные, посчитать вероятность события A?

Теорема 8 (Формула полной вероятности).

Пусть $H_1, H_2, \ldots$ — полная группа событий. Тогда вероятность любого события A может быть вычислена по формуле:

$$\mathsf{P}(A) = \sum_{i=1}^{\infty} \mathsf{P}(H_i)\mathsf{P}(A \big| H_i).$$

Доказательство. Заметим, что $A = A \cap \Omega = A \cap \left(\bigcup\limits_{i=1}^{\infty} H_i\right) = \bigcup\limits_{i=1}^{\infty} A \cap H_i$, и события $A \cap H_1, A \cap H_2, \ldots$ попарно несовместны. Поэтому (используем в первом равенстве σ-аддитивность вероятностной меры (***а что это?***), а во втором — теорему умножения)

$$\mathsf{P}(A) = \sum_{i=1}^{\infty} \mathsf{P}(A \cap H_i) = \sum_{i=1}^{\infty} \mathsf{P}(H_i)\mathsf{P}(A \big| H_i). \qquad \square$$

4.4 Формула Байеса

Теорема 9 (Формула Байеса).

Пусть $H_1, H_2, \ldots$ — полная группа событий и A — некоторое событие положительной вероятности. Тогда условная вероятность того, что имело место событие H_k, если в результате эксперимента наблюдалось событие A, может быть вычислена по формуле:

$$\mathsf{P}(H_k \big| A) = \frac{\mathsf{P}(H_k)\mathsf{P}(A \big| H_k)}{\sum_{i=1}^{\infty} \mathsf{P}(H_i)\mathsf{P}(A \big| H_i)}.$$

Доказательство. По определению условной вероятности,

$$\mathsf{P}(H_k \big| A) = \frac{\mathsf{P}(H_k \cap A)}{\mathsf{P}(A)} = \frac{\mathsf{P}(H_k)\mathsf{P}(A \big| H_k)}{\sum_{i=1}^{\infty} \mathsf{P}(H_i)\mathsf{P}(A \big| H_i)}.$$

Последнее равенство следует из теоремы умножения и формулы полной вероятности. $\square$

Пример 19. Вернемся к примеру **18**. Изделие выбирается наудачу из всей произведенной продукции. Рассмотрим три гипотезы: $H_i = \{$изделие изготовлено i-м заводом$\}$, $i = 1, 2, 3$. Вероятности этих событий даны: $\mathsf{P}(H_1) = 0.25$, $\mathsf{P}(H_2) = 0.35$, $\mathsf{P}(H_3) = 0.4$. Пусть $A = \{$изделие оказалось бракованным$\}$. Даны также условные вероятности $P(A \big| H_1) = 0.05$, $P(A \big| H_2) = 0.03$, $P(A \big| H_3) = 0.04$.

Убедиться, что полученные нами вероятности в (а) и (б) совпадают с вероятностями, вычисленными по формуле полной вероятности и формуле Байеса.

Пример 20. Два стрелка подбрасывают монетку и выбирают, кто из них стреляет по мишени (одной пулей). Первый стрелок попадает по мишени с вероятностью 1, второй стрелок — с вероятностью 0.00001. Можно сделать два предположения об эксперименте: $H_1 = \{$стреляет 1-й стрелок$\}$ и $H_2 = \{$стреляет 2-й стрелок$\}$. Априорные (a'priori — «до опыта») вероятности этих гипотез одинаковы: $\mathsf{P}(H_1) = \mathsf{P}(H_2) = 1/2$.

Рассмотрим событие $A = \{$пуля попала в мишень$\}$. Известно, что $\mathsf{P}(A \big| H_1) = 1, \mathsf{P}(A \big| H_2) = 0.00001$. Поэтому вероятность пуле попасть в мишень $\mathsf{P}(A) = 1/2 \cdot 1 + 1/2 \cdot 0.00001$. Предположим, что событие A произошло. Какова теперь апостериорная (a'posteriori — «после опыта») вероятность каждой из гипотез H_i? Очевидно, что первая из этих гипотез много вероятнее второй (а именно, в 100000 раз). Действительно,

$$\mathsf{P}(H_1 \big| A) = \frac{1/2 \cdot 1}{1/2 \cdot 1 + 1/2 \cdot 0.00001} = \frac{1}{1 + 0.00001}; \qquad \mathsf{P}(H_2 \big| A) = \frac{1/2 \cdot 0.00001}{1/2 \cdot 1 + 1/2 \cdot 0.00001} = \frac{0.00001}{1 + 0.00001}.$$

Раздел 5. Схема Бернулли

5.1 Распределение числа успехов в n испытаниях

Определение 19. *Схемой Бернулли* называется последовательность независимых испытаний, в каждом из которых возможны лишь два исхода — «успех» и «неудача», при этом «успех» в одном испытании происходит с вероятностью $p \in [0, 1]$, «неудача» — с вероятностью $q = 1 - p$.

Теорема 10 (Формула Бернулли).

Обозначим через ν_n число успехов в n испытаниях схемы Бернулли. Тогда для любого $k = 0, 1, \ldots, n$

$$\mathsf{P}(\nu_n = k) = C_n^k \; p^k \; (1-p)^{n-k} = C_n^k \; p^k \; q^{n-k}.$$

Доказательство. Событие $A = \{\nu_n = k\}$ означает, что в n испытаниях схемы Бернулли произошло ровно k успехов. Рассмотрим один из благоприятствующих событию A элементарных исходов: $(\underbrace{у, у, \ldots, у}_{k}, \underbrace{н, \ldots, н}_{n-k})$. Здесь буквами *«у»* и *«н»* обозначены, соответственно, успешный и неудачный результаты испытаний. Поскольку испытания независимы, вероятность такого элементарного исхода (первые k испытаний завершились успехом, остальные неудачей) равна $p^k \; (1-p)^{n-k}$.

Другие благоприятствующие событию A элементарные исходы отличаются от рассмотренного выше лишь расположением k успехов на n местах. Есть ровно C_n^k способов расположить k успехов на n местах. Поэтому событие A состоит из C_n^k элементарных исходов, вероятность каждого из которых равна $p^k \; (1-p)^{n-k}$.

$\square$

Определение 20. Набор чисел $\left\{C_n^k \; p^k \; (1-p)^{n-k}, k = 0, 1, \ldots, n\right\}$ называется *биномиальным распределением вероятностей* и обозначается $\mathbf{B}_{n,p}$ или $\mathbf{B}(n, p)$.

5.2 Наиболее вероятное число успехов

По формуле Бернулли, событие «произошло 0 успехов в n испытаниях» имеет вероятность q^n, 1 успех — вероятность $n\,p\,q^{n-1}$ и т.д. Какое же число успехов *наиболее вероятно*? Иначе говоря, при каком k достигается максимум $\mathsf{P}(\nu_n = k)$?

Чтобы выяснить это, сравним отношение $\mathsf{P}(\nu_n = k)$ и $\mathsf{P}(\nu_n = k-1)$ с единицей.

$$\frac{\mathsf{P}(\nu_n = k)}{\mathsf{P}(\nu_n = k-1)} = \frac{n!}{k!(n-k)!}\frac{(k-1)!(n-k+1)!}{n!}\frac{p^k q^{n-k}}{p^{k-1}q^{n-k+1}} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n-k+1)p}{kq} - 1 = 1 + \frac{np+p-k}{kq}.$$

Видим, что

(a) $\mathsf{P}(\nu_n = k) > \mathsf{P}(\nu_n = k-1)$ при $np+p-k>0$, то есть при $k<np+p$;

(b) $\mathsf{P}(\nu_n = k) < \mathsf{P}(\nu_n = k-1)$ при $np+p-k<0$, то есть при $k>np+p$;

(c) $\mathsf{P}(\nu_n = k) = \mathsf{P}(\nu_n = k-1)$ при $np+p-k=0$, что возможно лишь если $np+p$ — целое число.

Рассмотрим два случая: $np+p\in\mathbb{Z}$ и $np+p\notin\mathbb{Z}$. В первом случае пусть $k_0 = np+p$. Из полученных выше неравенств сразу следует, что

$$\ldots \; < \; \mathsf{P}(\nu_n = k_0-2) \overset{\textbf{(a)}}{<} \mathsf{P}(\nu_n = k_0-1) \overset{\textbf{(c)}}{=} \mathsf{P}(\nu_n = k_0) \overset{\textbf{(b)}}{>} \mathsf{P}(\nu_n = k_0+1) \; > \; \ldots$$

Во втором случае пусть $k_0 = [np+p]$ (целая часть числа $np+p$, то есть наибольшее целое число, не превосходящее $np+p$). Из неравенств **(a),(b)** следует, что

$$\ldots \; < \; \mathsf{P}(\nu_n = k_0-2) \overset{\textbf{(a)}}{<} \mathsf{P}(\nu_n = k_0-1) \overset{\textbf{(a)}}{<} \mathsf{P}(\nu_n = k_0) \overset{\textbf{(b)}}{>} \mathsf{P}(\nu_n = k_0+1) \; > \; \ldots$$

Действительно, неравенство $\mathsf{P}(\nu_n = k_0) > \mathsf{P}(\nu_n = k_0+1)$, например, следует из **(b)**, примененного для $k = k_0+1 > np+p$.

Видим, что в зависимости от того, является число $np+p$ целым или нет, имеется либо два равновероятных «наиболее вероятных» числа успехов $k_0 = np+p$ и $k_0-1 = np+p-1$, либо одно «наиболее вероятное» число успехов $k_0 = [np+p]$.

Сформулируем уже доказанное утверждение в виде теоремы.

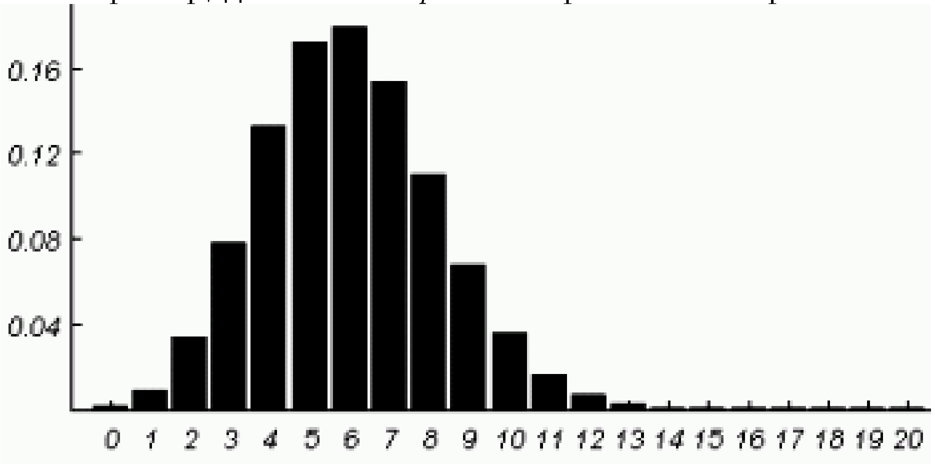
Теорема 11. *В n испытаниях схемы Бернулли с вероятностью успеха p наиболее вероятным числом успехов является*

а) единственное число $k_0 = [np+p]$, если число $np+p$ не целое;

б) два числа $k_0 = np+p$ и $k_0-1 = np+p-1$, если число $np+p$ целое.

Упражнение 5. Рассмотреть график вероятностей биномиального распределения и увидеть утверждение теоремы на графике.

Например, для $n = 30$ и $p = 0.2$ вероятности нарисованы ниже.



Пример 21. Если $p = q = 1/2$, то при четном числе испытаний n число $np + p = n/2 + 1/2 \notin \mathbb{Z}$ — не целое, так что наиболее вероятным является единственное число успехов $[n/2 + 1/2] = n/2$. Что совершенно понятно, так как есть нечетное число возможностей — получить $0, 1, \ldots, n$ успехов, причем вероятности получить k и $n-k$ успехов одинаковы.

При нечетном же числе испытаний n число $np + p = n/2 + 1/2 \in \mathbb{Z}$ — целое, так что наиболее вероятными (и одинаково вероятными) являются два числа успехов $n/2 + 1/2$ и $n/2 - 1/2$.

5.3 Номер первого успешного испытания

Рассмотрим схему Бернулли с вероятностью успеха p в одном испытании. Испытания проводятся до появления первого успеха. Введем величину τ, принимающую значения из $\{1, 2, 3, \ldots\}$, равную ***номеру первого успешного испытания***.

Теорема 12. *Вероятность того, что первый успех произойдет в испытании с номером $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$, равна*

$$\mathsf{P}(\tau = k) = p\ q^{k-1}.$$

Доказательство. Действительно, $\mathsf{P}(\tau = k) = \mathsf{P}(\underbrace{н,\ н,\ \ldots,\ н}_{k-1}, у) = p\ q^{k-1}$. □

Определение 21. Набор чисел $\{p\, q^{k-1},\ k = 1, 2, 3, \ldots\}$ называется ***геометрическим распределением вероятностей*** и обозначается $\mathbf{G}_p$ или $\mathbf{G}(p)$.

Геометрическое распределение вероятностей обладает интересным свойством, которое можно назвать свойством «нестарения». Пусть величина τ обозначает, скажем, время безотказной работы (измеряемое целым числом часов) некоторого устройства. Предположим, что для величины τ вероятность принять любое свое значение k в точности равна $p\, q^{k-1}$. Справедливо следующее утверждение.

Теорема 13. *Пусть* $\mathsf{P}(\tau = k) = p\, q^{k-1}$ *для любого* $k \in \mathbb{N}$. *Тогда для произвольных* $n, k \geqslant 0$

$$\mathsf{P}(\tau > n + k \mid \tau > n) = \mathsf{P}(\tau > k).$$

Данному равенству можно придать следующее звучание: *если известно, что устройство уже проработало без отказа n часов, то вероятность ему работать еще не менее k часов точно такая же, как вероятность проработать не менее k часов для нового устройства.*

Можно прочесть эту формулу и так: *вероятность работающему устройству проработать еще сколько-то часов не зависит от того момента, когда мы начали отсчет времени, или от того, сколько уже работает устройство* :-).

Доказательство. По определению условной вероятности,

$$\mathsf{P}(\tau > n + k \mid \tau > n) = \frac{\mathsf{P}(\tau > n + k, \tau > n)}{\mathsf{P}(\tau > n)} = \frac{\mathsf{P}(\tau > n + k)}{\mathsf{P}(\tau > n)}. \tag{7}$$

Последнее равенство следует из того, что событие $\{\tau > n + k\}$ влечет событие $\{\tau > n\}$, так что пересечение этих событий есть $\{\tau > n + k\}$. Найдем для произвольного $m \geqslant 0$ вероятность $\mathsf{P}(\tau > m)$.

$$\mathsf{P}(\tau > m) = \sum_{i=m+1}^{\infty} \mathsf{P}(\tau = i) = \sum_{i=m+1}^{\infty} p\, q^{i-1} = \frac{p\, q^m}{1-q} = q^m.$$

Можно также заметить, что событие $\{\tau > m\}$ означает, что в схеме Бернулли первые m испытаний завершились «неудачами», а это событие имеет вероятность как раз q^m.

Возвращаясь к (7), получим $$\mathsf{P}(\tau > n + k \mid \tau > n) = \frac{\mathsf{P}(\tau > n + k)}{\mathsf{P}(\tau > n)} = \frac{q^{n+k}}{q^n} = q^k = \mathsf{P}(\tau > k).$$ □

5.4 Приближение гипергеометрического распределения биномиальным

Рассмотрим урну, содержащую N шаров, из которых K шаров — белые, а оставшиеся $N - K$ шаров — черные. Из урны наудачу (без возвращения) выбираются n шаров. Вероятность $P_{N,K}(n, k)$ того, что будет выбрано ровно k белых и $n - k$ черных шаров, находится по формуле (см. определение **8** гипергеометрического распределения вероятностей):

$$P_{N,K}(n,k) = \frac{C_K^k\ C_{N-K}^{n-k}}{C_N^n}.$$

Если число шаров в урне очень велико, то извлечение одного, двух, трех шаров почти не меняет пропорцию белых и черных шаров в урне, так что вероятности $P_{N,K}(n,k)$ не очень отличаются от вероятностей в процедуре выбора *с возвращением*:

$$\mathsf{P}(\text{получить ровно } k \text{ белых шаров при выборе } n \text{ шаров с возвращением}) = C_n^k \left(\frac{K}{N}\right)^k \left(1 - \frac{K}{N}\right)^{n-k}.$$

Сформулируем и докажем нашу первую предельную теорему.

Теорема 14. *Если* $N \to \infty$ *и* $K \to \infty$ *так, что* $K/N \to p \in (0,1)$, *то для любых фиксированных* n, $0 \leqslant k \leqslant n$

$$P_{N,K}(n,k) = \frac{C_K^k\ C_{N-K}^{n-k}}{C_N^n} \ \to\ C_n^k\ p^k\ (1-p)^{n-k}.$$

Доказательство. Нам понадобятся следующие определение и свойство.

Определение 22. Говорят, что последовательности a_n и b_n асимптотически эквивалентны, и пишут $a_n \sim b_n$, если

$$\frac{a_n}{b_n} \to 1 \text{ при } n \to \infty.$$

Свойство 4. *Следующие последовательности асимптотически эквивалентны:* $C_K^k \sim \dfrac{K^k}{k!}$ *при* $K \to \infty$.

Доказательство. Действительно, рассмотрим отношение членов этих последовательностей

$$\frac{C_K^k\ k!}{K^k} = \frac{K!\ k!}{k!\ (K-k)!\ K^k} = \frac{K(K-1)\ldots(K-k+1)}{K^k}\ \to\ 1 \text{ при } K \to \infty,$$

поскольку предел произведения конечного числа k последовательностей, сходящихся к 1, равен 1. □

Следствие 3. $C_N^n \sim \dfrac{N^n}{n!}$ *при* $N \to \infty$, $C_{N-K}^{n-k} \sim \dfrac{(N-K)^{n-k}}{(n-k)!}$ *при* $N - K \to \infty$.

Упражнение 6. Почему $N - K$ стремится к бесконечности?

Воспользуемся теперь свойством **4** и следствием **3**:

$$P_{N,K}(n,k) = \frac{C_K^k\ C_{N-K}^{n-k}}{C_N^n}\ \sim\ \frac{K^k}{k!}\frac{(N-K)^{n-k}}{(n-k)!}\frac{n!}{N^n} = C_n^k\ \frac{K^k}{N^k}\ \frac{(N-K)^{n-k}}{N^{n-k}} = C_n^k\ \left(\frac{K}{N}\right)^k\ \left(1-\frac{K}{N}\right)^{n-k} \to\ C_n^k\ p^k\ (1-p)^{n-k}.$$

Мы получили, что $P_{N,K}(n,k)$ асимптотически эквивалентно выражению, сходящемуся к $C_n^k\ p^k\ (1-p)^{n-k}$ при стремлении N (и K в зависимости от N) к бесконечности. Осталось вспомнить и доказать свойство:

Свойство 5. *Пусть $a_n \sim b_n$ и существует $\lim\limits_{n\to\infty} b_n$. Тогда существует и $\lim\limits_{n\to\infty} a_n$, и эти пределы совпадают:*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

Упражнение 7. Доказать свойство **5**.

По свойству **5**, при $N \to \infty$ и $K \to \infty$ так, что $K/N \to p \in (0,1)$, существует $\lim P_{N,K}(n,k) = C_n^k\ p^k\ (1-p)^{n-k}$. □

5.5 Независимые испытания с несколькими исходами

Рассмотрим следующий пример, когда из двух очень похожих вопросов на один можно ответить, пользуясь формулой Бернулли, а для другого этой формулы оказывается недостаточно:

Пример 22. Игральная кость подбрасывается 15 раз. Найти вероятности следующих событий:
а) выпадет ровно 10 шестерок; б) выпадет ровно 10 шестерок и три единицы.

Р е ш е н и е:
а) есть 15 испытаний схемы Бернулли с вероятностью успеха 1/6 (выпадение шестерки). Вероятность десяти успехов в 15 испытаниях равна $C_{15}^{10}\left(\frac{1}{6}\right)^{10}\left(1-\frac{1}{6}\right)^5$;
б) здесь каждое испытание имеет три, а не два исхода: выпадение шестерки, выпадение единицы, выпадение остальных граней. Воспользоваться формулой для числа успехов в схеме Бернулли не удается — перед нами уже не схема Бернулли.

Осталось изобрести формулу для подсчета вероятности каждому исходу в нескольких независимых испытаниях выпасть нужное число раз, если в одном испытании возможно не два, а более исходов.

Пусть в одном испытании возможны m исходов. Обозначим их цифрами $1, 2, \ldots, m$. Пусть исход i в одном испытании случается с вероятностью p_i, $1 \leqslant i \leqslant m$, и $\sum_1^m p_i = 1$.
Обозначим через $P(n_1, \ldots, n_m)$ вероятность того, что в $n = n_1 + \ldots + n_m$ независимых испытаниях исход 1 появился n_1 раз, исход 2 — n_2 раз, ..., исход m — n_m раз.

Теорема 15. *Для любого n и любых целых $n_1 \geqslant 0, \ldots, n_m \geqslant 0$ таких, что $n_1 + \ldots + n_m = n$, верна формула:*

$$P(n_1, \ldots, n_m) = \frac{n!}{n_1! \ldots n_m!}\, p_1^{n_1} \cdot \ldots \cdot p_m^{n_m}.$$

Доказательство. Рассмотрим один элементарный исход, благоприятствующий выпадению n_1 единиц, n_2 двоек, ..., n_m раз m-ок:

$$(\underbrace{1, \ldots, 1}_{n_1}, \underbrace{2, \ldots, 2}_{n_2}, \ldots, \underbrace{m, \ldots, m}_{n_m})$$

— результат n экспериментов, когда все нужные исходы появились в некотором заранее заданом порядке. Вероятность такого результата n независимых испытаний равна $p_1^{n_1} \cdot \ldots \cdot p_m^{n_m}$.

Все остальные благоприятные исходы отличаются лишь расположением чисел $1, 2, \ldots, m$ на n местах. Число таких исходов равно числу способов расставить на n местах n_1 единиц, n_2 двоек, ..., n_m чисел m, то есть

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot C_{n-n_1-n_2}^{n_3} \cdot \ldots \cdot C_{n-n_1-\ldots-n_{m-1}}^{n_m} = \textit{проверить, что это так!} = \frac{n!}{n_1! \ \ldots \ n_m!} \qquad \square$$

Теперь мы можем вернуться к примеру **22**(б) и выписать ответ: так как вероятности выпадения шестерки и единицы равны 1/6, а вероятность третьего исхода (выпали любые другие грани) равна 4/6, то вероятность получить 10 шестерок, 3 единицы и еще 2 других очка равна

$$P(10, 3, 2) = \frac{15!}{10!\ 3!\ 2!}\ \frac{1}{6^{10}}\ \frac{1}{6^3}\ \left(\frac{4}{6}\right)^2.$$

5.6 Теорема Пуассона для схемы Бернулли

Предположим, нам нужна вероятность получить не менее десяти успехов в 1000 испытаний схемы Бернулли с вероятностью успеха 0.003. Вероятность этого события равна любому из следующих выражений:

$$\sum_{k=10}^{1000} C_{1000}^k\ (0.003)^k\ (0.997)^{1000-k}\ =\ 1 - \sum_{k=0}^{9} C_{1000}^k\ (0.003)^k\ (0.997)^{1000-k},$$

и вычисление даже одного слагаемого в каждом из этих выражений весьма проблематично.

Сформулируем теорему о приближенном вычислении вероятности какого-либо числа успехов в большом числе испытаний схемы Бернулли с маленькой вероятностью успеха. Термин «большое число» должен означать $n \to \infty$. Если при этом $p = p_n \not\to 0$, то, очевидно, вероятность получить любое конечное число успехов при растущем числе испытаний стремится к нулю. Необходимо чтобы вероятность успеха $p = p_n \to 0$ одновременно с ростом числа испытаний. Но от испытания к испытанию вероятность успеха меняться не может (см. определение схемы Бернулли).

Поэтому рассмотрим «схему серий»: есть

одно испытание	$\circ$	с вероятностью успеха p_1
два испытания	$\circ, \circ$	с вероятностью успеха p_2
...		...
n испытаний	$\circ, \ldots, \circ$	с вероятностью успеха p_n
...		...

Вероятность успеха меняется не внутри одной серии испытаний, а от серии к серии, когда меняется общее число испытаний. Обозначим через ν_n число успехов в n-й серии испытаний.

Теорема 16 (Теорема Пуассона).
Пусть $n \to \infty$, $p_n \to 0$ так, что $np_n \to \lambda > 0$. Тогда для любого $k \geqslant 0$ вероятность получить k успехов в n испытаниях схемы Бернулли с вероятностью успеха p_n стремится к величине $\dfrac{\lambda^k}{k!}\, e^{-\lambda}$:

$$\mathsf{P}(\nu_n = k) = C_n^k\, p_n^k\, (1-p_n)^{n-k} \;\to\; \frac{\lambda^k}{k!}\, e^{-\lambda} \;\; \textit{при} \;\; n \to \infty, p_n \to 0 \;\; \textit{так, что}\; np_n \to \lambda > 0.$$

Доказательство. Положим $\lambda_n = n \cdot p_n \to \lambda > 0$. По свойству **4**, $C_n^k \sim \dfrac{n^k}{k!}$ при фиксированном k и при $n \to \infty$. Тогда

$$C_n^k\, p_n^k\, (1-p_n)^{n-k} = C_n^k\, \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \sim \frac{\cancel{n^k}}{k!} \frac{\lambda_n^k}{\cancel{n^k}} \underbrace{\left(1-\frac{\lambda_n}{n}\right)^n}_{\substack{\downarrow \\ e^{-\lambda}}} \underbrace{\left(1-\frac{\lambda_n}{n}\right)^{-k}}_{\substack{\downarrow \\ 1}} \;\to\; \frac{\lambda^k}{k!}\, e^{-\lambda}. \tag{8}$$

В (8) мы использовали свойства $\lambda_n^k \to \lambda^k$ и $\left(1-\dfrac{\lambda_n}{n}\right)^n \to e^{-\lambda}$. Докажем последнее свойство:

$$\ln\left(1-\frac{\lambda_n}{n}\right)^n = n\ln\left(1-\frac{\lambda_n}{n}\right) = n\left(-\frac{\lambda_n}{n} + O\left(\frac{\lambda_n^2}{n^2}\right)\right) \to -\lambda.$$

Для доказательства теоремы осталось в формуле (8) воспользоваться свойством **5**. □

Определение 23. Пусть $\lambda > 0$ — некоторая постоянная. Набор чисел $\left\{\dfrac{\lambda^k}{k!}\, e^{-\lambda}, k = 0, 1, 2, \ldots\right\}$ называется *распределением Пуассона с параметром λ*.

Пользуясь теоремой **16**, можно приближенно посчитать вероятность получить не менее десяти успехов в 1000 испытаний схемы Бернулли с вероятностью успеха 0.003, с вычисления которой мы начали. Поскольку $n = 1000$ «велико», а $p_n = 0.003$ «мало», то, взяв $\lambda = np_n = 3$, можно написать приближенное равенство

$$1 - \sum_{k=0}^{9} C_{1000}^k \ (0.003)^k \ (0.997)^{1000-k} \approx 1 - \sum_{k=0}^{9} \frac{3^k}{k!} \ e^{-3} = \sum_{k=10}^{\infty} \frac{3^k}{k!} \ e^{-3} = \text{ табличное значение } \overline{\Pi}_3(10) \ \approx 0,001. \qquad (9)$$

Осталось решить, а достаточно ли $n = 10^3$ «велико», а $p_n = 0.003$ «мало», чтобы заменить точную вероятность $\mathsf{P}(\nu_n = k)$ на приближенное значение $\dfrac{\lambda^k}{k!} \ e^{-\lambda}$. Для этого нужно уметь оценивать разницу между этими двумя вероятностями.

Следующую очень полезную теорему мы докажем в конце курса.

Теорема 17 (Теорема Пуассона с оценкой погрешности).

Пусть $A \subseteq \{0, 1, 2, \ldots, n\}$ — произвольное множество целых неотрицательных чисел, ν_n — число успехов в n испытаниях схемы Бернулли с вероятностью успеха p, $\lambda = n \cdot p$. Тогда

$$\left| \mathsf{P}(\nu_n \in A) - \sum_{k \in A} \frac{\lambda^k}{k!} \ e^{-\lambda} \right| = \left| \ \sum_{k \in A} C_n^k \ p^k \ (1-p)^{n-k} - \sum_{k \in A} \frac{\lambda^k}{k!} \ e^{-\lambda} \right| \leqslant np^2 = \frac{\lambda^2}{n}.$$

Таким образом, теорема **17** предоставляет нам возможность самим решать, достаточно ли n «велико», а p «мало», руководствуясь полученной величиной погрешности. Какова же погрешность в формуле (9)?

$$\left| \mathsf{P}(\nu_{1000} \geqslant 10) - \sum_{k=10}^{\infty} \frac{3^k}{k!} \ e^{-3} \right| = \left| \ \left(1 - \sum_{k=0}^{9} C_{1000}^k \ (0.003)^k \ (0.997)^{1000-k}\right) - \sum_{k=10}^{\infty} \frac{3^k}{k!} \ e^{-3} \right| \leqslant np^2 = 0,009.$$

Погрешность не более 0,009 (при вероятности около 0,001 :-)). Во всяком случае, можно утверждать, что искомая вероятность никак не больше, чем 0,01 = 0,001 + 0,009.

Раздел 6. Случайные величины и их распределения

6.1 Случайные величины

Мы уже видели, что для очень многих экспериментов нет никаких различий в подсчете *вероятностей* событий, тогда как элементарные исходы в этих экспериментах очень различаются. Но нас и должны интересовать именно вероятности событий, а не структура пространства элементарных исходов. Поэтому пора во всех таких «похожих» экспериментах вместо самых разных элементарных исходов использовать, например, числа. То есть ввести соответствие (иначе говоря, отображение) между элементарными исходами и вещественными числами (с ними удобно работать).

Пусть имеется случайный эксперимент и задано вероятностное пространство $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle$.

Определение 24. Функция $\xi : \Omega \to \mathbb{R}$ называется *случайной величиной*, если для любого $x \in \mathbb{R}$ множество $\{\xi < x\} = \{\omega \ : \ \xi(\omega) < x\}$ является событием, то есть принадлежит σ-алгебре событий $\mathcal{F}$.

Замечание 9. Читатель, не желающий забивать себе голову абстракциями, связанными с σ-алгебрами событий и с измеримостью, может смело считать, что любое множество элементарных исходов есть событие, и, следовательно, случайная величина есть *произвольная* функция из Ω в $\mathbb{R}$. Никаких неприятностей на практике это обычно не влечет, так что все дальнейшее в этом параграфе можно пропустить. Полезно, тем не менее, помнить: каждая такая «уступка» себе существенно снижает ваши адаптивные способности к жизни.

Определение 25. Будем говорить, что функция $\xi : \Omega \to \mathbb{R}$ является $\mathcal{F}$-*измеримой*, если $\{\omega \ : \ \xi(\omega) < x\}$ принадлежит $\mathcal{F}$ для любого $x \in \mathbb{R}$.

Итак, случайная величина есть $\mathcal{F}$-измеримая функция, ставящая в соответствие каждому элементарному исходу $\omega \in \Omega$ число $\xi(\omega) \in \mathbb{R}$.

Пример 23. Подбрасываем 1 раз кубик. Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$, и две функции из Ω в $\mathbb{R}$ заданы так: $\xi(\omega) = \omega$, $\eta(\omega) = \omega^2$.

- Если $\mathcal{F}$ есть множество *всех* подмножеств Ω, то ξ и η являются случайными величинами, поскольку любое множество элементарных исходов принадлежит $\mathcal{F}$, в том числе и $\{\omega \ : \ \xi(\omega) < x\}$ или $\{\omega \ : \ \eta(\omega) < x\}$. Можно записать соответствие между значениями случайных величин ξ и η и вероятностями принимать эти значения в виде «таблицы распределения вероятностей» или, коротко, «таблицы распределения»:

ξ	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

η	1	4	9	16	25	36
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Здесь $\frac{1}{6} = \mathsf{P}(\xi = 1) = \ldots = \mathsf{P}(\xi = 6) = \mathsf{P}(\eta = 1) = \ldots = \mathsf{P}(\eta = 36)$.

- Пусть σ-алгебра событий $\mathcal{F}$ состоит всего из четырех множеств: $\mathcal{F} = \big\{\Omega, \varnothing, \{1,3,5\}, \{2,4,6\}\big\}$, то есть событием является, кроме достоверного и невозможного событий, выпадение четного (соответственно, нечетного) числа очков. Убедимся, что при такой «бедной» σ-алгебре ни ξ, ни η не являются случайными величинами, так как эти функции не $\mathcal{F}$-измеримы. Возьмем (например) $x = 3{,}967$. Видим, что $\{\omega \in \Omega \ : \ \xi(\omega) < 3{,}967\} = \{1,2,3\} \notin \mathcal{F}$ и $\{\omega \in \Omega \ : \ \eta(\omega) < 3{,}967\} = \{1\} \notin \mathcal{F}$.

  Упражнение. Описать класс всех функций, измеримых относительно σ-алгебры $\mathcal{F} = \big\{\Omega, \varnothing, \{1,3,5\}, \{2,4,6\}\big\}$.

- Пусть σ-алгебра событий $\mathcal{F}$ есть тривиальная σ-алгебра : $\mathcal{F} = \{\Omega, \varnothing\}$.

  Доказать, что ξ и η не являются случайными величинами, так как эти функции не $\mathcal{F}$-измеримы.

  Доказать, что измеримы относительно тривиальной σ-алгебры только функции вида $\xi(\omega) = c$ (постоянные).

Теперь попробуем понять, зачем нужна $\mathcal{F}$-измеримость и почему требуется, чтобы $\{\omega \ : \ \xi(\omega) < x\}$ являлось событием. Если задана случайная величина ξ, нам может потребоваться вычислить вероятности типа $\mathsf{P}(\xi = 5) = \mathsf{P}\{\omega : \xi(\omega) = 5\}$, $\mathsf{P}(\xi \in [-3,7])$, $\mathsf{P}(\xi \geqslant 3{,}2)$, $\mathsf{P}(\xi < 0)$ (и вообще самые разные вероятности попадания в различные множества на прямой). Это возможно только если множества, стоящие под знаком вероятности, являются событиями (напомню, что вероятность есть функция из σ-алгебры событий в [0,1]).

Но если потребовать, чтобы $A_x = \{\omega \ : \ \xi(\omega) < x\}$ было событием при любом x, то мы из свойств σ-алгебры сразу получим, что

$$\begin{aligned} &\text{и} \qquad \overline{A}_x = \{\omega \ : \ \xi(\omega) \geqslant x\} \text{ — событие, и } \{\omega \ : \ x_1 \leqslant \xi(\omega) < x_2\} = A_{x_2} \backslash A_{x_1} \text{ — событие,} \\ &\text{и} \qquad B_x = \{\omega \ : \ \xi(\omega) \leqslant x\} = \bigcap_{n=1}^{\infty} A_{x+\frac{1}{n}} \text{ — событие, и } \{\omega \ : \ \xi(\omega) = x\} = B_x \backslash A_x \text{ — событие,} \end{aligned} \tag{10}$$

и т.д., и т.п. (операции пересечения, объединения, дополнения событий не выводят из класса событий).

Можно потребовать в определении **24** чего-нибудь другого. Например, чтобы событием было попадание в любой интервал: $\{\omega \ : \ \xi(\omega) \in (a,b)\} \in \mathcal{F}$ для любых $a < b$. Или чтобы $\{\omega \ : \ \xi(\omega) \geqslant x\}$ было событием для любого x. Любое такое определение эквивалентно исходному.

Замечание 10. Те, кто не поленился прочесть про борелевскую σ-алгебру в разделе **3.3**, могут сформулировать все наши потребности так: мы хотим, чтобы попадание ξ в любое борелевское множество являлось событием. Мы могли это потребовать в определении, но ограничились эквивалентным условием, чтобы попадание в любой открытый интервал $(-\infty, x)$ было событием. Эти условия эквивалентны, поскольку борелевская σ-алгебра порождается интервалами, что мы еще раз показали в формулах (10).

Опишем различные типы распределений случайных величин. Под ***распределением*** случайной величины мы будем понимать соответствие

либо (чаще) «значение случайной величины $\leftrightarrow$ вероятность принимать это значение»,
«множество на прямой $\leftrightarrow$ вероятность случайной величине попасть в это множество».

6.2 Дискретные распределения

Определение 26. Говорят, что случайная величина ξ имеет *дискретное* распределение, если существует конечный или счетный набор чисел $\{a_1, a_2, \dots\}$ такой, что:

а) $p_i = \mathsf{P}(\xi = a_i) > 0$ для всех i; б) $\sum\limits_{i=1}^{\infty} p_i = 1$.

То есть случайная величина ξ имеет дискретное распределение, если она принимает не более чем счетное число значений.

Определение 27. Если случайная величина ξ имеет дискретное распределение, назовем *таблицей распределения* соответствие $a_i \leftrightarrow p_i$, которое чаще всего рисуют так:

ξ	a_1	a_2	a_3	$\dots$
P	p_1	p_2	p_3	$\dots$

Примеры дискретных распределений

Вырожденное распределение.

Говорят, что случайная величина ξ имеет вырожденное распределение с параметром a, и пишут $\xi \in \mathbf{I}_a$, если ξ принимает единственное значение a с вероятностью 1, то есть $\mathsf{P}(\xi = a) = 1$. Таблица распределения ξ имеет вид

ξ	a
P	1

Распределение Бернулли.

Говорят, что случайная величина ξ имеет распределение Бернулли с параметром p, и пишут $\xi \in \mathbf{B}_p$, если ξ принимает значения 1 и 0 с вероятностями p и $1-p$, соответственно. Случайная величина ξ с таким распределением равна *числу успехов в одном испытании схемы Бернулли с вероятностью успеха* p (0 успехов или 1 успех). Таблица распределения ξ имеет вид

ξ	0	1
P	$1-p$	p

Биномиальное распределение.

Говорят, что случайная величина ξ имеет биномиальное распределение с параметрами n и p, где $0 \leqslant p \leqslant 1$, и пишут $\xi \in \mathbf{B}_{n,p}$, если ξ принимает значения $0, 1, \dots, n$ с вероятностями $\mathsf{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}$. Случайная величина ξ с таким распределением имеет смысл *числа успехов в* n *испытаниях схемы Бернулли с вероятностью успеха* p.

Таблица распределения ξ имеет вид

ξ	0	1	...	k	...	n
P	$(1-p)^n$	$np(1-p)^{n-1}$	...	$C_n^k p^k (1-p)^{n-k}$	...	p^n

Геометрическое распределение.

Говорят, что случайная величина τ имеет геометрическое распределение с параметром p, где $0 \leqslant p \leqslant 1$, и пишут $\tau \Subset \mathbf{G}_p$, если τ принимает значения $1, 2, 3, \ldots$ с вероятностями $\mathsf{P}(\tau = k) = p(1-p)^{k-1}$. Случайная величина τ с таким распределением имеет смысл *номера первого успешного испытания в схеме Бернулли с вероятностью успеха* p.

Таблица распределения τ имеет вид

τ	1	2	...	k	...
P	p	$p(1-p)$	...	$p(1-p)^{k-1}$	...

Распределение Пуассона.

Говорят, что случайная величина ξ имеет распределение Пуассона с параметром λ, где $\lambda > 0$, и пишут $\xi \Subset \Pi_\lambda$, если ξ принимает значения $0, 1, 2, \ldots$ с вероятностями $\mathsf{P}(\xi = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.

Таблица распределения ξ имеет вид

ξ	0	1	...	k	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	...	$\dfrac{\lambda^k}{k!} e^{-\lambda}$	...

Гипергеометрическое распределение.

Говорят, что случайная величина ξ имеет гипергеометрическое распределение с параметрами n, N и K, где $K \leqslant N$, $n \leqslant N$, если ξ принимает целые значения от $\max\{0, N-K-n\}$ до $\min\{n, K\}$ с вероятностями $\mathsf{P}(\xi = k) = \dfrac{C_K^k C_{N-K}^{n-k}}{C_N^n}$. Случайная величина ξ с таким распределением имеет смысл *числа белых шаров среди* n *шаров, выбранных наудачу и без возвращения из урны, содержащей* K *белых шаров и* $N-K$ *не белых.*

Таблицу распределения ξ читатель может нарисовать самостоятельно.

Заметьте, что со всеми этими распределениями мы уже хорошо знакомы.

Но распределения случайных величин далеко не исчерпываются дискретными распределениями. Так, например, если точка бросается наудачу на отрезок [0,1], то можно задать случайную величину, равную координате этой точки. Но число значений этой случайной величины не счетно, так что ее распределение дискретным не является. Да и вероятность этой случайной величине принять каждое из своих возможных значений (попасть в точку) равна нулю. Так что не только таблица распределения не существует, но и соответствие «значение величины $\leftrightarrow$ вероятность его принять» ничего не говорит о распределении случайной величины.

Какими же характеристиками еще можно описать распределение?

Раздел 7. Функция распределения

Заметим, что на том же отрезке [0,1] вероятности попадания в множества положительной меры совсем не нулевые. И термин «наудачу» мы когда-то описывали как раз в терминах вероятностей попадания в множества.

Может быть, разумно описать распределение случайной величины, задав для любого множества вероятность принять значения из этого множества? Это действительно полная характеризация распределения, но уж очень трудно с ней работать — слишком много множеств на прямой.

Нельзя ли обойтись заданием вероятностей попадания в какой-нибудь меньший набор множеств на прямой? Оказывается, что можно ограничиться только вероятностями попадания в интервалы $(-\infty, x)$ для всех $x \in \mathbb{R}$, с помощью которых можно будет определить и вероятность попасть в любое другое множество.

Замечание 11. Можно с таким же успехом ограничиться набором вероятностей попадания в интервалы $(-\infty, x]$, или в (x, ∞), или в $[x, \infty)$, или в (x_1, x_2). Впрочем, последних уже слишком много.

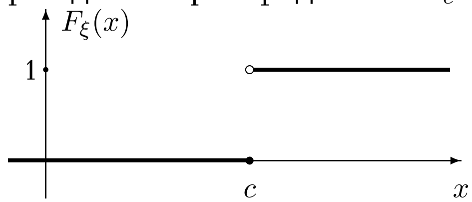
Определение 28.

Функцией распределения случайной величины ξ называется функция $F_\xi(x) : \mathbb{R} \to [0,1]$, при каждом $x \in \mathbb{R}$ равная

$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}\{\omega : \xi(\omega) < x\}.$$

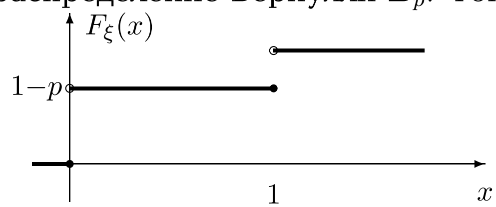
Пример 24. Случайная величина ξ имеет вырожденное распределение $\mathbf{I}_c$. Тогда

$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}(c < x) = \begin{cases} 0, & x \leqslant c; \\ 1, & x > c. \end{cases}$$



Пример 25. Случайная величина ξ имеет распределение Бернулли $\mathbf{B}_p$. Тогда

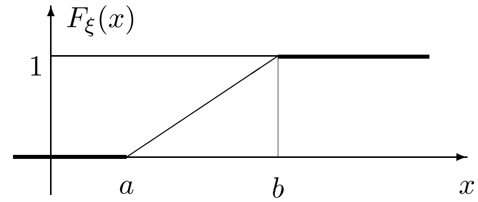
$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & x \leqslant 0; \\ 1-p, & 0 < x \leqslant 1 \\ 1, & x > 1. \end{cases}$$



Пример 26. Будем говорить, что случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$ и писать $\xi \in \mathbf{U}_{a,b}$ ("uniform"), если ξ — координата точки, брошенной наудачу на отрезок $[a, b]$ числовой прямой. Это

распределение можно задать и с помощью функции распределения:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \leqslant x \leqslant b \\ 1, & x > b. \end{cases}$$



Упражнение 8. Построить графики функций распределения для распределения Пуассона, биномиального и геометрического распределения.

7.1 Свойства функции распределения

Теорема 18.

Функция распределения $F_\xi(x)$ обладает следующими свойствами:

F1) *Функция распределения $F_\xi(x)$ не убывает: если $x_1 < x_2$, то $F_\xi(x_1) \leqslant F_\xi(x_2)$;*

F2) *Существуют пределы $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ и $\lim\limits_{x\to\infty} F_\xi(x) = 1$.*

F3) *Функция распределения $F_\xi(x)$ непрерывна слева: $F_\xi(x_0-0) = \lim_{x\to x_0-0} F_\xi(x) = F_\xi(x_0)$.*

Доказательство свойства (F1).
Если $x_1 < x_2$, то $\{\xi < x_1\} \subseteq \{\xi < x_2\}$. Поэтому $F_\xi(x_1) = \mathsf{P}\{\xi < x_1\} \leqslant \mathsf{P}\{\xi < x_2\} = F_\xi(x_2)$. □

Доказательство свойства (F2).

Замечание 12. Если ряд, составленный из неотрицательных слагаемых a_i, сходится, то есть существует $\sum\limits_{i=1}^{\infty} a_i \stackrel{\text{def}}{=} \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} a_i < \infty$, то «хвост» ряда стремится к нулю: $\lim\limits_{n\to\infty} \sum\limits_{i=n}^{\infty} a_i = 0$.

Замечание 13. Существование пределов в свойствах **(F2)**, **(F3)** вытекает из монотонности и ограниченности функции $F_\xi(x)$. Так что остается доказать равенства $\lim\limits_{x\to-\infty} F_\xi(x) = 0$, $\lim\limits_{x\to\infty} F_\xi(x) = 1$ и $\lim\limits_{x\to x_0-0} F_\xi(x) = F_\xi(x_0)$.

Замечание 14. Если существует $\lim\limits_{x\to a} f(x)$, то для произвольной последовательности $\{x_n\}$ такой, что $x_n \to a$ имеет место равенство $\lim\limits_{x\to\infty} f(x) = \lim\limits_{n\to\infty} f(x_n)$.

По замечанию **14**, для доказательства $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ достаточно доказать, что $F_\xi(-n) \to 0$ при $n \to \infty$.

Представим событие $\{\xi < -17\}$ (например) как счетное объединение событий:

$$\{\xi < -17\} = \ldots \cup \{-20 \leqslant \xi < -19\} \cup \{-19 \leqslant \xi < -18\} \cup \{-18 \leqslant \xi < -17\} = \bigcup_{i=17}^{\infty} \{-i-1 \leqslant \xi < -i\}.$$

Используя σ-аддитивность вероятности, и помня, что $\mathsf{P}\{\xi < -17\} \leqslant 1$, получим:

$$\mathsf{P}\{\xi < -17\} = \sum_{i=17}^{\infty} \mathsf{P}\{-i-1 \leqslant \xi < -i\} \leqslant 1, \quad \text{и, по замечанию } \mathbf{12}, \quad \sum_{i=n}^{\infty} \mathsf{P}\{-i-1 \leqslant \xi < -i\} \to 0.$$

Но $\sum\limits_{i=n}^{\infty} \mathsf{P}\{-i-1 \leqslant \xi < -i\} = \mathsf{P}\{\xi < -n\} = F_\xi(-n)$, и сходимость $F_\xi(x)$ к нулю при $x \to -\infty$ доказана.

Итого: *есть ряд, составленный из вероятностей, сумма которого тоже есть вероятность и, следовательно, конечна. А из того, что ряд сходится, по замечанию* **12** *вытекает сходимость «хвоста» ряда к нулю. Осталось посмотреть на этот хвост и убедиться, что он равен как раз той вероятности, сходимость к нулю которой нам нужно доказать. Точно так же докажем и остальные свойства.*

По замечанию **14**, для доказательства $\lim\limits_{x\to\infty} F_\xi(x) = 1$ достаточно доказать, что $F_\xi(n) \to 1$ при $n \to \infty$, или что $1 - F_\xi(n) = \mathsf{P}(\xi \geqslant n) \to 0$.

Представим событие $\{\xi \geqslant 11\}$ (например :-) как счетное объединение событий:

$$\{\xi \geqslant 11\} = \{11 \leqslant \xi < 12\} \cup \{12 \leqslant \xi < 13\} \cup \{13 \leqslant \xi < 14\} \cup \ldots = \bigcup_{i=11}^{\infty} \{i \leqslant \xi < i+1\}.$$

В силу σ-аддитивности вероятности,

$$\mathsf{P}\{\xi \geqslant 11\} = \sum_{i=11}^{\infty} \mathbf{P}\{i \leqslant \xi < i+1\} \leqslant 1, \quad \text{и, по замечанию } \mathbf{12}, \quad \sum_{i=n}^{\infty} \mathsf{P}\{i \leqslant \xi < i+1\} \to 0.$$

Но $\sum\limits_{i=n}^{\infty} \mathsf{P}\{i \leqslant \xi < i+1\} = \mathsf{P}\{\xi \geqslant n\} = 1 - F_\xi(n)$, и сходимость $F_\xi(x)$ к единице при $x \to \infty$ доказана.

$\square$

Доказательство свойства (F3).

Согласно замечанию **14**, достаточно доказать, что $F_\xi\left(x_0 - \dfrac{1}{n}\right) \to F_\xi(x_0)$ при $n \to \infty$. Или, что то же самое, доказать, что

$$F_\xi(x_0) - F_\xi\left(x_0 - \frac{1}{n}\right) = \mathsf{P}(\xi < x_0) - \mathsf{P}\left(\xi < x_0 - \frac{1}{n}\right) = \mathsf{P}\left(x_0 - \frac{1}{n} \leqslant \xi < x_0\right) \to 0. \tag{11}$$

Представим событие $\{\xi < x_0\}$ как счетное объединение событий:

$$\{\xi < x_0\} = \{\xi < x_0 - 1\} \cup \left\{x_0 - 1 \leqslant \xi < x_0 - \frac{1}{2}\right\} \cup \left\{x_0 - \frac{1}{2} \leqslant \xi < x_0 - \frac{1}{3}\right\} \cup \left\{x_0 - \frac{1}{3} \leqslant \xi < x_0 - \frac{1}{4}\right\} \cup \ldots =$$

$$= \{\xi < x_0 - 1\} \cup \bigcup_{i=1}^{\infty} \left\{x_0 - \frac{1}{i} \leqslant \xi < x_0 - \frac{1}{i+1}\right\}.$$

В силу σ-аддитивности вероятности,

$$\mathsf{P}\{\xi < x_0\} = \mathsf{P}\{\xi < x_0 - 1\} + \sum_{i=1}^{\infty} \mathsf{P}\left\{x_0 - \frac{1}{i} \leqslant \xi < x_0 - \frac{1}{i+1}\right\} \leqslant 1, \text{ поэтому снова } \sum_{i=n}^{\infty} \mathsf{P}\left\{x_0 - \frac{1}{i} \leqslant \xi < x_0 - \frac{1}{i+1}\right\} \to 0.$$

Но $\sum\limits_{i=n}^{\infty} \mathsf{P}\left\{x_0 - \dfrac{1}{i} \leqslant \xi < x_0 - \dfrac{1}{i+1}\right\} = \mathsf{P}\left\{x_0 - \dfrac{1}{n} \leqslant \xi < x_0\right\}$, и эта вероятность, как мы только что видели, стремится к нулю с ростом n. Тогда, по (11), $F_\xi(x) \to F_\xi(x_0)$ при $x \to x_0 - 0$ (непрерывность слева). □

Следующая теорема говорит о том, что три доказанных свойства полностью описывают класс функций распределения. То, что любая функция распределения ими обладает, мы с вами доказали, а теорема утверждает, что любая функция с такими свойствами есть функция распределения.

Теорема 19. *Если функция $F : \mathbb{R} \to [0, 1]$ удовлетворяет свойствам* **(F1)–(F3)**, *то F есть функция распределения некоторой случайной величины ξ, то есть найдется вероятностное пространство $\langle \Omega, \mathcal{F}, \mathsf{P}\rangle$ и случайная величина ξ на этом пространстве, что $F(x) \equiv F_\xi(x)$.*

Эту теорему мы доказывать не станем. Хотя ее можно попробовать доказать конструктивно — предъявив то вероятностное пространство (проще всего отрезок $\Omega = [0, 1]$ с σ-алгеброй борелевских множеств и мерой Лебега :-) и ту случайную величину, о существовании которых идет речь. Непременно попробуйте сделать это! Например, можно попробовать, не подойдет ли $\xi(\omega) = \sup\{x : F(x) < \omega\}$.

Прочие полезные свойства функций распределения

F4) В любой точке x_0 разница $F_\xi(x_0{+}0) - F_\xi(x_0)$ равна $\mathsf{P}(\xi = x_0)$:

$$F_\xi(x_0{+}0) - F_\xi(x_0) = \lim_{x\to x_0+0} F_\xi(x) - F_\xi(x_0) = \mathsf{P}(\xi = x_0), \quad \text{или, иначе,} \quad F_\xi(x_0{+}0) = \lim_{x\to x_0+0} F_\xi(x) = F_\xi(x_0) + \mathsf{P}(\xi = x_0) = \mathsf{P}(\xi \leqslant x_0).$$

Упражнение 9. Докажите сами (точно так же, как мы доказывали **(F2)** и **(F3)**).

Заметим, что разница $F_\xi(x_0{+}0) - F_\xi(x_0)$ между пределом при стремлении к x_0 справа и значением в точке x_0 есть величина скачка функции распределения, и равна нулю, если функция распределения непрерывна (справа) в точке x_0. Слева, напомню, функция распределения непрерывна всегда.

Следствие 4. *Если функция распределения $F_\xi(x)$ непрерывна в точке x_0, то $\mathsf{P}(\xi = x_0) = 0$.*

F5) Для любой случайной величины ξ имеет место равенство $\mathsf{P}(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a)$.

Если же функция распределения $F_\xi(x)$ непрерывна (для любого x, или только в точках a и b), то

$$\mathsf{P}(a \leqslant \xi < b) = \mathsf{P}(a < \xi < b) = \mathsf{P}(a \leqslant \xi \leqslant b) = \mathsf{P}(a < \xi \leqslant b) = F_\xi(b) - F_\xi(a).$$

Доказательство. Доказывать нужно только равенство $\mathsf{P}(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a)$, поскольку все остальные равенства следуют из него с учетом следствия **4**. Напомню, что этим равенством мы уже много раз пользовались, доказывая свойства **(F2)**, **(F3)**.

Заметим, что $\{\xi < a\} \cup \{a \leqslant \xi < b\} = \{\xi < b\}$, и первые два события несовместны. Поэтому

$$\mathsf{P}\{\xi < a\} + \mathsf{P}\{a \leqslant \xi < b\} = \mathsf{P}\{\xi < b\},$$

или $F_\xi(a) + \mathsf{P}\{a \leqslant \xi < b\} = F_\xi(b)$, что и требовалось доказать.

$\square$

Функция распределения дискретного распределения

Мы уже видели, как выглядят функции распределения некоторых дискретных распределений.

Из свойств **(F4)**, **(F5)** следует

Свойство 6. *Случайная величина ξ имеет дискретное распределение тогда и только тогда, когда функция распределения F_ξ — ступенчатая функция. При этом возможные значения ξ — точки a_i скачков F_ξ, и $p_i = \mathsf{P}(\xi = a_i) = F_\xi(a_i{+}0) - F_\xi(a_i)$ — величины скачков.*

Упражнение. Доказать, что любая функция распределения имеет не более чем счетное число точек разрыва (или «скачков»). Указание. Сколько скачков величиной более 1/2 может иметь функция распределения? А величиной более 1/3? Более 1/4?

В следующей главе мы рассмотрим случайные величины, функции распределения которых не удовлетворяют свойству **6** хотя бы потому, что они вовсе не имеют разрывов. Более того, мы выделим класс функций распределения, которые «восстанавливаются по своей производной» с помощью интегрирования (так называемые *абсолютно непрерывные функции*).

Раздел 8. Абсолютно непрерывные распределения

Определение 29.

Случайная величина ξ имеет *абсолютно непрерывное* распределение, если существует неотрицательная функция $f_\xi(x)$ такая, что для любого $x \in \mathbb{R}$ функция распределения $F_\xi(x)$ представима в виде

$$F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)\, dt.$$

При этом функция $f_\xi(x)$ называется *плотностью распределения* случайной величины ξ.

Теорема 20.

Плотность распределения обладает свойствами:

(f1) $f_\xi(x) \geqslant 0$ *для любого* x; **(f2)** $\int\limits_{-\infty}^{\infty} f_\xi(t)\, dt = 1$.

Доказательство. (f1) выполнено по определению плотности. Докажем (f2).

$$\int\limits_{-\infty}^{\infty} f_\xi(t)\, dt \stackrel{\text{def}}{=} \lim_{x\to\infty} \int\limits_{-\infty}^{x} f_\xi(t)\, dt = \lim_{x\to\infty} F_\xi(x) = 1$$

по свойству (F2) функций распределения. □

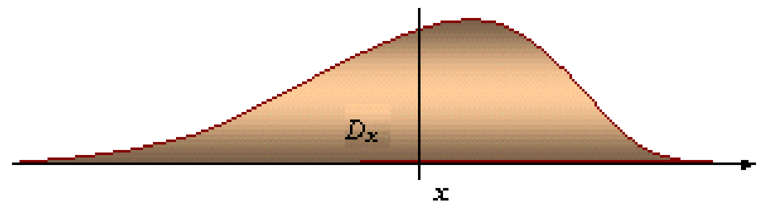
Эти два свойства полностью характеризуют класс плотностей:

Лемма 3. *Если функция f обладает свойствами* (f1) *и* (f2), *то существует вероятностное пространство и случайная величина ξ на нем, для которой f является плотностью распределения.*

Доказательство. Пусть Ω есть область, заключенная между осью абсцисс и графиком функции f («подграфик» функции f). Площадь области Ω равна 1 по свойству (f2). И пусть случайная величина ξ есть абсцисса точки, наудачу брошенной в эту область. Тогда (*вспомнить геометрическую вероятность*) для любого $x \in \mathbb{R}$

$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}(\text{точка попала в область } D_x) =$$

$$= \frac{\text{площадь } D_x}{\text{площадь } \Omega} = \int\limits_{-\infty}^{x} f(t)\ dt,$$



то есть f является плотностью распределения случайной величины ξ. □

Свойства плотностей

(f3) Если случайная величина ξ имеет абсолютно непрерывное распределение, то ее функция распределения всюду непрерывна.

Доказательство. Этот факт следует из представления $F_\xi(x) = \int\limits_{-\infty}^{x} f_\xi(t)\ dt$ и непрерывности интеграла как функции верхнего предела. □

Следствие 5. *Если случайная величина ξ имеет абсолютно непрерывное распределение, то $\mathsf{P}(\xi = x) = 0$ для любого $x \in \mathbb{R}$.*

(f4) Если случайная величина ξ имеет абсолютно непрерывное распределение, то ее функция распределения дифференцируема почти всюду, и $f_\xi(x) = F'_\xi(x) = \dfrac{d}{dx}F_\xi(x)$ для почти всех x.

Замечание 15. Термин для «почти всех» означает «для всех, кроме (возможно) x из некоторого множества нулевой меры (длины)». Заметьте, что стоящую под интегралом функцию можно изменить в одной точке (или на множестве нулевой длины), и интеграл («площадь подграфика») от этого не изменится.

(f5) Если случайная величина ξ имеет абсолютно непрерывное распределение, то

$$\mathsf{P}(a < \xi < b) = \mathsf{P}(a \leqslant \xi < b) = \mathsf{P}(a < \xi \leqslant b) = \mathsf{P}(a \leqslant \xi \leqslant b) = \int\limits_a^b f_\xi(t)\ dt.$$

Доказательство. Действительно, $\mathsf{P}(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a) = \int\limits_{-\infty}^{b} f_\xi(t)\ dt - \int\limits_{-\infty}^{a} f_\xi(t)\ dt$. Остальные равенства вытекают из следствия **5**. $\square$

8.1 Примеры абсолютно непрерывных распределений

Равномерное. Это распределение нам уже знакомо. Говорят, что ξ имеет равномерное распределение на отрезке $[a, b]$, и пишут $\xi \in \mathbf{U}_{a,b}$, если

$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \leqslant x \leqslant b \\ 1, & x > b, \end{cases} \qquad f_\xi(x) = \begin{cases} 0, & x < a; \\ \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x > b. \end{cases}$$

Заметьте, что в точках a и b функция распределения недифференцируема, и плотность можно задать как угодно.

Показательное. Говорят, что ξ имеет показательное распределение с параметром α, $\alpha > 0$ и пишут $\xi \in \mathbf{E}_\alpha$, если

$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\alpha x}, & x \geqslant 0, \end{cases} \qquad f_\xi(x) = \begin{cases} 0, & x < 0; \\ \alpha e^{-\alpha x}, & x \geqslant 0. \end{cases}$$

Показательное распределение является единственным абсолютно непрерывным распределением, для которого выполнено свойство «нестарения» (и в этом смысле оно является непрерывным аналогом дискретного геометрического распределения).

Теорема 21. *Свойство «нестарения». Пусть $\xi \in \mathbf{E}_\alpha$. Тогда для любых $x, y > 0$*

$$\mathsf{P}(\xi > x + y \mid \xi > x) = \mathsf{P}(\xi > y).$$

Упражнение 10. Доказать *«свойство нестарения»*.

Упражнение 11. * Доказать, что если неотрицательная случайная величина ξ имеет абсолютно непрерывное распределение и обладает свойством «нестарения», то есть для любых $x, y > 0$

$$\mathsf{P}(\xi > x + y \mid \xi > x) = \mathsf{P}(\xi > y),$$

то она имеет показательное распределение с некоторым параметром α.

Нормальное. Говорят, что ξ имеет нормальное распределение с параметрами a и σ^2, где $a \in \mathbb{R}$, $\sigma > 0$, и пишут $\xi \Subset \mathbf{N}_{a,\sigma^2}$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}} \quad \text{для любого} \quad x \in \mathbb{R}.$$

Убедимся, что $f_\xi(x)$ действительно является плотностью распределения. Так как $f_\xi(x) > 0$ для всех $x \in \mathbb{R}$, то свойство (f1) выполнено. Проверим выполнение (f2). Используем табличный интеграл (интеграл Пуассона)

$$\int\limits_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

Этот интеграл вычисляется так: $\int\limits_{-\infty}^{\infty} e^{-x^2/2}\, dx \int\limits_{-\infty}^{\infty} e^{-y^2/2}\, dy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy =$ *(цилиндрическая замена переменных* $x = r\cos\phi$, $y = r\sin\phi$, $dx\,dy = r\,dr\,d\phi$*)* $= \int\limits_{0}^{2\pi}\int\limits_{0}^{\infty} re^{-r^2/2}\, dr\, d\phi = \int\limits_{0}^{2\pi}\int\limits_{0}^{\infty} e^{-r^2/2}\, d(r^2/2)\, d\phi = 2\pi$.

$$\int\limits_{-\infty}^{\infty} f_\xi(x)\, dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx = \left[\begin{array}{l}\text{замена переменных}\\ t = \dfrac{x-a}{\sigma},\ dx = \sigma\, dt\end{array}\right] = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-t^2/2}\sigma\, dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-t^2/2}\, dt = 1.$$

Нормальное (иначе называемое гауссовским по имени Карла Гаусса, см. график плотности на купюре 10 DM) распределение играет исключительно важную роль в теории вероятностей, поэтому мы очень подробно изучим все свойства этого распределения.

8.2 Свойства нормального распределения

Нормальное распределение задается, как мы видим, с помощью плотности распределения. Связано это с тем, что нельзя выписать первообразную от функции e^{-x^2} иначе как в виде интеграла, поэтому функцию распределения этого закона можно записать лишь в таком виде:

$$F_\xi(x) = \Phi_{a,\sigma^2}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-a)^2}{2\sigma^2}}\, dt.$$

Мы часто будем использовать обозначение $\Phi_{a,\sigma^2}(x)$ для функции распределения нормального распределения с параметрами a и σ^2.

Исключительно полезно нарисовать график плотности и функции распределения (отметив точки экстремума, перегибов, посчитав значение в точке максимума плотности и расстояние между точками перегибов). График плотности и функции распределения нормального распределения можно также посмотреть здесь: http://www.nsu.ru/mmf/tvims/chernova/PlotDist.html.

Стандартное нормальное распределение

Нормальное распределение $\mathbf{N}_{a,\sigma^2}$ при $a = 0$ и $\sigma^2 = 1$ называется *стандартным нормальным распределением*. Плотность стандартного нормального распределения имеет вид $f_\xi(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ при любом $x \in \mathbb{R}$, а функция распределения $\Phi_{0,1}(x) = \int\limits_{-\infty}^{x} \dfrac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt$ табулирована (то есть ее значения вычислены при многих x) почти во всех математических справочниках. Установим связь между Φ_{a,σ^2} и $\Phi_{0,1}$.

Свойство 7. *Для любого $x \in \mathbb{R}$ справедливо соотношение $\Phi_{a,\sigma^2}(x) = \Phi_{0,1}\left(\dfrac{x-a}{\sigma}\right)$.*

Доказательство.

$$\Phi_{a,\sigma^2}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-a)^2}{2\sigma^2}}\, dt = \left[\begin{array}{l} \textit{замена переменных} \\ y = \dfrac{t-a}{\sigma}, \quad dt = \sigma\, dy \\ t = x \quad \mapsto \quad y = \dfrac{x-a}{\sigma} \end{array}\right] = \int\limits_{-\infty}^{\frac{x-a}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

□

То же самое на языке случайных величин можно сформулировать так:

Следствие 6. *Если $\xi \in \mathbf{N}_{a,\sigma^2}$, то $\eta = \dfrac{\xi - a}{\sigma} \in \mathbf{N}_{0,1}$.*

Доказательство.

$$F_\eta(x) = \mathsf{P}(\eta < x) = \mathsf{P}\left(\frac{\xi-a}{\sigma} < x\right) = \mathsf{P}(\xi < \sigma x + a) = \Phi_{a,\sigma^2}(\sigma x + a) = \Phi_{0,1}\left(\frac{\sigma x + a - a}{\sigma}\right) = \Phi_{0,1}(x).$$

□

Следствие 7. *Если* $\xi \in \mathbf{N}_{a,\sigma^2}$, *то* $\mathsf{P}(x_1 < \xi < x_2) = \Phi_{a,\sigma^2}(x_2) - \Phi_{a,\sigma^2}(x_1) = \Phi_{0,1}\left(\dfrac{x_2 - a}{\sigma}\right) - \Phi_{0,1}\left(\dfrac{x_1 - a}{\sigma}\right)$.

Как мы видим, вычисление любых вероятностей для нормально распределенной случайной величины сводится к вычислению функции распределения $\Phi_{0,1}$. Ее свойства (нарисовать их на графике ***плотности*** стандартного нормального распределения!!):

Свойство 8. $\Phi_{0,1}(0) = 0.5$.

Свойство 9. $\Phi_{0,1}(-x) = 1 - \Phi_{0,1}(x)$.

Свойство 10. *Если* $\xi \in \mathbf{N}_{0,1}$, *то* $\mathsf{P}(|\xi| < x) = 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1$.

Доказательство. $\mathsf{P}(|\xi| < x) = \mathsf{P}(-x < \xi < x) = \Phi_{0,1}(x) - \Phi_{0,1}(-x) =$ (по свойству **9**) $= 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1$. □

Свойство 11 («Правило трех сигм»). *Если* $\xi \in \mathbf{N}_{a,\sigma^2}$, *то* $\mathsf{P}(|\xi - a| > 3\sigma) = 0.0027$ (*мало, в общем* :).

Доказательство.

$$\mathsf{P}(|\xi - a| > 3\sigma) = 1 - \mathsf{P}(|\xi - a| < 3\sigma) = 1 - \mathsf{P}\left(\left|\frac{\xi - a}{\sigma}\right| < 3\right).$$

Но величина $\eta = \dfrac{\xi - a}{\sigma} \in \mathbf{N}_{0,1}$, и можно использовать свойство **10**: $1 - \mathsf{P}(|\eta| < 3) = 1 - (1 - 2\Phi_{0,1}(-3)) = 2\Phi_{0,1}(-3) = 2 \cdot 0.00135 = 0.0027$ (найти в таблице!). □

Смысла в запоминании числа 0.0027 нет никакого, а вот помнить, что почти вся масса нормального распределения сосредоточена в границах $[a - 3\sigma, a + 3\sigma]$, всегда полезно.

Раздел 9. Случайные вектора и их распределения

Определение 30. Если случайные величины $\xi_1, \ldots, \xi_n$ заданы на одном вероятностном пространстве, то вектор $(\xi_1, \ldots, \xi_n)$ мы будем называть *случайным вектором*.

Определение 31. Функция $F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \mathsf{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$ называется *функцией распределения* случайного вектора $(\xi_1, \ldots, \xi_n)$ или *функцией совместного распределения* случайных величин $\xi_1, \ldots, \xi_n$.

9.1 Свойства функции совместного распределения

Для простоты обозначений все дальнейшие рассуждения и формулировки приводятся в случае $n = 2$ для случайного вектора (ξ_1, ξ_2).

F0) $0 \leqslant F_{\xi_1,\xi_2}(x_1, x_2) \leqslant 1$.

F1) $F_{\xi_1,\xi_2}(x_1, x_2)$ не убывает по каждой координате вектора (x_1, x_2).

F2) Для любого $i = 1, 2$ существует $\lim\limits_{x_i \to -\infty} F_{\xi_1,\xi_2}(x_1, x_2) = 0$;

Для любого $i = 1, 2$ существует $\lim\limits_{x_i \to \infty} F_{\xi_1,\xi_2}(x_1, x_2)$. При этом

$$F_{\xi_1,\xi_2}(\infty, x_2) \stackrel{\text{def}}{=} \lim_{x_1 \to \infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_2}(x_2), \qquad F_{\xi_1,\xi_2}(x_1, \infty) \stackrel{\text{def}}{=} \lim_{x_2 \to \infty} F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_1}(x_1).$$

F3) Функция $F_{\xi_1,\xi_2}(x_1, x_2)$ по каждой координате вектора (x_1, x_2) непрерывна слева.

Доказательство этих свойств совершенно аналогично одномерному случаю.

Только теперь этих свойств оказывается недостаточно для описания класса функций совместного распределения. Иначе говоря, выполнение этих свойств для некоторой функции $F : \mathbb{R}^2 \to \mathbb{R}$ вовсе не гарантирует, что эта функция является функцией распределения некоторого случайного вектора.

Упражнение 12. Доказать, что функция

$$F(x_1, x_2) = \begin{cases} 0, & x_1 \leqslant 0 \text{ или } x_2 \leqslant 0 \text{ или } x_1 + x_2 \leqslant 1; \\ 1, & \text{иначе, то есть когда одновременно } x_1 > 0, x_2 > 0, x_1 + x_2 > 1. \end{cases}$$

а) удовлетворяет всем свойствам **(F0)-(F3)**;
б) не является функцией распределения никакого вектора (ξ_1, ξ_2) хотя бы потому, что, найдись такой вектор, найдется и прямоугольник $[a_1, b_1] \times [a_2, b_2]$, вероятность попасть в который (вычисленная с помощью этой «функции распределения») отрицательна: $\mathsf{P}(a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2) < 0$!

Как же связана вероятность вектору попасть в прямоугольник с функцией распределения этого вектора?

Упражнение 13. Доказать, что

$$\mathsf{P}(a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2) = F_{\xi_1,\xi_2}(b_1, b_2) - F_{\xi_1,\xi_2}(a_1, b_2) - F_{\xi_1,\xi_2}(b_1, a_2) + F_{\xi_1,\xi_2}(a_1, a_2). \tag{12}$$

Оказывается, если потребовать дополнительно от функции F, чтобы для всякого прямоугольника $[a_1, b_1] \times [a_2, b_2]$ вероятность $P(a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2)$, связанная с функцией F равенством (12), была неотрицательна, то любая функция, обладающая этим свойством и свойствами **(F0)-(F3)**, уже будет функцией распределения некоторого случайного вектора.

На самом деле существо свойства **(F2)** *в той его части, что касается предела на бесконечности, весьма туманно. Утверждает это свойство гораздо больше, чем просто «предел функции совместного распределения при стремлении одной координаты к бесконечности есть тоже функция распределения». Но как в общем случае проверить, что это не просто «некая функция распределения», но функция распределения оставшейся координаты вектора* (ξ_1, ξ_2)*? Если, не лукавя, рассмотреть в упражнении* **12** $F_1(x_1) = \lim_{x_2\to\infty} F(x_1, x_2)$ *и* $F_2(x_2) = \lim_{x_1\to\infty} F(x_1, x_2)$*, то обе эти функции являются функциями распределения (вырожденного закона, т.е. случайных величин* $\xi_1 = 0$ *и* $\xi_2 = 0$ *п.н.). Но две вырожденные случайные величины независимы, и их функция совместного распределения равна* 1 *в первом квадранте (включая его границу) и нулю в остальных квадрантах, но никак не равна* F*. Оставляю на суд читателя вопрос о том, выполнено ли все-таки условие* **(F2)** *для* F *из упражнения* **12**.

9.2 Типы многомерных распределений

Ограничимся рассмотрением только двух случаев, когда совместное распределение координат случайного вектора (ξ_1, ξ_2) либо *дискретно*, либо *абсолютно непрерывно*.

Дискретное совместное распределение

Определение 32. Говорят, что случайные величины ξ_1, ξ_2 имеют *дискретное* совместное распределение, если существует конечный или счетный набор $\{a_i, b_j\}$ такой, что $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = 1$.

Таблицу, на пересечении i-й строки и j-го столбца которой (или наоборот) стоит число $\mathsf{P}(\xi_1 = a_i, \xi_2 = b_j)$, называют *таблицей совместного распределения* случайных величин ξ_1 и ξ_2.

Замечание 16. Напомню, что таблицы распределения каждой из случайных величин ξ_1, ξ_2 в отдельности (таблицы частных, или *маргинальных* распределений) восстанавливаются по таблице *совместного* распределения с помощью очевидных формул:

$$\sum_{j=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = \mathsf{P}(\xi_1 = a_i), \quad \sum_{i=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = \mathsf{P}(\xi_2 = b_j).$$

Если эти формулы вам не представляются очевидными, необходимо вернуться к разделу **4** и перечитать определение **18** полной группы событий, обратив также внимание на доказательство теоремы **8** (формулы полной вероятности).

Абсолютно непрерывное совместное распределение

Определение 33. Говорят, что с.в. ξ_1, ξ_2 (заданные на одном вероятностном пространстве) имеют *абсолютно непрерывное совместное распределение*, если существует функция $f_{\xi_1,\xi_2}(x_1, x_2) \geqslant 0$ такая, что для любой точки $(x_1, x_2) \in \mathbb{R}^2$

$$F_{\xi_1,\xi_2}(x_1, x_2) = \mathsf{P}(\xi_1 < x_1, \xi_2 < x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f_{\xi_1,\xi_2}(s_1, s_2)\ ds_2 \right)\ ds_1.$$

Если такая функция $f_{\xi_1,\xi_2}(x_1, x_2)$ существует, она называется *плотностью совместного распределения* случайных величин ξ_1, ξ_2.

Замечание 17. Для всего дальнейшего более чем достаточно считать, что $\int\limits_{a_1}^{b_1} \left(\int\limits_{a_2}^{b_2} f(s_1, s_2)\ ds_2 \right)\ ds_1$ равняется объему под графиком функции f над областью интегрирования — прямоугольником $[a_1, b_1] \times [a_2, b_2]$.

Плотность совместного распределения обладает свойствами, аналогичными свойствам плотности распределения одной случайной величины:

(f1) $f_{\xi_1,\xi_2}(x_1, x_2) \geqslant 0$ для любых $x_1, x_2 \in \mathbb{R}$; **(f2)** $\int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_1, x_2)\ dx_2 \right)\ dx_1 = 1.$

Более того, любая функция, обладающая этими свойствами, является плотностью некоторого совместного распределения.

Если совместное распределение абсолютно непрерывно, то по функции совместного распределения его плотность находится как смешанная частная производная:

(f3) $f_{\xi_1,\xi_2}(x_1, x_2) = \dfrac{\partial^2}{\partial x_1 \partial x_2} F_{\xi_1,\xi_2}(x_1, x_2).$

Из свойства (F2) функции совместного распределения вытекает следующее утверждение. Для $n > 2$ это утверждение, как и свойство (F2), выглядит существенно иначе!

Теорема 22. *Если случайные величины ξ_1, ξ_2 имеют абсолютно непрерывное совместное распределение с плотностью $f(x_1, x_2)$, то ξ_1 и ξ_2 в отдельности также имеют абсолютно непрерывное распределение с плотностями:*

$$f_{\xi_1}(s_1) = \int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_2; \qquad f_{\xi_2}(s_2) = \int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_1.$$

Доказательство. По (F2),

$$F_{\xi_1}(x_1) \;=\; \lim_{x_2\to\infty} F_{\xi_1,\xi_2}(x_1, x_2) = \lim_{x_2\to\infty} \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f(s_1, s_2)\, ds_2 \right) ds_1 = \int\limits_{-\infty}^{x_1} \underbrace{\left(\int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_2 \right)}_{f_{\xi_1}(s_1)} ds_1; \quad \text{аналогично}$$

$$F_{\xi_2}(x_2) \;=\; \lim_{x_1\to\infty} F_{\xi_1,\xi_2}(x_1, x_2) = \lim_{x_1\to\infty} \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f(s_1, s_2)\, ds_2 \right) ds_1 = \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{x_2} f(s_1, s_2)\, ds_2 \right) ds_1 = \int\limits_{-\infty}^{x_2} \underbrace{\left(\int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_1 \right)}_{f_{\xi_2}(s_2)} ds_2.$$

$\square$

9.3 Независимость случайных величин

Определение 34. Случайные величины $\xi_1, \ldots, \xi_n$ ***независимы***, если для любого набора множеств $B_1 \subseteq \mathbb{R}, \ldots,$ $B_n \subseteq \mathbb{R}$ (из борелевской σ-алгебры — для тех, кто прочитал, что это такое, или произвольных — для тех, кто не прочитал) имеет место равенство:

$$\mathsf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathsf{P}(\xi_1 \in B_1) \cdot \ldots \cdot \mathsf{P}(\xi_n \in B_n).$$

Это определение можно сформулировать в терминах функций распределения:

Определение 35. Случайные величины $\xi_1, \ldots, \xi_n$ ***независимы***, если для любых $x_1, \ldots, x_n$ имеет место равенство:

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n).$$

Упражнение 14. Доказать, что из независимости в смысле определения **34** следует независимость в смысле определения **35** (доказательство в обратную сторону см. (только см.!) в §4 гл.3 учебника А.А.Боровкова «Теория вероятностей»).

Для случайных величин с дискретным распределением эквивалентное определение независимости выглядит так:

Определение 36. Случайные величины $\xi_1, \ldots, \xi_n$ с дискретным распределением ***независимы***, если для любых $a_1, \ldots, a_n$ имеет место равенство: $\mathsf{P}(\xi_1 = a_1, \ldots, \xi_n = a_n) = \mathsf{P}(\xi_1 = a_1) \cdot \ldots \cdot \mathsf{P}(\xi_n = a_n)$.

Упражнение 15. Доказать, что для случайных величин с дискретным распределением определения **34** и **36** эквивалентны.

Для случайных величин с абсолютно непрерывным совместным распределением определение независимости можно сформулировать так:

Определение 37. Случайные величины $\xi_1, \ldots, \xi_n$ с абсолютно непрерывным совместным распределением ***независимы***, если плотность совместного распределения равна произведению плотностей случайных величин $\xi_1, \ldots, \xi_n$, то есть для любых $x_1, \ldots, x_n$ имеет место равенство: $f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = f_{\xi_1}(x_1) \cdot \ldots \cdot f_{\xi_n}(x_n)$.

Доказательство. *Докажем эквивалентность* определений **35** и **37**. По теореме **22**, если совместное распределение $\xi_1, \ldots, \xi_n$ абсолютно непрерывно, то и в отдельности $\xi_1, \ldots, \xi_n$ также имеют абсолютно непрерывное распределение. Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы в смысле определения **35**, то есть для любых $x_1, \ldots, x_n$

$$\begin{aligned}\int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} f_{\xi_1,\ldots,\xi_n}(s_1, \ldots, s_n)\, ds_n \ldots\, ds_1 \;&=\; F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \left\{ \text{по опр. } \mathbf{35} \right\} = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n) = \\ &=\; \int\limits_{-\infty}^{x_1} f_{\xi_1}(s_1)\, ds_1 \cdot \ldots \cdot \int\limits_{-\infty}^{x_n} f_{\xi_n}(s_n)\, ds_n = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} f_{\xi_1}(s_1) \cdot \ldots \cdot f_{\xi_n}(s_n)\, ds_n \ldots\, ds_1.\end{aligned}$$

Равенство двух синих интегралов при всех значениях $x_1, \ldots, x_n$ влечет равенство подынтегральных выражений, то есть независимость в смысле определения **37**. Для доказательства в обратную сторону можно использовать те же равенства, но в другом порядке. $\square$

Раздел 10. Преобразования случайных величин

10.1 Преобразование одной случайной величины

Мы будем рассматривать только преобразования случайных величин с абсолютно непрерывными распределениями. Пусть с. в. ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$. Построим с помощью функции $g : \mathbb{R} \to \mathbb{R}$ случайную величину $\eta = g(\xi)$. Требуется найти функцию распределения и, если существует, плотность распределения η.

Замечание 18. Плотность распределения случайной величины $\eta = g(\xi)$ существует далеко не при любых функциях g. Так, если функция g кусочно-постоянна, то с. в. η имеет дискретное распределение, и плотность ее распределения не существует.

Упражнение 16. Привести пример случайной величины ξ с абсолютно непрерывным распределением и *непрерывной* функции g такой, что $\eta = g(\xi)$ имеет
а) дискретное распределение; б)* *невырожденное* дискретное распределение.

Плотность распределения $g(\xi)$ заведомо существует, если, например, функция g монотонна («строго монотонна»). Вспомним, что означает «найти плотность распределения η, если она существует».

По определению, если мы представим (для любого x) функцию распределения η в виде $F_\eta(x) = \int\limits_{-\infty}^{x} h(y)\,dy$, где подынтегральная функция $h(y)$ неотрицательна, то плотность распределения с. в. η существует и в точности равна подынтегральной функции: $f_\eta(x) = h(x)$.

Так что доказывать существование плотности распределения и находить ее мы будем одновременно, находя нужное интегральное представление для функции распределения.

Теорема 23. *Пусть ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$, и постоянная $a \neq 0$. Тогда случайная величина $\eta = a\xi + b$ имеет плотность распределения $f_\eta(x) = \dfrac{1}{|a|} \cdot f_\xi\left(\dfrac{x-b}{a}\right)$.*

Доказательство. Пусть сначала $a > 0$.

$$F_\eta(x) = F_{a\xi+b}(x) = \mathsf{P}(a\xi + b < x) = \mathsf{P}\left(\xi < \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right) = \int\limits_{-\infty}^{(x-b)/a} f_\xi(t)\,dt =$$

$$= \left[\begin{array}{l} \textit{Замена: } t = \dfrac{y-b}{a},\ dt = \dfrac{dy}{a} \\ t = \dfrac{x-b}{a} \mapsto y = x,\ t = -\infty \mapsto y = -\infty \end{array} \right] = \int\limits_{-\infty}^{x} \frac{1}{a} \cdot f_\xi\left(\frac{y-b}{a}\right)\, dy,$$

то есть при $a > 0$ случайная величина $\eta = a\xi + b$ имеет абсолютно непрерывное распределение с плотностью $f_\eta(x) = \dfrac{1}{a} \cdot f_\xi\left(\dfrac{x-b}{a}\right) = \dfrac{1}{|a|} \cdot f_\xi\left(\dfrac{x-b}{a}\right)$.

Пусть теперь $a < 0$.

$$F_\eta(x) = F_{a\xi+b}(x) = \mathsf{P}(a\xi + b < x) = \mathsf{P}\left(\xi > \frac{x-b}{a}\right) = \int\limits_{(x-b)/a}^{\infty} f_\xi(t)\, dt = \left[\begin{array}{l} \textit{Замена: } t = \dfrac{y-b}{a},\ dt = \dfrac{dy}{a} \\ t = \dfrac{x-b}{a} \mapsto y = x,\ t = \infty \mapsto y = -\infty \end{array} \right] =$$

$$= \int\limits_{x}^{-\infty} \frac{1}{a} \cdot f_\xi\left(\frac{y-b}{a}\right)\, dy = \int\limits_{-\infty}^{x} \frac{1}{|a|} \cdot f_\xi\left(\frac{y-b}{a}\right)\, dy,$$

то есть при $a < 0$ случайная величина $\eta = a\xi + b$ имеет абсолютно непрерывное распределение с плотностью $f_\eta(x) = \dfrac{1}{|a|} \cdot f_\xi\left(\dfrac{x-b}{a}\right)$. □

Для произвольной монотонной функции g (то есть такой, что для любых $x_1 < x_2$ либо $g(x_1) < g(x_2)$ (монотонно возрастающая функция), либо $g(x_1) > g(x_2)$ (монотонно убывающая функция)) справедливо аналогичное теореме **23** утверждение.

Теорема 24. *Пусть ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$, и функция $g : \mathbb{R} \to \mathbb{R}$ монотонна. Тогда случайная величина $\eta = g(\xi)$ имеет плотность распределения*

$$f_\eta(x) = \frac{1}{\left|g'\big(g^{-1}(x)\big)\right|} \cdot f_\xi\left(g^{-1}(x)\right).$$

Здесь g^{-1} — функция, обратная к g, а $g'\big(g^{-1}(x)\big) = \left(g^{-1}(x)\right)'$ — производная функции g в точке $g^{-1}(x)$, она же производная функции $g^{-1}(x)$.

Упражнение 17. Доказать теорему **24**.

Следующие утверждения сразу следуют из теоремы **23**. Первое из них мы уже доказывали непосредственно. *Доказать остальные.*

Следствие 8. *Если $\xi \in \mathbf{N}_{0,1}$, то $\eta = \sigma\xi + a \in \mathbf{N}_{a,\sigma^2}$.*

Доказательство. Действительно, $f_\eta(x) = \dfrac{1}{\sigma} f_\xi\left(\dfrac{x-a}{\sigma}\right) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$. □

Следствие 9. *Если $\eta \in \mathbf{N}_{a,\sigma^2}$, то $\xi = \dfrac{\eta - a}{\sigma} \in \mathbf{N}_{0,1}$.*

Следствие 10. *Если $\xi \in \mathbf{E}_\alpha$, то $\eta = \alpha\xi \in \mathbf{E}_1$.*

10.2 Функции от двух случайных величин

Пусть ξ_1, ξ_2 — случайные величины с плотностью совместного распределения $f_{\xi_1,\xi_2}(x_1, x_2)$, и функция $g : \mathbb{R}^2 \to \mathbb{R}$. Требуется найти функцию (а если существует, то и плотность) распределения случайной величины $\eta = g(\xi_1, \xi_2)$.

Пользуясь тем, что вероятность случайному вектору попасть в область можно вычислить как объем под графиком плотности распределения вектора над этой областью, сформулируем утверждение.

Теорема 25. *Пусть $x \in \mathbb{R}$, и область $D_x \subseteq \mathbb{R}^2$ состоит из точек (x_1, x_2) таких, что $g(x_1, x_2) < x$. Тогда случайная величина $\eta = g(\xi_1, \xi_2)$ имеет функцию распределения*

$$F_\eta(x) = \mathsf{P}\big(g(\xi_1, \xi_2) < x\big) = \mathsf{P}\big((\xi_1, \xi_2) \in D_x\big) = \iint\limits_{D_x} f_{\xi_1,\xi_2}(x_1, x_2)\ dx_2\ dx_1.$$

Всюду далее в этой главе предполагается, что случайные величины ξ_1 и ξ_2 независимы, то есть

$$f_{\xi_1,\xi_2}(x_1, x_2) = f_{\xi_1}(x_1) f_{\xi_2}(x_2).$$

Следствие 11 (Формула свертки).
Если с. в. ξ_1 и ξ_2 независимы и имеют абсолютно непрерывное распределение с плотностями $f_{\xi_1}(x_1)$ и $f_{\xi_2}(x_2)$, то плотность распределения суммы $\xi_1 + \xi_2$ равна «свертке» плотностей f_{ξ_1} и f_{ξ_2}:

$$f_{\xi_1+\xi_2}(t) = \int\limits_{-\infty}^{\infty} f_{\xi_1}(u)\ f_{\xi_2}(t-u)\ du = \int\limits_{-\infty}^{\infty} f_{\xi_2}(u)\ f_{\xi_1}(t-u)\ du \tag{13}$$

Доказательство. Воспользуемся утверждением теоремы **25** для функции $g(x_1,x_2)=x_1+x_2$. Итегрирование по области $D_x=\{(x_1,x_2)\ :\ x_1+x_2<x\}$ можно заменить последовательным вычислением двух интегралов: наружного — по переменной x_1, меняющейся от $-\infty$ до ∞, и внутреннего — по переменной x_2, которая при каждом x_1 должна быть меньше, чем $x-x_1$. То есть $D_x=\{(x_1,x_2)\ :\ x_1\in(-\infty,\infty), x_2\in(-\infty,x-x_1)\}$. Поэтому

$$F_{\xi_1+\xi_2}(x)=\iint\limits_{D_x} f_{\xi_1,\xi_2}(x_1,x_2)\ dx_2\ dx_1 \overset{\text{независ-ть}}{=} \iint\limits_{D_x} f_{\xi_1}(x_1)\ f_{\xi_2}(x_2)\ dx_2\ dx_1 = \int\limits_{-\infty}^{\infty}\left(\ \int\limits_{-\infty}^{x-x_1} f_{\xi_1}(x_1)\ f_{\xi_2}(x_2)\ dx_2\right)\ dx_1$$

Сделаем в последнем интеграле замену переменной x_2 на t так: $x_2=t-x_1$. При этом $x_2\in(-\infty,x-x_1)$ перейдет в $t\in(-\infty,x)$, $dx_2=dt$. В полученном интеграле меняем, наконец, порядок интегрирования:

$$F_{\xi_1+\xi_2}(x)=\int\limits_{-\infty}^{\infty}\left(\ \int\limits_{-\infty}^{x} f_{\xi_1}(x_1)\ f_{\xi_2}(t-x_1)\ dt\right)\ dx_1 = \int\limits_{-\infty}^{x}\underbrace{\left(\ \int\limits_{-\infty}^{\infty} f_{\xi_1}(x_1)\ f_{\xi_2}(t-x_1)\ dx_1\right)}_{f_{\xi_1+\xi_2}(t)} dt.$$

Итак, мы представили функцию распределения $F_{\xi_1+\xi_2}(x)$ в виде $\int\limits_{-\infty}^{x} f_{\xi_1+\xi_2}(t)\ dt$, где

$$f_{\xi_1+\xi_2}(t)=\int\limits_{-\infty}^{\infty} f_{\xi_1}(x_1)\ f_{\xi_2}(t-x_1)\ dx_1 = \int\limits_{-\infty}^{\infty} f_{\xi_1}(u)\ f_{\xi_2}(t-u)\ du.$$

Второе равенство получается либо из первого заменой переменных, либо если всюду в доказательстве поменять местами индексы $_1$ и $_2$. $\square$

Следствие **11** не только предлагает формулу для вычисления плотности распределения суммы, но и утверждает (заметьте!), что сумма двух независимых случайных величин с абсолютно непрерывными распределениями также имеет абсолютно непрерывное распределение. Для тех, кто уже ничему не удивляется, упражнение: *привести пример двух случайных величин с абсолютно непрерывными распределениями, таких что их сумма имеет вырожденное распределение*

Если даже одна из двух независимых случайных величин имеет дискретное, а вторая — абсолютно непрерывное распределение, то их сумма тоже имеет абсолютно непрерывное распределение, как показывает следующее упражнение.

Упражнение 18. Пусть с. в. ξ имеет таблицу распределения $\mathsf{P}(\xi=a_i)=p_i$, с. в. η имеет плотность распределения $f_\eta(x)$, и эти величины независимы. Доказать, что $\xi+\eta$ имеет плотность распределения $f_{\xi+\eta}(x)=\sum\limits_i p_i f_\eta(x-a_i)$.

10.3 Примеры использования формулы свертки

Пример 27. Пусть независимые случайные величины ξ и η имеют стандартное нормальное распределение. Докажем, что их сумма имеет нормальное распределение с параметрами 0 и 2.

Доказательство. По формуле свертки, плотность суммы равна

$$f_{\xi+\eta}(x)=\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\frac{1}{2}u^2}\,e^{-\frac{1}{2}(x-u)^2}du=\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\frac{1}{2}(2u^2+x^2-2xu)}du=\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\left(u^2+\frac{x^2}{2}-xu\right)}du.$$

Выделим полный квадрат по u в показателе экспоненты: $u^2+\frac{x^2}{2}-xu=\left(u-\frac{x}{2}\right)^2+\frac{x^2}{4}$. Тогда

$$f_{\xi+\eta}(x)=e^{\frac{x^2}{4}}\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\left(u-\frac{x}{2}\right)^2}du=e^{\frac{x^2}{4}}\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-v^2}dv=\frac{1}{2\sqrt{\pi}}\,e^{\frac{x^2}{4}}\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{\pi}}\,e^{-v^2}dv=\frac{1}{2\sqrt{\pi}}\,e^{\frac{x^2}{4}}.$$

Последнее равенство верно поскольку под интегралом стоит плотность нормального распределения с параметрами 0 и $\frac{1}{\sqrt{2}}$, так что интеграл по всей прямой равен 1. Итак, мы получили, что плотность суммы есть плотность нормального распределения с параметрами 0 и 2. □

Если сумма двух независимых случайных величин из одного и того же распределения (возможно, с разными параметрами) имеет такое же распределение, говорят, что это распределение ***устойчиво*** относительно суммирования.

В следующих утверждениях, доказать которые предлагается читателю, перечислены практически все устойчивые распределения. Еще с одним из них (распределением χ^2) читатель познакомится в курсе математической статистики.

Лемма 4. *Пусть случайные величины $\xi\in\Pi_\lambda$ и $\eta\in\Pi_\mu$ независимы. Тогда $\xi+\eta\in\Pi_{\lambda+\mu}$.*

Лемма 5. *Пусть случайные величины $\xi\in\mathbf{B}_{n,p}$ и $\eta\in\mathbf{B}_{m,p}$ независимы. Тогда $\xi+\eta\in\mathbf{B}_{n+m,p}$.*

Лемма 6. *Пусть случайные величины $\xi\in\mathbf{N}_{a_1,\sigma_1^2}$ и $\eta\in\mathbf{N}_{a_2,\sigma_2^2}$ независимы. Тогда $\xi+\eta\in\mathbf{N}_{a_1+a_2,\sigma_1^2+\sigma_2^2}$.*

Лемму **6** мы докажем позднее, используя аппарат характеристических функций, хотя при некотором терпении можно попробовать доказать ее с помощью формулы свертки, как в примере **27**.

Показательное распределение не устойчиво по суммированию, однако его можно считать частным случаем гамма-распределения, которое уже в некотором смысле устойчиво относительно суммирования.

Определение 38. Случайная величина ξ имеет ***гамма-распределение*** $\mathbf{\Gamma}_{\alpha,\lambda}$ с параметрами $\alpha > 0, \lambda > 0$, если она имеет плотность распределения

$$f_\xi(x) = \begin{cases} 0, & \text{при } x \leqslant 0, \\ c \cdot x^{\lambda-1} e^{-\alpha x}, & \text{при } x > 0, \end{cases}$$

где постоянная c вычисляется из условия $\int\limits_{-\infty}^{\infty} f_\xi(x)\ dx = c\int\limits_0^\infty x^{\lambda-1}e^{-\alpha x}\,dx = 1$, то есть $c = \dfrac{\alpha^\lambda}{\Gamma(\lambda)}$.

Здесь $\Gamma(\lambda) = \int\limits_0^\infty x^{\lambda-1}e^{-x}\,dx = (\lambda-1)\Gamma(\lambda-1)$ *— гамма-функция Эйлера, при* k *целых* $\Gamma(k) = (k-1)!$ *и* $\Gamma(1) = 1$*.*

Заметим, что показательное распределение $\mathbf{E}_\alpha$ есть гамма-распределение $\mathbf{\Gamma}_{\alpha,1}$.

Лемма 7. *Пусть независимые случайные величины* $\xi_1, \ldots, \xi_n$ *имеют показательное распределение* $\mathbf{E}_\alpha = \mathbf{\Gamma}_{\alpha,1}$. *Тогда* $\xi_1 + \cdots + \xi_n \in \mathbf{\Gamma}_{\alpha,n}$.

Доказательство. Докажем утверждение по индукции. При $n = 1$ оно верно в силу равенства $\mathbf{E}_\alpha = \mathbf{\Gamma}_{\alpha,1}$. Пусть утверждение леммы справедливо для $n = k-1$. Докажем, что оно верно и для $n = k$. По предположению индукции $S_{k-1} = \xi_1 + \cdots + \xi_{k-1} \in \mathbf{\Gamma}_{\alpha,k-1}$, то есть имеет плотность распределения

$$f_{S_{k-1}}(x) = \begin{cases} 0, & \text{при } x \leqslant 0, \\ \dfrac{\alpha^{k-1}}{(k-2)!} \cdot x^{k-2}e^{-\alpha x}, & \text{при } x > 0. \end{cases}$$

Тогда по формуле свертки плотность суммы $S_k = \xi_1 + \cdots + \xi_k$ равна

$$f_{S_k}(x) = \int\limits_{-\infty}^{\infty} f_{S_{k-1}}(u) f_{\xi_k}(x-u)\,du = \int\limits_0^\infty \frac{\alpha^{k-1}}{(k-2)!} \cdot u^{k-2}e^{-\alpha u} f_{\xi_k}(x-u)\,du.$$

Так как $f_{\xi_k}(x-u) = 0$ при $x-u < 0$, то есть при $u > x$, то плотность под интегралом отлична от нуля, если переменная интегрирования изменяется в пределах $0 \leqslant u \leqslant x$. Поэтому

$$f_{S_k}(x) = \int\limits_0^x \frac{\alpha^{k-1}}{(k-2)!} \cdot u^{k-2}e^{-\alpha u} \cdot \alpha e^{-\alpha(x-u)}\,du = e^{-\alpha x}\int\limits_0^x \frac{\alpha^k}{(k-2)!} \cdot u^{k-2}\,du = \frac{\alpha^k}{(k-1)!} \cdot x^{k-2}e^{-\alpha x}.$$

То есть $S_k \in \mathbf{\Gamma}_{\alpha,k}$, что и требовалось доказать. $\square$

«Если я имею одинаковые шансы на получение a или b, то цена моему ожиданию равна $(a+b)/2$».

Христиан Гюйгенс, *О расчетах в азартной игре* (1657)

Раздел 11. Числовые характеристики случайных величин

11.1 Математическое ожидание случайной величины

Определение 39. *Математическим ожиданием* $\mathsf{E}\,\xi$ (средним значением, первым моментом) случайной величины ξ с дискретным распределением, задаваемым таблицей $\mathsf{P}(\xi = a_i) = p_i$, $i \in \mathbb{Z}$, называется число

$$\mathsf{E}\,\xi = \sum_i a_i p_i = \sum_i a_i \mathsf{P}(\xi = a_i), \quad \text{если указанный ряд абсолютно сходится.}$$

Если же $\sum\limits_i |a_i| p_i = \infty$, то говорят, что математическое ожидание *не существует*.

Определение 40. *Математическим ожиданием* $\mathsf{E}\,\xi$ (средним значением, первым моментом) случайной величины ξ с абсолютно непрерывным распределением с плотностью распределения $f_\xi(x)$, называется число

$$\mathsf{E}\,\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x)\,dx, \quad \text{если указанный интеграл абсолютно сходится.}$$

Если же $\int\limits_{-\infty}^{\infty} |x| f_\xi(x)\,dx = \infty$, то говорят, что математическое ожидание *не существует*.

Математическое ожидание имеет простой физический смысл: если на прямой разместить единичную массу, поместив в точки a_i массу p_i (для дискретного распределения), или «размазав» ее с плотностью $f_\xi(x)$ (для абсолютно непрерывного распределения), то точка $\mathsf{E}\,\xi$ есть координата «центра тяжести» прямой.

Пример 28. Пусть случайная величина ξ равна числу очков, выпадающих при одном подбрасывании кубика. Тогда $\mathsf{E}\,\xi = \sum\limits_{k=1}^{6} k \cdot \frac{1}{6} = 3.5$: *в среднем при одном подбрасывании кубика выпадает* 3.5 *очка!*

Пример 29. Пусть случайная величина ξ — координата точки, брошенной наудачу на отрезок $[a, b]$. Тогда $\mathsf{E}\,\xi = \int\limits_a^b x \cdot \frac{1}{b-a}\,dx = \frac{a+b}{2}$: *центр тяжести равномерного распределения на отрезке есть середина отрезка!*

11.2 Свойства математического ожидания

Во всех свойствах предполагается, что рассматриваемые математические ожидания существуют.

E0. Математическое ожидание случайной величины есть ЧИСЛО!

E1. Для произвольной функции $g:\mathbb{R}\to\mathbb{R}$

$$\mathsf{E}\,g(\xi)=\begin{cases}\sum\limits_k g(a_k)\mathsf{P}(\xi=a_k), & \text{если распределение } \xi \text{ дискретно;}\\ \int\limits_{-\infty}^{\infty} g(x)f_\xi(x)\,dx, & \text{если распределение } \xi \text{ абсолютно непрерывно.}\end{cases}$$

Доказательство. Мы докажем это свойство (как и почти все дальнейшие) только для дискретного распределения. Пусть $g(\xi)$ принимает значения $c_1, c_2, \ldots$ с вероятностями $\mathsf{P}(g(\xi)=c_m)=\sum\limits_{k:g(a_k)=c_m}\mathsf{P}(\xi=a_k)$. Тогда

$$\mathsf{E}\,g(\xi)=\sum_m c_m\mathsf{P}(g(\xi)=c_m)=\sum_m c_m\sum_{k:g(a_k)=c_m}\mathsf{P}(\xi=a_k)=\sum_m\sum_{k:g(a_k)=c_m}g(a_k)\mathsf{P}(\xi=a_k)=\sum_k g(a_k)\mathsf{P}(\xi=a_k).\qquad\square$$

E2. *Математическое ожидание постоянной равно этой постоянной*: $\mathsf{E}\,c=c$.

E3. *Постоянную можно вынести за знак математического ожидания*: $\mathsf{E}\,(c\xi)=c\mathsf{E}\,\xi$.

Доказательство. Следует из свойства **E1** при $g(x)=cx$. $\square$

E4. *Математическое ожидание суммы любых случайных величин ξ и η равно сумме их математических ожиданий*:

$$\mathsf{E}\,(\xi+\eta)=\mathsf{E}\,\xi+\mathsf{E}\,\eta.$$

Доказательство. Для величин с дискретным распределением: пусть x_k и y_n — значения ξ и η, соответственно. Для функции $g:\mathbb{R}^2\to\mathbb{R}$ можно доказать свойство, аналогичное **E1** (*сделать это!*). Пользуясь этим свойством для $g(x,y)=x+y$, запишем:

$$\mathsf{E}\,(\xi+\eta)=\sum_{k,n}(x_k+y_n)\mathsf{P}(\xi=x_k,\eta=y_n)=\sum_k x_k\underbrace{\sum_n\mathsf{P}(\xi=x_k,\eta=y_n)}_{\mathsf{P}(\xi=x_k)}+\sum_n y_n\underbrace{\sum_k\mathsf{P}(\xi=x_k,\eta=y_n)}_{\mathsf{P}(\eta=y_n)}=\mathsf{E}\,\xi+\mathsf{E}\,\eta.$$

$\square$

E5.

- Если $\xi \geqslant 0$ п.н. (*«почти наверное», то есть с вероятностью* 1: $\mathsf{P}(\xi \geqslant 0) = 1$), то $\mathsf{E}\,\xi \geqslant 0$;
- Если $\xi \geqslant 0$ п.н., и при этом $\mathsf{E}\,\xi = 0$, то $\xi = 0$ п.н., то есть $\mathsf{P}(\xi = 0) = 1$.

Упражнение. Доказать для дискретного распределения!

Следствие 12.

- *Если $\xi \leqslant \eta$ п.н., то $\mathsf{E}\,\xi \leqslant \mathsf{E}\,\eta$.*
- *Если $\xi \leqslant \eta$ п.н., и при этом $\mathsf{E}\,\xi = \mathsf{E}\,\eta$, то $\xi = \eta$ п.н.*

E6. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий*: если ξ и η независимы, то $\mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta$.

Доказательство.

$$\mathsf{E}\,(\xi\eta) = \sum_{k,n}(x_k y_n)\mathsf{P}(\xi = x_k, \eta = y_n) = \sum_k x_k \sum_n y_n \mathsf{P}(\xi = x_k)\mathsf{P}(\eta = y_n) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta.$$

□

Замечание 19. Обратное утверждение к свойству **E6** неверно: из равенства $\mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta$ *не следует* независимость величин ξ и η.

Пример 30. Пусть $\varphi \in \mathbf{U}_{0,2\pi}$, $\xi = \cos\varphi$, $\eta = \sin\varphi$ — заведомо зависимые случайные величины (*доказать!*). Но математическое ожидание их произведения равно произведению их математических ожиданий: по свойству **E1**

$$\mathsf{E}\,\xi = \int_0^{2\pi} \frac{1}{2\pi}\,\cos x\,dx = 0, \quad \mathsf{E}\,\eta = \int_0^{2\pi} \frac{1}{2\pi}\,\sin x\,dx = 0, \quad \mathsf{E}\,\xi\eta = \int_0^{2\pi} \frac{1}{2\pi}\,\cos x \sin x\,dx = 0 = \mathsf{E}\,\xi\mathsf{E}\,\eta.$$

11.3 Моменты старших порядков. Дисперсия

Определение 41. Если $\mathsf{E}\,|\xi|^k < \infty$, то число

$\mathsf{E}\,\xi^k$ называется *моментом порядка* k (k-м моментом) случайной величины ξ;

$\mathsf{E}\,|\xi|^k$ называется *абсолютным моментом порядка* k (абсолютным k-м моментом) случайной величины ξ;

$\mathsf{E}\,(\xi - \mathsf{E}\,\xi)^k$ называется *центральным моментом порядка* k (центральным k-м моментом) случайной величины ξ;

$\mathsf{E}\,|\xi - \mathsf{E}\,\xi|^k$ называется *абсолютным центральным моментом порядка* k (абсолютным центральным k-м моментом) случайной величины ξ.

Число $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2$ (центральный момент порядка 2) называется *дисперсией* случайной величины ξ

Пример 31. Пусть, скажем, случайная величина ξ принимает значение 0 с вероятностью $1 - 10^{-5}$, и значение 100 с вероятностью 10^{-5}. Посмотрим, как моменты разных порядков реагируют на большие, но маловероятные значения случайной величины.

$$\mathsf{E}\,\xi = 0 \cdot (1 - 10^{-5}) + 100 \cdot 10^{-5} = 10^{-3}; \qquad \mathsf{E}\,\xi^2 = 0^2 \cdot (1 - 10^{-5}) + 100^2 \cdot 10^{-5} = 10^{-1};$$

$$\mathsf{E}\,\xi^4 = 0^4 \cdot (1 - 10^{-5}) + 100^4 \cdot 10^{-5} = 1000; \qquad \mathsf{E}\,\xi^6 = 0^6 \cdot (1 - 10^{-5}) + 100^6 \cdot 10^{-5} = 10000000.$$

Пример 32. Дисперсия $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2$ есть «среднее значение квадрата отклонения случайной величины ξ от своего среднего». Посмотрим, за что эта величина отвечает.

Пусть случайная величина ξ принимает значения ± 1 с вероятностью $1/2$, а случайная величина η — значения ± 10 с вероятностью $1/2$. Тогда $\mathsf{E}\,\xi = \mathsf{E}\,\eta = 0$, поэтому $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 = 1$, $\mathsf{D}\,\eta = \mathsf{E}\,\eta^2 = 100$. Говорят, что дисперсия характеризует *степень разброса значений случайной величины вокруг ее математического ожидания*.

Если говорить о распределении случайной величины, как о распределении единичной массы по невесомому стержню, то дисперсия есть в точности момент инерции этого стержня, закрепленного в центре тяжести.

Определение 42. Если дисперсия величины ξ конечна, то число $\sigma = \sqrt{\mathsf{D}\,\xi}$ называют *среднеквадратическим отклонением* случайной величины ξ.

Чтобы прояснить связь моментов разных порядков, докажем несколько неравенств.

Теорема 26 (Неравенство Йенсена).

Пусть функция $g : \mathbb{R} \to \mathbb{R}$ выпукла (вниз :-))). Тогда для любой случайной величины ξ с конечным первым моментом

$$\mathsf{E}\, g(\xi) \geqslant g(\mathsf{E}\, \xi).$$

Нам понадобится следующее свойство выпуклых функций (то есть таких, что для любых $a < b$ при всяком $\alpha \in [0,1]$ верно $\alpha g(a) + (1-\alpha)g(b) \geqslant g(\alpha a + (1-\alpha)b)$):

Лемма 8. *Пусть функция $g : \mathbb{R} \to \mathbb{R}$ выпукла. Тогда для всякого y найдется число $c(y)$ такое, что при всех x*

$$g(x) \geqslant g(y) + c(y)(x-y).$$

Доказательство (предложено Дебеловым Алексеем, гр.871) .

Положим $c(y) = \inf_{x>y} \dfrac{(g(x) - g(y)}{x-y} \geqslant -\infty$. Тогда $g(x) - g(y) \geqslant c(y)(x-y)$ при $x > y$. Докажем, что это неравенство верно и при $x < y$, и заодно покажем, что $c(y)$ конечно.

Пусть $x_1 < y$. Тогда y принадлежит отрезку $[x_1; x_2]$ для любого $x_2 > y$, то есть существует $\alpha \in [0,1]$ такое, что

$$y = \alpha \cdot x_1 + (1-\alpha) \cdot x_2, \quad \text{или} \quad x_2 = \frac{y - \alpha x_1}{1-\alpha}. \tag{14}$$

Но в силу выпуклости функции g

$$g(y) \leqslant \alpha \cdot g(x_1) + (1-\alpha) \cdot g(x_2), \quad \text{или} \quad g(x_2) \geqslant \frac{g(y) - \alpha \cdot g(x_1)}{1-\alpha}. \tag{15}$$

Вспомним, что $c(y) = \inf_{x_2>y} \dfrac{g(x_2) - g(y)}{x_2 - y}$ и подставим вместо $g(x_2)$ и x_2 выражения, стоящие в правой части формул (15) и (14), соответственно:

$$c(y) = \inf_{x_2>y} \frac{g(x_2) - g(y)}{x_2 - y} \geqslant \frac{\dfrac{g(y) - \alpha g(x_1)}{1-\alpha} - g(y)}{\dfrac{y - \alpha x_1}{1-\alpha} - y} = \frac{g(y) - g(x_1)}{y - x_1}.$$

Последнее выражение заведомо конечно, то есть $c(y) > -\infty$. Более того, мы получили, что требуемое неравенство выполнено и для $x < y$:

$$c(y) \geqslant \frac{g(y)-g(x_1)}{y-x_1} \text{ для любых } x_1 < y, \text{ то есть } g(x_1) \geqslant g(y) + c(y)(x_1 - y). \qquad \square$$

Доказательство теоремы 26. Возьмем в условиях леммы $y = \mathsf{E}\,\xi$, $x = \xi$. Тогда $g(\xi) \geqslant g(\mathsf{E}\,\xi) + c(\mathsf{E}\,\xi)(\xi - \mathsf{E}\,\xi)$. Вычислим математическое ожидание обеих частей неравенства. Так как $\mathsf{E}\,(\xi - \mathsf{E}\,\xi) = 0$, и неравенство между математическими ожиданиями сохраняется по следствию **12**, то $\mathsf{E}\,g(\xi) \geqslant g(\mathsf{E}\,\xi)$. $\square$

Следующее свойство показывает, что из существования моментов больших порядков следует существование моментов меньших порядков.

Следствие 13. *Если* $\mathsf{E}\,|\xi|^t < \infty$, *то для любого* $0 < s < t$

$$\left(\mathsf{E}\,|\xi|^s\right)^t \leqslant \left(\mathsf{E}\,|\xi|^t\right)^s$$

Доказательство. Поскольку $0 < s < t$, то $g(x) = |x|^{t/s}$ — выпуклая функция. По неравенству Йенсена для $\eta = |\xi|^s$,

$$g(\mathsf{E}\,\eta) = (\mathsf{E}\,\eta)^{t/s} = \left(\mathsf{E}\,|\xi|^s\right)^{t/s} \leqslant \mathsf{E}\,g(\eta) = \mathsf{E}\,|\eta|^{t/s} = \mathsf{E}\,|\xi|^{s\cdot t/s} = \mathsf{E}\,|\xi|^t.$$

Возведя обе части неравенства в степень s, получим требуемое неравенство. $\square$

В частности, конечность второго момента (или дисперсии) влечет существование математического ожидания.

Следствие 14. *Если* $\mathsf{E}\,|\xi|^k < \infty$ *при некотором* $k > 1$, *то*

$$\mathsf{E}\,|\xi| \leqslant \sqrt[k]{\mathsf{E}\,|\xi|^k}.$$

11.4 Свойства дисперсии

Все свойства дисперсии следуют из соответствующих свойств математического ожидания.

D1. $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2$.

Действительно, $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2 = \mathsf{E}\,\left(\xi^2 - 2\xi\mathsf{E}\,\xi + (\mathsf{E}\,\xi)^2\right) = \mathsf{E}\,\xi^2 - 2\mathsf{E}\,\xi\mathsf{E}\,\xi + (\mathsf{E}\,\xi)^2 = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2$. $\square$

D2. $\mathsf{D}\,(c\xi) = c^2\mathsf{D}\,\xi$. *Доказать!*

D3.
- $\mathsf{D}\,\xi \geqslant 0$;
- $\mathsf{D}\,\xi = 0$ если и только если $\xi = const$ п.н.

Доказательство.
Дисперсия есть всего-навсего математическое ожидание п.н. неотрицательной с.в.: $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2$, и неотрицательность дисперсии следует из свойства E5.
По тому же свойству, $\mathsf{D}\,\xi = 0$ если и только если $(\xi - \mathsf{E}\,\xi)^2 = 0$ п.н., то есть $\xi = \xi$ п.н. □

D4. Дисперсия не меняется от сдвига с.в. на постоянную: $\mathsf{D}\,(\xi + c) = \mathsf{D}\,\xi$. *Доказать!*

D5. Если ξ и η независимы, то $\mathsf{D}\,(\xi + \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta$.

Действительно,

$$\mathsf{D}\,(\xi + \eta) = \mathsf{E}\,(\xi + \eta)^2 - (\mathsf{E}\,(\xi + \eta))^2 = \mathsf{E}\,\xi^2 + \mathsf{E}\,\eta^2 + 2\mathsf{E}\,(\xi\eta) - (\mathsf{E}\,\xi)^2 - (\mathsf{E}\,\eta)^2 - 2\mathsf{E}\,\xi\mathsf{E}\,\eta = \mathsf{D}\,\xi + \mathsf{D}\,\eta,$$

так как математическое ожидание произведения независимых с.в. равно произведению их математических ожиданий. □

Замечание 20. См. замечание **19**.

D6. Минимум среднеквадратического отклонения случайной величины ξ от точек вещественной прямой есть среднеквадратическое отклонение ξ от своего математического ожидания: $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2 = \min\limits_a \mathsf{E}\,(\xi - a)^2$.

Наименьший момент инерции стержня с распределенной на нем единичной массой получится, если точка вращения — центр тяжести стержня, а не любая другая точка.

Доказательство.

$$\mathsf{E}\,(\xi - a)^2 = \mathsf{E}\,\big((\xi - \mathsf{E}\,\xi) + (\mathsf{E}\,\xi - a)\big)^2 = \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 + 2\,\underbrace{\mathsf{E}\,(\xi - \mathsf{E}\,\xi)}_{0}\big(\mathsf{E}\,\xi - a\big) = \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 \geqslant \mathsf{D}\,\xi,$$

причем равенство достигается только для $a = \mathsf{E}\,\xi$.

□

11.5 Математические ожидания и дисперсии стандартных распределений

Пример 33. *Распределение Бернулли* B_p

$\mathsf{E}\,\xi = 1\cdot p + 0\cdot q = p;\quad \mathsf{E}\,\xi^2 = 1^2\cdot p + 0^2\cdot q = p;\quad \mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = p - p^2 = pq.$

Пример 34. *Биномиальное распределение* $\mathrm{B}_{n,p}$

Воспользуемся свойством устойчивости биномиального распределения относительно суммирования — леммой **5**. Возьмем n независимых случайных величин $\xi_1,\ldots,\xi_n$, имеющих распределение Бернулли $\mathrm{B}_p = \mathrm{B}_{1,p}$.
Тогда их сумма $S_n = \xi_1+\ldots+\xi_n$ имеет распределение $\mathrm{B}_{n,p}$.

$\mathsf{E}\,S_n = \sum\limits_{i=1}^n \mathsf{E}\,\xi_i = n\mathsf{E}\,\xi_1 = np$, так как все ξ_i одинаково распределены и их математическое ожидание равно p;

$\mathsf{D}\,S_n = \sum\limits_{i=1}^n \mathsf{D}\,\xi_i = n\mathsf{D}\,\xi_1 = npq$, поскольку ξ_i независимы и дисперсия каждой равна pq.

Пример 35. *Геометрическое распределение* G_p

При $p\in(0,1)$

$$\mathsf{E}\,\xi = \sum_{k=1}^\infty k\cdot pq^{k-1} = p\cdot\sum_{k=1}^\infty kq^{k-1} = p\cdot\sum_{k=1}^\infty (q^k)' = p\cdot\left(\sum_{k=1}^\infty q^k\right)' \stackrel{*}{=} p\cdot\left(\sum_{k=0}^\infty q^k\right)' = p\cdot\left(\frac{1}{1-q}\right)' = p\left(\frac{1}{(1-q)^2}\right) = \frac{1}{p}.$$

Равенство (*) появилось из-за нежелания дифференцировать сумму геометрической прогрессии, начинающейся не с 1, а с q. Заметьте, что производная у добавленных слагаемых равна 0, так что производные от этих двух сумм равны.

$$\mathsf{E}\,\xi^2 = \sum_{k=1}^\infty k^2\cdot pq^{k-1} = p\cdot\sum_{k=1}^\infty k(k-1)q^{k-1} + p\cdot\sum_{k=1}^\infty kq^{k-1} = pq\cdot\sum_{k=1}^\infty k(k-1)q^{k-2} + \mathsf{E}\,\xi = pq\cdot\sum_{k=1}^\infty \frac{\partial^2}{\partial q^2}q^k + \mathsf{E}\,\xi =$$

$$= pq\cdot\left(\sum_{k=0}^\infty \frac{\partial^2}{\partial q^2}(q^k)\right) + \mathsf{E}\,\xi = pq\cdot\frac{\partial^2}{\partial q^2}\left(\frac{1}{1-q}\right) + \mathsf{E}\,\xi = 2pq\left(\frac{1}{(1-q)^3}\right) + \mathsf{E}\,\xi = \frac{2q}{p^2} + \frac{1}{p}.$$

Поэтому

$$\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q-1+p}{p^2} = \frac{q}{p^2}.$$

Пример 36. *Распределение Пуассона* Π_λ

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

Доказать, что $\mathsf{E}\,\xi^2=\lambda^2+\lambda$, *так что* $\mathsf{D}\,\xi=\lambda$.

Пример 37. *Равномерное распределение* $\mathbf{U}_{a,b}$

$$\mathsf{E}\,\xi=\int\limits_{-\infty}^{\infty}xf_\xi(x)\,dx=\int\limits_{a}^{b}x\,\frac{1}{b-a}\,dx=\frac{a+b}{2};$$

$$\mathsf{E}\,\xi^2=\int\limits_{-\infty}^{\infty}x^2f_\xi(x)\,dx=\int\limits_{a}^{b}x^2\,\frac{1}{b-a}\,dx=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3};\quad \mathsf{D}\,\xi=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=\frac{(b-a)^2}{12}.$$

Пример 38. *Стандартное нормальное распределение* $\mathbf{N}_{0,1}$

$$\mathsf{E}\,\xi=\int\limits_{-\infty}^{\infty}xf_\xi(x)\,dx=\int\limits_{-\infty}^{\infty}x\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=0,$$

поскольку под интегралом стоит нечетная функция, и сам интеграл абсолютно сходится (за счет быстро убывающей $e^{-x^2/2}$).

$$\mathsf{E}\,\xi^2=\int\limits_{-\infty}^{\infty}x^2\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=2\int\limits_{0}^{\infty}x^2\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=-2\int\limits_{0}^{\infty}x\,\frac{1}{\sqrt{2\pi}}\,de^{-x^2/2}=\underbrace{-2x\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\Big|_0^{\infty}}_{0}+\underbrace{2\int\limits_{0}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx}_{1}=1.$$

Последнее равенство следует из того, что $2\int\limits_0^\infty\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx=\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx$, а интеграл по всей прямой от плотности *любого* распределения равен 1. Поэтому $\mathsf{D}\,\xi=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=1-0=1$.

Пример 39. *Нормальное распределение* $\mathbf{N}_{a,\sigma^2}$

Мы знаем, что если $\xi \Subset \mathbf{N}_{a,\sigma^2}$, то $\eta = \dfrac{\xi - a}{\sigma} \Subset \mathbf{N}_{0,1}$, и $\mathsf{E}\,\eta = 0$, $\mathsf{D}\,\eta = 1$. Поэтому

$$\boxed{\mathsf{E}\,\xi = \mathsf{E}\,(\sigma\eta + a) = \sigma\mathsf{E}\,\eta + a = a; \qquad \mathsf{D}\,\xi = \mathsf{D}\,(\sigma\eta + a) = \sigma^2\mathsf{D}\,\eta = \sigma^2.} \tag{16}$$

Какими свойствами математического ожидания и дисперсии мы воспользовались в формуле (16)?

Пример 40. *Показательное (экспоненциальное) распределение* $\mathbf{E}_\alpha$

Найдем для произвольного $k \in \mathbb{N}$ момент порядка k.

$$\mathsf{E}\,\xi^k = \int\limits_{-\infty}^{\infty} x^k f_\xi(x)\,dx = \int\limits_{0}^{\infty} x^k\,\alpha\,e^{-\alpha x}\,dx = \frac{1}{\alpha^k}\int\limits_{0}^{\infty}(\alpha x)^k\,e^{-\alpha x}\,d(\alpha x) = \frac{k!}{\alpha^k}.$$

В последнем равенстве мы воспользовались гамма-функцией Эйлера: $\Gamma(k+1) = \int\limits_0^\infty u^k\,e^{-u}\,du = k!$ Соответственно,

$$\mathsf{E}\,\xi = \frac{1}{\alpha}, \quad \mathsf{E}\,\xi^2 = \frac{2}{\alpha^2}, \quad \mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = \frac{1}{\alpha^2}.$$

Пример 41. *Стандартное распределение Коши* $\mathbf{C}_{0,1}$

Распределение Коши. Говорят, что ξ имеет распределение Коши с параметрами a, σ^2, где $a \in \mathbb{R}$, $\sigma > 0$, и пишут (по крайней мере мы так будем писать) $\xi \Subset \mathbf{C}_{a,\sigma^2}$, если

$$f_\xi(x) = \frac{\sigma}{\pi(\sigma^2 + (x-a)^2)} \quad \text{для всех} \quad x \in \mathbb{R}.$$

Распределение Коши имеет, например, абсцисса точки пересечения луча, посланного из точки (a, σ) под наудачу выбранным углом $\varphi \Subset \mathbf{U}_{-\pi/2,\pi/2}$, с осью OX. *Это полезно доказать!*

Математическое ожидание для распределения Коши не существует, поскольку $\mathsf{E}\,|\xi| = \int\limits_{-\infty}^{\infty} |x|\dfrac{1}{\pi(1+x^2)}\,dx$ расходится (подынтегральная функция ведет себя на бесконечности как $1/x$).

Пример 42. *Распределение Парето*

Распределение Парето. Говорят, что ξ имеет распределение Парето с параметрами x_0, s, где $x_0 > 0$, $s > 0$, если

$$F_\xi(x) = \begin{cases} 1 - \left(\dfrac{x_0}{x}\right)^s, & x \geqslant x_0; \\ 0, & x < x_0. \end{cases} \qquad \text{или} \qquad f_\xi(x) = \begin{cases} s\,\dfrac{x_0^s}{x^{s+1}}, & x \geqslant x_0; \\ 0, & x < x_0. \end{cases}.$$

У распределения Парето существуют только моменты порядка $u < s$, поскольку $\mathsf{E}\,|\xi|^u = \int\limits_{x_0}^{\infty} x^u\, s\,\dfrac{x_0^s}{x^{s+1}}\,dx$ сходится при $u < s$, то есть когда подынтегральная функция на бесконечности бесконечно мала по сравнению с $1/x$.

Посчитать момент порядка $u < s$ распределения Парето. При каких s у этого распределения существует дисперсия?

Раздел 12. Числовые характеристики зависимости случайных величин

12.1 Чем отличается дисперсия суммы от суммы дисперсий?

Мы знаем, что для независимых с. в. с конечными вторыми моментами дисперсия их суммы равна сумме их дисперсий. Чему равна дисперсия суммы в общем случае?

$$\mathsf{D}\,(\xi+\eta)=\mathsf{E}\,(\xi+\eta)^2-\big(\mathsf{E}\,(\xi+\eta)\big)^2=\mathsf{E}\left(\xi^2+\eta^2+2\xi\eta\right)-\big(\mathsf{E}\,\xi\big)^2-\big(\mathsf{E}\,\eta\big)^2-2\mathsf{E}\,\xi\mathsf{E}\,\eta=\mathsf{D}\,\xi+\mathsf{D}\,\eta+2\left(\mathsf{E}\,(\xi\eta)-\mathsf{E}\,\xi\mathsf{E}\,\eta\right). \qquad (17)$$

Величина $\mathsf{E}\,(\xi\eta)-\mathsf{E}\,\xi\mathsf{E}\,\eta$ равняется нулю, если случайные величины ξ и η независимы (свойство **E6** математического ожидания). С другой стороны, из равенства ее нулю вовсе не следует независимость, как показывает пример **30**. Оказывается, что эту величину часто используют как «индикатор наличия зависимости» пары с. в.

Определение 43. *Ковариацией* $\mathrm{cov}(\xi,\eta)$ случайных величин ξ и η называется *число*

$$\mathrm{cov}(\xi,\eta)=\mathsf{E}\,\big((\xi-\mathsf{E}\,\xi)(\eta-\mathsf{E}\,\eta)\big).$$

Свойство 12. $\mathrm{cov}(\xi,\eta)=\mathsf{E}\,\big((\xi-\mathsf{E}\,\xi)(\eta-\mathsf{E}\,\eta)\big)=\mathsf{E}\,(\xi\eta)-\mathsf{E}\,\xi\mathsf{E}\,\eta.$

Упражнение 19. Доказать свойство **12**, пользуясь свойствами математического ожидания.

Свойство 13. а) $\mathrm{cov}(\xi,\xi)=\mathsf{D}\,\xi$; б) $\mathrm{cov}(\xi,\eta)=\mathrm{cov}(\eta,\xi)$.

Свойство 14. *Дисперсия суммы нескольких случайных величин вычисляется по любой из следующих формул:*

$$\mathsf{D}\,(\xi_1+\ldots+\xi_n) \;=\; \sum_{i=1}^{n}\mathsf{D}\,\xi_i \;+\; \sum_{i\neq j}\mathrm{cov}(\xi_i,\xi_j) \;=\; \sum_{i=1}^{n}\mathsf{D}\,\xi_i \;+\; 2\sum_{i<j}\mathrm{cov}(\xi_i,\xi_j) \;=\; \sum_{i,j}\mathrm{cov}(\xi_i,\xi_j).$$

Упражнение 20. Доказать свойство **14**, пользуясь равенствами

$$(a+b)^2 \;=\; a^2+b^2+ab+ba \;=\; a^2+b^2+2ab \;=\; aa+bb+ab+ba$$

и получив аналогичные равенства для квадрата суммы n слагаемых.

Обсудим достоинства и недостатки ковариации, как величины, характеризующей зависимость двух с. в.

1. Если ковариация $\mathrm{cov}(\xi,\eta)$ отлична от нуля, то величины ξ и η *зависимы*!

2. С гарантией о наличии зависимости мы можем судить, если знаем совместное распределение пары ξ и η, и можем проверить, равна ли (например) плотность совместного распределения произведению плотностей.
Но найти совместное распределение часто бывает сложнее, чем посчитать математическое ожидание произведения ξ и η. Если нам повезет, и математическое ожидание произведения ξ и η не будет равняться произведению их мат. ожиданий, мы скажем, что ξ и η зависимы *не находя* их совместного распределения!

Пример 43. Покажем, что с помощью ковариации можно судить о зависимости даже когда для вычисления соместного распределения недостаточно данных.

Пусть ξ и η — *независимые* случайные величины, и дисперсия ξ отлична от нуля (*то есть?*). Докажем, что ξ и $\xi+\eta$ зависимы.

$$\mathsf{E}\big(\xi(\xi+\eta)\big) = \mathsf{E}\,\xi^2 + \mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\eta; \quad \mathsf{E}\,\xi\,\mathsf{E}\,(\xi+\eta) = (\mathsf{E}\,\xi)^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\eta. \tag{18}$$

Поэтому $\operatorname{cov}(\xi,\xi+\eta) = \mathsf{E}\,\xi^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\eta - \big((\mathsf{E}\,\xi)^2 + \mathsf{E}\,\xi\,\mathsf{E}\,\eta\big) = \mathsf{D}\,\xi > 0$. Следовательно, ξ и $\xi+\eta$ зависимы.

3. Жаль, что величина $\operatorname{cov}(\xi,\eta)$ не является «безразмерной»: если ξ — объем газа в сосуде, а η — давление этого газа, то ковариация измеряется в кубометрах×Паскали :). Иначе говоря, при умножении одной из величин ξ, η на какое-нибудь число ковариация тоже умножается на это число. Но умножение на число не сказывается на «степени зависимости» величин (они от этого «более зависимыми» не становятся), так что большое значение ковариации не означает более сильной зависимости.
Нужно как-то нормировать ковариацию, получив из нее «безразмерную» величину, абсолютное значение которой
- а) не менялось бы при умножении или сдвиге случайных величин на число;
- б) свидетельствовало бы о «силе зависимости» с. в.

Говоря о «силе» зависимости между с. в., мы имеем ввиду следующее. Самая сильная зависимость — функциональная, а из функциональных — линейная зависимость, когда $\xi = a\eta + b$ п. н. Бывают гораздо более слабые зависимости. Так, если по последовательности независимых случайных величин $\xi_1, \xi_2, \ldots$ построить величины $\xi = \xi_1 + \ldots + \xi_{24} + \xi_{25}$ и $\eta = \xi_{25} + \xi_{26} + \ldots + \xi_{90}$, то эти величины зависимы, но очень «слабо зависимы»: через одно-единственное общее слагаемое ξ_{25}. Сильно ли зависимы число гербов в первых 25 подбрасываниях монеты и число гербов в той же серии, но в испытаниях с 25-го по 90-е?

Итак, следующая величина есть всего лишь ковариация, нормированная нужным образом.

12.2 Коэффициент корреляции

Определение 44. *Коэффициентом корреляции* $\rho(\xi,\eta)$ случайных величин ξ и η, дисперсии которых существуют и отличны от нуля, называется *число*

$$\rho(\xi,\eta)=\frac{\operatorname{cov}(\xi,\eta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\eta}}.$$

Замечание 21. Чтобы разглядеть «устройство» коэффициента корреляции, распишем по определению числитель и знаменатель:

$$\rho(\xi,\eta)=\frac{\mathsf{E}\left((\xi-\mathsf{E}\,\xi)(\eta-\mathsf{E}\,\eta)\right)}{\sqrt{\mathsf{E}\left(\xi-\mathsf{E}\,\xi\right)^2}\sqrt{\mathsf{E}\left(\eta-\mathsf{E}\,\eta\right)^2}}.$$

Здесь математикам уместно провести аналогии с «косинусом угла» между двумя элементами гильбертова пространства, образованного случайными величинами с конечным вторым моментом, снабженного скалярным произведением $\operatorname{cov}(\xi,\eta)$ *и «нормой», равной корню из дисперсии случайной величины, или корню из скалярного произведения* $\operatorname{cov}(\xi,\xi)$.

Пример 44. Рассмотрим продолжение примера **43**, но пусть ξ и η будут не только независимыми, но и одинаково распределенными случайными величинами, и их дисперсия отлична от нуля. Найдем коэффициент корреляции величин ξ и $\xi+\eta$. Согласно формуле (18),

$$\operatorname{cov}(\xi,\xi+\eta)=\mathsf{E}\,\xi^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta-\left((\mathsf{E}\,\xi)^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta\right)=\mathsf{E}\,\xi^2-(\mathsf{E}\,\xi)^2=\mathsf{D}\,\xi.$$

Поэтому

$$\rho(\xi,\xi+\eta)=\frac{\operatorname{cov}(\xi,\xi+\eta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,(\xi+\eta)}}=\frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\xi+\mathsf{D}\,\eta}}=\frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{2\mathsf{D}\,\xi}}=\frac{1}{\sqrt{2}}.$$

Иначе говоря, коэффициент корреляции величин ξ и $\xi+\eta$ равен косинусу угла $45°$, образованного «векторами» ξ и $\xi+\eta$, где «$\xi\perp\eta$» и их «длина» одинакова.

Упражнение 21. Чтобы аналогия не заходила слишком далеко, и у читателя не возникло искушения любые случайные величины рисовать стрелочками на плоскости и вместо подсчета математических ожиданий измерять углы, предлагаю убедиться, например, что коэффициент корреляции величин ξ и ξ^2 равен:

а) нулю, если ξ имеет нормальное распределение с нулевым средним;

б) $2/\sqrt{5}$, если ξ имеет показательное распределение с любым параметром.

Определение 45. Случайные величины ξ и η называют *некоррелированными*, если $\operatorname{cov}(\xi,\eta)=0$ (или если $\rho(\xi,\eta)=0$, — в том случае, когда коэффициент корреляции существует).

Замечание 22. Если одна из величин ξ и η — постоянная, то эти величины независимы (*проверить по определению!*), и $\operatorname{cov}(\xi, \eta) = 0$ (*проверить по определению!*). Естественно в этом случае тоже полагать, что ξ и η «некоррелированы», хотя коэффициент корреляции не определен (дисперсия постоянной равна 0).

Упражнение 22. А что будет, если доопределить коэффициент корреляции нулем, если хотя бы одна из величин — постоянная? Предлагаю подумать, какими достоинствами и недостатками обладает такое «раскрытие неопределенности типа $\frac{0}{0}$».

12.3 Свойства коэффициента корреляции

Всюду далее специально не оговаривается, но предполагается, что коэффициент корреляции существует.

Теорема 27.

Коэффициент корреляции обладает следующими свойствами.

1. *Если с. в. ξ и η независимы, то $\rho(\xi, \eta) = \operatorname{cov}(\xi, \eta) = 0$.*

2. $|\rho(\xi, \eta)| \leqslant 1$.

3. $|\rho(\xi, \eta)| = 1$, *если и только если с. в. ξ и η с вероятностью 1 линейно связаны, т.е. существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\eta = a\xi + b) = 1$.*

Доказательство.

1. Свойство **1** мы уже много раз (*сколько?*) упоминали и один раз доказали.

2. Для доказательства **2** нам понадобится одно преобразование, называемое *«стандартизацией»* случайной величины: с его помощью из с. в. с конечным вторым моментом (не постоянной) получают с. в. с нулевым математическим ожиданием (*«центрированную»*) и единичной дисперсией (*«нормированную»*).

Определение 46. Пусть $\mathsf{D}\xi$ конечна и отлична от нуля. Определим случайную величину $\tilde{\xi}$ так:

$$\tilde{\xi} = \frac{\xi - \mathsf{E}\xi}{\sqrt{\mathsf{D}\xi}}.$$

Преобразование $\xi \mapsto \tilde{\xi}$ называется *стандартизацией* случайной величины ξ, а сама с. в. $\tilde{\xi}$ называется *стандартизованной*, или (слэнг!) *центрированной и нормированной* версией с. в. ξ.

Упражнение 23. Обяснить, будет ли распределение $\tilde{\xi}$
а) нормальным, если ξ распределена по нормальному закону; б) равномерным, если ξ имеет равномерное распределение;
в) биномиальным, если ξ имеет биномиальное распределение; г) показательным, если ξ имеет показательное распределение; (и т.д.)

Свойство 15. *Стандартизованная с. в. $\tilde{\xi}$ имеет нулевое математическое ожидание и единичную дисперсию.*

Доказательство*.* Воспользуемся свойствами математического ожидания и дисперсии:

$$\mathsf{E}\,\tilde{\xi} = \mathsf{E}\left(\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}\right) = \frac{1}{\sqrt{\mathsf{D}\,\xi}}\;\mathsf{E}\,(\xi - \mathsf{E}\,\xi) = \frac{1}{\sqrt{\mathsf{D}\,\xi}}\;(\mathsf{E}\,\xi - \mathsf{E}\,\xi) = 0;$$

$$\mathsf{D}\,\tilde{\xi} = \mathsf{D}\left(\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}\right) = \frac{1}{\mathsf{D}\,\xi}\;\mathsf{D}\,(\xi - \mathsf{E}\,\xi) = \frac{1}{\mathsf{D}\,\xi}\;\mathsf{D}\,\xi = 1.$$

Не забудьте у каждого знака равенства написать, в силу какого свойства, утверждения или определения это равенство верно! □

Возвращаясь к доказательству **2**, заметим, что

$$\rho(\xi,\eta) = \frac{\mathsf{E}\big((\xi - \mathsf{E}\,\xi)(\eta - \mathsf{E}\,\eta)\big)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\eta}} = \mathsf{E}\left(\frac{(\xi - \mathsf{E}\,\xi)(\eta - \mathsf{E}\,\eta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\eta}}\right) = \mathsf{E}\left(\tilde{\xi}\,\tilde{\eta}\right),$$

где $\tilde{\xi} = \dfrac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}$ и $\tilde{\eta} = \dfrac{\eta - \mathsf{E}\,\eta}{\sqrt{\mathsf{D}\,\eta}}$ — стандартизованные версии с. в. ξ и η.

Теперь воспользуемся неравенством $0 \leqslant (a-b)^2 = a^2 - 2ab + b^2$, или $ab \leqslant \dfrac{1}{2}(a^2+b^2)$. Подставим $\tilde{\xi}$ вместо a, $\tilde{\eta}$ вместо b и возьмем математические ожидания от обеих частей неравенства:

$$\rho(\xi,\eta) = \mathsf{E}\left(\tilde{\xi}\,\tilde{\eta}\right) \leqslant \frac{1}{2}\;\mathsf{E}\left(\tilde{\xi}^2 + \tilde{\eta}^2\right) = \frac{1}{2}\;\left(\mathsf{D}\,\tilde{\xi} + \left(\mathsf{E}\,\tilde{\xi}\right)^2 + \mathsf{D}\,\tilde{\eta} + \left(\mathsf{E}\,\tilde{\eta}\right)^2\right) = \frac{1}{2}\cdot 2 = 1. \tag{19}$$

Пользуясь точно так же неравенством $0 \leqslant (a+b)^2 = a^2 + 2ab + b^2$, или $ab \geqslant -\dfrac{1}{2}(a^2+b^2)$, получим

$$\rho(\xi,\eta) = \mathsf{E}\left(\tilde{\xi}\,\tilde{\eta}\right) \geqslant -\frac{1}{2}\;\mathsf{E}\left(\tilde{\xi}^2 + \tilde{\eta}^2\right) = -\frac{1}{2}\cdot 2 = -1. \tag{20}$$

Таким образом, $|\rho(\xi,\eta)| \leqslant 1$, что и требовалось доказать.

3. В одну сторону утверждение проверяется непосредственно:

Воспользоваться свойствами математического ожидания и дисперсии и доказать, что

$$\rho(\xi, a\xi + b) = \begin{cases} 1, & a > 0; \\ -1, & a < 0. \end{cases}$$

Не забудьте, что** $\sqrt{a^2} = |a|$**, а не просто** a**!

Докажем вторую часть: если $|\rho(\xi,\eta)| = 1$, то существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\eta = a\xi + b) = 1$.

Рассмотрим сначала случай $\rho(\xi,\eta) = 1$. Это возможно только если единственное неравенство в формуле (19) превращается в равенство: $\mathsf{E}\left(\tilde{\xi}\,\tilde{\eta}\right) = \dfrac{1}{2}\,\mathsf{E}\left(\tilde{\xi}^2 + \tilde{\eta}^2\right)$, или

$$\mathsf{E}\left(\tilde{\xi} - \tilde{\eta}\right)^2 = 0.$$

Но по свойству **E5** математического ожидания равенство нулю мат. ожидания неотрицательной с. в. означает, что эта величина п.н. равна нулю:

$$\mathsf{P}\left(\tilde{\xi} - \tilde{\eta} = 0\right) \; = \; 1 \; = \; \mathsf{P}\left(\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = \frac{\eta - \mathsf{E}\,\eta}{\sqrt{\mathsf{D}\,\eta}}\right) = \mathsf{P}\Bigg(\eta = \underbrace{\frac{\sqrt{\mathsf{D}\,\eta}}{\sqrt{\mathsf{D}\,\xi}}}_{a}\;\xi \underbrace{- \frac{\sqrt{\mathsf{D}\,\eta}\,\mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} + \mathsf{E}\,\eta}_{b}\Bigg).$$

В случае $\rho(\xi,\eta) = -1$ нужно рассмотреть единственное неравенство в формуле (20) и повторить рассуждения. Тем самым теорема **27** доказана. □

Полезно знать следующие часто употребляемые термины.

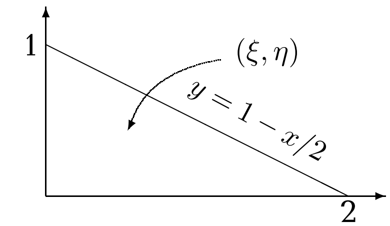
Определение 47. Говорят, что величины ξ и η *отрицательно коррелированы*, если $\rho(\xi,\eta) < 0$; говорят, что величины ξ и η *положительно коррелированы*, если $\rho(\xi,\eta) > 0$.

Смысл знака коэффициента корреляции особенно ясен в случае $\rho(\xi,\eta) = \pm 1$. Тогда знак ρ равен знаку a в равенстве $\eta = a\xi + b$ п.н. То есть $\rho(\xi,\eta) = 1$ означает, что чем больше ξ, тем больше и η. Напротив, $\rho(\xi,\eta) = -1$ означает, что чем больше ξ, тем меньше η. Похожим образом можно трактовать знак коэффициента корреляции и в случае, когда $|\rho(\xi,\eta)| < 1$, помня при этом, что зависимость величин ξ и η теперь уже не линейная и, возможно, даже не функциональная.

Так, величины ξ и $\xi + \eta$ в примерах **43** и **44** положительно коррелированы, но их зависимость не функциональная.

Пример 45.

Если с. в. ξ и η есть координаты точки, брошенной наудачу в треугольник с вершинами $(2,0)$, $(0,0)$ и $(0,1)$, то коэффициент корреляции $\rho(\xi,\eta)$ отрицателен. Это можно объяснить «на пальцах» так: *чем больше ξ, тем меньше у η возможностей быть большой.* :-) Предлагаю убедиться в этом, проверив справедливость следующих высказываний. Во-первых,



$$f_\xi(x)=\begin{cases}1-\dfrac{x}{2}, & 0\leqslant x\leqslant 2;\\ 0, & \text{иначе };\end{cases}\qquad \mathsf{E}\,\xi=\frac{2}{3};\qquad f_\eta(y)=\begin{cases}2-2y, & 0\leqslant y\leqslant 1;\\ 0, & \text{иначе };\end{cases}\qquad \mathsf{E}\,\eta=\frac{1}{3}.$$

Во-вторых,

совместное распределение координат точки, брошенной наудачу в произвольную (измеримую) область D на плоскости имеет постоянную плотность во всех точках области D.
Это связано с понятием «наудачу»: вероятность попасть в любую область $A\subset D$, с одной стороны, зависит только от площади A, и не зависит от формы и положения A внутри D, равняясь, с другой стороны, интегралу по области A от плотности совместного распределения координат точки.
Эти два качества возможно совместить, только если плотность совместного распределения постоянна внутри D.
Более того, эта постоянная, как легко видеть, есть просто $\dfrac{1}{\text{площадь } D}$ (хотя бы потому, что интеграл от нее по всей области D должен равняться вероятности попасть в D, или единице).

Распределение точки, брошенной наудачу в область (все равно где), называют ***равномерным*** распределением. Итак, плотность равномерного распределения в произвольной области на плоскости — постоянная, равная $(1/\textit{площадь области})$ для точек внутри области и нулю — вне. Поэтому (а также потому, что площадь этого треугольника равна 1)

$$\mathsf{E}\,(\xi\,\eta)=\iint\limits_{\triangleright} x\cdot y\cdot 1\,dy\,dx=\int\limits_0^2\left(\int\limits_0^{1-x/2} x\cdot y\,dy\right)dx=\text{(кажется)}\ \ \frac{1}{6}.$$

То есть ковариация (а с ней и коэффициент корреляции) отрицательна (*посчитать* $cov(\xi,\eta)$).

Упражнение 24. А верно ли, что коэффициент корреляции в примере **45** существует? Какие свойства случайных величин гарантируют конечность второго момента? А из *ограниченности* с. в. следует ли существование каких-нибудь моментов? Каких и почему?

Пример 46.

Найти коэффициент корреляции между числом выпадений единицы и числом выпадений шестерки при n подбрасываниях симметричного кубика.

Р е ш е н и е. Обозначим для $i = 1, 2, 3, 4, 5, 6$ через ξ_i случайную величину, равную числу выпадений грани с i очками при n подбрасываниях кубика. Посчитаем $\mathrm{cov}(\xi_1, \xi_6)$.

Каждая из случайных величин ξ_i имеет биномиальное распределение с параметрами n и $1/6$, поэтому $\mathsf{E}\,\xi_i = n/6$, $\mathsf{D}\,\xi_i = 5n/36$.

Заметим, что сумма $\xi_1 + \dots + \xi_6$ этих величин равна n. В силу симметрии кубика, все математические ожидания $\mathsf{E}\,\xi_1\xi_2$, $\mathsf{E}\,\xi_1\xi_3$, … , $\mathsf{E}\,\xi_1\xi_6$ одинаковы (но, скорее всего, отличаются от $\mathsf{E}\,\xi_1\xi_1 = \mathsf{E}\,\xi_1^2 = \mathsf{D}\,\xi_1 + (\mathsf{E}\,\xi_1)^2 = 5n/36 + n^2/36$).

Посчитаем $\mathsf{E}\,\xi_1(\xi_1 + \dots + \xi_6)$. С одной стороны, это равно

$$\mathsf{E}\,\xi_1(\xi_1 + \dots + \xi_6) = \mathsf{E}\,\xi_1 \cdot n = n^2/6,$$

с другой стороны,

$$\mathsf{E}\,\xi_1(\xi_1 + \dots + \xi_6) = \mathsf{E}\,\xi_1^2 + 5\mathsf{E}\,\xi_1\xi_6 = 5n/36 + n^2/36 + 5\mathsf{E}\,\xi_1\xi_6.$$

Отсюда $5\mathsf{E}\,\xi_1\xi_6 = n^2/6 - 5n/36 - n^2/36$, то есть $\mathsf{E}\,\xi_1\xi_6 = (n^2 - n)/36$.

Следовательно, искомый коэффициент корреляции равен

$$\rho(\xi_1, \xi_6) = \frac{\mathsf{E}\,\xi_1\xi_6 - \mathsf{E}\,\xi_1\mathsf{E}\,\xi_6}{\sqrt{\mathsf{D}\,\xi_1\mathsf{D}\,\xi_6}} = \frac{(n^2 - n)/36 - n^2/36}{5n/36} = -\frac{1}{5}.$$

Интересно, что полученный коэффициент корреляции не зависит от n.

Почему коэффициент корреляции $\rho(\xi_1, \xi_6)$ отрицателен?

... Откуда, наконец, вытекает то удивительное, по-видимому, следствие, что, если бы наблюдения над всеми событиями продолжать всю вечность, причем вероятность, наконец, перешла бы в полную достоверность, то было бы замечено, что в мире все управляется точными отношениями и постоянным законом изменений, так что даже в вещах, в высшей степени случайных, мы принуждены были бы признать как бы некоторую необходимость и, скажу я, рок.

Якоб Бернулли, *Ars conjectandi* (1713)

Раздел 13. Куда и как сходятся последовательности случайных величин

13.1 Сходимость «почти наверное» и «по вероятности»

Напомню, что случайная величина есть (измеримая) функция из некоторого абстрактного множества Ω в множество действительных чисел. Последовательность случайных величин есть, тем самым, последовательность функций (определенных на одном и том же пространстве элементарных исходов Ω). И если мы хотим говорить о сходимости последовательности случайных величин $\{\xi_n\}_{n=1}^{\infty}$, не будем забывать, что мы имеем дело не с последовательностью чисел, а с последовательностью ***функций***. Существуют разные виды сходимости последовательности функций. Всякий раз давать определение какой-либо сходимости мы будем, опираясь на ***сходимость числовых последовательностей*** как на уже известное основное понятие.

В частности, при каждом новом $\omega \in \Omega$ мы имеем новую ***числовую*** последовательность $\{\xi_n(\omega)\}_{n=1}^{\infty}$. Поэтому, во-первых, можно говорить о знакомой из математического анализа (почти) поточечной сходимости последовательностей функций: о сходимости «почти всюду», которую в теории вероятностей называют сходимостью «почти наверное».

Определение 48. Говорят, что последовательность с. в. $\{\xi_n\}$ ***сходится почти наверное*** к с. в. ξ при $n \to \infty$, и пишут: $\xi_n \to \xi$ п. н., если $\mathsf{P}\{\omega \,:\, \xi_n(\omega) \to \xi(\omega) \text{ при } n \to \infty\} = 1$. Иначе говоря, если $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$, кроме, возможно, $\omega \in A$, где множество (событие) A имеет нулевую вероятность.

Заметим сразу: чтобы говорить о сходимости «почти наверное», требуется (по крайней мере, по определению) знать, как устроены отображения $\omega \mapsto \xi_n(\omega)$. В задачах же теории вероятностей, как правило, известны ***не сами случайные величины, а лишь их распределения***. Известно, то есть, какова вероятность тех элементарных исходов ω, для которых $\xi_n(\omega)$ принимает значения в заданном множестве.

Можем ли мы, обладая только информацией о распределениях, говорить о какой-либо сходимости последовательности случайных величин $\{\xi_n\}$ к с. в. ξ?

Можно, например, потребовать, чтобы вероятность («доля») тех элементарных исходов ω, для которых $\xi_n(\omega)$ не попадает в «ε-окрестность» числа $\xi(\omega)$, уменьшалась до нуля с ростом n. Такая сходимость в функциональном анализе называется сходимостью «по мере», а в теории вероятностей — сходимостью «по вероятности».

Определение 49. Говорят, что последовательность с. в. $\{\xi_n\}$ *сходится по вероятности* к с. в. ξ при $n\to\infty$, и пишут: $\xi_n \xrightarrow{\text{p}} \xi$, если для любого $\varepsilon>0$

$$\mathsf{P}\left(|\xi_n-\xi|>\varepsilon\right)\to 0 \ \text{ при } \ n\to\infty \qquad \text{или} \qquad \mathsf{P}\left(|\xi_n-\xi|\leqslant\varepsilon\right)\to 1 \ \text{ при } \ n\to\infty.$$

Пример 47. Рассмотрим последовательность с. в. $\xi_1,\xi_2,\ldots$, в которой все величины имеют *разные* распределения: с. в. ξ_n, $n\geqslant 1$, принимает значения 0 и n^7 с вероятностями $\mathsf{P}\big(\xi_n=n^7\big)=1/n=1-\mathsf{P}(\xi_n=0)$. Докажем, что эта последовательность сходится по вероятности к случайной величине, равной нулю п. н. (к нулю, проще говоря).

Действительно, зафиксируем произвольное $\varepsilon>0$. Для всех n начиная с некоторого n_0 такого, что $n_0^7>\varepsilon$, верно равенство $(*)$ ниже

$$\mathsf{P}\big(|\xi_n-0|>\varepsilon\big) \overset{\xi_n\geqslant 0}{=} \mathsf{P}\big(\xi_n>\varepsilon\big) \overset{(*)}{=} \mathsf{P}\big(\xi_n=n^7\big) = \frac{1}{n} \to 0 \ \text{ при } \ n\to\infty.$$

Итак, случайные величины ξ_n с ростом n могут принимать все бо́льшие и бо́льшие значения, но со все меньшей и меньшей вероятностью.

А сходится ли данная последовательность к нулю «почти наверное»? Вопрос не слишком корректный, поскольку заданы не случайные величины, а лишь их распределения, и ответ на него, как правило, зависит от того, как сами величины взаимосвязаны. Если, скажем, $\xi_n(\omega)=0$ для $\omega\in[0,1-1/n]$ и $\xi_n(\omega)=n^7$ для $\omega\in(1-1/n,1]$, то сходимость «почти наверное» имеет место, так как для всякого ω начиная с некоторого n_0 все $\xi_n(\omega)$ равны нулю.
Попробуйте задать случайные величины ξ_n на $[0,1]$ так, чтобы сходимость «почти наверное» не имела место. Для этого нужно заставить отрезок длины $1/n$, на котором $\xi_n(\omega)=n^7$, «бегать» по отрезку $[0,1]$, чтобы любая точка $\omega\in[0,1]$ попадала внутрь этого отрезка бесконечное число раз. Воспользуйтесь тем, что гармонический ряд расходится. Если вам мешают концы отрезка, их можно склеить в окружность :)
Заметим однако, что если вероятности $\mathsf{P}(\xi_n=n^7)$ сходятся к нулю достаточно быстро (например, равны $1/n^2$), то сходимость к нулю п. н. всегда имеет место (см., например, теорему 2 §1 гл. 6 *на стр.* 134 *учебника А.А.Боровкова «Теория вероятностей»).*

Замечание 23. Сходимость по вероятности не обязательно сопровождается сходимостью математических ожиданий или моментов других порядков: из $\xi_n\xrightarrow{\text{p}}\xi$ *не следует*, что $\mathsf{E}\,\xi_n\to\mathsf{E}\,\xi$.

Действительно, в примере **47** имеет место сходимость $\xi_n\xrightarrow{\text{p}}\xi=0$, но $\mathsf{E}\,\xi_n=n^6\not\to\mathsf{E}\,\xi=0$.

Если вместо значения n^7 взять, скажем, n (с той же вероятностью $1/n$), получим $\mathsf{E}\,\xi_n=1\not\to\mathsf{E}\,\xi=0$.

А если ξ_n принимает значения 0 и $\sqrt{n}$ с теми же вероятностями, что и в примере **47**, то $\mathsf{E}\,\xi_n=1/\sqrt{n}\to\mathsf{E}\,\xi=0$, но уже вторые моменты сходиться ко второму моменту ξ не будут: $\mathsf{E}\,\xi_n^2=1\not\to\mathsf{E}\,\xi^2=0$.

Сходимость по вероятности обладает обычными для сходимостей свойствами. Например, такими.

Свойство 16. *Если* $\xi_n \xrightarrow{\text{p}} \xi$ *и* $\eta_n \xrightarrow{\text{p}} \eta$*, то*
1. $\xi_n + \eta_n \xrightarrow{\text{p}} \xi + \eta$;
2. $\xi_n \cdot \eta_n \xrightarrow{\text{p}} \xi \cdot \eta$.

Доказательство при первом прочтении можно пропустить.

1. В доказательстве мы будем пользоваться естественным свойством вероятности: если из события A следует событие B (*всегда, когда выполнено A, выполнено и B*), то вероятность A не превосходит вероятности B:

$$\text{если} \quad A \subseteq B, \quad \text{то} \quad \mathsf{P}(A) \leqslant \mathsf{P}(B).$$

Здесь я категорически требую остановиться и ответить на следующие «глупые вопросы»:

— *Верно ли, что модуль суммы не превосходит суммы модулей?*
— *Верно ли, что если $a > b$, и $c > a$, то $c > b$?*
— *Верно ли, что если $a + b > 2$, то хоть одно из чисел a, b больше единицы?*
— *Верно ли, что вероятность объединения двух событий не превосходит суммы их вероятностей?*
— *Верно ли, что вероятность пересечения двух событий не превосходит вероятности любого из них?*

Если на все вопросы вы ответили «да», можно двигаться дальше. Если не на все — ваш контрпример ошибочен. Если вы вообще не поняли, о чем это, лучше вернуться сюда ...

Пусть $\varepsilon > 0$. Требуется доказать, что $\mathsf{P}(|\xi_n + \eta_n - \xi - \eta| > \varepsilon) \to 0$ при $n \to \infty$. Но

а) $|\xi_n + \eta_n - \xi - \eta| \leqslant |\xi_n - \xi| + |\eta_n - \eta|$, поэтому

б) если $|\xi_n + \eta_n - \xi - \eta| > \varepsilon$, то и $|\xi_n - \xi| + |\eta_n - \eta| > \varepsilon$, и вероятность первого события не больше вероятности второго. Далее,

в) если $|\xi_n - \xi| + |\eta_n - \eta| > \varepsilon$, то хотя бы одно из слагаемых больше, чем $\varepsilon/2$.

Получаем следующую цепочку неравенств:

$$\begin{aligned} \mathsf{P}(|\xi_n + \eta_n - \xi - \eta| > \varepsilon) &\leqslant \mathsf{P}(|\xi_n - \xi| + |\eta_n - \eta| > \varepsilon) \leqslant \mathsf{P}\big(|\xi_n - \xi| > \varepsilon/2 \text{ или } |\eta_n - \eta| > \varepsilon/2\big) \leqslant \\ &\leqslant \mathsf{P}(|\xi_n - \xi| > \varepsilon/2) + \mathsf{P}(|\eta_n - \eta| > \varepsilon/2) \to 0 \end{aligned}$$

при $n \to \infty$, так как $\xi_n \xrightarrow{\text{p}} \xi$ и $\eta_n \xrightarrow{\text{p}} \eta$.

2. Нам понадобится «хорошее свойство»: для любой случайной величины ζ, просто по свойствам функций распределения, $\mathsf{P}(|\zeta| > M) \to 0$ при $M \to \infty$.

Представим $|\xi_n\eta_n - \xi\eta|$ как $|(\eta_n - \eta)(\xi_n - \xi) + \xi(\eta_n - \eta) + \eta(\xi_n - \xi)|$. Затем, как в **1**, получим

$$\mathsf{P}(|\xi_n\eta_n - \xi\eta| > \varepsilon) \leqslant \mathsf{P}(|\eta_n - \eta| \cdot |\xi_n - \xi| > \varepsilon/3) + \mathsf{P}(|\xi| \cdot |\eta_n - \eta| > \varepsilon/3) + \mathsf{P}(|\eta| \cdot |\xi_n - \xi| > \varepsilon/3).$$

Подумайте, что делать с первым слагаемым в правой части, а мы пока рассмотрим второе слагаемое (третье такое же). Обозначим за $A_n = \{|\xi| \cdot |\eta_n - \eta| > \varepsilon/3\}$ событие под знаком вероятности. Зафиксируем некоторое $M > 0$ и разобьем событие A_n по полной группе событий $\{|\xi| > M\}$ и $\{|\xi| \leqslant M\}$.

$$\mathsf{P}(A_n) = \mathsf{P}(|\xi| \cdot |\eta_n - \eta| > \varepsilon/3) = \mathsf{P}\big(A_n \cap \{|\xi| > M\}\big) + \mathsf{P}\big(A_n \cap \{|\xi| \leqslant M\}\big) \leqslant \ \ldots$$

Первую вероятность оцениваем в соответствии с последним «глупым вопросом», вторую — пользуясь тем, что из $|\xi| \cdot |\eta_n - \eta| > \varepsilon/3$ и $|\xi| \leqslant M$ следует, что $M \cdot |\eta_n - \eta| > \varepsilon/3$.

$$\ldots \ \leqslant \mathsf{P}(|\xi| > M) + \mathsf{P}\big(M \cdot |\eta_n - \eta| > \varepsilon/3\big) = \mathsf{P}(|\xi| > M) + \mathsf{P}\big(|\eta_n - \eta| > \varepsilon/3M\big).$$

Осталось для любого фиксированного $M > 0$ устремить n к бесконечности, получив для верхнего предела оценку $\overline{\lim\limits_{n\to\infty}} \mathsf{P}(A_n) \leqslant \mathsf{P}(|\xi| > M)$, после чего мы можем устремить к бесконечности M, пользуясь «хорошим свойством». $\square$

Упражнение 25. Восполнить все пропущенные подробности в доказательстве.

Сходимость по вероятности, так же как и любая другая сходимость, не портится под действием непрерывной функции.

Свойство 17.
Если $\xi_n \xrightarrow{\text{p}} \xi$ и g — непрерывная функция, то $g(\xi_n) \xrightarrow{\text{p}} g(\xi)$.
Если $\xi_n \xrightarrow{\text{p}} c$ и g непрерывна в точке c, то $g(\xi_n) \xrightarrow{\text{p}} g(c)$.

Доказательство. Простое доказательство первого утверждения можно предложить в двух случаях (которыми мы и ограничимся, предоставив все остальное читателю, знакомому, например, с теоремой Егорова): *если $\xi = c = const$* (и тогда достаточно, чтобы g была непрерывна в точке c) или *если функция g равномерно непрерывна* (*а что это значит?*).

И в том, и в другом случае для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для любого ω, удовлетворяющего условию $|\xi_n(\omega) - \xi(\omega)| < \delta$, выполняется неравенство $|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon$.

То есть событие $\big\{|\xi_n - \xi| < \delta\big\}$ влечет событие $\big\{|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon\big\}$. Следовательно, вероятность первого не больше, чем вероятность второго. Но, какое бы ни было $\delta > 0$, вероятность первого события стремится к единице по определению сходимости по вероятности:

$$1 \longleftarrow \mathsf{P}\big(|\xi_n - \xi| < \delta\big) \leqslant \mathsf{P}\big(|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon\big) \leqslant 1.$$

Следовательно, и вероятность второго события также стремится к единице.

Предлагаю поразмышлять на тему: в каком месте доказательства используется, что либо g равномерно непрерывна, либо ξ — постоянная. И над тем, как доказывать первую часть свойства **17** в общем случае. □

Чтобы доказывать сходимость по вероятности, можно просто уметь вычислять $\mathsf{P}\left(|\xi_n - \xi| > \varepsilon\right)$ при больших n. Но для этого нужно знать распределение ξ_n, что не всегда возможно. Скажем, ξ_n может быть суммой (или еще хуже :-) нескольких других с. в., распределения которых не устойчивы по суммированию, и вычислить распределение их суммы по формуле свертки или как-то еще бывает слишком сложно.

Если бы мы имели неравенства, позволяющие оценить $\mathsf{P}\left(|\xi_n - \xi| > \varepsilon\right)$ сверху чем-либо, что мы умеем устремлять к нулю и что проще вычисляется, то сходимость по вероятности мы получили бы по лемме о двух милиционерах: $0 \leqslant \mathsf{P}(...) \leqslant ... \to 0$. Итак, неравенства П. Л. Чебышёва.

13.2 Неравенства Чебышёва

Все неравенства в этом параграфе принято относить к одному классу, называемому «неравенствами Чебышёва». Следующее неравенство часто называют собственно неравенством Чебышёва, хотя в такой форме оно появилось впервые, видимо, в работах А. А. Маркова (например, *Исчисление вероятностей*, 1913 г.).

Теорема 28 (Неравенство Маркова).

Если $\mathsf{E}\,|\xi| < \infty$, то для любого положительного x

$$\mathsf{P}\big(|\xi| > x\big) \leqslant \frac{\mathsf{E}\,|\xi|}{x}.$$

Доказательство. Введем новую случайную величину ξ_x, называемую «срезкой» с. в. $|\xi|$ на уровне x:

$$\xi_x = \begin{cases} |\xi|, & \text{если } |\xi| \leqslant x, \\ x, & \text{если } |\xi| > x. \end{cases} \qquad \text{Для неё} \qquad \begin{array}{ll} 1) & \xi_x \leqslant |\xi|, \text{ и, следовательно,} \\ 2) & \mathsf{E}\,\xi_x \leqslant \mathsf{E}\,|\xi|. \end{array}$$

Нам потребуется следующее понятие.

Определение 50. Пусть A — некоторое событие. Назовем ***индикатором события*** A случайную величину $I(A)$, равную единице, если событие A произошло, и нулю, если A не произошло.

По определению, $I(A)$ имеет распределение Бернулли $\mathbf{B}_p$ с параметром $p = \mathsf{P}(I(A)=1) = \mathsf{P}(A)$, и ее математическое ожидание равно вероятности успеха $p = \mathsf{P}(A)$.

Случайную величину ξ_x можно представить в виде $\xi_x = |\xi| \cdot I(|\xi| \leqslant x) + x \cdot I(|\xi| > x)$ ***(проверьте!)***.

Тогда

$$\mathsf{E}\,\xi_x = \underbrace{\mathsf{E}\Big(|\xi| \cdot I\big(|\xi| \leqslant x\big)\Big)}_{\text{неотрицательно, отбросим}} + \mathsf{E}\Big(x \cdot I\big(|\xi| > x\big)\Big) \geqslant \mathsf{E}\Big(x \cdot I\big(|\xi| > x\big)\Big) = x \cdot \mathsf{P}\big(|\xi| > x\big). \tag{21}$$

Вспомним, что $\mathsf{E}\,|\xi| \geqslant \mathsf{E}\,\xi_x$, и оценим $\mathsf{E}\,\xi_x$ снизу согласно (21):

$$\mathsf{E}\,|\xi| \geqslant \mathsf{E}\,\xi_x \;\geqslant\; x \cdot \mathsf{P}\big(|\xi| > x\big).$$

Итак, $x \cdot \mathsf{P}\big(|\xi| > x\big) \leqslant \mathsf{E}\,|\xi|$, что и требовалось доказать. □

Следующее неравенство мы будем называть «обобщенным неравенством Чебышёва».

Следствие 15. *Пусть функция g монотонно возрастает и неотрицательна на $[0, \infty)$. Если $\mathsf{E}\,g(|\xi|) < \infty$, то для любого положительного x*

$$\mathsf{P}\big(|\xi| > x\big) \leqslant \frac{\mathsf{E}\,g(|\xi|)}{g(x)}.$$

Доказательство*.* Заметим, что $\mathsf{P}\big(|\xi| > x\big) = \mathsf{P}\big(g(|\xi|) > g(x)\big)$, поскольку функция g монотонно возрастает, и оценим последнюю вероятность согласно неравенству Маркова:

$$\mathsf{P}\big(g(|\xi|) > g(x)\big) \leqslant \frac{\mathsf{E}\,g(|\xi|)}{g(x)}. \qquad \square$$

В 1853 г. И. Бьенеме (I. Bienaymé) и в 1866 г., независимо от него, П. Л. Чебышёв прямыми методами доказали неравенство, которое нам будет удобно получить в качестве следствия из неравенства Маркова.

Следствие 16 (Неравенство Чебышёва-Бьенеме). *Если $\mathsf{E}\,\xi^2 < \infty$, то*

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| > x\big) \leqslant \frac{\mathsf{D}\,\xi}{x^2}.$$

Доказательство. Воспользуемся следствием **15** с функцией $g(x) = x^2$.

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| > x\big) \leqslant \frac{\mathsf{E}\left(\xi - \mathsf{E}\,\xi\right)^2}{x^2} = \frac{\mathsf{D}\,\xi}{x^2}. \qquad \square$$

В качестве следствия получим так называемое «правило трех сигм», которое формулируют, например, так: *Вероятность случайной величине отличаться от своего математического ожидания более, чем на три корня из дисперсии,* **мала**. Разумеется, для каждого распределения величина этой вероятности своя: для нормального распределения, например, эта вероятность равна 0,0027 — см. свойство **11**. Мы получим верную для всех распределений с конечной дисперсией оценку сверху для «вероятности с. в. отличаться от своего математического ожидания более, чем на три корня из дисперсии».

Следствие 17. *Если* $\mathsf{E}\,\xi^2 < \infty$, *то* $\quad \mathsf{P}\big(|\xi - \mathsf{E}\,\xi| > 3\sqrt{\mathsf{D}\,\xi}\big) \leqslant \dfrac{1}{9}.$

Доказательство. Согласно следствию **16**, $\quad \mathsf{P}\big(|\xi - \mathsf{E}\,\xi| > 3\sqrt{\mathsf{D}\,\xi}\big) \leqslant \dfrac{\mathsf{D}\,\xi}{\big(3\sqrt{\mathsf{D}\,\xi}\big)^2} = \dfrac{1}{9}.$ $\square$

Упражнение 26. Найти $\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| > 3\sqrt{\mathsf{D}\,\xi}\big)$, если с. в. ξ имеет
а) равномерное распределение на каком-нибудь отрезке;
б) показательное распределение с каким-нибудь параметром;
в) распределение Бернулли с параметром 1/2.

13.3 Законы больших чисел

Определение 51. Говорят, что последовательность с. в. $\{\xi_i\}_{i=1}^{\infty}$ с конечными первыми моментами *удовлетворяет закону больших чисел* (ЗБЧ), если

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{\mathsf{E}\,\xi_1 + \dots + \mathsf{E}\,\xi_n}{n} \xrightarrow{\text{p}} 0 \quad \text{при} \quad n \to \infty. \tag{22}$$

Законами больших чисел принято называть утверждения об условиях, при которых последовательность с. в. «удовлетворяет закону больших чисел».

Выясним сначала, что означает и когда выполнен ЗБЧ для последовательности *независимых и одинаково распределенных с. в.*

Заметим, что если с. в. одинаково распределены, то математические ожидания у них одинаковы (и равны, например, $\mathsf{E}\,\xi_1$), поэтому свойство (22) можно записать в виде $\dfrac{\xi_1+\dots+\xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1$.

Итак, законы больших чисел.

Теорема 29 (ЗБЧ в форме Чебышёва).

Для любой последовательности независимых и одинаково распределенных случайных величин с конечным вторым моментом $\mathsf{E}\,\xi_1^2 < \infty$ имеет место сходимость:

$$\frac{\xi_1+\dots+\xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1.$$

ЗБЧ утверждает, что среднее арифметическое большого числа случайных слагаемых «стабилизируется» с ростом этого числа. Как бы сильно каждая с. в. не отклонялась от своего среднего значения, при суммировании эти отклонения «взаимно гасятся», так что среднее арифметическое приближается к постоянной величине.

В дальнейшем мы увидим, что требование конечности второго момента (или дисперсии) связано исключительно со способом доказательства, и что утверждение остается верным если требовать существования только первого момента.

Доказательство. Обозначим через $S_n = \xi_1+\dots+\xi_n$ сумму первых n с. в., а через $\dfrac{S_n}{n} = \dfrac{\xi_1+\dots+\xi_n}{n}$ — их среднее арифметическое. Тогда

$$\mathsf{E}\left(\frac{S_n}{n}\right) = \frac{\mathsf{E}\,\xi_1+\dots+\mathsf{E}\,\xi_n}{n} = \frac{n\cdot\mathsf{E}\,\xi_1}{n} = \mathsf{E}\,\xi_1.$$

Пусть $\varepsilon>0$. Воспользуемся неравенством Чебышёва (следствие **16**):

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)\right| > \varepsilon\right) \leqslant \frac{\mathsf{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2} =$$

$$= \frac{\mathsf{D}\,S_n}{n^2\varepsilon^2} \overset{\textit{независ.}}{=} \frac{\mathsf{D}\,\xi_1+\dots+\mathsf{D}\,\xi_n}{n^2\varepsilon^2} \overset{\textit{од.распред.}}{=} \frac{n\,\mathsf{D}\,\xi_1}{n^2\varepsilon^2} = \frac{\mathsf{D}\,\xi_1}{n\varepsilon^2} \to 0 \tag{23}$$

при $n\to\infty$, поскольку $\mathsf{D}\,\xi_1$, по условию, конечна. □

Замечание 24. Мы не только доказали сходимость, но и получили оценку для вероятности среднему арифметическому любого числа независимых и одинаково распределенных величин отличаться от $\mathsf{E}\,\xi_1$ более чем на заданное число:

$$\mathsf{P}\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-\mathsf{E}\,\xi_1\right|>\varepsilon\right)\leqslant\frac{\mathsf{D}\,\xi_1}{n\varepsilon^2}. \tag{24}$$

Предлагаю, кроме того, читателям извлечь из неравенства (23) в доказательстве ЗБЧ Чебышёва доказательство следующего утверждения.

Следствие 18. *Последовательность с. в. $\{\xi_i\}_{i=1}^{\infty}$ с конечными вторыми моментами удовлетворяет ЗБЧ, то есть*

$$\frac{S_n}{n}-\mathsf{E}\left(\frac{S_n}{n}\right)=\frac{\xi_1+\cdots+\xi_n}{n}-\frac{\mathsf{E}\,\xi_1+\cdots+\mathsf{E}\,\xi_n}{n}\xrightarrow{\text{p}}0\quad\textit{при}\quad n\to\infty$$

при выполнении любого из следующих условий:

а) если $\mathsf{D}\,S_n=o(n^2)$, то есть $\dfrac{\mathsf{D}\,S_n}{n^2}\to 0$ при $n\to\infty$;

б) если $\xi_1,\xi_2,\ldots$ независимы и $\mathsf{D}\,S_n=\mathsf{D}\,\xi_1+\cdots+\mathsf{D}\,\xi_n=o(n^2)$, то есть $\dfrac{\mathsf{D}\,\xi_1+\cdots+\mathsf{D}\,\xi_n}{n^2}\to 0$ при $n\to\infty$;

в) если $\xi_1,\xi_2,\ldots$ независимы, одинаково распределены и имеют конечную дисперсию (ЗБЧ Чебышёва).

Скоро мы докажем (иными методами, чем А. Я. Хинчин) следующее утверждение.

Теорема 30 (ЗБЧ в форме Хинчина).

Для любой последовательности независимых и одинаково распределенных случайных величин с конечным ***первым*** *моментом $\mathsf{E}\,|\xi_1|<\infty$ имеет место сходимость:*

$$\frac{\xi_1+\cdots+\xi_n}{n}\xrightarrow{\text{p}}\mathsf{E}\,\xi_1.$$

Более того, в условиях теоремы **30** имеет место «почти наверное» сходимость $(\xi_1+\cdots+\xi_n)/n$ к $\mathsf{E}\,\xi_1$. Но этого мы уже доказывать не будем.

Получим в качестве следствия из ЗБЧ Чебышёва закон больших чисел Я. Бернулли (1713). В отличие от доказанного через полтора столетия ЗБЧ Чебышёва, описывающего предельное поведение среднего арифметического с. в. с *произвольными* распределениями, ЗБЧ Бернулли — утверждение только для *схемы Бернулли*.

Теорема 31 (ЗБЧ Бернулли).

Пусть A — событие, которое может произойти в любом из n независимых испытаний с одной и той же вероятностью $\mathsf{P}(A)$. Пусть $\nu_n(A)$ — число осуществлений события A в n испытаниях. Тогда $\dfrac{\nu_n(A)}{n} \xrightarrow{\text{p}} \mathsf{P}(A)$. При этом для любого $\varepsilon > 0$

$$\mathsf{P}\left(\left| \frac{\nu_n(A)}{n} - \mathsf{P}(A) \right| > \varepsilon\right) \leqslant \frac{\mathsf{P}(A)(1-\mathsf{P}(A))}{n\varepsilon^2}.$$

Доказательство. Заметим, что $\nu_n(A)$ есть сумма независимых, одинаково распределенных с. в., имеющих распределение Бернулли с параметром, равным вероятности успеха $\mathsf{P}(A)$ (индикаторов того, что в соответствующем испытании произошло A):

$$\nu_n(A) = \xi_1 + \cdots + \xi_n, \quad \xi_i = I_i(A) = \begin{cases} 1, & \text{если } A \text{ произошло в } i-\text{м испытании;} \\ 0, & \text{если } A \text{ не произошло в } i-\text{м испытании;} \end{cases} \qquad \mathsf{E}\,\xi_1 = \mathsf{P}(A),\ \mathsf{D}\,\xi_1 = \mathsf{P}(A)(1-\mathsf{P}(A)).$$

Осталось воспользоваться ЗБЧ в форме Чебышёва и неравенством (24) из замечания **24**. □

Рассмотрим примеры использования ЗБЧ в форме Чебышёва, вернее, неравенства (24).

13.4 Примеры использования ЗБЧ и неравенства Чебышёва

Пример 48.

З а д а ч а. Монета подбрасывается 10 000 раз. Оценить вероятность того, что частота выпадения герба отличается от вероятности более чем на одну сотую.

Р е ш е н и е. Требуется оценить $\mathsf{P}\left(\left|\dfrac{\nu_n}{n} - \dfrac{1}{2}\right| > 0{,}01\right)$, где $n = 10^4$, $\nu_n = \sum_{i=1}^n \xi_i$ — число выпадений герба, а ξ_i — независимые с. в., имеющие распределение Бернулли с параметром 1/2, равные «числу гербов, выпавших при i-м подбрасывании» (то есть единице, если выпал герб и нулю иначе, или индикатору того, что выпал герб).

Поскольку $\mathsf{D}\,\xi_1 = 1/2 \cdot 1/2 = 1/4$, искомая оценка сверху выглядит так:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > 0{,}01\right) \leqslant \frac{\mathsf{D}\,\xi_1}{n\ 0{,}01^2} = \frac{1}{4 \cdot 10^4 \cdot 10^{-4}} = \frac{1}{4}.$$

Иначе говоря, неравенство Чебышёва позволяет заключить, что, в среднем, не более чем в четверти случаев при 10 000 подбрасываниях монеты частота выпадения герба будет отличаться от 1/2 более чем на одну сотую. Мы увидим, насколько это грубая оценка, когда познакомимся с *центральной предельной теоремой*.

Пример 49.

З а д а ч а. Пусть $\xi_1, \xi_2, \ldots$ — последовательность случайных величин, дисперсии которых ограничены одной и той же постоянной C, а ковариации любых с. в. ξ_i и ξ_j $(i \neq j)$, не являющихся соседними в последовательности, равны нулю. Удовлетворяет ли эта последовательность ЗБЧ?

Р е ш е н и е. Воспользуемся неравенством (23) и свойством **14**:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)\right| > \varepsilon\right) \leqslant \frac{\mathsf{D}\left(\dfrac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathsf{D}\,S_n}{n^2\varepsilon^2}; \quad \mathsf{D}\,(\xi_1 + \ldots + \xi_n) = \sum_{i=1}^{n} \mathsf{D}\,\xi_i + 2\sum_{i<j} \operatorname{cov}(\xi_i, \xi_j).$$

Но для $i < j$, по условию, $\operatorname{cov}(\xi_i, \xi_j) = 0$, если $i \neq j-1$. Следовательно, в сумме $\sum_{i<j} \operatorname{cov}(\xi_i, \xi_j)$ равны нулю все слагаемые кроме, может быть, $\operatorname{cov}(\xi_1, \xi_2), \operatorname{cov}(\xi_2, \xi_3), \ldots, \operatorname{cov}(\xi_{n-1}, \xi_n)$ (их ровно $n-1$ штука).

Оценим каждое из них, используя одно из свойств коэффициента корреляции ***(какое?)***

$$\operatorname{cov}(\xi_i, \xi_j) \leqslant \sqrt{\mathsf{D}\,\xi_i}\sqrt{\mathsf{D}\,\xi_j} \leqslant \sqrt{C}\sqrt{C} = C,$$

так как для любого $1 \leqslant i \leqslant n$, по условию, $\mathsf{D}\,\xi_i \leqslant C$. Итак,

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)\right| > \varepsilon\right) \leqslant \frac{\mathsf{D}\,S_n}{n^2\varepsilon^2} = \frac{\sum\limits_{i=1}^{n} \mathsf{D}\,\xi_i + 2\sum\limits_{i<j} \operatorname{cov}(\xi_i, \xi_j)}{n^2\varepsilon^2} = \frac{\sum\limits_{i=1}^{n} \mathsf{D}\,\xi_i + 2\sum\limits_{i=1}^{n-1} \operatorname{cov}(\xi_i, \xi_{i+1})}{n^2\varepsilon^2} \leqslant \frac{nC + 2(n-1)C}{n^2\varepsilon^2} \to 0$$

при $n \to \infty$, то есть последовательность $\xi_1, \xi_2, \ldots$ удовлетворяет ЗБЧ.

Упражнение 27.

Привести пример последовательности с. в. $\xi_1, \xi_2, \ldots$ такой, что ковариации «несоседних» величин равны нулю. Привести пример последовательности с. в. $\xi_1, \xi_2, \ldots$ такой, что ковариации «несоседних» величин равны нулю, а ковариации соседних — не равны. Можно попробовать построить такую последовательность с помощью другой последовательности, составленной из независимых с. в.

... Из этой первой лекции по теории вероятностей я запомнил только полузнакомый термин «математическое ожидание». Незнакомец употреблял этот термин неоднократно, и каждый раз я представлял себе большое помещение, вроде зала ожидания, с кафельным полом, где сидят люди с портфелями и бюварами и, подбрасывая время от времени к потолку монетки и бутерброды, сосредоточенно чего-то ожидают. До сих пор я часто вижу это во сне. Но тут незнакомец оглушил меня звонким термином «предельная теорема Муавра — Лапласа» и сказал, что все это к делу не относится.

Аркадий и Борис Стругацкие, *Стажеры*

Раздел 14. ЦПТ (центральная предельная теорема)

14.1 Как быстро $\dfrac{S_n}{n}$ сходится к $\mathsf{E}\,\xi_1$?

Пусть, как в законе больших чисел в форме Чебышёва, $S_n = \xi_1 + \ldots + \xi_n$ — сумма n независимых и одинаково распределенных величин с конечной дисперсией. Тогда, в силу ЗБЧ, $\dfrac{S_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1$ с ростом n. Или, после приведения к общему знаменателю,

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{n} \xrightarrow{\text{p}} 0.$$

Если при делении на n мы получили в пределе нуль (в смысле некоторой, все равно какой, сходимости), резонно задать себе вопрос: а не слишком ли на «много» мы поделили? Нельзя ли поделить на что-нибудь, растущее к бесконечности медленнее, чем n, чтобы получить в пределе не нуль (и не бесконечность, само собой)?

Можно поставить этот вопрос по-другому. Вот последовательность, стремящаяся (как-то) к нулю. Можно ли ее домножить на что-либо растущее, чтобы «погасить» это стремление к нулю? Получив, тем самым, что-нибудь конечное и отличное от нуля в пределе?

Оказывается, что уже $\dfrac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}}$, или, что то же самое, $\sqrt{n} \cdot \dfrac{S_n - n\,\mathsf{E}\,\xi_1}{n}$, не сходится к нулю. Распределение этой, зависящей от n, случайной величины становится все более похоже на нормальное распределение! Можно считать, что такая последовательность сходится к случайной величине, имеющей нормальное распределение, но сходится не по вероятности, а только в смысле сходимости распределений, или «слабой сходимости».

14.2 Слабая сходимость

Пусть задана последовательность с. в. $\{\xi_n\}$, задано некоторое распределение $\mathcal{F}$ с функцией распределения F_ξ и ξ — произвольная с. в., имеющая распределение $\mathcal{F}$.

Определение 52. Говорят, что последовательность с. в. $\{\xi_n\}$ при $n \to \infty$ ***сходится слабо*** или ***по распределению*** к с. в. ξ, или говорят, что последовательность с. в. ***слабо сходится к распределению*** $\mathcal{F}$, или говорят, что ***распределения с. в.*** $\{\xi_n\}$ ***слабо сходятся к распределению*** $\mathcal{F}$, и пишут: $\xi_n \Rightarrow \xi$, или $F_{\xi_n} \Rightarrow F_\xi$, или $\xi_n \Rightarrow \mathcal{F}$, если для любого x такого, что функция распределения F_ξ непрерывна в точке x, имеет место сходимость $F_{\xi_n}(x) \to F_\xi(x)$ при $n \to \infty$.

Иначе говоря, слабая сходимость — это поточечная сходимость функций распределения во всех точках непрерывности предельной функции распределения.

Необходимо заметить, что запись «$\xi_n \Rightarrow \xi$» удобна, но не всегда разумна: если «предельную» с. в. ξ заменить на другую с. в. η с тем же распределением, ничего не изменится: в том же смысле $\xi_n \Rightarrow \eta$. Поэтому слабая сходимость все же не есть сходимость случайных величин, и ей нельзя оперировать как сходимостями п.н. и по вероятности, для которых предельная с.в. единственна (хотя бы с точностью до значений на множестве нулевой вероятности).

Следующее свойство очевидно. Если нет - вам нужно вернуться к разделу **7** и вспомнить, что такое функция распределения.

Свойство 18. *Если $\xi_n \Rightarrow \xi$, и функция распределения F_ξ непрерывна в точках a и b, то $\mathsf{P}(\xi_n \in [a,b]) \to \mathsf{P}(\xi \in [a,b])$ и т.д. (продолжить ряд). Наоборот, если во всех точках a и b непрерывности функции распределения F_ξ имеет место, например, сходимость $\mathsf{P}(\xi_n \in [a,b]) \to \mathsf{P}(\xi \in [a,b])$, то $\xi_n \Rightarrow \xi$.*

Следующее важное свойство уточняет отношения между сходимостями.

Свойство 19.
1. *Если $\xi_n \xrightarrow{\mathrm{p}} \xi$, то $\xi_n \Rightarrow \xi$.* **2.** *Если $\xi_n \Rightarrow c = const$, то $\xi_n \xrightarrow{\mathrm{p}} c$.*

Доказательство. Первое свойство мы доказывать не будем.
Докажем, что слабая сходимость к постояннной влечет сходимость по вероятности. Пусть

$$F_{\xi_n}(x) \to F_c(x) = \begin{cases} 0, & x \leqslant c; \\ 1, & x > c \end{cases}$$

при любом x, являющемся точкой непрерывности предельной функции $F_c(x)$, то есть при всех $x \neq c$.

Возьмем произвольное $\varepsilon > 0$ и докажем, что $\mathsf{P}(|\xi_n - c| \leqslant \varepsilon) \to 1$. Раскроем модуль:

$$\mathsf{P}(-\varepsilon \leqslant \xi_n - c \leqslant \varepsilon) = \mathsf{P}(c - \varepsilon \leqslant \xi_n \leqslant c + \varepsilon) \geqslant$$

(сужаем событие под знаком вероятности)

$$\geqslant \mathsf{P}(c - \varepsilon \leqslant \xi_n < c + \varepsilon) = F_{\xi_n}(c + \varepsilon) - F_{\xi_n}(c - \varepsilon) \to F_c(c + \varepsilon) - F_c(c - \varepsilon) = 1 - 0 = 1,$$

поскольку в точках $c \pm \varepsilon$ функция F_c непрерывна, и, следовательно, имеет место сходимость последовательности $F_{\xi_n}(c \pm \varepsilon)$ к $F_c(c \pm \varepsilon)$.

Осталось заметить, что $\mathsf{P}(|\xi_n - c| \leqslant \varepsilon)$ не бывает больше 1, так что по лемме о двух милиционерах $\mathsf{P}(|\xi_n - c| \leqslant \varepsilon) \to 1$.

$\square$

Следующее свойство приводит пример операций, которые можно применять к слабо сходящимся последовательностям — скажем, домножать их на последовательности, сходящиеся по вероятности к постоянным величинам.

Желание написать «если $\xi_n \Rightarrow \xi$ и $\eta_n \Rightarrow \eta$, то $\xi_n + \eta_n \Rightarrow \xi + \eta$» сразу проходит, стоит перевести это «свойство» на язык функций распределения и задуматься — что такое «функция распределения суммы $\xi + \eta$», когда вместо них можно брать любые другие $\tilde{\xi}$ и $\tilde{\eta}$ с теми же распределениями, как угодно зависимые. Иное дело — когда одно из предельных распределений вырождено. В этом случае функция распределения суммы или произведения определена однозначно.

Свойство 20.

1. *Если $\xi_n \xrightarrow{\mathrm{p}} c = const$ и $\eta_n \Rightarrow \eta$, то $\xi_n \cdot \eta_n \Rightarrow c\eta$.* **2.** *Если $\xi_n \xrightarrow{\mathrm{p}} c = const$ и $\eta_n \Rightarrow \eta$, то $\xi_n + \eta_n \Rightarrow c + \eta$.*

Доказательство. Заметим прежде всего, что если $\eta_n \Rightarrow \eta$, то $c\eta_n \Rightarrow c\eta$, $c + \eta_n \Rightarrow c + \eta$ (*доказать!*). Поэтому (и в силу соответствующих свойств сходимости по вероятности) достаточно доказать первое утверждение свойства **20** при $c = 1$, а второе утверждение — при $c = 0$.

Докажем второе утверждение, оставив первое читателю. Пусть $\xi_n \xrightarrow{\mathrm{p}} 0$ и $\eta_n \Rightarrow \eta$. Докажем, что $\xi_n + \eta_n \Rightarrow \eta$. Пусть x_0 — точка непрерывности функции распределения $F_\eta(x)$. Требуется доказать, что тогда имеет место сходимость $F_{\xi_n+\eta_n}(x_0) \to F_\eta(x_0)$. Зафиксируем достаточно маленькое $\varepsilon > 0$ такое, что $F_\eta(x)$ непрерывна в точках $x_0 \pm \varepsilon$.

$$F_{\xi_n+\eta_n}(x_0) = \mathsf{P}(\xi_n + \eta_n < x_0) = \mathsf{P}(\xi_n + \eta_n < x_0, |\xi_n| > \varepsilon) + \mathsf{P}(\xi_n + \eta_n < x_0, |\xi_n| \leqslant \varepsilon) = P_1 + P_2.$$

Оценим $P_1 + P_2$ сверху и снизу. Для P_1 имеем: $0 \leqslant P_1 = \mathsf{P}(\xi_n + \eta_n < x_0, |\xi_n| > \varepsilon) \leqslant \mathsf{P}(|\xi_n| > \varepsilon)$, и последняя вероятность может быть выбором n сделана сколь угодно малой.

Для P_2, с одной стороны,

$$P_2 = \mathsf{P}(\xi_n + \eta_n < x_0,\ -\varepsilon \leqslant \xi_n \leqslant \varepsilon) \leqslant \mathsf{P}(-\varepsilon + \eta_n < x_0) = \mathsf{P}(\eta_n < x_0 + \varepsilon).$$

Здесь первое неравенство следует из очевидного соображения:

если $-\varepsilon \leqslant \xi_n$ и $\xi_n + \eta_n < x_0$, то, тем более, $-\varepsilon + \eta_n < x_0$.

С другой стороны,

$$P_2 = \mathsf{P}(\xi_n + \eta_n < x_0,\ -\varepsilon \leqslant \xi_n \leqslant \varepsilon) \geqslant \mathsf{P}(\varepsilon + \eta_n < x_0, -\varepsilon \leqslant \xi_n \leqslant \varepsilon) \geqslant \mathsf{P}(\varepsilon + \eta_n < x_0) - \mathsf{P}(|\xi_n| > \varepsilon) = \mathsf{P}(\eta_n < x_0 - \varepsilon) - \mathsf{P}(|\xi_n| > \varepsilon).$$

Здесь первое неравенство объясняется включением $\{\varepsilon + \eta_n < x_0\} \cap \{-\varepsilon \leqslant \xi_n \leqslant \varepsilon\} \subset \{\xi_n + \eta_n < x_0\} \cap \{-\varepsilon \leqslant \xi_n \leqslant \varepsilon\}$ —

просто заменим в событии $\{\varepsilon + \eta_n < x_0\}$ число ε на ξ_n, так как $\xi_n \leqslant \varepsilon$. Второе неравенство следует из свойств:

$$\mathsf{P}(A\overline{B}) \leqslant \mathsf{P}(\overline{B}), \quad \text{поэтому} \quad \mathsf{P}(AB) = \mathsf{P}(A) - \mathsf{P}(A\overline{B}) \geqslant \mathsf{P}(A) - \mathsf{P}(\overline{B}).$$

Итак, мы получили оценки снизу и сверху для $P_1 + P_2$, то есть для $F_{\xi_n+\eta_n}(x_0)$:

$$\mathsf{P}(\eta_n < x_0 - \varepsilon) - \mathsf{P}(|\xi_n| > \varepsilon) \leqslant F_{\xi_n+\eta_n}(x_0) \leqslant \mathsf{P}(|\xi_n| > \varepsilon) + \mathsf{P}(\eta_n < x_0 + \varepsilon),$$

или

$$F_{\eta_n}(x_0 - \varepsilon) - \mathsf{P}(|\xi_n| > \varepsilon) \leqslant F_{\xi_n+\eta_n}(x_0) \leqslant \mathsf{P}(|\xi_n| > \varepsilon) + F_{\eta_n}(x_0 + \varepsilon).$$

Устремляя n к бесконечности, и вспоминая, что $x_0 \pm \varepsilon$ — точки непрерывности функции распределения F_η, получим

$$F_\eta(x_0 - \varepsilon) \leqslant \underline{\lim} F_{\xi_n+\eta_n}(x_0) \leqslant \overline{\lim} F_{\xi_n+\eta_n}(x_0) \leqslant F_\eta(x_0 + \varepsilon).$$

И поскольку эти неравенства верны для любого достаточно малого ε, а x_0 — точка непрерывности функции F_η, то, устремив ε к нулю, получим, что нижний и верхний пределы $F_{\xi_n+\eta_n}(x_0)$ при $n \to \infty$ совпадают и равны $F_\eta(x_0)$.

□

Несколько содержательных примеров слабой сходимости мы рассмотрим в следующей главе. Но основной источник слабо сходящихся последовательностей и необычайно мощное и универсальное средство для асимптотического анализа ***распределений сумм*** независимых и одинаково распределенных случайных величин предоставляет нам

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

14.3 Центральная предельная теорема

Мы будем называть следующее утверждение «ЦПТ А. М. Ляпунова» (1901), но сформулируем и докажем теорему Ляпунова только *в частном случае* — для последовательности независимых и одинаково распределенных случайных величин.

Теорема 32 (ЦПТ).

Пусть $\xi_1, \xi_2, \ldots$ — независимые и одинаково распределенные случайные величины с конечной и ненулевой дисперсией: $0 < \mathsf{D}\,\xi_1 < \infty$. Обозначим через S_n сумму первых n случайных величин: $S_n = \xi_1 + \ldots + \xi_n$. Тогда последовательность с. в. $\dfrac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}}$ слабо сходится к стандартному нормальному распределению.

Пользуясь определением и свойствами слабой сходимости, и заметив, что функция распределения $\Phi_{a,\sigma^2}(x)$ любого нормального закона непрерывна всюду на $\mathbb{R}$ (*почему?*), утверждение ЦПТ можно сформулировать любым из следующих способов:

Следствие 19. *Пусть $\xi_1, \xi_2, \ldots$ — независимые и одинаково распределенные случайные величины с конечной и ненулевой дисперсией. Следующие утверждения эквивалентны друг другу и равносильны утверждению ЦПТ.*

- *Для любых вещественных $x < y$ при $n \to \infty$ имеет место сходимость*

$$\mathsf{P}\left(x < \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} < y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\,dt;$$

- *Для любых вещественных $x < y$ при $n \to \infty$ имеет место сходимость*

$$\mathsf{P}\left(x \leqslant \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \leqslant y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\,dt;$$

- *Для любых вещественных $x < y$ при $n \to \infty$ имеет место сходимость*

$$\mathsf{P}\left(x \leqslant \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}} \leqslant y\right) \to \Phi_{0,\mathsf{D}\,\xi_1}(y) - \Phi_{0,\mathsf{D}\,\xi_1}(x) = \frac{1}{\sqrt{\mathsf{D}\,\xi_1}}\int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\,dt;$$

- *Если η — произвольная с. в. со стандартным нормальным распределением, то*

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \Rightarrow \eta \Subset \mathbf{N}_{0,1}, \quad \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}} \Rightarrow \sqrt{\mathsf{D}\,\xi_1}\cdot\eta \Subset \mathbf{N}_{0,\mathsf{D}\,\xi_1}, \quad \sqrt{n}\left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) \Rightarrow \sqrt{\mathsf{D}\,\xi_1}\cdot\eta \Subset \mathbf{N}_{0,\mathsf{D}\,\xi_1}.$$

Замечание 25. Еще раз напомним, что функция распределения стандартного нормального закона ищется либо по соответствующей таблице в справочнике, либо с помощью какого-либо программного обеспечения, но никак не путем нахождения первообразной.

Мы докажем ЦПТ и ЗБЧ в форме Хинчина несколькими главами позднее. Нам потребуется для этого познакомиться с мощным математическим инструментом, который в математике обычно называют «преобразованиями Фурье», а в теории вероятностей — «характеристическими функциями».

14.4 Предельная теорема Муавра — Лапласа

Получим в качестве следствия из ЦПТ предельную теорему Муавра — Лапласа (P. S. Laplace, 1812; A. de Moivre, 1730). Подобно ЗБЧ Бернулли, предельная теорема Муавра — Лапласа — утверждение только для *схемы Бернулли*.

Теорема 33 (Предельная теорема Муавра — Лапласа).

> *Пусть A — событие, которое может произойти в любом из n независимых испытаний с одной и той же вероятностью $p = \mathsf{P}(A)$. Пусть $\nu_n(A)$ — число осуществлений события A в n испытаниях. Тогда $\dfrac{\nu_n(A) - np}{\sqrt{np(1-p)}} \Rightarrow \mathbf{N}_{0,1}$. Иначе говоря, для любых вещественных $x < y$ при $n \to \infty$ имеет место сходимость*
>
> $$\mathsf{P}\left(x \leqslant \frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \leqslant y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt;$$

Доказательство. По-прежнему $\nu_n(A)$ есть сумма независимых, одинаково распределенных с. в., имеющих распределение Бернулли с параметром, равным вероятности успеха $p = \mathsf{P}(A)$:

$$\nu_n(A) = \xi_1 + \dots + \xi_n, \quad \xi_i = I_i(A) = \begin{cases} 1, & \text{если } A \text{ произошло в } i-\text{м испытании;} \\ 0, & \text{если } A \text{ не произошло в } i-\text{м испытании;} \end{cases}$$

$$\mathsf{E}\,\xi_1 = \mathsf{P}(A) = p,\ \mathsf{D}\,\xi_1 = \mathsf{P}(A)(1 - \mathsf{P}(A)) = p(1-p).$$

Осталось воспользоваться ЦПТ. □

14.5 Примеры использования ЦПТ

Пример 50.

З а д а ч а из примера **48**. Монета подбрасывается 10 000 раз. Оценить вероятность того, что частота выпадения герба отличается от вероятности более чем на одну сотую.

Р е ш е н и е. Требуется найти $\mathsf{P}\left(\left|\dfrac{\nu_n}{n} - \dfrac{1}{2}\right| > 0{,}01\right)$, где $n = 10^4$, $\nu_n = \sum_{i=1}^n \xi_i = S_n$ — число выпадений герба, а ξ_i — независимые с. в., имеющие одно и то же распределение Бернулли с параметром 1/2. Домножим обе части неравенства под знаком вероятности на $\sqrt{n} = 100$ и поделим на корень из дисперсии $\sqrt{\mathsf{D}\,\xi_1} = 1/2$ одного слагаемого.

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| > 0{,}01\right) = 1 - \mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \leqslant 0{,}01\right) = 1 - \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left|\frac{S_n}{n} - \mathsf{E}\,\xi_1\right| \leqslant 0{,}01\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\right) = 1 - \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left|\frac{S_n}{n} - \mathsf{E}\,\xi_1\right| \leqslant 2\right).$$

Согласно ЦПТ или предельной теореме Муавра — Лапласа, последовательность

$$\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) = \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}}$$

слабо сходится к стандартному нормальному распределению. Рассмотрим произвольную с. в. η, имеющую распределение $\mathbf{N}_{0,1}$.

$$1 - \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left|\frac{\nu_n}{n} - \mathsf{E}\,\xi_1\right| \leqslant 2\right) \approx 1 - \mathsf{P}\,(|\eta| \leqslant 2) = 1 - (1 - 2\Phi_{0,1}(-2)) = 2\Phi_{0,1}(-2) = 2 \cdot 0.0228 = 0.0456.$$

Равенство $\mathsf{P}\,(|\eta| \leqslant 2) = 1 - 2\Phi_{0,1}(-2)$ следует из свойства **10**.

Замечание 26. Центральной предельной теоремой пользуются для приближенного вычисления вероятностей, связанных с суммами большого числа независимых и одинаково распределенных величин. При этом распределение центрированной и нормированной суммы заменяют на стандартное нормальное распределение. Насколько велика ошибка при такой замене (погрешность приближения)?

Упражнение 28. Какие еще предельные теоремы для схемы Бернулли вы знаете? Что такое теорема Пуассона? Найти ее. Какова погрешность пуассоновского приближения? Вычислить ее. Объяснить, исходя из полученной величины, почему **теорема Пуассона** не применима в задаче из примера **50**.

В примере **50** мы вычислили искомую вероятность тоже не точно, а приближенно — взгляните на равенство «≈» и спросите себя: насколько мы ошиблись? Стоит ли доверять ответу «0.0456»? Что, если разница между вероятностями в приближенном равенстве «≈» превосходит ответ на порядок? Следующий результат позволяет оценить погрешность приближения в ЦПТ.

Теорема 34 (Неравенство Берри – Эссе́ена).
В условиях ЦПТ для любого $x \in \mathbb{R}$ (то есть равномерно по x)

$$\left|\ \mathsf{P}\left(\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} < x\right) \ - \ \Phi_{0,1}(x)\ \right| \ \leqslant \ C \cdot \frac{\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3}{\sqrt{n}\left(\sqrt{\mathsf{D}\,\xi_1}\right)^3}.$$

Замечание 27. Про постоянную C известно, что:
а) в общем случае C не превышает 0.7655 (И. С. Шиганов),
б) погрешность приближения наиболее велика, если слагаемые ξ_i имеют распределение Бернулли, и C в этом

случае не меньше, чем $\dfrac{\sqrt{10}+3}{6\sqrt{2\pi}} \approx 0.4097$ (C. G. Esseen, Б. А. Рогозин),

в) как показывают расчеты, можно смело брать в качестве C число 0.4 — даже для слагаемых с распределением Бернулли, особенно при малых n, когда и это значение постоянной оказывается слишком грубой оценкой.

Подробный обзор можно найти в монографии В. М. Золотарева «Современная теория суммирования независимых случайных величин», стр. 264—291.

Продолжение примера 50. Проверьте, что для с. в. ξ_1 с распределением Бернулли

$$\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3 = |0-p|^3\ \mathsf{P}(\xi_1=0) + |1-p|^3\ \mathsf{P}(\xi_1=1) = p^3q + q^3p = pq(p^2+q^2).$$

Поэтому разница между левой и правой частями приближенного равенства «$\approx$» в примере **50** при $n = 10^4$ и $p = q = 1/2$ не превышает величины

$$C \cdot \frac{pq(p^2+q^2)}{\sqrt{n}\big(\sqrt{pq}\big)^3} = C \cdot \frac{p^2+q^2}{\sqrt{n}\sqrt{pq}} \leqslant 0.4 \cdot \frac{1}{100} = 0.004,$$

так что искомая вероятность $\mathsf{P}\left(\left|\dfrac{\nu_n}{n} - \dfrac{1}{2}\right| > 0{,}01\right)$ не больше, чем $0.0456 + 0.004.$ Уместно сравнить этот ответ с оценкой, полученной с помощью ЗБЧ в примере **48**.

Следующая проблема связана с распространеннейшим на ЭФ и ММФ НГУ заблуждением, которое можно образно передать афоризмом:

$$\mathsf{P}\left(\frac{S_n}{n} < \mathsf{E}\,\xi_1\right) \underset{n\to\infty}{\longrightarrow} \mathsf{P}\,(\mathsf{E}\,\xi_1 < \mathsf{E}\,\xi_1) = 0, \quad \text{но} \quad \mathsf{P}\left(\frac{S_n}{n} \leqslant \mathsf{E}\,\xi_1\right) \underset{n\to\infty}{\longrightarrow} \mathsf{P}\,(\mathsf{E}\,\xi_1 \leqslant \mathsf{E}\,\xi_1) = 1.$$

Пример 51.

З а д а ч а. Пусть $\xi_1, \xi_2, \ldots$ — независимые и одинаково распределенные случайные величины с конечной и ненулевой дисперсией, $S_n = \xi_1 + \cdots + \xi_n$ — сумма первых n случайных величин. При каких c имеет или не имеет место сходимость

$$\mathsf{P}\left(\frac{S_n}{n} < c\right) \to \mathsf{P}\,(\mathsf{E}\,\xi_1 < c)\,?$$

Решение. Согласно ЗБЧ, последовательность $\dfrac{S_n}{n}$ сходится по вероятности (а, следовательно, и *слабо*) к $\mathsf{E}\,\xi_1$.

Слабая сходимость означает, что последовательность функций распределения $F_n(c) = \mathsf{P}\left(\dfrac{S_n}{n} < c\right)$ сходится к функции распределения $F(c) = \mathsf{P}\,(\mathsf{E}\,\xi_1 < c)$, если $F(x)$ непрерывна в точке c (и ничего не означает, если $F(x)$ разрывна в точке c). Но

$$F(c) = \mathsf{P}\,(\mathsf{E}\,\xi_1 < c) = \begin{cases} 0, & c \leqslant \mathsf{E}\,\xi_1; \\ 1, & c > \mathsf{E}\,\xi_1 \end{cases}$$

есть функция распределения вырожденного закона и непрерывна в любой точке c, кроме $c = \mathsf{E}\,\xi_1$. Итак, первый вывод: сходимость $\mathsf{P}\left(\dfrac{S_n}{n} < c\right) \to \mathsf{P}\,(\mathsf{E}\,\xi_1 < c)$ имеет место для любого c, кроме, возможно, $c = \mathsf{E}\,\xi_1$. Убедимся, что для $c = \mathsf{E}\,\xi_1$ такой сходимости быть не может. Пусть $\eta \in \mathbf{N}_{0,1}$. Согласно ЦПТ,

$$\mathsf{P}\left(\frac{S_n}{n} < \mathsf{E}\,\xi_1\right) = \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) < 0\right) \to \mathsf{P}(\eta < 0) = \Phi_{0,1}(0) = \frac{1}{2} \neq \mathsf{P}\,(\mathsf{E}\,\xi_1 < \mathsf{E}\,\xi_1) = 0.$$

Аналогично, кстати, ведет себя и вероятность $\mathsf{P}\left(\dfrac{S_n}{n} \leqslant \mathsf{E}\,\xi_1\right)$. Она тоже стремится к $\dfrac{1}{2}$, а не к $\mathsf{P}\,(\mathsf{E}\,\xi_1 \leqslant \mathsf{E}\,\xi_1) = 1$.

И изящное упражнение на ту же тему:

Упражнение 29. Доказать, что

$$\lim_{n\to\infty} \int\limits_0^{0.999999n} \frac{1}{(n-1)!} y^{n-1} e^{-y}\, dy = 0; \quad \lim_{n\to\infty} \int\limits_0^{n} \frac{1}{(n-1)!} y^{n-1} e^{-y}\, dy = \frac{1}{2}; \quad \lim_{n\to\infty} \int\limits_0^{1.000001n} \frac{1}{(n-1)!} y^{n-1} e^{-y}\, dy = 1.$$

Указание. Каждый из интегралов равен функции распределения суммы независимых случайных величин с каким-то показательным распределением в некоторой точке. Вспомнить, что такое гамма-распределение и что такое «устойчивость относительно суммирования».

Раздел 15. Характеристические функции

Всюду в этой главе $i = \sqrt{-1}$ — мнимая единица, t — вещественная переменная, $e^{it} = \cos t + i \sin t$ — формула Эйлера, $\mathsf{E}(\eta + i\zeta) = \mathsf{E}\eta + i\,\mathsf{E}\zeta$ — способ вычисления математического ожидания комплекснозначной случайной величины $\eta + i\zeta$, если математические ожидания ее действительной (η) и мнимой (ζ) частей существуют.

Как всегда, модулем комплексного числа $z = x + iy$ называется $|z| = \sqrt{x^2 + y^2}$, так что $|e^{it}| = 1$.

Определение 53. Функция $\varphi_\xi(t) = \mathsf{E}\,e^{it\xi}$ называется ***характеристической функцией*** случайной величины ξ.

15.1 Примеры вычисления

Пример 52. Пусть с. в. ξ имеет распределение Бернулли с параметром p. Ее характеристическая функция (х. ф.) равна

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi} = e^{it\cdot 0}\,\mathsf{P}(\xi = 0) + e^{it\cdot 1}\,\mathsf{P}(\xi = 1) = 1 - p + pe^{it}.$$

Пример 53. Пусть с. в. ξ имеет биномиальное распределение с параметрами n и p. Ее х. ф. равна

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi} = \sum_{k=0}^{n} e^{it\cdot k}\,\mathsf{P}(\xi = k) = \sum_{k=0}^{n} e^{it\cdot k}\,C_n^k\,p^k\,(1-p)^{n-k} = \sum_{k=0}^{n} C_n^k\,\left(pe^{it}\right)^k (1-p)^{n-k} = \left(1 - p + pe^{it}\right)^n.$$

Последнее равенство суть бином Ньютона.

Пример 54. Пусть с. в. ξ имеет распределение Пуассона с параметром λ. Ее х. ф. равна

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi} = \sum_{k=0}^{\infty} e^{it\cdot k}\,\mathsf{P}(\xi = k) = \sum_{k=0}^{\infty} e^{it\cdot k}\,\frac{\lambda^k}{k!}\,e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda}e^{\lambda e^{it}} = \exp\{\lambda\left(e^{it} - 1\right)\}.$$

Пример 55. Пусть с. в. ξ имеет показательное распределение с параметром α. Ее х. ф. функция равна

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi} = \int\limits_0^\infty e^{it\cdot x}\,f_\xi(x)\,dx = \int\limits_0^\infty e^{itx}\,\alpha e^{-\alpha x}\,dx = \int\limits_0^\infty \alpha e^{-x(\alpha - it)}\,dx = \frac{\alpha}{\alpha - it}\left(-e^{-x(\alpha - it)}\Big|_0^\infty\right) = \frac{\alpha}{\alpha - it},$$

поскольку при $x \to \infty$ модуль величины $e^{-x(\alpha - it)} = e^{-\alpha x}\cdot e^{itx}$ стремится к нулю: $\left|e^{-x(\alpha - it)}\right| = e^{-\alpha x} \to 0$.

Пример 56. Пусть с. в. ξ имеет гамма-распределение с параметрами α, λ. Ее х. ф. равна

$$\varphi_\xi(t) = \mathsf{E}\, e^{it\xi} = \int\limits_0^\infty e^{it\cdot x}\, f_\xi(x)\, dx = \int\limits_0^\infty e^{itx}\, \frac{\alpha^\lambda}{\Gamma(\lambda)}\, x^{\lambda-1}\, e^{-\alpha x}\, dx = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int\limits_0^\infty x^{\lambda-1}\, e^{-x(\alpha-it)}\, dx = \left(\frac{\alpha}{\alpha-it}\right)^\lambda = \left(1-\frac{it}{\alpha}\right)^{-\lambda}.$$

Интеграл мы вычислили с помощью гамма-функции Эйлера:

$$\int\limits_0^\infty x^{\lambda-1}\, e^{-x(\alpha-it)}\, dx = \frac{1}{(\alpha-it)^\lambda} \int\limits_0^\infty (x(\alpha-it))^{\lambda-1}\, e^{-x(\alpha-it)}\, dx(\alpha-it) = \frac{\Gamma(\lambda)}{(\alpha-it)^\lambda}.$$

Пример 57. Пусть с. в. ξ имеет стандартное нормальное распределение. Ее х. ф. равна

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty e^{itx}\, e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty e^{-t^2/2}\, e^{-(x-it)^2/2}\, dx = e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty e^{-(x-it)^2/2}\, d(x-it) = e^{-t^2/2}.$$

Как всегда, при интегрировании мы выделили полный квадрат в показателе экспоненты и вспомнили, чему равен интеграл по всей прямой от функции $\dfrac{1}{\sqrt{2\pi}}\, e^{-u^2/2}$. *А чему он равен?*

Самое время остановиться и спросить: "Ну и что? Зачем нам эти функции и какой от них прок?" Приглашаю читателя познакомиться с замечательными свойствами х. ф.

15.2 Свойства характеристических функций

Ф1. Характеристическая функция всегда существует: по неравенству Йенсена

$$|\varphi_\xi(t)| = |\mathsf{E}\, e^{it\xi}| \leqslant \mathsf{E}\, |e^{it\xi}| = \mathsf{E}\, 1 = 1$$

Полезно вспомнить, что обычные математические ожидания существуют не у всех распределений.

Ф2. По характеристической функции однозначно восстанавливается распределение (функция распределения, а также плотность или таблица распределения). То есть если две с. в. имеют одинаковые х. ф., то и распределения этих с. в. совпадают.

Формулы, с помощью которых это делается, в анализе называют формулами «обратного преобразования Фурье». Например, если модуль х. ф. интегрируем на всей прямой, то у с. в. есть плотность распределения, и она находится по формуле (проверьте на примере примера **57**)

$$f_\xi(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-itx}\,\varphi_\xi(t)\,dt.$$

Ни одна из формул обратного преобразования Фурье нам не понадобится.

Ф3. Характеристическая функция с. в. $a+b\xi$ связана с х. ф. случайной величины ξ равенством

$$\varphi_{a+b\xi}(t) = \mathsf{E}\,e^{it(a+b\xi)} = e^{ita}\,\varphi_\xi(tb).$$

Пример 58. Вычислим х. ф. случайной величины ξ, имеющей нормальное распределение с параметрами a и σ^2. Мы знаем, что у стандартизованной с. в. $\zeta = \dfrac{\xi - a}{\sigma}$ характеристическая функция равна $\varphi_\zeta(t) = e^{-t^2/2}$. Тогда х. ф. случайной величины $\xi = a + \sigma\zeta$ равна

$$\varphi_\xi(t) = \varphi_{a+\sigma\zeta}(t) = e^{ita}\,\varphi_\zeta(t\sigma) == e^{ita}\,e^{-(t\sigma)^2/2} = \exp\left\{ita - \frac{t^2\sigma^2}{2}\right\}.$$

Ф4. Характеристическая функция *суммы независимых с. в.* равна произведению характеристических функций слагаемых: если с. в. ξ и η независимы, то, по свойству **E6** математических ожиданий

$$\varphi_{\xi+\eta}(t) = \mathsf{E}\,e^{it(\xi+\eta)} = \mathsf{E}\,e^{it\xi}\,\mathsf{E}\,e^{it\eta} = \varphi_\xi(t)\,\varphi_\eta(t).$$

Этим замечательным свойством мы сразу же воспользуемся, как обещали, для доказательства леммы **6**, утверждающей устойчивость нормального распределения относительно суммирования.

Пример 59. **Доказательство леммы 6.**
Пусть случайные величины $\xi \in \mathbf{N}_{a_1,\sigma_1^2}$ и $\eta \in \mathbf{N}_{a_2,\sigma_2^2}$ независимы. Характеристическая функция суммы $\xi+\eta$ равна

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\,\varphi_\eta(t) = \exp\left\{ita_1 - \frac{t^2\sigma_1^2}{2}\right\}\exp\left\{ita_2 - \frac{t^2\sigma_2^2}{2}\right\} = \exp\left\{it(a_1+a_2) - \frac{t^2(\sigma_1^2+\sigma_2^2)}{2}\right\}.$$

То есть х. ф. суммы есть характеристическая функция нормального распределения с параметрами a_1+a_2, $\sigma_1^2+\sigma_2^2$. Тогда $\xi+\eta \in \mathbf{N}_{a_1+a_2,\sigma_1^2+\sigma_2^2}$ по свойству **Ф 2**. □

Пример 60. Докажем свойства устойчивости по суммированию биномиального распределения, распределения Пуассона и гамма-распределения (леммы **4**, **5**, **7**), используя вычисленные в примерах **52**−**56** характеристические функции.

Для независимых с. в. с распределениями Пуассона Π_λ и Π_μ х. ф. суммы

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\,\varphi_\eta(t) = \exp\left\{\lambda\left(e^{it}-1\right)\right\}\,\exp\left\{\mu\left(e^{it}-1\right)\right\} = \exp\left\{(\lambda+\mu)\left(e^{it}-1\right)\right\}$$

равна характеристической функции распределения Пуассона с параметром $\lambda+\mu$.

Для независимых с. в. с биномиальными распределениями $\mathbf{B}_{n,p}$ и $\mathbf{B}_{m,p}$ х. ф. суммы

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\,\varphi_\eta(t) = \left(1-p+pe^{it}\right)^n\left(1-p+pe^{it}\right)^m = \left(1-p+pe^{it}\right)^{n+m}.$$

равна характеристической функции биномиального распределения с параметрами $n+m$, p.

Для n независимых с. в. с показательным распределением $\mathbf{E}_\alpha$ х. ф. суммы

$$\varphi_{\xi_1+\dots+\xi_n}(t) = (\varphi_{\xi_1}(t))^n = \left(\frac{\alpha}{\alpha-it}\right)^n = \left(1-\frac{it}{\alpha}\right)^{-n}$$

равна характеристической функции гамма-распределения с параметрами α, n.

Ф5. Пусть существует момент порядка $k=1,2,\ldots$ случайной величины ξ, то есть $\mathsf{E}|\xi|^k<\infty$. Тогда ее характеристическая функция $\varphi_\xi(t)$ непрерывно дифференцируема k раз, и ее k-я производная ***в нуле*** связана с моментом порядка k равенством:

$$\varphi_\xi^{(k)}(0) = \left(\frac{d^k}{d\,t^k}\,\mathsf{E}\,e^{it\xi}\right)\Bigg|_{t=0} = \left(\mathsf{E}\,i^k\,\xi^k\,e^{it\xi}\right)\Bigg|_{t=0} = i^k\,\mathsf{E}\,\xi^k.$$

Существование и непрерывность k-й производной, равно как и законность переноса производной под знак математического ожидания мы доказывать не будем.

Упражнение 30. Доказать, что для с. в. ξ со стандартным нормальным распределением момент четного порядка $2k$ равен $\mathsf{E}\,\xi^{2k} = (2k-1)!! \stackrel{\text{def}}{=} (2k-1)\cdot(2k-3)\cdot\ldots\cdot 3\cdot 1$.

Доказать по определению, что все моменты нечетных порядков стандартного нормального распределения существуют и равны нулю.

Как только появились производные высших порядков, самое время разложить функцию в ряд Тейлора.

Ф6. Пусть существует момент порядка $k=1,2,\ldots$ случайной величины ξ, то есть $\mathsf{E}|\xi|^k<\infty$. Тогда ее характеристическая функция $\varphi_\xi(t)$ в окрестности точки $t=0$ разлагается в ряд Тейлора

$$\varphi_\xi(t) = \varphi_\xi(0) + \sum_{j=1}^{k}\frac{t^j}{j!}\,\varphi_\xi^{(j)}(0) + o(|t^k|) = 1 + \sum_{j=1}^{k}\frac{i^jt^j}{j!}\,\mathsf{E}\,\xi^j + o(|t^k|) = 1 + it\,\mathsf{E}\,\xi - \frac{t^2}{2}\,\mathsf{E}\,\xi^2 + \ldots + \frac{i^kt^k}{k!}\,\mathsf{E}\,\xi^k + o(|t^k|).$$

Ряды Тейлора, как правило, возникают при предельном переходе. Следующее основное свойство характеристических функций потребуется нам для доказательства предельных теорем, и это свойство — последняя теорема, оставленная нами без доказательства.

Теорема 35 (Теорема о непрерывном соответствии). *Случайные величины ξ_n слабо сходятся к с. в. ξ тогда и только тогда, когда для любого t характеристические функции $\varphi_{\xi_n}(t)$ сходятся к характеристической функции $\varphi_\xi(t)$.*

Сформулированная теорема устанавливает непрерывное соответствие между классами $\langle F_\xi, \Rightarrow \rangle$ функций распределения со слабой сходимостью и $\langle \varphi_\xi, \to \rangle$ характеристических функций со сходимостью в каждой точке. «Непрерывность» этого соответствия — в том, что пределу в одном классе относительно заданной в этом классе сходимости соответствует предел в другом классе относительно сходимости, заданной в этом классе.

Осталось воспользоваться теоремой о непрерывном соответствии и доказать ЗБЧ в форме Хинчина и ЦПТ.

15.3 Доказательство ЗБЧ Хинчина

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых и одинаково распределенных случайных величин с конечным ***первым*** моментом $\mathsf{E}|\xi_1| < \infty$. Обозначим через a математическое ожидание $\mathsf{E}\xi_1$. Требуется доказать, что

$$\frac{S_n}{n} = \frac{\xi_1 + \cdots + \xi_n}{n} \xrightarrow{\text{p}} a.$$

По свойству **19** сходимость по вероятности ***к постоянной*** эквивалентна слабой сходимости. Так как a — постоянная, достаточно доказать слабую сходимость $\dfrac{S_n}{n} = \dfrac{\xi_1 + \cdots + \xi_n}{n} \Rightarrow a$. По теореме о непрерывном соответствии, эта сходимость имеет место, если и только если для любого $t \in \mathbb{R}$ сходятся характеристические функции

$$\varphi_{S_n/n}(t) \to \varphi_a(t) = \mathsf{E}\, e^{ita} = e^{ita}.$$

Найдем характеристическую функцию с. в. $\dfrac{S_n}{n}$. Пользуясь свойствами **Ф3** и **Ф4**, получим

$$\varphi_{S_n/n}(t) \overset{\textbf{Ф3}}{=} \varphi_{S_n}\left(\frac{t}{n}\right) \overset{\textbf{Ф4}}{=} \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n.$$

Вспомним, что первый момент ξ_1 существует, поэтому свойство **Ф6** позволяет разложить $\varphi_{\xi_1}(t)$ в ряд Тейлора в окрестности нуля $\varphi_{\xi_1}(t) = 1 + it\,\mathsf{E}\,\xi_1 + o(|t|) = 1 + ita + o(|t|)$. В точке t/n, соответственно,

$$\varphi_{\xi_1}\left(\frac{t}{n}\right) = 1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right), \qquad \varphi_{S_n/n}(t) = \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right)\right)^n.$$

При $n \to \infty$, пользуясь «замечательным пределом» $\left(1+\dfrac{x}{n}\right)^n \to e^x$, получим

$$\varphi_{S_n/n}(t) \;=\; \left(1+\frac{ita}{n}+o\left(\left|\frac{t}{n}\right|\right)\right)^n \;\to\; e^{ita},$$

что и требовалось доказать. $\square$

15.4 Доказательство центральной предельной теоремы

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых и одинаково распределенных случайных величин с конечной и ненулевой дисперсией. Обозначим через a математическое ожидание $\mathsf{E}\xi_1$ и через σ^2 — дисперсию $\mathsf{D}\xi_1$. Требуется доказать, что

$$\frac{S_n - na}{\sigma\sqrt{n}} = \frac{\xi_1+\cdots+\xi_n - na}{\sigma\sqrt{n}} \Rightarrow \mathbf{N}_{0,1}.$$

Введем стандартизованные случайные величины $\zeta_i = \dfrac{\xi_i - a}{\sigma}$ — независимые с. в. с нулевыми математическими ожиданиями и единичными дисперсиями. Пусть Z_n есть их сумма $Z_n = \zeta_1 + \cdots + \zeta_n = (S_n - na)/\sigma$. Требуется доказать, что

$$\frac{Z_n}{\sqrt{n}} \Rightarrow \mathbf{N}_{0,1}.$$

Характеристическая функция величины $Z_n/\sqrt{n}$ равна

$$\varphi_{Z_n/\sqrt{n}}(t) \overset{\textbf{Ф3}}{=} \varphi_{Z_n}\left(\frac{t}{\sqrt{n}}\right) \overset{\textbf{Ф4}}{=} \left(\varphi_{\zeta_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n. \tag{25}$$

Характеристическую функцию с. в. ζ_1 можно разложить в ряд Тейлора, в коэффициентах которого использовать известные моменты $\mathsf{E}\zeta_1 = 0$, $\mathsf{E}{\zeta_1}^2 = \mathsf{D}\zeta_1 = 1$. Получим

$$\varphi_{\zeta_1}(t) \overset{\textbf{Ф6}}{=} 1 + it\,\mathsf{E}\zeta_1 - \frac{t^2}{2}\,\mathsf{E}{\zeta_1}^2 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2).$$

Подставим это разложение, взятое в точке $t/\sqrt{n}$, в равенство (25) и устремим n к бесконечности. Еще раз воспользуемся замечательным пределом.

$$\varphi_{Z_n/\sqrt{n}}(t) = \left(\varphi_{\zeta_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to \exp\left\{-\frac{t^2}{2}\right\} \quad \text{при} \quad n \to \infty.$$

В пределе получили характеристическую функцию стандартного нормального закона. По теореме о непрерывном соответствии можно сделать вывод о слабой сходимости

$$\frac{Z_n}{\sqrt{n}} = \frac{S_n - na}{\sigma\sqrt{n}} \Rightarrow \mathbf{N}_{0,1}$$

распределений стандартизованных сумм к стандартному нормальному распределению, что и утверждается в ЦПТ. $\square$

Попробуйте теперь сами:

Упражнение 31. Пусть при любом $\lambda > 0$ случайная величина ξ_λ имеет распределение Пуассона с параметром λ. Используя теорему о непрерывном соответствии, доказать, что случайные величины

$$\frac{\xi_\lambda - \lambda}{\sqrt{\lambda}}$$

слабо сходятся к стандартному нормальному распределению при $\lambda \to \infty$.

Характеристическая функция с. в. ξ_λ вычислена в примере **54**.

Раздел 16. Доказательство теоремы Пуассона

Нам осталось доказать теорему Пуассона. В доказательство будут использоваться только свойства устойчивости биномиального и пуассоновского распределений относительно операции суммирования. Никакие разделы, связанные с числовыми характеристиками с. в., сходимостями или характеристическими функциями, нам в доказательстве не понадобятся.

Вспомним утверждение, которое мы собрались доказывать. Теперь, когда мы знакомы с термином «распределение», можно сформулировать теорему Пуассона так:

Теорема Пуассона с оценкой погрешности

Пусть $A \subseteq \{0, 1, 2, \ldots, n\}$ — произвольное множество целых неотрицательных чисел, случайная величина ν_n имеет биномиальное распределение $\mathbf{B}_{n,p}$ с параметрами n и p, случайная величина μ_n имеет распределение Пуассона с параметром $\lambda = np$. Тогда

$$\left|\ \mathsf{P}(\nu_n \in A) - \ \mathsf{P}(\mu_n \in A)\ \right| = \left|\ \sum_{k \in A} C_n^k\ p^k\ (1-p)^{n-k} - \sum_{k \in A} \frac{\lambda^k}{k!}\ e^{-\lambda}\ \right| \leqslant np^2 = \frac{\lambda^2}{n}.$$

Иначе говоря, требуется доказать, что

$$\sup_A \left|\ \mathsf{P}(\nu_n \in A) - \ \mathsf{P}(\mu_n \in A)\ \right| \leqslant np^2.$$

Доказательство проведем, используя так называемый «метод одного вероятностного пространства». Нам нужно оценить сверху разницу между двумя распределениями, а именно: доказать, что для любых множеств $A \subseteq \{0, 1, 2, \ldots, n\}$ разницу между вероятностями попадания в множество A биномиальной (с параметрами n и p) и пуассоновской (с параметром np) случайных величин можно оценить величиной np^2.

Заметим, прежде всего, что разность

$$\left|\ \mathsf{P}(\nu_n \in A) - \ \mathsf{P}(\mu_n \in A)\ \right| = \left|\ \sum_{k \in A} C_n^k\ p^k\ (1-p)^{n-k} - \sum_{k \in A} \frac{\lambda^k}{k!}\ e^{-\lambda}\ \right|$$

никак не зависит от того, каким образом величины ν_n и μ_n взаимосвязаны и на каком вероятностном пространстве заданы, если только одна из этих величин имеет биномиальное, а вторая — пуассоновское распределение с нужными параметрами. Совместное распределение этих величин тут никак не участвует, поэтому данная разность не изменится, если заменить ν_n и μ_n на *другие* случайные величины с теми же распределениями.

Первое, что мы сделаем — докажем, что для двух случайных величин ξ и η (где угодно заданных) «расстояние между распределениями», то есть $\sup_A |\,\mathsf{P}(\xi \in A) - \mathsf{P}(\eta \in A)\,|$, не больше, чем вероятность $\mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right)$ двум ***произвольным*** случайным величинам $\tilde{\xi}, \tilde{\eta}$ с данными распределениями не совпадать. Понятно, что эти новые с. в. должны быть заданы на одном вероятностном пространстве, и наилучшая оценка сверху получится, если нам удастся ***так задать на одном вероятностном пространстве с. в. $\tilde{\xi}$, распределенную как ξ, и $\tilde{\eta}$, распределенную как η, чтобы вероятность $\mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right)$ была наименьшей***.

Лемма 9 (Неравенство каплинга). *Пусть ξ и η — произвольные с. в. Пусть случайная величина $\tilde{\xi}$ одинаково распределена с ξ, случайная величина $\tilde{\eta}$ одинаково распределена с η, и величины $\tilde{\xi}, \tilde{\eta}$ заданы на одном вероятностном пространстве. Тогда*

$$\sup_{A \subseteq \mathbb{R}} |\,\mathsf{P}(\xi \in A) - \mathsf{P}(\eta \in A)\,| \leqslant \mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right).$$

Замечание 28. *Каплингом* (*coupling*) двух с. в. ξ и η называют задание на одном вероятностном пространстве случайных величин $\tilde{\xi}$, распределенной как ξ, и $\tilde{\eta}$, распределенной как η.

Доказательство неравенства каплинга. Воспользуемся равенством $\mathsf{P}(C) = \mathsf{P}(C \cap B) + \mathsf{P}(C \cap \overline{B})$, а также тем, что вероятность пересечения двух событий не превосходит вероятности любого из них. Для любого множества $A \subseteq \mathbb{R}$

$$\begin{aligned}\mathsf{P}(\xi \in A) = \mathsf{P}\left(\tilde{\xi} \in A\right) &= \mathsf{P}\left(\tilde{\xi} \in A, \tilde{\xi} = \tilde{\eta}\right) + \mathsf{P}\left(\tilde{\xi} \in A, \tilde{\xi} \neq \tilde{\eta}\right) = \mathsf{P}\left(\tilde{\eta} \in A, \tilde{\xi} = \tilde{\eta}\right) + \mathsf{P}\left(\tilde{\xi} \in A, \tilde{\xi} \neq \tilde{\eta}\right) \leqslant \\ &\leqslant \mathsf{P}\left(\tilde{\eta} \in A\right) + \mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right) = \mathsf{P}\left(\eta \in A\right) + \mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right),\end{aligned}$$

то есть

$$\mathsf{P}(\xi \in A) - \mathsf{P}\left(\eta \in A\right) \leqslant \mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right).$$

Поменяем местами ξ и η и получим, что для любого множества $A \subseteq \mathbb{R}$

$$|\,\mathsf{P}(\xi \in A) - \mathsf{P}(\eta \in A)\,| \leqslant \mathsf{P}\left(\tilde{\xi} \neq \tilde{\eta}\right).$$

□

Займемся заданием на одном вероятностном пространстве величин $\tilde{\nu}_n$ и $\tilde{\mu}_n$, распределенных как ν_n и μ_n, соответственно.

Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, имеющие распределение Бернулли с параметром p. Тогда их сумма $\tilde{\nu}_n = \xi_1 + \ldots + \xi_n$ имеет биномиальное распределение с параметрами n и p, то есть одинаково распределена с ν_n.

Пусть $\eta_1, \ldots, \eta_n$ — независимые случайные величины, имеющие распределение Пуассона с параметром p. Тогда их сумма $\tilde{\mu}_n = \eta_1 + \ldots + \eta_n$ также имеет распределение Пуассона с параметром, равным сумме параметров слагаемых, то есть np, и одинаково распределена с μ_n. Мы будем считать, что эти наборы с. в. сразу заданы на одном вероятностном пространстве, и позже построим их.

Тогда, в силу неравенства каплинга,

$$| \mathsf{P}(\nu_n \in A) - \mathsf{P}(\mu_n \in A) | \leqslant \mathsf{P}\left(\tilde{\nu}_n \neq \tilde{\mu}_n\right) = \mathsf{P}\left(\sum_{i=1}^{n} \xi_i \neq \sum_{i=1}^{n} \eta_i\right).$$

Заметим теперь, что если две суммы с неотрицательными слагаемыми не равны друг другу, то хотя бы одно слагаемое в первой сумме отличается от соответствующего слагаемого в другой сумме (*иначе...*). Поэтому

$$| \mathsf{P}(\nu_n \in A) - \mathsf{P}(\mu_n \in A) | \leqslant \mathsf{P}\left(\sum_{i=1}^{n} \xi_i \neq \sum_{i=1}^{n} \eta_i\right) \leqslant \mathsf{P}\left(\bigcup_{i=1}^{n} \{\xi_i \neq \eta_i\}\right) \leqslant \sum_{i=1}^{n} \mathsf{P}\left(\xi_i \neq \eta_i\right). \qquad (26)$$

В последнем неравенстве использовано, что вероятность объединения не превосходит суммы вероятностей.

Осталось теперь так задать на одном вероятностном пространстве ξ_i и η_i, чтобы минимизировать $\mathsf{P}\left(\xi_i \neq \eta_i\right)$.

Пусть множество элементарных исходов Ω есть n-мерный куб, стороны которого — отрезки $[0,1]$ на осях координат, вероятность есть просто мера Лебега, заданная на σ-алгебре борелевских множеств.

Вот ровно сейчас тот, кто поленился о них прочитать, должен об этом пожалеть!

То есть мы наудачу выбираем точку $\omega = (\omega_1, \ldots, \omega_n)$ в кубе, или, что то же самое, каждую из координат ω_i выбираем наудачу и независимо от других на $[0,1]$.

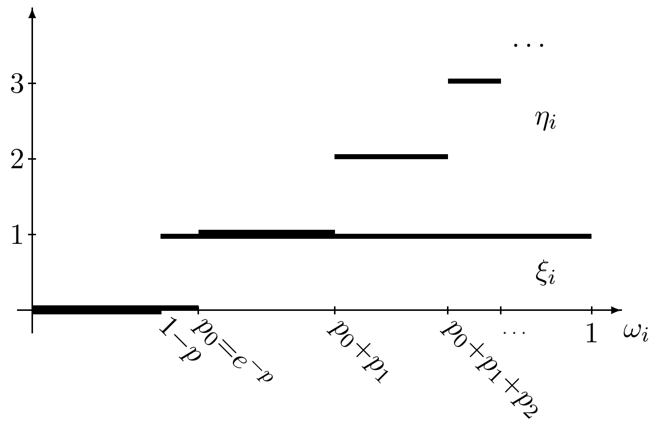
Построим для каждого $i = 1, \ldots, n$ по ω_i случайные величины $\xi_i = \xi_i(\omega_i)$ и $\eta_i = \eta_i(\omega_i)$ с нужными распределениями, чтобы они, к тому же, совпадали с большой вероятностью. Положим

$$\xi_i(\omega_i) = \begin{cases} 0, & \text{если } 0 \leqslant \omega_i < 1-p, \\ 1, & \text{если } 1-p \leqslant \omega_i \leqslant 1. \end{cases}$$

Эта с. в. имеет распределение Бернулли: $\mathsf{P}(\xi_i = 0) = \mathsf{P}(0 \leqslant \omega_i < 1-p) = 1-p$, $\mathsf{P}(\xi_i = 1) = \mathsf{P}(1-p \leqslant \omega_i \leqslant 1) = p$.

Случайная величина η_i должна иметь распределение Пуассона с параметром p, то есть $p_k = \mathsf{P}(\eta_i = k) = \dfrac{p^k}{k!}\, e^{-p}$ при $k = 0, 1, 2, \ldots$. Сумма этих вероятностей равна 1, поэтому можно разбить ***тот же самый*** отрезок $[0,1]$ на отрезки, длина k-го из которых равна p_k при $k = 0, 1, 2, \ldots$, и положить $\eta_i = k$, если ω_i принадлежит отрезку с номером k:

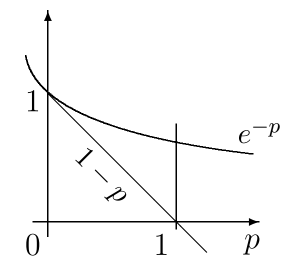
$$\eta_i(\omega_i) = \begin{cases} 0, & \text{если } 0 \leqslant \omega_i < p_0, \\ 1, & \text{если } p_0 \leqslant \omega_i < p_0 + p_1, \\ 2, & \text{если } p_0 + p_1 \leqslant \omega_i < p_0 + p_1 + p_2, \\ \ldots \\ k, & \text{если } p_0 + \ldots + p_{k-1} \leqslant \omega_i < p_0 + \ldots + p_k, \\ \ldots \end{cases}$$



С очевидностью, получим с. в. с распределением Пуассона:

$$\mathsf{P}(\eta_i = k) = \mathsf{P}(p_0 + \ldots + p_{k-1} \leqslant \omega_i < p_0 + \ldots + p_{k-1} + p_k) = p_k, \quad k = 0, 1, 2, \ldots.$$

Отметим, что $p_0 = \frac{p^0}{0!}\, e^{-p} = e^{-p}$. Докажем, что $e^{-p} \geqslant 1 - p$ при $p \geqslant 0$. Действительно,



а) при $p = 0$ значения функций совпадают: $e^{-0} = 1 - 0 = 1$;

б) производные в нуле у e^{-p} и $1 - p$ также совпадают (и равны -1)

в) при $p = 1$ левая часть больше правой: $e^{-1} > 1 - 1 = 0$;

г) функция e^{-p} выпукла (ее производная отрицательна всюду), так что коснувшись однажды прямой $1 - p$, она ее не перекает нигде, оставаясь всегда больше.

Посмотрим, с какой вероятностью с. в. ξ_i и η_i не совпадают. Это происходит при $1 - p \leqslant \omega_i < e^{-p}$ — на этом интервале $\xi_i = 1$, а $\eta_i = 0$, а также при $\omega_i \geqslant p_0 + p_1 = e^{-p} + pe^{-p}$ — на этом интервале $\xi_i = 1$, а $\eta_i \geqslant 2$. Поэтому

$$\mathsf{P}(\xi_i \neq \eta_i) = \mathsf{P}(1 - p \leqslant \omega_i < e^{-p} \text{ или } e^{-p} + pe^{-p} \leqslant \omega_i \leqslant 1) = \big(e^{-p} - (1 - p)\big) + \big(1 - \big(e^{-p} + pe^{-p}\big)\big) = p\left(1 - e^{-p}\right) \leqslant p^2.$$

В последнем неравенстве мы снова воспользовались тем, что $e^{-p} \geqslant 1 - p$, или $1 - e^{-p} \leqslant p$.

Итак, при каждом $i = 1, \ldots, n$ мы построили пару с. в. ξ_i, η_i, отличающихся с вероятностью не более p^2. При разных i эти с. в. независимы, так как построены по независимым координатам точки, выбранной наудачу в кубе. Окончательно, из неравенства (26) получим:

$$\left|\ \mathsf{P}(\nu_n \in A) - \ \mathsf{P}(\mu_n \in A)\ \right| \leqslant \sum_{i=1}^{n} \mathsf{P}\left(\xi_i \neq \eta_i\right) \leqslant np^2.$$

$\square$

Приложение. Знакомьтесь — максимум и минимум

В этом разделе, который никогда не будет прочитан на лекциях, хотя бы потому, что эта тема подробно разбирается на практических занятиях, мы поговорим о «максимуме и минимуме из n случайных величин». Вдумчивый читатель уже смог догадаться, что к клубу «Максимин» ЭФ НГУ данная тема касательства не имеет.

Пусть с. в. $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы и одинаково распределены, $F_{\xi_1}(x)$ — их общая функция распределения.

Определение № N. Случайную величину $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$ назовем ***максимумом***, а случайную величину $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$ — ***минимумом*** из n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Замечание № N. Заметим на всякий случай, что $\varphi_n(\omega) = \max\{\xi_1(\omega), \ldots, \xi_n(\omega)\}$, так что φ_n на каждом элементарном исходе совпадает ***с одной из*** ξ_i, $1 \leqslant i \leqslant n$, но ***ни с одной*** из них не совпадает ***при всех*** ω (если с. в. независимы).

Упражнение № N. Доказать, что вероятность максимуму из первых n ***независимых и одинаково распределенных*** случайных величин, имеющих, к примеру, равномерное распределение, равняться первой из них (или любой другой), есть $1/n$:

$$\mathsf{P}(\max\{\xi_1, \ldots, \xi_n\} = \xi_1) = \frac{1}{n} = \mathsf{P}(\xi_1 > \xi_2, \ldots, \xi_1 > \xi_n),$$

то есть «в среднем в $1/n$ случаев максимум совпадает с выбранной вами среди $\xi_1, \ldots, \xi_n$ величиной».

Для доказательства воспользоваться соображениями симметрии, разбив все пространство Ω на не(сколько?) равновероятных событий типа $\{\xi_1 > \xi_2, \ldots, \xi_1 > \xi_n\}$ и несколько событий нулевой вероятности, включающих возможные равенства. Вспомнить, с какой вероятностью две (или больше) из $\{\xi_i\}$ совпадают (нарисовать событие $\{\xi_1 = \xi_2\}$ в квадрате на плоскости).

Лемма № N. *Функции распределения случайных величин $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$ и $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$ равны, соответственно,*

$$F_{\varphi_n}(x) = \big(F_{\xi_1}(x)\big)^n \quad \text{и} \quad F_{\psi_n}(x) = 1 - \big(1 - F_{\xi_1}(x)\big)^n.$$

Доказательство. Найдем функцию распределения $F_{\varphi_n}(x)$. Максимум из n величин меньше x тогда и только тогда, когда каждая из этих величин меньше x.

$$\begin{aligned} F_{\varphi_n}(x) &= \mathsf{P}\big(\max\{\xi_1,\ldots,\xi_n\}<x\big) = \mathsf{P}\big(\xi_1<x,\ldots,\xi_n<x\big) \overset{\text{независ.}}{=} \\ &= \mathsf{P}\big(\xi_1<x\big)\cdot\ldots\cdot\mathsf{P}\big(\xi_n<x\big) \overset{\text{од.распред.}}{=} \Big(\mathsf{P}\big(\xi_1<x\big)\Big)^n = \big(F_{\xi_1}(x)\big)^n. \end{aligned}$$

Найдем функцию распределения $F_{\psi_n}(x)$. Минимум из n величин ***не меньше*** x тогда и только тогда, когда каждая из этих величин не меньше x.

$$\begin{aligned} F_{\psi_n}(x) &= \mathsf{P}\big(\min\{\xi_1,\ldots,\xi_n\}<x\big) = 1-\mathsf{P}\big(\min\{\xi_1,\ldots,\xi_n\}\geqslant x\big) = 1-\mathsf{P}\big(\xi_1\geqslant x,\ldots,\xi_n\geqslant x\big) \overset{\text{независ.}}{=} \\ &= 1-\mathsf{P}\big(\xi_1\geqslant x\big)\cdot\ldots\cdot\mathsf{P}\big(\xi_n\geqslant x\big) \overset{\text{од.распред.}}{=} 1-\Big(\mathsf{P}\big(\xi_1\geqslant x\big)\Big)^n = 1-\big(1-F_{\xi_1}(x)\big)^n. \end{aligned}$$

□

Пример № N. Пусть с. в. $\xi_1,\xi_2,\ldots,\xi_n,\ldots$ независимы и имеют равномерное распределение на отрезке $[0,1]$. Докажем, что последовательность с. в. $\varphi_1=\xi_1$, $\varphi_2=\max\{\xi_1,\xi_2\}$, … , $\varphi_n=\max\{\xi_1,\ldots,\xi_n\}$, … сходится по вероятности к правому концу отрезка — к 1. Не употребляя термин «последовательность», можно произнести это утверждение так: «максимум из первых n случайных величин с ростом n сходится к единице по вероятности».

Есть как минимум два способа доказательства:

Способ 1. По определению. Пусть $\varepsilon>0$. Найдем $\mathsf{P}(|\varphi_n-1|>\varepsilon)$. Заметим, что $\varphi_n\leqslant 1$, поскольку это максимум из с. в., принимающих значения на отрезке $[0,1]$ (крайняя правая из «координат n точек, брошенных наудачу на [0,1] независимо друг от друга»). Поэтому

$$\mathsf{P}\big(|\varphi_n-1|>\varepsilon\big) = \mathsf{P}\big(1-\varphi_n>\varepsilon\big)$$

Для того, чтобы установить сходимость последней вероятности к нулю, можно ее либо найти, либо оценить с помощью неравенства Чебышёва (Маркова).

1(a). Найдем эту вероятность.

$$\mathsf{P}\big(1-\varphi_n>\varepsilon\big) = \mathsf{P}\big(1-\varepsilon>\varphi_n\big) = \mathsf{P}\big(\varphi_n<1-\varepsilon\big) = F_{\varphi_n}(1-\varepsilon).$$

Для равномерного распределения на отрезке $[0,1]$

$$F_{\xi_1}(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1, \end{cases} \quad \text{поэтому} \quad F_{\varphi_n}(x) = \bigl(F_{\xi_1}(x)\bigr)^n = \begin{cases} 0, & x < 0; \\ x^n, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1. \end{cases}$$

А если еще заметить, что $1-\varepsilon < 1$, то

$$\mathsf{P}\bigl(|\varphi_n - 1| > \varepsilon\bigr) = F_{\varphi_n}(1-\varepsilon) = \begin{cases} 0, & 1-\varepsilon < 0, \ \text{ то есть } \ \varepsilon > 1; \\ (1-\varepsilon)^n, & 0 \leqslant 1-\varepsilon < 1, \ \text{ то есть } \ 0 < \varepsilon \leqslant 1 \end{cases} \longrightarrow 0 \ \text{ при } \ n \to \infty.$$

1(б). Оценим вероятность сверху. Поскольку $1-\varphi_n \geqslant 0$, по неравенству Маркова

$$\mathsf{P}\bigl(1-\varphi_n > \varepsilon\bigr) \leqslant \frac{\mathsf{E}\,(1-\varphi_n)}{\varepsilon} = \frac{1-\mathsf{E}\,\varphi_n}{\varepsilon}. \tag{27}$$

Найдем плотность распределения с. в. φ_n и математическое ожидание $\mathsf{E}\,\varphi_n$.

$$f_{\varphi_n}(x) = \bigl(F_{\varphi_n}(x)\bigr)' = \begin{cases} 0, & x < 0; \\ nx^{n-1}, & 0 \leqslant x \leqslant 1 \\ 0, & x > 1; \end{cases} \qquad \mathsf{E}\,\varphi_n = \int\limits_0^1 x\, nx^{n-1}\, dx = \frac{n}{n+1}.$$

Подставляя математическое ожидание в неравенство (27), получим

$$\mathsf{P}\bigl(1-\varphi_n > \varepsilon\bigr) \leqslant \frac{1-\mathsf{E}\,\varphi_n}{\varepsilon} = \frac{1-\dfrac{n}{n+1}}{\varepsilon} = \frac{1}{(n+1)\,\varepsilon} \ \to \ 0 \text{ при } n \to \infty.$$

Способ 2. Используем связь со слабой сходимостью. Сходимость по вероятности *к константе* равносильна слабой сходимости (свойство **19**). Докажем поэтому, что φ_n слабо сходится к единице. Требуется доказать, что функция распределения $F_{\varphi_n}(x)$ сходится к $F_1(x) = \mathsf{P}(1 < x)$ для любого $x \neq 1$ (***почему кроме 1??***).

При любом $x \neq 1$ имеем:

$$F_{\varphi_n}(x) = \begin{cases} 0, & x < 0; \\ x^n, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1 \end{cases} \quad \to \quad F_1(x) = \begin{cases} 0, & x \leqslant 1 \\ 1, & x > 1, \end{cases}$$

и только при $x = 1$ сходимости нет: $F_{\varphi_n}(1) = 1$, тогда как $F_1(1) = 0$.

Таким образом, φ_n сходится слабо к единице, и, следовательно, сходится к ней же по вероятности.

Упражнение № N+1. Доказать (способами 1(а), 1(б) и 2), что, в условиях примера **N**, последовательность $\psi_1, \psi_2, \ldots, \psi_n, \ldots$ сходится по вероятности к нулю (мы будем говорить «минимум из первых n случайных величин с ростом n сходится к нулю по вероятности»).

Красивых задач, связанных с максимумом и минимумом, слишком много, Предлагаю вам решить, например, следующие:

Пусть с. в. $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы и имеют равномерное распределение на отрезке $[a, b]$, $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$, $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$. Доказать, что

1) $\dfrac{n}{b-a}(b - \varphi_n)$ при $n \to \infty$ слабо сходится к показательному распределению с параметром 1;

2) точно так же себя ведет последовательность $\dfrac{n}{b-a}(\psi_n - a)$;

3) это не удивительно, поскольку с. в. $b - \varphi_n$ и $\psi_n - a$ одинаково распределены;

4) посчитав вероятность $\mathsf{P}(\psi_n \geqslant x,\ \varphi_n < y) = \big(\mathsf{P}(x \leqslant \xi_1 < y)\big)^n$, можно легко найти функцию совместного распределения с. в. ψ_n, φ_n и с ее помощью, например, доказать зависимость этих с. в. (последнее и так очевидно, не правда ли?)

Теперь, наконец,

ПОРА ИЗУЧАТЬ МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ!

Рекомендуемая литература

[1] *Гнеденко Б. В.* Курс теории вероятностей. М., 1988.

[2] *Чистяков В. П.* Курс теории вероятностей. М., 1982.

[3] *Боровков А. А.* Теория вероятностей. М., 1986.

[4] *Колемаев В. А., Калинина В. Н.* Теория вероятностей и математическая статистика. М., Инфра-М, 1997.

На практических занятиях потребуется задачник:

[5] *Коршунов Д. А., Фосс С. Г.* Сборник задач и упражнений по теории вероятностей. Новосибирск, 1997.

Для самостоятельной работы можно использовать также задачники с ответами и указаниями:

[6] *Севастьянов Б. А., Чистяков В. П., Зубков А. М.* Сборник задач по теории вероятностей. М., 1986.

[7] *Вентцель Е. С., Овчаров Л. А.* Теория вероятностей (Избранные главы высшей математики для инженеров и студентов втузов, задачи и упражнения). М., 1973.