Содержание

Часть I

Теория вероятности

Ты спросишь меня, кого я люблю больше: тебя или теорию вероятности. Я отвечу, что почти наверное тебя и ты уйдешь, так и не узнав, что почти наверное значит с вероятностью 1.

1 Вероятностное пространство

Методы теории вероятности работают в ситуациях, называемых стохастическими. Для них характерны три свойства:

1. Непредсказуемость

2. Воспроизводимость

3. Устойчивость частот

Для описания стохастических ситуаций ситуаций необходимо определить функцию вероятности. Её область определения назывется множеством событий. В свою очередь событие (такое как, например, выпадание чётного числа на кубике) могут являться совокупностью неких более простых событий, описывающих стохастическую ситуацию (число, выпавшее на кубике). Последнее множество называется множетсвом элементарных исходов и обозначается Ω.

Множество событий, обозначаемое $\mathscr{A}$, должно обладать следующими интуитивными свойствами:

1. Отрицание события есть событие (если «пойдет дождь» событие, то «не пойдет дождь» также событие)

2. Объединение событий есть событие («пойдет дождь» или «пойдет снег»)

3. Все множество элементарных исходов является событие («Что-нибудь да произойдет»)

Формализуя эти свойства, получаем определение алгебры.

Определение 1.1 *Семейство $\mathscr{A}$ подмножеств множества Ω называется <u>алгеброй</u>, если*

1. $\forall A \in \mathscr{A}, B \in \mathscr{A} \Rightarrow A \bigcup B \in \mathscr{A}$

2. $\forall A \in \mathscr{A} \Rightarrow \overline{A} \in \mathscr{A}$

3. $\Omega \in \mathscr{A}$

Из аксиом алгебры и формулы $A \bigcap B = \overline{\overline{A} \bigcup \overline{B}}$ следует, что пересечений событий явялется событием.

Наименьшей возможной алгеброй является $\{\Omega, \varnothing\}$

Определение 1.2 *Семейство $\mathscr{A}$ подмножеств множества Ω называется <u>σ-алгеброй</u>, если*

1. $\forall A_1, \ldots, A_n, \ldots \in \mathscr{A} \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathscr{A}$

2. $\forall A \in \mathscr{A} \Rightarrow \overline{A} \in \mathscr{A}$

3. $\Omega \in \mathscr{A}$

Определение 1.3 *Пусть $\mathscr{K}$ - класс подмножеств Ω. σ-алгебра $\sigma(\mathscr{K})$, порожденная классом $\mathscr{K}$ — наименьшая σ-алгебра, содержащая $\mathscr{K}$, то есть любая σ-алгебра, содержащая $\mathscr{K}$, содержит и $\sigma(\mathscr{K})$.*

Пример σ-алгеброй, порожденной $\mathscr{K} = A$, будет являться $\sigma(A) = \left\{\varnothing, A, \bar(A), \Omega\right\}$.

σ-алгебра является более узким понятием, нежели алгебра, то есть любая σ-алгебра является алгеброй, а обратное, вообще говоря, неверно.

Пример Пусть $\Omega = \mathbb{R}$, $\mathscr{A}$ содержит конечные подмножества Ω и их дополнения. Для такого множества выполнены все аксиомы алгебры: $\Omega = \overline{\varnothing} \in \mathscr{A}$, объединение конечных множеств есть конечное множество, объединение конечного множества с дополнением к конечному множеству так же является дополнением к некоторому множеству. То же можно сказать и об объединении двух дополнений. Таким образом, $\mathscr{A}$ является алгеброй. Все элементы $\mathscr{A}$ либо конечны, либо континуальны, поэтому $\mathscr{A}$ не содержит $\mathbb{N}$. Но $\mathbb{N} = \bigcup\limits_{i=1}^{\infty}\{i\}$, то есть не выполнено свойство счетной аддитивности из определения σ-алгебры.

Определение 1.4 *Пара $(\Omega, \mathscr{A})$ называется измеримым пространством, если $\mathscr{A}$ является σ-алгеброй. Если же $\mathscr{A}$ - алгебра, то $(\Omega, \mathscr{A})$ — измеримое пространство в широком смысле.*

Определение 1.5 *Вероятностью называется функция $\mathbb{P}\colon \mathscr{A} \to \mathscr{R}_+$, удовлетворяющая свойстам*

1. $\forall A \in \mathscr{A} \quad \mathbb{P}(A) \geq 0$
2. $\forall A_1, \ldots, A_n, \ldots \in \mathscr{A} \quad A_i \bigcap A_j = \varnothing\ (i \neq j) \Rightarrow \mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} \mathbb{P}(A_i)$
3. $\mathbb{P}(\Omega) = 1$

Определение 1.6 *Вероятностным пространством $(\Omega, \mathscr{A}, \mathbb{P})$ называется измеримое пространство $(\Omega, \mathscr{A})$, снабженное вероятностью $\mathbb{P}$.*

 Кому вообще нужна σ-алгебра событий, и зачем весь этот огород, если можно рассматривать множество всех подмножеств множества Ω? Когда-то кто-то доказал, что в случае очень большого множества элементарных исходов, например, континуального, множество 2^Ω будет иметь такую крокодильски большую мощность, что вся теория сломается. Таким образом, алгебры нужны для того, чтобы вероятность имела хорошую область определения.

Свойства вероятности

1. $\mathbb{P}(\varnothing) = 0$
2. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
3. $A \subseteq B \quad \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
4. $\mathbb{P}(A) \leq 1$

5. $\mathbb{P}(A \bigcup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$

6. $\mathbb{P}(A \bigcup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

7. $\mathbb{P}(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{k=1}^{n} \sum\limits_{i_1 < \dots < i_k} (-1)^{k+1} \mathbb{P}(A_{i_1} A_{i_2} \dots A_{i_k})$

8. $\mathbb{P}(\bigcap\limits_{i=1}^{n} A_i) \geq 1 - \sum\limits_{i=1}^{n} \mathbb{P}(\overline{A_i})$ - неравенство Бонферрони

Второй пункт в определении вероятностной меры нельзя заменить аналогичным с конечными объединением и суммой. Однако если добавить к данному требованию так называемое свойство непрерывности вероятностной меры, т.е

$$\forall B_1, B_2, \ldots \in \mathscr{A} \quad B_{n+1} \subseteq B_n \Rightarrow \lim_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(B)$$

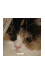
, то они вместе будут эквивалентны 2. из определения вероятности.

Утверждение 1.1 $\forall A_1, \ldots, A_n, \ldots \in \mathscr{A} \quad A_i \bigcap A_j = \varnothing \ (i \neq j) \Rightarrow \mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} A_i \Leftrightarrow (\forall A_1, \ldots, A_n \in \mathscr{A} \quad A_i \bigcap A_j = \varnothing \ (i \neq j) \Rightarrow \mathbb{P}(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} A_i) \wedge (\forall B_1, B_2, \ldots \in \mathscr{A} \quad B_{n+1} \subseteq B_n \Rightarrow \lim\limits_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(B))$

Доказательство
$\Rightarrow$
Обозначим $C_n = B_n \setminus B_n + 1$. Множества $B, C_1, C_2, \ldots$ не имеют общих точек.
$\forall n \quad B_n = \bigcup\limits_{k=n}^{\infty} C_k \bigcup B$. Тогда $\mathbb{P}(B_1) = \mathbb{P}(B) + \sum\limits_{k=1}^{\infty} \mathbb{P}(C_k)$. Отсюда следует, что ряд в правой части сходится, так как имеет конечную сумму. $\mathbb{P}(B_n) = \mathbb{P}(B) + \sum\limits_{k=n}^{\infty} \mathbb{P}(C_k)$. При $n \to \infty$ сумма ряда стремится к нулю как остаточный член ряда из предыдущего выражения. В предельном переходе получаем свойство непрерывности.
$\Leftarrow$
Рассмотрим произвольный набор $A_1, A_2, \ldots \in \mathscr{A} \quad A_i A_j = \varnothing$.
$\mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i) = \mathbb{P}(\bigcup\limits_{i=1}^{n} A_i) + \mathbb{P}(\bigcup\limits_{i=n+1}^{\infty} A_i) = \sum\limits_{i=1}^{n} \mathbb{P}(A_i) + \mathbb{P}(\bigcup\limits_{i=n+1}^{\infty} A_i)$.
Обозначим $B_n = \bigcup\limits_{i=n+1}^{\infty} A_i, \quad B_{n+1} \subseteq B_n \quad \forall n, \quad \bigcap\limits_{n=1}^{\infty} B_n = \varnothing$
$\sum\limits_{i=1}^{\infty} \mathbb{P}(A_i) = \lim\limits_{n \to \infty} (\mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i) - \mathbb{P}(B_n)) = \mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i) - \lim\limits_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i)$ ■

Теорема 1.1 (Каратеодори) *Пусть* $(\Omega, \mathscr{A})$ *— измеримое пространство в широком смысле, а некоторая функция* $\mathbb{P}$ *обладает свойствами вероятностной меры.Тогда на измеримом пространстве* $(\Omega, \sigma(\mathscr{A}))$ $\exists! \ \mathbb{P}' : \forall A \in \mathscr{A} \quad \mathbb{P}(A) = \mathbb{P}'(A)$

Доказательство отсутсвует ■

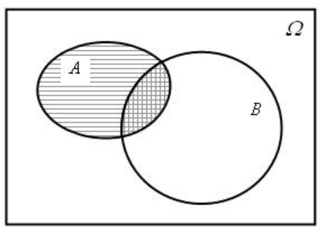
 Зачем это нужно? Теорема Каратеодори говорит о том, что любую вероятностную меру, заданную на алгебре, можно однозначно продолжить на σ-алгебру, то есть расширить область ее определения. При этом значения функции на алгебре не изменятся. Теорема будет использоваться при определении интеграла Лебега.

2 Условная вероятность. Независимость событий

Рассмотрим произвольное $B \in \mathscr{A}: \quad \mathbb{P}(B) > 0.$

Определение 2.1 *Условной вероятностью событии $A \in \mathscr{A}$ при условии B называется $\frac{\mathbb{P}(AB)}{\mathbb{P}(B)} =: \mathbb{P}(A|B) = \mathbb{P}_B(A)$*

Что это означает на пальцах? Условная вероятность $\mathbb{P}(A|B)$ — это веротяность того, что произойдет событие A, если мы точно знаем, что произошло событие B.



Графически это означает, что, когда произошло событие B, мы оказались в круге B. Тогда формула $\frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$ есть просто вероятность попасть в AB.

Из определения следует так называемый «Закон умножения вероятностей»:

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(AB)$$

Легко проверяется, что $(B, \mathscr{A}_B, \mathbb{P}_B)$, где $\mathscr{A}_B = \{A \bigcap B \colon A \in \mathscr{A}\}$, так же является вероятностным пространством.

Зачем нужно требование $\mathbb{P}(B) > 0$, если можно в случае $\mathbb{P}(B) = 0$ доопределить условную вероятность нулем как вероятность при условии невозможного события? При таком доопределении нарушится аксиома 3. вероятности $\mathbb{P}_B$, поскольку $\mathbb{P}_B(B)$ по доопределению будет равно 0.

Определение 2.2 *События $A, B \in \mathscr{A}$ называются независимыми, если*

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

.

Для независимых событий

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

.

Пример Являются ли несовместные события ($AB = \varnothing$) независимыми? Нет, пусть $A, B \in \mathscr{A}: \quad \mathbb{P}(A) > 0,\ \mathbb{P}(B) > 0$. Тогда $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) = 0$, что является противоречием. По-простому, если произошло одно из несовместных событий, то второе уже не может произойти, и его условная веротяность равна 0, а не вероятности самого события, что требуется для независимости.

Следующее определение обобщает понятие независимости на произвольное количество событий.

Определение 2.3 *События $A_1, A_2, \ldots, A_n$ называются независимыми в совокупности, если*

$$\forall m = 2, \ldots, n \quad \forall 1 \leq j_1 < \ldots < j_m \leq n \quad \mathbb{P}(\bigcap_{k=1}^{m} A_{j_k}) = \prod_{k=1}^{m} \mathbb{P}(A_{j_k})$$

Пример На примере тетраэдра Бернштейна можно убедиться в том, что попарной независимости событий недостаточно для независимости в совокупности. Рассмотри тетраэдр, у которого три стороны покрашены в красный, синий и зеленый, а четвертая содержит все три цвета. События $\{$выпадет красный$\}=\{$К$\}$, $\{$выпадет синий$\}=\{$С$\}$, $\{$выпадет зеленый$\}=\{$З$\}$ попарно независимы (например, вероятность события $\{\text{С}\}\bigcap\{\text{К}\}$ равна веротяности выпадения четвертой грани, т. е. $\frac{1}{4}$, в то время как выпадения каждого цвета равна $\frac{1}{2}$). Однако $\mathbb{P}(\{\text{С}\}\bigcap\{\text{К}\}\bigcap\{\text{З}\}) = \frac{1}{4} \neq (\frac{1}{2})^3$.

3 Формула полной веротяности. Формула Байеса

Определение 3.1 *$B_1, \ldots, B_n$ образуют <u>полную группу</u>, если выполнены следующие условия:*

1. $\mathbb{P}(B_i) > 0 \quad \forall i = 1, \ldots, n$
2. $B_i B_j = \varnothing \quad (i \neq j)$
3. $\bigcup\limits_{i=1}^{n} B_i = \Omega$

Теорема 3.1 *Пусть $B_1, \ldots, B_n$ образуют полную группу. Вероятность события $A \in \mathscr{A}$ можно вычислить по <u>формуле полной вероятности</u>:*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Доказательство
$A = \bigcup\limits_{i=1}^{n} AB_i, \quad AB_i \bigcap AB_j = \varnothing \quad (i \neq j)$
$\mathbb{P}(A) = \sum\limits_{i=1}^{n} \mathbb{P}(AB_i) = \sum\limits_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$
Последний переход следует из закона умножения вероятностей. ∎

Первое требование определения полной группы необходимо для возможности определить условную веротяность, второе позволяет разбить множество A на непересекающиеся части. Третье требование, вообще говоря, можно ослабить, потребовав, чтобы $A \subseteq \bigcup\limits_{i=1}^{n} B_i$. Доказательство при этом не изменится.

Пример Проиллюстрировать формулу полной вероятности можно обычным экзаменом: A - {студент сдал экзамен}, B_i - {студент попал к преподавателю i}. Как и в любой другой лотерее, можно оценить вероятность попадания к преподавателю i, то есть $\mathbb{P}(B_i)$, а трезво оценивая свои силы можно прикинуть и вероятность сдать тому или иному преподавателю $\mathbb{P}(A|B_i)$. Зная все вышеперечисленое, несложно по формуле вычислить вероятность успешной сдачи.

Формула полной веротяности используется для вычисления априорной вероятности, т.е. вероятности события, которое еще не произошло. Пусть теперь $\mathbb{P}(A) > 0$. Тогда $\mathbb{P}(B_i|A) = \frac{\mathbb{P}(AB_i)}{\mathbb{P}(A)}$. Используя формулу полной вероятности, получаем <u>формулу Байеса</u>:

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum\limits_{j=1}^{n} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

Формула Байеса используется для вычисления апостериорной вероятности. То есть уже известно, что произошло некоторое событие A, и нужно вычислить вероятность того, что произошло некоторое B_i. В примере с экзаменом, например, может быть известно, что студент не сдал экзамен, и хочется вычислить вероятность того, что он сдавал преподавателю «Р».

4 Случайные величины

Случайные события — это хорошо, но с события типа «на монетке выпал герб» плохо формализуемы, а мы хотим формальности и математичности. Поэтому вместо всяких событий хочется работать с числами. Вот этим и займемся. При рассмотрении случайных событий мы ввели вероятностное пространство, которе выглядит так:

$$(\Omega, \mathscr{A}, \mathbb{P}),$$

где Ω — множество элементарных событий, $\mathscr{A}$ — σ-алгебра подмножеств множества элементарных событий, а $\mathbb{P}$ — вероятность. Мы же будем рассамтривать теперь тройку

$$(\mathbb{R}, \mathscr{B}, \mathbb{P}),$$

где $\mathbb{R}$ — действительная прямая, $\mathscr{B}$ — борелевская σ-алгебра, а $\mathbb{P}$ — вероятность. Поясним.

Определение 4.1 *Борелевской σ-алгеброй называется минимальная σ-алгебра, содержащая все открытые подмножества топологического пространства. Элементы борелевской σ-алгебры называются борелевскими множествами.*

Мы будем рассматривать только топологическое пространство $\mathbb{R}$, так что это стремное словосочетание можно прямо сейчас забыть и понимать открытое множество как открытое множество из матана (все точки внутренние).
Пример Покажем, что все «хорошие» множества являются Борелевскими.

1. Все открытые интревалы входят по определению.

2. Отрезок вида $[a, b]$ входит как $\overline{(-\infty, a) \bigcup (b, +\infty)}$.

3. Точка ходит как вырожденный отрезок $[a, a]$.

4. Счетное объединение таких множеств входит по поределению.

Теперь формально введем понятие случайной величины (может использоваться сокращение с.в.).

Определение 4.2 *Пусть $(\Omega, \mathscr{A}, \mathbb{P})$ — вероятностное пространство.*
Тогда случайной величиной ξ называется функция $\xi : \Omega \to \mathbb{R}$, измеримая относительно $\mathscr{A}$ и $\mathscr{B}$. По-другому, ξ — случайная величина, если

$$\forall B \in \mathscr{B} \quad \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathscr{A}$$

.

Таким финтом ушами мы, по сути, сопоставили каждому событию какое-то «хорошее» множество на числовой прямой, и можем рассматривать не вероятности событий, а вероятности попадания в эти «хорошие» подмножества числовой прямой.

Введем еще несколько бесполезных определений, которые в дальнейшем использоваться не будут, но знать их не вредно.

Определение 4.3 *С каждой случайной величиной свяжем два вероятностных пространства: первое — $(\Omega, \mathscr{A}_\xi, \mathbb{P})$ — вероятностное пространство, порожденное ξ. Здесь $\mathscr{A}_\xi$ - наименьшая σ-алгебра, для которой выполняется свойство измеримости. Второе — $(\mathbb{R}, \mathscr{B}, \mathbb{P}_\xi)$, где $\mathbb{P}_\xi(B) = \mathbb{P}(\xi^{-1}(B)) \quad \forall B \in \mathscr{B}$ и называется распределением вероятностей ξ.*

Идем дальше в сторону упрощения работы со случайностями. Вместо того чтобы рассматривать произвольные борелевские множества, мы будем рассамтривать только множества вида $(-\infty, x)$. Действительно, интервал (a, b) получается из полупрямых так: $(a, b) = (-\infty, b) \setminus (-\infty, a]$ Таким образом, мы можем рассматривать случайные величины только на таких множествах. Здесь имеется в виду, что для удовлетворения определению случайной величины достаточно измеримости только на полупрямых, что следует из следующих свойств полного прообраза: прообраз объединения есть объединение прообразов, прообраз пересечения есть пересечение прообразов, прообраз отрицания есть отрицание прообраза. Выше показано, что из полупрямых с помощью этих операций можно получить интервалы, а из интервалов и все $\mathscr{B}$.

Теперь несколько полезных утверждений. Пусть ξ — случайная величина. Тогда $-\xi$ также случайная величина, так как её прообраз от любой полупрямой является прообразом ξ от симметричной полупрямой, то есть лежит в $\mathscr{A}$. $\xi + c$ также будет случайной величиной, поскольку ее прообразом для любой полупрямой будет прообраз ξ для полупрямой, сдвинутой на c, то есть будет лежать в $\mathscr{A}$.

Утверждение 4.1 *Пусть ξ, η — случайные величины. Тогда множество $\{\omega \colon \xi(\omega) < \eta(\omega)\}$ является событием.*

Доказательство $\{\omega \colon \xi(\omega) < \eta(\omega)\} = \bigcup\limits_{r \in \mathbb{Q}} \{\omega \colon \xi(\omega) < r, \eta(\omega) > r\}$. Заметим, что $\{\omega \colon \xi(\omega) < r)\}$ является событием. Аналогично для η. Выражение, написанное выше, является счетным объединением пересечений двух событий, то есть событием. ■

Похожими махинациями, а также с использованием этого утверждения, доказывается, что $\xi^2, \xi + \eta, \xi\eta$ являются случайными величинами. Более того, если $\xi_1, \ldots, \xi_n$ — с.в., а функция $\phi(x_1, \ldots, x_n)$ является непрерывной на множестве их значений, то $\phi(\xi_1, \ldots, \xi_n)$ будет случайной величиной. Это не доказывалось.

Определение 4.4 *Рассмотрим вероятностное пространство $(\Omega, \mathscr{A}, \mathbb{P})$ и определенную на нем случайную величину ξ. Тогда её функцией распределения $F_\xi(x)$ называется функция $F_\xi : \mathbb{R} \to \mathbb{R}$*

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

.

Запись $\mathbb{P}(\xi < x)$ является в некотором смысле жаргонной, так как аргументов вероятности должно быть событие из $\mathscr{A}$. Но $\xi < x$ мы в дальнейшем будем отождествлять с оъединением элементарных событий, образ которых меньше x. Из определения случайной величины получаем, что это объединение является событием, поэтому применение к нему функции вероятности корректно.

Функция распределения (сокращение ф.р.) является очень полезной штукой, поскольку имеет достаточно простой вид и несет в себе всю информацию о распределении, то есть однозначно определяет $\mathbb{P}_\xi$.

Рассмотрим основные свойства функции распределения:

1. $F_\xi(x) \in [0, 1]$
2. $\lim\limits_{x\to-\infty} F_\xi(x) = 0$
3. $\lim\limits_{x\to+\infty} F_\xi(x) = 1$
4. $F_\xi(x)$ монотонно не убывает.
5. $F_\xi(x)$ непрерывна слева.

Вероятность попадания с.в. в полуинтервал $\mathbb{P}_\xi[a, b) = F_\xi(b) - F_\xi(a)$. При стремлении $b \to a$ получим $\mathbb{P}(\xi = a) = F_\xi(a+) - F_\xi(a)$, то есть вероятность попадания в точку равна скачку функции распределения в этой точке.

Определение 4.5 *Точка x_0 называется <u>точкой роста</u> $F_\xi(x)$, если $\forall\epsilon > 0 \quad \mathbb{P}(x_0 - \epsilon \le \xi < x_0 + \epsilon) > 0$*

Пример Это очень полезный пример, который будет использоваться в матстате и который очень любят спрашивать. Пусть ξ — случайная величина. $\eta = F_\xi(\xi)$. Чему равна $F_\eta(x)$? По определению $F_\eta(x) = \mathbb{P}(\eta < x) = \mathbb{P}(F_\xi(\xi) < x) = \mathbb{P}(\xi < F_\xi^{-1}(x)) = F_\xi(F_\xi^{-1}(x)) = x$. Вообще, тут было бы неплохо сказать, что F_ξ непрерывна и строго монотонна, чтобы со спокойной совестью использовать обратную функцию. Таким образом η имеет равномерное распределение.

5 Понятие меры и интеграла Лебега

Мера Лебега вводится потихонечку: сначала для полуинтервала, потом для борелевских множеств, а затем и для всей числовой прямой. Основная идея — обобщение понятия длины на страшные, ненормальные множества.

Вспомним сначала аксиомы меры μ из функана (хех): это функция на множестве B (в нашем случае $[0,1)$), $\mu\colon B \to \mathbb{R}$ со свойствами

1. $\forall A \in B \quad \mu(A) \geq 0$
2. $\forall A_1, A_2, \ldots \in B\colon A_iA_j = \varnothing \quad (i \neq j) \Rightarrow \mu(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} \mu(A_i)$

Введем сначала меру на борелевских множествах полуинтервала $[0,1)$, то есть на $\mathscr{B}_{[0,1)}$. Итак,

- На полуинтервалах (a,b) введем меру как $\mu_{[0,1)}((a,b)) = b - a$
- Тогда мера одной точки равна нулю, и мы можем не обращать внимание на концы множеств
- Любое конечное объединение, пересечение, отрицание таких интервалов есть конечное объединение интервалов (и, возможно, конечное число точек-концов) На них введем меру как сумму мер этих интервалов.
- Теперь осталось определить меру на бесконечных объединениях/пересечениях. Для этого воспользуемся теоремой Каратеодори, согласно которой можно продолжить меру на σ-алгебре.

Аналогичным образом определим меру $\mu_{[k,k+1)}$ на всех полуинтервалах вида $[k, k+1) \quad (k \in \mathbb{Z})$. Осталось доопределить меру на всей $\mathscr{B}$. $\forall A \in \mathscr{B}$

$$\mu(A) = \sum_{i=-\infty}^{\infty} \mu_{[i,i+1)}(A \bigcap [i, i+1))$$

Теперь введем интеграл интеграл Лебега (интеграл по мере Лебега). В отличие от интеграла Римана, где происходит разбиение области определения и выбирается значение функции из образа элемента разбиения, в интеграле Лебега разбивается область значений (то есть Oy), и некоторое значение из элемента разбиения умножается на меру прообраза этого элемента, затем все благополучно складывается. Введем теперь это формально.

Будем рассматривать функции $f\colon \mathbb{R} \to \mathbb{R}$ такие, что $\forall c \in \mathbb{R} \quad \{x\colon f(x) < c\} \in \mathscr{B}$, то есть прообразы полупрямых являются измеримыми множествами. А как было показано выше, отсюда следует, что прообразы всех борелевских множеств являются измеримыми.

Введем сначала понятие индикатора. Индикатор события A - это случайная величина, принимающая значение 1, если событие произошло, и 0 в противном случае. Таким образом

$$\mathbb{1}(x) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}$$

Определим интеграл Лебега на полуинтервале $[0,1)$

- Пусть функция имеет вид $f(x) = \sum\limits_{i=1}^{k} y_i \mathbb{1}(A_i) \quad (i \neq j \Rightarrow A_i A_j = \varnothing,\ \bigcup_{i=1}^{k} A_i = [0,1))$. Функции такого вида называются <u>финитными</u>. В этом случае

$$\int\limits_0^1 f(x)\mu(dx) = \sum_{i=1}^{k} y_i \mu(A_i)$$

  То есть берется мера кусочка, на котором функция принимает значение y_i, и умножается на это значение. Получается нечто, напоминающее площадь под графиком.

- Пусть теперь $f(x) \geq 0$ на $[0,1)$. Будем приближать функцию финитной. Возьмем отрезок области значений $[0,n], n \in \mathbb{N}$. Разобьем этот отрезок на $n2^n$ кусочков $[\frac{k-1}{2^n}, \frac{k}{2^n}], k = 1, 2, \ldots, n2^n$. На каждом кусочке скажем, что значение функции равно $\frac{k-1}{2^n}$. Осталось приблизить $[n, +\infty)$. Будем считать это одной частью, на которой функция принимает значение n. Осталось устремить n к бесконечности: тогда, так как размер каждого кусочка первой части равен $\frac{1}{2^n}$, их длина станет бесконечно малой, а та часть, которую мы «обрубили» сверху (то есть $[n, +\infty)$), уйдет в бесконечность. Итак,

$$\int\limits_0^1 f(x)\mu(dx) = \lim_{n\to\infty}\left[\sum_{k=1}^{n2^n} \frac{k-1}{2^n}\mu\left(\left\{x\colon \frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}\right\}\right) + n\mu(\{x\colon f(x) \geq n\})\right]$$

- Остался случай произвольной функции $f(x)$. Разобьем ее на две: одна функция совпадает с $f(x)$ там, где та положительна, и равняется 0 в остальных случаях, другая равна $|f(x)|$ там, где $f(x)$ отрицательна, и 0 иначе.

$$f^+(x) = \max\{0, f(x)\}, \quad f^-(x) = -\min\{0, f(x)\}$$

  Для каждой из этих функций интеграл определен по предыдущему пункту. Пользуясь тем, что $f(x) = f^+(x) - f^-(x)$, определим интеграл так:

$$\int\limits_0^1 f(x)\mu(dx) = \int\limits_0^1 f^+(x)\mu(dx) - \int\limits_0^1 f^-(x)\mu(dx)$$

  В случае, если оба интеграла в правой части расходятся, интеграл от $f(x)$ не определен. Так как $|f(x)| = f^+(x) + f^-(x)$, сходимость интеграла Лебега от модуля функции (абсолютная сходимость) эквивалента сходимости интеграла от самой функции (то есть обычной сходимости). Это следует из того, что интеграл сходится только в случае конечности обоих интегралов правой части (иначе он не определен), откуда следует конечность их суммы и разности. Таким образом, для интеграла Лебега не существует условно сходящихся функций.

Аналогично вводим интеграл на каждом полуинтервале $[i, i+1), \quad i \in \mathbb{Z}$. Тогда на всей прямой интеграл по множеству $A \subset \mathbb{R}$ будет определяться так:

$$\int\limits_A f(x)\mu(dx) = \sum_{i=-\infty}^{+\infty} \int\limits_{A\bigcap[i,i+1)} f(x)\mu(dx)$$

Пример С помощью интеграла Лебега можно считать интегралы от функций, об интегрировании которых раньше было страшно даже подумать. Например, от функции Дирихле:

$$D_{[0,1)}(x) = \begin{cases} 1, x \in [0,1) \setminus \mathbb{Q} \\ 0, x \in \mathbb{Q} \end{cases}$$

Данная функция является финитной, а именно $D_{[0,1)}(x) = \mathbb{1}([0,1) \setminus \mathbb{Q})$. Поэтому по первому пунккту

$$\int\limits_0^1 D_{[0,1)}(x)\mu(dx) = 1$$

В дальнейшем $\mu(dx)$ будет опускаться обозначаться просто как dx или dy чтобы подчеркнуть, что считается именно интеграл Лебега.

6 Виды распределений

Распределения случайных величин можно разделить на 3 типа: непрерывные, дискретные и сингулярные.

Определение 6.1 *Случайная величина ξ называется <u>абсолютно непрерывной</u>, если существует интегрируемая функция $p_\xi(x) \geq 0, \ x \in \mathbb{R}$ такая, что функция распределения ξ является почти всюду (за исключением не более, чем счетного числа точек) дифференцируемой функцией и представима в виде*

$$F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(y)dy$$

Отсюда следует, что функция распределения непрерывна на $\mathbb{R}$. $p_\xi(x)$ называется <u>плотностью распределения</u>, и почти всюду выполнено $p_\xi(x) = F'_\xi(x)$. Плотность, вообще говоря, определена не однозначно.

Определение 6.2 *Случайная величина ξ называется <u>дискретной</u>, если множество точек роста не более, чем счетно, но распределение не является сингулярным, или, другими словами $\exists B = \{x_1, x_2, \ldots\} \colon \mathbb{P}(\xi \in B) = 1$.*

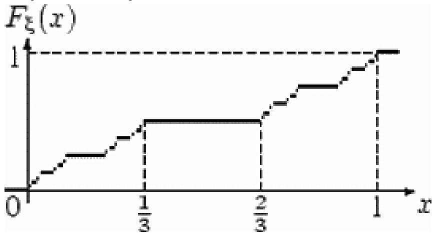
Определение 6.3 *Случайная величина ξ называется <u>сингулярной</u>, если F_ξ непрерына, и $\exists B \in \mathscr{B} \colon \mu(B) = 0, \ \mathbb{P}(\xi \in B) = 1$, то есть множество значений случайной величины имеет меру 0, но вероятность попасть в каждую точку этого множества так же нулевая.*

Пара слов о жизненном смысле определений: непрерывная случайная величина имеет областью значений континуальное множество, при этом вероятность попасть в отдельно взятую точку нулевая. Пример: равномерное распределение по отрезку. Плотность же отражает вероятность попасть в ту или иную область: интеграл по области равен этой вероятности. Дискретная случайная величина принимает конечное или счетное множество значений, вследствие этого имеет ступенчатую функцию распределения, например, бросок монетки имеет дискретное распределение. Сингулярное распределение — это крокодил, который не встречается в жизни и будет рассмотрен отдельно.

Утверждение 6.1 *Дискретная случайная величина имеет не более, чем счетное число скачков.*

Доказательство Из свойств функции распределения следует, что дискретная величина имеет не больше двух скачков величины больше $\frac{1}{2}$. Аналогично, скачков величины больше $\frac{1}{3}$ не больше 3. То есть скачков величины больше $\frac{1}{n}$ не более n. Для любого скачка можно указать $n \in \mathbb{N}$ такое, что величина, этого скачка больше $\frac{1}{n}$. Значит,каждому скачку можно поставить в соответствие n, множество которых счетно. При этом для каждого n существует не более чем счетное число скачков, ему соответсвующих (величины $> \frac{1}{n}$). А так как объединение не более, чем счетного числа не более, чем счетных множеств, не более, чем счетно, получаем требуемое. ■

Пример Для полного счастья приведем пример сингулярной случайной величины. Пусть функция распределения - так называемая лестница Кантора (см. рисунок).



Посчитаем меру множества, на котором функция константа, то есть точки этого множества не будут точками роста: сначала это одна ступенька длины 1/3, потом две длины 1/9, и т.д.

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} = \frac{1}{3}\sum_{k=1}^{\infty}(\frac{2}{3})^{k-1} = 1$$

Тогда множество точек роста имеет меру 0 в силу свойства аддитивности меры.

Вообще говоря, существуют менее изысканные примеры сингулярных распределений. Например, при стрельбе из лука в круглую мишень распределение будет сингулярным, если стрелок попадает только в точки одной прямой. В самом деле, двумерная мера прямой равна 0, как и вероятность попасть в каждую отдельную точку.

Теорема 6.1 (Лебега) *Любую случайную величину можно представить в виде суммы дискретной, абсолютно непрерывной и сингулярной случайной величины. То есть*

$$F(x) = \alpha_d F_d(x) + \alpha_c F_c(x) + \alpha_s F_s(x), \quad \alpha_d + \alpha_c + \alpha_s = 1$$

Доказательство вышло и не вернулось ■

7 Характеристики случайных величин

Математическое ожидание (обозначается $\mathbb{E}$) обобщает понятие среднего арфиметического для произвольной случайной величины и показывает, какие значения в среднем принимает случайная величина. Оно, как и интеграл Лебега, вводится в несколько этапов. В этом определении вероятность $\mathbb{P}$ играет роль Лебеговой меры μ.

- Если $\xi(\omega) = \sum\limits_{i=1}^{k} x_i \mathbb{1}(A_i) \quad (i \neq j \Rightarrow A_i A_j = \varnothing,\ \bigcup_{i=1}^{k} A_i = \Omega)$.

$$\mathbb{E}\xi = \int\limits_{\Omega} \xi(\omega)\mathbb{P}(d\omega) \equiv \sum_{i=1}^{k} x_i \mathbb{P}(\{\omega\colon \xi(\omega) = x_i\}) = \sum_{i=1}^{k} x_i \mathbb{P}(A_i)$$

- $\xi(\omega) \geq 0$. В этом случае аналогично интегралу Лебега

$$\mathbb{E}\xi = \int\limits_{\Omega} \xi d\mathbb{P} = \lim_{n\to\infty}\left[\sum_{k=1}^{n2^n} \frac{k-1}{2^n}\mathbb{P}\left(\left\{\omega\colon \frac{k-1}{2^n} \leq \xi(\omega) < \frac{k}{2^n}\right\}\right) + n\mathbb{P}(\{\omega\colon \xi(\omega) \geq n\})\right]$$

- Для произвольной $\xi(\omega)$ вводятся

$$\xi^{+}(\omega) = \max\{0, \xi(\omega)\}, \quad \xi^{-}(\omega) = -\min\{0, \xi(\omega)\}$$

$$\mathbb{E}\xi = \mathbb{E}\xi^{+} - \mathbb{E}\xi^{-}$$

Вспомним, что любая случайная величина ξ индуцирует вероятностное пространство $(\mathbb{R}, \mathscr{B}_\xi, \mathbb{P}_\xi)$. Тогда, поскольку $\mathbb{P}_\xi(dx)$ есть вероятность попасть в dx, выразим ее через функцию распределения $\mathbb{P}_\xi(dx) = F_\xi(x+dx) - F_\xi(x) = dF_\xi(x)$ Тогда перепишем матожидание в более привычной форме

$$\mathbb{E}\xi = \int\limits_{\Omega} \xi(\omega)\mathbb{P}(d\omega) = \int\limits_{-\infty}^{+\infty} x\mathbb{P}_\xi(dx) = \int\limits_{-\infty}^{+\infty} x dF_\xi(x)$$

Перечислим некоторые свойства математического ожидания:

1. $\mathbb{E}(\xi + a) = \mathbb{E}\xi + a \quad \forall a \in \mathbb{R}$
2. $\mathbb{E}(a\xi) = a\mathbb{E}\xi \quad \forall a \in \mathbb{R}$
3. $\mathbb{E}(\xi + \eta) = \mathbb{E}\xi + \mathbb{E}\eta$ (здесь подразумевается, что существуют два из трех математических ожидания, из чего следует существование третьего)

Пример Рассмотрим дискретную случайную величину ξ, которая принимает значения n с вероятностью $\frac{c}{n^2}$ $(c = \frac{6}{\pi^2})$. По определению

$$\mathbb{E}\xi = \sum_{n=1}^{\infty} n\frac{c}{n^2}$$

Данный ряд, очевидно, расходится. Сиутацию не спасет даже рассмотрение случайной величины η, принимающей значения $\pm n$ с вероятностью $\frac{c}{2n^2}$, которая имеет среднее значение 0, поскольку $\mathbb{E}\eta = \sum\limits_{n=1}^{\infty}\frac{c}{2n} - \sum\limits_{n=1}^{\infty}\frac{c}{2n}$, что не определено, поскольку интегралы Лебега от η^+ и η^- расходятся.

Пример Другим примером является распределение Коши с плотностью $p_\xi(x) = \frac{1}{\pi(1+x^2)}$. График этой функции симметричен относительно 0 и похож на горку, из чего методом пристального взгляда можно сделать вывод, что средним значением должно быть 0. Однако $\int\limits_{-\infty}^{+\infty} xdF_\xi(x) = \int\limits_{-\infty}^{+\infty} xp_\xi(x)dx$ расходится, поэтому математического ожидания не существует.

Определение 7.1 *Моментом порядка k случайной величины ξ называется $\mathbb{E}\xi^k$*
Абсолютным моментом порядка k случайной величины ξ называется $\mathbb{E}|\xi|^k$
Центральным моментом порядка k случайной величины ξ называется $\mathbb{E}(\xi - \mathbb{E}\xi)^k$

Определение 7.2 *Квантилью случайной величины ξ порядка q называется величина $l_\xi(q)$:*

$$l_\xi(q)\colon \begin{cases} \mathbb{P}(\xi \leq l_\xi(q)) \geq q \\ \mathbb{P}(\xi \geq l_\xi(q) \geq 1-q \end{cases}$$

В случае $q = \frac{1}{2}$ квантиль называется медианой и обозначается $\operatorname{med}\xi$.

Если $q = \frac{1}{4}$, то l_ξ называется квартилью, если $q = \frac{1}{10}$, децилью, а если $q = \frac{1}{100}$ — перцентилью.

Жизненный смысл медианы заключается в том, что она является точкой, вероятность попасть левее которой равна пероятности попасть правее нее. Аналогично можно сказать про квантиль любого порядка. Квантиль определена не единственным образом: пусть ξ принимает значения $\{0, 1\}$ с вероятностями $\frac{1}{2}$. Тогда медианой ξ может быть любая точка из отрезка $[0, 1]$.

Определение 7.3 *Интерквантильный размах — величина $R_\xi = l_\xi(\frac{3}{4}) - l_\xi(\frac{1}{4})$ — длина отрезка, вероятность попасть в который равна $\frac{1}{2}$.*

С матожиданием и медианой связана задача о «деловых людях». (здесь будет ссылка)

Определение 7.4 *Мода случайной величины ξ — это наиболее вероятное значение случайной величины. Оюозначается $\operatorname{mod}\xi$.*

При наблюдении случайной величины важно знать не только её среднее значение (матожидание), но и то, как сильно она от него отклоняется (например, измерение линейнкой в среднем дает правильный результат, однаком необходимо знать погрешность измерения). В связи с этим вводится понятие дисперсии.

Определение 7.5 *Пусть для случайной величины ξ существуют конечный $\mathbb{E}\xi$ и $\mathbb{E}\xi^2$. Дисперсией назыветтся величина, равная*

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

Эту формулу можно привести к более простому для вычисления виду:

$$\mathbb{D}\xi = \mathbb{E}(\xi^2 - 2\xi\mathbb{E}\xi + (\mathbb{E}\xi)^2) = \mathbb{E}\xi^2 - 2\mathbb{E}\xi\mathbb{E}\xi - (\mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$$

Перечислим некоторые свойства дисперсии. ξ, η - случайные величины, $c \in \mathbb{R}$.

1. $\mathbb{D}\xi \geq 0$ как матожидание от неотрицательной функции
2. $\mathbb{D}c\xi = c^2\mathbb{D}\xi$ — следует из определения и линейности матожидания.
3. $\mathbb{D}(\xi + c) = \mathbb{E}(\xi + c - \mathbb{E}(\xi + c))^2 = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{D}\xi$
4. $\mathbb{P}(\xi = c) = 1 \Leftrightarrow \mathbb{D}\xi = 0$ — отклонение равно нулю для константы
5. $\mathbb{D}(\xi + \eta) = \mathbb{E}(\xi + \eta)^2 - (\mathbb{E}(\xi + \eta))^2 = \mathbb{E}(\xi^2 + 2\xi\eta + \eta^2) - (\mathbb{E}\xi)^2 - 2\mathbb{E}\xi\mathbb{E}\eta - (\mathbb{E}\eta)^2 = \mathbb{D}\xi + \mathbb{D}\eta + 2(\mathbb{E}\xi\eta - \mathbb{E}\xi\mathbb{E}\eta)$

Определение 7.6 *Ковариация двух случайных величин — это величина, равная*

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)$$

Ковариация положительна, если случайные величины одновременно отклоняются в одну сторону и отрицательная, если в разные. Формулу ковариации так же можно упростить, раскрыв скобки в определении:

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}\xi\eta - \mathbb{E}\xi\mathbb{E}\eta$$

Таким образом пятое свойство дисперсии можно переписать так:

5. $\mathbb{D}(\xi \pm \eta) = \mathbb{D}\xi + \mathbb{D}\eta \pm 2\operatorname{cov}(\xi, \eta)$

Определение 7.7 *Среднеквадратическое отклонение — величина $\sigma = \sqrt{\mathbb{D}\xi}$.*

8 Независимость случайных величин

Определение 8.1 *Случайные величины ξ, η называются <u>независимыми</u>, если*

$$\forall B_1, B_2 \in \mathscr{B} \quad \mathbb{P}(\xi \in B_1, \eta \in B_2) = \mathbb{P}(\xi \in B_1)\mathbb{P}(\eta \in B_2)$$

Утверждение 8.1 *Для независимых случайных величин ξ, η выполнено*

$$\mathbb{E}\xi\eta = \mathbb{E}\xi\mathbb{E}\eta$$

Доказательство проведем только для случая дискреьных случайных величин. Пусть

$$\xi = \begin{cases} x_1, x_2, \dots \\ p_1, p_2, \dots \end{cases}$$

То есть ξ приминмает значение x_i с вероятностью p_i.

$$\eta = \begin{cases} y_1, y_2, \dots \\ q_1, q_2, \dots \end{cases}$$

$$\mathbb{E}\xi\eta = \sum_{i,j} x_i y_j \mathbb{P}(\xi = x_i, \eta = y_j) = \sum_{i,j} x_i y_j \mathbb{P}(\xi = x_i)\mathbb{P}(\eta = y_j) = \sum_i x_i \mathbb{P}(\xi = x_i) \sum_j y_j \mathbb{P}(\eta = y_j) = \mathbb{E}\xi\mathbb{E}\eta$$

■

Для независимых случайных величин

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}\xi\eta - \mathbb{E}\xi\mathbb{E}\eta = 0$$

Обратное, вообще говоря, неверно: пусть мы равновероятно выбираем одну из точек $(-1, 0), (0, 1), (1, 0), (0, -1)$. Каждая координата принимает значения $-1, 0, 1$, но координаты зависимы, так как $\mathbb{P}(x = 0, y = 0) = 0$ (никогда не выбираем (0,0)), а $\mathbb{P}(x = 0)\mathbb{P}(y = 0) = \frac{1}{2}\frac{1}{2} \neq 0$. Однако

$$\mathbb{E}x = \mathbb{E}y = \mathbb{E}xy = 0$$

в силу симметрии задачи. Отсюда $\operatorname{cov}(x, y) = 0$.

Таким образом коварияция показывает зависимость величин, однако не дает представления, насколько они зависимы. Для это вводится понятие коэффициента корреляции —- нормированная ковариация.

Определение 8.2 *<u>Коэффициент корреляции</u> — величина, описываемая формулой*

$$\rho(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathbb{D}\xi\mathbb{D}\eta}}$$

Коэффициент корреляции является псевдоскалярным произведением, то есть выполнены все аксиомы скалярного произведения, кроме половины четвертой. Поэтому для нее выполнено неравенство Коши-Буняковского.

Некоторые свойства коэффициента корреляции:

1. $|\rho(\xi, \eta)| \leq 1$ — неравенство Коши-Буняковского
2. $|\rho(\xi, \eta)| = 1 \Leftrightarrow \exists a, b\colon \quad \mathbb{P}(\xi = a\eta + b) = 1$ — равенство достигается, если случайные величины линейно зависимы
3. для независимых случайных величин $\rho(\xi, \eta) = 0$

9 Вероятностные неравенства

Лемма 9.1 *Для любой неотрицательной неубывающей функции* $g(x)$ *выполнено неравенство*

$$\mathbb{P}(|\xi| > x) \leq \frac{\mathbb{E}g(|\xi|)}{g(x)}$$

Доказательство

$$\mathbb{E}g(|\xi|) = \mathbb{E}g(|\xi|)\mathbb{1}(|\xi| \geq x) + \mathbb{E}g(|\xi|)\mathbb{1}(|\xi| < x) \geq \mathbb{E}g(|\xi|)\mathbb{1}(|\xi| \geq x) \geq g(x)\mathbb{E}\mathbb{1}(|\xi| \geq x) = g(x)\mathbb{P}(|\xi| > x)$$

Здесь мы воспользовались представлением $1 = \mathbb{1}(A) + \mathbb{1}(\overline{A})$, затем неотрицательностью функции и, следовательно, ее матожидания. Далее использовалась монотонность функции, и в последнем переходе тождество $\mathbb{E}\mathbb{1}(A) = 1\mathbb{P}(A) + 0\mathbb{P}(\overline{A}) = \mathbb{P}(A)$ ■

Из этой леммы слкдуют два полезных неравенства.

Теорема 9.1 (неравенство Маркова) *Для любой случайной величины* ξ*, имеющей конечное* $\mathbb{E}|\xi|$*, выполнено*

$$\mathbb{P}(|\xi| > \epsilon) \leq \frac{\mathbb{E}|\xi|}{\epsilon}$$

Доказательство
В неравенстве леммы возьмем $g(x) = x$. ■

Теорема 9.2 (неравенство Чебышева) *Для любой случайной величины* ξ*, имеющей конечный первый и второй момент, выполнено*

$$\mathbb{P}(|\xi - \mathbb{E}\xi| > \epsilon) \leq \frac{\mathbb{D}\xi}{\epsilon^2}$$

Доказательство
В неравенстве леммы возьмем $g(x) = x^2$. ■

10 Испытания Бернулли

Определение 10.1 *Дискретная случайная величина ξ имеет распределение Бернулли, если*

$$\mathbb{P}(\xi = x_1) = p,\ \mathbb{P}(\xi = x_2) = q = 1 - p, \quad x_1 \neq x_2$$

Определение 10.2 *Схема Бернулли — последовательность испытаний, удовлетворяющих следующим условиям:*

1. *Дихотомичность — у каждого испытания два исхода, называемые «успехом» и «неудачей» или, сокращенно У/Н.*
2. *Независимость — результаты испытаний являются независимыми событиями.*
3. *Однородность — вероятности успеха в каждом испытании равны.*

Из определения следует, что одно испытание имеет распределение Бернулли. Элементарным исходом в схеме Бернулли из n испытаний будет являться

$$\omega = (x_1, x_2, \ldots, x_n),$$

где x_i — результат испытания i. Пусть в элементарном исходе k успехов. Тогда

$$\mathbb{P}(\omega) = p^k q^{n-k}$$

Покажем, что веорятность, введенная таким образом, удовлетворяет всем аксиомам вероятностной меры. Неочевидной здесь является только проверка нормировки, то есть надо доказать, что

$$\sum_{\omega} \mathbb{P}(\omega) = 1$$

Для этого просуммируем все исходы по числу успехов (обозначим μ_n).

$$\sum_{\omega} \mathbb{P}(\omega) = \sum_{k=0}^{n} \sum_{\omega\colon \mu_n = k} \mathbb{P}(\omega) = \sum_{k=0}^{n} \sum_{\omega\colon \mu_n = k} p^k q^{n-k} = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1$$

Рассмотрим теперь некоторые важные распределения, связанные со схемой Бернулли.

Уже рассмотренная величина μ_n, равная числу успехов в n испытаниях Бернулли, является случайной величиной с распределением

$$\forall k = 0, 1, \ldots, n \quad \mathbb{P}(\mu_n = k) = C_n^k p^k q^{n-k}$$

Это следует из того, что исходов с k успехами ровно C_n^k, а вероятность каждого равна $p^k q^{n-k}$. Такое распределение называется биномиальным и обозначается $Bi(n, p)$.

Рассимотрим случайные величины $X_i = \mathbb{1}(A_i) \quad i = 1, \ldots, n$, где A_i — успех в i-м испытании. Каждая такая величина имеет распределение Бернулли. Тогда число успехов можно представить так:

$$\mu_n = \sum_{i=1}^{n} X_i$$

Найдем матожидание и дисперсию μ_n.

$$\mathbb{E}\mu_n = \mathbb{E}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n}\mathbb{E}X_i = \sum_{i=1}^{n}(1p+0q) = np$$

$$\mathbb{E}X_i^2 = 1p + 0q = p$$
$$\mathbb{D}X_i = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = p - p^2 = pq$$

В силу независимости испытаний дисперсия линейна относительно сложения

$$\mathbb{D}\mu_n = \mathbb{D}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n}\mathbb{D}X_i = \sum_{i=1}^{n} pq = npq$$

Теорема 10.1 (Бернулли) *Для случайной величины с распределением* $Bi(n,p)$

$$\mathbb{P}(|\frac{\mu_n}{n} - p| \geq \epsilon) \leq \frac{npq}{n^2\epsilon^2}$$

Доказательство
Домножим обе части неравенства на n и воспользуемся неравенством Чебышева:

$$\mathbb{P}(|\mu_n - pn| \geq n\epsilon) \leq \frac{\mathbb{D}(\mu_n)}{(n\epsilon)^2} = \frac{npq}{n^2\epsilon^2}$$

■

Пусть $f(x) \in C[0,1]$.

Определение 10.3 *Многочленом Бернштейна* *называется функция*

$$B_n(x,f) = \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(\frac{k}{n}), \quad x \in [0,1]$$

Заметим, что $B_n(x,f) = \mathbb{E}f(\frac{\mu_n}{n}), \quad \mu_n \sim Bi(n,x)$. (запись $\xi \sim$ *destrname* означает, что случайная величина ξ имеет распределение *destrname*).

Утверждение 10.1
$$B_n(x,f) \rightrightarrows f(x), \quad x \in [0,1]$$

Доказательство
Пользуясь тем, что $\sum\limits_{k=0}^{n} C_n^k x^k (1-x)^{n-k} = 1$ получим

$$|B_n(x,f) - f(x)| = |\sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(\frac{k}{n}) - \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(x)| \leq \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} |f(\frac{k}{n}) - f(x)|$$

Разобьем данную сумму на две:

$$\sum_{k=0}^{n} = \sum_{|\frac{k}{n} - x| < \delta} + \sum_{|\frac{k}{n} - x| \geq \delta}$$

Выберем δ так, чтобы первая сумма была меньше $\frac{\epsilon}{2}$. Это всегда можно сделать, так как функция $f(x)$ непрерывна, а $\sum\limits_{|\frac{k}{n}-x|<\delta} C_n^k x^k(1-x)^{n-k} \leq 1$. Во второй сумме ограничим модуль числом $M = 2\sup\limits_{[0,1]} f(x)$ (1-я теорема Вейерштрасса), а к оставшейся сумме применим неравенство Чебышева, поскольку она равна вероятности $\mathbb{P}(|\mu_n - nx| \geq n\delta)$.

$$\sum_{|\frac{k}{n}-x|\geq\delta} \ldots \leq 2M \sum_{|\frac{k}{n}-x|\geq\delta} C_n^k x^k (1-x)^{n-k} \leq 2M\frac{1}{n\delta^2}$$

Выбором n сделаем вторую сумму меньше $\frac{\epsilon}{2}$, доказав, тем самым, равномерную сходимость. ■

Пусть в схеме Бернулли с вероятностью успеха $0 < p \leq 1$ величина η равна номеру первого успеха. η является случайной величиной, принимающей натуральные значения. Найдем распределение η: серия, в которой первый успех появляется в k-м испытании выглядит так: НН...НУ. Отсюда

$$\forall k \in \mathbb{N} \quad \mathbb{P}(\eta = k) = (1-p)^{k-1}p = q^{k-1}p$$

Так как $\Omega = \left\{\omega_k = \underbrace{0\ldots0}_{k}1\colon k \in \mathbb{N}\right\}$, то

$$\sum_{k=1}^{n} \mathbb{P}(\omega_k) = \sum_{k=1}^{n} p(1-p)^{k-1} = 1$$

Такое распределение называется <u>геометрическим с параметром p</u>.

Пусть ξ имеет геометрическое распределение с параметром p. Тогда

$$\mathbb{E}\xi = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -p\sum_{k=1}^{\infty}\frac{d}{dp}(1-p)^k = -p\frac{d}{dp}(\frac{1-p}{p}) = \frac{1}{p}$$

Аналогично, дифференцируя степенные ряды, получим диспресию

$$\mathbb{D}\xi = \frac{q}{p^2}$$

Пусть теперь θ — число неудач до r-го успеха. Тогда

$$\forall k \in \mathbb{Z}_0 \quad \mathbb{P}(\theta = k) = p^r q^k C_{k+r-1}^k$$

Это следует из того, что всего испытаний было $r + k$, на последнем месте успех, а до него как-то располагаются k неудач и $r - 1$ успех. Такое распределение называется <u>отрицательным биномиальным</u> с параметрами r, p.

11 Предельные теоремы

Теорема 11.1 (Пуассон) *Пусть $\lambda = np$. Тогда при малых p и больших n можно использовать приближение*

$$\mathbb{P}(\mu_n = k) = C_n^k p^k q^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}$$

Доказательство Докажем индукцией по k. При $k = 0$

$$\mathbb{P}(\mu_n = 0) = (1-p)^n = (1-\frac{\lambda}{n})^n \longrightarrow e^{-\lambda} \quad (n \to \infty)$$

$$\frac{\mathbb{P}(\mu_n = k)}{\mathbb{P}(\mu_n = k-1)} = \frac{n!}{k!(n-k)!}p^k q^{n-k}\frac{(k-1)!(n-k+1)!}{n!}\frac{1}{p^{k-1}q^{n-k+1}} = \frac{(n-k+1)p}{kq}$$

$$\mathbb{P}(\mu_n = k) = \frac{(n-k+1)p}{kq}\mathbb{P}(\mu_n = k-1) = \frac{(n-k+1)\frac{\lambda}{n}}{k(1-\frac{\lambda}{n})}\mathbb{P}(\mu_n = k-1)$$

Воспользуемся предположением индукции для $k-1$.

$$\mathbb{P}(\mu_n = k) \to \frac{(n-k+1)\frac{\lambda}{n}}{k(1-\frac{\lambda}{n})}\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \longrightarrow \frac{e^{-\lambda}\lambda^k}{k!}(n \to \infty)$$

■

Следующая оценка формализует «малость» p и «величину» n:

$$\sup_k |\mathbb{P}(\mu_n = k) - \frac{e^{-\lambda}\lambda^k}{k!}| \leq 2np^2$$

Таким образом, зная n, p всегда можно оценить сверху погрешность аппроксимации.

При очень большом числе испытаний ни один нормальный компьютер не способен вычислить $C_n^k p^k q^{n-k}$, а уж тем более сложить их. Данная теорема позволяет сводить вычисление таких сложных вещей к вычислению экспонент.

Распределение величины ξ такое, что $\forall k \in \mathbb{Z}_+ \quad \mathbb{P}(\xi = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, называется распределением Пуассона и обозначается $\mathrm{Pois}\,(\lambda)$

Одно из приложений распределения Пуассона — это пуассоновские потоки. Пусть во времени происходят некоторые события, которые мы фиксируем. λ — интенсивность потока — показывает среднее число событий за единицу времени. Тогда число произошедших событий на отрезке $[0, t]$ равно ξ_t: $\mathbb{P}(\xi_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$.

Свойства пуассоновского распределения. ($\xi \sim \mathrm{Pois}\,(\lambda)$)

1. $\mathbb{E}\xi = \lambda$
2. $\mathbb{D}\xi = \lambda$
3. $\xi_i \sim \mathrm{Pois}\,(\lambda_i) \quad i = 1, \dots, n \Rightarrow \sum\limits_{i=1}^{n}\xi_i \sim \mathrm{Pois}\,(\sum\limits_{i=1}^{n}\lambda_i)$

Утверждение 11.1 *Пусть независимые случайные величины $\xi_i \sim \mathrm{Pois}\,(\lambda_i), \quad i = 1, 2$. Тогда условное распределение ξ_1 при условии $\xi_1 + \xi_2 = n$ имеет биномиальное распределение с параметрами $n, \frac{\lambda_1}{\lambda_1+\lambda_2}$, то есть*

$$\mathbb{P}(\xi_1 = k | \xi_1 + \xi_2 = n) = C_n^k p^k (1-p)^{n-k}, \quad p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Доказательство

$$\mathbb{P}(\xi_1 = k | \xi_1 + \xi_2 = n) = \frac{\mathbb{P}(\xi_1 = k, \xi_1 + \xi_2 = n}{\mathbb{P}(\xi_1 + \xi + 2 = n)} = \frac{\mathbb{P}(\xi_1 = k, \xi_2 = n - k}{\mathbb{P}(\xi_1 + \xi + 2 = n)} =$$

В числителе воспользуемся независимостью ξ_1, ξ_2, а в знаменателе свойством 3 пуассоновского распределения:

$$= \frac{e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}} = C_n^k (\frac{\lambda_1}{\lambda_1 + \lambda_2})^k (\frac{\lambda_2}{\lambda_1 + \lambda_2})^{n-k}$$

■

Теорема 11.2 (Муавра-Лапласа) *Пусть* $\sqrt{npq}$ *велико. Тогда*

$$\mathbb{P}(\mu_n = k) = \mathbb{P}(\frac{\mu_n - np}{\sqrt{npq}} = \frac{k - np}{\sqrt{npq}}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + o(1), \quad x \equiv \frac{k - np}{\sqrt{npq}}$$

Доказательство не было и не должно быть. ■

Рассмотрим функции

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int\limits_{-\infty}^{x} \phi(u) du$$

Заметим, что

$$\int\limits_{-\infty}^{+\infty} \phi(u) du = 1$$

Это верно, так как данный интеграл сводится заменой $t = \frac{x}{\sqrt{2}}$ к известному интегралу Пуассона.

Так как $\phi(x)$ неотрицательна, и интеграл от нее по всей прямой равен 1, она является плотностью некоторого абсолютно непрерывного распределения, которое называется стандартным нормальным распределением и обозначается $N(0, 1)$.

Пусть $\xi \sim N(0, 1)$.

$$\mathbb{E}\xi^k = \int\limits_{-\infty}^{+\infty} x^k \phi(x) dx$$

$$e^{h\xi} = 1 + h\xi + \frac{(h\xi)^2}{2!} + \dots$$

$$\mathbb{E}e^{h\xi} = 1 + h\mathbb{E}\xi + \frac{h^2}{2!}\mathbb{E}\xi^2 + \dots \equiv \psi(h)$$

Функция $\psi(h)$ называется производящей функцией моментов.

$$\psi(h) = \int\limits_{-\infty}^{+\infty} e^{hx} \phi(x) dx = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{hx - \frac{x^2}{2}} dx = \frac{e^{\frac{h^2}{2}}}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{(x-h)^2}{2}} dx = e^{\frac{h^2}{2}}$$

$$e^{\frac{h^2}{2}} = 1 + \frac{h^2}{2} + \frac{1}{2!}(\frac{h^2}{2})^2 + \dots$$

Приравнивая это равенство к предыдущему разложению получим, что все нечетные центральные моменты равны 0.

$$\mathbb{E}\xi^{2n-1} = 0$$

$$\frac{1}{n!2^n} = \frac{\mathbb{E}\xi^{2n}}{(2n)!} \Rightarrow \mathbb{E}\xi^{2n} = \frac{(2n)!}{n!2^n} = \frac{2n(2n-1)\dots n \dots 1}{n!2^n} = (2n-1)!!$$

Последняя формула позволяет очень быстро получить нужный момент, не считая интеграл по частям много раз.

Теорема 11.3 (Интегральная теорема Муавра-Лапласа) *Пусть $\sqrt{npq}$ велико. Тогда*

$$\mathbb{P}(m_1 \leq \mu_n \leq m_2) = \mathbb{P}(\frac{m_1 - np}{\sqrt{npq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}}) \approx \int\limits_{x_1}^{x_2} \phi(x)dx, \quad x1 \equiv \frac{m_1 - np}{\sqrt{npq}}, x2 \equiv \frac{m_2 - np}{\sqrt{npq}}$$

С данной теоремой связана «Задача о докторе Споке». (Добавить ссылку)

Рассмотрим теперь случайную величину ξ с геометрическим распределением. Поскольку $\mathbb{E}\xi = \frac{1}{p}$,

$$p\xi = \frac{\xi}{\mathbb{E}\xi}$$

При стремлении вероятности успеха к нулю, номер первого успеха будет стремиться к бесконечности, как и матожидание ξ. Однако их отношение будет иметь конечный предел:

$$\forall x > 0 \quad \mathbb{P}(p\xi > x) = \mathbb{P}(\xi > \frac{x}{p}) = \sum_{k=[\frac{x}{p}]+1}^{\infty} p(1-p)^{k-1} = \frac{p(1-p)^{[\frac{x}{p}]}}{p} = (1-p)^{[\frac{x}{p}]}$$

$$\lim_{p \to 0} \mathbb{P}(p\xi > x) = \lim_{p \to 0} (1-p)^{[\frac{x}{p}]} = e^{-x}$$

В последнем переходе от дробной части можно избавиться, так как она не вносит никакого вклада в предел. Итак, получили, что

$$\lim_{p \to 0} \mathbb{P}(p\xi < x) = 1 - e^{-x}$$

Это частный случай <u>показательного</u> (экспоненциального) распределения. В общем случае оно выглядит так:

$$F(x) = 1 - e^{-\lambda x}$$

$$p(x) = \lambda e^{-\lambda x}$$

Матожидание случайной величины, распределенной показательно, имеет вид

$$\int\limits_0^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

Показательно распределение обладает интересным свойством: свойством отсутвия последействия (или отсутствия памяти). Предположим, что вы ждете автобус на остановке, а время между автобусами ξ имеет показательное распределение. Тогда вероятность того, что вы прождете автобус еще t никак не зависит от того, как долго (τ) вы уже ждете. Формализуем это:

$$\mathbb{P}(\xi > t + \tau | \xi > \tau) = \mathbb{P}(\xi > t)$$

По определению условной вероятности

$$\mathbb{P}(\xi > t + \tau | \xi > \tau) = \frac{\mathbb{P}(\xi > t + \tau, \xi > \tau)}{\mathbb{P}(\xi > \tau)} = \frac{\mathbb{P}(\xi > t + \tau)}{\mathbb{P}(\xi > \tau)} = \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}} = e^{-\lambda t}$$

Показательное распределение является единственным обладающим таким свойством в классе абсолютно непрерывных распределений. В классе дискретных распределений таким свойством обладает только геометрическое распределение (введенное как число неудач до первого успеха).

12 Векторные случайные величины

Векторную случайную величину можно определить двумя способами:

1. Сказать, что $\xi = (\xi_1, \ldots, \xi_n)$ является векторной случайной величиной, если ее координаты являются случайными величинами.

2. Дать определение аналогично определению одномерной случайной величины, то есть рассмотреть $(\mathbb{R}^n, \mathscr{B}_n)$, где $\mathscr{B}_n$ — n-мерная борелевская σ-алгебра, то есть минимальная σ-алгебра, содержащая все параллелепипеды (аналогично интервалам в одномерном случае). Тогда ξ — случайная величина, если
$$\forall B \in \mathscr{B}_n \colon \xi^{-1}(B) \in \mathscr{A}$$
.

Определение 12.1 *Функция распределения*

$$F_\xi(x_1, \ldots, x_n) = \mathbb{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$$

Свойства функции распределения:

1. Пусть $x, y \in \mathbb{R}^n$ такие, что $x_i < y_i \quad i = 1, \ldots, n$. Тогда $F_\xi(x) \leq F_\xi(y)$.

2. $0 \leq F_\xi(x) \leq 1$

3. $\lim\limits_{x_i \to -\infty} F_\xi(x) = 0$. То есть при стремлении одной координаты к $-\infty$ функция распределения стремится к 0, поскольку вероятность для соответсвующей координаты стремится к 0.

4. Если же какую-то из координат устремить к бесконечности, то она не будет учитываться в вероятности, поэтому $\lim\limits_{x_i \to +\infty} F_\xi(x) = F_{(\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n)}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

5. Непрерывна слева по каждому аргументу.

Определение 12.2 *Абсолютно непрерывная случайная величина ξ — такая случайная величина, что*

$$\forall B \in \mathscr{B}_n \quad \mathbb{P}(\xi \in \mathscr{B}_n) = \int\limits_B p_\xi(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$\xi_1, \ldots, \xi_n$ независимы в совокупности $\Leftrightarrow$ функцию распределения векторной случайной величины $\xi = (\xi_1, \ldots, \xi_n)$ можно представить произведением функций распределений координат, то есть

$$F_\xi(x) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n)$$

В случае абсолютно непрерынвой векторной случайной величины это эквивалентно аналогичному выражению для плотностей:

$$p_\xi(x) = p_{\xi_1}(x_1) \ldots p_{\xi_n}(x_n)$$

13 Распределение функций от случайных величин

Рассмотрим отображение $f\colon \mathbb{R}^n \to \mathbb{R}^n$ с якобианом $J_f = \frac{D(f_1(x),\dots,f_n(x))}{D(x_1,\dots,x_n)}$. Если он отличен от 0, то существует обратная функция, и выполнено соотношение

$$J_f J_{f^{-1}} = 1$$

Теорема 13.1 *Пусть $f\colon \mathbb{R}^n \to \mathbb{R}^n$ — достаточно гладкая функция с ненулевым якобианом. $\xi = (\xi_1, \dots, \xi_n) \sim p_\xi(x)$. Рассмотрим случайную величину $\eta = f(\xi)$. Тогда*

$$p_\eta(x) = p_\xi(f^{-1}(x))|J_{f^{-1}}(x)|$$

Доказательство
Вспомним формулу замены переменных в интеграле:

$$\int\limits_B \phi(x)dx = \int\limits_{f^{-1}(B)} \phi(f(y))|J_f(y)|dy$$

Тогда $\forall B \in \mathscr{B}_n$

$$\int\limits_B p_\xi(f^{-1}(x))|J_{f^{-1}(x)}|dx = \int\limits_{f^{-1}(B)} p_\xi(f(f^{-1}(B)))|J_{f^{-1}(x)}||J_{f(x)}|dx =$$

$$\int\limits_{f^{-1}(B)} p_\xi(x)dx = \mathbb{P}(\xi \in f^{-1}(B)) = \mathbb{P}(f(\xi) \in B) = \mathbb{P}(\eta \in B)$$

■

Теперь найдем формулу для плотности суммы случайных величин. Пусть $\xi = (\xi_1, \xi_2) \sim p_\xi(x)$ Рассмотрим $f\colon (x_1, x_2) \mapsto (x_1 + x_2, x_2)$. Тогда

$$f^{-1}\colon (x_1, x_2) \mapsto (x_1 - x_2, x_2), \quad J_{f^{-1}} = 1$$

Обозначим $\eta = f(\xi)$. Тогда по только что доказанной теореме $p_\eta(x_1, x_2) = p_\xi(x_1 - x_2, x_2)$ Тогда по свойству 4 функции распределения векторной случайной величины

$$p_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} p_\xi(x - x_2, x_2)dx_2 = \{\text{независимость } \xi_1, \xi_2\} = \int\limits_{-\infty}^{+\infty} p_{\xi_1}(x - x_2)p_{\xi_2}(x_2)dx_2$$

В силу симметрии

$$p_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} p_{\xi_1}(x_1)p_{\xi_2}(x - x_2)dx_1$$

Эти формулы называются формулами свертки. Существует аналогичная формула для функция распределения, но мы ее не доказываем, потому что это какой-то функан:

$$F_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} F_{\xi_1}(x - y)dF_{\xi_2}(y) = \int\limits_{-\infty}^{+\infty} F_{\xi_2}(x - y)dF_{\xi_1}(y)$$

14 Виды сходимости случайных величин

В этом разделе мы введем вагон непонятных определений, а потом постараемся запутаться еще больше, доказывая, какая сходимость круче, и ковыряясь в контрпримерах. Но для начала вспомним наших старых знакомых: измеримое пространство $(\Omega, \mathscr{A})$, вероятностное пространство $(\Omega, \mathscr{A}, \mathbb{P})$ и случайную величину $\xi\colon \Omega \to \mathbb{R}$, которая обладает свойством

$$\forall B \in \mathscr{B} \quad \xi^{-1}(B) = \{\omega \in \Omega\colon \xi(\omega) \in B\} \in \mathscr{A}$$

Пусть теперь $\xi, \xi_1, \xi_2, \ldots$ — случайные величины на $(\Omega, \mathscr{A})$.

Определение 14.1 *Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ почти наверное (с вероятностью 1), если вероятность множества тех элементарных событий, где она не сходится, равно нулю.*

$$\{\xi_n\} \xrightarrow{\text{п.н.}} \xi, \textit{если} \quad \mathbb{P}(\{\omega\colon \xi_n(\omega) \nrightarrow \xi\}) = 0$$

Пример Пусть ξ_n принимает значение n в рациональных точках числовой прямой и 0 — в иррациональных. Тогда $\{\xi_n\} \xrightarrow{\text{п.н.}} 0$, так как мера множества рациональных чисел, где случайная величина расходится, — ноль.

Определение 14.2 *Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ в среднем порядка r, если r-тый момент их разности сходится к нулю.*

$$\{\xi_n\} \xrightarrow{(r)} \xi, \textit{если} \quad \mathbb{E}(\xi_n - \xi)^r \xrightarrow[n\to+\infty]{} 0$$

Пример Возьмем в качестве множества Ω окружность длины 1, событиями будут борелевские множества, а вероятность введем как меру Лебега. События A_n введем таким образом: A_1 — дуга длины $1/2$, отложенная от какой-то точки против часовой стрелки. A_n — дуга длины $\frac{1}{n+1}$, отложенная против часовой стрелки от конца дуги A_{n-1}. Введем случайную величину: $\xi_n(\omega) = \mathbb{1}(\omega \in A_n)$ Покажем, что $\{\xi_n\}$ сходится к 0 в среднем любого положительного порядка.

$$\mathbb{E}(\xi_n - \xi)^r = \mathbb{E}\xi_n^r = \mathbb{E}\left(\mathbb{1}(\omega \in A_n)\right)^r = \mathbb{P}(A_n) = 1/n \xrightarrow[n\to+\infty]{} 0$$

Утверждение 14.1 *Из сходимости в среднем, вообще говоря, не следует сходимость почти наверное.*

Доказательство В рассмотренном выше примере $\xi_n(\omega)$ не сходится ни в одной точке окружности. Действительно, так как ряд $\sum_{n=1}^{\infty} \frac{1}{n}$ расходится, то для любой точки ω на окружности мы можем указать бесконечное число номеров n_k, таких что $\omega \in A_{n_k}$. ■

Утверждение 14.2 *Из сходимости почти наверное, вообще говоря, не следует сходимость в среднем.*

Доказательство Рассмотрим отрезок $[0,1]$, с событиями, являющимися борелевскими множествами и вероятностью, введенной как мера Лебега. Положим $\xi_n(\omega) = e^n\mathbb{1}(\omega \in [0,1/n])$. Тогда $\xi_n \xrightarrow{\text{п.н.}} 0$, но при этом $\forall r > 0 \quad \mathbb{E}\xi_n^r = e^{np}\mathbb{E}\mathbb{1}(\omega \in [0,1/n]) = e^{np}/n$, а эта величина стремится к бесконечности, значит, сходимости в среднем нет. ■

Введем еще один вид сходимости.

Определение 14.3 *Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ по вероятности, если для любого сколь угодно малого положительного ε вероятность таких событий, что модуль разности ξ_n и ξ больше ε, стремится к 0.*

$$\{\xi_n\} \xrightarrow{\mathbb{P}} \xi, \textit{если} \quad \forall\varepsilon > 0 \quad \mathbb{P}(\{\omega\colon |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}) \xrightarrow[n\to+\infty]{} 0$$

 На этом моменте может немного поплавиться мозг в попытках понять, чем сходимость по вероятности отличается от сходимости с вероятностью 1. Действительно, и там и там мы говорим, что вероятность тех событий, на которых случайная величина не сходится, равна нулю. Но разница в том, что в случае сходимости почти наверное мы сначала устремляем n к бесконечности, а потом считаем вероятность, событий, когда не сходится, а в случае сходимости по вероятности мы сначала посчитали вероятность для какого-то фиксированного n, а потом устремились к бесконечности.

Пример Докажем, что та жуткая последовательность случайных величин на окружности сходится по вероятности к нулю. Действительно,

$$\mathbb{P}(\{\omega\colon |\xi_n(\omega) - 0| > \varepsilon\}) = \mathbb{P}(\{\omega\colon \xi_n(\omega) > \varepsilon\}) = \mathbb{P}(\{\omega\colon \xi_n(\omega) = 1\}) = 1/n \xrightarrow[n\to+\infty]{} 0$$

Утверждение 14.3 *Только что рассмотренным примером мы доказали, что из сходимости по вероятности, вообще говоря, не следует сходимость почти наверное.*

Утверждение 14.4 *Из сходимости по вероятности, вообще говоря, не следует сходимость в среднем.*

Доказательство Для доказательства воспользуемся примером про отрезок. Докажем, что $\{\xi_n\} \xrightarrow{\mathbb{P}} 0$. Действительно, $\mathbb{P}\{\omega\colon \xi_n(\omega) > \varepsilon\} = 1/n \xrightarrow[n\to+\infty]{} 0$. Отсутствие сходимости в среднем мы уже доказали. ■

Теорема 14.1 *Из сходимости почти наверное следует сходимость по вероятности.*

Доказательство Сначала докажем, что $\{\xi_n\} \xrightarrow{\text{п.н.}} \xi \quad \Leftrightarrow \quad \mathbb{P}\left(\sup\limits_{k\geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right) \to 0.$

Положим $A_n^\varepsilon = \{\omega\colon |\xi_n - \xi| \geqslant \varepsilon\}$, $A^\varepsilon = \overline{\lim}\, A_n^\varepsilon \equiv \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k\geqslant n} A_k^\varepsilon$. Тогда:

$$\mathbb{P}\left(\{\omega\colon \xi_n(\omega) \nrightarrow \xi\}\right) = 0 \quad \Leftrightarrow \quad \mathbb{P}\left(\bigcup_{\varepsilon>0} A^\varepsilon\right) = 0 \quad \Leftrightarrow \quad \mathbb{P}\left(\bigcup_{m=1} A^{1/m}\right) = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \mathbb{P}(A^{1/m}) = 0, m \geqslant 1 \quad \Leftrightarrow \quad \mathbb{P}(A^{\varepsilon}) = 0, \varepsilon > 0 \quad \Leftrightarrow \quad \mathbb{P}\left(\bigcup_{k \geqslant n} A_k^{\varepsilon}\right) \xrightarrow[n\to\infty]{} 0, \varepsilon > 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \mathbb{P}\left(\sup_{k\geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \quad \Rightarrow \quad \mathbb{P}\left(|\xi_k - \xi| \geqslant \varepsilon\right) \to 0.$$

А последнее утверждение — это определение сходимости по вероятности. ■

 Совершенно жуткое доказательство, понимается методом вглядывания: если достаточно долго медитировать над каждой импликацией, то все переходы в конце концов станут понятными.

Теорема 14.2 *Из сходимости в среднем следует сходимость по вероятности.*

Доказательство Утвеждение теоремы практически сразу же следует из обобщенного неравентва Чебышева:

$$\mathbb{P}\left(|\xi_n - \xi| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{E}|\xi_n - \xi|^r}{\varepsilon^r}.$$

Переходя в неравенстве к пределу при $n \to \infty$, получаем требуемое. ■

Определение 14.4 *Последовательность случайных величин $\{\xi_n\}$ слабо сходится к случайной величине ξ, если для любой непрерывной и ограниченной функции последовательность мат. ожиданий функций от ξ_n сходится к мат. ожиданию функции от ξ.*

$$\{\xi_n\} \xrightarrow{w} \xi, \textit{если} \quad \forall f(x)\colon f(x) \in C \textit{ и } |f(x)| \leqslant M \quad \mathbb{E}f(\xi_n) \to \mathbb{E}f(\xi)$$

Утверждение 14.5 *Из сходимости по вероятности следует слабая сходимость.*

Доказательство Пусть $f(x)\colon |f(x)| \leqslant M$. Также $\forall \varepsilon > 0$ выберем N так, чтобы $\mathbb{P}|\xi| > N \leqslant \dfrac{\varepsilon}{4M}$. Выберем δ так, чтобы $\forall |x| \leqslant N$ и $|x - y| \leqslant \delta$ было выполнено неравенство $|f(x) - f(y)| \leqslant \varepsilon/2$. Тогда:

$$\begin{aligned}
\mathbb{E}|f(\xi_n) - f(\xi)| = \mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta; |\xi| \leqslant N) + \\
\mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta; |\xi| > N) + \mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| > \delta)) \leqslant \\
\varepsilon/2 + \varepsilon/2 + 2M * \mathbb{P}(|\xi_n - \xi| > \delta) = \varepsilon + \mathbb{P}(|\xi_n - \xi| > \delta).
\end{aligned}$$

Но из сходимости по веротности следует, что $\mathbb{P}(|\xi_n - \xi| > \delta) \to 0$, значит при $n \to \infty \quad \mathbb{P}(|\xi_n - \xi| > \delta) < \varepsilon$, тогда $\mathbb{E}|f(\xi_n) - f(\xi)| < 2\varepsilon$, откуда в силу произвольного выбора ε и в силу того, что модуль числа не превосходит самого числа, получаем требуемое. ■

Определение 14.5 *Последовательность слечайных величин $\{\xi_n\}$ сходится к случайной величине ξ по распределению, если функции распределения $\{\xi_n\}$ сходятся к функции распределения ξ во всех точках, в которых предельная функция распределения непрерывна.*

$$\{\xi_n\} \xrightarrow{d} \xi, \textit{если} \quad F_{\xi_n}(x) \to F_{\xi} \quad \forall x\colon F_{\xi}(x) \in C(x)$$

Пример Пусть $\{\xi_n\}$ принимает значения 0 и $1 - \dfrac{1}{n}$ с вероятностями $\dfrac{1}{2}$. Тогда функция распределения для случайной величины ξ_i выглядит так (Рис. 1):

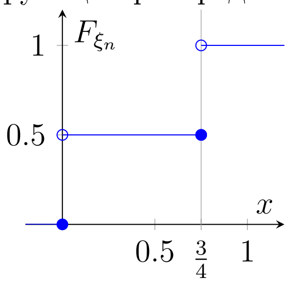


Рис. 1: Функция распределения ξ_n при $n = 4$

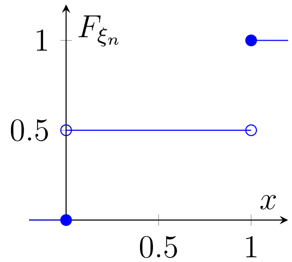


Рис. 2: Предельная функция

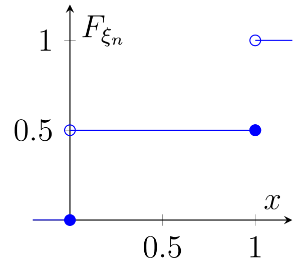


Рис. 3: Функция распределения ξ

Тогда предельная функция будет выглядеть так (Рис. 2). Заметим, что она вообще не является функцией распределения, так как в точке 1 нарушено условие непрерывности слева. Тогда рассмотрим случайную величину ξ, принимающую значение 0 и 1 с вероятностями 1/2, тогда ее функция распределения имеет вид (Рис. 3). Заметим, что предельная функция и функция распределения ξ различаются только в точке 1, в которой $F_\xi(x)$ не является непрерывной, значит $\{\xi_n\} \xrightarrow{d} \xi$.

Утверждение 14.6 *Сходимость по распределению эквивалентна слабой сходимости.*

Доказательство __**TODO:**__ Чет я не знаю, как это доказать =(■

Утверждение 14.7 *Если последовательность случайных величин $\{\xi_n\}$ сходится по распределению к вырожденной случайной величине $\xi \overset{\text{п.н.}}{\equiv} a$, то $\{\xi_n\}$ сходится к ξ по вероятности.*

Доказательство Ы ■

А теперь докажем теорему, ради которой мы и городили все эти огороды сходимостей. Эта теорема впоследствии будет использоваться при доказательстве центральной предельной теоремы.

Теорема 14.3 (теорема Леви о непрерывности)

1. *Пусть $\{\xi_n\} \xrightarrow{d} \xi$ и $f_n(t) := \mathbb{E}e^{it\xi_n}$. Тогда $\{f_n(t)\} \to f(t)$, где $f(t) = \mathbb{E}e^{it\xi}$.*
2. *Пусть теперь $f_n(t) := \mathbb{E}e^{it\xi_n}$. Тогда если $f_n(t) \to f(t) \quad \forall t \in \mathbb{R}, \quad f(t) \in C(0)$, то $f(t)$ является характеристической функцией некоторой случайной величины ξ, такой что $\{\xi_n\} \xrightarrow{d} \xi$.*

Доказательство

1. Как было показано выше, сходимость по распределению эквивалентна слабой сходимости. Положим в определении слабой сходимость функции $\phi(x) := Re(e^{itx})$ и $\psi(x) := Im(e^{itx})$, тогда $\mathbb{E}\phi(\xi_n) \to \mathbb{E}\phi(\xi)$ и $\mathbb{E}\psi(\xi_n) \to \mathbb{E}\psi(\xi)$, откуда следует утверждение теоремы.

2. Достаточно грустное доказательство, но, если у меня будет настроение, я его напишу.

■

А вот картинка, на которой схематично показано, что из чего следует.

$$
\begin{tikzcd}
\boxed{\{\xi_n\} \xrightarrow{\text{п.н.}} \xi} \arrow[dr] & & & \\
& \boxed{\{\xi_n\} \xrightarrow{(\mathbb{P})} \xi} \arrow[r] & \boxed{\{\xi_n\} \xrightarrow{(d)} \xi} \arrow[l, bend right, "\xi \overset{\text{п.н.}}{\equiv} a"'] \arrow[r, dotted, dash, very thick] \arrow[dd, leftrightarrow] & \boxed{\text{т. о непрерывности}} \\
\boxed{\{\xi_n\} \xrightarrow{(r)} \xi} \arrow[ur] & & & \\
& & \boxed{\{\xi_n\} \xrightarrow{(w)} \xi} &
\end{tikzcd}
$$

15 Закон больших чисел

В этой части рассмотрим закон больших чисел. Причем сначала скажем, как он формулируется, а потом подоказываем его для различных исходных.

Определение 15.1 *Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин, а $S_n = \sum\limits_{k=1}^{n} \xi_n$. Говорят, что $\{\xi_n\}$ удовлетворяет закону больших чисел (ЗБЧ), если $\dfrac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_i$.*

Пример Пусть ξ_k независимы и равномерно распределены на отрезке $[-1, 1]$. Тогда матожидание каждого слагаемого равно 0. Тогда

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{D}S_n}{\varepsilon^2} = \frac{\mathbb{D}\xi_1}{n\varepsilon^2} \to 0.$$

То есть $\dfrac{S_n}{n}$ сходится по вероятности к своему матожиданию, значит для данной последовательности слуайных величин выполнен ЗБЧ.

Теорема 15.1 (ЗБЧ в форме Хинчина) *Если $\{\xi_k\}$ — независимые одинаково распределенные случайные величины с конечным матождиданием, то для них выполнен ЗБЧ.*

Доказательство Первое желание, которое может возникнуть при виде данной теоремы — записать неравенство Чебышева, но для неравентва Чебышева нужно не только конечное матожидание, но и конечный второй момент, поэтому так не прокатит. Тогда мы воспользуемся тем, что если последовательность случайных величин сходится к числу, то сходимость по вероятности эквивалентна слабой сходимости, и в качестве функции в определении слабой сходимости положим характеристическую функцию.

$$f(t) := \mathbb{E}e^{it\xi_k} \quad \forall k = \overline{1 \ldots n}.$$

Здесь нам не важно, для какой случайной величины считать харфункцию или матожидание, так как они все одинаково распределены.

$$f_n(t) := \mathbb{E}e^{it\frac{S_n}{n}} = \mathbb{E}e^{it\frac{\xi_1 + \ldots + \xi_n}{n}} = \prod_{k=1}^{n} \mathbb{E}e^{\frac{it\xi_k}{n}} = f^n\left(\frac{t}{n}\right).$$

Далее вспомним важной свойство харфункции: $a := f'(0) = \mathbb{E}\xi_k$. Тогда:

$$f^n\left(\frac{t}{n}\right) = \left(1 + \frac{ita}{n} + \overline{o}\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{ita}.$$

Заметим, что e^{ita} непрерывна на всей действительной прямой и обращается в 1 при $t = 0$. Значит по теореме о непрерывности e^{ita} представляет собой харфункцию некоторой случайной величины, к которой слабо сходится $\dfrac{S_n}{n}$. Действительно, эта случайная величина $\xi \overset{\text{п.н.}}{\equiv} a$. Таким образом, имеем, что $\frac{S_n}{n} \xrightarrow{w} \mathbb{E}\xi_k$, но если предельная случайная величина вырождена, то слабая сходимость эквивалентна сходимости по вероятности, значит $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_k$. ■

Лулзов ради закинем еще парочку ЗБЧ.

Теорема 15.2 (ЗБЧ в форме Чебышева) *Пусть $\{\xi_k\}$ — последовательность некоррелированных случайных величин с равномерно ограниченной дисперсией. То есть $\sup\limits_{k\in\mathbb{N}} \mathbb{D}\xi_k \leqslant C < \infty$. Тогда для данной последовательности случайых величин выполнен ЗБЧ.*

Доказательство Последовательно воспользуемся неравенством Чебышева, некоррелированостью случайных величин (а значит дисперсия суммы равна сумме дисперсий) и равномерной ограниченностью дисперсий.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\xi_k\right| \geqslant \varepsilon\right) = \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{D}S_n}{(n\varepsilon)^2} = \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n}\mathbb{D}\xi_k \leqslant \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

■

Замечание Условия теоремы можно ослабить: вместо некоррелированности потребовать, чтобы сумма ковариаций обращалась в 0, а вместо равномерной ограниченности константой можно потребовать, чтобы дисперсии росли строго медленнее, чем линейная функция.

Теорема 15.3 (Усиленный ЗБЧ (УЗБЧ) в форме Колмогорова) *Если $\{\xi_k\}$ — независимые одинаково распределенные случайные величины с конечным матожиданием, то $\dfrac{S_n}{n} \xrightarrow{\text{п.н.}} \mathbb{E}\xi_k$.*

Замечание Условия те же, что и на ЗБЧ в форме Хинчина, но сходимость уже не по вероятности, а почти наверное.
Доказательство Все верно. Отвечаю. ■

 Собственно, а к чему будут сходиться средние арифметические, если нет матожидания?

Пример Распределим $\{\xi_k\}$ по Коши с параметрами $C(0,\gamma)$ и посмотрим, что будет происходить со средними арифмитическими.

$$p_{\xi_k}(x) = \frac{1}{\pi\gamma\left(1 + \left(\dfrac{x}{\gamma}\right)^2\right)}, \quad f(t) := \mathbb{E}e^{it\xi_k} = e^{-\gamma|t|}.$$

$$f_n(t) := \mathbb{E}S_n = \mathbb{E}e^{it\sum\limits_{k=1}^{n}\xi_k} = \prod_{k=1}^{n}\mathbb{E}e^{it\xi_k} = f^n(t) = e^{-\gamma n|t|} \implies$$

$$p_{S_n}(t) = \frac{1}{\pi n\gamma\left(1 + \left(\dfrac{x}{n\gamma}\right)^2\right)}.$$

Теперь предположим, что последовательность $\{S_n\}$ сходится по вероятности к чему-то, и придем к противоречию. Пусть $\mathbb{P}\left(\left|\frac{S_n}{n}-a\right|\geqslant\varepsilon\right)\to 0$. Тогда:

$$\mathbb{P}\left(\left|\frac{S_n}{n}-a\right|\geqslant\varepsilon\right)=\int\limits_{n(a+\varepsilon)}^{+\infty}p(x)dx+\int\limits_{-\infty}^{n(a-\varepsilon)}p(x)dx=\frac{1}{\pi}\operatorname{arctg}\left(\frac{x}{\gamma n}\right)\bigg|_{n(a+\varepsilon)}^{+\infty}+\frac{1}{\pi}\operatorname{arctg}\left(\frac{x}{\gamma n}\right)\bigg|_{-\infty}^{n(a-\varepsilon)}=$$

$$=\frac{1}{\pi}\left(\operatorname{arctg}\left(\frac{a-\varepsilon}{\gamma}\right)-\operatorname{arctg}\left(\frac{a+\varepsilon}{\gamma}\right)\right)\cancel{\xrightarrow[n\to\infty]{}}0.$$

Таким образом, последовательность средних арифметических случайных величин, распределенный по Коши, не сходится по вероятности вообще ни к чему.

16 Центральная предельная теорема

С места в карьер:

Определение 16.1 *Последовательность случайных величин* $\{\xi_k\}$ *удовлетворяет ЦПТ, если* $\exists\{a_k\} \in \mathbb{R}, \{b_k\} > 0$, *такие что*

$$\frac{\xi_k - a_k}{b_k} \xrightarrow{d} \xi \sim \mathcal{N}(0, 1).$$

Разомнемся на чем-нибудь попроще.

Теорема 16.1 *Пусть* $\{\xi_k\}$ *— норсв с конечным матожиданием и конечной ненулевой дисперсией. Тогда* $\{\xi_k\}$ *удовлетворяет ЦПТ, причем* $a_n = \mathbb{E}S_n$, *а* $b_n = \sqrt{\mathbb{D}S_n}$, *где* $S_n = \sum\limits_{k=1}^{n} \xi_k$.

Замечание Здесь распределение $\dfrac{\xi_k - \mathbb{E}\xi_k}{\sqrt{\mathbb{D}\xi_k}}$ сходится к стандартному нормальному не абы как, а очень даже равномерно по x на всей действительной прямой.
Доказательство Без ограничения общности будем считать, что $\mathbb{E}\xi_k = 0$ и $\mathbb{D}\xi_k = 1$. Тогда

$$f(t) := \mathbb{E}e^{it\xi_k} = 1 + \cancelto{0}{ita} - \frac{t^2}{2} + \overline{o}\left(\frac{1}{n}\right).$$

$$f_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E}e^{it\frac{S_n}{\sqrt{n}}} = \left(f\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + \overline{o}\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-\frac{t^2}{2}}.$$

Таким образом, харфункция центрированной нормированной случайной величины сходится слабо к харфункции стандартного нормального распределения, а значит, центрированная нормарованная случайная величина сходится по распределению к $\mathcal{N}(0, 1)$. ■

 Если вам уже стало плохо, то переживать не стоит — дальше будет еще хуже

С этого момента положим:

- Случайные величины $\{\xi_k\}$ независимы и определены на одном вероятностном пространстве
- $F_k(x) = \mathbb{P}(\xi_k < x)$
- $S_n = \sum\limits_{k=1}^{n} \xi_k$
- $a_k = \mathbb{E}\xi_k$, $A_n = \sum\limits_{k=1}^{n} a_k$
- $b_k^2 = \mathbb{D}\xi_k$, $B_n^2 = \sum\limits_{k=1}^{n} b_k^2$

Теорема 16.2 (Ляпунова) *Пусть* $\mu_k^3 := \mathbb{E}\left|\xi_k - a_k\right|^3 < \infty, \quad M_n^3 := \sum\limits_{k=1}^{n} \mu_k^3$. *Пусть выполнено* <u>*условие Ляпунова*</u>*:*

$$\frac{M_n^3}{B_n^3} \xrightarrow[n\to\infty]{} 0 \tag{16.1}$$

Тогда $\{\xi_k\}$ *удовлетворяет ЦПТ.*

Теорема 16.3 (Линдеберга) *Пусть* $\varepsilon > 0$. *Тогда если* $a_k < \infty$, $b_k < \infty$ *и выполнено* <u>*условие Линдеберга*</u>*:*

$$L_n(\varepsilon) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x-a_k|>\varepsilon B_n} (x-a_k)^2 \, dF_k(x) \xrightarrow[n\to\infty]{} 0, \tag{16.2}$$

То $\{\xi_k\}$ *удовлетворяет ЦПТ.*

Замечание Если случайные величины одинаково распределены, то условие Линдеберга эквивалентно существованию дисперсии.

Определение 16.2 $\{\xi_k\}$ *удовлетворяют* <u>*условию Феллера*</u>*, если*

$$\forall \varepsilon > 0 \quad \max_{1\leqslant k\leqslant n} \mathbb{P}\left(\frac{|\xi_k - a_k|}{B_n}\right) > \varepsilon \xrightarrow[n\to\infty]{} 0. \tag{16.3}$$

Теорема 16.4

$$\begin{cases} \text{ЦПТ} \\ (16.3) \end{cases} \iff (16.2).$$

Доказательство Покажем, что из (16.2) следует (16.3).

$$\max_{1\leqslant k\leqslant n} \mathbb{P}\left(\frac{|\xi_k - a_k|}{B_n} > \varepsilon\right) \leqslant \max_{1\leqslant k\leqslant n} \frac{b_k^2}{\varepsilon^2 B_n^2} = \max_{1\leqslant k\leqslant n} \frac{b_k^2}{\varepsilon^2\left(b_1^2 + \ldots + b_n^2\right)}$$

$$\max_{1\leqslant k\leqslant n} \mathbb{P}\left(|\xi_k - a_k| > \varepsilon B_n\right) \leqslant \sum_{k=1}^{n} \mathbb{P}\left(|\xi_k - a_k| > \varepsilon B_n\right) = \sum_{k=1}^{n} \int\limits_{|x-a_k|>\varepsilon B_n} 1 dF_k(x) =$$

$$\sum_{k=1}^{n} \int\limits_{|x-a_k|>\varepsilon B_n} \frac{(x-a_k)^2}{(x-a_k)^2} dF_k(x) \leqslant \frac{1}{\varepsilon^2 B_n^2} \int\limits_{|x-a_k|>\varepsilon B_n} (x-a_k)^2 dF_k(x) \xrightarrow[n\to\infty]{} 0.$$

Таким образом, условие Линдеберга влечет за собой условие Феллера. ■

Теорема 16.5 *Условие Ляпунова (16.1) влечет за собой условие Линдеберга (16.2).*

Доказательство

$$L_n(\varepsilon) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x-a_k|>\varepsilon B_n} (x-a_k)^2 dF_k(x) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x-a_k|>\varepsilon B_n} \frac{|x-a_k|^3}{|x-a_k|} dF_k(x) \leqslant$$

$$\leqslant \frac{1}{\varepsilon B_n^3} \sum_{k=1}^{n} \int\limits_{-\infty}^{+\infty} \left|x-a_k\right|^3 dF_k(x) = \frac{M_n^3}{\varepsilon B_n^3} \xrightarrow[n\to\infty]{} 0.$$

■

Теорема 16.6 (Неравенство Берри—Эссена) $\sup\limits_x \left| \mathbb{P}\left(\frac{S_n - A_n}{B_n} < x \right) - \Phi(x) \right| \leqslant C \frac{M_n^3}{B_n^3}$.
Если случайные величины одинаково распределены, то правую часть можно переписать так:

$$C \frac{M_n^3}{B_n^3} = C_0 \frac{n\mu_k^3}{n^{3/2} b_k^3} = C_0 \frac{\mu_k^3}{b_k^3 \sqrt{n}}.$$

Константу C_0 *уточняли, уточняли, уточняли и в конце концов доуточнялись до* $C_0 \geqslant \frac{\sqrt{10}+3}{6\sqrt{2\pi}}$. *А потом доказали, что здесь вообще стоит строгое равенство.*