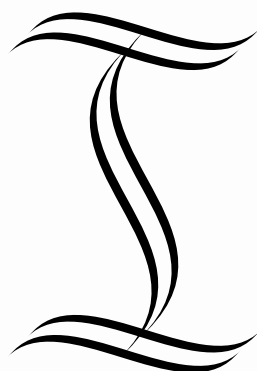


ВМК МГУ & ФУПМ МФТИ

# Теория вероятности. Математическая статистика

*Игорь Тао, Дима Сотников, Дима Гущин*

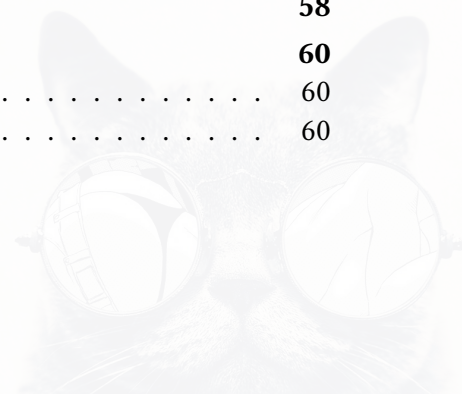


27 сентября 2019 г.



# Содержание

<b>Часть I Теория вероятности</b>	<b>4</b>
<b>1 Вероятностное пространство</b>	<b>5</b>
1.1 Стохастические ситуации	5
1.2 Измеримое пространство	5
1.3 Вероятность. Вероятностное пространство	8
<b>2 Понятие меры и интеграла Лебега</b>	<b>11</b>
2.1 Мера Лебега	11
2.2 Интеграл Лебега	12
<b>3 Связь между событиями</b>	<b>15</b>
3.1 Условная вероятность	15
3.2 Независимость событий. Независимость в совокупности	16
3.3 Формула полной вероятности. Формула Байеса	17
<b>4 Случайные величины</b>	<b>19</b>
4.1 Случайные величины: определение	19
4.2 Функция распределения	19
4.3 Виды распределений	21
4.4 Векторные случайные величины	23
4.5 Условное математическое ожидание	24
4.5.1 Конструктивные определения	24
4.5.2 Аксиоматическое определение	26
4.5.3 Свойства и красивые картинки	26
<b>5 Характеристики случайных величин</b>	<b>29</b>
5.1 Математическое ожидание	29
5.2 Моменты, дисперсия, ковариация и прочее	30
5.3 Независимость случайных величин	31
5.4 Неравенство Маркова, неравенство Чебышева	33
<b>6 Распределения и предельные теоремы</b>	<b>35</b>
6.1 Распределение Бернулли	35
6.2 Многочлен Бернштейна	36
6.3 Распределение Пуассона. Теорема Муавра-Лапласа	38
6.4 Характеристические функции	42
6.5 Распределение функций от случайных величин	45
<b>7 Сходимость случайных величин</b>	<b>47</b>
7.1 Виды сходимости	47
7.2 Закон больших чисел	51
7.3 Центральная предельная теорема	54
<b>Часть II Математическая статистика</b>	<b>57</b>
<b>1 Введение</b>	<b>58</b>
<b>2 Точечное оценивание</b>	<b>60</b>
2.1 Эмпирическая функция распределения и её свойства	60
2.1.1 Вариационный ряд	60



2.1.2	Эмпирическая ф.р.	60
2.1.3	Ядерные оценки	63
2.2	Выборочные моменты	65
2.3	Параметрическая статистика. Оценки	67
2.3.1	Определения	67
2.3.2	Примеры	67
2.4	Функция правдоподобия	69



## Предисловие

Здорово! Раз вы читаете эти строки, значит вы желаете обновить или заложить свои знания в области Теории вероятности и Математической статистике.

Мы придерживались следующего порядка изложения: сначала дали неформальное объяснение всего материала, который дается в книге, затем рассказываем поверхностно о мере Лебега и переходим к теоретико-вероятностным выкладкам, которые являются инструментом в прикладных аспектах этого направления.



---

## ***Часть № I. Теория вероятности***

---

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

Игорян



---

## §1. Вероятностное пространство

### 1.1. Стохастические ситуации

Часть информации, изложенной дальше, взято из книги [1] В. Ю. Королева, так что любознательный читатель может ознакомиться с первоисточником при особом желании. Когда можно применять теорию вероятности — поговорим об этом.

Следуя В. Ю. Королеву, будем говорить о *случайности* как о принципиально неустранимой неопределенности; это как девушку ждать на свидании: точное время, когда она придет, — это неустранимая неопределенность. Чтобы понять, как с ней же работать, необходимо выделить те случаи, когда мы можем сказать: «А здесь нам поможет теория вероятностей», — и не облажаться.

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

- ❖ **1: Непредсказуемость:** исход ситуации нельзя предсказать абсолютно, то есть без какой-либо погрешности. В ином случае использовать теорию вероятностей — что по воробьям из пушки палить!
- ❖ **2: Воспроизводимость:** у испытателей есть возможность (хотя бы теоретическая) повторить ситуацию сколь угодно много раз при неизменных условиях.
- ❖ **3: Устойчивость частот:** при многократном повторении ситуации частота исхода, то есть отношение числа опытов, в которых мы получили этот исход, к их общему числу, должна колебаться около некоторого значения, приближаясь к нему все ближе и ближе. Другими словами, частота исхода должна иметь предел при стремлении числа опытов к  $\infty$ .

А где мы можем встретить такие ситуации, а? Правильно! В казино! Исторически теория вероятности создавалась, чтобы играть в азартные игры с опорой на науку, отсюда и типичные задачи: бросаем кости, гоняем шары в лототроне, делим ставку и т. д.

Чтобы превратить бытовую ситуацию в математическую науку, необходимо внести некий математический эквивалент словам «а ему чаще везет». И таким математическим объектом будет функция вероятности. Её область определения называется *множеством событий*, то есть это некое математическое описание результатов, которые вы можете получить. В свою очередь, событие (такое как, например, выпадание чётного числа на кубике) может являться совокупностью неких более простых событий, описывающих стохастическую ситуацию (число, выпавшее на кубике). Последнее множество называется *множеством элементарных исходов* и обозначается  $\Omega$ .

Есть разные теории описывающие вероятностные пространства, но самой используемой является теория, созданная Колмогоровым в начале прошлого века. Перейдем к ее описанию, то есть к нестрогой формализации понятий: множество событий, множество элементарных исходов.

### 1.2. Измеримое пространство

Здесь только суть. Быстрым темпом! Спринт по началам теории меры, которые и приведут нас к определению вероятностного пространства. Строго излагать их мы,

конечно, не будем. Поэтому любители обмазываться формальностями могут пройти сюда — [2].

Вероятностное пространство, которое мы дальше определим, должно содержать в себе описание трех позиций, важных в эксперименте: во-первых, оно должно содержать множество элементарных исходов  $\Omega$ , чтобы было предельно понятно, что является исходом; во-вторых, на этом множестве мы будем рассматривать некоторую структуру подмножеств, которая и будет множеством событий, которые имеют для нас смысл (т.е. если мы бросаем кубик и нам важна четность выпавшего числа, то не имеет смысла рассматривать исход «выпало 1 очко» в отдельности от исходов «выпало 3 очка» и «выпало 5 очков»); в-третьих, на этом множестве мы зададим функцию, которая и будет вероятностью, то есть некоторой характеристикой, описывающей частоту того или иного исхода.

Множество событий, обозначаемое  $\mathcal{A}$  (используют наравне с этим еще обозначение  $\mathcal{S}$ ), 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 событием («что-нибудь да произойдет»).

Если мы посадим эти три свойства в одну лодку с Математикой, то получим следующее

**Определение 1.1.** Семейство  $\mathcal{A}$  подмножеств множества  $\Omega$  называется алгеброй, если верны три аксиомы:

- (i)  $\forall A \in \mathcal{A}, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$  (если объединить два множества из семейства, то обязательно получим множество из семейства),
- (ii)  $\forall A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$  (множество и его дополнение одновременно: либо лежат в нашем семействе, либо нет),
- (iii)  $\Omega \in \mathcal{A}$  (все множество точно лежит в  $\mathcal{A}$ ).

↳ Из аксиом алгебры и формулы  $A \cap B = \overline{\overline{A} \cup \overline{B}}$  следует, что пересечение событий является событием. Подобным образом можно показать, что и разность, и симметрическая разность двух множеств из семейства  $\mathcal{A}$  тоже дадут множество из  $\mathcal{A}$ .

**Пример.** Наименьшей алгеброй является  $\{\Omega, \emptyset\}$ .

Косяк определения, которое мы только что дали, в том, что часто встречаются бесконечные множества элементарных исходов, поэтому полезно, чтобы множество событий было замкнуто не только относительно объединения, но и относительно счетного объединения.



**Определение 1.2.** Семейство  $\mathcal{A}$  подмножеств множества  $\Omega$  называется  $\sigma$ -алгеброй (читается как сигма-алгебра), если

- (i)  $\forall A_1, \dots, A_n, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ ,
- (ii)  $\forall A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ ,
- (iii)  $\Omega \in \mathcal{A}$ .

Не всегда по задаче понятно, какую сигма-алгебру надо выбрать, однако большой бонус ее аксиом заключается в том, что мы можем пересекать различные сигма-алгебры и получать снова сигма-алгебру. Из-за этого, если нам известно неполное множество событий  $\mathcal{K}$ , т.е.  $\mathcal{K} \subseteq \mathcal{A}$ , то мы можем восстановить сигма-алгебру с точностью до тех событий, которых нет в  $\mathcal{K}$ . Иными словами, мы можем найти такую минимальную сигма-алгебру, которая содержит в себе  $\mathcal{K}$  как подмножество.

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

**Пример.**  $\sigma$ -алгеброй, порожденной  $\mathcal{K} = A$  (класс состоит из одного элементарного события), будет являться  $\sigma(A) = \{\emptyset, A, \bar{A}, \Omega\}$ .

Сигма-алгебра является более узким понятием, нежели алгебра, то есть любая  $\sigma$ -алгебра является алгеброй, а обратное, вообще говоря, неверно. То есть это такая «элитка» среди всевозможных алгебр: не всем алгебрам дано быть «элиткой».

**Пример.** Пусть  $\Omega = \mathbb{R}$ ,  $\mathcal{A}$  содержит конечные подмножества  $\Omega$  и их дополнения. Для такого множества выполнены все аксиомы алгебры:  $\Omega = \bar{\emptyset} \in \mathcal{A}$ , 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,  $\mathcal{A}$  является алгеброй. Все элементы  $\mathcal{A}$  либо конечны, либо континуальны, поэтому  $\mathcal{A}$  не содержит  $\mathbb{N}$ . Но  $\mathbb{N} = \bigcup_{i=1}^{\infty} \{i\}$ , то есть не выполнено свойство счетной аддитивности из определения  $\sigma$ -алгебры. Вот видите: это не достойное нашего внимания семейство, оно не настолько элитное, чтобы называться сигма-алгеброй.

Вот мы и подошли к первому основополагающему определению.

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

По сути измеримое пространство нам говорит о том, какие у нас есть элементарные исходы и что мы считаем за событие. Но если при бросании кости ты будешь использовать кость со смещенным центром тяжести, то игроки что-то заподозрят и



популярно пояснят тебе, что ты неправ (поверь, тебе это не понравится), так что имеет смысл определить не только множество событий, но и вероятность того или иного события. Мы этим и займемся дальше.

### 1.3. Вероятность. Вероятностное пространство

Ну чё, народ, погнажи на!.. На новые баррикады! Они не такие высокие, как те, которые мы уже перепрыгнули, так что следующие пару страниц вам покажутся легкой прогулкой теплым осенним вечером по желтеющему бульвару, залитому солнечными лучами заходящего солнца... Кхм-Кхм... Это из другой пьесы. Так о чем это мы? Да, теперь пора определить вероятность.

**Определение 1.5.** Пусть  $(\Omega, \mathcal{A})$  — измеримое пространство ( $\mathcal{A}$  —  $\sigma$ -алгебра). Вероятностью называется функция  $\mathbb{P}: \mathcal{A} \rightarrow \mathbb{R}$ , удовлетворяющая свойствам

1.  $\forall A \in \mathcal{A} \mapsto \mathbb{P}(A) \geq 0$ ,
2.  $\forall A_1, \dots, A_n, \dots \in \mathcal{A}, \quad A_i \cap A_j = \emptyset \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ ,
3.  $\mathbb{P}(\Omega) = 1$ .

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.

$$\forall B_1, B_2, \dots \in \mathcal{A}, \quad B_{n+1} \subseteq B_n \Rightarrow \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}(B), \quad B = \bigcap_{n=1}^{\infty} B_n$$

то они вместе будут эквивалентны 2 из определения вероятности. Покажем это.

**Утверждение 1.6.**

$$\left[ \forall A_1, \dots, A_n, \dots \in \mathcal{A}, \quad A_i \cap A_j = \emptyset \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \right] \Leftrightarrow$$

$$\left[ \forall A_1, \dots, A_n \in \mathcal{A} \quad A_i \cap A_j = \emptyset \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) \right] \wedge$$

$$\left[ \forall B_1, B_2, \dots \in \mathcal{A}, \quad B_{n+1} \subseteq B_n \Rightarrow \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}(B), \quad B = \bigcap_{n=1}^{\infty} B_n \right].$$

**Доказательство.**

$\Rightarrow$

Обозначим  $C_n = B_n \setminus B_{n+1}$ . Множества  $B, C_1, C_2, \dots$  не имеют общих точек.

$\forall n \quad B_n = \left( \bigcup_{k=n}^{\infty} C_k \right) \cup B$ . Тогда  $\mathbb{P}(B_1) = \mathbb{P}(B) + \sum_{k=1}^{\infty} \mathbb{P}(C_k)$ . Отсюда следует, что ряд в

правой части сходится, так как имеет конечную сумму.  $\mathbb{P}(B_n) = \mathbb{P}(B) + \sum_{k=n}^{\infty} \mathbb{P}(C_k)$ .  
 При  $n \rightarrow \infty$  сумма ряда стремится к нулю как остаточный член ряда из предыдущего выражения. В предельном переходе получаем свойство непрерывности.

⇐

Рассмотрим произвольный набор  $A_1, A_2, \dots \in \mathcal{A}$   $A_i \cap A_j = \emptyset$ .

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

Обозначим  $B_n = \bigcup_{i=n+1}^{\infty} A_i$ ,  $B_{n+1} \subseteq B_n$   $\forall n$ ,  $\bigcap_{n=1}^{\infty} B_n = \emptyset$

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \lim_{n \rightarrow \infty} \left( \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) - \mathbb{P}(B_n) \right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) - \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right). \blacksquare$$

**Определение 1.7.** Вероятностным пространством  $(\Omega, \mathcal{A}, \mathbb{P})$  называется измеримое пространство  $(\Omega, \mathcal{A})$ , с заданной на нём вероятностью  $\mathbb{P}$ .

Ничего сложного, если знать определения и уметь вертеть математическими понятиями и логикой. Главное не перепутайте: вертеть, а не класть.

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Кому вообще нужна  $\sigma$ -алгебра событий и зачем весь этот огород и маленькая тележка, если можно рассматривать множество всех подмножеств множества событий  $\Omega$ ? Когда-то давно кто-то доказал, что в случае очень большого множества элементарных исходов, например, континуального (это много, поверьте наслову), множество  $2^{\Omega}$  будет иметь такую крокодильски, или алигаторски, большую мощность, что вся теория сломается. Таким образом, алгебры нужны для того, чтобы вероятность имела хорошую область определения.

Перечислим свойства вероятности. Доказательства их можно найти в любом из классических учебников по теории вероятностей или можно их придумать самому: большинство из них тривиальны. Да, реально тривиальны, а не как на матеше в школе.

**Свойства вероятности:**

1.  $\mathbb{P}(\emptyset) = 0$ ,
2.  $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$ ,
3.  $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ ,
4.  $\mathbb{P}(A) \leq 1$ ,
5.  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$ ,
6.  $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ ,
7.  $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \sum_{i_1 < \dots < i_k} (-1)^{k+1} \mathbb{P}(A_{i_1} A_{i_2} \dots A_{i_k})$ ,
8.  $\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n \mathbb{P}(\overline{A_i})$  — неравенство Бонферрони.

Примите к факту, что в силу простоты этих определений, математики используют их на уровне интуиции, то есть вы можете не заметить, а при этом в паре строк уже будет использовано несколько этих свойств. В некотором роде это и достоинство теории вероятности, так как происходит подобное из-за ясной бытовой аналогии, навязанной всем распространенностью азартных игр. С другой стороны, есть отличный способ избежать возможности быть посаженным на что-нибудь за отсутствие формализма: для этого стоит каждый раз, как перед тобой будут выписаны математические выкладки, пробовать осмыслить каждый переход, который в них сделан. Если это не получается сделать, то не стремись хвататься за голову: возможно, причина в том, что используются какие-то свойства, которые автор посчитал очевидными. Тогда стоит поискать в других источниках доказательства этого факта. Вполне вероятно, что это поможет найти такое доказательство, в котором будет все предельно понятно или в тяжелых моментах будет указано, что стоит перечитать и повторить.

В заключение вырежем клинышком в этой глиняной массе

**Теорема 1.8** (Каратеодори). Пусть  $(\Omega, \mathcal{A})$  — измеримое пространство в широком смысле, а некоторая функция  $\mathbb{P}$  обладает свойствами вероятностной меры. Тогда на измеримом пространстве  $(\Omega, \sigma(\mathcal{A}))$

$$\exists! \mathbb{P}': \forall A \in \mathcal{A} \mapsto \mathbb{P}(A) = \mathbb{P}'(A).$$

**Доказательство.** Ушло и не вернулось. Последний раз его видели здесь [2]. ■

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



## §2. Понятие меры и интеграла Лебега

### 2.1. Мера Лебега

Не пугайтесь. Сейчас мы введем меру Лебега, но не так быстро, как вводили вероятностное пространство. Представьте, что задание меры Лебега — это как надуть несколько воздушных шариков внутри друг друга. Самый маленький шарик — это полуинтервалы. Средний — борелевские множества (о них будет дальше). Большой — измеримые множества на прямой. На каждом из этих множеств вводится своя мера, а потом продолжается на большие множества.

Дадим определение семейства, которое нас будет сильно интересовать дальше.

**Определение 2.1.** Борелевской  $\sigma$ -алгеброй называется минимальная  $\sigma$ -алгебра, содержащая все открытые подмножества топологического пространства. Элементы борелевской  $\sigma$ -алгебры называются борелевскими множествами.

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** 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 интервалы входят по определению.
2. Отрезок вида  $[a, b]$  входит как  $\overline{(-\infty, a) \cup (b, +\infty)}$ .
3. Точка ходит как вырожденный отрезок  $[a, a]$ .
4. Счетное объединение таких множеств входит по определению.

**Определение 2.2.** Теперь вспомним сначала аксиомы счетно-аддитивной меры  $\mu$  из функана (злодейский смех): это функция на множестве  $B$ ,  $\mu: B \rightarrow \mathbb{R}$ , со свойствами

1.  $\forall A \in B \mapsto \mu(A) \geq 0$  (неотрицательность),
2.  $\forall A_1, A_2, \dots \in B: A_i \cap A_j = \emptyset \ (i \neq j) \Rightarrow \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$  ( $\sigma$ -аддитивность).

Начнем использовать это определение в самом простом случае. Введем сначала меру на борелевских множествах полуинтервала  $[0, 1)$ , то есть на  $\mathcal{B}_{[0,1)}$ . Итак,

- На интервалах  $(a, b)$  введем меру как  $\mu_{[0,1)}((a, b)) = b - a$ , то есть примерно если бы ваш сосед пришел бы с рулеткой к прямой и, померив, грубым голос добавил: «Здесь на вскидку  $b - a$  получается.»
- Мера одной точки равна нулю, и мы можем не обращать внимание на концы множеств. Это так пусть будет по определению.

- На любых множествах, которые можно представить как конечное объединение интервалов и какого-то числа точек, мы зададим меру как сумму мер каждого из составляющих. Это как если бы вам надо было добраться по проспекту от точки  $A$  до точки  $B$ : прошли до остановки метров 100, сели в автобус, проехали пару километров, вышли, пересели в другой автобус, проехали на нем, вышли, и прошли до места назначения метров 200. Тогда множество точек проспекта (прямой), где вы шли или ждали автобус, будет иметь меру (длину) равную 300 метрам:

$$300 = 100 + 0 + 200.$$

- Теперь осталось определить меру на бесконечных объединениях/пересечениях. Для этого воспользуемся теоремой Каратеодори (1.8), согласно которой можно продолжить меру на  $\sigma$ -алгебру.

Аналогичным образом определим меру  $\mu_{[k, k+1)}$  на всех полуинтервалах с целыми концами  $[k, k+1)$ , ( $k \in \mathbb{Z}$ ). Осталось доопределить меру на всей  $\mathcal{B}$ .

$$\forall A \in \mathcal{B} \quad \mu(A) = \sum_{i=-\infty}^{\infty} \mu_{[i, i+1)}(A \cap [i, i+1)).$$

## 2.2. Интеграл Лебега

Теперь введем *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  $f: \mathbb{R} \rightarrow \mathbb{R}$  такие, что  $\forall c \in \mathbb{R}$  множество  $A_c \{x: f(x) < c\}$  борелевское, то есть прообразы полупрямых являются борелевскими множествами.

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  $f(x) = \sum_{i=1}^k y_i \mathbb{1}(A_i)$ , где  $A_i$  не пересекаются и покрывают весь полуинтервал  $[0, 1)$ . Функции такого вида называются *финитными*. В этом случае

$$\int_0^1 f(x) \mu(dx) = \sum_{i=1}^k y_i \mu(A_i).$$

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  $[\frac{k-1}{2^n}, \frac{k}{2^n})$ ,  $k = 1, 2, \dots, n2^n$ . 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  $n$ . Устремим  $n$  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,

$$\int_0^1 f(x) \mu(dx) = \lim_{n \rightarrow \infty} \left[ \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mu \left( \left\{ x : \frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n} \right\} \right) + n \mu(\{x : f(x) \geq n\}) \right].$$

- 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:

$$\int_0^1 f(x) \mu(dx) = \int_0^1 f^+(x) \mu(dx) - \int_0^1 f^-(x) \mu(dx).$$

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

$$\int_A f(x) \mu(dx) = \sum_{i=-\infty}^{+\infty} \int_{A \cap [i, i+1)} f(x) \mu(dx).$$

**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 пункту

$$\int_0^1 D_{[0,1)}(x) \mu(dx) = 1.$$

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

Все. На этом введение в теорию меры заканчивается. Перейдем к более интересным разделам: связанным уже с вероятностями.





## §3. Связь между событиями

### 3.1. Условная вероятность

Условно мы будем тут работать с вероятностью нескольких событий. Ведь бывают разные игры: и в карты, и в кости. Есть и такие, в которых придется не единожды просить судьбу «выдать козырь», поэтому выгодно понимать, сколько в действительности у тебя шансов дёрнуть судьбу так, как тебя тогда на рынке дёрнули.

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

**Определение 3.1.** Условной вероятностью события  $A \in \mathcal{A}$  при условии  $B$  называется

$$\frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \stackrel{\text{def}}{=} \mathbb{P}(A|B) = \mathbb{P}_B(A).$$

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** 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.

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

**Пример.** Представь, что ты бросаешь 2 кубика. Событие  $A$  — у тебя выпало суммарно меньше 5 очков. А  $B$ , например, означает, что у тебя выпало 6 очков на первом кубике. Очевидно, что если на одном выпало 6, то суммарно у тебя меньше 6 быть не может, так что вероятность тоже должна быть равна нулю, что согласуется с определением:

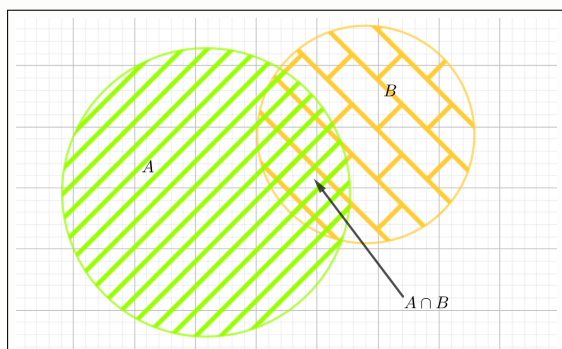
$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{0}{P(B)} = 0.$$

Допустим теперь, что  $B$  — у тебя выпало на первом кубике 1 очко. Тогда понятно, что вероятность события  $A$  — будет равна  $1/2$ , так как при выпадении 1, 2, 3 на втором кубике событие произойдет, а при выпадении других чисел — нет. Вычислим теперь по определению.

Вероятность выпадения 1 на первом кубике, притом чтобы сумма была меньше пяти, равна  $1/12$ , так как всего 36 комбинаций на кубиках, а подходит только 3:  $(1, 1), (1, 2), (1, 3)$ . Вероятность выпадения любого из чисел на кубике —  $1/6$ , так что

$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{1/12}{1/6} = \frac{1}{2}.$$

Опа! Америка-Европа! Все сошлось с интуитивными предположениями.



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  $A \cap B$ .

Из определения следует так называемый «Закон умножения вероятностей»:

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

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

### 3.2. Независимость событий. Независимость в совокупности

Да, и в теории вероятности есть независимость! Но это уже другая независимость, а не та, про которую так много новостей нынче.

**Определение 3.2.** События  $A, B \in \mathcal{A}$  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 ( $A \cap B = \emptyset$ ) независимыми? Ответ: нет, пусть  $A, B \in \mathcal{A} : \mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ . 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 вероятность равна 0, а не вероятности самого события, что требуется для независимости.

Обобщим понятие независимости на произвольное количество событий.

**Определение 3.3.** События  $A_1, A_2, \dots, A_n$  называются независимыми в совокупности, если

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

**Пример.** На примере тетраэдра Бернштейна можно убедиться в том, что попарной независимости событий недостаточно для независимости в совокупности. Рассмотрим 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  $\{С\} \cap \{К\}$  равна вероятности выпадения четвертой грани, т. е.  $1/4$ , в то время как вероятность выпадения каждого цвета равна  $1/2$ ). Однако

$$\mathbb{P}(\{С\} \cap \{К\} \cap \{З\}) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3.$$

### 3.3. Формула полной вероятности. Формула Байеса

Чуть выше мы рассмотрели условную вероятность, то есть поговорили о том, что некоторые события могут пересекаться и когда в таких случаях можно сказать, что они независимы. Часто встречается следующая группа событий.

**Определение 3.4.**  $B_1, \dots, B_n$  образуют полную группу, если выполнены следующие условия:

1.  $\mathbb{P}(B_i) > 0 \quad \forall i = 1, \dots, n,$
2.  $B_i B_j = \emptyset \quad (i \neq j),$
3.  $\bigcup_{i=1}^n B_i = \Omega.$

**Теорема 3.5.** Пусть  $B_1, \dots, B_n$  образуют полную группу. Вероятность события  $A \in \mathcal{A}$  можно вычислить по *формуле полной вероятности*:

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

**Доказательство.**

$$A = \bigcup_{i=1}^n AB_i, \quad AB_i \cap AB_j = \emptyset \quad (i \neq j),$$

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

Последний переход следует из закона умножения вероятностей. ■

Первое требование определения полной группы необходимо для возможности определить условную вероятность, второе позволяет разбить множество  $A$  на непересекающиеся части. Третье требование, вообще говоря, можно ослабить, потребовав, чтобы  $A \subseteq \bigcup_{i=1}^n B_i$ . 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 вышеперечисленное, несложно по формуле вычислить вероятность успешной сдачи.

Формула полной вероятности используется для вычисления априорной вероятности, т.е. вероятности события, которое еще не произошло. Пусть теперь  $\mathbb{P}(A) > 0$ . Тогда

$\mathbb{P}(B_i|A) = \mathbb{P}(AB_i)/\mathbb{P}(A)$ . Используя формулу полной вероятности, получаем *формулу Байеса*:

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

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

### 4.1. Случайные величины: определение

Случайные события — это хорошо, но 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, которое выглядит так:

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

где  $\Omega$  — множество элементарных событий,  $\mathcal{A}$  —  $\sigma$ -алгебра подмножеств множества элементарных событий, а  $\mathbb{P}$  — вероятность. Мы же будем рассматривать теперь тройку

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

где  $\mathbb{R}$  — действительная прямая,  $\mathcal{B}$  — борелевская  $\sigma$ -алгебра, а  $\mathbb{P}$  — вероятность.

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

**Определение 4.1.** Пусть  $(\Omega, \mathcal{A}, \mathbb{P})$  — вероятностное пространство. Тогда случайной величиной  $\xi$  называется функция  $\xi : \Omega \rightarrow \mathbb{R}$ , измеримая относительно  $\mathcal{A}$  и  $\mathcal{B}$ . По-другому,  $\xi$  — случайная величина, если

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

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** 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.

### 4.2. Функция распределения

Введем еще несколько «полезных» определений.

**Определение 4.2.** С каждой случайной величиной свяжем два вероятностных пространства: первое —  $(\Omega, \mathcal{A}_\xi, \mathbb{P})$  — вероятностное пространство, порожденное  $\xi$ . Здесь  $\mathcal{A}_\xi$  — наименьшая  $\sigma$ -алгебра, для которой выполняется свойство измеримости. Второе —  $(\mathbb{R}, \mathcal{B}, \mathbb{P}_\xi)$ , где  $\mathbb{P}_\xi(B) = \mathbb{P}(\xi^{-1}(B)) \quad \forall B \in \mathcal{B}$  и называется распределением вероятностей  $\xi$ .

Идем дальше в сторону упрощения работы со случайностями. Вместо того чтобы рассматривать произвольные борелевские множества, мы будем рассматривать 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  $\mathcal{B}$ .

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

**Утверждение 4.3.** Пусть  $\xi, \eta$  — случайные величины. Тогда множество  $\{\omega : \xi(\omega) < \eta(\omega)\}$  является событием.

**Доказательство.**  $\{\omega : \xi(\omega) < \eta(\omega)\} = \bigcup_{r \in \mathbb{Q}} \{\omega : \xi(\omega) < r, \eta(\omega) > r\}$ . Заметим, что  $\{\omega : \xi(\omega) < r\}$  является событием. Аналогично для  $\eta$ . 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  $\xi_1, \dots, \xi_n$  — случайные величины, а функция  $\varphi(x_1, \dots, x_n)$  является непрерывной на множестве их значений, то  $\varphi(\xi_1, \dots, \xi_n)$  будет случайной величиной.

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

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

Запись  $\mathbb{P}(\xi < x)$  является в некотором смысле жаргонной, так как аргументов вероятности должно быть событие из  $\mathcal{A}$ . Но  $\xi < x$  мы в дальнейшем будем отождествлять с объединением 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.  $\lim_{x \rightarrow -\infty} F_\xi(x) = 0$
3.  $\lim_{x \rightarrow +\infty} F_\xi(x) = 1$
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  $b \rightarrow a$  получим  $\mathbb{P}(\xi = a) = F_\xi(a + 0) - F_\xi(a)$ , то есть вероятность попадания в точку равна скачку функции распределения в этой точке.

**Определение 4.5.** Точка  $x_0$  называется точкой роста  $F_\xi(x)$ , если  $\forall \varepsilon > 0$   
 $\mathbb{P}(x_0 - \varepsilon \leq \xi < x_0 + \varepsilon) > 0$

**Пример.** Это очень полезный пример, который будет использоваться в матстате и который очень любят спрашивать. Пусть  $\xi$  — случайная величина.  $\eta = F_\xi(\xi)$ . Чему равна  $F_\eta(x)$ ? По определению

$$F_\eta(x) = \mathbb{P}(\eta < x) = \mathbb{P}(F_\xi(\xi) < x) = \mathbb{P}(\xi < F_\xi^{-1}(x)) = F_\xi(F_\xi^{-1}(x)) = x \quad (4.1)$$

Вообще, тут было бы неплохо сказать, что  $F_\xi$  непрерывна и строго монотонна, чтобы со спокойной совестью использовать обратную функцию. Таким образом  $\eta$  имеет равномерное распределение.

### 4.3. Виды распределений

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

**Определение 4.6.** Случайная величина  $\xi$  называется абсолютно непрерывной, если существует интегрируемая функция  $p_\xi(x) \geq 0$ ,  $x \in \mathbb{R}$  такая, что функция распределения  $\xi$  является почти всюду (за исключением не более, чем счетного числа точек) дифференцируемой функцией и представима в виде

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

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

**Определение 4.7.** Случайная величина  $\xi$  называется дискретной, если множество точек роста не более, чем счетно, но распределение не является сингулярным, или, другими словами  $\exists B = \{x_1, x_2, \dots\} : \mathbb{P}(\xi \in B) = 1$ .

**Определение 4.8.** Случайная величина  $\xi$  называется сингулярной, если  $F_\xi$  непрерывна, и  $\exists B \in \mathcal{B} : \mu(B) = 0$ ,  $\mathbb{P}(\xi \in B) = 1$ , 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.

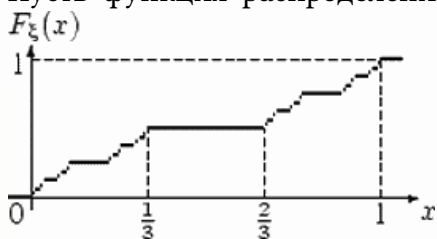


Плотность же отражает вероятность попасть в ту или иную область: интеграл по области равен этой вероятности. Дискретная случайная величина принимает конечное или счетное множество значений, вследствие этого имеет ступенчатую функцию распределения, например, бросок монетки имеет дискретное распределение. Сингулярное распределение — это крокодил, который не встречается в жизни и будет рассмотрен отдельно.

**Утверждение 4.9.** 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 соответствующих (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.



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

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 4.10 (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.** Не вошло в содержание и не вышло с публикацией. Ищите в других учебниках. ■

#### 4.4. Векторные случайные величины

Векторную случайную величину можно определить двумя способами:

1. Сказать, что  $\xi = (\xi_1, \dots, \xi_n)$  является векторной случайной величиной, если ее координаты являются случайными величинами.
2. Дать определение аналогично определению одномерной случайной величины, то есть рассмотреть  $(\mathbb{R}^n, \mathcal{B}_n)$ , где  $\mathcal{B}_n$  —  $n$ -мерная борелевская  $\sigma$ -алгебра, то есть минимальная  $\sigma$ -алгебра, содержащая все параллелепипеды (аналогично интервалам в одномерном случае). Тогда  $\xi$  — случайная величина, если

$$\forall B \in \mathcal{B}_n: \xi^{-1}(B) \in \mathcal{A}.$$

**Определение 4.11.** Функция распределения векторной случайной величины

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

Свойства функции распределения:

1. Пусть  $x, y \in \mathbb{R}^n$  такие, что  $x_i < y_i \quad i = 1, \dots, n$ . Тогда  $F_\xi(x) \leq F_\xi(y)$ .
2.  $0 \leq F_\xi(x) \leq 1$ .
3.  $\lim_{x_i \rightarrow -\infty} F_\xi(x) = 0$ . То есть при стремлении одной координаты к  $-\infty$  функция распределения стремится к 0, поскольку вероятность для соответствующей координаты стремится к 0.
4. Если же какую-то из координат устремить к бесконечности, то она не будет учитываться в вероятности, поэтому

$$\lim_{x_i \rightarrow +\infty} F_\xi(x) = F_{(\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n)}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

5. Непрерывна слева по каждому аргументу.

**Определение 4.12.** Абсолютно непрерывная случайная величина  $\xi$  — такая случайная величина, что

$$\forall B \in \mathcal{B}_n \quad \mathbb{P}(\xi \in B) = \int_B p_\xi(x_1, \dots, x_n) dx_1 \dots dx_n.$$

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

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

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

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

## 4.5. Условное математическое ожидание

### 4.5.1. Конструктивные определения

**Зачем?** Уберите от экрана детей, беременных женщин и людей со слабой нервной системой. Сейчас будем вводить условное математическое ожидание. Для начала надо понять, а где тут вообще проблема и почему нельзя ввести условную плотность как отношение плотностей и тупо брать по ней интеграл. Во-первых, что такое матожидание? Матожидание — это простое обычное число. А теперь рассмотрим две случайные величины:  $\xi$  и  $\eta$  и условное матожидание  $\mathbb{E}(\xi|\eta)$ . Пусть для наглядности  $\xi$  — число, выпавшее на игральной кости, а  $\eta$  — остаток от деления на 2 выпавшего числа. Тогда матожидание выпавшего числа будет равно 4, если известно, что выпало четное число, и 3 — в противном случае. То есть матожидание зависит от условия, которое является случайной величиной, а значит и само матожидание является случайной величиной. Таким образом, условное математическое ожидание мы не можем тупо определить как интеграл от условной плотности, так как интеграл — это число. Поэтому мы введем условное матожидание (УМО) отдельно для дискретных и непрерывных случайных величин, и в каждый раз будем вводить этого крокодила в 2 этапа. Сначала введем  $\mathbb{E}(\xi|\eta = y)$  — матожидание относительно конкретной реализации случайной величины. Это будет какая-то функция от  $y$ . Потом скажем, что если в эту функцию подставить случайную величину  $\eta$ , то мы получим  $\mathbb{E}(\xi|\eta)$  — УМО относительно случайной величины. А потом еще для расширения сознания введем УМО относительно сигма-алгебры и введем все то же, но по-другому, аксиоматически, а не конструктивно.

Пусть  $\xi$  принимает значения  $\{x_i\}$ , а  $\eta - \{y_i\}$ , тогда

$$\textbf{Определение 4.13. } \mathbb{E}(\xi|\eta = y_j) := \sum_i x_i \mathbb{P}(\xi = x_i | \eta = y_j) = \sum_i x_i \frac{\mathbb{P}(\xi = x_i, \eta = y_j)}{\mathbb{P}(\eta = y_j)} = f(y), \quad \text{где } y = y_j.$$

**Определение 4.14.**  $\mathbb{E}(\xi|\eta) = f(\eta)$ , где  $f$  — функция, полученная с помощью предыдущего определения.

Так как УМО — случайная величина, то мы можем взять матожидание этой случайной величины и посмотреть, а что будет.

$$\begin{aligned} \forall B \in \mathcal{B} \quad \mathbb{E}[\mathbb{E}(\xi|\eta) \mathbb{1}(\eta \in B)] &= \sum_{j: y_j \in B} f(y_j) \mathbb{P}(\eta = y_j) = \\ &= \sum_{j: y_j \in B} \left[ \sum_k x_k \frac{\mathbb{P}(\xi = x_k, \eta = y_j)}{\mathbb{P}(\eta = y_j)} \right] \mathbb{P}(\eta = y_j) = \\ &= \sum_{j: y_j \in B} \sum_k \mathbb{P}(\xi = x_k, \eta = y_j) = \mathbb{E}[\xi \mathbb{1}(\eta \in B)]. \quad (4.2) \end{aligned}$$

И в частности, если в качестве  $B$  выбрать  $\mathbb{R}$ , то индикатор всегда будет равен 1 и тогда  $\mathbb{E}[\mathbb{E}(\xi|\eta)] = \mathbb{E}\xi$ .

Теперь рассмотрим абсолютно непрерывные случайные величины  $\xi, \eta \sim p_{\xi,\eta}(x, y)$ . Попробуем подступиться так же, как и в дискретном случае, через условную вероятность.

$$\mathbb{P}(\xi < x | \eta = y) = \frac{\mathbb{P}(\xi < x, \eta = y)}{\mathbb{P}(\eta = y)} = \frac{0}{0}. \quad (4.3)$$

Проблемка. Попробуем тогда у условия сказать, что  $\eta \in [y, y + \varepsilon)$ , и устремить  $\varepsilon$  к нулю.

$$\begin{aligned} \mathbb{P}(\xi < x | y \leq \eta < y + \varepsilon) &= \frac{\mathbb{P}(\xi < x | y \leq \eta < y + \varepsilon)}{\mathbb{P}(y \leq \eta < y + \varepsilon)} = \\ &= \frac{\int_{-\infty}^x \int_y^{y+\varepsilon} p_{\xi,\eta}(x, y) dy dx}{\int_y^{y+\varepsilon} p_{\eta}(y) dy} = \left\{ \begin{array}{l} \text{считаем, что} \\ \text{можно приме-} \\ \text{нить т. о среднем} \end{array} \right\} = \frac{\int_{-\infty}^x p_{\xi,\eta}(x, y') dx}{\int_y^{y+\varepsilon} p_{\eta}(y'') dy} = \\ &= \int_{-\infty}^x \frac{p_{\xi,\eta}(x, y')}{p_{\eta}(y'')} dx \xrightarrow{\varepsilon \rightarrow 0} \int_{-\infty}^x \frac{p_{\xi,\eta}(x, y)}{p_{\eta}(y)} dx. \quad (4.4) \end{aligned}$$

Теперь можем вводить определение условной плотности:

**Определение 4.15.**  $p_{\xi|\eta=y}(x) = \frac{p_{\xi,\eta}(x, y)}{p_{\eta}(y)}$ .

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Условная плотность — функция от  $x$ , но параметризованная  $y$ .

**Определение 4.16.**  $\mathbb{E}(\xi|\eta = y) := \int x p_{\xi|\eta=y}(x) dx = f(y)$ .

Теперь, как и в дискретном случае, конструктивно введем определение УМО относительно случайной величины.

**Определение 4.17.**  $\mathbb{E}(\xi|\eta) := f(\eta)$ .

И снова посчитаем МО УМО:

$$\begin{aligned} \forall B \in \mathcal{B} \quad \mathbb{E}[\mathbb{E}(\xi|\eta) \mathbb{1}(\eta \in B)] &= \mathbb{E}[f(\eta) \mathbb{1}(\eta \in B)] = \int_B f(y) p_{\eta}(y) dy = \\ &= \int_B \int x \frac{p_{\xi,\eta}(x, y)}{p_{\eta}(y)} dx p_{\eta}(y) dy = \int_B \int x p_{\xi,\eta}(x, y) dx dy = \mathbb{E}[\xi \mathbb{1}(\eta \in B)]. \quad (4.5) \end{aligned}$$

## 4.5.2. Аксиоматическое определение

Теперь потихоньку будем вводить УМО через всякие там приколы с мерой. Вспомним наше любимое вероятностное пространство  $(\Omega, \mathcal{A}, \mathbb{P})$  и рассмотрим некоторую под- $\sigma$ -алгебру  $\mathcal{F} \subseteq \mathcal{A}$ .

**Определение 4.18.** Случайная величина  $\eta$  называется  $\mathcal{F}$ -измеримой, если

$$\forall B \in \mathcal{B} \quad \eta^{-1}(B) = \{\omega : \eta(\omega) \in B\} \in \mathcal{F}.$$

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Поясним: рассмотрим игральную кость. Тогда  $\Omega = \{1, \dots, 6\}$ ,  $\mathcal{A} = \{2^\Omega\}$ . Теперь положим  $\mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ .

Теперь рассмотрим случайную величину  $\eta$ , которая равна чётности выпавшего числа. По определению легко проверяется, что  $\eta$  является  $\mathcal{F}$ -измеримой.

**Определение 4.19.** Условным математическим ожиданием случайной величины  $\xi$  относительно  $\sigma$ -алгебры  $\mathcal{F}$  называется случайная величина  $\mathbb{E}(\xi|\mathcal{F})$ , обладающая следующими свойствами:

1.  $\mathbb{E}(\xi|\mathcal{F})$   $\mathcal{F}$ -измерима.
2.  $\forall A \in \mathcal{F} \quad \mathbb{E}[\mathbb{E}(\xi|\mathcal{F})\mathbb{1}(A)] = \mathbb{E}[\xi\mathbb{1}(A)]$

Введем теперь определение порожденной  $\sigma$ -алгебры.

**Определение 4.20.**  $\sigma(\eta) := \bigcup_{\forall B \in \mathcal{B}} \{\omega : \eta(\omega) \in B\}$ .

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Это объединение всех множеств, которые мы получили следующим образом: берем борелевское множество, смотрим, какие события туда «падают», если перегнать из событий в  $\mathbb{R}$ .

И теперь, когда мы определили порожденную  $\sigma$ -алгебру, можно вводить аксиоматические определения для УМО относительно случайных величин.

**Определение 4.21.**  $\mathbb{E}(\xi|\eta) := \mathbb{E}(\xi|\sigma(\eta))$ .

Вообще по понятиям, если вводится такого рода определения, то надо доказывать корректность таких определений. То есть мы определили какой-то новый объект, но еще не факт, что такие крокодилы водятся в природе. Мы тут этим заниматься не будем, чтобы не опухнуть от функана. Ключевые слова, чтобы казаться умным: мера, абсолютно непрерывная относительно другой меры, теорема Радона–Никодима, заряд.

## 4.5.3. Свойства и красивые картинки

Для начала запишем несколько свойств УМО (доказывать их я, конечно, не буду):

1.  $\forall B \in \mathcal{B} \quad \mathbb{P}(\xi \in B|\mathcal{F}) = \mathbb{E}[\mathbb{1}(\xi \in B)|\mathcal{F}]$ .
2.  $\mathbb{E}(a\xi_1 + b\xi_2|\mathcal{F}) \stackrel{\text{н.н.}}{=} a\mathbb{E}(\xi_1|\mathcal{F}) + b\mathbb{E}(\xi_2|\mathcal{F})$ .
3. Если  $\eta$   $\mathcal{F}$ -измерима, то  $\mathbb{E}(\xi\eta|\mathcal{F}) \stackrel{\text{н.н.}}{=} \eta\mathbb{E}(\xi|\mathcal{F})$ .



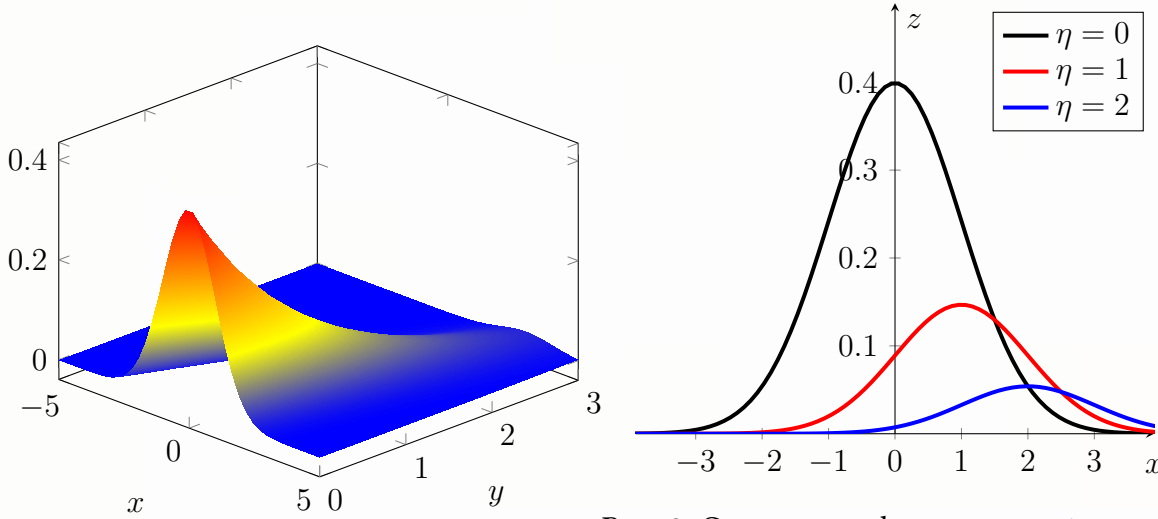


Рис. 1: Совместная плотность (4.6).

 Рис. 2: Сечения графика на рис. 1 при различных реализациях  $\eta$ .

4. Если  $\xi$  и  $\eta$  независимы, то  $\mathbb{E}(\xi|\eta) \stackrel{\text{п.н.}}{=} \mathbb{E}(\xi)$ .

5.  $\mathbb{E}[\mathbb{E}(\xi|\mathcal{F})] = \mathbb{E}\xi$ .

Теперь картинки. Мы придумаем какую-нибудь плотность  $p_{\xi, \eta}(x, y)$ , а потом построим график плотности условного матожидания и попытаемся понять его жизненный смысл.

Чтобы было легко и просто мы представим плотность в таком виде:

$$p_{\xi, \eta}(x, y) = p_{\xi|\eta}(x)p_{\eta}(y). \quad (4.6)$$

Затем положим:

$$p_{\xi|\eta}(x) = \mathcal{N}(\eta, 1) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \eta)^2}{2} \right\}, \quad (4.7)$$

$$p_{\eta}(y) = \exp(1) = e^{-y} \mathbb{1}(y \geq 0). \quad (4.8)$$

Теперь по понятиям. Что значит математическое ожидание  $\xi$  при условии  $\eta$ ? Это МО, если считать, что  $\eta$  уже свалилось к нам с неба и не меняется. В силу построения условной плотности  $\mathbb{E}(\xi|\eta) = \eta$  просто из свойства нормального распределения. И полученная случайная величина, утверждается, что распределена экспоненциально. Попробуем увидеть это на графиках.

Теперь если мы будем рассматривать сечения этого графика плоскостями  $y = t, t \geq 0$ , то будем получать каждый раз почти графики плотности нормального распределения с матожиданием  $t$ . Почему почти? Потому что если проинтегрировать такой график, то получится число, меньшее 1, а именно равное вероятности (плотность имеет смысл вероятности) того, что матожидание будет именно таким. Так что в данном конкретном случае УМО можно увидеть на графике, рассматривая сечения при различных  $y$ . А вероятность попасть в то или иное УМО — это значение функции в максимуме при конкретном значении  $y$  с поправкой на множитель  $\frac{1}{\sqrt{2\pi}}$ . Если приглядеться, то можно увидеть, что вероятность попасть в матожидание убывает экспоненциально с

ростом  $y$ , и так оно и должно быть в силу выбора  $p_\eta(y)$ . Конечно, не всегда все так красиво, это просто жалкая попытка показать на картинке жизненный смысл всего этого безобразия.





## §5. Характеристики случайных величин

### 5.1. Математическое ожидание

*Математическое ожидание* (обозначается  $\mathbb{E}$ ) обобщает понятие среднего арифметического 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  $\mu$ .

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

$$\mathbb{E}\xi = \int_{\Omega} \xi(\omega) \mathbb{P}(d\omega) \equiv \sum_{i=1}^k x_i \mathbb{P}(\{\omega: \xi(\omega) = x_i\}) = \sum_{i=1}^k x_i \mathbb{P}(A_i).$$

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

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

- 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  $\xi$  индуцирует вероятностное пространство  $(\mathbb{R}, \mathcal{B}_{\xi}, \mathbb{P}_{\xi})$ . 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 математическое ожидание в более привычной форме

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

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  $\xi$ , 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. Ситуацию не спасет даже рассмотрение случайной величины  $\eta$ , принимающей значения  $\pm n$  с вероятностью  $\frac{c}{2n^2}$ , которая имеет среднее значение 0, поскольку  $\mathbb{E}\eta = \sum_{n=1}^{\infty} \frac{c}{2n} - \sum_{n=1}^{\infty} \frac{c}{2n}$ , что не определено, поскольку интегралы Лебега от  $\eta^+$  и  $\eta^-$  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  $\int_{-\infty}^{+\infty} x dF_{\xi}(x) = \int_{-\infty}^{+\infty} x p_{\xi}(x) dx$  расходится, поэтому математического ожидания не существует.

## 5.2. Моменты, дисперсия, ковариация и прочее

**Определение 5.1.** Моментом порядка  $k$  случайной величины  $\xi$  называется  $\mathbb{E}\xi^k$ . Абсолютным моментом порядка  $k$  случайной величины  $\xi$  называется  $\mathbb{E}|\xi|^k$ . Центральным моментом порядка  $k$  случайной величины  $\xi$  называется  $\mathbb{E}(\xi - \mathbb{E}\xi)^k$ .

**Определение 5.2.** Квантилью случайной величины  $\xi$  порядка  $q$  называется величина  $l_{\xi}(q)$ :

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

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

Если  $q = \frac{1}{4}$ , то  $l_{\xi}$  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 вероятности попасть правее нее. Аналогично можно сказать про квантиль любого порядка. Квантиль определена не единственным образом: пусть  $\xi$  принимает значения  $\{0, 1\}$  с вероятностями  $\frac{1}{2}$ . Тогда медианой  $\xi$  может быть любая точка из отрезка  $[0, 1]$ .

**Определение 5.3.** 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 5.4.** Мода случайной величины  $\xi$  — это наиболее вероятное значение случайной величины. Обозначается  $\text{mod } \xi$ .

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

**Определение 5.5.** Пусть для случайной величины  $\xi$  существуют конечный  $\mathbb{E}\xi$  и  $\mathbb{E}\xi^2$ . Дисперсией называется 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.  $\xi, \eta$  - 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 5.6.** Ковариация двух случайных величин — это величина, равная

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

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:

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

Таким образом пятое свойство дисперсии можно переписать так:

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

**Определение 5.7.** Среднеквадратическое отклонение — величина  $\sigma = \sqrt{\mathbb{D}\xi}$ .

### 5.3. Независимость случайных величин

**Определение 5.8.** Случайные величины  $\xi, \eta$  называются независимыми, если

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

**Утверждение 5.9.** Для независимых случайных величин  $\xi, \eta$  выполнено

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

**Доказательство.** Проведем только для случая дискретных случайных величин.  
Пусть

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

то есть  $\xi$  принимает значение  $x_i$  с вероятностью  $p_i$ .

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

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

■

Для независимых случайных величин

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

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  $\mathbb{P}(x = 0)\mathbb{P}(y = 0) = \frac{1}{2} \cdot \frac{1}{2} \neq 0$ . Однако

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

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

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

**Определение 5.10.** Коэффициент корреляции — величина, описываемая формулой

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

Некоторые свойства коэффициента корреляции:

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

Некоторые пояснения к первому свойству. Ковариация является псевдоскалярным произведением, то есть выполнены все аксиомы скалярного произведения, кроме половины четвертой. Поэтому для нее выполнено неравенство Коши-Буняковского:

$$\text{cov}(\xi, \eta)^2 \leq \text{cov}(\xi, \xi) \text{cov}(\eta, \eta). \quad (5.1)$$

Теперь поделим на произведение ковариаций, извлечем корень и получим нужное свойство.

#### 5.4. Неравенство Маркова, неравенство Чебышева

**Лемма 5.11.** Для любой неотрицательной неубывающей функции  $g(x)$  выполнено неравенство

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

**Доказательство.**

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

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

Из этой леммы следуют два полезных неравенства.

**Теорема 5.12** (неравенство Маркова). Для любой случайной величины  $\xi$ , имеющей конечное  $\mathbb{E}|\xi|$ , выполнено

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

**Доказательство.**

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

**Теорема 5.13** (неравенство Чебышева). Для любой случайной величины  $\xi$ , имеющей конечный первый и второй момент, выполнено

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



**Доказательство.**

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





## §6. Распределения и предельные теоремы

### 6.1. Распределение Бернулли

**Определение 6.1.** Дискретная случайная величина  $\xi$  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 6.2.** 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

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

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

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

Покажем, что вероятность, введенная таким образом, удовлетворяет всем аксиомам вероятностной меры. Не очевидной 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  $\mu_n$ )

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

Рассмотрим теперь некоторые важные распределения, связанные со схемой Бернулли.

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

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

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  $Bi(n, p)$ .



Рассмотрим случайные величины  $X_i = \mathbb{1}(A_i)$   $i = 1, \dots, n$ , 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  $\mu_n$

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

$$\mathbb{E}X_i^2 = 1p + 0q = p,$$

$$\mathbb{D}X_i = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2 = p - p^2 = pq.$$

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 6.3** (Бернулли). Для случайной величины с распределением  $Bi(n, p)$

$$\mathbb{P} \left( \left| \frac{\mu_n}{n} - p \right| \geq \varepsilon \right) \leq \frac{npq}{n^2 \varepsilon^2}.$$

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

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

## 6.2. Многочлен Бернштейна

Пусть  $f(x) \in C[0, 1]$ .

**Определение 6.4.** Многочленом Бернштейна называется функция

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

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

**Утверждение 6.5.**

$$B_n(x, f) \Rightarrow f(x), \quad x \in [0, 1].$$

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

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

Разобьем данную сумму на две:

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

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

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

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

Пусть в схеме Бернулли с вероятностью успеха  $0 < p \leq 1$  величина  $\eta$  равна номеру первого успеха.  $\eta$  является случайной величиной, принимающей натуральные значения. Найдем распределение  $\eta$ : 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  $\Omega = \left\{ \omega_k = \underbrace{0 \dots 0}_k 1 : k \in \mathbb{N} \right\}$ , то

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

Такое распределение называется *геометрическим с параметром  $p$* .

Пусть  $\xi$  имеет геометрическое распределение с параметром  $p$ . Тогда

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

Аналогично, дифференцируя степенные ряды, получим дисперсию

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

Пусть теперь  $\theta$  — 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 *отрицательным биномиальным* с параметрами  $r, p$ .

### 6.3. Распределение Пуассона. Теорема Муавра-Лапласа

**Теорема 6.6** (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  $k = 0$

$$\mathbb{P}(\mu_n = 0) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda} \quad (n \rightarrow \infty).$$

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

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

Воспользуемся предположением индукции для  $k - 1$ .

$$\mathbb{P}(\mu_n = k) \rightarrow \frac{(n - k + 1) \frac{\lambda}{n}}{k \left(1 - \frac{\lambda}{n}\right)} \frac{e^{-\lambda} \lambda^{k-1}}{(k - 1)!} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!} \quad (n \rightarrow \infty).$$

■

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

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

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

**Зочем?** 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.

**Определение 6.7.** Распределение величины  $\xi$  такое, что  $\forall k \in \mathbb{Z}_+ \quad \mathbb{P}(\xi = k) = \frac{e^{-\lambda} \lambda^k}{k!}$ , называется распределением Пуассона и обозначается  $\text{Pois}(\lambda)$ .

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

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

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

**Утверждение 6.8.** Пусть независимые случайные величины  $\xi_i \sim \text{Pois}(\lambda_i)$ ,  $i = 1, 2$ . Тогда условное распределение  $\xi_1$  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.**

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

В числителе воспользуемся независимостью  $\xi_1, \xi_2$ , а в знаменателе свойством 3 пуассоновского распределения:

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

■

**Теорема 6.9** (Муавра-Лапласа). Пусть  $\sqrt{npq}$  велико. Тогда

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

**Доказательство.** Не было и не должно быть.

■

Рассмотрим функции

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

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int_{-\infty}^x \varphi(u) du.$$

Заметим, что

$$\int_{-\infty}^{+\infty} \varphi(u) du = 1.$$

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

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

Пусть  $\xi \sim \mathcal{N}(0, 1)$ .

$$\mathbb{E}\xi^k = \int_{-\infty}^{+\infty} x^k \varphi(x) dx.$$

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

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

**Определение 6.10.** Функция  $\psi(h)$  называется *производящей функцией моментов*.

Производящая функция моментов позволяет легко найти любой момент, просто взяв производную от функции в нуле.

$$\psi(h) = \int_{-\infty}^{+\infty} e^{hx} \varphi(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{hx - \frac{x^2}{2}} dx = \frac{e^{\frac{h^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-h)^2}{2}} dx = e^{\frac{h^2}{2}}.$$

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

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

$$\mathbb{P}(m_1 \leq \mu_n \leq m_2) = \mathbb{P}\left(\frac{m_1 - np}{\sqrt{npq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}}\right) \approx \int_{x_1}^{x_2} \varphi(x) dx,$$

$$x_1 \equiv \frac{m_1 - np}{\sqrt{npq}}, \quad x_2 \equiv \frac{m_2 - np}{\sqrt{npq}}.$$

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

Рассмотрим теперь случайную величину  $\xi$  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  $\xi$ . Однако их отношение будет иметь конечный предел:

$$\forall x > 0 \quad \mathbb{P}(p\xi > x) = \mathbb{P}\left(\xi > \frac{x}{p}\right) = \sum_{k=\lceil \frac{x}{p} \rceil}^{\infty} p(1-p)^{k-1} = \frac{p(1-p)^{\lceil \frac{x}{p} \rceil}}{p} = (1-p)^{\lceil \frac{x}{p} \rceil}.$$

$$\lim_{p \rightarrow 0} \mathbb{P}(p\xi > x) = \lim_{p \rightarrow 0} (1-p)^{\lceil \frac{x}{p} \rceil} = e^{-x}.$$

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

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

Это частный случай *показательного* (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

$$\int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

Показательно распределение обладает интересным свойством: свойством отсутствия последствия (или отсутствия памяти). Предположим, что вы ждете автобус на остановке, а время между автобусами  $\xi$  имеет показательное распределение. Тогда вероятность того, что вы прождете автобус еще  $t$  никак не зависит от того, как долго ( $\tau$ ) 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).

## 6.4. Характеристические функции

Введем еще какие-нибудь понятия, которые что-то там нам облегчат (наверное).

**Определение 6.12.** Пусть  $\xi$  — случайная величина. Тогда будем говорить, что  $f_\xi(x)$  — характеристическая функция  $\xi$ , если:

$$f_\xi(t) = \mathbb{E}e^{it\xi}.$$

Поясним раз.  $\mathbb{E}e^{it\xi}$ , чтобы не сломать мозг раньше времени, понимаем в смысле  $\mathbb{E}e^{it\xi} = \mathbb{E} \cos t\xi + i\mathbb{E} \sin t\xi$ .

Поясним два. Если  $\xi$  — дискретная случайная величина, то она задается рядом распределения  $\mathbb{P}(\xi = x_k)$ , тогда  $f_\xi(t) = \sum_k e^{itx_k} \mathbb{P}(\xi = x_k)$ . Если же  $\xi$  — абсолютно непрерывная случайная величина, то  $f_\xi(t) = \int e^{itx} dF_\xi(x)$ , где  $F_\xi(x)$  — функция распределения  $\xi$ , а интеграл как всегда берется по всему пространству, да еще и Лебегов.

Прежде чем мы немного поковыряем свойства харфункции и посмотрим примерчики, введем еще пару определений, ведь без определений так скучно жить!

**Определение 6.13.** Функция  $f(x)$  называется неотрицательно определенной, если:

$$\forall n \in \mathbb{N} \quad \forall t_1, \dots, t_n \in \mathbb{R} \quad \forall z_1, \dots, z_n \in \mathbb{C} \quad \sum_{i,j=1}^n f(t_i - t_j) z_i \overline{z_j} \geq 0.$$

**Определение 6.14.** Случайная величина  $\xi$  имеет решетчатое распределение, если

$\exists a, b: \sum_{k=-\infty}^{+\infty} \mathbb{P}(\xi = a + bk) = 1$ . Тогда число  $b$  называется шагом распределения.

Свойства (под  $f(t)$  подразумевается  $f_\xi(t)$  для какой-то случайной величины  $\xi$ ):

1.  $|f(t)| \leq 1$ .
2.  $f(0) = 1$ .
3.  $f(-t) = \overline{f(t)}$
4.  $\forall t \in \mathbb{R} \quad f_\xi(t) \in \mathbb{R} \iff \xi$  распределена симметрично (тривиальное следствие из предыдущего).

5.  $f(t)$  равномерно непрерывна на  $\mathbb{R}$ .
6. Если  $\eta = a\xi + b$ , то  $f_\eta(t) = e^{itb}f_\xi(at)$ .
7. Если  $\xi_1, \dots, \xi_n$  — независимые случайные величины, и  $\eta = \xi_1 + \dots + \xi_n$ , то  $f_\eta(x) = \prod_{k=1}^n f_{\xi_k}(x)$ .
8. Если  $\mathbb{E}|\xi^k| < \infty$ , то  $\mathbb{E}\xi^k = i^k f_\xi^{(k)}(0)$ . Если  $k$  четно, то верно и обратное утверждение.
9. Верна

**Теорема 6.15** (Бохнера-Хинчина).  $f(t)$  является характеристической функцией  $\iff f(0) = 1$  и  $f(t)$  обладает свойством неотрицательной определенности.

10.  $|f_\xi(t)|$  интегрируема  $\implies p_\xi(x) \xrightarrow{|x| \rightarrow \infty} 0$ , где  $p_\xi(x)$  — плотность распределения.
11. Случайная величина  $\xi$  имеет решетчатое распределение с шагом  $b \iff \left| f_\xi\left(\frac{2\pi}{b}\right) \right| = 1$ .

Не расслабляться! Сейчас докажем некоторые утверждения.

**Доказательство. (5)**

$$\begin{aligned} |f(t+h) - f(t)| &= \left| \int_{-\infty}^{\infty} e^{i(t+h)x} dF(x) - \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |e^{itx}(e^{ihx} - 1)| dF(x) \leq \\ &\leq \int_{-\infty}^{\infty} |e^{ihx} - 1| dF(x) = \underbrace{\int_{|x| \leq M} |e^{ihx} - 1| dF(x)}_{I_1} + \underbrace{\int_{|x| > M} |e^{ihx} - 1| dF(x)}_{I_2}. \end{aligned}$$

Оценим теперь интегралы  $I_1$  и  $I_2$ . Функция, непрерывная на компакте равномерно непрерывна на нем, а значит  $\forall \varepsilon > 0 \exists h$ , такой что  $I_1 < \frac{\varepsilon}{2}$ . Для второго интеграла имеем:  $|e^{ihx} - 1| \leq 2 \implies I_2 \leq 2\mathbb{P}(|\xi| > M) < \frac{\varepsilon}{2}$  за счет выбора  $M$ . ■

**Доказательство. (7)**

Воспользуемся независимостью случайных величин, чтобы разбить одно большое матожидание на много маленьких.

$$f_{\xi_1 + \dots + \xi_n} = \mathbb{E}e^{it(\xi_1 + \dots + \xi_n)} = \mathbb{E}(e^{it\xi_1} \dots e^{it\xi_n}) = \mathbb{E}e^{it\xi_1} \dots \mathbb{E}e^{it\xi_n} = \prod_{k=1}^n f_{\xi_k}(x).$$

**Доказательство. (8)**

Первую производную посчитаем по определению:

$$f'_\xi(t) = \lim_{h \rightarrow 0} \frac{f_\xi(t+h) - f_\xi(t)}{h} = \lim_{h \rightarrow 0} \int_{-\infty}^{\infty} e^{ixt} \frac{e^{ixh} - 1}{h} dF_\xi(x) = \dots$$

Тут у нас  $\frac{e^{ixh} - 1}{h} \leq |x|$ , поэтому интеграл сходится равномерно (признак Вейерштрасса), и мы можем поменять местами предел и интеграл и ничего нам за это не будет:

$$\dots = \int_{-\infty}^{\infty} e^{ixt} \lim_{h \rightarrow 0} \frac{e^{ixh} - 1}{h} dF_\xi(x) = \dots$$

Теперь надо пристально посмотреть на последнее подынтегральное выражение и заметить там 3-ий замечательный предел:

$$\frac{e^{ixh} - 1}{h} \xrightarrow{h \rightarrow 0} ix.$$

Тогда продолжаем:

$$\dots = i \int_{-\infty}^{\infty} x e^{itx} dF_\xi(x).$$

При  $t = 0$   $e^{itx} = 1$ , поэтому

$$f'_\xi(0) = i \int_{-\infty}^{\infty} x dF(x) = i\mathbb{E}\xi.$$

По индукции показываем справедливость для производных старших порядков. Обратное утверждение для четных  $k$  доказывать не будем. Но там немного попопалитесь и доказать это все безобразие по индукции. Можно залезть в Ширяева и удовлетворить свое любопытство. ■

**Зочем?** Мая. Если  $\mathbb{E}|\xi|^n < \infty$ , то мы можем записать харфункцию в виде суммы:

$$f_\xi(t) = \sum_{k=0}^n \frac{t^k}{k!} f_\xi^{(k)}(0) + \bar{o}(t^k) = 1 + \sum_{k=1}^n \frac{(it)^k}{k!} \mathbb{E}\xi^k + \bar{o}(t^k).$$

Довольно удобно считать моменты, если мы уверены в их существовании.

Теперь рассмотрим некоторые примеры:

**Пример.** Найдём характеристическую функцию для стандартного нормального рас-

предела  $\mathcal{N}(0, 1)$  с плотностью  $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ .

$$\begin{aligned} f(t) = \mathbb{E}e^{it\xi} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} e^{-\frac{t^2}{2}} dx = \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} d(x-it) = e^{-\frac{t^2}{2}}. \end{aligned}$$

**Пример.** Вот представь ситуацию:ходишь ты в хату, а тебе пахан кидает под ноги кидает  $\cos t^2$  и говорит: «а найди-ка нам случайную величину, для которой это выражение будет харфункцией». Тут главное — не зашквариться и по понятиям пояснить, что если бы  $\cos t^2$  была бы харфункцией, то по свойству (5) она была бы равномерно непрерывной на  $\mathbb{R}$ , а это не так: если  $t_1^2 = 2\pi k - \frac{\pi}{2}$ , а  $t_2^2 = 2\pi k$ , то  $\cos t_2^2 - \cos t_1^2 = 1$ , а

$$t_2 - t_1 = \sqrt{2\pi k} - \sqrt{2\pi k - \frac{\pi}{2}} = \frac{\frac{\pi}{2}}{\sqrt{2\pi k} + \sqrt{2\pi k - \frac{\pi}{2}}} \xrightarrow{k \rightarrow \infty} 0.$$

То есть для любого наперед заданного  $\delta > 0$  мы можем найти  $k$  достаточно большое, чтобы разность между  $t_1$  и  $t_2$  была меньше  $\delta$ , а  $\cos t_2^2 - \cos t_1^2 = 1$ , что противоречит условию равномерной непрерывности.

**Пример.** А теперь найдем случайную величину  $\xi$ , такую что  $\mathbb{E}e^{it\xi} = \cos t$ . Для этого вспомним, что если  $\xi$  — дискретная случайная величина, то  $\mathbb{E}g(\xi) = \sum_k g(k)\mathbb{P}(\xi = k)$

$$\cos t = \frac{1}{2}(e^{it} + e^{-it}) \implies \xi = \begin{cases} -1, & \mathbb{P}(\xi = -1) = \frac{1}{2} \\ 1, & \mathbb{P}(\xi = 1) = \frac{1}{2} \end{cases}.$$

## 6.5. Распределение функций от случайных величин

Рассмотрим отображение  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  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 6.16.** Пусть  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  — 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:

$$\int_B \varphi(x) dx = \int_{f^{-1}(B)} \varphi(f(y)) |J_f(y)| dy$$

Тогда  $\forall B \in \mathcal{B}_n$

$$\int_B p_\xi(f^{-1}(x))|J_{f^{-1}(x)}|dx = \int_{f^{-1}(B)} p_\xi(f(f^{-1}(B)))|J_{f^{-1}(x)}||J_{f(x)}|dx = .$$

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

■

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

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

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

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

В силу симметрии

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

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

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



## §7. Сходимость случайных величин

### 7.1. Виды сходимости

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  $(\Omega, \mathcal{A})$ , вероятностное пространство  $(\Omega, \mathcal{A}, \mathbb{P})$  и случайную величину  $\xi: \Omega \rightarrow \mathbb{R}$ , которая обладает свойством

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

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

**Определение 7.1.** Последовательность случайных величин  $\{\xi_n\}$  сходится к случайной величине  $\xi$  почти наверное (с вероятностью 1), если вероятность множества тех элементарных событий, где она не сходится, равно нулю.

$$\{\xi_n\} \xrightarrow{\text{п.н.}} \xi, \text{ если } \mathbb{P}(\{\omega: \xi_n(\omega) \not\rightarrow \xi\}) = 0.$$

**Пример.** Пусть  $\xi_n$  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 7.2.** Последовательность случайных величин  $\{\xi_n\}$  сходится к случайной величине  $\xi$  в среднем порядка  $r$ , если  $r$ -ый момент их разности сходится к нулю.

$$\{\xi_n\} \xrightarrow{(r)} \xi, \text{ если } \mathbb{E}(\xi_n - \xi)^r \xrightarrow{n \rightarrow +\infty} 0.$$

**Пример.** Возьмем в качестве множества  $\Omega$  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.

$$\mathbb{E}(\xi_n - \xi)^r = \mathbb{E}\xi_n^r = \mathbb{E}(\mathbb{1}(\omega \in A_n))^r = \mathbb{P}(A_n) = 1/n \xrightarrow{n \rightarrow +\infty} 0.$$

**Утверждение 7.3.** 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  $\omega$  на окружности мы можем указать бесконечное число номеров  $n_k$ , таких что  $\omega \in A_{n_k}$ . ■

**Утверждение 7.4.** 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  $\forall r > 0 \quad \mathbb{E} \xi_n^r = e^{nr} \mathbb{E} \mathbb{1}(\omega \in [0, 1/n]) = e^{nr}/n$ , а эта величина стремится к бесконечности, значит, сходимости в среднем нет. ■

Введем еще один вид сходимости.

**Определение 7.5.** Последовательность случайных величин  $\{\xi_n\}$  сходится к случайной величине  $\xi$  по вероятности, если для любого сколь угодно малого положительного  $\varepsilon$  вероятность таких событий, что модуль разности  $\xi_n$  и  $\xi$  больше  $\varepsilon$ , стремится к 0.

$$\{\xi_n\} \xrightarrow{\mathbb{P}} \xi, \text{ если } \forall \varepsilon > 0 \quad \mathbb{P}(\{\omega: |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}) \xrightarrow{n \rightarrow +\infty} 0.$$

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** На этом моменте может немного поплавать 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,

$$\mathbb{P}(\{\omega: |\xi_n(\omega) - 0| > \varepsilon\}) = \mathbb{P}(\{\omega: \xi_n(\omega) > \varepsilon\}) = \mathbb{P}(\{\omega: \xi_n(\omega) = 1\}) = 1/n \xrightarrow{n \rightarrow +\infty} 0.$$

**Утверждение 7.6.** Только что рассмотренным примером мы доказали, что из сходимости по вероятности, вообще говоря, не следует сходимость почти наверное.

**Утверждение 7.7.** 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,  $\mathbb{P}\{\omega: \xi_n(\omega) > \varepsilon\} = 1/n \xrightarrow{n \rightarrow +\infty} 0$ . Отсутствие сходимости в среднем мы уже доказали. ■

**Теорема 7.8.** Из сходимости почти наверное следует сходимость по вероятности.

**Доказательство.** Сначала докажем, что  $\{\xi_n\} \xrightarrow{\text{п.н.}} \xi \Leftrightarrow \mathbb{P}\left(\sup_{k \geq n} |\xi_k - \xi| \geq \varepsilon\right) \rightarrow 0$ . Положим  $A_n^\varepsilon = \{\omega: |\xi_n - \xi| \geq \varepsilon\}$ ,  $A^\varepsilon = \overline{\lim} A_n^\varepsilon \equiv \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k^\varepsilon$ . Тогда:

$$\begin{aligned}
& \mathbb{P}(\{\omega: \xi_n(\omega) \not\rightarrow \xi\}) = 0 \Leftrightarrow \mathbb{P}\left(\bigcup_{\varepsilon>0} A^\varepsilon\right) = 0 \Leftrightarrow \mathbb{P}\left(\bigcup_{m=1} A^{1/m}\right) = 0 \Leftrightarrow . \\
& \Leftrightarrow \mathbb{P}(A^{1/m}) = 0, m \geq 1 \Leftrightarrow \mathbb{P}(A^\varepsilon) = 0, \varepsilon > 0 \Leftrightarrow \mathbb{P}\left(\bigcup_{k \geq n} A_k^\varepsilon\right) \xrightarrow{n \rightarrow \infty} 0, \varepsilon > 0 \Leftrightarrow . \\
& \Leftrightarrow \mathbb{P}\left(\sup_{k \geq n} |\xi_k - \xi| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \mathbb{P}(|\xi_k - \xi| \geq \varepsilon) \rightarrow 0.
\end{aligned}$$

А последнее утверждение — это определение сходимости по вероятности. ■

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Совершенно жуткое доказательство, понимается методом вглядывания: если достаточно долго медитировать над каждой импликацией, то все переходы в конце концов станут понятными.

**Теорема 7.9.** Из сходимости в среднем следует сходимость по вероятности.

**Доказательство.** Утверждение теоремы практически сразу же следует из обобщенного неравенства Чебышева:

$$\mathbb{P}(|\xi_n - \xi| \geq \varepsilon) \leq \frac{\mathbb{E}|\xi_n - \xi|^r}{\varepsilon^r}.$$

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

**Определение 7.10.** Последовательность случайных величин  $\{\xi_n\}$  слабо сходится к случайной величине  $\xi$ , если для любой непрерывной и ограниченной функции последовательность мат. ожиданий функций от  $\xi_n$  сходится к мат. ожиданию функции от  $\xi$ .

$$\{\xi_n\} \xrightarrow{w} \xi, \text{ если } \forall f(x): f(x) \in C \text{ и } |f(x)| \leq M \quad \mathbb{E}f(\xi_n) \rightarrow \mathbb{E}f(\xi).$$

**Утверждение 7.11.** Из сходимости по вероятности следует слабая сходимость.

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

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

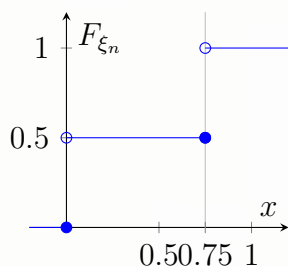
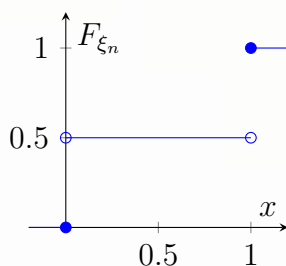
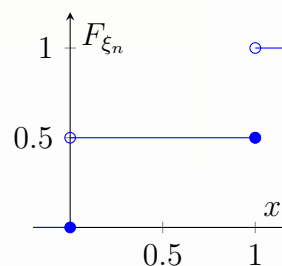
Рис. 3: Функция распределения  $\xi_n$  при  $n = 4$ 

Рис. 4: Предельная функция

Рис. 5: Функция распределения  $\xi$ 

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

**Определение 7.12.** Последовательность случайных величин  $\{\xi_n\}$  сходится к случайной величине  $\xi$  по распределению, если функции распределения  $\{F_{\xi_n}\}$  сходятся к функции распределения  $F_{\xi}$  во всех точках, в которых предельная функция распределения непрерывна.

$$\{\xi_n\} \xrightarrow{d} \xi, \text{ если } F_{\xi_n}(x) \rightarrow F_{\xi}(x) \quad \forall x: F_{\xi}(x) \in C(x).$$

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

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

**Утверждение 7.13.** Сходимость по распределению эквивалентна слабой сходимости.

**Доказательство.** Гдето шляется, мы искренне надеемся, что оно сюда вернется. ■

**Утверждение 7.14.** Если последовательность случайных величин  $\{\xi_n\}$  сходится по распределению к вырожденной случайной величине  $\xi \stackrel{\text{п.н.}}{\equiv} a$ , то  $\{\xi_n\}$  сходится к  $\xi$  по вероятности.

**Доказательство. No.**

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

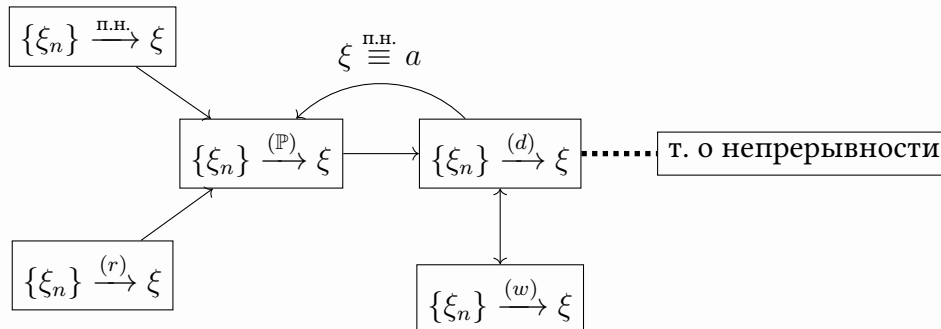
**Теорема 7.15** (теорема Леви о непрерывности). 1. Пусть  $\{\xi_n\} \xrightarrow{d} \xi$  и  $f_n(t) := \mathbb{E}e^{it\xi_n}$ . Тогда  $\{f_n(t)\} \rightarrow f(t)$ , где  $f(t) = \mathbb{E}e^{it\xi}$ .

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

**Доказательство.**

1. Как было показано выше, сходимость по распределению эквивалентна слабой сходимости. Положим в определении слабой сходимости функции  $\varphi(x) := \operatorname{Re}(e^{itx})$  и  $\psi(x) := \operatorname{Im}(e^{itx})$ , тогда  $\mathbb{E}\varphi(\xi_n) \rightarrow \mathbb{E}\varphi(\xi)$  и  $\mathbb{E}\psi(\xi_n) \rightarrow \mathbb{E}\psi(\xi)$ , откуда следует утверждение теоремы.
2. Ребята, не стоит вскрывать эту тему...

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** А вот картинка, на которой схематично показано, что из чего следует.

**7.2. Закон больших чисел**

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

**Определение 7.16.** Пусть  $\{\xi_n\}$  — последовательность независимых одинаково распределенных случайных величин, а  $S_n = \sum_{k=1}^n \xi_k$ . Говорят, что  $\{\xi_n\}$  удовлетворяет закону больших чисел (ЗБЧ), если  $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_1$ .

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

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\mathbb{D}S_n}{\varepsilon^2} = \frac{\mathbb{D}\xi_1}{n\varepsilon^2} \rightarrow 0.$$

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

**Теорема 7.17** (ЗБЧ в форме Хинчина). Если  $\{\xi_k\}$  — независимые одинаково распределенные случайные величины с конечным матожиданием, то для них выполнен ЗБЧ.

**Доказательство.** Первое желание, которое может возникнуть при виде данной теоремы — записать неравенство Чебышева, но для неравенства 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 сходимости по распределению, и воспользуемся теоремой Леви о непрерывности.

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

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

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

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

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

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 сходится по распределению  $\frac{S_n}{n}$ . Действительно, эта случайная величина  $\xi \stackrel{\text{п.н.}}{=} a$ . Таким образом, имеем, что  $\frac{S_n}{n} \xrightarrow{w} \mathbb{E}\xi_k$ , но если предельная случайная величина вырождена, то сходимость по распределению эквивалентна сходимости по вероятности, значит  $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_k$ . ■

Лулзов ради закинем еще парочку ЗБЧ.

**Теорема 7.18** (ЗБЧ в форме Чебышева). Пусть  $\{\xi_k\}$  — последовательность некоррелированных случайных величин с равномерно ограниченной дисперсией. То есть  $\sup_{k \in \mathbb{N}} \mathbb{D}\xi_k \leq C < \infty$ . Тогда для данной последовательности случайных величин выполнен ЗБЧ.

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

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\xi_k\right| \geq \varepsilon\right) = \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\mathbb{D}S_n}{(n\varepsilon)^2} = \frac{1}{n^2\varepsilon^2} \sum_{k=1}^n \mathbb{D}\xi_k \leq \frac{C}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

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

**Замечание** Условия те же, что и на ЗБЧ в форме Хинчина, но сходимость уже не по вероятности, а почти наверное.

**Доказательство.** Все верно. Отвечаю.

**ЧО?! ЧЁ ЭТО ТАКОЕ?! 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 арифметическими.

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

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

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

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

$$\begin{aligned} \mathbb{P}\left(\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right) &= \int_{n(a+\varepsilon)}^{+\infty} p(x)dx + \int_{-\infty}^{n(a-\varepsilon)} p(x)dx = \frac{1}{\pi} \operatorname{arctg}\left(\frac{x}{\gamma n}\right) \Big|_{n(a+\varepsilon)}^{+\infty} + \frac{1}{\pi} \operatorname{arctg}\left(\frac{x}{\gamma n}\right) \Big|_{-\infty}^{n(a-\varepsilon)} = \\ &= \frac{1}{\pi} \left( \operatorname{arctg}\left(\frac{a-\varepsilon}{\gamma}\right) - \operatorname{arctg}\left(\frac{a+\varepsilon}{\gamma}\right) \right) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$



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

### 7.3. Центральная предельная теорема

С места в карьер:

**Определение 7.20.** Последовательность случайных величин  $\{\xi_k\}$  удовлетворяет ЦПТ, если  $\exists \{a_k\} \in \mathbb{R}, \{b_k\} > 0$ , такие что

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

Разомнемся на чем-нибудь попроще.

**Теорема 7.21.** 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  $S_n = \sum_{k=1}^n \xi_k$ .

**Замечание** Здесь распределение  $\frac{S_k - \mathbb{E}\xi_k}{\sqrt{\mathbb{D}\xi_k}}$  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

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

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

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

**Зочем?** 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.
- $F_k(x) = \mathbb{P}(\xi_k < x)$ .
- $S_n = \sum_{k=1}^n \xi_k$ .
- $a_k = \mathbb{E}\xi_k, A_n = \sum_{k=1}^n a_k$ .

$$\bullet \quad b_k^2 = \mathbb{D}\xi_k, \quad B_n^2 = \sum_{k=1}^n b_k^2.$$

**Теорема 7.22** (Ляпунова). Пусть  $\mu_k^3 := \mathbb{E}|\xi_k - a_k|^3 < \infty$ ,  $M_n^3 := \sum_{k=1}^n \mu_k^3$ . Пусть выполнено условие Ляпунова:

$$\frac{M_n^3}{B_n^3} \xrightarrow{n \rightarrow \infty} 0 \quad (7.1)$$

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

**Теорема 7.23** (Линдберга). Пусть  $\varepsilon > 0$ . Тогда если  $a_k < \infty$ ,  $b_k < \infty$  и выполнено условие Линдберга:

$$L_n(\varepsilon) = \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k| > \varepsilon B_n} (x - a_k)^2 dF_k(x) \xrightarrow{n \rightarrow \infty} 0, \quad (7.2)$$

то  $\{\xi_k\}$  удовлетворяет ЦПТ.

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

**Определение 7.24.**  $\{\xi_k\}$  удовлетворяют условию Феллера, если

$$\forall \varepsilon > 0 \quad \max_{1 \leq k \leq n} \mathbb{P} \left( \frac{|\xi_k - a_k|}{B_n} > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0. \quad (7.3)$$

**Теорема 7.25.**

$$\begin{cases} \text{ЦПТ} \\ (7.3) \end{cases} \iff (7.2).$$

**Доказательство.** Покажем, что из (7.2) следует (7.3).

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

$$\begin{aligned} \max_{1 \leq k \leq n} \mathbb{P}(|\xi_k - a_k| > \varepsilon B_n) &\leq \sum_{k=1}^n \mathbb{P}(|\xi_k - a_k| > \varepsilon B_n) = \sum_{k=1}^n \int_{|x-a_k| > \varepsilon B_n} 1 dF_k(x) = \\ &= \sum_{k=1}^n \int_{|x-a_k| > \varepsilon B_n} \frac{(x - a_k)^2}{(x - a_k)^2} dF_k(x) \leq \frac{1}{\varepsilon^2 B_n^2} \int_{|x-a_k| > \varepsilon B_n} (x - a_k)^2 dF_k(x) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

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

**Теорема 7.26.** Условие Ляпунова (7.1) влечет за собой условие Линдберга (7.2).

**Доказательство.**

$$\begin{aligned}
 L_n(\varepsilon) &= \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k| > \varepsilon B_n} (x-a_k)^2 dF_k(x) = \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k| > \varepsilon B_n} \frac{|x-a_k|^3}{|x-a_k|} dF_k(x) \leq \\
 &\leq \frac{1}{\varepsilon B_n^3} \sum_{k=1}^n \int_{-\infty}^{+\infty} |x-a_k|^3 dF_k(x) = \frac{M_n^3}{\varepsilon B_n^3} \xrightarrow{n \rightarrow \infty} 0.
 \end{aligned}$$

■

**Теорема 7.27** (Неравенство Берри—Эссена).  $\sup_x \left| \mathbb{P} \left( \frac{S_n - A_n}{B_n} < x \right) - \Phi(x) \right| \leq C \frac{M_n^3}{B_n^3}$ .

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  $C_0 \geq \frac{\sqrt{10} + 3}{6\sqrt{2\pi}}$ . А потом доказали, что здесь вообще стоит строгое равенство.



---

## ***Часть № II. Математическая статистика***

---



## §1. Введение

В прошлой части мы упарывались с теорией вероятности, а теперь будем флексить над математической статистикой. Отличительной особенностью теории вероятности было то, что нам дана была вероятностная мера и мы должны были что-то там с ней сделать. Например, найти вероятность какой-то сложного события или матожидание какой-нибудь случайной величины. А в матстате вероятностная мера не задана в условии, и надо как-то извратиться и найти ее или хотя бы оценить. Просто так с потолка что-либо оценить довольно трудно, поэтому оценивают обычно на основе нескольких реализаций случайных событий. Теперь любимые определения:

**Определение 1.1.** Множество случайных величин  $\mathbb{X} = (X_1, \dots, X_n)$  называется выборкой.

**Определение 1.2.** Выборка называется независимой, если случайные величины независимы в совокупности.

**Определение 1.3.** Выборка называется однородной, если случайные величины одинаково распределены.

В дальнейшем под выборкой мы будем подразумевать независимую однородную выборку, т. е. случайные величины будут всегда (если не оговорено отдельно) норсв. Теперь, раз уж у нас есть много случаных величин, то можно рассматривать выборку как векторную случайную величину, определенную на измеримом пространстве  $(\mathbb{R}_n, \mathcal{B}_n)$ . Но, как мы помним, вероятностной меры мы не знаем, поэтому на данном вероятностном пространстве зададим сразу семейство вероятностных мер  $\mathcal{P}$  и получится такой крокодил, как

**Определение 1.4.** Статистическая структура —  $(\mathbb{R}_n, \mathcal{B}_n, \mathcal{P})$ .

Есть 3 больших класса задач, которые решает матстат:

1. *Точечное оценивание.* В задачах точечного оценивания мы считаем, что наше вероятностное семейство параметризовано некоторым параметром  $\theta$ , изменяющемся на множестве  $\Theta$ :  $\mathcal{P} = \{P_\theta: \theta \in \Theta\}$ . В этом классе задач от нас хотят, чтобы мы сказали, что  $\theta \approx \theta_0 \in \Theta$ , где примерное равенство обычно понимают в смысле математического ожидания. Иногда так же полезно найти разброс (дисперсию) нашей оценки. Например, мы знаем, что выборка сгенерирована из нормального распределение с дисперсией 1 и неизвестным матожиданием, которое надо найти.
2. *Интервальное оценивание.* Здесь у нас тоже вероятность параметризована, но мы уже не просто тыкаем в какое-то конкретное значение  $\theta_0$ , а говорим, что  $\theta_1 < \theta < \theta_2$  с такой-то вероятностью.
3. *Проверка гипотез.* Гипотеза — это какое-то наше предположение относительно вероятностных мер. Обозначается  $H$  (возможно, с индексом) и формализуется

---

каждый раз по-разному. Например, гипотеза о том, что выборка из равномерного распределения или о том, что выборка из гамма-распределения с параметром формы, не превосходящем 13.





## §2. Точечное оценивание

### 2.1. Эмпирическая функция распределения и её свойства

#### 2.1.1. Вариационный ряд

Как известно, функция распределения  $F(x) = \mathbb{P}(X < x)$  однозначно определяет вероятностную меру. Попробуем построить что-то похожее на функцию распределения. Для этого мы возьмем нашу выборку  $\mathbb{X} = (X_1, \dots, X_n)$  и отсортируем по неубыванию. Получим

**Определение 2.1.** Вариационный ряд  $\tilde{\mathbb{X}} = (X_{(1)}, \dots, X_{(n)})$  — выборка, составленная из элементов исходной выборки, упорядоченных по неубыванию. Элементы вариационного ряда называются порядковыми статистиками.

**Замечание** Тут уже мы понимаем выборку в широком смысле, так как элементы вариационного ряда не являются независимыми и тем более не являются одинаково распределенными. Самое простое: если  $X_{(1)} = 2$ , то остальные элементы ряда заведомо не меньше 2, то есть зависят от  $X_{(1)}$ . Дальше зависимость элементов показана аккуратно.

Теперь выведем функцию распределения  $X_{(k)}$ . Наша цель — найти  $F_{(k)}(x) = \mathbb{P}(X_{(k)} < x)$ .  $X_{(k)} < x$  тогда и только тогда, когда хотя бы  $k$  элементов выборки меньше  $x$ . То есть возможны варианты для всех  $i$  от  $k$  до  $n$ , в каждом из которых ровно  $i$  элементов меньше  $x$  и ровно  $n - i$  элементов не меньше  $x$ .

$$F_{(k)}(x) = \mathbb{P}(X_{(k)} < x) = \sum_{i=k}^n C_n^i (F(x))^i (1 - F(x))^{n-i} \quad (2.1)$$

Напоминает схему испытаний Бернулли. Это она и есть.

#### 2.1.2. Эмпирическая ф.р.

**Определение 2.2.** Эмпирическая (выборочная) функция распределения

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i < x). \quad (2.2)$$

Выборочная функция распределения приближает истинную функцию распределения. Смысл, в котором эмпирическая функция приближает истинную, раскрывается в пачке следующих теорем и утверждений.

**Утверждение 2.3.**

$$\mathbb{E}F_n(x) = F(x), \quad (2.3)$$

где  $F(x)$  — истинная функция распределения.

**Доказательство.** Зафиксируем  $x$  и рассмотрим процесс генерации нашей выборки. Каждое испытание — генерация следующего значения из распределения. Будем считать, что  $k$ -е испытание успешное, если  $X_k < x$ . Тогда веро-

ятность успеха в  $k$ -ом испытании  $\mathbb{P}(X_k < x) = \mathbb{P}(X < x) = F(x)$ , а сумма индикаторов  $nF_n(x)$  — число успехов в  $n$  испытаниях. Поэтому мы можем записать, что  $nF_n(x) \sim Bi(n, F(x))$ . И из свойств биномиального распределения  $\mathbb{E}[nF_n(x)] = nF(x) \Rightarrow \mathbb{E}F_n(x) = F(x)$ . ■

Теперь надо собрать волю в кулак и вспомнить усиленный закон больших чисел Колмогорова (УЗБЧ). Он говорит, что для норсв с конечным МО  $\frac{S_n}{n} \xrightarrow{\text{п.н.}} \mathbb{E}\xi$ . В нашем случае, он выглядит так:

$$F_n(x) \xrightarrow{\text{п.н.}} F(x). \quad (2.4)$$

А что это тут у нас тако-о-ое? А это же функциональная последовательность! А когда у нас функциональная последовательность куда-то сходится, что мы от неё хотим? Пра-а-авильно, равномерной сходимости. Так вот есть такая

**Теорема 2.4** (Гливленко). Эмпирическая функция распределения сходится к истинной равномерно с вероятностью 1.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0\right) = 1. \quad (2.5)$$

**Доказательство.** Для простоты докажем для абсолютно непрерывного случая. Пусть наша выборка генерируется из распределения с ф.р.  $F(x)$ . Тогда зафиксируем число  $m$ , и для него найдем числа  $-\infty = x_0 < x_1 < \dots < x_{m-1} < x_m = \infty$ , такие что  $F(x_j) - F(x_{j-1}) = \frac{1}{m}$ ,  $j = \overline{1, m}$ . Теперь  $\forall x \exists j = \overline{1, m}$ , такое что  $x \in [x_{j-1}, x_j]$ . Тогда

$$F_n(x) - F(x) \leq F_n(x_j) - F(x_{j-1}) = F_n(x_j) - F(x_j) + \frac{1}{m}, \quad (2.6)$$

$$F_n(x) - F(x) \geq F_n(x_{j-1}) - F(x_j) = F_n(x_{j-1}) - F(x_{j-1}) - \frac{1}{m}. \quad (2.7)$$

Комбинируя (2.6) и (2.7) и ещё там всякие модули, супремумы, получаем

$$\sup_x |F_n(x) - F(x)| \leq \max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| + \frac{1}{m} \quad (2.8)$$

Теперь в силу (2.4)  $\max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| \xrightarrow{\text{п.н.}} 0$ , поэтому  $\forall m \exists N$ , такой что  $\forall n \geq N \sup_x |F_n(x) - F(x)| \leq \frac{1}{m} + \varepsilon$ . Наконец, полагая  $m \rightarrow \infty$  и  $\varepsilon \rightarrow 0$ , получаем утверждение теоремы. ■

Хорошо. Вот мы выяснили, что все хорошо и выборочная ф.р. благополучно сходится к теоретической. Но сразу встает вопрос, а вот какой размер выборки мы можем считать достаточным? Иными словами, нам интересна не только сам факт сходимости, но и скорость сходимости эмпирической ф.р. к теоретической. А для этого на понадобится вот этот (4.1) факт. Обозначим  $D_n = \sup_x |F_n(x) - F(x)|$

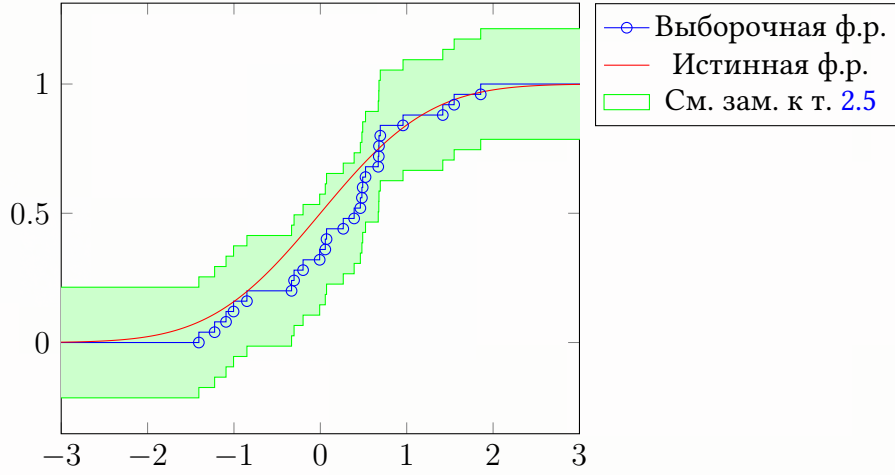


Рис. 6: Эмпирическая и истинная функции распределения. В качестве примера взята выборка размера 25 из стандартного нормального распределения.

**Теорема 2.5** (Колмогорова).

$$\mathbb{P}(\sqrt{n}D_n < x) = K_n(x) \quad \forall \text{ ф.р. } F.$$

$$\lim_{n \rightarrow \infty} K_n(x) = K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 x^2} \mathbb{1}(x \geq 0).$$

**Доказательство.** Докажем только первую часть теоремы, а именно, что распределение  $D_n$  не зависит от теоретической функции распределения. На  $\sqrt{n}$  мы без зазрения совести забываем, потому что он влияет только на вид распределения, но не на зависимость или независимость от  $F$ .

$$\begin{aligned} \mathbb{P}(D_n < x) &= \mathbb{P}\left(\sup_y |F_n(y) - F(y)| < x\right) = \\ &= \left\{ \begin{array}{l} \text{Считаем, что } F(x) \text{ непре-} \\ \text{рывна и строго монотон-} \\ \text{на и полагаем } y = F^{-1}(z) \end{array} \right\} = \mathbb{P}\left(\sup_{0 \leq z \leq 1} |F_n(F^{-1}(z)) - F(F^{-1}(z))| < z\right) = \\ &= \mathbb{P}\left(\sup_{0 \leq z \leq 1} |F_n(F^{-1}(z)) - z| < z\right) = \mathbb{P}\left(\sup_{0 \leq z \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i < F^{-1}(z)) - z \right| < z\right) = \\ &= \mathbb{P}\left(\sup_{0 \leq z \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}(F(X_i) < z) - z \right| < z\right) = \left\{ \begin{array}{l} U_i := F(X_i), \\ U_i \sim U[0, 1] \text{ (4.1)} \end{array} \right\} = \\ &= \mathbb{P}\left(\sup_{0 \leq z \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}(U_i < z) - z \right| < z\right) \quad (2.9) \end{aligned}$$

**Замечание** Какой первый вопрос, который возникает у каждого трушного прикладного математика, когда он видит подобную теорему? Конечно! Всем интересно, чему

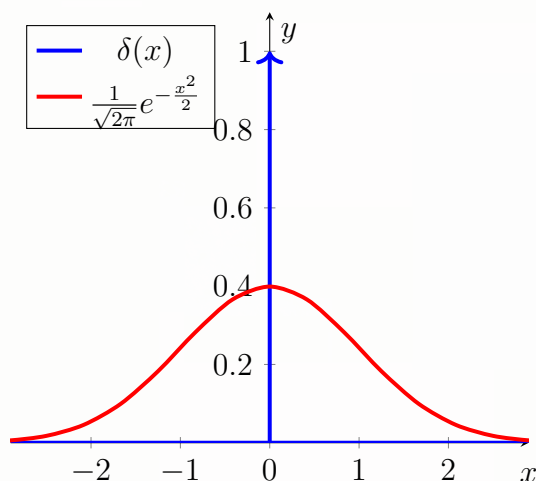


Рис. 7: Различные варианты ядра.

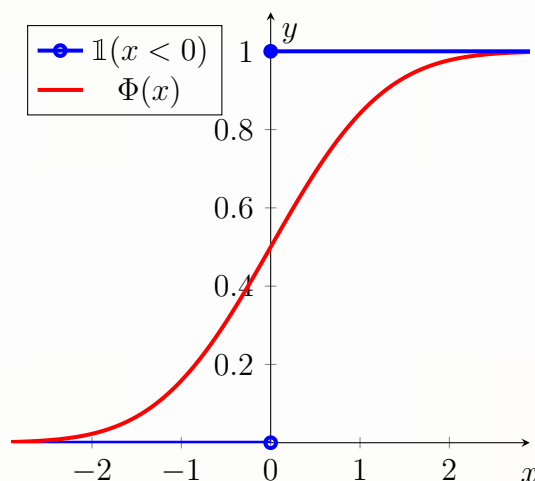


Рис. 8: Соответствующие функции распределения.

равна бесконечность. Иными словами, при каком размере выборки мы можем благополучно использовать  $K(x)$  вместо унылых  $K_n(x)$ ? Так вот для этой теоремы считается, что  $\infty = 20$ . Как обычно пользуются этой теоремой? Как-то так:

1. Определение границ, в которых с заданной вероятностью находится теоретическая функция распределения. Давайте для нашей выборки (рис. 6) построим границы, в которой с вероятностью 0,8 находится теоретическая ф.р. 0,8-квантиль распределения Колмогорова примерно равна 1,07. То есть в нашем случае  $0,8 = \mathbb{P}(\sqrt{25}D_{25} < 1,07) = \mathbb{P}(D_{25} < 0,214)$ . И вот таким вот макаром мы получили зелёного крокодила на рис. 6.
2. Проверка гипотез. Эту тему мы будем мусолить позже, пока кратко. Вот кинули в нас выборкой, и сказали: «а ну ка отвечай, может это быть выборкой из такого-то распределения с такими-то параметрами?» Тут можно построить выборочную функцию распределения и посчитать, насколько максимум выборочная ф.р. отличается от теоретической, которую мы хотим проверить, и посмотреть вероятность, с которой можно встретить такое отклонение. Если вероятность маленькая, то, скорее всего, нам присунули какую-то левую выборку, а если большая, то надо кастовать магическое заклинание «данные не противоречат», о котором будет позже.

### 2.1.3. Ядерные оценки

Эмпирическая функция распределения великолепна и достойна всяческого восхищения. Но у неё есть один недостаток: она разрывна и не строго монотонна, а мы любим, когда все хорошо, монотонно и непрерывно. Что делать? Введем функцию  $k(x)$ , которую заставим обладать следующими свойствами:

$$k(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad (2.10)$$

$$\int_{-\infty}^{\infty} k(x) dx = 1, \quad (2.11)$$

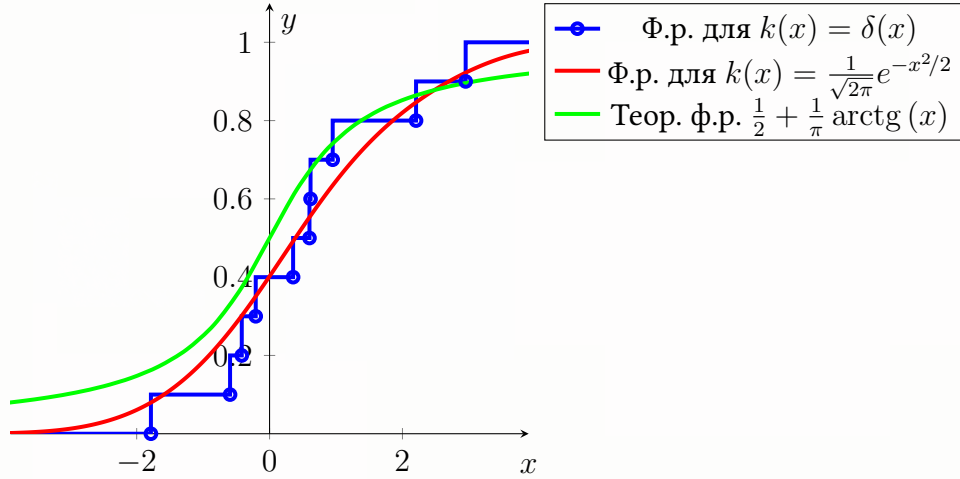


Рис. 9: Приближение теоретической функции распределения выборочной и ядерная оценка с использованием функции плотности нормального распределения.

$$\int_{-\infty}^{\infty} xk(x) dx = 0, \quad (2.12)$$

$$\int_{-\infty}^{\infty} x^2k(x) dx = 1. \quad (2.13)$$

Что сейчас только что произошло? Свойствами (2.10) и (2.11) мы сказали, что хотим корректную функцию плотности, а свойствами (2.12) и (2.13) мы пронормировали это безобразие на нулевое матожидание и единичную дисперсию. Теперь введем для нашего ядра функцию распределения:

$$K(x) := \int_{-\infty}^x k(x) d. \quad (2.14)$$

И в выражение для выборочной ф.р. подставим в общем случае ф.р. ядра:

$$F_n^K(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{x - X_i}{h_i}\right), \quad (2.15)$$

где  $X_i$  символизирует матожидание, то есть, что мы пляшем вокруг конкретного элемента выборки, а  $h_i$  позволяет управлять крутизной функции распределения и символизирует дисперсию. Теперь, если мы в качестве функции ядра выберем  $\delta$ -функцию (с оговоркой, что говорить про дисперсию будет совсем некорректно), то мы получим классическую выборочную функцию распределения, как в определении 2.2. Но поскольку самое пацанское распределение — это нормальное распределение, то в качестве ядра можно выбрать и  $k(x) = (2\pi)^{-1/2} e^{-x^2/2}$ , тогда получим оценку теоретической функции распределения, как на графике на рис 9.

## 2.2. Выборочные моменты

Все с детского садика знают, что чтобы посчитать что-то среднее, надо сложить все вместе и поделить на количество. А почему это так работает? Сейчас будем выяснять. Вводные такие: есть функция распределения  $F(x)$ , и из неё нам нагенерировали выборку  $\mathbb{X} = (X_1, \dots, X_n)$ . Пусть для этого распределения существуют матожидание и дисперсия:

$$a \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x dF(x), \quad (2.16)$$

$$\sigma^2 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x^2 dF(x) - a^2. \quad (2.17)$$

Теперь введём понятия выборочного среднего и выборочной дисперсии.

**Определение 2.6.** Выборочным средним называется  $\bar{X} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n X_i$ .

**Определение 2.7.** Выборочной дисперсией называется  $S^2 \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ .

Теперь исследуем, каким образом выборочное среднее приближаем матожидание. Переформулировав усиленный закон больших чисел, получаем:

$$\bar{X} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \mathbb{E}X_i. \quad (2.18)$$

Собственно, на этом мои полномочия всё. Про именно матожидание больше ничего интересного нет, будем ковырять дисперсию. Посчитаем  $\mathbb{E}S^2$ :

$$\begin{aligned} \mathbb{E}S^2 &= \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E} (X_i - a + a - \bar{X})^2 = \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} (X_i - a)^2 - \frac{2}{n} \sum_{i=1}^n \mathbb{E} (X_i - a) (\bar{X} - a) + \frac{1}{n} \sum_{i=1}^n \mathbb{E} (\bar{X} - a)^2 = \\ &= \sigma^2 - 2\mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n X_i - a \right) (\bar{X} - a) \right] + \mathbb{E} (\bar{X} - a)^2 = \\ &= \sigma^2 - \mathbb{E} (\bar{X} - a)^2 = \left\{ \begin{array}{l} \text{2-е слагае-} \\ \text{мое - дис-} \\ \text{персия} \end{array} \right\} = \sigma^2 - \mathbb{D}\bar{X} = \\ &= \left\{ \begin{array}{l} \text{Для норсв дисперсия сум-} \\ \text{мы равна сумме диспер-} \\ \text{сий и константа выносит-} \\ \text{ся с квадратом.} \end{array} \right\} = \sigma^2 - \frac{1}{n^2} \mathbb{D} \left[ \sum_{i=1}^n X_i \right] = \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2 \quad (2.19) \end{aligned}$$

О как. Немного флекса и мы получили так называемую смещенную оценку дисперсии. Наша оценка стабильно врет в  $\frac{n-1}{n}$  раз. Поэтому мы её немного поправим и получим



**Определение 2.8.** Исправленной (несмещенной) выборочной дисперсией называется

$$\bar{S}^2 \stackrel{\text{def}}{=} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Тут можно запутаться и забыть, там минус один или плюс. Поясним. Если выборка состоит из одного элемента, то говорить об *отклонении* (а дисперсия — это среднеквадратичное отклонение от среднего) не по понятиям. Поэтому формула и не работает, когда  $n = 1$ . А вообще этот  $n - 1$  в знаменателе лезет везде, где мы пытаемся оценить что-то, что выглядит как дисперсия или по смыслу напоминает дисперсию.

Теперь надо успеть рассмотреть выборочные моменты до того, как поедет кукуха. Обозначим  $\mu_k \stackrel{\text{def}}{=} \mathbb{E}X_i^k$  и попробуем их приблизить.

**Определение 2.9.** Выборочным моментом  $k$ -го порядка называется число

$$m_k \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Тут уже мы не будем особо ковыряться во всём этом безобразии и просто запишем ЦПТ, откуда уже можно получить примерное выражение для вероятности  $\mathbb{P}(m_k < x)$ . Естественно, матожиданием выборочного момента является теоретический момент.

$$\frac{\sqrt{n}(m_k - \mu_k)}{\sqrt{\mathbb{D}X_i^k}} = \frac{\sum_{i=1}^n X_i^k - n\mu_k}{\sqrt{n(\mu_{2k} - \mu_k^2)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.20)$$



## 2.3. Параметрическая статистика. Оценки

### 2.3.1. Определения

Сначала вспомним, что есть такая штука, как статистическая структура  $(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P})$ . Теперь немного уточним, что представляет собой  $\mathcal{P}$ . В точечных оценках и в доверительных интервалах мы будем считать, что  $\mathcal{P} = \mathbb{P}_\theta$ . То есть мы считаем, что наша семейство распределений параметризовано каким-то векторным параметром  $\theta \in \Theta \subset \mathbb{R}^m$ . Например,  $\mathbb{P}_\theta = \mathcal{N}(\theta_1, \theta_2^2)$ , тогда пространство параметров  $\Theta = (-\infty, \infty) \times (0, \infty)$ . Теперь, собственно, каким образом мы будем это оценивать.

**Определение 2.10.** Статистикой называется любая измеримая вектор-функция от выборки, не зависящая от  $\theta$ .

$$T(\mathbb{X}) = (T_1(\mathbb{X}), \dots, T_n(\mathbb{X}))^T: \mathbb{R}^n \mapsto \mathbb{R}^m.$$

**Определение 2.11.** Оценкой называется статистика, множество значений которой совпадает с  $\Theta$

**Определение 2.12.** Оценка называется несмещённой, если  $\mathbb{E}T(\mathbb{X}) = \theta$ .

**Определение 2.13.** Оценка называется состоятельной, если  $T(\mathbb{X}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\theta} \theta \quad \forall \theta \in \Theta$ . Оценка называется сильно состоятельной, если сходимость почти наверное.

Теперь надо примеров каких-нибудь накидать.

### 2.3.2. Примеры

**Пример.** В предыдущем параграфе мы ввели выборочное среднее (см. опр. 2.6). Пройдемся по каждому из этих 4-х пунктов, если положить  $T(\mathbb{X}) = \bar{X}$ , а  $\theta$  имеет смысл математического ожидания.

1. **Измеримость.** Да, выборочное среднее измеримо как композиция сложения и деления, а эти функции являются измеримыми.
2. **Множество значений.** Тут мы можем выстрелить себе в ногу или в какое другое место и искусственно ввести ограничение на пространство параметров, но зачем? Если не вводить никаких дурацких ограничений (например, оцениваем МО выборки из нормального распределения, но говорим, что МО обязательно не меньше 300), то все ОК.

3. **Несмещённость.**

$$\mathbb{E}\bar{X} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \mathbb{E}X_i$$

Следовательно, оценка является несмещённой.

4. **Состоятельность** сразу следует из ЗБЧ, а сильная состоятельность — из УЗБЧ.

**Пример.** Чуть менее подробно рассмотрим выборочную дисперсию  $\sigma^2$  (см. опр. 2.7).  $T(\mathbb{X}) = S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ . Как было показано ранее (2.19),  $T(\mathbb{X})$  не является несмещённой оценкой, но является состоятельной и даже сильно состоятельной. Хочется не париться и сказать, что вот здесь (2.19)  $\frac{n-1}{n}\sigma^2 \rightarrow \sigma^2$ , но не тут-то было. Там мы считали матожидание оценки, а нам нужна сходимость самой оценки. Тогда может к той сумме прикрутить УЗБЧ и всё получится? Снова нет, так как в УЗБЧ мы генерировали значения из одного распределения, а тут для разных  $n$  случайная величина  $(X_i - \bar{X})^2$ , возможно, будет распределена по-разному ( $\bar{X}$  зависит от  $n$ ). УЗБЧ тут применить можно, но не сразу, надо сначала немного повертеть буквами. Из этой (2.19) выкладки следует, что:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - (\bar{X} - a)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i - a \right)^2 \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \mathbb{E}(X_i - a)^2 - (a - a)^2 = \mathbb{D}X_i. \quad (2.21)$$

Как-то так. Из простейших свойств пределов (не, ну тут правда очевидно) следует, что исправленная (несмещённая) выборочная дисперсия также является состоятельной оценкой теоретической дисперсии.

**Пример.** Последнее, что хотелось бы тут рассмотреть и поковырять — несмещённую, но не состоятельную оценку. Тут придётся натянуть сову на глобус и рассмотреть выборку из  $\mathcal{N}(a, 1)$  и оценку матожидания  $T(\mathbb{X}) = \frac{n-1}{n}X_1 + \frac{1}{n(n-1)} \sum_{i=2}^n X_i$ . Докажем несмещённость:

$$\mathbb{E}T(\mathbb{X}) = \frac{n-1}{n}\mathbb{E}X_1 + \frac{1}{n}\mathbb{E}X_i = \frac{n-1}{n}a + \frac{1}{n}a = a. \quad (2.22)$$

Снова воспользовались нашим любимым ЗБЧ. А теперь несостоятельность:

$$\begin{aligned} \mathbb{P}(|T(\mathbb{X}) - a| < \varepsilon) &= \mathbb{P}\left(\left|\frac{n-1}{n}(X_1 - a) + \frac{1}{n(n-1)} \sum_{i=2}^n (X_i - a)\right| < \varepsilon\right) = \\ &= \mathbb{P}(|X_1 - a| < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) \xrightarrow[n \rightarrow \infty]{} 0. \end{aligned} \quad (2.23)$$

Мы воспользовались тем, что  $X_i$  одинаково распределены, а потом тем, что  $(X_i - a) \sim \mathcal{N}(0, 1)$ , и подставили соответствующую вероятность.

## 2.4. Функция правдоподобия

Вот мы рассмотрели оценки матожидания и дисперсии. Они хорошие: несмещённые, состоятельные и вообще милашки. Но что если нам нужны не они? Откуда можно брать оценки? Для этого придумали функцию правдоподобия. Сначала дадим пацанское определение функции правдоподобия, потом поковыряем парочку простых примеров, потом поймём, что с пацанским определением возникают некоторые проблемы, и дадим страшное и жуткое определение через лютый функан. Погнали!

**Определение 2.14.** Пусть дана выборка  $\mathbb{X} = (X_1, \dots, X_n)$  из распределения с плотностью  $p(x, \theta)$ , где  $\theta \in \Theta$  — параметр. Тогда функцией правдоподобия называется произведение плотностей.

$$L(\mathbb{X}, \theta) \stackrel{\text{def}}{=} \prod_{i=1}^n p(X_i, \theta).$$

**ЧО?! ЧЁ ЭТО ТАКОЕ?!** Жизненный смысл функции правдоподобия такой: вероятность наблюдать то, что мы наблюдали. Как ей пользоваться? Для начала отметим, что эта вероятность зависит от параметра. Тогда мы говорим, что параметр в распределении такой, что вероятность наблюдать то, что мы наблюдали, наибольшая. То есть, кто мы такие, чтобы наблюдать редкие события? Да никто! Недостойны мы такой чести. Поэтому полагаем параметр таким, что событие  $\mathbb{X}$  наиболее вероятное.

**Пример.** Пусть  $X_i \sim \mathcal{N}(\theta, 1)$ . Мы хотим оценить матожидание. Запишем функцию правдоподобия:

$$L(\mathbb{X}, \theta) = \prod_{i=1}^n p(X_i, \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(X_i - \theta)^2}{2} \right\}. \quad (2.24)$$

Дальше надо её максимизировать по  $\theta$ , но работать с произведением — себя не уважать, поэтому будем максимизировать не саму функцию правдоподобия, а её логарифм. Тут все законно, потому что логарифм — монотонная функция, и максимум логарифма достигается при том же  $\theta$ , при котором достигается максимум самой функции правдоподобия. Да и вообще, первое, что обычно делают с функцией правдоподобия — берут от неё логарифм и произведение становится суммой.

$$\ln L(\mathbb{X}, \theta) = - \sum_{i=1}^n \frac{(X_i - \theta)^2}{2} + \text{const}. \quad (2.25)$$

В  $\text{const}$  мы запихали сумму логарифмов  $\frac{1}{\sqrt{2\pi}}$ . Все равно мы будем сейчас максимизировать (= брать производную) по  $\theta$  и все константы убьются.

$$\frac{d}{d\theta} L(\mathbb{X}, \theta) = 0 \iff \sum_{i=1}^n (X_i - \theta) = 0 \iff \theta = \frac{1}{n} \sum_{i=1}^n X_i. \quad (2.26)$$

Паразитально! Оценили матожидание с помощью функции правдоподобия и получили выборочное матожидание. Кстати, параметр  $\theta$ , на котором достигается максимум функции правдоподобия, обычно обозначают  $\theta_{ML}$ , от слов «maximum likelihood».

**Пример.** Теперь пусть  $X_i \sim \mathcal{N}(a, \theta^2)$ . Попробуем теперь кастануть ММП (метод максимального правдоподобия).

$$L(\mathbb{X}, \theta) = \prod_{i=1}^n \frac{1}{\theta \sqrt{2\pi}} \exp \left\{ -\frac{(X_i - a)^2}{2\theta^2} \right\}$$

$$\frac{d}{d\theta} \ln L(\mathbb{X}, \theta) = -\sum_{i=1}^n \frac{1}{\theta} + \sum_{i=1}^n \frac{(X_i - a)^2}{\theta^3} \quad (2.27)$$

$$\left. \frac{d}{d\theta} \ln L(\mathbb{X}, \theta) \right|_{\theta=\theta_{ML}} = 0 \implies \theta_{ML} = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2.$$

Получили выборочную дисперсию. Вспомним, что выборочная дисперсия смещённо оценивает теоретическую дисперсию. Но никто и не обещал, что ММП даёт несмещённые оценки. Было бы неплохо в этих двух примерах показать, что мы получили именно максимум, а не какую-то левую стационарную точку, но мы не будем.

**Пример.** А теперь рассмотрим пример, в котором непонятно, какую оценку брать.  $X_i \sim U[0, \theta]$  — равномерное распределение на отрезке, с неизвестным верхним концом. Ну по классике:

$$L(\mathbb{X}, \theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbb{1}(X_i \in [0, \theta]) = \theta^{-n} \mathbb{1}(X_{(n)} \leq \theta). \quad (2.28)$$

Как мы так упростили произведение индикаторов? Так как по условию  $X_i$  распределена от 0 до  $\theta$ , то  $X_i \geq 0$  заведомо, и эту часть индикаторов можно выкинуть. Теперь произведение индикаторов того, что все  $X_i \leq \theta$  равносильно индикатору того, что каждая  $X_i \leq \theta$ , а это уже равносильно тому, что максимум не больше  $\theta$ . Переходим к максимизации этого безобразия. Тут нельзя просто так взять и продифференцировать из-за индикаторов. Попробуем по-другому: тут стоит  $\theta^{-n}$ . Эта штука большая (максимизируем же), когда  $\theta$  маленькая. Но если  $\theta$  будет меньше максимального элемента выборки  $X_{(n)}$ , то вся наша максимизация пойдёт лесом, так как один из индикаторов обнулится. Поэтому  $\theta_{ML} = X_{(n)}$ .

Всё, хватит пока примеров. Казалось, бы, всё просто шоколадно, но что если нам дадут выборку из случайных величин, которые принимают значения 1 и 0 с вероятностями  $p$  и  $1 - p$ ? Какую плотность взять для функции правдоподобия? Правильный ответ — любую удобную нам, для которой интеграл Лебега по считающей мере по произвольному борелевскому множеству будет выдавать нужные вероятности. Сейчас введём всю нужную математику, а потом вернёмся к этому примеру и придумаем, как тут быть.

**Определение 2.15.** Пусть на измеримом пространстве  $(\Omega, \mathcal{A})$  заданы меры  $\mu$  и  $\nu$ . Мера  $\nu$  называется абсолютно непрерывной относительно меры  $\mu$ , если  $\forall A \in \mathcal{A}, \mu(A) = 0 \implies \nu(A) = 0$ . Обозначается  $\nu \ll \mu$ .

**Теорема 2.16** (Радóна—Никóдима). Если на измеримом пространстве заданы меры  $\nu$  и  $\mu$  и при этом  $\nu \ll \mu$ , то существует почти всюду единственная измеримая функция  $X(\omega)$ , такая что

$$\forall A \in \mathcal{A} \implies \nu(A) = \int_A X(\omega) \mu(d\omega).$$



**ЧО?! ЧЁ ЭТО ТАКОЕ?! Эта теорема обобщает понятие производной на всякие различные кривые функции.**

**Зочем?** А вот зачем: пусть у нас на измеримом пространстве вдруг задана вероятность событий, причем распределение дискретно. А мы хотим что-то оценить методом максимального правдоподобия. Тогда теорема говорит, что если мы найдем такую меру  $\mu$ , относительно которой наша вероятностная мера будет абсолютно непрерывна, то существует плотность распределения, которая определена однозначно с точностью до значений на множестве  $S$ , такого что  $\mu(S) = 0$ .

**Определение 2.17.** Статистическая структура допускает функцию правдоподобия, если существует  $\mu$ , определенная на  $\sigma$ -алгебре событий  $\mathcal{A}$ , относительно которой вероятностные меры нашей статистической структуры абсолютно непрерывны.

**Определение 2.18.** Пусть задано не более чем счётное число точек  $A = \{a_1, \dots, a_n, \dots\}$ . Мера называется считающей, если для любого борелевского множества его мера равна числу точек из  $A$ , принадлежащих этому множеству.

**Пример.** Теперь вернёмся к тому примеру с дискретной случайной величиной:  $\mathbb{P}(\xi = 1) = p$ ,  $\mathbb{P}(\xi = 0) = 1 - p$ . Сначала найдем меру, относительно которой вероятностная мера будет абсолютно непрерывна. Это будет считающая мера  $\mu_A(B)$ , где  $A = \{0, 1\}$  — множество, по которому мы будем считать. Эта мера сопоставляет каждому борелевскому множеству число 0, если в него не попала ни одна из точек  $\{0, 1\}$ , 1, если попала ровно одна из этих точек и 2, если попали обе точки. Проверим условие абсолютной непрерывности. Действительно, вероятность только трех событий больше нуля: 0, 1 и оба сразу. Соответственно, вероятностная мера выдаст ноль на тех и только на тех множествах, которые не содержат ни одну из этих точек. Наша считающая мера выдает ноль на тех же множествах. А значит  $\forall B \in \mathcal{B}$ , если  $\mu(B) = 0$ , то и  $\mathbb{P}(B) = 0$ , и вероятностная мера является абсолютно непрерывной относительно меры  $\mu$ .

Теперь будем разбираться с правдоподобием. По теореме Радона—Никодима функция правдоподобия почти всюду единственна. В нашем случае почти всюду — это в двух точках, мера которых отлична от нуля: 0 и 1. Причем должно быть выполнено:

$$\begin{cases} L(0, p) = 1 - p, \\ L(1, p) = p, \end{cases}$$

чтобы интеграл от неё действительно давал нужную вероятность. Лулзов ради накидаем пяток подходящих функций, а потом выберем удобную:

$$L(x, p) = 1 - p + (2p - 1)x, \quad (2.29a)$$

$$L(x, p) = p\mathbb{1}(x = 1) + (1 - p)\mathbb{1}(x = 0) \quad (2.29b)$$

$$L(x, p) = (1 - p) \cos\left(\frac{\pi x}{2}\right) + p \sin\left(\frac{\pi x}{2}\right) \quad (2.29c)$$

$$L(x, p) = 1 - p + \frac{8p - 4}{\pi} \arctg(x), \quad (2.29d)$$

$$L(x, p) = p^x(1 - p)^{1-x}. \quad (2.29e)$$



В принципе, все эти функции подойдут, и в случае, когда выборка состоит из одного элемента, можно использовать любую, но пусть у нас теперь  $\mathbb{X} = (X_1, \dots, X_n)$ . Сейчас будет понятно, почему (2.29e) — самый удобный вариант.

$$L(\mathbb{X}, p) = \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i} (1-p)^{n - \sum_{i=1}^n X_i} \quad (2.30)$$

$$\frac{d}{dp} \ln L(\mathbb{X}, p) = \frac{1}{p} \sum_{i=1}^n X_i - \frac{1}{1-p} \left( n - \sum_{i=1}^n X_i \right) = 0 \quad (2.31)$$

$$p = \frac{1}{n} \sum_{i=1}^n X_i. \quad (2.32)$$

Получили ту самую оценку МО в схеме бернулли. Самые отбитые могут попробовать проверить это с другой функцией правдоподобия — будет долго и больно.



## Список литературы

- [1] Королев В. Ю. Теория вероятностей и математическая статистика: учеб. М.: ТК Вебли, Изд-во Проспект, 2006. с. 160.
- [2] Гусев Н. А. Заметки к курсу: мера и интеграл Лебега. 2019. <https://www.ngusev.ru/mt/mt-lectures.pdf>.

