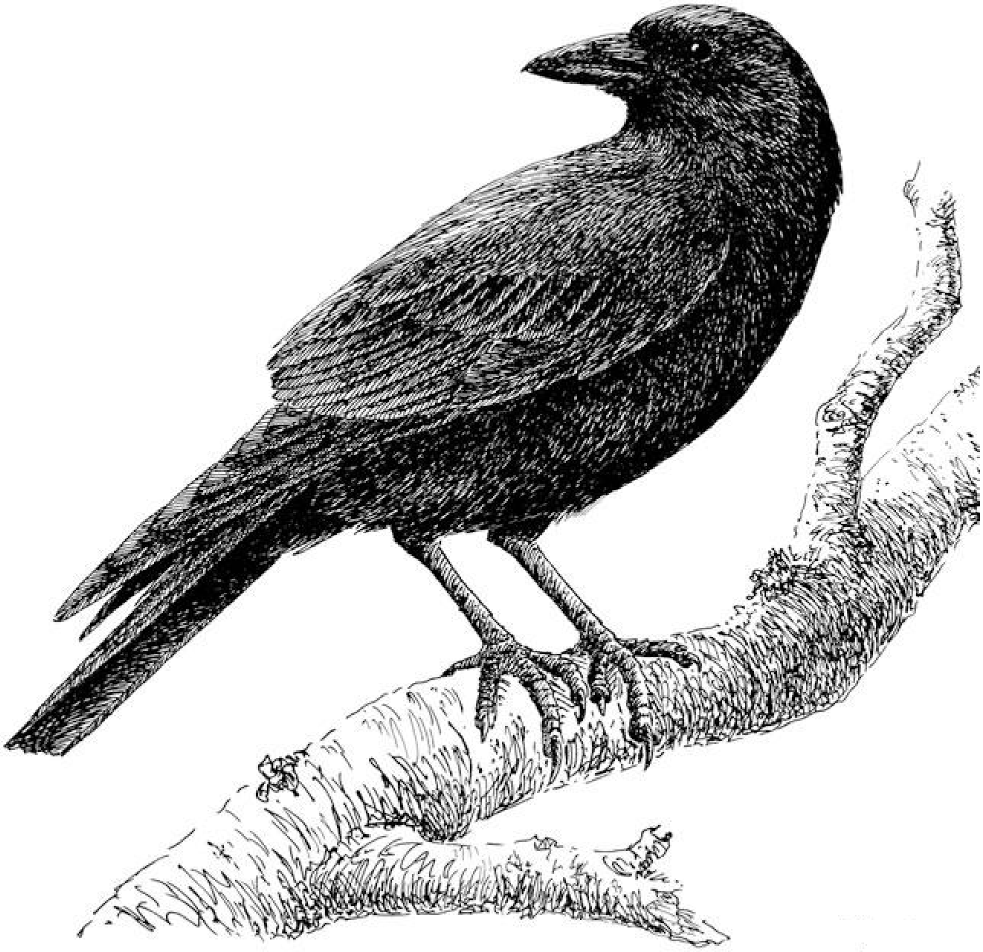


Теория вероятностей

Конспекты лекций и семинаров

Лектор: Д.А. Шабанов

Конспекты вели Денис Беляков, Никита Попов и Алексей Хачиянц

НИУ ВШЭ, 2016-2017

Оглавление

Глава 1

Лекции

1.1 Лекция от 09.09.2016

1.1.1 Введение. Принцип устойчивости частоты

Чем занимается теория вероятностей? Она изучает *случайные* явления. Допустим, мы провели какой-либо эксперимент. Можем ли мы что-то заранее сказать о результате?

- Если да, то результат называют *детерминированным*. Пример такого эксперимента — выбрасывание кирпича из окна. Очевидно, что кирпич упадёт на землю[1] и результат предопределён. Такие задачи изучают в той же линейной алгебре или где-либо ещё, но не в теории вероятностей.

- А теперь предположим, что заранее сказать, каков будет результат, невозможно. Например, точно сказать, какой стороной упадёт подброшенная монетка, вряд ли получится. Тогда результат называют *недетерминированным*. Именно задачи с недетерминированным результатом и изучаются в теории вероятностей.

Небольшое историческое отступление — вообще говоря, теория вероятностей появилась в связи с изучением азартных игр наподобие рулетки ещё в средних веках. Но тогда она представляла собой скорее набор эмпирических фактов, чем полноценную науку. Теория вероятностей стала такой, какой она является сейчас, лишь в XX веке благодаря трудам А.Н. Колмогорова.

Хорошо, а как изучаются случайные процессы? Ну выпала решка, и что? На самом деле теория вероятностей не о единичных экспериментах, а об *асимптотике*. Это значит следующее: если проводить серию одинаковых экспериментов, то теория вероятностей поможет предсказать частоту, с которой будет появляться какой-либо ответ.

Теория вероятностей держится на крайне важном *принципе устойчивости частоты*. Перед тем, как ввести формальное определение, рассмотрим пару экспериментов, связанных с подбрасыванием монетки:

- В XVIII веке Жорж-Луи Леклерк де Бюффон провёл эксперимент, подбросив монетку 4040 раз. Из них в 2048 бросках выпал герб. В итоге частота составила около 0.506.

- В XIX веке пошли ещё дальше — Карл Пирсон подбросил монетку 24000 раз.[2] У него получилось так, что герб выпал 12012 раз. В итоге частота составила 0.5005.

[1] Если его не запустили с первой космической скоростью, конечно.

[2] Оставим вопрос о том, как он не поленился провернуть это, без ответа.

Отсюда видно, что эксперименты дают частоту, близкую к $1/2$.

Неформально говоря, принцип формулируется так: если мы проводим серию одинаковых экспериментов, то количество появлений одного определённого ответа при делении на число экспериментов сходится к некоторому числу $p \in [0,1]$. Теперь можно ввести формальное определение.

Принцип устойчивости частоты. *Пусть A — некоторое событие, а $v_n(A)$ — число экспериментов, в которых происходит событие A среди первых n. Тогда*

$$\lim_{n\to\infty} \frac{v_n(A)}{n} = p, \quad p \in [0,1].$$

Получаемое число p называют *вероятностью* события A и обозначают $\mathsf{P}(A)$. Например, $\mathsf{P}(\text{встретить живого динозавра на улице}) = 0$, так как они все вымерли.

1.1.2 Вероятностное пространство. Простейшие свойства вероятности

Именно после введения этого понятия Колмогоровым теория вероятностей перестала быть прежней. Введём определение для дискретного случая (общий оставим на потом):

Определение 1. *Дискретным вероятностным пространством* называется пара (Ω, P), где Ω — *множество элементарных исходов*, а P — *вероятность* на Ω.

Множество элементарных исходов Ω — некоторое конечное или счётное множество. Элемент $\omega \in \Omega$ называют *элементарным исходом.* Полагается, что в случайном эксперименте обязательно получается один и только один элементарный исход.

Примеры множеств элементарных исходов:

1. $\Omega = \{\text{О}, \text{Р}\}$ — бросок монеты.
2. $\Omega = \{\omega_1, \ldots, \omega_6\}$, где $\omega_i = \{\text{выпало } i \text{ очков}\}$ — бросок игрального кубика.
3. $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$, где $\omega_i = \{\text{на данный момент горит } i \text{ зданий}\}$ — предсказание пожаров в городе.

Определение 2. Подмножество $A \subseteq \Omega$ называется *событием* на вероятностном пространстве (Ω, P).

Пример события: пусть подбрасывают игральную кость, и $A = \{\text{выпало чётное число очков}\}$. Тогда $A = \{\omega_2, \omega_4, \omega_6\}$.

Определение 3. Отображение $\mathsf{P} : \Omega \to [0,1]$ называют *вероятностью*[3], если $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$. В случае счётного множества Ω данный ряд должен сходиться абсолютно.

Пусть у нас есть некоторое событие A на вероятностном пространстве (Ω, P). Как посчитать его вероятность?

Определение 4. *Вероятностью события $A \subseteq \Omega$* называют $\mathsf{P}(A) = \sum\limits_{\omega\in A} \mathsf{P}(\omega)$.

[3] В общем случае вероятность ещё могут называть *вероятностной мерой.*

Определение 5. Пусть A — некое событие. Тогда *дополнением* к событию A называют событие $\overline{A} = \Omega \setminus A$.

Перед тем, как идти дальше, напомним определение дизъюнктного объединения.

Определение 6. Пусть есть множества $A_1, A_2, \ldots, A_n$. Тогда *дизъюнктным объединением* множеств называют объединение попарно непересекающихся "копий" множеств:

$$\bigsqcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} \{(x, i) \mid x \in A_i\}.$$

В нашем случае полагается, что если пишут дизъюнктное объединение, то множества попарно не пересекаются.

Рассмотрим некоторые свойства вероятности.

Теорема 1 (Простейшие свойства вероятности)**.** *Для любого дискретного вероятностного пространства* (Ω, P) *выполняется следующее:*

1. $\mathsf{P}(\Omega) = 1$, $\mathsf{P}(\varnothing) = 0$.
2. *Конечная аддитивность:* $\mathsf{P}\left(\bigsqcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i)$.
3. $\mathsf{P}(A) + \mathsf{P}(\overline{A}) = 1$.
4. $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$.
5. *Для любого набора событий* $A_1, A_2, \ldots, A_n$ $\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathsf{P}(A_i)$.
6. *Счётная аддитивность:* $\mathsf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$.

Последнее свойство выполняется только для счётного Ω.

Доказательство. Распишем доказательства для каждого пункта по отдельности:

1. $\mathsf{P}(\Omega) = 1$ следует из определения вероятности, а $\mathsf{P}(\varnothing) = 0$ следует из определения вероятности события.
2. Случай с конечным множеством Ω очевиден. Тогда положим, что Ω счётно, то есть $\Omega = \{\omega_i \mid i \in \mathbb{N}\}$.

   Пусть есть некоторое событие A. Тогда представим его вероятность в удобном для нас виде:
   $$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega) = \sum_{i\,:\,\omega_i \in A} \mathsf{P}(\omega_i) = \lim_{N \to \infty} \sum_{\substack{i\,:\,\omega_i \in A \\ i < N}} \mathsf{P}(\omega_i).$$

   Теперь распишем вероятность дизъюнктного объединения событий $A_1, A_2, \ldots, A_n$:
   $$\mathsf{P}\left(\bigsqcup_{i=1}^{n} A_i\right) = \sum_{\omega \in \bigsqcup A_i} \mathsf{P}(\omega) = \lim_{N \to \infty} \left(\sum_{\substack{i\,:\,\omega_i \in \bigsqcup A_i \\ i < N}} \mathsf{P}(\omega_i) \right) = \lim_{N \to \infty} \sum_{i=1}^{n} \sum_{\substack{j\,:\,\omega_j \in A_i \\ j < N}} \mathsf{P}(\omega_j) =$$
   $$= \sum_{i=1}^{n} \lim_{N \to \infty} \sum_{\substack{j\,:\,\omega_j \in A_i \\ j < N}} \mathsf{P}(\omega_j) = \sum_{i=1}^{n} \mathsf{P}(A_i).$$

3. Согласно второму пункту, $1 = \mathsf{P}(\Omega) = \mathsf{P}(A \cup \overline{A}) = \mathsf{P}(A) + \mathsf{P}(\overline{A})$.

4. Так как $A \cup B = A \cup (B \setminus A)$ и $A \cap (B \setminus A) = \varnothing$, то $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$. Далее, заметим, что $B \setminus A = B \setminus (A \cap B)$. Тогда $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus (A \cap B))$.

   Рассмотрим второй член. Заметим, что $\mathsf{P}(A \cap B) + \mathsf{P}(B \setminus (A \cap B)) = \mathsf{P}(B)$. Тогда получаем, что $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$.

5. Докажем это утверждение по индукции. База была доказана в пункте 4 (так как $\mathsf{P}(A \cap B) \geqslant 0$). Теперь рассмотрим шаг индукции. Пусть утверждение верно для какого-то m. Тогда $\mathsf{P}\left(\bigcup\limits_{i=1}^{m+1} A_i\right) \leqslant \mathsf{P}(A_{m+1}) + \mathsf{P}\left(\bigcup\limits_{i=1}^{m} A_i\right) \leqslant \sum\limits_{i=1}^{m+1} \mathsf{P}(A_i)$.

6. За доказательством этого пункта обращайтесь к учебнику матанализа на тему частичных сумм и абсолютной сходимости.[4]

□

1.1.3 Классическая модель. Примеры

Пусть (Ω, P) — некоторое конечное вероятностное пространство, при этом все элементарные исходы равновероятны. Тогда легко посчитать вероятность элементарного исхода: $\mathsf{P}(\omega) = \frac{1}{|\Omega|}$ для всех $\omega \in \Omega$. Такую модель называют *классической*.

Как посчитать вероятность события в классической модели? Очень просто: $\mathsf{P}(A) = \frac{|A|}{|\Omega|}$ для всех $A \subseteq \Omega$.

Рассмотрим некоторые примеры классических моделей.

1. Бросок монетки. В таком случае $\Omega = \{\text{О}, \text{Р}\}$ и $\mathsf{P}(\text{О}) = \mathsf{P}(\text{Р}) = \frac{1}{2}$.

2. Бросок двух монеток. С этой моделью связано одно заблуждение Д'Аламбера. Он рассуждал следующим образом: так как $\Omega = \{\text{ОО}, \text{РР}, \text{ОР}\}$, то $\mathsf{P}(\text{ОО}) = \mathsf{P}(\text{РР}) = \mathsf{P}(\text{ОР}) = \frac{1}{3}$. Но это опровергается экспериментами. И как это исправить? Есть два варианта:

   (a) Можно сказать, что модель не является классической и поправить вероятности: $\mathsf{P}(\text{ОО}) = \mathsf{P}(\text{РР}) = \frac{1}{4}$, $\mathsf{P}(\text{ОР}) = \frac{1}{2}$.

   (b) А можно просто изменить множество элементарных исходов. Начнём учитывать порядок выпадения: $\Omega = \{\text{ОО}, \text{РР}, \text{ОР}, \text{РО}\}$. Такая модель уже является классической.

   Рассуждая в стиле Д'Аламбера, можно прийти к выводу, что вероятность встретить живого динозавра на улице равна $\frac{1}{2}$, ведь его можно либо встретить, либо не встретить.

3. Бросок n монет. В таком случае вероятностное пространство будет устроено следующим образом: $\Omega = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) \mid \omega_i \in \{\text{О}, \text{Р}\}\}$. Легко понять, что в данной модели 2^n элементарных исходов.

   Примечание. Вероятностное пространство такого вида называют *симметрической схемой Бернулли.*

[4] На самом деле я попробую найти доказательство. Когда-нибудь. Но не сейчас. (А.Х.)

Но данная модель является классической только тогда, когда монетки “честные”, то есть которые падают орлом или решкой вверх равновероятно. Если же это не так, то вероятность элементарного исхода $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ задаётся следующей формулой: $\mathsf{P}(\omega) = p^{\sum\limits_{i=1}^{n}\omega_i}(1-p)^{n-\sum\limits_{i=1}^{n}\omega_i}$.

4. Урновые схемы (размещение частиц по ячейкам). Пусть есть n различных шаров в ящике. Мы случайным образом вынимаем m шаров. Вопрос: каков размер множества элементарных исходов? Сначала приведём ответ, после чего докажем его.

Порядок? / Возврат?	Упорядоченный набор	Неупорядоченный набор
С возвратом	n^m	$C_{n+m-1}^m = \dfrac{(n+m-1)!}{m!(n-1)!}$
Без возврата	$A_n^m = \dfrac{n!}{(n-m)!}$	$C_n^m = \dfrac{n!}{m!(n-m)!}$

Доказательство. Будем доказывать утверждения от верхнего левого против часовой стрелки.

(a) Пусть набор упорядочен и можно возвращать. Тогда любой элемент набора можно получить n способами (так как все элементы можно вернуть). Отсюда получаем n^m.

(b) Теперь положим, что набор упорядочен, но возвращать нельзя. Тогда первый элемент можно выбрать n способами, второй — $n-1$ способом и так далее до m-го элемента, который можно выбрать $n-m+1$ способом. По правилу умножения получаем $\frac{n!}{(n-m)!} = A_n^m$.

(c) Рассмотрим случай, когда набор неупорядочен и возвращать нельзя. Тогда необходимо посчитать количество способов выбрать k шаров из m. Достаточно логично, что это равно $\frac{A_n^m}{m!} = \frac{n!}{m!(n-m)!} = C_n^m$, так как в последовательности нам не важен порядок.

(d) Осталось рассмотреть последний случай — неупорядоченный набор с возвратом. В этом случае нам достаточно указать, сколько раз мы выбрали каждый шар. Как это сделать? Воспользуемся методом точек и перегородок. Пусть есть m точек и нужно распределить их по n группам. Для этого нужно использовать $n-1$ перегородку. Тогда задача сводится к нахождению количества способов выбрать m элементов из $n+m-1$. А это равно $C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}$.

□

1.1.4 Условная вероятность. Формула полной вероятности

Пусть (Ω, P) — дискретное вероятностное пространство.

Определение 7. Пусть $A \subseteq \Omega$ — некоторое событие и $B \subseteq \Omega$ — другое событие, причём $\mathsf{P}(B) > 0$. Тогда *условной вероятностью события A при условии B* называют

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

Если $\mathsf{P}(B) = 0$, то положим, что $\mathsf{P}(A \mid B) = 0$ для любого события $A \subseteq \Omega$.

Условную вероятность можно воспринимать следующим образом: сузим множество элементарных исходов до B и посчитаем вероятность события A на полученном множестве.

Примечание. Если $\mathsf{P}(B) > 0$, то $\tilde{\mathsf{P}}(A) = \mathsf{P}(A \mid B)$ тоже является вероятностью на Ω.

Определение 8. Пусть $B_1, B_2, \ldots, B_n$ — некоторые события на Ω такие, что $\bigsqcup\limits_{i=1}^{n} B_i = \Omega$. Тогда этот набор событий называется (конечным) *разбиением* Ω.

Теперь докажем важную формулу:

Формула полной вероятности. *Пусть $B_1, B_2, \ldots, B_n$ — разбиение Ω. Тогда для любого события $A \subseteq \Omega$ верно, что $\mathsf{P}(A) = \sum\limits_{i=1}^{n} \mathsf{P}(A \mid B_i)\,\mathsf{P}(B_i)$.*

Доказательство. Так как $A \cap \Omega = A$ и $\bigsqcup\limits_{i=1}^{n} B_i = \Omega$, то $\mathsf{P}(A) = \mathsf{P}\left(A \cap \left(\bigsqcup\limits_{i=1}^{n} B_i\right)\right)$. Заметим, что $A \cap \left(\bigsqcup\limits_{i=1}^{n} B_i\right) = \bigsqcup\limits_{i=1}^{n}(A \cap B_i)$. Тогда $\mathsf{P}(A) = \sum\limits_{i=1}^{n} \mathsf{P}(A \cap B_i) = \sum\limits_{i=1}^{n} \mathsf{P}(A \mid B_i)\,\mathsf{P}(B_i)$. □

Заметим, что формула полной вероятности работает и в случае, когда $\mathsf{P}(B_i) = 0$ для какого-то i.

1.2 Лекция от 16.09.2016

1.2.1 Классические задачи теории вероятностей

Задача 1 (Задача о сумасшедшей старушке)**.** *Есть самолёт на n мест, в который садятся n пассажиров. Первой в него заходит безумная старушка, которая садится на случайное место. Каждый следующий пассажир действует по следующему правилу: садится на своё место, если оно свободно, и на случайное, если своё занято. С какой вероятностью*

1. *последний пассажир сядет на своё место?*
2. *предпоследний пассажир сядет на своё место?*
3. *и последний, и предпоследний пассажир сядут на свои места?*

Решение. У первого пункта есть элементарное решение. Пусть при некоторой рассадке пассажиров последний пассажир сел не на свое место (такую рассадку назовем неудачной). Тогда до прихода последнего пассажира его место было занято пассажиром A (A может быть и сумасшедшей старушкой). В момент прихода пассажира A перед ним стоит выбор — какое место занять. В рассматриваемой рассадке он занимает место последнего пассажира. Но с той же вероятностью он мог занять и место старушки, и в дальнейшем все пассажиры, включая последнего, займут свои собственные места. (Конечно, нужно еще пояснить, почему в момент прихода пассажира A старушкино место все еще свободно. Но это действительно так — нетрудно проследить, что пока старушкино место свободно, среди всех еще не вошедших пассажиров есть ровно один, чье место уже занято. Как только очередной пассажир занимает старушкино место, все остальные будут садиться только на свои места.) Таким образом, каждой неудачной рассадке соответствует удачная, которая может случиться с той же вероятностью. Это говорит о том, что ровно в половине случаев рассадка будет неудачной.

Теперь рассмотрим формальное решение для первого пункта. Пусть $A = \{$последний сядет на своё место$\}$. Если[5] $n = 2$, то $\mathsf{P}(A) = \frac{1}{2}$. Теперь рассмотрим случай $n = 3$. Пусть $B_i = \{$бабушка есла на i-е место$\}$. По формуле полной вероятности

$$\mathsf{P}(A) = \mathsf{P}(A \mid B_1)\,\mathsf{P}(B_1) + \mathsf{P}(A \mid B_2)\,\mathsf{P}(B_2) + \mathsf{P}(A \mid B_3)\,\mathsf{P}(B_3) = \frac{1}{3}\left(1 + \frac{1}{2} + 0\right) = \frac{1}{2}.$$

Намечается закономерность. Попробуем сформулировать гипотезу и доказать её:

Гипотеза. *Для любого n вероятность того, что последний пассажир сядет на своё место, равна $\frac{1}{2}$.*

Доказательство. По индукции. База ($n = 2$) была доказана ранее. Теперь предположим, что утверждение верно для всех $k < n$. Тогда докажем, что утверждение верно для $k = n$. Опять же, распишем вероятность по формуле полной вероятности:

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A \mid B_i)\,\mathsf{P}(B_i).$$

Заметим, что $\mathsf{P}(B_i) = \frac{1}{n}$. Теперь посмотрим на значения условных вероятностей:

$$\mathsf{P}(A \mid B_i) = \begin{cases} 1, & i = 1 \\ 0, & i = n \\ \frac{1}{2}, & 2 \leqslant i \leqslant n - 1 \end{cases}$$

Последнее условие на значение условной вероятности следует из шага индукции, так как i-й пассажир “становится” бабушкой. Тогда $\mathsf{P}(A) = \frac{1}{n}\left(1 + \frac{1}{2} + \ldots + \frac{1}{2} + 0\right) = \frac{1}{2}$. □

Перейдём ко второму пункту задачи. Пусть $C = \{$предпоследний пассажир сел на своё место$\}$. Тогда рассмотрим $\mathsf{P}(C)$ в случае, когда $n = 3$. Он сможет сесть на своё место тогда и только тогда, когда бабушка не села на его место. Тогда $\mathsf{P}(C) = \frac{2}{3}$. Попробуем доказать гипотезу, аналогичную случаю с последним пассажиром.

Гипотеза. *Для любого n вероятность того, что предпоследний пассажир сядет на своё место, равна $\frac{2}{3}$.*

Доказательство. Доказательство практически такое же, как и для последнего пассажира, с тем отличием, что значения условной вероятности будут несколько другие:

$$\mathsf{P}(C \mid B_i) = \begin{cases} 1, & i = 1, n \\ 0, & i = n - 1 \\ \frac{2}{3}, & 2 \leqslant i \leqslant n - 2 \end{cases}$$

Тогда $\mathsf{P}(C) = \frac{1}{n}\left(1 + \frac{2}{3} + \ldots + \frac{2}{3} + 0 + 1\right) = \frac{2}{3}$ □

[5] Это уже вертолёт, скорее. (Д.А. Шабанов)

Аналогичными рассуждениями можно доказать, что вероятность того, что i-й с конца пассажир сядет на своё место, равна $\frac{i}{i+1}$.

Теперь приступим к третьему пункту. Докажем следующее утверждение:

Гипотеза. *Вероятность этого события (обозначим его за D) равна $\frac{1}{3}$.*

Доказательство. Опять же, по индукции. Базой служит случай $n = 3$. В таком случае условие выполнимо тогда и только тогда, когда бабушка сядет на своё место. Тогда вероятность равна $\frac{1}{3}$ и база верна.

Перейдём к шагу индукции. Рассуждаем абсолютно аналогично: по формуле полной вероятности $\mathsf{P}(D) = \sum\limits_{i=1}^{n} \mathsf{P}(D \mid B_i)\,\mathsf{P}(B_i)$. Значения условных вероятностей равны

$$\mathsf{P}(D \mid B_i) = \begin{cases} 1, & i = 1 \\ 0, & i = n-1, n \\ \frac{1}{3}, & 2 \leqslant i \leqslant n-2 \end{cases}$$

и $\mathsf{P}(D) = \frac{1}{n}\left(1 + \frac{1}{3} + \ldots + \frac{1}{3} + 0 + 0\right) = \frac{1}{3}$. □

Заметим, что $\mathsf{P}(D) = \mathsf{P}(A)\,\mathsf{P}(C)$. Это ~~счастливое совнадение~~ *неспроста.* □

Задача 2 (Задача об удачливом студенте)**.** *Студент начал готовиться к экзамену слишком поздно и выучил только k билетов из n. Студент решил схитрить и выбрать место в очереди такое, чтобы вероятность получить выученный билет была максимальной. Какое место ему выбрать?*

Решение. Пусть $A_s = \{$студент вытащил хороший билет, если он встал s-тым в очереди$\}$. Докажем следующую гипотезу:

Гипотеза. $\mathsf{P}(A_s) = \frac{k}{n}$.

Доказательство. Введём разбиение $B_0, B_1, \ldots, B_k$ пространства Ω, где $B_i = \{$до студента взяли ровно i хороших билетов$\}$. Тогда по формуле полной вероятности

$$\mathsf{P}(A_s) = \sum_{i=0}^{k} \mathsf{P}(A_s \mid B_i)\,\mathsf{P}(B_i).$$

Посчитаем $\mathsf{P}(B_i)$. Для определения количества успешных исходов нужно выбрать i билетов из k хороших и $s-i-1$ из $n-k$ плохих. Тогда

$$\mathsf{P}(B_i) = \frac{\binom{k}{i}\binom{n-k}{s-i-1}}{\binom{n}{s-1}}.$$

Теперь надо определить значение условной вероятности $\mathsf{P}(A_s \mid B_i)$. Так как есть $n-s+1$ невыбранный билет и $k-i$ из них изучены, то $\mathsf{P}(A_s \mid B_i) = \frac{k-i}{n-s+1}$. Тогда

$$\mathsf{P}(A) = \sum_{i=0}^{k} \frac{k-i}{n-s+1} \frac{\binom{k}{i}\binom{n-k}{s-i-1}}{\binom{n}{s-1}} = \frac{k}{n} \sum_{i=0}^{k} \frac{\binom{k-1}{i}\binom{n-k}{s-i-1}}{\binom{n-1}{s-1}}.$$

Теперь заметим, что сумма справа равна 1, так как она соответствует разбиению пространства в случае, когда всего задач $n-1$. Тогда $\mathsf{P}(A_s) = \frac{k}{n}$. □

В итоге как ни вставай — всё равно никакой разницы не будет.[6] □

[6]Ибо по закону подлости попадётся невыученный билет.

1.2.2 Теорема Байеса

Теорема 2 (Байес)**.** *Пусть* $\{B_1, B_2, \ldots, B_n\}$ *— разбиение* Ω*, причём* $\mathsf{P}(B_i) > 0$ *для всех* $i \in \{1, 2, \ldots, n\}$*. Тогда для любого события* A *такого, что* $\mathsf{P}(A) > 0$*, выполняется*

$$\mathsf{P}(B_i \mid A) = \frac{\mathsf{P}(A \mid B_i)\,\mathsf{P}(B_i)}{\sum\limits_{j=1}^{n} \mathsf{P}(A \mid B_j)\,\mathsf{P}(B_j)}$$

Доказательство. По определению условной вероятности $\mathsf{P}(B_i \mid A) = \frac{\mathsf{P}(A \cap B_i)}{\mathsf{P}(A)} = \frac{\mathsf{P}(A|B_i)\,\mathsf{P}(B_i)}{\mathsf{P}(A)}$. Тогда, применяя формулу полной вероятности для $\mathsf{P}(A)$, получаем желаемое. □

1.2.3 Независимость событий

Определение 9. События A и B на вероятностном пространстве (Ω, P) называются *независимыми*, если $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$. Иногда используется обозначение $A \perp\!\!\!\perp B$.

Рассмотрим некоторые примеры независимых событий:

1. Задача про сумасшедшую бабушку: События $A = \{$последний сядет на своё место$\}$ и $B = \{$предпоследний сядет на своё место$\}$ независимы, как было доказано ранее.

2. Бросок игральной кости. Пусть $A = \{$выпало чётное число очков$\}$, а $A = \{$выпало число очков, кратное 3$\}$. Докажем, что они независимы. Заметим, что $A \cap B = \{$число очков кратно 6$\}$. Тогда $\mathsf{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathsf{P}(A)\,\mathsf{P}(B)$.

Данное определение работает только для двух событий. Можно ли как-то его обобщить? Попробуем ввести аналогично:

Определение 10. События $A_1, \ldots, A_n$ называются *попарно независимыми*, если для любых i и j таких, что $i \neq j$, A_i независимо от A_j.

Однако обобщение обычно вводят по другому:

Определение 11. События $A_1, a_2, \ldots, A_n$ называются *независимыми по совокупности*, если для любого множества $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$ выполняется, что $\mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod\limits_{j=1}^{k} \mathsf{P}(A_{i_j})$.

Примечание. Если говорят про независимость событий и не указывают тип, то обычно подразумевают независимость по совокупности.

Стоит заметить, что попарная независимость — гораздо более слабое условие, чем независимость по совокупности. Приведём пример. Пусть есть тетраэдр, грани которого покрашены следующим образом: первая грань покрашена в красный, вторая — в синий, третья — в зелёный, а четвёртая — во все три цвета сразу. Его подбрасывают. Введём три события: $A_{\text{К}} = \{$на нижней грани есть красный цвет$\}$, $A_{\text{С}} = \{$на нижней грани есть синий цвет$\}$ и $A_{\text{З}} = \{$на нижней грани есть зелёный цвет$\}$. Очевидно, что вероятность любого события — $\frac{1}{2}$, а любой пары событий — $\frac{1}{4}$. Однако вероятность объединения всех трёх событий равна $\frac{1}{4}$, а не $\frac{1}{8}$. Следовательно, события попарно независимы, но не независимы по совокупности.

Упражнение. Приведите пример n событий таких, что любой набор из $n-1$ события независим, а все n событий вместе зависимы.

Для независимости выполняются следующие свойства:

1. $A \perp\!\!\!\perp A \iff \mathsf{P}(A) = 0$ или $\mathsf{P}(A) = 1 \iff$ для любого события B $A \perp\!\!\!\perp B$.

2. $A \perp\!\!\!\perp B \implies \overline{A} \perp\!\!\!\perp B$.

3. Если $A_1, A_2, \ldots, A_n$ независимы в совокупности, то любой набор $B_1, B_2, \ldots, B_n$ такой, что $B_i = A_i$ или $B_i = \overline{A_i}$, тоже независим.

1.2.4 Случайные величины в дискретных вероятностных пространствах

Определение 12. Пусть (Ω, P) — дискретное вероятностное пространство. Тогда любое отображение $\xi : \Omega \to \mathbb{R}$ называется *случайной величиной*.

Примечание. В литературе случайную величину часто сокращают до с.в.

Рассмотрим некоторые примеры случайных величин:

1. Широко распространённым примером случайной величины служит индикатор какого-либо события. Введём определение.

   Определение 13. Пусть $A \in \Omega$ — событие. Тогда *индикатором* события A называют называется случайная величина, равная

$$I_A(\omega) = \begin{cases} 1, w \in A; \\ 0, w \notin A; \end{cases}$$

2. Бросок игральной кости. Тогда случайной величиной будет число выпавших очков.

3. Схема Бернулли: $\Omega = \{\omega = (\omega_1 \ldots \omega_n),\ \omega_i \in \{0, 1\}\}$. В данном случае вводится случайная величина, равная $\xi(\omega) = \sum\limits_{i=1}^{n} \omega_i$. Ещё её называют *числом успехов в схеме Бернулли*.

1.3 Лекция от 23.09.2016

1.3.1 Распределение случайной величины. Примеры распределений

Рассмотрим некоторое дискретное пространство (Ω, P). Тогда случайная величина ξ на этом пространстве принимает не более, чем счетное множество значений. Пусть $X = (x_1, x_2, \ldots, x_n, \ldots)$ — множество значений случайной величины ξ. Введём событие $A_i = \{\omega : \xi(\omega) = x_i\}$. Его можно интерпретировать, как событие $\{\xi$ приняло значение $x_i\}$. Для удобства будем использовать обозначение $A_i := \{\xi = x_i\}$. Также введем обозначение p_i для вероятности события A_i: $p_i = \mathsf{P}(A_i) = \mathsf{P}(\xi = x_i)$.

Определение 14. Вместе множество значений $X = (x_1, x_2, \ldots, x_n, \ldots)$ и набор вероятностей $(p_1, p_2, \ldots, p_n, \ldots)$ образуют то, что называется *распределением случайной величины* ξ.

Понятно, что каждому числу x_i сопоставлено число p_i. Легко заметить, что A_i образуют разбиение пространства Ω. Тогда $\sum\limits_{i=0}^{|X|} p_i = 1$.

Рассмотрим некоторые известные примеры распределений:

1. Распределение Бернулли.

   Определение 15. Случайная величина ξ имеет *распределение Бернулли*, если она принимает всего два значения, 1 или 0, с заранее известными вероятностями p и $q \equiv 1 - p$. Обозначение: $\xi \sim \mathrm{Bern}(p)$.

   Легко понять, что $X = \{0, 1\}$, $\mathsf{P}(\xi = 1) = p$, $\mathsf{P}(\xi = 0) = q$. Принято говорить, что событие $\{\xi = 1\}$ соответствует "успеху", а событие $\{\xi = 0\}$ — "неудаче". Эти названия достаточно условные, и в зависимости от конкретной задачи могут быть заменены на противоположные.

2. Биномиальное распределение.

   Определение 16. *Биномиальное распределение* в теории вероятностей — распределение количества "успехов" ξ в последовательности из n независимых случайных экспериментов, таких, что вероятность "успеха" в каждом из них постоянна и равна p. Обозначение: $\xi \sim \mathrm{Bin}(n, p)$.

   Пусть $x_1, \ldots, x_n$ — последовательность независимых случайных величин с одинаковым распределением Бернулли с параметром p. Тогда $\xi = \sum\limits_{i=1}^{n} x_i$. Посчитаем $\mathsf{P}(\xi = k)$. Для этого необходимо выбрать k исходов из n и сказать, что они успешны. Тогда $\mathsf{P}(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

3. Пуассоновское распределение.

   Определение 17. Говорят, что случайная величина ξ имеет *распределение Пуассона*, если она принимает любое значение[7] $k \in \mathbb{N}$ с вероятностью $\frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda > 0$ — некоторый параметр. Обозначение: $\xi \sim \mathrm{Pois}(\lambda)$.

   Почему оно вводится именно так? Об этом будет рассказано позднее.

4. Геометрическое распределение.

   Определение 18. Говорят, что случайная величина ξ имеет *геометрическое распределение*, если она принимает любое значение $k \in \mathbb{N}$ с вероятностью $p(1-p)^{k-1}$. Обозначение: $\xi \sim \mathrm{Geom}(p)$.

[7] Есть путаница, откуда считать — с нуля или же с единицы. Будем считать, что начинаем с единицы.

1.3.2 Независимость случайных величин

Далее множество значений случайной величины ξ на вероятностном пространстве (Ω, P) будем обозначать как $\xi(\Omega)$.

Определение 19. Пусть ξ, η — случайные величины на вероятностном пространстве (Ω, P). Будем говорить, что эти случайные величины *независимы*, если для любых $a \in \xi(\Omega)$, $b \in \eta(\Omega)$ события $\{\xi = a\}$ и $\{\eta = b\}$ независимы.

По опредлению независимых событий получаем, что

$$\mathsf{P}(\{\xi = a\} \cap \{\eta = b\}) = \mathsf{P}(\xi = a)\,\mathsf{P}(\eta = b).$$

Примечание. Для простоты левую часть часто обозначают за $\mathsf{P}(\xi = a, \eta = b)$.

Теперь обобщим это понятие на произвольное количество случайных величин:

Определение 20. Пусть $\xi_1, \ldots, \xi_n$ — случайные величины на вероятностном пространстве (Ω, P) и известно, что ξ_i принимает значения $(a_1^{(i)}, a_2^{(i)}, \ldots, a_n^{(i)})$. Тогда будем говорить, что случайные величины $\xi_1, \ldots, \xi_n$ *независимы (в совокупности)*, если $\forall j_1, \ldots, j_n$ выполнено:

$$\mathsf{P}(\xi_1 = a_{j_1}^{(1)}, \ldots, \xi_n = a_{j_n}^{(n)}) = \prod_{k=1}^{n} \mathsf{P}(\xi_k = a_{j_k}^{(k)}).$$

Упражнение. Пусть A, B — некоторые события над (Ω, P). Показать, что A и B независимы тогда и только тогда, когда их индикаторы I_A и I_B независимы.

Упражнение. Показать, что случайные величины $\xi_1, \ldots, \xi_n$ независимы тогда и только тогда, когда

$$\forall \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R} \quad \mathsf{P}(\xi_1 = \alpha_1, \ldots, \xi_n = \alpha_n) = \prod_{i=1}^{n} \mathsf{P}(\xi_i = \alpha_i)$$

1.3.3 Математическое ожидание. Свойства математического ожидания

Опять же, зафиксируем некоторое дискретное вероятностное пространство (Ω, P).

Определение 21. *Математическим ожиданием* случайной величины ξ называют величину $\mathsf{E}[\xi] = \sum\limits_{\omega \in \Omega} \xi(\omega)\,\mathsf{P}(\omega)$.

Примечание. Если Ω счётно, то ряд $\sum\limits_{\omega \in \Omega} \xi(\omega)\,\mathsf{P}(\omega)$ должен сходиться абсолютно; иначе сумма ряда не определена однозначно, так как порядок перебора ω не задан (см. теорему Римана о перестановке членов условно сходящегося ряда).

Смысл у математического ожидания случайной величины простой — его можно понимать как среднее значение этой случайной величины.

Рассмотрим некоторые свойства математического ожидания.

Теорема 3 (Простейшие свойства математического ожидания)**.** *Пусть ξ, η — некоторые случайные величины. Тогда выполняются следующие свойства:*

1. *Матожидание линейно:* $\forall a, b \in \mathbb{R}\ \ \mathsf{E}[a\xi + b\eta] = a\,\mathsf{E}[\xi] + b\,\mathsf{E}[\eta]$;

2. *Оно сохраняет относительный порядок: если* $\xi \leqslant \eta$ *(то есть для любого* $\omega \in \Omega$ $\xi(\omega) \leqslant \eta(\omega)$*), то* $\mathsf{E}[\xi] \leqslant \mathsf{E}[\eta]$*;*

3. *Модуль математического ожидания меньше математического ожидания модуля:* $|\mathsf{E}[\xi]| \leqslant \mathsf{E}\left[|\xi|\right]$*;*

4. $\mathsf{E}[\xi] = \sum\limits_{a\in\xi(\Omega)} a\,\mathsf{P}(\xi = a)$*;*

5. *Для любой функции* $\varphi(x)$ *выполняется* $\mathsf{E}[\varphi(\xi)] = \sum\limits_{a\in\xi(\Omega)} \varphi(a)\,\mathsf{P}(\xi = a)$*;*

6. *Если* $\mathsf{P}(\xi = c) = 1$ *для некоторой константы* c*, то* $\mathsf{E}[\xi] = c$*;*

7. *Если* $\xi \geqslant 0$*, то* $\mathsf{E}[\xi] \geqslant 0$*;*

8. *Если* $E[\xi] = 0$ *и* $\xi \geqslant 0$*, то* $\mathsf{P}(\xi = 0) = 1$*;*

9. *Если* ξ *и* η *независимы, то* $\mathsf{E}[\xi\eta] = \mathsf{E}[\xi]\,\mathsf{E}[\eta]$.

Доказательство. Докажем пункты по порядку:

1. $\mathsf{E}[a\xi + b\eta] = \sum\limits_{\omega\in\Omega} (a\xi + b\eta)(\omega)\,\mathsf{P}(\omega) = \sum\limits_{\omega\in\Omega} a\xi(\omega)\,\mathsf{P}(\omega) + \sum\limits_{\omega\in\Omega} b\eta(\omega)\,\mathsf{P}(\omega) = a\,\mathsf{E}[\xi] + b\,\mathsf{E}[\eta]$.

2. $\mathsf{E}[\xi] = \sum\limits_{\omega\in\Omega} \xi(\omega)\,\mathsf{P}(\omega) \leqslant \sum\limits_{\omega\in\Omega} \eta(\omega)\,\mathsf{P}(\omega) = \mathsf{E}[\eta]$.

3. $\mathsf{E}[-|\xi|] \leqslant \mathsf{E}[\xi] \leqslant \mathsf{E}[|\xi|] \implies -\mathsf{E}[|\xi|] \leqslant \mathsf{E}[\xi] \leqslant \mathsf{E}[|\xi|] \implies |\mathsf{E}[\xi]| \leqslant \mathsf{E}\left[|\xi|\right]$

4. $$\mathsf{E}[\xi] = \sum_{\omega\in\Omega} \xi(\omega)\,\mathsf{P}(\omega) = \sum_{a\in\xi(\Omega)} \sum_{\omega\,:\,\xi(\omega)=a} \xi(\omega)\,\mathsf{P}(\omega) = \sum_{a\in\xi(\Omega)} \left(a \sum_{\omega\,:\,\xi(\omega)=a} \mathsf{P}(\omega) \right) = \\ = \sum_{a\in\xi(\Omega)} a\,\mathsf{P}(\xi = a)$$

5. $$\mathsf{E}[\varphi(\xi)] = \sum_{\omega\in\Omega} \varphi(\xi(\omega))\,\mathsf{P}(\omega) = \sum_{a\in\xi(\Omega)} \sum_{\omega\,:\,\xi(\omega)=a} \varphi(\xi(\omega))\,\mathsf{P}(\omega) = \\ = \sum_{a\in\xi(\Omega)} \left(\varphi(a) \sum_{\omega\,:\,\xi(\omega)=a} \mathsf{P}(\omega) \right) = \sum_{a\in\xi(\Omega)} \varphi(a)\,\mathsf{P}(\xi = a)$$

6. $$\mathsf{E}[\xi] = \sum_{\omega\in\Omega} \xi(\omega)\,\mathsf{P}(\omega) = \sum_{\omega:\xi(\omega)=c} \xi(\omega)\,\mathsf{P}(\omega) + \sum_{\omega:\xi(\omega)\neq c} \xi(\omega)\,\mathsf{P}(\omega) = c\,\mathsf{P}(\xi = c) + \\ + \sum_{\omega:\xi(\omega)\neq c} \xi(\omega) \cdot 0 = c.$$

7. $\xi \geqslant 0 \implies \mathsf{E}[\xi] \geqslant \mathsf{E}[0] \geqslant 0$.

8. $\mathsf{E}[\xi] = \sum\limits_{\omega\in\Omega} \underbrace{\xi(\omega)}_{\geqslant 0} \underbrace{\mathsf{P}(\omega)}_{\geqslant 0} = 0 \implies \forall\omega \in \Omega\ \xi(\omega)\,\mathsf{P}(\omega) = 0$. Тогда если $\xi(\omega) \neq 0$, то $\mathsf{P}(\omega) = 0$. Тогда $\mathsf{P}(\xi = 0) = 1$.

   Заметим, что если величина может быть и отрицательной, то это свойство не выполняется. Предположим, что ξ равновероятно принимает значения 1 и -1. Тогда $\mathsf{E}[\xi] = 0$, но $\mathsf{P}(\xi = 0) = 0$.

9. $$\mathsf{E}[\xi\eta] = \sum_{\omega\in\Omega} \xi(\omega)\eta(\omega)\,\mathsf{P}(\omega) = \left\{\begin{matrix} \xi(\Omega) = (a_1, a_2, \ldots) \\ \eta(\Omega) = (b_1, b_2, \ldots) \end{matrix}\right\} = \sum_{i,j} \sum_{\substack{\omega: \\ \xi(\omega)=a_i \\ \eta(\omega)=b_j}} \xi(\omega)\eta(\omega)\,\mathsf{P}(\omega) = \\ = \sum_{i,j} a_i b_j\,\mathsf{P}(\xi = a_i, \eta = b_j)$$

   Так как $\xi \perp\!\!\!\perp \eta$, то $\mathsf{P}(\xi = a_i, \eta = b_j) = \mathsf{P}(\xi = a_i)\,\mathsf{P}(\eta = b_j)$ и сумма равна

   $$\sum_{i,j} a_i b_j\,\mathsf{P}(\xi = a_i)\,\mathsf{P}(\eta = b_j) = \left(\sum_i a_i\,\mathsf{P}(\xi = a_i)\right)\left(\sum_j b_j\,\mathsf{P}(\eta = b_j)\right) = \mathsf{E}[\xi]\,\mathsf{E}[\eta]$$

$\square$

Примеры использования математического ожидания:

1. Посчитаем математическое ожидание индикатора события A:
   $\mathsf{E}[I_A] = \sum\limits_{\omega\in\Omega} I_A(\omega)\,\mathsf{P}(\omega) = \sum\limits_{\omega\in A} \mathsf{P}(\omega) = \mathsf{P}(A)$.

2. Если мы рассмотрим классическую модель, т.е. такую модель, где все исходы равновероятны, то $\mathsf{E}[\xi] = \frac{1}{|\Omega|} \sum\limits_{\omega\in\Omega} \xi(\omega)$. Иначе говоря, математическое ожидание в классической модели равно среднему арифметическому возможных значений ξ.

3. Пусть $\xi \sim \mathrm{Bin}(n, p)$. Тогда

   $$\mathsf{E}[\xi] = \sum_{k=0}^{n} k\,\mathsf{P}(\xi = k) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \\ = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{t=0}^{n-1} \binom{n-1}{t} p^t (1-p)^{n-1-t}.$$

1.3.4 Дисперсия. Ковариация. Их свойства

Определение 22. Пусть ξ — некоторая случайная величина над (Ω, P). Тогда *дисперсией* ξ называется $\mathsf{D}[\xi] = \mathsf{E}[(\xi - \mathsf{E}[\xi])^2]$.

Дисперсию случайной величины можно понимать как среднеквадратическое отклонение этой случайной величины от её среднего значения (математического ожидания).

Примечание. Вероятность незначительного отклонения величины от математического ожидания, то есть $\mathsf{P}(\mathsf{E}[\xi] - 2\sqrt{\mathsf{D}[\xi]} \leqslant \xi \leqslant \mathsf{E}[\xi] + 2\sqrt{\mathsf{D}[\xi]})$, всегда велика! Это нам пока не пригодится, но на будущее стоит запомнить.

Определение 23. Пусть ξ и η — две случайные величины. Тогда *ковариацией* этих величин называется $\mathsf{cov}(\xi, \eta) = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])(\eta - \mathsf{E}[\eta])\right]$.

Ковариацию стоит воспринимать как меру зависимости двух случайных величин.

Определение 24. Две случайные величины ξ и η называют *некоррелированными*, если $\mathsf{cov}(\xi, \eta) = 0$.

Теорема 4 (Простейшие свойства дисперсии и ковариации)**.** *Пусть ξ,η и χ — некоторые случайные величины. Тогда выполняются следующие свойства:*

1. *Ковариация билинейна:* $\mathsf{cov}(\xi, a\eta + b\chi) = a\,\mathsf{cov}(\xi, \eta) + b\,\mathsf{cov}(\xi, \chi)$*;*
2. $\mathsf{D}[\xi] = \mathsf{cov}(\xi, \xi)$*;*
3. *Для любого $c \in \mathbb{R}$ верно, что* $\mathsf{D}[c\xi] = c^2\,\mathsf{D}[\xi]$ *и* $\mathsf{D}[\xi + c] = \mathsf{D}[\xi]$*;*
4. *Дисперсия неотрицательна:* $\mathsf{D}[\xi] \geqslant 0$*. При этом* $\mathsf{D}[\xi] = 0 \iff \mathsf{P}(\xi = E[\xi]) = 1$*;*
5. *Если ξ и η некоррелированные, то* $\mathsf{D}[\xi + \eta] = \mathsf{D}[\xi] + \mathsf{D}[\eta]$*;*
6. *Связь с матожиданием:* $\mathsf{D}[\xi] = \mathsf{E}[\xi^2] - (\mathsf{E}[\xi])^2$, $\mathsf{cov}(\xi, \eta) = \mathsf{E}[\xi\eta] - \mathsf{E}[\xi]\,\mathsf{E}[\eta]$

Доказательство. Докажем все пункты по порядку:

1. По определению ковариации $\mathsf{cov}(\xi, a\eta + b\chi) = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])((a\eta + b\chi) - \mathsf{E}[a\eta + b\chi])\right]$. Распишем вторую скобку:

   $$(a\eta + b\chi) - \mathsf{E}[a\eta + b\chi] = (a\eta + b\chi) - a\,\mathsf{E}[\eta] - b\,\mathsf{E}[\chi] = a(\eta - \mathsf{E}[\eta]) + b(\chi - \mathsf{E}[\chi]).$$

   Тогда

   $$\begin{aligned} \mathsf{cov}(\xi, a\eta + b\chi) &= \mathsf{E}\left[(\xi - \mathsf{E}[\xi])(a(\eta - \mathsf{E}[\eta]) + b(\chi - \mathsf{E}[\chi]))\right] \\ &= \mathsf{E}\left[a(\xi - \mathsf{E}[\xi])(\eta - \mathsf{E}[\eta]) + b(\xi - \mathsf{E}[\xi])(\chi - \mathsf{E}[\chi])\right] \\ &= a\,\mathsf{E}[(\xi - \mathsf{E}[\xi])(\eta - \mathsf{E}[\eta])] + b\,\mathsf{E}[(\xi - \mathsf{E}[\xi])(\eta - \mathsf{E}[\eta])] \\ &= a\,\mathsf{cov}(\xi, \eta) + b\,\mathsf{cov}(\xi, \chi). \end{aligned}$$

2. $\mathsf{cov}(\xi, \xi) = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])(\xi - \mathsf{E}[\xi])\right] = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])^2\right] = \mathsf{D}[\xi]$.
3. Для произведения $\mathsf{D}[c\xi] = \mathsf{cov}(c\xi, c\xi) = c^2\,\mathsf{cov}(\xi, \xi) = c^2\,\mathsf{D}[\xi]$. Для суммы $\mathsf{D}[\xi + c] = \mathsf{E}[((\xi + c) - \mathsf{E}[\xi + c])^2] = \mathsf{E}[(\xi + c - \mathsf{E}[\xi] - c)^2] = \mathsf{E}[(\xi - \mathsf{E}[\xi])^2] = \mathsf{D}[\xi]$.
4. По определению дисперсии, $D[\xi] = \mathsf{E}[(\xi - \mathsf{E}[\xi])^2]$, а $(\xi - \mathsf{E}[\xi])^2 \geqslant 0$. Тогда $\mathsf{E}[(\xi - \mathsf{E}[\xi])^2] \geqslant 0$.
5. Воспользуемся выражением дисперсии через ковариацию и билинейностью ковариации: $\mathsf{D}[\xi + \eta] = \mathsf{cov}(\xi + \eta, \xi + \eta) = \mathsf{cov}(\xi, \xi) + \mathsf{cov}(\eta, \eta) + 2\,\mathsf{cov}(\xi, \eta)$. Так как ξ и η некоррелированные, то $\mathsf{cov}(\xi, \eta) = 0$ и $\mathsf{D}[\xi + \eta] = \mathsf{cov}(\xi, \xi) + \mathsf{cov}(\eta, \eta) = \mathsf{D}[\xi] + \mathsf{D}[\eta]$.
6. $\mathsf{cov}(\xi, \eta) = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])(\eta - \mathsf{E}[\eta])\right] = \mathsf{E}[\xi\eta - \xi\,\mathsf{E}[\eta] - \eta\,\mathsf{E}[\xi] + \mathsf{E}[\xi]\,\mathsf{E}[\eta]]$. По линейности матожидания $\mathsf{cov}(\xi, \eta) = \mathsf{E}[\xi\eta] - 2\,\mathsf{E}[\xi]\,\mathsf{E}[\eta] + \mathsf{E}[\xi]\,\mathsf{E}[\eta] = \mathsf{E}[\xi\eta] - \mathsf{E}[\xi]\,\mathsf{E}[\eta]$. Для дисперсии воспользуемся тем, что $\mathsf{D}[\xi] = \mathsf{cov}(\xi, \xi)$.

$\square$

Теорема 5. *Если случайные величины независимы, то они* некоррелированные. *Обратное, вообще говоря, неверно.*

Доказательство. Докажем, что из независимости следует некоррелируемость. Пусть ξ и η — независимые случайные величины. Тогда $\mathsf{cov}(\xi, \eta) = \mathsf{E}[\xi\eta] - \mathsf{E}[\xi]\,\mathsf{E}[\eta]$. Но по свойству матожидания из независимости случайных величин следует, что $\mathsf{E}[\xi\eta] = \mathsf{E}[\xi]\,\mathsf{E}[\eta]$. Тогда $\mathsf{cov}(\xi, \eta) = 0$ и эти величины некоррелированные.

Теперь покажем, что из некоррелированности не обязательно следует независимость. Пусть случайная величина ξ равновероятно принимает значения из множества $\{0, 1, -1\}$. Возьмем случайную величину $\eta = \xi^2$. По определению можно проверить, что величины η и ξ *некоррелированные*: $\mathsf{cov}(\xi, \eta) = \mathsf{E}[\xi\eta] - \mathsf{E}[\xi]\,\mathsf{E}[\eta] = \mathsf{E}[\xi^3] - \mathsf{E}[\xi]\,\mathsf{E}[\xi^2] = 0 - 0 = 0$. Но ξ и η *не являются независимыми*, что проверяется опять же по определению: $\mathsf{P}(\xi = 0, \eta = 0) = \frac{1}{3} \neq \frac{1}{9} = \mathsf{P}(\xi = 0)\,\mathsf{P}(\eta = 0)$. □

1.3.5 Неравенства Маркова и Чебышёва

Под конец лекции обсудим два неравенства, которые сами по себе являются весьма полезными.

Теорема 6 (Неравенство Маркова). *Пусть $\xi \geqslant 0$ — неотрицательная случайная величина на (Ω, P). Тогда для любого положительного a*

$$\mathsf{P}(\xi \geqslant a) \leqslant \frac{\mathsf{E}[\xi]}{a}.$$

Доказательство. Как было доказано ранее, матожидание индикатора события равно вероятности события. Тогда $\mathsf{P}(\xi \geqslant a) = \mathsf{E}[I_{\xi \geqslant a}]$. Далее, заметим, что $I_{\xi \geqslant a} \leqslant \frac{\xi}{a} I_{\xi \geqslant a}$. Тогда

$$\mathsf{P}(\xi \geqslant a) \leqslant \mathsf{E}\left[\frac{\xi}{a} I_{\xi \geqslant a}\right] \leqslant \mathsf{E}\left[\frac{\xi}{a}\right] = \frac{\mathsf{E}[\xi]}{a}.$$ □

Теорема 7 (Неравенство Чебышёва). *Пусть ξ — случайная величина на (Ω, P) такая, что $\mathsf{D}[\xi] < \infty$. Тогда для любого положительного a*

$$\mathsf{P}\left(|\xi - E[\xi]| \geqslant a\right) \leqslant \frac{\mathsf{D}[\xi]}{a^2}.$$

Доказательство. $\mathsf{P}(|\xi - \mathsf{E}[\xi]| \geqslant a) = \mathsf{P}\left((\xi - E[\xi])^2 \geqslant a^2\right)$. По неравенству Маркова

$$\mathsf{P}\left((\xi - E[\xi])^2 \geqslant a^2\right) \leqslant \frac{\mathsf{E}[(\xi - E[\xi])^2]}{a^2} = \frac{\mathsf{D}[\xi]}{a^2}.$$ □

1.4 Лекция от 30.09.2016

1.4.1 Сходимость случайных величин. Закон больших чисел.

Сначала введём пару определений:

Определение 25. Пусть $\{\xi_n\}_{n=1}^{\infty}$ — последовательность случайных величин, заданных на вероятностном пространстве (Ω, P). Тогда будем говорить, что последовательность *сходится по вероятности* к случайной величине ξ, если для любого положительного α выполнено, что

$$\lim_{n\to\infty} \mathsf{P}(|\xi_n - \xi| \geqslant \alpha) = \lim_{n\to\infty} \mathsf{P}(\{\omega : |\xi_n(\omega) - \xi(\omega)| \geqslant \alpha\}) = 0.$$

Обозначение: $\xi_n \xrightarrow{\mathsf{P}} \xi$.

Определение 26. Случайные величины ξ и η называют *одинаково распределёнными*, если

$$\begin{cases} \xi(\Omega) = \eta(\Omega); \\ \forall x \in \xi(\Omega), \mathsf{P}(\xi = x) = \mathsf{P}(\eta = x). \end{cases}$$

Говоря простым языком, у этих случайных величин совпадают возможные значения и их вероятности.

Обозначение: $\xi \overset{d}{=} \eta$.

Теорема 8 (Закон больших чисел)**.** *Пусть есть последовательность попарно независимых и одинаково распределённых случайных величин $\{\xi_n\}_{n=1}^{\infty}$ и существует число C такое, что $\mathsf{D}[\xi_n] \leqslant C$ для всех $n \in \mathbb{N}$.*

Введём обозначение $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. Тогда

$$\frac{S_n - \mathsf{E}[S_n]}{n} \xrightarrow{\mathsf{P}} 0.$$

Доказательство. Для начала заметим, что $\mathsf{E}\left[\frac{S_n - \mathsf{E}[S_n]}{n}\right] = 0$. Тогда по неравенству Чебышёва выполнено, что:

$$\mathsf{P}\left(\left|\frac{S_n - \mathsf{E}[S_n]}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\left[\frac{S_n - \mathsf{E}[S_n]}{n}\right]}{\varepsilon^2}.$$

Так как $\mathsf{D}\left[\frac{S_n - \mathsf{E}[S_n]}{n}\right] = \frac{1}{n^2}\mathsf{D}[S_n]$ для фиксированного n, то

$$\mathsf{P}\left(\left|\frac{S_n - \mathsf{E}[S_n]}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}[S_n]}{n^2\varepsilon^2}.$$

Распишем дисперсию в виде суммы ковариаций:

$$\mathsf{D}[S_n] = \mathsf{cov}(\xi_1 + \ldots + \xi_n, \xi_1 + \ldots + \xi_n) = \sum_{i,j=1}^{n} \mathsf{cov}(\xi_i, \xi_j).$$

Если $i \neq j$, то случайные величины ξ_i и ξ_j независимы. Но тогда они некоррелированы и $\mathsf{cov}(\xi_i, \xi_j) = 0$. Следовательно,

$$\mathsf{D}[S_n] = \sum_{i=1}^{n} \mathsf{cov}(\xi_i, \xi_i) \leqslant Cn.$$

Подставляя выражение дисперсии, получаем, что

$$\mathsf{P}\left(\left|\frac{S_n - \mathsf{E}[S_n]}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0. \qquad \square$$

Примечание. Вообще, в законе больших чисел не обязательна независимость случайных величин — достаточно попарной некоррелированности.

Следствие. *Пусть* $\{\xi_n\}_{n=1}^{\infty}$ *— последовательность попарно независимых случайных величин. Пусть существуют* α *и* C *такие, что* $\mathsf{E}[\xi_n] = \alpha$ *и* $\mathsf{D}[\xi_n] \leqslant C$ *для всех* $n \in \mathbb{N}$. *Тогда*

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\mathsf{P}} \alpha.$$

Доказательство. Из условия независимости случайных величин нам хватит следствия о том, что они являются попарно некоррелированными. Из закона больших чисел следует, что $\frac{S_n - \mathsf{E}[S_n]}{n} \xrightarrow{\mathsf{P}} 0$. Так как $\mathsf{E}[S_n] = n\alpha$, то $\frac{S_n}{n} - \alpha \xrightarrow{\mathsf{P}} 0$. Также заметим, что в силу определения сходимости, $\frac{S_n}{n} \xrightarrow{\mathsf{P}} \alpha$. □

Давайте попробуем осознать, в чем заключается смысл Закона больших чисел. Пусть случайная величина задается индикатором происхождения события, $\xi_i = I\{$событие A произошло в i-ом эксперименте$\}$, при этом все эксперименты проводятся независимо друг от друга (ξ_n является независимой случайной величиной). Обозначим за $\vartheta(A)$ частоту появления события A в n-м количестве экспериментов, $\vartheta_n(A) = \frac{\xi_1 + \ldots + \xi_n}{n}$. Тогда по Закону больших чисел $\vartheta_n(A) \xrightarrow{\mathsf{P}} \mathsf{E}[\xi_i] = \mathsf{E}[I\{$событие A произошло в i-ом эксперименте$\}] = \mathsf{P}(A)$. Выходит, что Закон больших чисел является теоретическим обоснованием принципа устойчивости частот.

Определение 27. Последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ сходится к случайной величине ξ с вероятностью 1 (или *почти наверное*[8]), если

$$\mathsf{P}\Big(\lim_{n\to\infty} \xi_n = \xi\Big) = \mathsf{P}\Big(\Big\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\Big\}\Big) = 1.$$

Обозначение: $\xi_n \xrightarrow{\text{п.н.}} \xi$ (сходится *почти наверное*).

Примечание. Данное определение можно записать следующим образом: для любого $\omega \in \Omega$ такого, что $\mathsf{P}(\omega) > 0$, выполнено $\xi_n(\omega) \to \xi(\omega)$.

Теорема 9. *В дискретных вероятностных пространствах* $\xi_n \xrightarrow{\mathsf{P}} \xi \iff \xi_n \xrightarrow{\text{п.н.}} \xi$.

Доказательство. Докажем по очереди в обе стороны:

[⇒] Допустим, что $\xi_n \xrightarrow{\mathsf{P}} \xi$. Докажем, что тогда для любого $\omega \in \Omega$, для которого $\mathsf{P}(\omega) > 0$, выполнено $\xi_n(\omega) \to \xi(\omega)$ (стремится в понимании предела последовательности).

Пусть есть элементарный исход ω_0 такой, что сходимости нет: $\xi_n(\omega_0) \nrightarrow \xi(\omega_0)$. В таком случае существует положительное α и подпоследовательность $\{\xi_{n_k}(\omega_0)\}_{k=1}^{\infty}$ такие, что $\forall k \in \mathbb{N}\, |\xi_{n_k}(\omega_0) - \xi(\omega_0)| \geqslant \alpha$. Но тогда $\mathsf{P}(|\xi_{n_k} - \xi| \geqslant \alpha) \geqslant \mathsf{P}(\omega_0) > 0$. Но по определению сходимости по вероятности $\mathsf{P}(|\xi_{n_k} - \xi| \geqslant \alpha) \to 0$. Получаем противоречие. Следовательно, $\xi_n \xrightarrow{\text{п.н.}} \xi$.

[⇐] Пусть $\xi_n \xrightarrow{\text{п.н.}} \xi$. Согласно определению, $\mathsf{P}(\lim\limits_{n\to\infty} \xi_n = \xi) = 1$. Это равносильно тому, что для любого $\omega \in \Omega$, для которого $\mathsf{P}(\omega) > 0$, выполнено $\xi_n(\omega) \to \xi(\omega)$.

[8]Название “сходимость почти наверное” произошло от применяющегося в англоязычной литературе термина “almost sure convergence”.

Докажем, что выполнено условие сходимости *по вероятности*: $\mathsf{P}(|\xi_n - \xi| \geqslant \alpha) \to 0$ для любого положительного α. Пусть $\Omega = \{\omega_i \mid i \in \mathbb{N}\}$. Зафиксируем произвольное $\delta > 0$ и выберем номер $N = N(\delta)$ такой, что $\sum\limits_{i>N} \mathsf{P}(\omega_i) \leqslant \delta$. Тогда

$$\begin{aligned}
\mathsf{P}(|\xi_n - \xi| \geqslant \alpha) &= \sum_{i\,:\,|\xi_n(\omega_i)-\xi(\omega_i)|\geqslant\alpha} \mathsf{P}(\omega_i) \\
&= \sum_{\substack{i\,:\,i\leqslant N \\ |\xi_n(\omega_i)-\xi(\omega_i)|\geqslant\alpha}} \mathsf{P}(\omega_i) + \sum_{\substack{i\,:\,i>N \\ |\xi_n(\omega_i)-\xi(\omega_i)|\geqslant\alpha}} \mathsf{P}(\omega_i) \\
&\leqslant \sum_{\substack{i\,:\,i\leqslant N \\ |\xi_n(\omega_i)-\xi(\omega_i)|\geqslant\alpha}} \mathsf{P}(\omega_i) + \sum_{i>N} \mathsf{P}(\omega_i) \\
&\leqslant \sum_{\substack{i\,:\,i\leqslant N \\ |\xi_n(\omega_i)-\xi(\omega_i)|\geqslant\alpha}} \mathsf{P}(\omega_i) + \delta.
\end{aligned}$$

Теперь нужно перейти к пределу при $n \to \infty$. Но чтобы избежать ситуации, когда предела не существует, перейдем к верхнему пределу для тех же n. Также заметим, что так как $\xi_n(\omega_i) \to \xi(\omega_i)$, начиная с некоторого n, ω_i перестает входить в сумму. Тогда первая сумма стремится к нулю и

$$\overline{\lim_{n\to\infty}} \mathsf{P}(|\xi_n - \xi| \geqslant \alpha) \leqslant 0 + \delta.$$

В силу произвольности $\delta > 0$ можно говорить о том, что этот предел стремится к нулю. □

Допустим, что $S_n \sim \mathrm{Bin}(n, p)$. Тогда сформулируем два вопроса:
Вопрос 1: Как оценить скорость сходимости $\frac{S_n}{n} - a$ к нулю?
Вопрос 2: Как оценить скорость сходимости $\mathsf{P}\big(\big|\frac{S_n}{n} - a\big| \geqslant \varepsilon\big)$ к нулю?

1.4.2 Схема испытаний Бернулли

Описание: Схема испытаний Бернулли подразумевает много однородных экспериментов, в каждом из которых возможно два исхода — “успех” и “неудача”; при этом нас интересует распределение числа успехов.

Математическая модель: Пусть $\{\xi_i\}$ — независимые *одинаково распределённые* случайные величины (например, последовательные броски монеты); $\xi_i(\Omega) = \{0; 1\}$; пусть также $\mathsf{P}(\xi_i = 1) = p$ — вероятность успеха каждого отдельного эксперимента. Введём также обозначение $S_n = \sum\limits_{i=1}^{n} \xi_i$ — случайную величину, обозначающую количество успехов.

Распределение S_n кажется очевидным: $S_n \sim \mathrm{Bin}(n, p)$, т.е. $\mathsf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k}$.

Упражнение. Докажите, что распределение S_n и вправду биномиально.

Упражнение. В качестве упражнения, читателю предлагается оценить, сколько нужно попыток, чтобы $\frac{\text{число успехов}}{\text{число попыток}}$ было примерно равно p, т.е. как быстро наблюдаемая вероятность сходится к реальной. Для этого предлагается оценить следующую вероятность:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) =?$$

Кстати говоря, p совершенно не обязана быть константой и может являться функцией от n: $p(n)$.

Тогда, вообще говоря, есть несколько отдельных случаев:

- $np(n) \xrightarrow[n\to\infty]{} \lambda > 0$.

  Теорема 10 (Пуассон). *Пусть* $np(n) \xrightarrow[n\to\infty]{} \lambda > 0$. *Тогда*

$$\forall k \in \mathbb{Z}_+, \mathsf{P}(S_n = k) \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!}.$$

  Доказательство.

$$C_n^k p^k (1-p)^{n-k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p(n)^k (1-p(n))^{n-k} \sim$$
$$\sim \frac{n^k}{k!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} =$$
$$= \frac{1}{k!}\left(\lambda + o\left(1\right)\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

  □

  Следствие: $P(S_n = k) \xrightarrow{d} P(\xi = k)$, $\xi \sim \mathrm{Pois}(\lambda)$ — сходимость *по распределению.*

- $np(n) \xrightarrow[n\to\infty]{} \infty$.

  Сделаем ещё одно отступлление в сторону и рассмотрим величину $\frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}}$. Пусть $\xi_1, \dots, \xi_n$ — n бернуллиевских случайных величин с $\mathsf{P}(\xi_i = 1) = p$, $\mathsf{P}(\xi_i = 0) = q$, $p + q = 1$. Тогда для $S_n = \xi_1 + \ldots + \xi_n$ находим, что

$$\mathsf{E}[S_n] = np.$$

  Несложно также показать, что

$$\mathsf{D}[S_n] = \sum \mathsf{D}[\xi_i] + \sum \mathsf{cov}[x_i, x_j] = \sum \mathsf{D}[\xi_i] = n\,\mathsf{D}[\xi_1] =$$
$$= n\left(\mathsf{E}[\xi_1^2] - (\mathsf{E}[\xi_1])^2\right) = n(p - p^2) = npq.$$

  Теорема 11 (Муавра-Лапласа). *Пусть вероятность p того, что во время эксперимента произойдёт событие, постоянна и находится в интервале* $(0, 1)$. *Допустим, что было проведено n независимых испытаний. Обозначим за S_n количество испытаний, в которых произошло событие (это равносильно тому, что $S_n \sim \mathrm{Bin}(n, p)$). Введём следующее обозначение:*

$$\mathsf{P}_n(a,b) := \mathsf{P}\left(a < \frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}} \leqslant b\right).$$

  Тогда

$$\sup_{a<b}\left|\mathsf{P}_n(a,b) - \int\limits_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right| \xrightarrow[n\to\infty]{} 0.$$

1.5 Лекция от 07.10.2016

1.5.1 Скорость сходимости закона больших чисел

Ранее мы формулировали два вопроса, связанные со скоростью сходимости закона больших чисел. Ответим на них.

1. В первом пункте от нас требуется найти скорость сходимости $\frac{S_n}{n} - p$ к нулю, где $S_n \sim \mathrm{Bin}(n, p)$. Для этого вспомним теорему Муавра-Лапласа:

$$\mathsf{P}\left(a < \frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}} \leqslant b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x = c, c \in (0,1).$$

Преобразуем условие вероятности, пользуясь тем, что $\mathsf{E}[S_n] = np$, а $\mathsf{D}[S_n] = np(1-p)$:

$$a < \frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}} \leqslant b \implies \frac{a\sqrt{p(1-p)}}{\sqrt{n}} < \frac{S_n}{n} - p \leqslant \frac{b\sqrt{p(1-p)}}{\sqrt{n}}.$$

Возьмём $\delta \in (0,1)$ такое, что $1 - \delta < c$. Тогда при достаточно больших n выполнено (по определению предела):

$$\mathsf{P}\left(\frac{c_1}{\sqrt{n}} < \frac{S_n}{n} - p \leqslant \frac{c_2}{\sqrt{n}}\right) > 1 - \delta.$$

Увеличивая расстояние между a и b и увеличивая n, получаем следующий результат:

$$\lim_{n, |a-b| \to \infty} \mathsf{P}\left(\frac{S_n}{n} - p = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right) = 1.$$

Отсюда получаем, что $\frac{S_n}{n} = p + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

2. Теперь приступим ко второму вопросу. Для него достаточно применить неравенство Чебышёва:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\left[\frac{S_n}{n}\right]}{\varepsilon^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \mathcal{O}\left(\frac{1}{n}\right).$$

1.5.2 Неравенство Чернова

Следующая теорема показывает то, насколько вероятно отклонение случайной величины от её математического ожидания.

Теорема 12 (Неравенство Чернова)**.** *Пусть $S_n \sim \mathrm{Bin}(n, p)$, а $\lambda = \mathsf{E}[S_n] = np$. Тогда для любого $t > 0$ выполнено следующее:*

$$\mathsf{P}(S_n \geqslant \lambda + t) \leqslant \exp\left(-\frac{t^2}{2(\lambda + \frac{t}{3})}\right), \qquad \mathsf{P}(S_n \leqslant \lambda - t) \leqslant \exp\left(-\frac{t^2}{2\lambda}\right).$$

Доказательство. Крайний случай для первого неравенства — $t = n - \lambda$ — особого интереса не вызывает. Поэтому положим, что $t < n - \lambda$.

Заметим, что для любого положительного u выполнено следующее:

$$\mathsf{P}(S_n \geqslant \lambda + t) = \mathsf{P}\big(e^{uS_n} \geqslant e^{u(\lambda+t)}\big) \leqslant \{\text{по неравенству Маркова}\} \leqslant \frac{\mathsf{E}\big[e^{uS_n}\big]}{e^{u(\lambda+t)}}.$$

Посчитаем $\mathsf{E}\big[e^{uS_n}\big]$:

$$\mathsf{E}\big[e^{uS_n}\big] = \sum_{k=1}^{n} e^{uk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n}\binom{n}{k}(pe^u)^k(1-p)^{n-k} = (1-p+pe^u)^n.$$

Положим $x = e^u$. Тогда

$$\mathsf{P}(S_n \geqslant \lambda + t) \leqslant x^{-(\lambda+t)}(1-p+px)^n.$$

Минимизируем это выражение по x. Для этого посчитаем производную и приравняем её к нулю:

$$(x^{-(\lambda+t)}(1-p+px)^n)' = -(\lambda+t)x^{-(\lambda+t+1)}(1-p+px)^n + npx^{-(\lambda+t)}(1-p+px)^{n-1} = 0.$$

Отсюда получаем, что $-(\lambda+t)(1-p+px) + npx = 0$ и $x = \frac{(\lambda+t)(1-p)}{p(n-\lambda-t)}$. Тогда получаем следующую оценку сверху:

$$\begin{aligned}
\mathsf{P}(S_n \geqslant \lambda + t) &\leqslant \left(\frac{p(n-\lambda-t)}{(\lambda+t)(1-p)}\right)^{\lambda+t}\left(1-p+p\frac{(\lambda+t)(1-p)}{p(n-\lambda-t)}\right)^n \\
&= \left(\frac{p(n-\lambda-t)}{(\lambda+t)(1-p)}\right)^{\lambda+t}\left(\frac{n(1-p)}{n-\lambda-t}\right)^n \\
&= \left(\frac{np}{\lambda+t}\right)^{\lambda+t}\left(\frac{n-np}{n-\lambda-t}\right)^{n-\lambda-t} \\
&= \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t}\left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t}
\end{aligned}$$

Представим эту оценку в следующем виде:

$$\left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t}\left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} = \exp\left\{-(\lambda+t)\ln\left(1+\frac{t}{\lambda}\right) - (n-\lambda-t)\ln\left(1-\frac{t}{n-\lambda}\right)\right\}.$$

Введём функцию $\varphi(x) = (1+x)\ln(1+x) - x$ для $x > -1$. Тогда (проверьте):

$$\mathsf{P}(S_n \geqslant \lambda + t) \leqslant \exp\left\{-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(-\frac{t}{n-\lambda}\right)\right\}.$$

Теперь заметим, что если $x > -1$, то $\varphi(x) > 0$. Тогда будет верно следующее (вторую оценку можно получить абсолютно аналогично первой, только взяв $-u$ вместо u):

$$\begin{aligned}
\mathsf{P}(S_n \geqslant \lambda + t) &\leqslant \exp\left\{-\lambda\varphi\left(\frac{t}{\lambda}\right)\right\}, \\
\mathsf{P}(S_n \leqslant \lambda - t) &\leqslant \exp\left\{-\lambda\varphi\left(\frac{-t}{\lambda}\right)\right\}.
\end{aligned}$$

А теперь достаём бубен и начинаем оптимизировать:

[$\leqslant$] Заметим, что $\varphi(0) = 0$ и $\varphi'(x) = 1 - 1 + \ln(x+1) = \ln(x+1) \leqslant x$. Тогда

$$-\varphi(y) = \int\limits_y^0 \varphi'(x)\,\mathrm{d}x \leqslant \int\limits_y^0 x\,\mathrm{d}x = -\frac{y^2}{2}.$$

Отсюда получаем, что $\varphi(y) \geqslant \frac{y^2}{2}$ и $\mathsf{P}(S_n \leqslant \lambda - t) \leqslant \exp\left\{-\frac{t^2}{2\lambda}\right\}$.

[$\geqslant$] Дальше идёт полное шаманство (sic!). Заметим, что

$$\varphi''(x) = \frac{1}{1+x} \geqslant \frac{1}{\left(1+\frac{x}{3}\right)^3} = \left(\frac{x^2}{2\left(1+\frac{x}{3}\right)}\right)'' \implies \varphi(x) \geqslant \frac{x^2}{2\left(1+\frac{x}{3}\right)}.$$

Отсюда получаем, что $\mathsf{P}(S_n \geqslant \lambda + t) \leqslant \exp\left\{-\frac{t^2}{2\left(\lambda+\frac{t}{3}\right)}\right\}$.

$\square$

К счастью, на экзамене будет достаточно понимать идею, и не требуется точное воспроизведение всех вычислений.

Теперь, используя неравенство Чернова, оценим скорость сходимости $\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right)$ к нулю по-другому:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) = \mathsf{P}(|S_n - \lambda| \geqslant n\varepsilon) \leqslant \exp\left\{-\frac{n\varepsilon^2}{2p}\right\}.$$

Заметим, что оценка через неравенство Чебышёва давало гораздо оценку $\frac{1}{\text{полином}}$, а неравенство Чернова даёт оценку $\frac{1}{\text{экспонента}}$, что сходится гораздо быстрее.

1.5.3 Алгебры событий

Определение 28. Пусть $\mathcal{A}$ — система событий на (Ω, P). Она называется *алгеброй*, если

1. $\Omega \in \mathcal{A}$,
2. Если $A \in \mathcal{A}$, то и $\overline{A} \in \mathcal{A}$,
3. Если $A \in \mathcal{A}$ и $B \in \mathcal{A}$, то и $A \cap B \in \mathcal{A}$.

Упражнение. Докажите, что алгебра замкнута по основным операциям: $\cup$, $\setminus$, $\triangle$.

Примеры:

- $\{\varnothing, \Omega\}$;
- 2^Ω;
- $\left\{\varnothing, \Omega, A, \overline{A}\right\}$ — алгебра, порождённая A;
- $\alpha(A_1, \ldots, A_n)$ — минимальная алгебра, содержащая $A_1, \ldots, A_n$.

1.6 Лекция от 14.10.2016

1.6.1 Локальная лемма Ловаса: несимметричный и симметричный случаи

Предупреждение: в данном случае мы решили несколько изменить доказательство, отойдя от сигма-алгебр к графам зависимости. Идейно эти принципы ничем не отличаются, но на коллоквиуме лучше рассказывать через алгебры событий.

Введём понятие *графа зависимости*.

Определение 29. *Графом зависимости* $D = (V, E)$ для событий $A_1, A_2, \dots, A_n$ называют ориентированный граф такой, что $V = \{1, 2, \dots, n\}$, а для любого $1 \leqslant i \leqslant n$ событие A_i независимо от событий A_j, а также от всевозможных объединений и пересечений данных событий, если $(i, j) \notin E$.

Теорема 13 (Несимметричный случай локальной леммы Ловаса)**.** *Пусть $A_1, \dots, A_n$ — события на произвольном вероятностном пространстве (Ω, P). Предположим, что*

1. *$D = (V, E)$ является орграфом зависимости для определенных выше событий;*
2. *Существуют действительные числа $x_1, \dots x_n \in [0, 1)$ такие, что для любого $1 \leqslant i \leqslant n$ выполняется*
$$\mathsf{P}(A_i) \leqslant x_i \prod_{(i,j)\in E} (1 - x_j).$$

Тогда вероятность того, что ни одно из событий $A_1, \dots, A_n$ не произойдёт, можно ограничить снизу положительным числом:

$$\mathsf{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geqslant \prod_{i=1}^{n}(1 - x_i) > 0.$$

Перед тем, как приступать к доказательству этой леммы, докажем следующую формулу:

Теорема 14 (Формула умножения вероятностей)**.** *Пусть $B_1, B_2, \dots, B_n$ — события на вероятностном пространстве (Ω, P). Тогда верно следующее:*

$$\mathsf{P}(B_1 \cap \dots \cap B_m) = \mathsf{P}(B_1 \mid B_2 \cap \dots \cap B_m)\, \mathsf{P}(B_2 \mid B_3 \cap \dots \cap B_m) \cdots \mathsf{P}(B_{m-1} \mid B_m)\, \mathsf{P}(B_m).$$

Доказательство. Будем последовательно применять определение условной вероятности:

$$\begin{aligned}
\mathsf{P}(B_1 \cap \dots \cap B_m) &= \mathsf{P}(B_1 \mid B_2 \cap \dots \cap B_m)\, \mathsf{P}(B_2 \cap \dots \cap B_m)\\
\mathsf{P}(B_2 \cap \dots \cap B_m) &= \mathsf{P}(B_2 \mid B_3 \cap \dots \cap B_m)\, \mathsf{P}(B_3 \cap \dots \cap B_m)\\
&\dots\\
\mathsf{P}(B_{m-1} \cap B_m) &= \mathsf{P}(B_{m-1} \mid B_m)\, \mathsf{P}(B_m).
\end{aligned}$$

Постепенными подстановками получим желаемое. □

Теперь перейдём к доказательству леммы.

Доказательство. Разобьем доказательство на подпункты.

1. Докажем индукцией по t, что $\forall i = 1, \ldots, n$, для любого набора чисел $T \in \{1, \ldots, n\} \setminus \{i\}$, $t = |T| < n$, справедливо неравенство $\mathsf{P}(A_i \mid \bigcap\limits_{j \in T} \overline{A_j}) \leqslant x_i$.

   Пусть $t = 1$. Тогда неравенство примет следующий вид:

$$\mathsf{P}(A_i \mid \overline{A_j}) = \begin{cases} \mathsf{P}(A_i), & (i,j) \notin E \\ \frac{\mathsf{P}(A_i \cap \overline{A_j})}{\mathsf{P}(\overline{A_j})} \leqslant \frac{\mathsf{P}(A_i)}{\mathsf{P}(\overline{A_j})} \leqslant \frac{x_i(1-x_j)}{1-x_j}, & (i,j) \in E \end{cases}$$

   Так как по условию $\mathsf{P}(A_i) \leqslant x_i \prod\limits_{(i,j) \in E} (1 - x_j) \leqslant x_i(1 - x_j)$, вероятность $\mathsf{P}(\overline{A_j}) = 1 - \mathsf{P}(A_j) \geqslant 1 - x_j$.

   Также, если $T = \varnothing$, из условия следует, что $\mathsf{P}(A_i) \leqslant x_i$.

   Предполагая, что неравенство выполняется для всех $t' < t$, докажем для его для t. Положим $T_1 = \{j \in T, (i,j) \in E\}, T_2 = T \setminus T_1$. Тогда:

$$\mathsf{P}\left(A_i \mid \bigcap_{j \in T} \overline{A_j}\right) = \frac{\mathsf{P}\left(A_i \cap (\bigcap\limits_{j \in T_1} \overline{A_j}) \mid \bigcap\limits_{k \in T_2} \overline{A_k}\right)}{\mathsf{P}\left(\bigcap\limits_{j \in T_1} \overline{A_j} \mid \bigcap\limits_{k \in T_2} \overline{A_k}\right)} \leqslant \frac{\mathsf{P}\left(A_i \mid \bigcap\limits_{k \in T_2} \overline{A_k}\right)}{\mathsf{P}\left(\bigcap\limits_{j \in T_1} \overline{A_j} \mid \bigcap\limits_{k \in T_2} \overline{A_k}\right)}.$$

   Для оценки числителя заметим, что событие A_i взаимно независимо с событиями $\{A_k : k \in T_2\}$. Следовательно, $\mathsf{P}\left(A_i \mid \bigcap\limits_{k \in T_2} \overline{A_k}\right) = \mathsf{P}(A_i) \leqslant x_i \prod\limits_{(i,j) \in E} (1 - x_j)$.

   Предположим $T_1 = \{j_1, \ldots, j_r\}$. При $r = 0$ неравенство $\mathsf{P}\left(A_i \mid \bigcap\limits_{j \in T} \overline{A_j}\right) \leqslant x_i$. заведомо выполняется, так как в нашем графе нет ребер, связывающих вершину i с другими вершинами, а значит, событие A_i не зависит ни от одного события A_j, где $j \in T$.

   Для оценки знаменателя применим предположение индукции. В случаях, когда $r > 0$, используя формулу умножения вероятностей, получаем

$$\mathsf{P}\left(\bigcap_{j \in T_1} \overline{A_j} \mid \bigcap_{k \in T_2} \overline{A_k}\right) = \mathsf{P}\left(\overline{A_{j_1}} \mid \bigcap_{s=2}^{r} \overline{A_{j_s}} \cap \bigcap_{k \in T_2} \overline{A_k}\right) \mathsf{P}\left(\overline{A_{j_2}} \mid \bigcap_{s=3}^{r} \overline{A_{j_s}} \cap \bigcap_{k \in T_2} \overline{A_k}\right) \ldots$$
$$\ldots \mathsf{P}\left(\overline{A_{j_r}} \mid \bigcap_{k \in T_2} \overline{A_k}\right) \geqslant (1 - x_{j_1}) \ldots (1 - x_{j_r}) \geqslant \prod_{(i,j) \in E} (1 - x_j).$$

   Следовательно, $\mathsf{P}\left(A_i \mid \bigcap\limits_{j \in T} \overline{A_j}\right) \leqslant \frac{x_i \prod\limits_{(i,j) \in E} (1-x_j)}{\prod\limits_{(i,j) \in E} (1-x_j)} = x_i$.

2. Запишем окончательный результат рассуждений выше с использованием формулы полной вероятности и полученной в первом пункте оценки.

$$\mathsf{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) = \mathsf{P}\left(\overline{A_1} \mid \bigcap_{i=2}^{n} \overline{A_i}\right) \mathsf{P}\left(\overline{A_2} \mid \bigcap_{i=3}^{n} \overline{A_i}\right) \ldots \mathsf{P}\left(\overline{A_n}\right) \geqslant (1 - x_1)(1 - x_2) \ldots (1 - x_n).$$

□

Упражнение. Сформулируйте и докажите несимметрический случай локальной леммы Ловаса, используя алгебры событий вместо графа зависимости.

Однако на практике гораздо чаще используется более слабая версия леммы.

Теорема 15 (Симметричный случай локальной леммы Ловаса)**.** *Пусть* $A_1, A_2, \ldots, A_n$ *— события в произвольном вероятностном пространстве* (Ω, P)*. Если выполнены следующие условия:*

1. *существует* $p \in [0, 1)$ *такое, что* $\mathsf{P}(A_i) \leqslant p$ *для любого* $1 \leqslant i \leqslant n$*;*
2. *каждое событие* A_i *взаимно независимо со всеми событиями за исключением не более, чем* d *событий;*
3. $ep(d+1) \leqslant 1$,

то

$$\mathsf{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

Доказательство. Для начала рассмотрим экстремальный случай — $d = 0$. В таком случае все события независимы и утверждение теоремы выполнено.

Теперь перейдём к общему случаю. Рассмотрим граф зависимости $D = (V, E)$ для событий $A_1, A_2, \ldots, A_n$, в котором для любого $i \in \{1, 2, \ldots, n\}$ выполнено $|\{j \mid (i,j) \in E\}| \leqslant d$. Далее, для всех i положим $x_i = \frac{1}{d+1}$. Покажем, что в этом случае выполняется требование несимметричного случая локальной леммы Ловаса:

$$x_i \prod_{(i,j) \in E} (1 - x_j) \geqslant \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d \geqslant \frac{1}{e(d+1)} \geqslant p \geqslant \mathsf{P}(A_i).$$

Как видно, оно выполняется. Тогда по локальной лемме получаем, что

$$\mathsf{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$ □

1.6.2 Задача k-SAT. Теорема Мозеса-Тардоша

Сейчас мы обсудим так называемую задачу k-SAT и её методы решения, основанные на вероятностном подходе. Но начнём мы с формулировки задачи.

Определение 30. Пусть есть некоторая булева формула $f(x_1, x_2, \ldots, x_n)$. Тогда *k-КНФ* называют представление этой формулы в виде

$$f(x_1, x_2, \ldots, x_n) = \bigwedge_{i=1}^{m} (l_{i1} \vee l_{i2} \vee \cdots \vee l_{ik}),$$

где $l_{ij} \in \{x_1, \ldots, x_n\} \cup \{\overline{x_1}, \overline{x_2}, \ldots, \overline{x_n}\}$, причём в дизъюнкте все переменные разные (то есть нет повторов).

Задача 1 (k-SAT)**.** *Пусть есть некоторая булева функция $f(x_1, x_2, \ldots, x_n)$, представленная в виде k-КНФ. Выполнима ли функция f, то есть существует ли такой набор переменных $v_1, v_2, \ldots, v_n$, что $f(v_1, v_2, \ldots, v_n) = 1$?*

Докажем следующее утверждение:

Теорема 16. *Если k-КНФ для функции $f(x_1, x_2, \ldots, x_n)$ содержит меньше, чем 2^k дизъюнктов, то f выполнима.*

Доказательство. Рассмотрим ситуацию, когда переменным равновероятно и независимо присваивается либо 0, либо 1. Тогда $\mathsf{P}(x_i = 0) = \mathsf{P}(x_i = 1) = \frac{1}{2}$ для $i \in \{1, 2, \ldots, n\}$.

Чему равна вероятность того, что дизъюнкт $l_{i1} \vee l_{i2} \vee \cdots \vee l_{ik}$ будет равен 0? Она равна

$$\mathsf{P}(l_{i1} = 0, l_{i2} = 0, \ldots, l_{ik} = 0) = \mathsf{P}(l_{i1} = 0)\,\mathsf{P}(l_{i2} = 0)\ldots\mathsf{P}(l_{ik} = 0) = \frac{1}{2^k}.$$

Теперь посмотрим вероятность того, что вся КНФ будет равна 0:

$$\begin{aligned}\mathsf{P}(f(x_1, x_2, \ldots, x_n) = 0) &= \mathsf{P}\left(\bigcup_{i=1}^{m} \{l_{i1} \vee l_{i2} \vee \cdots \vee l_{ik} = 0\}\right) \\ &\leqslant \sum_{i=1}^{m} \mathsf{P}(l_{i1} \vee l_{i2} \vee \cdots \vee l_{ik} = 0) = \frac{m}{2^k} < 1.\end{aligned}$$

Отсюда получаем, что вероятность того, что КНФ будет равна 1, больше нуля. А это и означает выполнимость f. □

Теперь посмотрим на следующий вопрос. Пусть каждая переменная (или её отрицание) лежит не более, чем в d скобках. При каком d мы можем гарантировать выполнимость? Ответ на этот вопрос даёт следующая теорема:

Теорема 17. *Пусть булева формула $f(x_1, x_2, \ldots, x_n)$ записана в виде k-КНФ. Также пусть любая переменная или ее отрицание входят в не более, чем $\frac{2^{k-1}}{ek}$ дизъюнкций. Тогда f выполнима.*

Доказательство. Так же, как и в прошлой теореме, рассмотрим ситуацию, когда переменным равновероятно и независимо присваивается либо 0, либо 1. Тогда $\mathsf{P}(x_i = 0) = \mathsf{P}(x_i = 1) = \frac{1}{2}$ для $i \in \{1, 2, \ldots, n\}$.

Рассмотрим событие $A_i = \{i\text{-й дизъюнкт равен } 0\}$. Исходя из того, что каждый дизъюнкт содержит k литералов, а также того, что все значения переменных равновероятны, мы можем оценить вероятность каждого события следующим образом:

$$\mathsf{P}(A_i) = \begin{cases} 0, & \text{в дизъюнкте одновременно есть } y \text{ и } \overline{y}, \\ \frac{1}{2^k}, & \text{иначе.} \end{cases}$$

Отсюда видно, что $\mathsf{P}(A_i) \leqslant \frac{1}{2^k}$ для всех $i \in \{1, 2, \ldots, m\}$.

Теперь посмотрим на количество событий, зависимых с A_i. Для этого рассмотрим зависимость на каком-либо примере. Пусть i-й дизъюнкт равен $x_1 \vee \overline{x_2} \vee \overline{x_3} \vee \cdots \vee \overline{x_k}$. Тогда A_i будет независимо с теми событиями, соответствующие дизъюнкты которых не содержат элементов из $\{x_1, x_2, \ldots, x_k\} \cup \{\overline{x_1}, \overline{x_2}, \ldots, \overline{x_k}\}$.

Тогда количество событий, зависимых с A_i равно количеству дизъюнктов, совпадающих хотя бы по одной переменной (с точностью до отрицания). Обозначим его за S_i. Попробуем ограничить его сверху. Выберем произвольную позицию, которая будет общей.

Так как переменная в этой позиции входит в не более, чем $\frac{2^{k-1}}{ek}$ дизъюнкций и переменные должны совпадать с точностью до отрицания, то есть не более, чем $2k\frac{2^{k-1}}{ek} = \frac{2^k}{e}$. Однако мы посчитали и сам дизъюнкт. Поэтому

$$|S_i| \leqslant \frac{2^k}{e} - 1 = d.$$

Теперь покажем, что выполняется и последнее условие симметричного случая локальной леммы Ловаса: $ep(d+1) \leqslant \frac{e}{2^k}\frac{2^k}{e} = 1$.

Отсюда по симметричной локальной лемме Ловаса получаем, что $\mathsf{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) > 0$. Это означает, что вероятность того, что при заданном наборе переменных функция равна 1 больше нуля, что доказывает выполнимость f. □

Из общего случая локальной леммы можно вывести следующую оценку для задачи выше.

$$\mathsf{P}\left(\bigcap_{i=1}^m \overline{A_i}\right) \geqslant \prod_{i=1}^m (1-x_i) = \left(\frac{d}{d+1}\right)^m = \left(1 - \frac{e}{2^k}\right)^m = q.$$

Теперь можно задаться вполне резонным вопросом: а как все же искать такое выполняющее означивание? Рассмотрим два подхода к решению этой проблемы.

1. *Наивный подход*: подставляем случайные значения $(x_1, x_2, \ldots, x_n)$. Если $f = 0$, тогда независимо подставим новые значения $(y_1, y_2, \ldots, y_n)$. Так будем повторять до победного конца.

   Пусть X — число подстановок новых пар значений переменных (*resampling steps*). Легко понять, что $X \sim \mathrm{Geom}(q)$ и[9]

   $$\mathsf{E}[X] = \sum_{k=0}^{\infty} kq(1-q)^k = \frac{1}{q} = \mathcal{O}\left(\left(\frac{2^k}{2^k - e}\right)^m\right).$$

2. *Продвинутый подход*: подставляем случайные значения $(x_1, x_2, \ldots, x_n)$. Если $f = 0$, то смотрим, какие именно дизъюнкты равны 0. Если $l_{i1} \vee l_{i2} \vee \cdots \vee l_{ik} = 0$, то меняем значения переменных $l_{i1}, l_{i2}, \ldots, l_{ik}$ на случайные независимым образом. Данная операция и будет служить *resampling step*. Продолжаем это до тех пор, пока не получим $f = 1$.

Алгоритм в продвинутом подходе крепко связан с локальной леммой, ибо это есть ни что иное, как алгоритмическая версия локальной леммы Ловаса. Данный алгоритм называется *алгоритмом Мозера-Тардоша* и для него верна следующая теорема:[10]

Теорема 18. *Пусть Y_i — количество resampling steps для события $A_i = \{i$-й дизъюнкт равен $0\}$ В условиях задачи k-SAT для любых $1 \leqslant i \leqslant m$, где m — это количество дизъюнктов в формуле, $\mathsf{E}[Y_i] \leqslant \frac{1}{d} = \frac{e}{2^k - e}$.*

Пусть $Y = \sum_{i=1}^m Y_i$ — общее число resampling steps. Тогда $\mathsf{E}[Y] \leqslant \frac{m}{d} = \mathcal{O}(m)$.

[9] За доказательством этого обращайтесь ко второй задаче 4-го семинара.

[10] – А мы не доказываем эту теорему? – Нет. Это тяжело. (Д.А. Шабанов)

1.6.3 Геометрические вероятности. Задача о встрече

Определение 31. Пусть вероятностное пространство $\Omega \in \mathbb{R}^n$ таково, что у Ω определен n-мерный объем, $0 < \mu(\Omega) < \infty$. Тогда в геометрических вероятностях для любого $A \in \Omega$, у которого тоже определен объем $\mu(A)$, полагаем вероятность $\mathsf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}$. Это естественное обобщение классической модели на непрерывный случай.

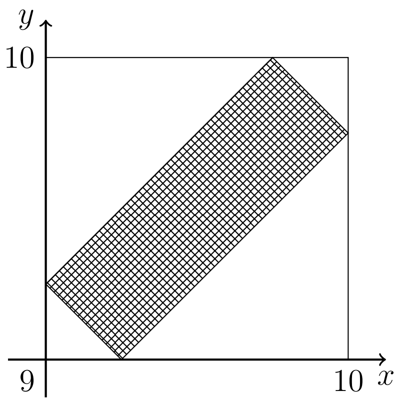
Примечание. Естественно оно по той простой причине, что аналогично дискретно случаю. Да и вообще вполне логично.

Для чего вводится такое определение? Бывают ситуации, когда конечным числом элементарных исходов обойтись нельзя. Ярким примером этому служит классическая *задача о встрече*

Задача 2 (Задача о встрече). *Два друга договорились встретиться на автобусной остановке между 9-ю и 10-ю часами утра. Они договорились, что встретятся в течение этого часа, но не условились о конкретном времени встречи. При этом, если, прождав 15 минут с момента своего прихода, один из друзей не видит второго, он уходит, а значит встреча не состоялась. Вопрос: с какой вероятностью друзья встретятся?*

Решение. Вообще, то, что конечности вероятностного пространства не хватит, понятно сразу, ведь моменты времени, в которые люди приходят на остановку, конечно же, не дискретны. Отложим на оси x и на оси y по отрезку длины 1, которые будут символизировать час между 9-ю и 10-ю часами, в течение которого на остановку приходит первый и второй человек соответственно. Итак, вероятностное пространство $\Omega = [9, 10] \times [9, 10]$.

Событию A, что встреча произойдет, удовлетворяют все точки получившегося квадрата, координаты которых отличатся не более, чем на четверть. То есть, моменты времени прихода друзей отличаются не более, чем на 15 минут. $A = \{\text{встреча произойдет}\} = \{(x, y) \mid |x - y| \leqslant \frac{1}{4}\}$. Изобразим это:



Довольно интуитивно хочется посчитать вероятность $\mathsf{P}(A)$ как отношение площадей закрашенной области (всех удовлетворяющих исходов) ко всей площади квадрата (множество всех исходов). Это и будет геометрическая вероятность.

$$\mathsf{P}(A) = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

□

1.7 Лекция от 21.10.2016

Будем дальше двигаться в сторону непрерывных вероятностей. Наша задача состоит в определении вероятностного пространства в общем случае.

1.7.1 Понятие вероятностного пространства в общем случае

Определение 32. *Вероятностным пространством* (или *тройкой Колмогорова*) называют тройку $(\Omega, \mathcal{F}, \mathsf{P})$, где Ω — множество элементарных исходов, $\mathcal{F}$ — множество событий, а P — вероятностная мера.

Разберем по отдельности каждый символ из этой "тройки".

- Ω — произвольное множество элементарных исходов. Как и в дискретном случае, во время эксперимента возможно получить один и только один элементарный исход.

  Например, если событие — это выстрел в мишень, то множество элементарных исходов будет задаваться плоскостью: $\Omega = \mathbb{R}^2$.

- $\mathcal{F}$ — это совокупность подмножеств Ω, называемых *событиями*. В дискретном случае множество событий содержало все подмножества Ω ($\mathcal{F} = 2^\Omega$) и не было никаких проблем. Однако в общем случае возникают некоторые ограничения на $\mathcal{F}$. Обсудим, какие же. Перед этим напомним определение *алгебры*:

  Определение 33. Пусть задано некоторое множество Ω. Система $\mathcal{F}$ подмножеств Ω называется *алгеброй*, если выполняются следующие требования:

  1. $\varnothing \in \mathcal{F}$.
  2. Если $A \in \mathcal{F}$, то и его дополнение $\overline{A} = \Omega \setminus A \in \mathcal{F}$.
  3. Если $A, B \in \mathcal{F}$, то их пересечение и объединение тоже лежит в $\mathcal{F}$.

  Вообще говоря, достаточно требовать либо пересечение, либо объединение, так как по закону де Моргана $A \cup B = \overline{\overline{A} \cap \overline{B}}$.

  Пример. Конечные объединения множеств вида $(-\infty, a]$, $(b, c]$, $(d, +\infty)$ образуют алгебру.

  Далее, несколько усилим это определение, добавив дополнительное условие:

  Определение 34. Пусть задано некоторое множество Ω. Система $\mathcal{F}$ подмножеств Ω называется *σ-алгеброй*, если выполняются следующие требования:

  1. $\mathcal{F}$ является алгеброй.
  2. Для любой последовательности событий $\{A_n\}_{n=1}^{\infty}$ их пересечение $\bigcap\limits_{n=1}^{\infty} A_n$ и объединение $\bigcup\limits_{n=1}^{\infty} A_n$ лежат в $\mathcal{F}$.

  В данном случае опять же можно обойтись одним из двух — либо пересечением, либо объединением.

  В общем случае полагается, что $\mathcal{F}$ обязана быть σ-алгеброй.[11]

  Покажем, что среди σ-алгебр над одним множеством (их, вообще говоря, может быть много) есть минимальная.

[11] Вообще, это требуется для того, чтобы всё не поломалось: http://stats.stackexchange.com/questions/199280/why-do-we-need-sigma-algebras-to-define-probability-spaces

Лемма. *Пусть M — некоторая система подмножеств Ω. Тогда существует минимальная (по включению) σ-алгебра, содержащая в себе все подмножества из M.*

Идея доказательства. Очевидно, что существуют σ-алгебры, содержащие M (та же 2^{Ω}). Тогда рассмотрим все σ-алгебры, содержащие M. Их пересечение тоже является σ-алгеброй и, при этом, минимально по включению. □

Минимальную σ-алгебру, содержащую M, принято обозначать через $\sigma(M)$.

Над $\mathbb{R}$ принято вводить особую σ-алгебру.

Определение 35. *Борелевской σ-алгеброй* на множестве $\mathbb{R}$ называется минимальная σ-алгебра, содержащая все полуинтервалы вида $(a, b]$, где $a, b \in \mathbb{R}$:

$$\mathcal{B}(\mathbb{R}) = \sigma \left\{ (a, b] \mid a, b \in \mathbb{R}, a < b \right\}.$$

Упражнение. $\mathcal{B}(\mathbb{R})$ можно определить через интервалы, лучи, отрезки, открытые и замкнутые множества — это даст одну и ту же σ-алгебру.

Можно показать, что $\mathcal{B}(\mathbb{R}) \neq 2^{\mathbb{R}}$ (например, множество Витали[12] не входит в $\mathcal{B}(\mathbb{R})$).

Определение 36. Пара $(\Omega, \mathcal{F})$, где $\mathcal{F}$ — σ-алгебра подмножеств Ω, называется *измеримым пространством.*

- Наконец, мы подошли к рассмотрению последнего члена “тройки Колмогорова”.

Определение 37. Отображение P из $\mathcal{F}$ в $[0;\ 1]$ ($(\mathsf{P} : \mathcal{F} \to [0; 1])$) называется *вероятностной мерой* на $(\Omega, \mathcal{F})$, если оно удовлетворяет следующим двум условиям:

1. $\mathsf{P}(\Omega) = 1$.
2. *Счётная аддитивность:* для любой последовательности попарно непересекающихся событий $\{A_n\}_{n=1}^{\infty}$ выполнено, что

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

В дискретных вероятностных пространствах свойство счётной аддитивности доказывалось. Здесь его сразу постулируют.

Лемма (Свойства вероятности)**.** *Вероятностные меры обладают следующими свойствами:*

1. $\mathsf{P}(\varnothing) = 0$.
2. *Конечная аддитивность. Если $\{A_1, A_2, \ldots, A_n\}$ — попарно не пересекающиеся события, то*

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

3. *Если $A, B \in \mathcal{F}$ и $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.*

[12] https://en.wikipedia.org/wiki/Vitali_set

4. $\mathsf{P}(A) + \mathsf{P}(\overline{A}) = 1$.

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

6. Для любого набора событий $A_1, A_2, \ldots, A_n$ $\mathsf{P}\left(\bigcup\limits_{n=1}^{m} A_n\right) \leqslant \sum\limits_{n=1}^{m} \mathsf{P}(A_n)$.

Доказательство. Докажем только первые два свойства, так как остальные доказываеются ровно так же, как и в дискретном случае. За этими доказательствами можно обратиться к первой лекции.

1. Рассмотрим последовательность $\{A_n\}_{n=1}^{\infty}$, где для любого $n \in \mathbb{N}$ $A_n = \varnothing$. Тогда $A_i \cap A_j = \varnothing$ для любых $i, j \in \mathbb{N}$ и по свойству счётной аддитивности

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \begin{cases} \mathsf{P}\left(\bigcup\limits_{n=1}^{\infty} \varnothing\right) = \mathsf{P}(\varnothing) \\ \sum\limits_{n=1}^{\infty} \mathsf{P}(A_n) = \sum\limits_{n=1}^{\infty} \mathsf{P}(\varnothing) \end{cases}$$

   Отсюда получаем, что

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

   У этого уравнения есть два решения — 0 и ∞. Но так как $\mathsf{P}(\varnothing) \in [0, 1]$, то $\mathsf{P}(\varnothing) = 0$.

2. Положим $A_n = \varnothing$ при $n > m$. Тогда по свойству счётной аддитивности:

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(A_n) = \sum_{n=1}^{m} \mathsf{P}(A_n) + \sum_{n=m+1}^{\infty} \mathsf{P}(\varnothing) = \sum_{n=1}^{m} \mathsf{P}(A_n). \qquad \square$$

Теорема 19 (О непрерывности вероятностной меры)**.** *Пусть* P *— конечно-аддитивная функция на* σ*-алгебре событий* $\mathcal{F}$, $\mathsf{P} : \mathcal{F} \to [0; 1]$, $\mathsf{P}(\Omega) = 1$. *Тогда следующие четыре условия эквивалентны:*

(a) P *является счетно-аддитивной функцией.*

(b) P *непрерывна в "нуле", то есть для любых множеств* $A_1, A_2, \ldots \in \mathcal{F}$ *таких, что* $A_{n+1} \subset A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$, *выполняется:*

$$\lim_{n} \mathsf{P}(A_n) = 0.$$

(c) P *непрерывна сверху, то есть для любых множеств* $A_1, A_2, \ldots \in \mathcal{F}$ *таких, что* $A_n \subset A_{n+1}$, $\bigcup\limits_{n=1}^{\infty} A_n = A \in \mathcal{F}$, $(A_n \uparrow A)$, *выполняется:*

$$\lim_{n} \mathsf{P}(A_n) = \mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(d) P *непрерывна снизу, то есть для любых множеств* $A_1, A_2, \ldots \in \mathcal{F}$ *таких, что* $A_{n+1} \subset A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = A \in \mathcal{F}$, $(A_n \downarrow A)$, *выполняется:*

$$\lim_n \mathsf{P}(A_n) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Доказательство. Докажем эквивалентность всех этих утверждений в несколько шагов:

$[a \Rightarrow c]$ Покажем, что счётная аддитивность влечёт непрерывность сверху. Заметим, что $\bigcup_{n=1}^{\infty} A_n$ можно представить в виде счётного объединения попарно непересекающихся множеств:

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \ldots .$$

Действительно, $A_{i+1} \setminus A_i$ содержит элементы, лежащие только в A_{i+1}. Тогда по счётной аддитивности

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathsf{P}(A_1) + \mathsf{P}(A_2 \setminus A_1) + \mathsf{P}(A_3 \setminus A_2) + \ldots =$$
$$= \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1) + \mathsf{P}(A_3) - \mathsf{P}(A_2) + \ldots = \lim_{n\to\infty} \mathsf{P}(A_n).$$

$[c \Rightarrow d]$ Покажем, что непрерывность сверху влечёт непрерывность снизу. Рассмотрим некоторую последовательность $A_n \downarrow A$. Тогда для любого $n \geqslant 1$ выполнено, что $A_n \subseteq A_1$ и $A_n = A_1 \setminus (A_1 \setminus A_n)$. Следовательно,

$$\mathsf{P}(A_n) = \mathsf{P}(A_1 \setminus (A_1 \setminus A_n)) = \mathsf{P}(A_1) - \mathsf{P}(A_1 \setminus A_n).$$

Заметим, что последовательность $\{B_n\}_{n=1}^{\infty}$, где $B_n = A_1 \setminus A_n$ является неубывающей и $B_n \uparrow A_1 \setminus A$. Тогда, согласно пункту c,

$$\lim_{n\to\infty} \mathsf{P}(A_1 \setminus A_n) = \mathsf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right).$$

А это означает, что

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}(A_1) - \lim_{n\to\infty} \mathsf{P}(A_1 \setminus A_n) = \mathsf{P}(A_1) - \mathsf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right)$$
$$= \mathsf{P}(A_1) - \mathsf{P}\left(A_1 \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)\right) = \mathsf{P}(A_1) - \mathsf{P}(A_1) + \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$$
$$= \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

$[d \Rightarrow b]$ Достаточно взять последовательность $A_n \downarrow \varnothing$. Тогда из непрерывности снизу будет следовать непрерывность в нуле.

$[b \Rightarrow a]$ Покажем, что непрерывность в нуле даёт счётную аддитивность. Рассмотрим некоторую последовательность попарно непересекающихся событий $\{A_n\}_{n=1}^{\infty}$. Заметим, что

$$\mathsf{P}\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \mathsf{P}\left(\bigsqcup_{i=1}^{n} A_i\right) + \mathsf{P}\left(\bigsqcup_{i=n+1}^{\infty} A_i\right).$$

Осталось только заметить, что $\bigsqcup\limits_{i=n+1}^{\infty} A_i \downarrow \varnothing$ при $n \to \infty$. Тогда

$$\begin{aligned}\sum_{i=1}^{\infty} \mathsf{P}(A_i) &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{P}(A_i) = \lim_{n\to\infty} \mathsf{P}\left(\bigsqcup_{i=1}^{n} A_i\right) \\ &= \lim_{n\to\infty}\left(\mathsf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) - \mathsf{P}\left(\bigsqcup_{i=n+1}^{\infty} A_i\right)\right) \\ &= \mathsf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) - \lim_{n\to\infty} \mathsf{P}\left(\bigsqcup_{i=n+1}^{\infty} A_i\right) = \mathsf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right).\end{aligned}$$

□

1.7.2 Вероятностные меры на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Пусть P — некоторая вероятностная мера.

Определение 38. *Функцией распределения* вероятностной меры P называют функцию $F : \mathbb{R} \to [0, 1]$ такую, что $F(x) = \mathsf{P}((-\infty, x])$.

Лемма (Свойства функции распределения)**.**

1. $F(x)$ *неубывающая;*
2. $\lim\limits_{x\to+\infty} F(x) = 1$;
3. $\lim\limits_{x\to-\infty} F(x) = 0$;
4. $F(x)$ *непрерывна справа.*

Доказательство.

1. Если $y > x$, то в силу аддитивности P, $F(y) - F(x) = \mathsf{P}((x, y]) \geqslant 0$.
2. Пусть $x_n \uparrow +\infty$; тогда $(-\infty, x_n] \uparrow \mathbb{R}$. Значит, $\lim\limits_{x_n\to+\infty} F(x_n) = \lim\limits_{x_n\to+\infty} \mathsf{P}((-\infty, x_n]) = \mathsf{P}(\mathbb{R}) = 1$
3. Аналогично.
4. Пусть $x_n \downarrow x$; тогда, в силу непрерывности вероятностной меры, $F(x_n) \to F(x)$.

□

Примеры:

- Функция распределения константной случайной величины:

$$F(x) = \begin{cases} 1, x \geqslant c; \\ 0, x < c \end{cases}$$

- Функция распределения случайной величины, равномерно распределённой на $[0,1]$:

$$F(x) = \begin{cases} 1, x > 1; \\ x, x \in [0,1]; \\ 0, x < 0 \end{cases}$$

Такой функции распределения соответствует вероятностная мера P такая, что $\forall a < b \in [0,1]$, $\mathsf{P}((a,b]) = b - a$. Эту меру также называют *мерой Лебега* [13].

1.8 Лекция от 01.11.2016

1.8.1 Взаимно-однозначное соответствие функции распределения и вероятностной меры

Определение 39. Пусть функция $F(x), x \in \mathbb{R}$, удовлетворяет свойствам из леммы, то есть:

1. $F(x)$ неубывающая;
2. $\lim\limits_{x\to+\infty} F(x) = 1$;
3. $\lim\limits_{x\to-\infty} F(x) = 0$;
4. $F(x)$ непрерывна справа.

Тогда такую функцию будем называть *функцией распределения* на прямой.

Следующая фундаментальная и очень важная теорема будет введена без доказательства.

Теорема 20 (Теорема Каратеодори о продолжении меры)**.** *Пусть Ω — некоторое множество, $\mathcal{A}$ – алгебра подмножеств Ω. Пусть вероятностная мера $\mathsf{P}_0 : \mathcal{A} \mapsto [0;1]$ удовлетворяет следующим свойствам:*

1. $\mathsf{P}_0(\Omega) = 1$;

2. P_0 *— счетно-аддитивна на* $\mathcal{A}$.

Тогда существует (и притом единственна) вероятностная мера P на $\sigma(\mathcal{A})$ такая, что мера P является продолжением меры P_0, иными словами $\forall A \in \mathcal{A}$ $\mathsf{P}_0(\mathcal{A}) = \mathsf{P}(\mathcal{A})$.

Теорема 21 (Взаимно-однозначное соответствие функций распределения и мер на прямой)**.** *Пусть $F(x), x \in \mathbb{R}$ — функция распределения на прямой. Тогда существует (и притом единственна) вероятностная мера P на $\mathcal{B}$, такая что $F(x)$ — её функция распределения, то есть $\forall x \in \mathbb{R}$ $\mathsf{P}((-\infty; x]) = F(x)$.*

[13] Говоря простым языком, вероятность попадания в полуинтервал пропорциональна его длине.

Доказательство. Рассмотрим алгебру $\mathcal{A}$, состоящую из конечных объединений непересекающихся полуинтервалов вида $(a;b]$, $\forall A \in \mathcal{A}$ имеет вид

$$A = \bigcup_{k=1}^{n} (a_k; b_k], \; -\infty \leqslant a_1 < b_1 < a_2 < b_2 < \ldots < b_n \leqslant +\infty.$$

Зададим на $\mathcal{A}$ меру P_0 по правилу

$$\mathsf{P}_0(A) = \sum_{k=1}^{n} \left(F(b_k) - F(a_k) \right).$$

Тогда $\mathsf{P}_0(\mathbb{R}) = F(+\infty) - F(-\infty) = 1 - 0 = 1$ и по построению P_0 будет конечно-аддитивной мерой.

Для того чтобы воспользоваться теоремой Каратеодори, вероятностная мера P_0 должна обладать свойством счетной аддитивности. Но вместо того, чтобы доказывать это свойство напрямую, воспользуемся *теоремой о непрерывности вероятностной меры* из предыдущей лекции и докажем эквивалентное условие. Покажем, что мера P_0 является непрерывной в нуле.

Итак, нужно проверить, выполняется ли

$$\lim_{n \to \infty} \mathsf{P}_0(A_n) = 0$$

для любых множеств $A_1, A_2, \ldots \in \mathcal{A}$ таких, что $A_{n+1} \subset A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$

Заметим, что для любого полуинтервала $(a, b]$ и для любого $\delta > 0$ можно взять такое $a' > a$, что:

$$\mathsf{P}_0\left((a;b]\right) - \mathsf{P}_0\left((a';b]\right) \leqslant \delta$$

$$\mathsf{P}_0\left((a;b]\right) - \mathsf{P}_0\left((a';b]\right) = \left(F(b) - F(a)\right) - \left(F(b) - F(a')\right) = F(a') - F(a).$$

Зафиксируем произвольное $\varepsilon > 0$. В силу непрерывности справа такое значение a' можно подобрать. Тогда

$$\forall A_n \exists B_n \in \mathcal{A}, \; \mathsf{P}_0(A_n) - \mathsf{P}_0(B_n) \leqslant \varepsilon 2^{-n}, \; B_n \in A_n \text{ **и** } [B_n] \in A_n, \text{ где } [B_n] \text{ есть замыкание } B_n.$$

То есть если B_n имеет вид $(a, b]$, то $[B_n]$, являющееся замыканием B_n, имеет вид $[a, b]$. Мы добавили в множество B_n его граничные точки.

Пусть сначала все A_n лежат внутри $[-N; N]$ для некоторого $N \in \mathbb{N}$. Мы знаем, что $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$. Следовательно, $\bigcap\limits_{n=1}^{\infty} [B_n] = \varnothing$.

По принципу компактности существует n_0 такой, что из $\bigcap\limits_{n=1}^{n_0} [B_n] = \varnothing$ следует $\bigcap\limits_{n=1}^{n_0} B_n = \varnothing$. Тогда:

$$\mathsf{P}_0(A_{n_0}) = \mathsf{P}_0\left(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n \right)$$

Следующий переход основан на том, что если $\omega \in \mathcal{A}_n \setminus \bigcap\limits_{n=1}^{n_0} B_n$, то существует такой номер k, что $\omega \notin B_k$, а значит $\omega \in \mathcal{A}_{n_0} \setminus B_k$.

$$\mathsf{P}_0\left(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n\right) \leqslant \sum_{k=1}^{n_0} \mathsf{P}_0\left(A_{n_0} \setminus B_k\right) \leqslant \sum_{k=1}^{n_0} \mathsf{P}_0\left(A_k \setminus B_k\right) \leqslant$$

$$\leqslant \sum_{k=1}^{n_0} \left(\mathsf{P}_0\left(A_k\right) - \mathsf{P}_0\left(b_k\right)\right) \leqslant \sum_{k=1}^{n_0} \varepsilon 2^{-k} \leqslant \varepsilon.$$

Получили, что $\mathsf{P}_0(A_{n_0}) \leqslant \varepsilon$. Значит,

$$\forall n > n_0 \;\; \mathsf{P}_0(A_n) \leqslant \varepsilon \implies \lim_{n\to\infty} \mathsf{P}_0(A_n) = 0.$$

Если в A_1 есть бесконечные полуинтервалы, то выберем полуинтервал $(-N; N]$, что $\mathsf{P}_0\left((-N; N]\right) \geqslant 1 - \frac{\varepsilon}{2}$ и рассмотрим $A'_n = A_n \cap (-N; N]$.

По доказанному выше $\mathsf{P}_0(A'_n) \leqslant \frac{\varepsilon}{2}$ при $n > n_0(\varepsilon)$ и

$$\mathsf{P}_0(A_n) = \mathsf{P}_0\left(A'_n \cup (A_n \cup (-\infty; -N]) \cup (A_n \cap (N; +\infty))\right) \leqslant \mathsf{P}_0\left(A'_n\right) + \mathsf{P}_0\left(\mathbb{R} \setminus (-N; N]\right) \leqslant \varepsilon.$$

По теореме Каратеодори искомая P существует и единственна. □

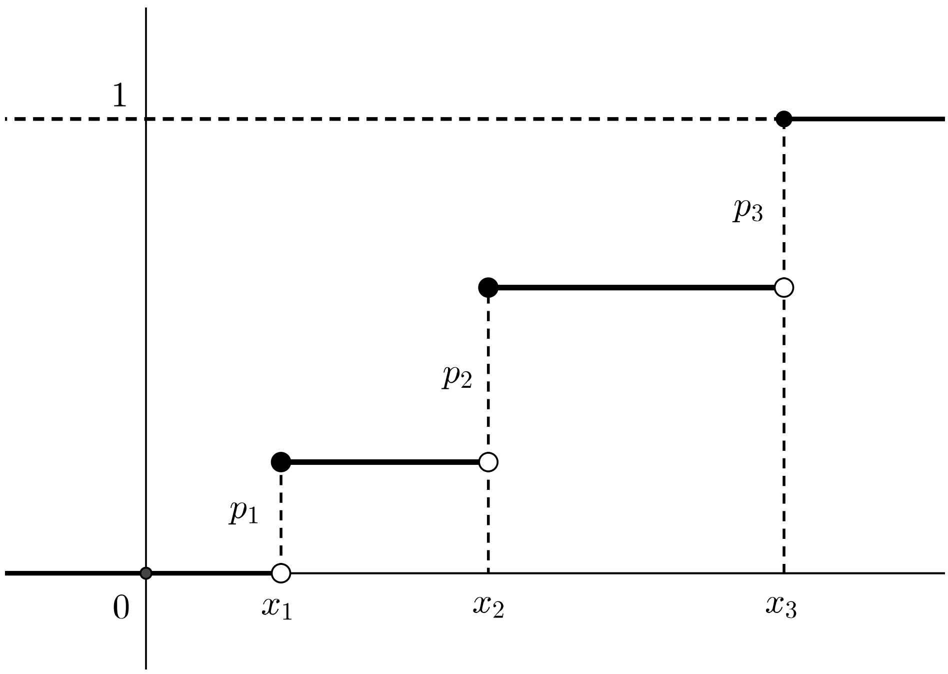
1.8.2 Классификация функций распределения на $\mathbb{R}$

Дискретные функции распределения

Вероятностная мера P на $\mathcal{B}(\mathbb{R})$ называется дискретной, если $\exists X = \{x_n\}$ — не более чем счётное множество, такое что $\mathsf{P}(\mathbb{R} \setminus X) = 0$ и $\forall x_n \in X, \mathsf{P}(x_n) > 0$. Между прочим, в этом случае говорят, что P *сосредоточена* в X.

Но нас интересует функция распределения. Она будет выглядеть так:

$$F(x) = \mathsf{P}((-\infty, x]) = \sum_{k: x_k \in X} \mathsf{P}(x_k)$$



Примеры:

- Равномерное дискретное распределение:

$$x = \{1 \dots n\}, \mathsf{P}_k = \frac{1}{n}$$

- Распределение Бернулли:

$$x = \{0, 1\}, \mathsf{P}_1 = p, \mathsf{P}_0 = 1 - p$$

- Биномиальное распределение:

$$\mathsf{P}_k = \binom{n}{k} p^k (1-p)^{n-k}$$

- Пуассоновское распределение:

$$\mathsf{P}_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

Абсолютно непрерывные функции распределения

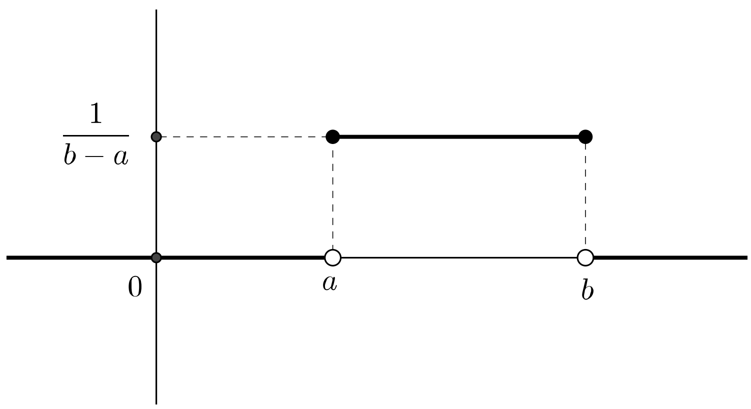
Рассмотрим некоторую функцию $p(t)$, неотрицательную на $\mathbb{R}$, такую, что $\int\limits_{-\infty}^{\infty} p(t)dt = 1$.

Определение 40. Функция распределения $F(x)$ называется абсолютно непрерывной с плотностью $p(t)$, если $\forall x \in \mathbb{R}, F(x) = \int\limits_{-\infty}^{x} p(t)dt$

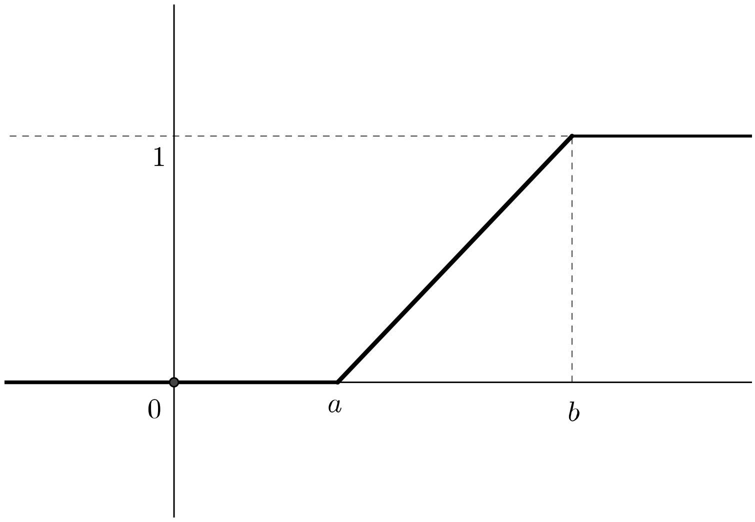
Примеры:

- Равномерное распределение на отрезке:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

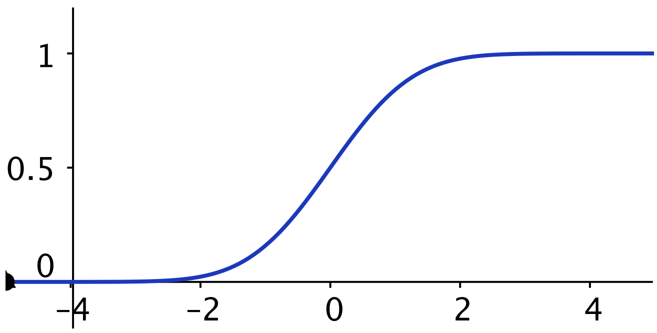


$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

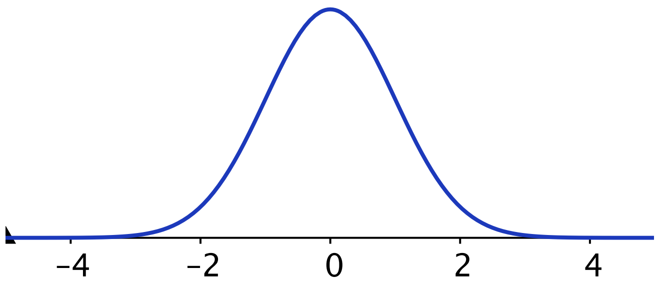


- Нормальное распределение: $N(a, \sigma^2)$. Самое известное и распространённое распределение, встречающееся всюду; например погрешности измерений, как правило, подчиняются этому распределению.

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

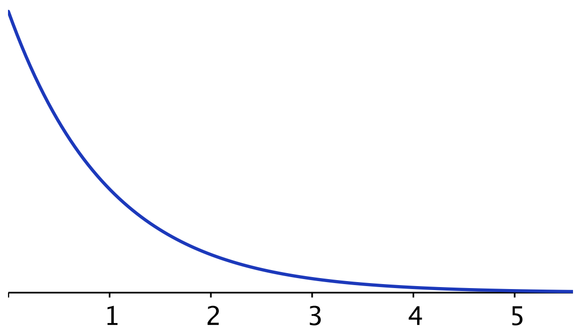


$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right), \text{где } \Phi(x) \text{ — функция Лапласа}$$

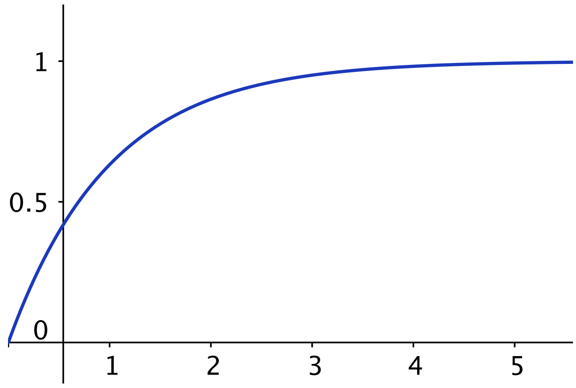


- Экспоненциальное распределение: $\mathrm{Exp}(\alpha)$.

$$p(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0, & otherwise \end{cases}$$



$$F(x) = \begin{cases} 1 - e^{-\alpha x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$$



Сингулярные функции распределения

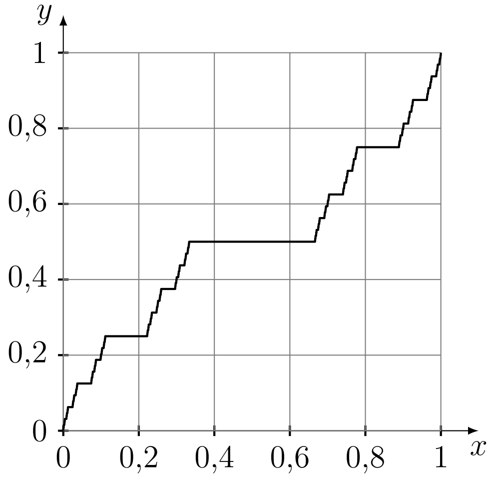
Определение 41. Точка y называется точкой роста функции $F(x), x \in \mathbb{R}$, если $\forall \varepsilon > 0, F(y+\varepsilon) - F(y-\varepsilon) > 0$.

Определение 42. $\mathcal{A} \subset \mathbb{R}$ имеет Лебегову меру 0, если $\forall \varepsilon$ существует набор интервалов $(a_n; b_n)$ таких что $\sum(b_k - a_k) < \varepsilon$ и $\mathcal{A} \subset \bigcup(a_k; b_k)$.

Определение 43. Функция распределения $F(x)$ называется *сингулярной*, если она непрерывна и множество её точек роста имеет Лебегову меру 0.

Пример:

- Канторова лестница:



1.9 Лекция от 08.11.2016

1.9.1 Канторова лестница. Продолжение.

Заметим, что канторова лестница — неубывающая непрерывная функция, которая возрастает от 0 до 1. Давайте рассмотрим множество её точек роста:

Легко понять, что точками роста могут быть только точки, не попавшие ни в один заполняемый интервал. Давайте оценим меру Лебега множества точек роста, пользуясь аддитивностью меры и тем, что мера $[0, 1]$ равна 1.

Сумма длин интервалов на i-ом шаге — $\frac{1}{3} \cdot \left(\frac{2}{3}\right)^i$; тогда всего эти интервалы заполняют множество меры $\sum\limits_{i=0}^{\infty} \frac{1}{3} \cdot \left(\frac{2}{3}\right)^i = 1$. Значит, мера множества точек роста равна $1 - 1 = 0$, значит, канторова лестница является сингулярной функцией распределения.

Казалось бы, а зачем вообще нужны сингулярные функции? Оказывается, что дискретные, абсолютно непрерывные и синуглярные функции покрывают все функции распределения.

Теорема 22 (Лебег)**.** *Пусть $F(x)$ — функция распределения на $\mathbb{R}$. Тогда имеет место представление $F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x) + \alpha_3 F_3(x)$, где*

- *F_1 — дискретная;*
- *F_2 — абсолютно непрерывная;*
- *F_3 — сингулярная.*

При этом $a_i \geqslant 0$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Доказательство. Это хорошая теорема. Доказывать её мы, конечно, не будем. □

1.9.2 Случайные величины и векторы

В дискретном случае мы говорили, что ξ — это произвольное отображение из Ω в R. Но в дискретном случае никаких сложностей не возникает, так как всегда можно посчитать любую вероятность. В общем случае такой красоты нет. Тогда нужно вводить некоторые ограничения.

Однако, было бы неплохо ввести обобщение так, чтобы дискретный случай был частным случаем общего (как и должно быть). То есть, если на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ задана некоторая случайная величина ξ, то мы хотим уметь считать вероятности вида $\mathsf{P}(\xi \leqslant x)$, $\mathsf{P}(\xi = x)$, $\mathsf{P}(a \leqslant \xi \leqslant b)$ и так далее. Поэтому придётся ввести одно ограничение.

Определение 44. Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathsf{P})$. Будем называть *случайной величиной* отображение $\xi : \Omega \mapsto \mathbb{R}$, если оно удовлетворяет свойству *измеримости*: для любого $x \in \mathbb{R}$ событие $\{\xi \leqslant x\} \equiv \{\omega \in \Omega \mid \xi(\omega) \leqslant x\}$ принадлежит $\mathcal{F}$.

Смысл измеримости описан ранее и скрыт в названии: если функция измерима, то мы можем её измерить.

Если выбрать случайную точку на той же плоскости, то как её описать? Введём понятие *случайного вектора*:

Определение 45. *Случайный вектор* — вектор, состоящий из случайных величин.

Казалось бы, почему мы вводим только такие события? Оказывается, этого более, чем достаточно и для случайных величин, и для случайных векторов. Однако для того, чтобы доказать это, нужно ввести ещё одно определение:

Определение 46. Борелевской σ-алгеброй в $\mathbb{R}^n$ называют минимальную σ-алгебру сожержащую все $(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]$:

$$\mathcal{B}(\mathbb{R}^n) = \sigma \left\{(a_1, b_1] \times \cdots \times (a_n, b_n] \mid a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}, a_i < b_i\right\}.$$

Примечание. Можно показать, что в $\mathcal{B}(\mathbb{R}^n)$ лежит много чего: все прямоугольники вида $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, все открытые и замкнутые множества, элементы вида $(-\infty, b_1] \times (-\infty, b_2] \times \cdots \times (-\infty, b_n]$ и так далее.

Теперь докажем то, что нашего определения достаточно:

Лемма. *Пусть $(\Omega, \mathcal{F}, \mathsf{P})$ — вероятностное пространство, и на нём заданы случайная величина ξ и случайный вектор $\eta = (\eta_1, \ldots, \eta_n)$. Тогда*

1. *ξ является случайной величиной тогда и только тогда, когда для любого $B \in \mathcal{B}(\mathbb{R})$ событие $\xi^{-1}(B) = \{\xi \in B\}$ лежит в $\mathcal{F}$.*

2. *ξ является случайным вектором тогда и только тогда, когда для любого $B \in \mathcal{B}(\mathbb{R}^n)$ событие $\eta^{-1}(B) = \{\eta \in B\}$ лежит в $\mathcal{F}$.*

Примечание. Если $B \in \mathcal{B}(\mathbb{R}^n)$, то его ещё (неформально) называют *борелевским множеством.*

Доказательство. Для начала докажем это для случайных величин.

[$\Leftarrow$] Заметим, что $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ для любого $x \in \mathbb{R}$. Тогда

$$\xi^{-1}\left((-\infty, x]\right) = \{\xi \in (-\infty, x]\} \in \mathcal{F}.$$

А это и есть определение измеримости. Следовательно, ξ — случайная величина.

[$\Rightarrow$] Рассмотрим следующее множество:

$$D = \left\{B \in \mathcal{B}(\mathbb{R}) \mid \xi^{-1}(B) \in \mathcal{F}\right\}.$$

Покажем, что оно является σ-алгеброй. Так как полный прообраз сохраняет все теоретико-множественные операции, то в D будут находиться пересечения, дополнения, объединения и счётные объединения. Тогда D действительно является σ-алгеброй.

Однако, согласно определению случайной величины, все лучи вида $(-\infty, x]$ лежат в D. Тогда D есть σ-алгебра, содержащая все лучи. Но тогда она содержит в себе $\mathcal{B}(\mathbb{R})$. Так как по построению $D \subseteq \mathcal{B}(\mathbb{R})$, то $D = \mathcal{B}(\mathbb{R})$.

Идейно доказательство для случайных векторов ничем не отличается.

[$\Leftarrow$] Заметим, что $(-\infty, x_1] \times \cdots \times (-\infty, x_n] \in \mathcal{B}(\mathbb{R}^n)$ для любых $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Тогда

$$\eta^{-1}\left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right) = \{\eta_1 \in (-\infty, x_1]\} \cap \cdots \cap \{\eta_n \in (-\infty, x_n]\} \in \mathcal{F}.$$

Тогда все $\eta_1, \ldots, \eta_n$ являются случайными величинами и η действительно является случайным вектором.

[⇒] Рассмотрим следующее множество:

$$D = \left\{B \in \mathcal{B}(\mathbb{R}) \mid \eta^{-1}(B) \in \mathcal{F}\right\}.$$

Покажем, что оно является σ-алгеброй. Так как полный прообраз сохраняет все теоретико-множественные операции, то в D будут находиться пересечения, дополнения, объединения и счётные объединения. Тогда D действительно является σ-алгеброй.

Однако, согласно определению случайной величины, все элементы вида $(-\infty, x_1] \times \dots \times (-\infty, x_n]$ лежат в D. Тогда D содержит в себе $\mathcal{B}(\mathbb{R})$. Так как по построению $D \subseteq \mathcal{B}(\mathbb{R})$, то $D = \mathcal{B}(\mathbb{R})$.

□

1.9.3 Действия над случайными величинами

Допустим, у нас есть несколько случайных величин $\xi_1, \xi_2, \dots, \xi_n$, и мы хотим с ними что-то сделать — перемножить, например. Другими словами, мы хотим взять *функцию* от случайных величин $f(\xi_1, \xi_2, \dots, \xi_n)$. Но можно ли сказать, что это действительно будет случайной величиной? Не всегда. Однако есть класс функций, называемых *борелевскими*, для которых это точно верно. Введём определение:

Определение 47. Пусть $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ — некоторая функция. Будем называть её *борелевской*, если выполнено следующее условие:

$$\forall B \in \mathcal{B}(\mathbb{R}^m)\ f^{-1}(B) \equiv \{x : f(x) \in B\} \in \mathcal{B}(\mathbb{R}^n).$$

Вообще, почти все "обычные" функции являются борелевскими. Например, непрерывные функции являются борелевскими, так как они переводят открытые множества в открытые. Можно так же доказать, что если у функции множество точек разрыва имеет лебегову меру 0, то она тоже является борелевской.[14]

Теперь покажем, что борелевские функции действительно ничего не ломают.

Теорема 23. *Пусть $\xi_1, \dots, \xi_n$ — случайные величины, а $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ — борелевская функция. Тогда $f(\xi_1, \xi_2, \dots, \xi_n)$ является случайной величиной.*

Доказательство. Рассмотрим произвольное борелевское множество $B \in \mathcal{B}(\mathbb{R}^m)$. Тогда

$$\{f(\xi_1, \dots, \xi_n) \in B\} = \left\{(\xi_1, \xi_2, \dots, \xi_n) \in f^{-1}(B)\right\}.$$

Так как $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$, то $\{f(\xi_1, \dots, \xi_n) \in B\} \in \mathcal{F}$. □

Следствие. *Если ξ и η — это случайные величины, то $\xi \pm \eta$, $\alpha\xi$ для любого $\alpha \in \mathbb{R}$, $\xi\eta$ и ξ/η при $\eta \neq 0$ тоже являются случайными величинами.*

В дальнейшем у нас будут возникать последовательности случайных величин. Хотелось бы понять — а будут ли те же пределы последовательностей случайными величинами? Ответ — да.

[14] Доказательство можно найти по следующей ссылке: http://math.stackexchange.com/questions/369268/a-function-with-countable-discontinuities-is-borel-measurable

Теорема 24. *Пусть $\{\xi_n\}_{n=1}^{\infty}$ — последовательность случайных величин. Тогда*

$$\varlimsup_{n\to\infty} \xi_n,\ \varliminf_{n\to\infty} \xi_n, \sup_{n\in\mathbb{N}} \xi_n,\ \inf_{n\in\mathbb{N}} \xi_n$$

тоже являются случайными величинами.

Доказательство. Покажем, что все они измеримы:

$$\left\{\sup_{n\in\mathbb{N}} \xi_n > x\right\} = \bigcup_{n=1}^{\infty} \underbrace{\{\xi_n > x\}}_{\in\mathcal{F}} \in \mathcal{F}$$

$$\left\{\inf_{n\in\mathbb{N}} \xi_n < x\right\} = \bigcup_{n=1}^{\infty} \underbrace{\{\xi_n < x\}}_{\in\mathcal{F}} \in \mathcal{F}$$

$$\varlimsup_{n\in\mathbb{N}} \xi_n = \inf_{n\in\mathbb{N}} \sup_{k\geqslant n} \xi_k$$

$$\varliminf_{n\in\mathbb{N}} \xi_n = \sup_{n\in\mathbb{N}} \inf_{k\geqslant n} \xi_k$$

□

1.9.4 Простые случайные величины. Матожидание в общем случае

Напомним определение индикатора:

Определение 48. Пусть $(\Omega, \mathcal{F}, \mathsf{P})$ — вероятностное пространство и $A \in \mathcal{F}$ — некоторое событие. Тогда *индикатором* события A будем называть функцию $\mathbf{I}_A : \Omega \mapsto \{0, 1\}$, устроенную следующим образом:

$$\mathbf{I}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

Иногда индикатор обозначают через $\mathbf{I}\{A\}$.

Покажем, что индикатор является случайной величиной. Действительно,

$$\{\mathbf{I}_A \leqslant x\} = \begin{cases} \varnothing, & x < 0 \\ \overline{A}, & 0 \leqslant x < 1 \\ \Omega, & x \geqslant 1 \end{cases}$$

Так как каждое из трёх подмножеств лежит в $\mathcal{F}$, то $\mathbf{I}_A$ является случайной величиной, что и требовалось доказать.

Индикатор является один из примеров так называемых *простых* случайных величин. Введём определение:

Определение 49. *Простая случайная величина* — случайная величина, принимающая конечное число значений.

Если $x_1, x_2, \ldots, x_n$ — все значения простой случайной величины ξ, то её можно представить в следующем виде:

$$\xi = \sum_{k=1}^{n} x_k \mathbf{I}\{\xi = x_k\}$$

Теперь о самом матожидании:

- Пусть ξ — простая случайная величина. Тогда, по аналогии с дискретным случаем, $\mathsf{E}[\xi]$ будет равно

  $$\mathsf{E}[\xi] \equiv \sum_{k=1}^{n} x_k \, \mathsf{P}(\xi = x_k).$$

  В дискретном случае похожая формула доказывалась, однако в общем случае она постулируется.

- Теперь допустим, что ξ — это произвольная неотрицательная случайная величина. Как ввести её матожидание? Допустим, что есть последовательность $\{\xi_n\}_{n=1}^{\infty}$ простых случайных величин такая, что $\xi_n(\omega) \uparrow \xi(\omega)$ для любого $\omega \in \Omega$. Тогда положим, что

  $$\mathsf{E}[\xi] \equiv \lim_{n\to\infty} \mathsf{E}[\xi_n].$$

  Однако возникает вопрос: а существует ли она? Докажем, что такая последовательность ξ_n действительно существует.

  Рассмотрим последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$, устроенную следующим образом:

  $$\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{I}\left\{\frac{k-1}{2^n} \leqslant \xi < \frac{k}{2^n}\right\} + n\,\mathbf{I}\left\{\xi \geqslant n\right\}.$$

  Непосредственная проверка убеждает в том, что $\xi_n(\omega) \uparrow \xi(\omega)$ для любого $\omega \in \Omega$.

1.10 Лекция от 15.11.2016

1.10.1 Свойства математического ожидания для простых случайных величин.

1. Линейность.

   Если ξ, η — простые случайные величины, есть некоторые константы $a, b \in \mathbb{R}$, то

   $$\mathsf{E}[a\xi + b\eta] = a\,\mathsf{E}[\xi] + b\,\mathsf{E}[\eta].$$

   Доказательство. Введем обозначение $\delta = a\xi + b\eta$. Пусть $x_1, \dots, x_k$ и $y_1, \dots, y_m$ — все значения случайных величин ξ и η соответственно, а $z_1, \dots, z_n$ — все значения δ.

Тогда:

$$\mathsf{E}[a\xi + b\eta] = \mathsf{E}[\delta] = \sum_{i=1}^{n} z_i \, \mathsf{P}(\delta = z_i) = \sum_{i=1}^{n} z_i \, \mathsf{P}\left(\bigcup_{\substack{t,j:\\ ax_t+by_j=z_i}} \{\xi = x_t, \eta = y_j\}\right) =$$

$$= \sum_{i=1}^{n} z_i \sum_{\substack{t,j:\\ ax_t+by_j=z_i}} \mathsf{P}(\xi = x_t, \eta = y_j) =$$

$$= \sum_{i=1}^{n} \sum_{\substack{t,j:\\ ax_t+by_j=z_i}} (ax_t + by_j) \, \mathsf{P}(\xi = x_t, \eta = y_j) = \sum_{t=1}^{k}\sum_{j=1}^{m}(ax_t + by_j) \, \mathsf{P}(\xi = x_t, \eta = y_j) =$$

$$= \sum_{t=1}^{k}\sum_{j=1}^{m} ax_t \, \mathsf{P}(\xi = x_t, \eta = y_j) + \sum_{t=1}^{k}\sum_{j=1}^{m} by_j \, \mathsf{P}(\xi = x_t, \eta = y_j) =$$

$$= \sum_{t=1}^{k} ax_t \, \mathsf{P}(\xi = x_t) + \sum_{j=1}^{m} by_j \, \mathsf{P}(\eta = y_j) = a \, \mathsf{E}[\xi] + b \, \mathsf{E}[\eta].$$

□

2. Сохранение относительного порядка:

   Если ξ, η — простые случайные величины, тогда:

   (a) $\xi \geqslant 0 \implies \mathsf{E}[\xi] \geqslant 0$

   (b) $\xi \geqslant \eta \implies \mathsf{E}[\xi] \geqslant \mathsf{E}[\eta]$.

   Доказательство. Докажем свойства по порядку:

   (a) Если $x_1, \ldots, x_k$ — все значения $\xi \geqslant 0$, то $\forall i = 1, \ldots, k \; x_i \geqslant 0$. Значит, $\mathsf{E}[\xi] = \sum\limits_{i=1}^{k} x_i \, \mathsf{P}(\xi = x_i) \geqslant 0$.

   (b) Введем величину $\delta = \xi - \eta$. Так как, $\xi \geqslant \eta$, $\delta \geqslant 0$. Тогда по первому пункту $\mathsf{E}[\delta] \geqslant 0 = \mathsf{E}[\xi] - \mathsf{E}[\eta]$.

   □

3.

   Лемма. *Пусть* $\eta, \{\xi_n, n \in \mathbb{N}\}$ *— простые случайные величины, причем* $\xi_n \uparrow \xi$, $\xi_n \geqslant 0 \; \forall n \in \mathbb{N}$, $\eta \leqslant \xi$, $\eta \geqslant 0$. *Тогда*

$$\mathsf{E}[\eta] \leqslant \lim_{n} \mathsf{E}[\xi_n].$$

   Доказательство. Возьмем любое $\alpha > 0$ и положим

$$A_n = \{\omega : \xi_n(\omega) \geqslant \eta(\omega) - \alpha\}.$$

Раз $\xi_n \uparrow \xi$, то $A_n \uparrow \Omega$. По теореме о непрерывности вероятностной меры $\mathsf{P}(A_n) \to 1$. Рассмотрим математическое ожидание случайной величины ξ_n:

$$\mathsf{E}[\xi_n] = \mathsf{E}[\xi_n I_{A_n}] + \mathsf{E}\left[\xi_n I_{\overline{A_n}}\right] \geqslant \mathsf{E}[\xi_n I_{A_n}] \geqslant \mathsf{E}[(\eta - \alpha) I_{A_n}] = \mathsf{E}[\eta] - \mathsf{E}\left[\eta I_{\overline{A_n}}\right] - \mathsf{E}[\alpha I_{A_n}].$$

Введем обозначение: $c = \max\limits_{\omega \in \Omega} \eta(\omega)$, тогда

$$\mathsf{E}[\eta] - \mathsf{E}\left[\eta I_{\overline{A_n}}\right] - \mathsf{E}[\alpha I_{A_n}] \geqslant \mathsf{E}[\eta] - c\,\mathsf{E}\left[I_{\overline{A_n}}\right] - \alpha\,\mathsf{E}[I_{A_n}] \geqslant \mathsf{E}[\eta] - c\,\mathsf{P}\left(\overline{A_n}\right) - \alpha.$$

Раз для любого n, $\xi_n \leqslant \xi_{n+1}$, то $\mathsf{E}[\xi_n] \leqslant \mathsf{E}[\xi_{n+1}]$. Следовательно, существует $\lim\limits_n \mathsf{E}[\xi_n]$. Перейдем к пределу в неравенстве выше:

$$\lim_n \mathsf{E}[\xi_n] \geqslant \mathsf{E}[\eta] - \alpha.$$

В силу произвольности $\alpha > 0$, $\lim\limits_n \mathsf{E}[\xi_n] \geqslant \mathsf{E}[\eta]$. □

Следствие (Корректность определения матожидания.)**.** *Для неотрицательных случайных величин, матожидание определено корректно.*

Доказательство. Нужно показать, что предел не зависит от выбора последовательности простых случайных величин $0 \leqslant \xi_n \uparrow \xi$. Пусть есть другая такая последовательность, $0 \leqslant \eta \uparrow \xi$. Тогда по лемме, для любого натурального m $\lim\limits_n \mathsf{E}[\xi_n] \geqslant \mathsf{E}[\eta_m]$. Значит, $\lim\limits_n \mathsf{E}[\xi_n] \geqslant \lim\limits_m \mathsf{E}[\eta_m]$. С другой стороны, для любых натуральных n,

$$\lim_m \mathsf{E}[\eta_m] \geqslant \mathsf{E}[\xi_n] \implies \lim_m \mathsf{E}[\eta_m] \geqslant \lim_n \mathsf{E}[\xi_n].$$

Следовательно, $\lim\limits_m \mathsf{E}[\eta_m] = \lim\limits_n \mathsf{E}[\xi_n]$. □

1.10.2 Свойства математического ожидания в общем случае.

Пусть ξ — произвольная случайная величина. Введем обозначения: $\xi^+ = \max(\xi, 0)$ и $\xi^- = -\min(\xi, 0)$.

Определение 50. Говорят, что математическое ожидание $\mathsf{E}[\xi]$ случайной величины ξ *существует*, или *определено*, если по крайней мере одна из величин $\mathsf{E}[\xi^+]$ или $\mathsf{E}[\xi^-]$ конечна. В этом случае по *определению* полагают $\mathsf{E}[\xi] = \mathsf{E}[\xi^+] - \mathsf{E}[\xi^-]$. Есть четыре комбинации значений $\mathsf{E}[\xi^+]$ и $\mathsf{E}[\xi^-]$. Разберем их:

1. Если $\mathsf{E}[\xi^+]$ и $\mathsf{E}[\xi^-]$ — конечны, то $\mathsf{E}[\xi] = \mathsf{E}[\xi^+] - \mathsf{E}[\xi^-]$.
2. Если $\mathsf{E}[\xi^+] = \infty$, $\mathsf{E}[\xi^-]$ — конечно, то $\mathsf{E}[\xi] = +\infty$.
3. Если $\mathsf{E}[\xi^+]$ — конечно, $\mathsf{E}[\xi^-] = \infty$, то $\mathsf{E}[\xi] = -\infty$.
4. Если $\mathsf{E}[\xi^+] = \infty$ и $\mathsf{E}[\xi^-] = \infty$, то $\mathsf{E}[\xi]$ *не определено.*

Итак, перейдем к рассмотрению свойств математического ожидания случайных величин в общем случае. Но перед этим введем еще одно важное понятие, которое будет использовать для целой группы утверждений в следующей теореме.

Определение 51. Будем говорить, что некоторое свойство A выполнено P *−почти наверное*, если существует такое множество N с $\mathsf{P}(N) = 0$ такое, что это свойство A выполнено для каждой точки $\omega = \Omega \setminus N$. Также можно определить, что событие A происходит *почти наверное*, если $\mathsf{P}(A) = 1$. Обозначение: $A_{п.н.}$.

Теорема 25 (Свойства математического ожидания в общем случае.)**.** *Если* $\mathsf{E}[\xi]$ *— математическое ожидание случайной величины* ξ*, то выполняются следущие свойства:*

1. *Если* $c \in \mathbb{R}$, $\mathsf{E}[\xi]$ *— конечно, тогда*
$$\mathsf{E}[c\xi] = c\,\mathsf{E}[\xi].$$
2. *Пусть* $\xi \leqslant \eta$ *и* $\mathsf{E}[\xi], \mathsf{E}[\eta]$ *— конечны. Тогда*
$$\mathsf{E}[\xi] \leqslant \mathsf{E}[\eta].$$
3. *Если* $\mathsf{E}[\xi]$ *и* $\mathsf{E}[\eta]$ *— конечны, то* $\mathsf{E}[\xi + \eta]$ *— конечно и*
$$\mathsf{E}[\xi + \eta] = \mathsf{E}[\eta] + \mathsf{E}[\xi].$$
4. *Если* $\mathsf{E}[\xi]$ *— существует, тогда*
$$|\mathsf{E}[\xi]| \leqslant \mathsf{E}[|\xi|].$$
5. (a) *Пусть* $0 \leqslant \xi \leqslant \eta$ *и* $\mathsf{E}[\eta]$ *— конечно, тогда* $\mathsf{E}[\xi]$ *также конечно.*
   (b) *Пусть* $|\xi| \leqslant \eta$ *и* $\mathsf{E}[\eta]$ *– конечно, тогда* $\mathsf{E}[\xi]$ *также конечно.*
6. *Если* $\xi = 0_{п.н.}$*, то* $\mathsf{E}[\xi] = 0$.
7. *Если* $\xi = \eta_{п.н.}$*, и* $\mathsf{E}[\eta]$ *— конечно, то* $\mathsf{E}[\xi] = \mathsf{E}[\eta]$.
8. *Если* $\xi \geqslant 0$ *и* $\mathsf{E}[\xi] = 0$*, то* $\xi = 0_{п.н.}$.
9. *Пусть* $\forall A \in \mathcal{F}$, $\mathsf{E}[\xi I_A] \leqslant \mathsf{E}[\eta I_A]$ *и* $\mathsf{E}[\xi]$, $\mathsf{E}[\eta]$ *— конечны. Тогда* $\xi \leqslant \eta_{п.н.}$.

Доказательство. Докажем первую группу свойств:

1. Для простых случайных величин было доказано в первой части лекции. В общем же случае надо рассмотреть представление $\xi = \xi^+ - \xi^-$ и заметить, что для $c \geqslant 0$, $(c\xi)^+ = c\xi^+$, $(c\xi)^- = c\xi^-$, а для $c < 0$, $(c\xi)^+ = -c\xi^-$, $(c\xi)^- = -c\xi^+$.

2. Для простых случайных величин было доказано в первой части лекции. Рассмотрим общий случай. Пусть теперь $\mathsf{E}[\xi] > -\infty$, тогда $\mathsf{E}[\xi^-] < \infty$. Если $\xi \leqslant \eta$, то $\xi^+ \leqslant \eta^+$, $\xi^- \geqslant \eta^-$. Тогда $\mathsf{E}[\eta^-] \leqslant \mathsf{E}[\xi^-] < \infty$, следовательно, $\mathsf{E}[\eta]$ определено и $\mathsf{E}[\xi] = \mathsf{E}[\xi^+] - \mathsf{E}[\xi^-] \leqslant \mathsf{E}[\eta^+] - \mathsf{E}[\eta^-] = \mathsf{E}[\eta]$. Аналогичным образом рассматривается случай, когда $\mathsf{E}[\eta] < \infty$.

3. Случай с простыми случайными величинами разбирался в первой части лекции. В общем же случае, когда $\mathsf{E}[|\xi|] < \infty$, $\mathsf{E}[|\eta|] < \infty$, сводится к рассмотренному, если воспользоваться тем, что $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, $\xi^+ \leqslant |\xi|$, $\xi^- \leqslant |\xi|$, $\eta^+ \leqslant |\eta|$, $\eta^- \leqslant |\eta|$.

4. Поскольку $-|\xi| \leqslant \xi \leqslant |\xi|$, по первым двум свойствам $-\mathsf{E}[|\xi|] \leqslant \mathsf{E}[\xi] \leqslant \mathsf{E}[|\xi|]$.

Теперь докажем вторую группу свойств, связанных с понятием “P *−почти навеное*”:

6. В самом деле, если ξ – простая случайная величина, тогда она по определению разбивается в следующую сумму: $\xi = \sum x_k I_{A_k}(\omega)$ и $x_k \neq 0$. По условию, $\xi = 0_{п.н.}$, следовательно для любых событий $A_k = \{\xi = x_k\}$, $\mathsf{P}(A_k) = 0$, а значит, $\mathsf{E}[\xi] = 0$. Если же $\xi \geqslant 0$ — неотрицательная случайная величина, то для любой последовательности $\{\xi_n\}$, $0 \leqslant \xi_n \uparrow \xi$,
$$\mathsf{P}(\xi_n) \geqslant \mathsf{P}(\xi = 0) = 1 \implies \xi_n = 0_{п.н.}.$$
Значит,
$$\mathsf{E}[\xi_n] = 0, \forall n \implies \mathsf{E}[\xi] = \lim_n \mathsf{E}[\xi_n] = 0.$$
Если $\xi = 0_{п.н.}$, то
$$\xi^+ = 0_{п.н.},\ \xi^- = 0_{п.н.},\ \implies \mathsf{E}\big[\xi^+\big] = 0_{п.н.},\ \mathsf{E}\big[\xi^-\big] = 0_{п.н.} \implies \mathsf{E}[\xi] = \mathsf{E}\big[\xi^+\big] - \mathsf{E}\big[\xi^-\big] = 0.$$

7. В самом деле, пусть $N = \{\omega : \xi \neq \eta\}$. Тогда $\mathsf{P}(N) = 0$ и $\xi = \xi I_N + \xi I_{\overline{N}}$, $\eta = \eta I_N + \eta I_{\overline{N}} = \eta I_N + \xi I_{\overline{N}}$. По свойствам (3) и (6) $\mathsf{E}[\xi] = \mathsf{E}[\xi I_N] + \mathsf{E}[\xi I_{\overline{N}}] = \mathsf{E}[\xi I_N] = \mathsf{E}[\eta I_{\overline{N}}]$. Но $\mathsf{E}[\eta I_N] = 0$, поэтому по свойству (3) $\mathsf{E}[\xi] = \mathsf{E}[\eta I_{\overline{N}}] + \mathsf{E}[\eta I_N] = \mathsf{E}[\eta]$.

8. Пусть ξ — простая случайная величина со значениями $x_1, \dots, x_k$. Тогда $x_i \geqslant 0$ и $\mathsf{E}[\xi] = \sum\limits_{i=1}^{k} x_i \mathsf{P}(\xi = x_i) = 0$. Следовательно, для любого i, такого что $x_i \neq 0$ выполнено $\mathsf{P}(\xi = x_i) = 0 \implies \mathsf{P}(\xi = 0) = 1$. Если $\xi \geqslant 0$, то $\forall n \in \mathbb{N}$ рассмотрим $A_n = \{\xi \geqslant \frac{1}{n}\}$.
$$\mathsf{P}(A_n) = \mathsf{E}[I_{A_n}] \leqslant \mathsf{E}[n\xi I_{A_n}] \leqslant \mathsf{E}[n\xi] \leqslant n\,\mathsf{E}[\xi] = 0.$$
Но $A_n \downarrow \{\xi > 0\} \implies \mathsf{P}(\xi > 0) = \lim_n \mathsf{P}(A_n) = 0$.

9. В самом деле, пусть $B = \{\omega : \xi(\omega) > \eta(\omega)\}$. Тогда $\mathsf{E}[\eta I_B] \leqslant \mathsf{E}[\xi I_M] \leqslant \mathsf{E}[\eta I_B]$ и, значит, $\mathsf{E}[\xi I_B] = \mathsf{E}[\eta I_b]$. В силу свойства аддитивности (3) $\mathsf{E}[(\xi - \eta) I_B] = 0$ и по свойству (8) $(\xi - \eta) I_B = 0_{п.н.}$, откуда $\mathsf{P}(B) = 0$.

$\square$

1.11 Лекция от 22.11.2016

1.11.1 Классификация случайных величин. Формулы подсчёта математического ожидания

На предыдущих лекциях мы ввели понятие математического ожидания. Но мы пока что не знаем, как его считать. Разберёмся с этим.

Перед этим введём пару важных определений:

Определение 52. Пусть ξ — случайная величина на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$. Тогда *функцией распределения* случайной величины ξ называют функцию $F_\xi : \mathbb{R} \to [0, 1]$, определяемую следующим образом:
$$F_\xi(x) = \mathsf{P}(\xi \leqslant x).$$

Определение 53. Пусть ξ — случайная величина на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$. Тогда *распределением* случайной величины ξ называется вероятностная мера P_ξ на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ такая, что для любого $B \in \mathcal{B}(\mathbb{R})$

$$\mathsf{P}_\xi(B) = \mathsf{P}(\xi \in B).$$

Примечание. P_ξ действительно является вероятностной мерой, а F_ξ действительно является функцией распределения.

Теперь введём следующую класификацию случайных величин по типу функции распределения:

1. Если F_ξ дискретна, то будем называть случайную величину ξ *дискретной*.
2. Если F_ξ абсолютно непрерывна, то будем называть случайную величину ξ *абсолютно непрерывной*. В данном случае можно ввести понятие *плотности распределения*:

   Определение 54. Пусть ξ — асболютно непрерывная случайная величина, и существует неотрицательная функция $p_\xi(x)$ такая, что для любого $x \in \mathbb{R}$ выполнено, что
   $$F_\xi(x) = \mathsf{P}(\xi \leqslant x) = \int\limits_{-\infty}^{x} p_\xi(t)\,\mathrm{d}t.$$
   Тогда $p_\xi(x)$ называют *плотностью распределения* случайной величины ξ.
3. Если же $F_\xi(x)$ сингулярна, то ξ называют *сингулярной*.

Начнём разбираться с математическим ожиданием с дискретных случайных величин. Вообще, если ξ дискретна, то множество значений $X = \xi(\Omega)$ не более, чем счётно, причем $\mathsf{P}_\xi(X) = 1$.

Теорема 26. *Пусть ξ — дискретная случайная величина, а $f(x)$ — некоторая борелевская функция, определённая на множестве значений X. Тогда*

$$\mathsf{E}[f(\xi)] = \sum_{x \in X} f(x)\,\mathsf{P}(\xi = x).$$

Доказательство. Для начала предположим, что X конечно. Но в этом случае ξ — это простая случайная величина. Следовательно, $f(\xi)$ — это тоже простая случайная величина. Тогда по определению математического ожидания для простой величины:

$$\mathsf{E}[f(\xi)] = \sum_{z \in f(X)} f(z)\,\mathsf{P}(f(\xi) = z) = \sum_{z \in f(X)} \left(\sum_{\substack{x \in X \\ f(x) = z}} z\,\mathsf{P}(\xi = x) \right)$$
$$= \sum_{z \in f(X)} \left(\sum_{\substack{x \in X \\ f(x) = z}} f(x)\,\mathsf{P}(\xi = x) \right) = \sum_{x \in X} f(x)\,\mathsf{P}(\xi = x).$$

Теперь рассмотрим случай счётного X. Каким-либо образом занумеруем элементы X: $X = \{x_n\}_{n=1}^{\infty}$.

Предположим, что $f(x)$ неотрицательна. Тогда введём последовательность случайных величин $\{\eta_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\eta_n = \sum_{i=1}^{n} f(x_i) I\{\xi = x_i\}.$$

Тогда $\eta_n \uparrow f(\xi)$. Следовательно, $\mathsf{E}[\eta_n]$ монотонно возрастает и сходится к $\mathsf{E}[f(\xi)]$. Однако, в силу линейности математического ожидания простой случайной величины получаем, что

$$\begin{aligned}\mathsf{E}[f(\xi)] = \lim_{n\to\infty} \mathsf{E}[\eta_n] &= \lim_{n\to\infty} \left(\sum_{i=1}^{n} f(x_i)\, \mathsf{E}[I\{\xi = x_i\}] \right) \\ &= \lim_{n\to\infty} \left(\sum_{i=1}^{n} f(x_i)\, \mathsf{P}(\xi = x_i) \right) = \sum_{i=1}^{\infty} f(x)\, \mathsf{P}(\xi = x_i).\end{aligned}$$

В общем случае же представим $f(x)$ в виде $f^+(x) - f^-(x)$ и пользуемся доказанным ранее. $\square$

Теперь приступим к абсолютно непрерывным случайным величинам.

Теорема 27. *Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p_\xi(x)$. Тогда*

$$\mathsf{E}[\xi] = \int_{-\infty}^{+\infty} x p_\xi(x)\, \mathrm{d}x.$$

Доказательство. Для начала рассмотрим неотрицательную случайнуй величину ξ. Введём последовательность простых случайных величин $\{\xi_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\xi_n = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n}\, I\left\{ \frac{k-1}{2^n} < \xi \leqslant \frac{k}{2^n} \right\}.$$

Заметим, что данная последовательность будет сходиться к ξ снизу: $\xi_n \uparrow \xi$. Однако

$$\begin{aligned}\mathsf{E}[\xi_n] &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n}\, \mathsf{P}\left(\frac{k-1}{2^n} < \xi \leqslant \frac{k}{2^n} \right) \\ &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \left(F_\xi\left(\frac{k}{2^n}\right) - F_\xi\left(\frac{k-1}{2^n}\right) \right) = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} p_\xi(x)\, \mathrm{d}x\end{aligned}$$

Далее заметим, что

$$\int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \left(x - \frac{1}{2^n} \right) p_\xi(x)\, \mathrm{d}x \leqslant \frac{k-1}{2^n} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} p_\xi(x)\, \mathrm{d}x \leqslant \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} x p_\xi(x)\, \mathrm{d}x.$$

Тогда, суммируя все такие элементы и вынося константу в ограничени снизу, получаем, что

$$\int_0^n x p_\xi(x)\, \mathrm{d}x - \frac{1}{2^n} \leqslant \int_0^n \left(x - \frac{1}{2^n} \right) p_\xi(x)\, \mathrm{d}x \leqslant \mathsf{E}[\xi_n] \leqslant \int_0^n x p_\xi(x)\, \mathrm{d}x \leqslant \int_0^\infty x p_\xi(x)\, \mathrm{d}x.$$

Отсюда получаем, что

$$\mathsf{E}[\xi] = \int\limits_0^\infty x p_\xi(x)\,\mathrm{d}x.$$

Если же случайная величина ξ неположительна, то аналогичными рассуждениями получаем, что

$$\mathsf{E}[\xi] = \int\limits_{-\infty}^0 x p_\xi(x)\,\mathrm{d}x.$$

В общем случае представим ξ в виде суммы неотрицательной и неположительной случайных величин. Тогда получим желаемое. □

Хорошо, мы научились считать матожидание абсолютно непрерывной случайной величины. Но хотелось бы научиться считать матожидание от функции от случайной величины. Следующая теорема говорит нам, как его вычислять.

Теорема 28. *Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p(x)$ и $f(x)$ — некоторая борелевская функция. Тогда*

$$\mathsf{E}[f(x)] = \int\limits_{-\infty}^{+\infty} f(x) p_\xi(x)\,\mathrm{d}x.$$

Набросок доказательства. В случае монотонной функции $f(x)$ рассуждение работает. Теперь предположим более слабое условие — неотрицательность $f(x)$. Опять же, построим последовательность простых случайных величин $\{\eta_n\}_{n=1}^\infty$, устроенную следующим образом:

$$\sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} < f(\xi) \leqslant \frac{k}{2^n}\right\}.$$

Тогда $\eta_n \uparrow f(\xi)$ и $\mathsf{E}[\eta_n]$ монотонно сходится к $\mathsf{E}[f(\xi)]$. Однако

$$\mathsf{E}[\eta_n] = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \mathsf{P}\left(\frac{k-1}{2^n} < f(\xi) \leqslant \frac{k}{2^n}\right) = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \mathsf{P}\left(\xi \in f^{-1}\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]\right)$$

Обозначим $f^{-1}\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$ за $B_{k,n}$. Тогда

$$\mathsf{E}[\eta_n] = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \int\limits_{B_{k,n}} p_\xi(x)\,\mathrm{d}x.$$

Дальше рассуждаем в стиле того, как мы доказывали формулу для $\mathsf{E}[\xi]$. □

В принципе, для вычисления математического ожидания функции от случайной величины ξ достаточно знать её функцию распределения.

Теорема 29. [15] *Пусть ξ — это случайная величина на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ с функцией распределения F_ξ, а $f(x)$ — некая борелевская функция. Тогда математическое ожидание равно*

$$\mathsf{E}[f(\xi)] = \int\limits_{-\infty}^{+\infty} f(x)\,\mathrm{d}F_\xi(x).$$

Рассмотрим какой-нибудь пример.

Определение 55. *Гамма-функцией* $\Gamma(\lambda)$ $(\lambda > 0)$ называют следующую функцию

$$\Gamma(\lambda) = \int\limits_{0}^{+\infty} x^{\lambda-1} e^{-x}\,\mathrm{d}x.$$

У гамма-функции есть несколько простых свойств:

- $\Gamma(n) = (n-1)!$ для любого $n \in \mathbb{N}$,
- $\Gamma(\lambda+1) = \lambda\Gamma(\lambda)$ для любого $\lambda > 0$,
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ (это не такое простое, ладно).

Определение 56. Пусть ξ — некоторая случайная величина, про которую известно, что

$$p_\xi(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \text{ при } x \geqslant 0, \alpha, \beta > 0.$$

В таком случае говорят, что ξ подчиняется *гамма-распределению* с параметрами α и β. Обозначается $\xi \sim \mathrm{Gamma}(\alpha, \beta)$.

Пусть у нас есть случайная величина $\xi \sim \mathrm{Gamma}(\alpha, \beta)$. Найдём $\mathsf{E}[\xi]$:

$$\mathsf{E}[\xi] = \int\limits_{0}^{+\infty} \frac{\beta^\alpha x^\alpha e^{-\beta x}}{\Gamma(\alpha)}\,\mathrm{d}x = \left\{ u = x\beta, \mathrm{d}x = \frac{\mathrm{d}u}{\beta} \right\} = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} = \frac{\alpha}{\beta}.$$

1.11.2 Независимость случайных величин и векторов

Определение 57. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми в совокупности*, если для любых $x_1, x_2, \ldots, x_n \in \mathbb{R}$

$$\mathsf{P}(\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n) = \prod_{i=1}^{n} \mathsf{P}(\xi_i \leqslant x_i).$$

Примечание. Если $\xi_1.\xi_2, \ldots, \xi_n$ независимы в совокупности, то любой поднабор тоже независим в совокупности.

Теперь возникает вопрос: а можно ли заменить знак $\leqslant$ на что-либо другое? Да.

[15] Это вовсе не теорема, а определение математического ожидания. И здесь не интеграл Римана, а интеграл Лебега-Стильтеса. Попытки избежать теорию меры и функциональный анализ проваливаются с треском. (А.Х.)

Теорема 30. *Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы тогда и только тогда, когда для любых множеств $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ выполнено, что*

$$\mathsf{P}(\xi_1 \in B_1, \xi_2 \in B_2, \ldots, \xi_n \in B_n) = \prod_{i=1}^{n} \mathsf{P}(\xi_i \in B_i).$$

Пользуясь этой «теоремой», докажем один полезный факт.

Лемма. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, а $f_1, f_2, \ldots, f_n$ — некоторые борелевские функции. Тогда случайные величины $f_1(\xi_1), f_2(\xi_2), \ldots, f_n(\xi_n)$ тоже независимы.*

Доказательство. Проверим то, что выполнено определение:

$$\mathsf{P}(f_1(\xi_1) \leqslant x_1, \ldots, f_n(\xi_n) \leqslant x_n) = \mathsf{P}\big(\xi_1 \in f_1^{-1}(-\infty, x_1], \ldots, \xi_n \in f_n^{-1}(-\infty, x_n]\big)$$

Заметим, что $f_i^{-1}(-\infty, x_i] \in \mathcal{B}(\mathbb{R})$. Тогда

$$\mathsf{P}\big(\xi_1 \in f_1^{-1}(-\infty, x_1], \ldots, \xi_n \in f_n^{-1}(-\infty, x_n]\big) = \prod_{i=1}^{n} \mathsf{P}\big(\xi_i \in f_i^{-1}(-\infty, x_i]\big).$$

Отсюда получаем, что

$$\mathsf{P}(f_1(\xi_1) \leqslant x_1, \ldots, f_n(\xi_n) \leqslant x_n) = \prod_{i=1}^{n} \mathsf{P}(f_i(\xi_i) \leqslant x_i).$$ □

1.11.3 Математическое ожидание произведения независимых случайных величин

В дискретном вероятностном пространстве было верно следующее утверждение: если случайные величины независимы, то матожидание произведения равно произведению матожиданий. Верно ли оно в общем случае? Докажем, что да.

Теорема 31. *Пусть случайные величины ξ и η независимы, а их матожидания конечны. Тогда матожидание случайной величины $\xi\eta$ тоже конечно и*

$$\mathsf{E}[\xi\eta] = \mathsf{E}[\xi]\,\mathsf{E}[\eta].$$

Доказательство. Для начала рассмотрим случай простых случайных величин. Тогда их можно представить в следующем виде:

$$\xi = \sum_{i=1}^{n} x_i I\{\xi = x_i\}$$
$$\eta = \sum_{j=1}^{m} y_j I\{\eta = y_j\}$$

Теперь распишем матожидание произведения:

$$\begin{aligned}
\mathsf{E}[\xi\eta] &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j\,\mathsf{E}[I\{\xi = x_i\} I\{\eta = y_j\}] = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j\,\mathsf{P}(\xi = x_i, \eta = y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j\,\mathsf{P}(\xi = x_i)\,\mathsf{P}(\eta = y_j) = \left(\sum_{i=1}^{n} x_i\,\mathsf{P}(\xi = x_i)\right)\left(\sum_{j=1}^{m} y_j\,\mathsf{P}(\eta = y_j)\right) \\
&= \mathsf{E}[\xi]\,\mathsf{E}[\eta].
\end{aligned}$$

Теперь перейдём к случаю неотрицательных случайных величин. Построим последовательности случайных величин $\{\xi_n\}_{n=1}^{\infty}$ и $\{\eta_n\}_{n=1}^{\infty}$ такие, что $\xi_n \uparrow \xi$ и $\eta_n \uparrow \eta$. Тогда $\xi_n\eta_n \uparrow \xi\eta$ и по определению

$$\mathsf{E}[\xi\eta] = \lim_{n\to\infty} \mathsf{E}[\xi_n\eta_n].$$

Но нужно следать так, чтобы $\mathsf{E}[\xi_n\eta_n] = \mathsf{E}[\xi_n]\,\mathsf{E}[\eta_n]$. Для этого скажем, что ξ_n и η_n есть какие-то функции от ξ и η соответственно. Тогда ξ_n и η_n независимы. Следовательно,

$$\lim_{n\to\infty} \mathsf{E}[\xi_n\eta_n] = \lim_{n\to\infty} \mathsf{E}[\xi_n]\,\mathsf{E}[\eta_n] = \mathsf{E}[\xi]\,\mathsf{E}[\eta].$$

В общем же случае разобъём обе случайные величины следующим образом:

$$\begin{aligned} \xi &= \xi^+ - \xi^-, \\ \eta &= \eta^+ - \eta^-. \end{aligned}$$

Тогда $\xi\eta = (\xi\eta)^+ - (\xi\eta)^-$, где

$$\begin{aligned} (\xi\eta)^+ &= \xi^+\eta^+ + \xi^-\eta^-, \\ (\xi\eta)^- &= \xi^+\eta^- + \xi^-\eta^+. \end{aligned}$$

Заметим, что $\xi^{\pm}$ независимо от $\eta^{\pm}$, так как это функции от независимых случайных величин. Тогда, пользуясь формулой для неотрицательных случайных величин, получаем, что

$$\begin{aligned} \mathsf{E}\left[(\xi\eta)^+\right] &= \mathsf{E}\left[\xi^+\right]\mathsf{E}\left[\eta^+\right] + \mathsf{E}\left[\xi^-\right]\mathsf{E}\left[\eta^-\right] \\ \mathsf{E}\left[(\xi\eta)^-\right] &= \mathsf{E}\left[\xi^+\right]\mathsf{E}\left[\eta^-\right] + \mathsf{E}\left[\xi^-\right]\mathsf{E}\left[\eta^+\right] \\ \mathsf{E}[\xi\eta] &= \left(\mathsf{E}\left[\xi^+\right] - \mathsf{E}\left[\xi^-\right]\right)\left(\mathsf{E}\left[\eta^+\right] - \mathsf{E}\left[\eta^-\right]\right) = \mathsf{E}[\xi]\,\mathsf{E}[\eta]. \end{aligned}$$

□

1.12 Лекция от 29.11.2016

1.12.1 Многомерный случай

> Многомерный случай — это одномерный случай.
>
> Д.А. Шабанов

Ранее мы рассматривали только одну случайную величину за раз. Но порой возникают ситуации, когда их нужно рассматривать по несколько одновременно. Введём необходимую теорию для этого.

Определение 58. Пусть P — это некоторая вероятностная мера на $\mathbb{R}^n$ $(n > 1)$. Тогда назовём *функцией распределения* $F : \mathbb{R}^n \to [0,1]$ такую, что для любых $x_1, x_2, \ldots, x_n \in \mathbb{R}$ верно следующее:

$$F(x_1, x_2, \ldots, x_n) = \mathsf{P}((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]).$$

Примечание. Для удобства это определение можно записать следующим образом (в данном случае $\vec{x} = (x_1, x_2, \ldots, x_n)$):

$$F(\vec{x}) = \mathsf{P}((-\infty, \vec{x}]).$$

Вообще, запись $(-\infty, \vec{x}]$ не несёт особого смысла, и её стоит понимать, как сокращение для $(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$.

Как и в одномерном случае, фукнция распределения обладает несколькими свойствами. Непрервыность справа и предельные значения легко обобщить, а как быть с неубыванием? Для этого это свойство несколько изменяют. В итоге получается следующая лемма.

Лемма (Основные свойства функции распределения)**.** *Функция распределения в многомерном случае обладает следующими свойствами:*

1. *Пусть есть некоторая точка $\vec{x}$ и последовательность точек $\{\vec{x}^{(m)}\}_{m=1}^{\infty}$, про которую известно, что $\vec{x}^{(m)} \downarrow \vec{x}$ и $x_i^{(m+1)} \leqslant x_i^{(m)}$ для всех $i \in \{1, 2, \ldots, n\}$. Тогда*

$$\lim_{m\to\infty} F(\vec{x}^{(m)}) = F(\vec{x}).$$

Данное свойство выступает аналогом непрерывности справа в одномерном случае.

2.

$$\lim_{\substack{x_1 \to +\infty \\ \ldots \\ x_n \to +\infty}} F(x_1, x_2, \ldots, x_n) = 1.$$

В данном случае этот предел множно понимать, как повторный.

3. *Для любого $i \in \{1, 2, \ldots, n\}$ выполнено*

$$\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_n) = 0.$$

4. *Введём оператор $\Delta_{a,b}^i F(x_1, \ldots, x_n)$, равный*

$$F(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n).$$

Тогда для любого набора чисел $a_1, \ldots, a_n, b_1, \ldots, b_n$ такого, что $a_i \leqslant b_i$ для любого $i \in \{1, 2, \ldots, n\}$, выполнено

$$\Delta_{a_1,b_1}^1 \Delta_{a_2,b_2}^2 \ldots \Delta_{a_n,b_n}^n F(x_1, \ldots, x_n) \geqslant 0.$$

Данное свойство обобщает неубывание на многомерное пространство.

Доказательство. В принципе, доказательство первых трёх пунктов идейно ничем не отличается от одномерного случая. Но всё равно докажем их.

1. Если $\{\vec{x}^{(m)}\}_{m=1}^{\infty}$ монотонно сходится к $\vec{x}$ сверху, то (все сходимости монотонные):

$$\begin{Bmatrix} \left(-\infty, x_1^{(m)}\right] \downarrow (-\infty, x_1] \\ \ldots \\ \left(-\infty, x_n^{(m)}\right] \downarrow (-\infty, x_n] \end{Bmatrix} \implies \left(-\infty, \vec{x}^{(m)}\right] \downarrow (-\infty, \vec{x}]\,.$$

Следовательно, пользуясь непрерывностью меры, получаем, что

$$\lim_{m\to\infty} F(\vec{x}^{(m)}) = \lim_{m\to\infty} \mathsf{P}\left(\left(-\infty, \vec{x}^{(m)}\right]\right) = \mathsf{P}((-\infty, \vec{x}]) = F(\vec{x}).$$

2. Для начала заметим следующее (все сходимости монотонные):

$$\left.\begin{cases} (-\infty, x_1] \xrightarrow[x_1 \to +\infty]{} (-\infty, +\infty) \\ \dots \\ (-\infty, x_n] \xrightarrow[x_n \to +\infty]{} (-\infty, +\infty) \end{cases}\right\} \implies (-\infty, \vec{x}] \to \mathbb{R}^n.$$

Тогда по непрерывности меры получаем, что

$$\lim_{\substack{x_1 \to +\infty \\ \dots \\ x_n \to +\infty}} F(x_1, x_2, \dots, x_n) = \mathsf{P}(\mathbb{R}^n) = 1.$$

3. Если $x_i \to -\infty$, то $(-\infty, x_i] \downarrow \varnothing$. Тогда

$$(-\infty, \vec{x}] \text{ сходится к } (-\infty, x_1] \times \dots \times \varnothing \times \dots \times (-\infty, x_n] = \varnothing.$$

Тогда по непрерывности меры получаем, что

$$\lim_{x_i \to -\infty} F(x_1, x_2, \dots, x_n) = \mathsf{P}(\varnothing) = 0.$$

4. Следующее свойство достаточно муторно доказывать в общем случае, так что разберём его для двумерного пространства — идея доказательства та же. Распишем это, пользуясь определением оператора $\Delta^i_{a,b}$:

$$\begin{aligned} \Delta^1_{a_1,b_1}\Delta^2_{a_2,b_2}F(x_1,x_2) &= \Delta^1_{a_1,b_1}\left(F(x_1,b_2) - F(x_1,a_2)\right) \\ &= F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2). \end{aligned}$$

Теперь воспользуемся определением функции распределения[16]:

$$\begin{aligned} F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) = \mathsf{P}((-\infty,b_1] \times (-\infty,b_2]) - \\ - \mathsf{P}((-\infty,a_1] \times (-\infty,b_2]) - \mathsf{P}((-\infty,b_1] \times (-\infty,a_2]) + \mathsf{P}((-\infty,a_1] \times (-\infty,a_2]) = \\ = \mathsf{P}((a_1,b_1] \times (-\infty,b_2]) - \mathsf{P}((a_1,b_1] \times (-\infty,a_2]) = \mathsf{P}((a_1,b_1] \times (a_2,b_2]) \geqslant 0. \end{aligned}$$

Можно заметить, что эта сумма есть ни что иное, как формула включений-исключений. Тогда в общем случае

$$\Delta^1_{a_1,b_1}\Delta^2_{a_2,b_2}\dots\Delta^n_{a_n,b_n}F(x_1,\dots,x_n) = \mathsf{P}((a_1,b_1] \times \dots \times (a_n,b_n]) \geqslant 0. \qquad \square$$

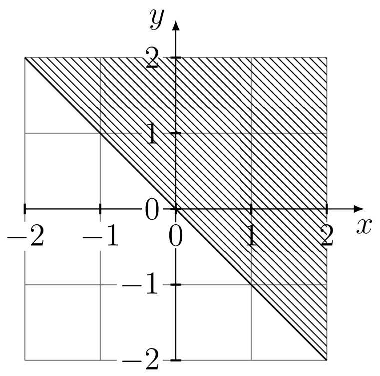
Последнее свойство на первый взгляд кажется странным. Возникает вопрос: а нельзя ли его заменить на неубывание по всем координатам? Увы, но нет.

Пример. Рассмотрим следующую функцию:

$$F(x,y) = \begin{cases} 1, & x + y \geqslant 0 \\ 0, & \text{иначе} \end{cases}$$

Изобразим её следующим образом: в заштрихованной области значение функции равно 1, в незаштрихованной — 0.

[16]На это выражение можно посмотреть и графически, но это роскошь двумерного случая.



Данная функция удовлетворяет свойствам 1-3 и не убывает по каждой координате. Проверим, выполняется ли свойство 4. Рассмотрим следующее значение:

$$\Delta^1_{-1,1}\Delta^2_{-1,1}F(x_1,x_2) = F(1,1) - F(-1,1) - F(1,-1) + F(-1,-1) = -1 < 0.$$

Как видим, оно не выполняется.

В одномерном случае между функцией распределения и вероятностной мерой есть биекция. Следующая теорема утверждает, что это верно и в многомерном случае:

Теорема 32. *Пусть $F(x) : \mathbb{R}^n \to [0,1]$ — некоторая функция, удовлетворяющая основным свойствам функции распределения. Тогда существует единственная вероятностная мера P на $\mathbb{R}^n$ такая, что F — это её функция распределения. Другими словами, для любого $\vec{x} \in \mathbb{R}^n$*

$$\mathsf{P}((-\infty, \vec{x}]) = F(\vec{x}).$$

Теорема хорошая, доказывать мы её, конечно, не будем.

Рассмотрим несколько примеров многомерных распределений.

Самый простое, что может быть — это произведение функций распределения для одномерных случаев. Другими словами, функция распределения определяется следующим образом: пусть есть n одномерных функций распределения $F_1, F_2, \ldots, F_n$. Тогда

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i).$$

Проверим, что это действительно будет функцией распределения. Легко проверить, что свойства $1-3$ выполнены. Теперь проверим последнее свойство. Заметим, что

$$\Delta^1_{a_1,b_1}\Delta^2_{a_2,b_2}\ldots\Delta^n_{a_n,b_n}F(x_1,\ldots,x_n) = \prod_{i=1}^{n} \Delta^1_{a_i,b_i}F_i(x_i) \geqslant 0.$$

Теперь рассмотрим несколько другой случай. Пусть есть неотрицательная функция $p(x_1, x_2, \ldots, x_n)$ такая, что

$$\int\limits_{\mathbb{R}^n} p(t_1, t_2, \ldots, t_n)\,\mathrm{d}t_1\ldots\mathrm{d}t_n = 1.$$

Тогда введём следующую функцию:

$$F(x_1, x_2, \ldots, x_n) = \int\limits_{-\infty}^{x_1}\cdots\int\limits_{-\infty}^{x_n} p(t_1, t_2, \ldots, t_n)\,\mathrm{d}t_1\ldots\mathrm{d}t_n$$

Покажем, что это функция распределения. Первые три свойства легко проверяются. Теперь проверим четвёртое:

$$\Delta^1_{a_1,b_1}\Delta^2_{a_2,b_2}\ldots\Delta^n_{a_n,b_n}F(x_1,\ldots,x_n)=\int\limits_{a_1}^{b_1}\cdots\int\limits_{a_n}^{b_n}p(t_1,t_2,\ldots,t_n)\,\mathrm{d}t_1\ldots\mathrm{d}t_n\geqslant 0.$$

В этом случае $p(x_1,x_2,\ldots,x_n)$ принято называть *плотностью* функции распределения F.

Допустим, что есть некоторая функция распределения $F(x_1,x_2,\ldots,x_n)$. В каком случае она имеет плотность? Ответ даёт теорема Радона-Никодима[17]: функция должна быть абсолютно непрерывной. Что это значит?

Определение 59. Вероятностная мера P называется *абсолютно непрерывной*, если для любого множества $B\in\mathcal{B}(\mathbb{R}^n)$ с лебеговой мерой 0 выполнено $\mathsf{P}(B)=0$.

Если функция F (а значит, и вероятностная мера P) абсолютно непрерывна, то плотность можно получить следующим образом:

$$p(x_1,x_2,\ldots,x_n)=\frac{\partial^n}{\partial x_1\partial x_2\ldots\partial x_n}F(x_1,\ldots,x_n).$$

1.12.2 Случайные величины в многомерном случае

Для одномерного случая были введены функции распределения и распределения случайных величин. Обобщим это.

Определение 60. Пусть $\xi_1,\xi_2,\ldots,\xi_n$ — некоторые случайные величины. Тогда их *совместной функцией распределения* называется следующая функция

$$F_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n)=\mathsf{P}(\xi_1\leqslant x_1,\ldots,\xi_n\leqslant x_n).$$

Определение 61. *Совместным распределением* случайных величин $\xi_1,\xi_2,\ldots,\xi_n$ называется вероятностная мера $\mathsf{P}_{\xi_1,\ldots,\xi_n}$ на $\mathbb{R}^n$, заданная следующим правилом: для любого множества $B\in\mathcal{B}(\mathbb{R}^n)$

$$\mathsf{P}_{\xi_1,\ldots,\xi_n}(B)=\mathsf{P}((\xi_1,\ldots,\xi_n)\in B).$$

Заметим следующее:

1. $F_{\xi_1,\ldots,\xi_n}$ — это настоящая функция распределения.
2. $\mathsf{P}_{\xi_1,\ldots,\xi_n}$ — это настоящая вероятностная мера.
3. $F_{\xi_1,\ldots,\xi_n}$ — это функция распределения для $\mathsf{P}_{\xi_1,\ldots,\xi_n}$.

В каких случаях совместная функция распределения хорошо считается? Например, в случае независимых случайных величин.

Теорема 33. *Если случайные величины $\xi_1,\xi_2,\ldots,\xi_n$ независимы (в совокупности) и функция распределения ξ_i равна F_{ξ_i} для всех $i\in\{1,2,\ldots,n\}$, то*

$$F_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n)=\prod_{i=1}^n F_{\xi_i}(x_i).$$

[17] Эта теорема гласит о том, что если функция абсолютно непрерывна, то она представима в нужном нам виде. За доказательством обращайтесь к учебнику по функциональному анализу.

Доказательство. Распишем совместную функцию распределения по определению:

$$F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \mathsf{P}(\xi_1 \leqslant x_1,\dots,\xi_n \leqslant x_n) = \prod_{i=1}^{n} \mathsf{P}(\xi_i \leqslant x_i) = \prod_{i=1}^{n} F_{\xi_i}(x_i). \qquad \square$$

Теперь, по аналогии со вторым примером функции распределения, введём плотность совместного распределения.

Определение 62. Пусть есть случайные величины $\xi_1,\dots,\xi_n$ с функцией совместного распределения $F_{\xi_1,\dots,\xi_n}$. Если у $F_{\xi_1,\dots,\xi_n}$ есть плотность $p_{\xi_1,\dots,\xi_n}$, то $p_{\xi_1,\dots,\xi_n}$ называют *совместной плотностью распределения.* В данном случае

$$F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} p_{\xi_1,\dots,\xi_n}(t_1,t_2,\dots,t_n)\,\mathrm{d}t_1\dots\mathrm{d}t_n.$$

Теперь докажем один весьма полезный факт.

Теорема 34. *Если у случайных величин $\xi_1,\xi_2,\dots,\xi_n$ есть совместная плотность, то в каждой из случайных величин есть плотность.*

Доказательство. Вспомним определение плотности для одномерной величины: если $p_\xi(x)$ — плотность, то для любого $x \in \mathbb{R}$ выполнено

$$\mathsf{P}(\xi \leqslant x) = \int\limits_{-\infty}^{x} p_\xi(t)\,\mathrm{d}t.$$

Попробуем получить нечто подобное, используя совместную плотность. По определениям совместной плотности и совместной функции распределения:

$$\mathsf{P}(\xi_1 \leqslant x_1,\dots,\xi_n \leqslant x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} p_{\xi_1,\dots,\xi_n}(t_1,t_2,\dots,t_n)\,\mathrm{d}t_1\dots\mathrm{d}t_n.$$

Допустим, что мы хотим получить плотность ξ_i. Тогда устремим все x_k, $k \in \{1,2,\dots,n\} \setminus \{i\}$ к бесконечности. Тогда все условия, кроме i-го будут гарантированно выполнены. Изменяя порядок интегрирования (так как плотность абсолютно непрерывна, то это легально), получаем, что

$$\mathsf{P}(\xi_i \leqslant x_i) = \int\limits_{-\infty}^{x_i} \left(\int\limits_{-\infty}^{+\infty} \cdots \int\limits_{-\infty}^{+\infty} p_{\xi_1,\dots,\xi_n}(t_1,t_2,\dots,t_n)\,\mathrm{d}t_1\dots\mathrm{d}t_n \right) \mathrm{d}t_i.$$

Сравнивая это с определением плотности в одномерном случае, получаем, что

$$p_{\xi_i}(x_i) = \int\limits_{-\infty}^{+\infty} \cdots \int\limits_{-\infty}^{+\infty} p_{\xi_1,\dots,\xi_n}(t_1,t_2,\dots,t_n)\,\mathrm{d}t_1\dots\mathrm{d}t_{i-1}\,\mathrm{d}t_{i+1}\dots\mathrm{d}t_n. \qquad \square$$

А что можно сказать про плотности независимых случайных величин? Ответ на этот вопрос даёт следующая теорема.

Теорема 35. *Пусть есть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ с совместной плотностью распределения $p_{\xi_1,\ldots,\xi_n}$, и для каждой из них существует плотность p_{ξ_i}. Тогда эти случайные величины независимы в совокупности тогда и только тогда, когда*

$$p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{\xi_i}(x_i).$$

Доказательство.

[⇒] Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы. Тогда

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{\xi_i}(x_i) = \prod_{i=1}^{n} \left(\int\limits_{-\infty}^{x_i} p_\xi(t_i)\, \mathrm{d}t_i \right).$$

Заметим, что произведение интегралов можно представить в следующем виде:

$$\prod_{i=1}^{n} \left(\int\limits_{-\infty}^{x_i} p_\xi(t_i)\, \mathrm{d}t_i \right) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} \left(\prod_{i=1}^{n} p_{\xi_i}(t_i) \right) \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

Однако

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} p_{\xi_1,\ldots,\xi_n}(t_1, t_2, \ldots, t_n)\, \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

Сравнивая формулы, получаем, что

$$p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{\xi_i}(x_i).$$

[⇐] Пусть известно, что

$$p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{\xi_i}(x_i).$$

Тогда

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} \left(\prod_{i=1}^{n} p_{\xi_i}(t_i) \right) \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

Разобъём его в произведение интегралов:

$$\int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} \left(\prod_{i=1}^{n} p_{\xi_i}(t_i) \right) \mathrm{d}t_1 \ldots \mathrm{d}t_n = \prod_{i=1}^{n} \left(\int\limits_{-\infty}^{x_i} p_\xi(t_i)\, \mathrm{d}t_i \right).$$

Отсюда получаем, что

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(\int\limits_{-\infty}^{x_i} p_\xi(t_i)\, \mathrm{d}t_i \right) = \prod_{i=1}^{n} F_{\xi_i}(x_i).$$

А это означает независимость $\xi_1, \xi_2, \ldots, \xi_n$.

□

1.12.3 Математическое ожидание в многомерном случае

Простой вопрос: а зачем мы ввели всё это? Ответ тоже прост — для подсчёта математического ожидания от функции многих случайных величин. Поставим вопрос формально.

Задача 1. *Пусть есть n случайных величин $\xi_1, \xi_2, \dots, \xi_n$ с совместной плотностью распределения $p_{\xi_1,\dots,\xi_n}$ и некоторая борелевская функция $f(x_1, \dots, x_n)$. Чему равно*

$$\mathsf{E}[f(\xi_1, \xi_2, \dots, \xi_n)]?$$

Ответ на этот вопрос даёт следующая теорема, которую мы сформулируем без доказательства:

Теорема 36. *Пусть $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины с совместной плотностью распределения $p_{\xi_1,\dots,\xi_n}$, а $f(x_1, x_2, \dots, x_n)$ — некоторая борелевская функция. Тогда*

$$\mathsf{E}[f(\xi_1, \xi_2, \dots, \xi_n)] = \int\limits_{\mathbb{R}^n} f(x_1, x_2, \dots, x_n) p_{\xi_1,\dots,\xi_n}(x_1, x_2, \dots, x_n)\,\mathrm{d}x_1 \dots \mathrm{d}x_n.$$

В данном случае мы пишем теорему без доказательства по той причине, потому что нам не хватает знаний по функциональному анализу и теории меры. Однако мы можем доказать этот результат в некоторых частных случаях. Попробуем доказать формулу для матожидания произведения двух случайных величин.

Теорема 37. *Пусть ξ и η — некоторые (не обязательно независимые) неотрицательные случайные величины с совместной плотностью распределения $p_{\xi,\eta}$. Тогда*

$$\mathsf{E}[\xi\eta] = \int\limits_{\mathbb{R}^2} xy p_{\xi,\eta}(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Доказательство. Введём последовательность простых случайных величин $\{\delta_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\delta_n = \sum_{i=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{i-1}{2^n} \frac{j-1}{2^n} \mathrm{I}\left\{\frac{i-1}{2^n} < \xi \leqslant \frac{i}{2^n}, \frac{j-1}{2^n} < \eta \leqslant \frac{j}{2^n}\right\}.$$

Заметим, что $\delta_n \uparrow \xi\eta$ и $\delta_{n+1} \geqslant \delta_n$. Тогда по определению математического ожидания

$$\mathsf{E}[\xi\eta] = \lim_{n\to\infty} \mathsf{E}[\delta_n].$$

Теперь посчитаем $\mathsf{E}[\delta_n]$:

$$\begin{aligned}
\mathsf{E}[\delta_n] &= \sum_{k=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{k-1}{2^n} \frac{j-1}{2^n} \mathsf{P}\left(\frac{k-1}{2^n} < \xi \leqslant \frac{k}{2^n}, \frac{j-1}{2^n} < \eta \leqslant \frac{j}{2^n}\right) \\
&= \sum_{k=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{k-1}{2^n} \frac{j-1}{2^n} \left(\int\limits_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \int\limits_{\frac{j-1}{2^n}}^{\frac{j}{2^n}} p_{\xi,\eta}(x, y)\,\mathrm{d}x\,\mathrm{d}y \right).
\end{aligned}$$

Теперь заметим, что

$$\int_0^n \int_0^n p_{\xi,\eta}(x,y)\,\mathrm{d}x\,\mathrm{d}y - \mathcal{O}\left(\frac{n}{2^n}\right) \leqslant \mathsf{E}[\delta_n] \leqslant \int_0^n \int_0^n p_{\xi,\eta}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Отсюда получаем, что

$$\mathsf{E}[\xi\eta] = \int_0^{+\infty} \int_0^{+\infty} p_{\xi,\eta}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

□

1.13 Лекция от 06.12.2016

1.13.1 Полезные формулы

Пусть есть некоторый случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ с плотностью распределения $p_\xi(x)$. Докажем следующие три утверждения:

Теорема 38. *Для любого $B \in \mathcal{B}(\mathbb{R}^n)$ выполнено, что*

$$\mathsf{P}(\xi \in B) = \int \cdots \int_B p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Доказательство. Введём функцию $f(\mathbf{x}) = \mathbf{I}_B(\mathbf{x})$. Тогда

$$\mathsf{P}(\xi \in B) = \mathsf{E}[f(\xi)] = \int \cdots \int_{\mathbb{R}^n} \mathbf{I}_B(\mathbf{x}) p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \cdots \int_B p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$ □

Теорема 39. *Для любой борелевской функции $f : \mathbb{R} \to \mathbb{R}$ и любого $i \in \{1, 2, \ldots, n\}$ выполнено*

$$\mathsf{E}[f(\xi_i)] = \int \cdots \int_{\mathbb{R}^n} f(x_i) p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Доказательство. Введём функцию $g(\mathbf{x}) = f(x_i)$. Тогда

$$\mathsf{E}[f(\xi_i)] = \mathsf{E}[g(\xi)] = \int \cdots \int_{\mathbb{R}^n} g(\mathbf{x}) p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \cdots \int_{\mathbb{R}^n} f(x_i) p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$ □

Теорема 40. *Для любой борелевской функции $f : \mathbb{R}^n \to \mathbb{R}^n$ выполнено, что*

$$\mathsf{P}(f(\xi) \in B) = \int \cdots \int_{\mathbf{x}\colon f(\mathbf{x}) \in B} p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Доказательство. Так как функция f борелевская, то для любого $B \in \mathcal{B}(\mathbb{R}^n)$ $f^{-1}(B)$ лежит в $\mathcal{B}(\mathbb{R}^n)$. Тогда, пользуясь рассуждениями, аналогичными теоремам выше, получаем, что

$$\mathsf{P}(f(\xi) \in B) = \mathsf{E}\big[\mathbf{I}_{f^{-1}(B)}(\xi)\big] = \int \cdots \int_{f^{-1}(B)} p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \cdots \int_{\mathbf{x}\colon f(\mathbf{x}) \in B} p_\xi(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$ □

1.13.2 Формула замены переменных в кратном интеграле

Допустим, что у нас есть случайная величина ξ. Рассмотрим следующую функцию: $f(\xi) = (f_1(\xi), f_2(\xi), \ldots, f_n(\xi))$. Как посчитать плотность полученного вектора?

Для ответа на этот вопрос нужно ввести формулу замены переменной в n-мерном интеграле. Сделаем это.

Определение 63. *Матрицей Якоби* $\mathbf{J}_f$ для функции $f : \mathbb{R}^n \to \mathbb{R}^m$ матрица частных производных

$$\mathbf{J}_f = \mathbf{J}(f_1, \ldots, f_m) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}_{m \times n}$$

Якобианом называют называют определитель матрицы Якоби (если он существует).

Теорема 41 (Замена переменной)**.** *Пусть задано биективное отображение $\psi : \mathbb{R}^n \to \mathbb{R}^n$, переводящее область D' в D по следующему правилу:*

$$\mathbf{x} = \psi(\mathbf{y}) \iff \begin{cases} x_1 = \psi_1(y_1, \ldots, y_n) \\ x_2 = \psi_2(y_1, \ldots, y_n) \\ \ldots \\ x_n = \psi_n(y_1, \ldots, y_n) \end{cases}$$

где $x_1, \ldots, x_n$ — "старые" координаты, а $y_1, \ldots, y_n$ — "новые" координаты. Далее, предположим, что все функции, задающее отображение, гладкие (имеют непрерывные производные первого порядка) на D', а так же $0 < |\det \mathbf{J}(\psi_1, \ldots, \psi_n)| < \infty$. Тогда, если существует интеграл

$$\int\limits_D \cdots \int f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int\limits_D \cdots \int f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n,$$

то верна следующая формула

$$\int\limits_D \cdots \int f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int\limits_{D'} \cdots \int f(\psi(\mathbf{y}))|\det \mathbf{J}(\psi_1, \ldots, \psi_n)|\, \mathrm{d}\mathbf{y}.$$

Заметим, что $f^{-1}(f(\mathbf{x})) = \mathbf{x}$, если функция f обратима. Тогда сделаем замену $\mathbf{x} = f^{-1}(\mathbf{t})$. Тогда по формуле замены переменных

$$\mathsf{P}(f(\xi) \in B) = \int\limits_{\mathbf{x}:\, f(\mathbf{x}) \in B} \cdots \int p_\xi(\mathbf{x})\, \mathrm{d}\mathbf{t} = \int\limits_B \cdots \int p_\xi(f^{-1}(\mathbf{t}))|\det \mathbf{J}(f_1^{-1}, \ldots, f_n^{-1})|\, \mathrm{d}\mathbf{t}.$$

Тогда $p_\xi(f^{-1}(\mathbf{t}))|\det \mathbf{J}(f_1^{-1}, \ldots, f_n^{-1})|$ будет плотностью функции $f(\xi)$ в точке $(t_1, \ldots, t_n)$. Докажем следующую формулу:

Теорема 42 (Формула свёртки)**.** *Пусть ξ и η — независимые случайные величины с плотностями p_ξ и p_η соответственно. Тогда*

$$p_{\xi+\eta}(x) = \int\limits_{-\infty}^{+\infty} p_\xi(y) p_\eta(x-y)\,\mathrm{d}y = \int\limits_{-\infty}^{+\infty} p_\xi(x-y) p_\eta(y)\,\mathrm{d}y.$$

Доказательство. Распишем функцию распределения суммы, пользуясь независимостью случайных величин ξ и η:

$$\mathsf{P}(\xi+\eta \leqslant z) = \iint\limits_{(x,y)\colon x+y\leqslant z} p_\xi(x)p_\eta(y)\,\mathrm{d}x\,\mathrm{d}y = \int\limits_{-\infty}^{z} p_{\xi+\eta}(y)\,\mathrm{d}y.$$

Сделаем замену переменных:

$$\begin{cases} u = y \\ v = x+y \end{cases} \iff \begin{cases} y = u \\ x = v-u \end{cases}$$

Получаем, что

$$\det \mathbf{J}(x,y) = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} = 1.$$

Тогда

$$\iint\limits_{(x,y)\colon x+y\leqslant z} p_\xi(x)p_\eta(y)\,\mathrm{d}x\,\mathrm{d}y = \iint\limits_{(u,v)\colon u\leqslant z} p_\xi(v)p_\eta(v-u)\,\mathrm{d}u\,\mathrm{d}v.$$

Далее, по теореме Фубини

$$\iint\limits_{(u,v)\colon u\leqslant z} p_\xi(v)p_\eta(v-u)\,\mathrm{d}u\,\mathrm{d}v = \int\limits_{-\infty}^{z}\int\limits_{-\infty}^{+\infty} p_\xi(v)p_\eta(v-u)\,\mathrm{d}v\,\mathrm{d}u.$$

Отсюда (по теореме Радона-Никодима) получаем, что почти везде выполнено, что

$$p_{\xi+\eta}(x) = \int\limits_{-\infty}^{+\infty} p_\xi(y)p_\eta(x-y)\,\mathrm{d}y.$$

□

1.13.3 Дисперсия и ковариация.

Пусть $(\Omega, \mathcal{F}, \mathsf{P})$ — вероятностное пространство и ξ — случайная величина, для которой определено математическое ожидание $\mathsf{E}[\xi]$.

Определение 64. *Дисперсией* случайной величины ξ называется величина

$$\mathsf{D}[\xi] = \mathsf{E}\left[(\xi - \mathsf{E}[\xi])^2\right].$$

Определение 65. Пусть (ξ, η) — пара случайных величин. Их *ковариацией* называется

$$\mathsf{cov}(\xi,\eta) = \mathsf{E}[(\xi - \mathsf{E}[\xi])\,(\eta - \mathsf{E}[\eta])].$$

Определение 66. Случайные величины ξ, η называются *некоррелированными*, если их ковариация равна нулю: $\mathsf{cov}(\xi, \eta) = 0$.

Определение 67. Если $\mathsf{D}[\xi] > 0$, $\mathsf{D}[\eta] > 0$, то величина

$$\rho(\xi, \eta) = \frac{\mathsf{cov}(\xi, \eta)}{\sqrt{\mathsf{D}[\xi]\,\mathsf{D}[\eta]}}$$

называется *коэффициентом корреляции* случайных величин ξ и η.

Лемма (Свойства дисперсии и ковариации)**.** *Пусть ξ, η, χ — некоторые случайные величины. Тогда верны следующие утверждения:*

1. *Ковариация билинейна:* $\mathsf{cov}(\xi, a\eta + b\chi) = a\,\mathsf{cov}(\xi, \eta) + b\,\mathsf{cov}(\xi, \chi)$;
2. $\mathsf{D}[\xi] = \mathsf{cov}(\xi, \xi)$;
3. *Для любого $c \in \mathbb{R}$ верно, что* $\mathsf{D}[c\xi] = c^2\,\mathsf{D}[\xi]$ *и* $\mathsf{D}[\xi + c] = \mathsf{D}[\xi]$;
4. *Связь с матожиданием:* $\mathsf{D}[\xi] = \mathsf{E}[\xi^2] - (\mathsf{E}[\xi])^2$, $\mathsf{cov}(\xi, \eta) = \mathsf{E}[\xi\eta] - \mathsf{E}[\xi]\,\mathsf{E}[\eta]$;
5. *Неравенство Коши-Буняковского*
$$\mathsf{E}[\xi\eta] \leqslant \sqrt{\mathsf{E}[\xi^2]\,\mathsf{E}[\eta^2]}.$$
6. $|\rho(\xi, \eta)| \leqslant 1$ *и принимает значения ± 1 тогда и только тогда, когда ξ и η являются линейно зависимыми почти наверное.*

Доказательство. Первые четыре свойства доказываются точно так же, как и в дискретном случае, так что повторно доказывать мы их не будем. Докажем последние два:

5. Посмотрим на следущее матожидание $\mathsf{E}[(\xi + \lambda\eta)^2]$. Заметим, что оно неотрицательно как матожидание от неотрицательной случайной величины. Распишем его, используя свойство линейности:
$$0 \leqslant \mathsf{E}\left[(\xi + \lambda\eta)^2\right] = \mathsf{E}\left[\xi^2\right] + 2\lambda\,\mathsf{E}[\xi\eta] + \lambda^2\,\mathsf{E}\left[\eta^2\right].$$
Теперь рассморим его с точки зрения многочлена от λ. Зная, что он неотрицателен, можно сказать, что дискриминант не превосходит нуля:
$$\frac{D}{4} = (\mathsf{E}[\xi\eta])^2 - \mathsf{E}\left[\xi^2\right]\mathsf{E}\left[\eta^2\right] \leqslant 0.$$
Откуда и получаем желаемое.

6. $|\rho(\xi, \eta)| \leqslant 1$ получаем, использзуя неравнство Коши-Буняковского к случайным величинам $\alpha = \frac{\xi - \mathsf{E}[\xi]}{\sqrt{\mathsf{D}[\xi]}}$, $\beta = \frac{\eta - \mathsf{E}[\eta]}{\sqrt{\mathsf{D}[\eta]}}$:
$$|\rho(\xi, \eta)| = \left|\frac{\mathsf{cov}(\xi, \eta)}{\sqrt{\mathsf{D}[\xi]\,\mathsf{D}[\eta]}}\right| = \left|\mathsf{E}\left[\left(\frac{\xi - \mathsf{E}[\xi]}{\sqrt{\mathsf{D}[\xi]}}\right)\left(\frac{\eta - \mathsf{E}[\eta]}{\sqrt{\mathsf{D}[\eta]}}\right)\right]\right| \leqslant$$
$$\leqslant \sqrt{\mathsf{E}\left[\left(\frac{\xi - \mathsf{E}[\xi]}{\sqrt{\mathsf{D}[\xi]}}\right)^2\right]\mathsf{E}\left[\left(\frac{\eta - \mathsf{E}[\eta]}{\sqrt{\mathsf{D}[\eta]}}\right)^2\right]} = 1.$$
Тогда следующие условия эквивалентны:

- $|\rho(\xi, \eta)| = 1$;
- Дискриминант в неравенстве Коши-Буняковского равен нулю;
- Сущействует корень уравнения $\mathsf{E}[(\alpha + \lambda\beta)^2] = 0$ (обозначим его за λ_0);
- $\alpha + \lambda_0\beta = 0$ почти наверное;
- $\exists a,b \in \mathbb{R} : \xi = a\eta + b$ почти наверное.

$\square$

Следствие. *Если случайные величины $\xi_1, \ldots, \xi_n$ являются попарно некоррелированными, то*

$$\mathsf{D}[\xi_1 + \ldots + \xi_n] = \mathsf{D}[\xi_1] + \ldots + \mathsf{D}[\xi_n].$$

Доказательство. Распишем дисперсию, как сумму ковариаций:

$$\mathsf{D}[\xi_1 + \ldots + \xi_n] = \mathsf{cov}(\xi_1 + \ldots + \xi_n, \xi_1 + \ldots + \xi_n) = \sum_{i,j=1}^{n} \mathsf{cov}(\xi_i, \xi_j).$$

Так как ξ_i и ξ_j некоррелированные при $i \neq j$, то в полученной сумме ненулевыми могут быть только члены вида $\mathsf{cov}(\xi_i, \xi_i)$. Тогда

$$\mathsf{D}[\xi_1 + \ldots + \xi_n] = \sum_{i=1}^{n} \mathsf{cov}(\xi_i, \xi_i) = \mathsf{D}[\xi_1] + \ldots + \mathsf{D}[\xi_n]. \qquad \square$$

Следовательно, для независимых случайных величин это также выполняется.

Определение 68. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор. Тогда математическим ожиданием случайного вектора ξ будем называть вектор, составленный из матожиданий компонент:

$$\mathsf{E}[\xi] = (\mathsf{E}[\xi_1], \ldots, \mathsf{E}[\xi_n]).$$

Назовем *матрицей ковариации* вектора матрицу порядка $n \times n$

$$\mathsf{D}[\xi] = \|\mathsf{cov}(\xi_i, \xi_j)\|_{i,\,j=1}^{n}.$$

Утверждение. Матрица ковариаций любого случайного вектора является *симметрической* и *неотрицательно определенной.*

Доказательство. Исходя из того, что $\mathsf{cov}(\xi_i, \xi_j) = \mathsf{cov}(\xi_j, \xi_i)$, матрица является симметрической.

Теперь докажем второе свойство. Пусть $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Тогда

$$\langle \mathsf{D}[\xi]\vec{x}, \vec{x} \rangle = \sum_{i,j=1}^{n} \mathsf{cov}(\xi_i, \xi_j) x_i x_j = \sum_{i,j=1}^{n} \mathsf{cov}(\xi_i x_i, \xi_j x_j).$$

Теперь распишем по линейности:

$$\sum_{i,j=1}^{n} \mathsf{cov}(\xi_i x_i, \xi_j x_j) = \sum_{j=1}^{n} \mathsf{cov}\left(\sum_{i=1}^{n} x_i \xi_i, x_j \xi_j\right) = \mathsf{cov}\left(\sum_{i=1}^{n} x_i \xi_i, \sum_{j=1}^{n} x_j \xi_j\right) = \mathsf{D}\left[\sum_{i=1}^{n} x_i \xi_i\right] \geqslant 0.$$

$\square$

1.14 Лекция от 13.12.2016

1.14.1 Три неравенства

Для случайных величин в общем случае выполнены те же неравенства, что и для дискретного случая.

Теорема 43 (Неравенство Маркова)**.** *Пусть ξ — неотрицательная случайная величина. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}(\xi \geqslant \varepsilon) \leqslant \frac{\mathsf{E}[\xi]}{\varepsilon}$$

Доказательство.

$$\mathsf{P}(\xi \geqslant \varepsilon) = \mathsf{E}\big[\mathbf{1}_{\{\xi\geqslant\varepsilon\}}\big] \leqslant \mathsf{E}\left[\frac{\xi}{\varepsilon}\mathbf{1}_{\{\xi\geqslant\varepsilon\}}\right] \leqslant \mathsf{E}\left[\frac{\xi}{\varepsilon}\right] = \frac{\mathsf{E}[\xi]}{\varepsilon}$$ □

Из неравенства Маркова легко получить неравенство Чебышёва:

Теорема 44 (Неравенство Чебышёва)**.** *Пусть ξ — случайная величина. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}(|\xi - \mathsf{E}[\xi]| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}[\xi]}{\varepsilon^2}$$

Доказательство. Просто воспользуемся неравенством Маркова:

$$\mathsf{P}(|\xi - \mathsf{E}[\xi]| \geqslant \varepsilon) = \mathsf{P}\big((\xi - \mathsf{E}[\xi])^2 \geqslant \varepsilon^2\big) \leqslant \frac{\mathsf{E}\big[(\xi - \mathsf{E}[\xi])^2\big]}{\varepsilon^2} = \frac{\mathsf{D}[\xi]}{\varepsilon^2}$$ □

Теперь докажем ещё одно неравенство:

Теорема 45 (Неравенство Йенсена)**.** *Пусть ξ — случайная величина, а φ — некоторая выпуклая вниз функция. Тогда $\varphi(\mathsf{E}[\xi]) \leqslant \mathsf{E}[\varphi(\xi)]$.*

Доказательство. Как известно, для выпуклых вниз функций выполнено следующее свойство: для любого $x_0 \in \mathbb{R}$ существует $\lambda(x_0)$ такое, что для любого $x \in \mathbb{R}$

$$\varphi(x) \geqslant \varphi(x_0) + \lambda(x_0)(x - x_0)$$

По сути, это говорит, что график лежит выше касательной. Теперь подставим $x = \xi$, $x_0 = \mathsf{E}[\xi]$:

$$\varphi(\xi) \geqslant \varphi(\mathsf{E}[\xi]) + \lambda(\mathsf{E}[\xi])(\xi - \mathsf{E}[\xi])$$

Осталось взять матожидание обеих частей и заметить, что $\mathsf{E}[\xi - \mathsf{E}[\xi]] = 0$:

$$\mathsf{E}[\varphi(\xi)] \geqslant \varphi(\mathsf{E}[\xi])$$ □

Если же функция $\varphi(x)$ выпукла вверх, то знак меняется на противоположный.

Пример. Если $\xi \geqslant 0$, то $\mathsf{E}\big[e^{\xi}\big] \geqslant e^{\mathsf{E}[\xi]}$. Далее, $\mathsf{E}[\xi^3] \geqslant (\mathsf{E}[\xi])^3$.

Примечание. В данном доказательстве была опущена одна весьма важная деталь: мы не требовали, чтобы φ была борелевской, а ведь надо было. Однако любая выпуклая функция является борелевской. Это следует из того, что любую выпуклую функцию можно представить, как супремум счётного набора линейных функций, то есть существуют последовательности $\{a_n\}$, $\{b_n\}$ такие, что $\varphi(x) = \sup(a_n x + b_n)$. Так как супремум от счётного множества измерим, то и $\varphi(x)$ измерима, а, следовательно, является борелевской.

1.14.2 Сходимости случайных величин

В дискретном случае мы уже вводили сходимости случайных величин. Почему бы и не обобщить этот результат? Рассмотрим некоторое вероятностное пространство $(\Omega, \mathcal{F}, \mathsf{P})$ и $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность случайных величин на этом пространстве. Далее, пусть ξ — это некоторая случайная величина.

Определение 69. Будем говорить, что ξ_n сходится к ξ *почти наверное* (с вероятностью 1), если

$$\mathsf{P}\Big(\Big\{\omega \in \Omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\Big\}\Big) = 1$$

Обозначение: $\xi_n \xrightarrow{\text{п.н.}} \xi$.

Смысл: сходимость почти наверное — это поточечная сходимость.

Определение 70. Будем говорить, что ξ_n сходится к ξ *по вероятности*, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) = 0$$

Обозначение: $\xi_n \xrightarrow{\mathsf{P}} \xi$.

Смысл: ξ_n мало отклоняются от ξ.

Определение 71. Будем говорить, что ξ_n сходится к ξ *в среднем порядка* p, если

$$\lim_{n\to\infty} \mathsf{E}[|\xi_n - \xi|^p] = 0$$

Обозначение: $\xi_n \xrightarrow{L^p} \xi$.

Смысл: сходимость последовательности в пространстве $L^p = \{\xi : \mathsf{E}\,|\xi|^p < +\infty\}$.

Определение 72. Будем говорить, что ξ_n сходится к ξ *по распределению*, если для всех x точек непрерывности функции $F_\xi(x)$

$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$$

Обозначение: $\xi_n \xrightarrow{\mathrm{d}} \xi$.

Смысл: аппроксимация распределения.

Теперь напомню вам несколько результатов, связанных с сходимостью случайных величин. Для простоты введём следующее: если $\xi_n \to \xi$, где ξ обладает каким-то распределением T, то будем писать, что $\xi_n \to T$.

Теорема 46 (Пуассон)**.** *Пусть* $\xi_n \sim \mathrm{Bin}(n, p(n))$*, где* $p(n) = \frac{\lambda}{n}$*,* $\lambda > 0$*. Тогда* $\xi_n \xrightarrow{\mathrm{d}} \mathrm{Pois}(\lambda)$*.*

Теорема 47 (Муавр, Лаплас)**.** *Пусть* $\xi_n \sim \mathrm{Bin}(n, p)$*, где* $p \in (0, 1)$ *— фиксированное число. Тогда*

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{\xi_n - \mathsf{E}[\xi]}{\sqrt{\mathsf{D}[\xi]}} \leqslant x\right) = \Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,\mathrm{d}y$$

где $\Phi(x)$ *— это функция распределения* $\mathcal{N}(0, 1)$*. По-другому это можно записать так:*

$$\frac{\xi_n - \mathsf{E}[\xi]}{\sqrt{\mathsf{D}[\xi]}} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1)$$

Теорема 48 (закон больших чисел в форме Чебышёва)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}$ *— последовательность попарно независимых случайных величин таких, что все дисперсии конечны: то есть, существует константа* $C \in \mathbb{R}$ *такая, что для любого натурального* n $\mathsf{D}[\xi_n] \leqslant C$. *Теперь, введём обозначение* $S_n = \xi_1 + \dots + \xi_n$. *Тогда*

$$\frac{S_n - \mathsf{E}[S_n]}{n} \xrightarrow{\mathsf{P}} 0$$

Теперь вернёмся к новому материалу.

Определение 73. Пусть $\{A_n \mid n \in \mathbb{N}\}$ — это счётная последовательность событий. Тогда событием $\{A_n \text{ б.ч.}\}$ (*событием, происходящим бесконечное число раз*) назовём событие, устроенное следующим образом:

$$\omega \in \{A_n \text{ б.ч.}\} \iff \exists \{A_{n_k} \mid k \in \mathbb{N}\} : \forall k \in \mathbb{N}\ \omega \in A_{n_k}$$

Формально это событие можно записать следующим образом:

$$\{A_n \text{ б.ч.}\} = \bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_m$$

Лемма (Критерий сходимости почти наверное)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}$, ξ *— случайные величины. Тогда* $\xi_n \xrightarrow{\text{п.н.}} \xi$ *тогда и только тогда, когда для любого* $\varepsilon > 0$

$$\mathsf{P}(\{A_n^\varepsilon \text{ б.ч.}\}) = 0, \text{ где } A_n^\varepsilon = \{|\xi_n - \xi| \geqslant \varepsilon\}$$

Или же, что равносильно:

$$\forall \varepsilon > 0 \lim_{n \to \infty} \mathsf{P}\left(\sup_{m \geqslant n} |\xi_m - \xi| \geqslant \varepsilon\right) = 0.$$

Доказательство. Из определения предела последовательности следует, что $\xi_n(\omega) \to \xi(\omega)$ равносильно тому, что для любого $\varepsilon > 0$ событий вида $\{|\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\}$ лишь конечное число. Следовательно, $\mathsf{P}(\xi_n \not\to \xi$ почти наверное$) = 0$ можно записать следующим образом:

$$\mathsf{P}\left(\bigcup_{m=1}^{\infty} \{A_n^{1/m} \text{ б.ч.}\}\right) = 0 \iff \forall m \in \mathbb{N}\, \mathsf{P}\big(\{A_n^{1/m} \text{ б.ч.}\}\big) = 0$$

Так как для любого $\varepsilon > \frac{1}{m}$ $A_n^\varepsilon \subseteq A_n^{1/m}$, то это равносильно тому, что

$$\forall \varepsilon > 0\, \mathsf{P}(\{A_n^\varepsilon \text{ б.ч.}\}) = 0 \iff \forall \varepsilon > 0\, \mathsf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_m^\varepsilon\right) = 0$$

Теперь введём события

$$B_n^\varepsilon = \bigcup_{m=n}^{\infty} A_m^\varepsilon$$

Несложно заметить, что $B_n^\varepsilon \supseteq B_{n+1}^\varepsilon$. Тогда по непрерывности вероятностной меры

$$\lim_{n \to \infty} \mathsf{P}(B_n^\varepsilon) = 0 = \lim_{n \to \infty} \mathsf{P}\left(\bigcup_{m=n}^{\infty} A_m^\varepsilon\right)$$

Осталось заметить две вещи:

1. Так как $\{\bigcup_{m=n}^{\infty} |\xi_m - \xi| \geqslant \varepsilon\} \subseteq \{\sup_{m\geqslant n} |\xi_m - \xi| \geqslant \varepsilon\}$, то

$$\mathsf{P}\left(\sup_{m\geqslant n} |\xi_m - \xi| \geqslant \varepsilon\right) = 0 \implies \mathsf{P}\left(\bigcup_{m=n}^{\infty} |\xi_m - \xi| \geqslant \varepsilon\right) = 0$$

2. Однако:

$$\mathsf{P}\left(\bigcup_{m=n}^{\infty} |\xi_m - \xi| \geqslant \varepsilon\right) = 0 \implies \mathsf{P}\left(\sup_{m\geqslant n} |\xi_m - \xi| \geqslant 2\varepsilon\right) = 0$$

Тем самым получаем желаемое. □

Следствие. *Если для любого $\varepsilon > 0$ ряд $\sum \mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon)$ сходится, то $\xi_n \xrightarrow{\text{п.н.}} \xi$.*

Доказательство. Как мы показывали, $\xi_n \xrightarrow{\text{п.н.}} \xi$ равносильно тому, что для любого $\varepsilon > 0$ $\lim\limits_{n\to\infty} \mathsf{P}(B_n^\varepsilon) = 0$. Однако,

$$\mathsf{P}(B_n^\varepsilon) = \mathsf{P}\left(\bigcup_{m=n}^{\infty} |\xi_m - \xi| \geqslant \varepsilon\right) \leqslant \sum_{m=n}^{\infty} \mathsf{P}(|\xi_m - \xi| \geqslant \varepsilon) \xrightarrow{n\to\infty} 0 \text{ (как остаток ряда)} \qquad \square$$

Казалось бы, зачем столько видов сходимостей? Они настолько разные? Оказывается, что да.

Теорема 49 (о взаимоотношении сходимостей)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные величины. Тогда*

1. $\xi_n \xrightarrow{\text{п.н.}} \xi \implies \xi_n \xrightarrow{\mathsf{P}} \xi$
2. $\xi_n \xrightarrow{L^p} \xi \implies \xi_n \xrightarrow{\mathsf{P}} \xi$
3. $\xi_n \xrightarrow{\mathsf{P}} \xi \implies \xi_n \xrightarrow{\mathrm{d}} \xi$

В обратную сторону утверждения в общем случае не верны. Про взаимосвязь сходимости почти наверное и в среднем ничего сказать нельзя.

Доказательство. Сперва докажем, что из сходимости почти наверное следует сходимость по вероятности. Действительно, пусть $\xi_n \xrightarrow{\text{п.н.}} \xi$. По доказанному выше критерию

$$\forall \varepsilon > 0 \lim_{n\to\infty} \mathsf{P}\left(\sup_{m\geqslant n} |\xi_m - \xi| \geqslant \varepsilon\right) = 0.$$

Однако

$$\mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) \leqslant \mathsf{P}\left(\sup_{m\geqslant n} |\xi_m - \xi| \geqslant \varepsilon\right) \xrightarrow{n\to\infty} 0$$

Теперь покажем, что из сходимости в среднем порядка p следует сходимость по вероятности. На самом деле, по неравенству Маркова

$$\mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon) = \mathsf{P}(|\xi_n - \xi|^p \geqslant \varepsilon^p) \leqslant \frac{\mathsf{E}\,|\xi_n - \xi|^p}{\varepsilon^p} \xrightarrow{n\to\infty} 0.$$

Теперь приступим к самому сложному: доказательству того, что из сходимости по вероятности следует сходимость по распределению. Пусть $\xi_n \xrightarrow{\mathsf{P}} \xi$ и x — это точка непрерывности F_ξ. Тогда

$$F_{\xi_n}(x) = \mathsf{P}(\xi_n \leqslant x) = \mathsf{P}(\xi_n \leqslant x, \xi - \xi_n \leqslant \varepsilon) + \mathsf{P}(\xi_n \leqslant x, \xi - \xi_n > \varepsilon)$$
$$\leqslant \mathsf{P}(\xi \leqslant x + \varepsilon) + \mathsf{P}(|\xi - \xi_n| > \varepsilon)$$

Так как $\xi_n \xrightarrow{\mathsf{P}} \xi$, то вторая вероятность стремится к 0 и

$$\varlimsup_{n\to\infty} F_{\xi_n}(x) \leqslant F_\xi(x+\varepsilon)$$

Аналогичными рассуждениями получаем, что

$$\varlimsup_{n\to\infty}(1 - F_{\xi_n}(x)) = 1 - \varliminf_{n\to\infty} F_{\xi_n}(x) \leqslant 1 - F_\xi(x-\varepsilon)$$

Тогда

$$F_\xi(x-\varepsilon) \leqslant \varliminf_{n\to\infty} F_{\xi_n}(x) \leqslant \varlimsup_{n\to\infty} F_{\xi_n}(x) \leqslant F_\xi(x+\varepsilon)$$

Так какx — точка непрерывности, то, устремляя ε к нулю, получаем, что

$$\exists \lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$$

На тему остальных связей можно посмотреть книжку Стоянова "Контрпримеры в теории вероятностей". □

Однако в некторых случаях можно перейти от более слабой к более сильной сходимости.

Утверждение. Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные величины. Если $\xi_n \xrightarrow{\mathsf{P}} \xi$, то существует подпоследовательность $\{\xi_{n_k} \mid k \in \mathbb{N}\}$ такая, что $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$.

Доказательство. Для каждого $k \in \mathbb{N}$ выберем n_k такое, что

$$\forall n > n_k\, \mathsf{P}\left(|\xi_n - \xi| \geqslant \frac{1}{k}\right) \leqslant \frac{1}{2^k}$$

Но тогда

$$\sum_{k=1}^{\infty} \mathsf{P}(|\xi_{n_k} - \xi| \geqslant \varepsilon) = \left|K > \frac{1}{\varepsilon}\right| = O\left(\sum_{k=K}^{\infty} \mathsf{P}(|\xi_{n_k} - \xi| \geqslant \varepsilon)\right)$$
$$= O\left(\sum_{k=K}^{\infty} \mathsf{P}\left(|\xi_{n_k} - \xi| \geqslant \frac{1}{k}\right)\right) < +\infty$$

Это и означает, что $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$. □

1.15 Лекция от 20.12.2016

1.15.1 Немного о законе больших чисел

Как вы помните, в том законе больших чисел, который мы доказывали, была сходимость по вероятности. Это так называемый *закон больших чисел в форме Чебышёва.* Но его можно усилить в смысле вида сходимости:

Теорема 50 (Усиленный закон больших чисел)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}$ *— последовательность независимых и одинаково распределённых случайных величин, причём четвёртый центральный момент* ξ_1 *ограничен, то есть существует такая константа* $C \in (0, +\infty)$*, что* $\mathsf{E}[(\xi_1 - \mathsf{E}[\xi_1])^4] \leqslant C$*. Как обычно, обозначим через* S_n *сумму первых* n *случайных величин:* $S_n = \xi_1 + \dots + \xi_n$*. Тогда*

$$\frac{S_n - \mathsf{E}[S_n]}{n} \xrightarrow{\text{п.н.}} 0$$

Доказательство. Без ограничения общности будем считать, что $\mathsf{E}[\xi_1] = 0$. Если это не так, то введём в рассмотрение случайные величины $\eta_i = \xi_i - \mathsf{E}[\xi_i]$.

Мы хотим показать, что $\frac{S_n}{n} \xrightarrow{\text{п.н.}} 0$. Это равносильно тому, что для любого $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathsf{P}\left(\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right) < +\infty$$

Оценим член суммы с помощью неравенства Маркова:

$$\mathsf{P}\left(\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right) = \mathsf{P}\left(\left|\frac{S_n}{n}\right|^4 \geqslant \varepsilon^4\right) \leqslant \frac{\mathsf{E}\left[\frac{S_n^4}{n^4}\right]}{\varepsilon^4} = \frac{\mathsf{E}[S_n^4]}{\varepsilon^4 n^4}$$

Теперь оценим $\mathsf{E}[S_n^4]$:

$$\mathsf{E}\left[S_n^4\right] = \sum_{i,j,k,l=1}^{n} \mathsf{E}[\xi_i \xi_j \xi_k \xi_k]$$

Заметим, что если какая-то случайная величина входит в матожидание с первой степенью, то по независимости это матожидание обернётся в ноль. Тогда

$$\mathsf{E}\left[S_n^4\right] = \sum_{i=1}^{n} \mathsf{E}\left[\xi_i^4\right] + 6 \sum_{i,j=1}^{n} \mathsf{E}\left[\xi_i^2\right] \mathsf{E}\left[\xi_j^2\right]$$

Осталось заметить две вещи: $\mathsf{E}[\xi_i^4] \leqslant C$, и, так как $\mathsf{D}[\xi_i^2] \geqslant 0$, то $\mathsf{E}[\xi_i^2] \leqslant \sqrt{\mathsf{E}[\xi_i^4]} \leqslant \sqrt{C}$. Следовательно,

$$\mathsf{E}\left[S_n^4\right] \leqslant Cn + 6Cn^2 \implies \frac{\mathsf{E}[S_n^4]}{\varepsilon^4 n^4} = O\left(\frac{1}{n^2}\right)$$

Тогда по признаку сравнения ряд сходится и мы получаем желаемое. □

Следствие. *В условиях уЗБЧ*

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} \mathsf{E}[\xi_1]$$

Впрочем, в этом усилении появляется весьма странное ограничение на четвёртый центральный момент. Можно ли от него избавиться? Можно:

Теорема 51 (уЗБЧ в форме Колмогорова)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}$ *— последовательность независимых и одинаково распределённых случайных величин, причём* $\mathsf{E}[\xi_1] < \infty$*. Тогда*

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} \mathsf{E}[\xi_1]$$

В дальнейшем мы будем пользоваться именно этим вариантом закона больших чисел. Смысл уЗБЧ не изменился — он является теоретическим обоснованием принципа устойчивости частот.

1.15.2 Предельный переход под знаком матожидания

Допустим, что у нас есть последовательность случайных величин $\{\xi_n \mid n \in \mathbb{N}\}$ и случайная величина ξ такая, что $\xi_n \xrightarrow{\text{п.н.}} \xi$. Можно ли сказать, что $\mathsf{E}[\xi_n] \to \mathsf{E}[\xi]$? Ответ на этот вопрос даёт следующая

Теорема 52 (о монотонной сходимости)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ *— случайные величины.*

1. *Пусть* $0 \leqslant \xi_n \uparrow \xi$*. Тогда* $\mathsf{E}[\xi_n] \to \mathsf{E}[\xi]$*.*
2. *Пусть* $\xi_n \uparrow \xi$ *и* $\xi_n \geqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда* $\mathsf{E}[\xi_n] \to \mathsf{E}[\xi]$*.*
3. *Пусть* $\xi_n \downarrow \xi$ *и* $\xi_n \leqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда* $\mathsf{E}[\xi_n] \to \mathsf{E}[\xi]$*.*

Доказательство. Второй и третий пункт являются простыми следствиями первого, так что начнём с первого пункта.

Для каждого $n \in \mathbb{N}$ построим последовательность простых случайных величин $\{\xi_n^{(k)} \mid k \in \mathbb{N}\}$ такую, что $0 \leqslant \xi_n^{(k)} \uparrow \xi_n$. Далее, построим последовательность $\{\delta_k \mid k \in \mathbb{N}\}$ по следующему правилу:

$$\delta_k = \max_{1 \leqslant n \leqslant k} \xi_n^{(k)}$$

Несложно заметить, что все δ_k — это простые случайные величины и $0 \leqslant \delta_k \leqslant \delta_{k+1}$. Следовательно, существует случайная величина δ такая, что $\delta_k \uparrow \delta$. Теперь заметим две вещи:

1. С одной стороны, $\delta_k = \max\limits_{1 \leqslant n \leqslant k} \xi_n^{(k)} \leqslant \max\limits_{1 \leqslant n \leqslant k} \xi_n \leqslant \xi$. Из этого следует, что $\delta \leqslant \xi$.
2. С другой стороны, для всех k верно, что $\delta_k \geqslant \xi_n^{(k)}$. Возьмём предел по k: $\delta \geqslant \xi_n$. Отсюда получаем, что для любого натурального n $\delta \geqslant \xi_n$. Возьмём предел по n: $\delta \geqslant \xi$.

Из этого следует, что $\delta = \xi$ почти наверное. Осталось найти $\mathsf{E}[\xi]$. Из показанного следует, что оно равно $\mathsf{E}[\delta]$. По определению, $\mathsf{E}[\delta_n] \to \mathsf{E}[\delta]$. Снова заметим две вещи:

1. $\delta_k = \max\limits_{1 \leqslant n \leqslant k} \xi_n^{(k)} \leqslant \max\limits_{1 \leqslant n \leqslant k} \xi_n = \xi_k$. Тогда $\mathsf{E}[\xi] = \mathsf{E}[\delta] = \lim\limits_{k \to \infty} \mathsf{E}[\delta_k] \leqslant \lim\limits_{k \to \infty} \mathsf{E}[\xi_k]$.
2. Однако $\xi_k \leqslant \xi$. Тогда $\lim\limits_{k \to \infty} \mathsf{E}[\xi_k] \leqslant \mathsf{E}[\xi]$.

Отсюда следует, что $\mathsf{E}[\xi] = \lim\limits_{k \to \infty} \mathsf{E}[\xi_k]$.

Теперь докажем второй пункт. Рассмотрим последовательность $\delta_k = \xi_k - \eta$. Понятно, что $0 \leqslant \delta_k \uparrow \xi - \eta$. Тогда

$$\lim_{k \to \infty} (\mathsf{E}[\xi_k] - \mathsf{E}[\eta]) = \lim_{k \to \infty} \mathsf{E}[\delta_k] = \mathsf{E}[\xi - \eta] = \mathsf{E}[\xi] - \mathsf{E}[\eta]$$

Доказательство же третьего пункта абсолютно аналогично доказательству второго, только нужно изменить знак на противоположный. □

Вообще говоря, монотонная сходимость — штука весьма редкая. Но можно ли что-либо сказать про матожидания, если её нет и в помине? Уже поменьше, но всё равно можно:

Лемма (Фату)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}, \eta$ *— случайные величины.*

1. *Пусть для всех натуральных* n $\xi_n \geqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда*

$$\mathsf{E}\left[\varliminf_{n\to\infty} \xi_n\right] \leqslant \varliminf_{n\to\infty} \mathsf{E}[\xi_n]$$

2. *Пусть для всех натуральных* n $\xi_n \leqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда*

$$\mathsf{E}\left[\varlimsup_{n\to\infty} \xi_n\right] \geqslant \varlimsup_{n\to\infty} \mathsf{E}[\xi_n]$$

3. *Пусть для всех натуральных* n $|\xi_n| \leqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда*

$$\mathsf{E}\left[\varliminf_{n\to\infty} \xi_n\right] \leqslant \varliminf_{n\to\infty} \mathsf{E}[\xi_n] \leqslant \varlimsup_{n\to\infty} \mathsf{E}[\xi_n] \leqslant \mathsf{E}\left[\varlimsup_{n\to\infty} \xi_n\right]$$

Доказательство. Сразу скажем, что третий пункт — это просто объединение первых двух, поэтому мы его доказывать не будем. Второй же пункт абсолютно аналогичен первому с точностью до замены знака. Так что докажем только первый пункт.

Введём последовательность случайных величин $\{\psi_n \mid n \in \mathbb{N}\}$ следующим образом:

$$\psi_n = \inf_{k\geqslant n} \xi_k$$

Тогда $\psi_n \uparrow \varliminf_{n\to\infty} \xi_n$. Так как $\xi_n \geqslant \eta$, то $\psi_n \geqslant \eta$. По теореме о монотонной сходимости

$$\mathsf{E}\left[\varliminf_{n\to\infty} \xi_n\right] = \lim_{n\to\infty} \mathsf{E}[\psi_n] = \varliminf_{n\to\infty} \mathsf{E}[\psi_n] \leqslant \varliminf_{n\to\infty} \mathsf{E}[\xi_n]$$ □

Из леммы Фату можно получить один факт, который мы будем часто использовать в дальнейшем:

Теорема 53 (Лебега о мажорируемой сходимости)**.** *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}, \xi, \eta$ *— случайные величины такие, что для всех натуральных* n $|\xi_n| \leqslant \eta$*, причём* $\mathsf{E}[\eta]$ *конечно. Тогда, если* $\xi_n \xrightarrow{\text{п.н.}} \xi$*, то*

$$\mathsf{E}[\xi_n] \xrightarrow[n\to\infty]{} \mathsf{E}[\xi] \text{ и } \mathsf{E}[|\xi_n - \xi|] \xrightarrow[n\to\infty]{} 0$$

Доказательство. По лемме Фату

$$\mathsf{E}\left[\varliminf_{n\to\infty} \xi_n\right] \leqslant \varliminf_{n\to\infty} \mathsf{E}[\xi_n] \leqslant \varlimsup_{n\to\infty} \mathsf{E}[\xi_n] \leqslant \mathsf{E}\left[\varlimsup_{n\to\infty} \xi_n\right]$$

Однако $\xi_n \xrightarrow{\text{п.н.}} \xi$. Из этого следует, что

$$\varliminf_{n\to\infty} \xi_n = \xi = \varlimsup_{n\to\infty} \xi_n \implies \mathsf{E}[\xi] = \lim_{n\to\infty} \mathsf{E}[\xi_n]$$

Второй пункт следует из первого и того, что $|\xi_n - \xi| \leqslant 2\eta$. □

1.15.3 Характеристические функции

Для исследования поведения случайных величин принять вводить так называемые *характеристические функции.*

Определение 74. *Характеристической функцией* случайной величины ξ называется

$$\varphi_\xi(t) = \mathsf{E}\left[e^{it\xi}\right]$$

Примечание. По формуле Эйлера и линейности матожидания можно записать, что

$$\varphi_\xi(t) = \mathsf{E}[\cos(t\xi)] + i\,\mathsf{E}[\sin(t\xi)]$$

Казалось бы, зачем её вводить именно так? Это станет несколько понятней, когда будут пройдены преобразования Фурье. А пока что рассмотрим пару примеров.

Задача 1. *Пусть $\xi \sim \mathrm{Bin}(n,p)$, а $\eta \sim \mathrm{Exp}(\lambda)$. Найдите их характеристические функции.*

Решение. Просто посчитаем по определению:

$$\varphi_\xi(t) = \mathsf{E}\left[e^{it\xi}\right] = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{itk} = (e^{it}p + 1 - p)^n$$
$$\varphi_\eta(t) = \mathsf{E}\left[e^{it\eta}\right] = \int_0^\infty \lambda e^{-(\lambda - it)x}\,\mathrm{d}x = \frac{\lambda}{\lambda - it}$$ □

Теперь опишем свойства характеристической функции. Их, как обычно, немало.

Свойство 1. *Для любого $t \in \mathbb{R}$ $|\varphi_\xi(t)| \leqslant 1$ и $\varphi_\xi(0) = 1$.*

Доказательство. На самом деле, $\varphi_\eta(0) = \mathsf{E}[1] = 1$ и $\left|\mathsf{E}\left[e^{it\xi}\right]\right| \leqslant \mathsf{E}\left|e^{it\xi}\right| = 1$. □

Свойство 2. *Пусть $\eta = a\xi + b$. Тогда $\varphi_\eta(t) = e^{itb}\varphi_\xi(at)$.*

Доказательство. $\varphi_\eta(t) = \mathsf{E}\left[e^{it(a\xi+b)}\right] = e^{itb}\,\mathsf{E}\left[e^{iat\xi}\right] = e^{itb}\varphi_\xi(at)$. □

Свойство 3. *Характеристическая функция равномерно непрерывна на $\mathbb{R}$.*

Доказательство. Рассмотрим модуль разности значений функций:

$$|\varphi_\xi(t+\Delta t) - \varphi_\xi(t)| = \left|\mathsf{E}\left[e^{i(t+\Delta t)\xi}\right] - \mathsf{E}\left[e^{it\xi}\right]\right| = \left|\mathsf{E}\left[e^{it\xi}\left(e^{i\Delta t\xi} - 1\right)\right]\right| \leqslant \mathsf{E}\left|e^{i\Delta t\xi} - 1\right|$$

Однако $e^{i\Delta t\xi} - 1 \xrightarrow{\text{п.н.}} 0$ при $\Delta t \to 0$ (просто поточечная сходимость) и $\left|e^{i\Delta t\xi} - 1\right| \leqslant 2$. Тогда по теореме Лебега о мажорируемой сходимости $|\varphi_\xi(t+\Delta t) - \varphi_\xi(t)| \to 0$ при $\Delta t \to 0$. А это и означает равномерную непрерывность. □

Свойство 4. $\overline{\varphi_\xi(t)} = \varphi_\xi(-t) = \varphi_{-\xi}(t)$.

Доказательство. Очевидно. □

Свойство 5 (Теорема единственности)**.** *Если $\varphi_\xi(t) = \varphi_\eta(t)$ для всех $t \in \mathbb{R}$, то $\xi \overset{d}{=} \eta$.*

Доказательство. Без доказательства. □

Свойство 6. *$\varphi_\xi(t)$ действительнозначна тогда и только тогда, когда распределение ξ симметрично, то есть $\xi \overset{d}{=} -\xi$.*

[$\Rightarrow$] $\varphi_\xi(t) = \overline{\varphi_\xi(t)} \implies \varphi_\xi(t) = \varphi_{-\xi}(t)$. Тогда по теореме единственности $\xi \stackrel{d}{=} -\xi$.

[$\Leftarrow$] Так как $\xi \stackrel{d}{=} -\xi$, то по теореме единственности $\varphi_\xi(t) = \varphi_{-\xi}(t)$. Но по свойству 4 $\varphi_\xi(t) = \overline{\varphi_\xi(t)}$.

Свойство 7. *Пусть ξ и η — независимые случайные величины. Тогда*

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t)$$

Доказательство. $\varphi_{\xi+\eta}(t) = \mathsf{E}\left[e^{it(\xi+\eta)}\right] = \mathsf{E}\left[e^{it\xi}e^{it\eta}\right] = \mathsf{E}\left[e^{it\xi}\right]\mathsf{E}\left[e^{it\eta}\right] = \varphi_\xi(t)\varphi_\eta(t)$. □

Рассмотрим пример применения этих свойств:

Задача 2. *Пусть $\xi_1 \sim \mathrm{Pois}(\lambda_1)$, $\xi_2 \sim \mathrm{Pois}(\lambda_2)$. Найдите распределение $\xi_1 + \xi_2$.*

Решение. Для начала найдём характеристическую функцию $\xi \sim \mathrm{Pois}(\lambda)$:

$$\varphi_\xi(t) = \sum_{n=1}^{\infty} e^{itn}\frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}e^{\lambda e^{it}} = \exp\left\{\lambda(e^{it}-1)\right\}$$

Следовательно, по независимости ξ_1 и ξ_2

$$\varphi_{\xi_1+\xi_2} = \exp\left\{\lambda_1(e^{it}-1)\right\}\exp\left\{\lambda_2(e^{it}-1)\right\} = \exp\left\{(\lambda_1+\lambda_2)(e^{it}-1)\right\}$$

Отсюда получаем, что $\xi_1 + \xi_2 \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$. □

1.16 Лекция от 13.01.2017

1.16.1 Задача математической статистики

Сегодняшняя лекция будет посвящена такой науке, как математическая статистика, а точнее, тому, чем она занимается и в чём её отличие от теории вероятностей.

Для начала вспомним, чем вообще занимается теория вероятностей. Вообще, основной вопрос теории вероятности состоит в следующем: пусть есть случайная величина ξ. Какова вероятность того, что $\xi \in [a, b]$? Чему равно её математическое ожидание и дисперсия?

Математическая статистика же занимается обратным вопросом: допустим, что нам известны некоторые наблюдения случайной величины ξ. Как она себя ведёт?

Рассмотрим несколько задач математической статистики:

Пример (Вопрос эмпирического выбора). Пусть в городе проживает N человек, среди которых M больно гриппом.

Примером задачи **теории вероятности** может служить следующий вопрос: допустим, мы выбрали n человек. Какова вероятность того, что среди них будет m заболевших? Стоит заметить, что для этой задачи M предполагается известным.

Задачей же **математической статистики** можно назвать следующую: допустим, что на осмотр к врачу пришло n человек и оказалось, что из них m больны. Сколько человек в городе болеет гриппом?

Посмотрим на возможный путь решения. Можно сказать, что выполнена следующая сходимость: $\frac{m}{n}N \xrightarrow{\mathsf{P}} M$ при $n \to \infty$. Далее, можно найти $\varepsilon > 0$ такой, что $\mathsf{P}\left(|M - \frac{m}{n}N| > \varepsilon\right) \leqslant 0.05$. В таком случае интервал $(\frac{m}{n}N - \varepsilon, \frac{m}{n}N + \varepsilon)$ называется *доверительным.*

Пример (Регрессионная модель)*.* Пусть была запущена ракета, и мы хотим оценить траекторию. При этом у нас есть данные вида

$$x_i = at_i^2 + bt_i + \varepsilon_i, \quad i \in \{1, 2, \dots, N\}$$

где x_i — положение ракеты в момент времени t_i, а $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$[18] — ошибка вычислений. По этим данным нужно оценить a и b. Однако никто не обещал, что данные будут однородны (т.е. распределение не обязано совпадать).

Пример (Проверка однородности)*.* Иногда мы хотим просто сказать нечто про распределение, а не искать его. Допустим, что есть два набора $(x_1, \dots, x_N)$ и $(y_1, \dots, y_N)$. Можно ли сказать, что для всех i выполнено, что $x_i \overset{d}{=} y_i$?

Пример (Проверка независимости)*.* Опять же, предположим, что у нас есть два набора $(x_1, \dots, x_N)$ и $(y_1, \dots, y_N)$. Можно ли сказать, что для всех i выполнено, что x_i независимо с y_i? На практике примером такой задачи служит связь цвета волос и цвета глаз.

1.16.2 И снова характеристические функции

Теперь же вернёмся к теории вероятностей, а точнее — к характеристическим функциям. С ними связано пара интересных теорем.

Задача 1 (Вейерштрасс)**.** *Пусть $f(x)$ — непрерывная на $[-a, a]$ функция ($a > 0$), причём $f(-a) = f(a)$. Тогда для любого $\varepsilon > 0$ существует линейная тригонометрическая комбинация вида*

$$g_\varepsilon(x) = \sum_{k=1}^{K} \left(a_k \cos \frac{\pi k x}{a} + b_k \sin \frac{\pi k x}{a} \right)$$

такая, что для любого $x \in [-a, a]$ $|f(x) - g_\varepsilon(x)| < \varepsilon$.

Доказательство. Доказательство можно найти в курсе матанализа. □

Теорема 54 (единственности)**.** *Пусть ξ и η — это случайные величины. Тогда $\xi \overset{d}{=} \eta$ тогда и только тогда, когда $\varphi_\xi(t) = \varphi_\eta(t)$ для всех $t \in \mathbb{R}$.*

Доказательство. Несложно понять, что если $\xi \overset{d}{=} \eta$, то для любой борелевской функции f $\mathsf{E}[f(\xi)] = \mathsf{E}[f(\eta)]$ и $\varphi_\xi(x) = \varphi_\eta(x)$ для всех $x \in \mathbb{R}$. Поэтому сосредоточимся на доказательстве обратного утверждения.

Зафиксируем произвольные $a < b$ и для любого $\varepsilon > 0$ введём функцию $f^\varepsilon(x)$, устроенную следующим образом:

$$f^\varepsilon(x) = \begin{cases} 0, & , x < a \\ \frac{x-a}{\varepsilon}, & a \leqslant x \leqslant a + \varepsilon \\ 1, & a + \varepsilon \leqslant x < b \\ \frac{b-x+\varepsilon}{\varepsilon}, & b \leqslant x < b + \varepsilon \\ 0, & x \geqslant b + \varepsilon \end{cases}$$

[18]Такое предположение делается для упрощения вычислений

По сути, $f^\varepsilon(x)$ — это просто непрерывное приближение $\mathbf{1}_{[a,b]}(x)$ с точностью до ε. Теперь возьмём натуральное n такое, что $[a, b+\varepsilon] \subset [-n, n]$. Тогда по теореме Вейерштрасса существует функция $f_n^\varepsilon(x)$ такая, что

$$f_n^\varepsilon(x) = \sum_{k=1}^{K} c_k e^{i\frac{\pi k}{n}x}, \quad c_k \in \mathbb{C}$$
$$\forall x \in [-n, n]\ |f^\varepsilon(x) - f_n^\varepsilon(x)| < \frac{1}{n}$$

Так как $f_n^\varepsilon(x)$ периодична, то $|f_n^\varepsilon(x)| \leqslant 2$ для всех $x \in R$.

Теперь пришло время оценок. Так как $\varphi_\xi(x) = \varphi_\eta(x)$ для всех $x \in \mathbb{R}$, то $\mathsf{E}[f_n^\varepsilon(\xi)] = \mathsf{E}[f_n^\varepsilon(\eta)]$. Тогда

$$\begin{aligned}|\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f^\varepsilon(\eta)]| &= |\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f_n^\varepsilon(\xi)] + \mathsf{E}[f_n^\varepsilon(\eta)] - \mathsf{E}[f^\varepsilon(\eta)]| \\ &\leqslant |\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f_n^\varepsilon(\xi)]| + |\,\mathsf{E}[f^\varepsilon(\eta)] - \mathsf{E}[f_n^\varepsilon(\eta)]|\end{aligned}$$

Осталось оценить $|\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f_n^\varepsilon(\xi)]|$. Заметим, что

$$\begin{aligned}|\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f_n^\varepsilon(\xi)]| &= |\,\mathsf{E}[(f^\varepsilon(\xi)] - \mathsf{E}\big[f_n^\varepsilon(\xi))(\mathbf{1}_{\{|\xi|\leqslant n\}} + \mathbf{1}_{\{|\xi|>n\}})\big]| \\ &\leqslant \mathsf{E}\big[|f^\varepsilon(\xi) - f_n^\varepsilon(\xi)|\mathbf{1}_{\{|\xi|\leqslant n\}}\big] + \mathsf{E}\big[|f^\varepsilon(\xi) - f_n^\varepsilon(\xi)|\mathbf{1}_{\{|\xi|>n\}}\big]\end{aligned}$$

Осталось заметить, что оба матожидания можно ограничить сверху:

$$\mathsf{E}\big[|f^\varepsilon(\xi) - f_n^\varepsilon(\xi)|\mathbf{1}_{\{|\xi|\leqslant n\}}\big] \leqslant \frac{1}{n}\,\mathsf{P}(|\xi| \leqslant n) \leqslant \frac{1}{n}$$
$$\mathsf{E}\big[|f^\varepsilon(\xi) - f_n^\varepsilon(\xi)|\mathbf{1}_{\{|\xi|>n\}}\big] \leqslant 2\,\mathsf{P}(|\xi| > n)$$

В итоге несложно понять, что при $n \to \infty$ $|\,\mathsf{E}[f^\varepsilon(\xi)] - \mathsf{E}[f^\varepsilon(\eta)]| \to 0$. Следовательно, для любого $\varepsilon > 0$ $\mathsf{E}[f^\varepsilon(\xi)] = \mathsf{E}[f^\varepsilon(\eta)]$.

Осталось дело за малым: скажем, что $|f^\varepsilon(x)| \leqslant 1$ и воспользуемся теоремой Лебега о мажорируемой сходимости:

$$\lim_{\varepsilon\to 0} \mathsf{E}[f^\varepsilon(\xi)] = \mathsf{E}\big[\mathbf{1}_{[a,b]}(\xi)\big] = F_\xi(b) - F_\xi(a)$$

Тогда для любых $a < b$ выполнено, что

$$F_\xi(b) - F_\xi(a) = F_\eta(b) - F_\eta(a)$$

Устремляя a к $-\infty$, получаем, что для любого $b \in \mathbb{R}$ $F_\xi(b) = F_\eta(b)$, что и требовалось получить. □

Теперь что-нибудь посчитаем.

Задача 2. *Найдите характеристическую функцию стандартного нормального распределения.*

Решение. Посчитаем характеристическую функцию $\xi \sim \mathcal{N}(0,1)$ по определению:

$$\varphi_\xi(t) = \mathsf{E}\big[e^{it\xi}\big] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{itx} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \cos(tx) e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

Продифференцируем обе части:

$$\varphi_\xi'(t) = \mathsf{E}\big[(i\xi)e^{it\xi}\big] = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} ixe^{itx}e^{-\frac{x^2}{2}}\,\mathrm{d}x = -\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} x\sin(tx)e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

Полученный интеграл возьмём по частям:

$$-\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} x\sin(tx)e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\sin(tx)e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}\cos(tx)e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

Несложно заметить, что мы получили дифференциальное уравнение:

$$\varphi_\xi'(t) = -t\varphi_\xi(t)$$

Теперь заметим, что

$$(\ln\varphi_\xi(t))' = \frac{\varphi_\xi'(t)}{\varphi_\xi(t)} = -t \implies \ln\varphi_\xi(t) = -\frac{t^2}{2} + C$$

Пользуясь тем, что $\varphi_\xi(0) = 1$, получаем, что $\varphi_\xi(t) = e^{-\frac{t^2}{2}}$. □

Следствие. *Если* $\xi \sim \mathcal{N}(a, \sigma^2)$, *то* $\varphi_\xi(t) = \exp\{iat - \frac{1}{2}\sigma^2 t^2\}$.

Доказательство. Как известно, $\xi = a + \sigma\eta$, где $\eta \sim \mathcal{N}(0,1)$. Тогда $\varphi_{a+\sigma\eta}(t) = e^{ita}\varphi_\eta(\sigma t) = \exp\{iat - \frac{1}{2}\sigma^2 t^2\}$. □

Следствие. *Пусть* $\xi_1 \sim \mathcal{N}(a_1, \sigma_1^2)$ *и* $\xi_2 \sim \mathcal{N}(a_2, \sigma_2^2)$ *— независимые случайные величины. Тогда* $\xi_1 + \xi_2 \sim \mathcal{N}(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

Доказательство. Пользуясь независимостью, распишем характеристическую функцию суммы:

$$\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t) = \exp\left\{i(a_1+a_2)t - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right\}$$

По теореме единственности $\xi_1 + \xi_2 \sim \mathcal{N}(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$. □

Теперь введём без доказательства одну весьма важную теорему:

Теорема 55 (Леви об обращении)**.** *Пусть* ξ *— случайная величина с функцией распределения* F_ξ *и характеристической функцией* φ_ξ. *Тогда верны два утверждения:*

1. *Для любых точек непрерывности* F_ξ $a < b$ *верно, что*

$$F_\xi(b) - F_\xi(a) = \frac{1}{2\pi}\lim_{c\to+\infty}\int\limits_{-c}^{c}\frac{e^{-ita} - e^{-itb}}{it}\varphi_\xi(t)\,\mathrm{d}t$$

2. *Если*

$$\int\limits_{-\infty}^{+\infty}|\varphi_\xi(t)|\,\mathrm{d}t < +\infty,$$

то ξ *имеет плотность* p_ξ, *которая равна*

$$p_\xi(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{-itx}\varphi_\xi(t)\,\mathrm{d}t$$

1.17 Лекция от 20.01.2017

1.17.1 Дифференцирование характеристических функций

Продолжим обсуждать характеристические функции. Вообще говоря, характеристическая функция — это вполне себе функция.[19] Поэтому хотелось бы уметь, например, брать производные от характеристической функции. Но как они будут устроены? Ответ на это даёт следующая

Теорема 56 (о производных характеристической функции)**.** *Пусть ξ — это некоторая случайная величина и $\mathsf{E}[|\xi|^r] < +\infty$ для всех $r \in \overline{1,\ldots,n}$. Тогда для всех $r \in \{1, 2, \ldots, n\}$ существуют $\varphi_\xi^{(r)}(t)$, причём*

$$\varphi_\xi^{(r)}(t) = \mathsf{E}\big[(i\xi)^r e^{it\xi}\big].$$

Более того,

$$\varphi_\xi^{(r)}(0) = i^r\,\mathsf{E}[\xi^r]$$
$$\varphi_\xi(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!}\,\mathsf{E}\big[\xi^k\big] + \frac{(it)^n}{n!}\mathcal{E}_n(t)$$

где $|\mathcal{E}_n(t)| \leqslant 3\,\mathsf{E}[|\xi|^n]$ и $\lim\limits_{t\to 0}\mathcal{E}_n(t) = 0$.

Доказательство. Докажем первое утверждение только для первой производной, ибо утверждение для старших степеней легко обобщается по индукции аналогичный образом. Рассмотрим производную функции $\varphi_\xi(t)$ в точке t_0:

$$\varphi_\xi'(t_0) = \lim_{\Delta t\to 0}\frac{\varphi_\xi(t_0+\Delta t) - \varphi_\xi(t_0)}{\Delta t} = \lim_{\Delta t\to 0}\mathsf{E}\left[e^{i\xi t_0}\left(\frac{e^{i\xi\Delta t}-1}{\Delta t}\right)\right].$$

Теперь заметим, что $\frac{e^{i\xi\Delta t}-1}{\Delta t} \xrightarrow{\text{п.н.}} i\xi$ (просто поточечная сходимость). Далее,

$$\begin{aligned}|e^{i\xi\Delta t} - 1| &= |\cos(\xi\Delta t) + i\sin(\xi\Delta t) - 1|\\ &\leqslant |\cos(\xi\Delta t) - 1| + |\sin(\xi\Delta t)| \leqslant 2|\xi\Delta t|\end{aligned}$$

Следовательно, по теореме Лебега можно считать и предел математического ожидания:

$$\lim_{\Delta t\to 0}\mathsf{E}\left[e^{i\xi t_0}\left(\frac{e^{i\xi\Delta t}-1}{\Delta t}\right)\right] = \mathsf{E}\big[e^{i\xi t_0}(i\xi)\big].$$

Если подставить в полученную формулу 0, то получим, что

$$\varphi_\xi^{(r)}(0) = \mathsf{E}\big[(i\xi)^r e^0\big] = \mathsf{E}[(i\xi)^r] = i^r\,\mathsf{E}[\xi^r].$$

Теперь докажем последнее утверждение. Разложим e^{iy} в ряд Тейлора-Маклорена с остаточным членом в форме Лагранжа до $(n-1)$-го порядка:

$$e^{iy} = \cos(y) + i\sin(y) = \sum_{k=0}^{n-1}\frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}(\cos(\theta_1 y) + i\sin(\theta_2 y)),$$

[19] Да ладно?

где $\theta_1, \theta_2 \in [0,1]$.

Почему мы берём именно до $(n-1)$-го порядка? Если взять до n-го порядка, то в остаточном члене будет $(n+1)$-я степень, а матожидание $\mathsf{E}[\xi^{n+1}]$ может быть бесконечным или вообще не существовать.

Теперь возьмём матожидание от этого выражения в точке $y = t\xi$:

$$\mathsf{E}\left[e^{it\xi}\right] = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathsf{E}\left[\xi^k\right] + \frac{(it)^n}{n!} \mathsf{E}[\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi))]$$

Теперь добавим и вычтем $\frac{(it)^n}{k!} \mathsf{E}[\xi^n]$:

$$\mathsf{E}\left[e^{it\xi}\right] = \sum_{k=0}^{n} \frac{(it)^k}{k!} \mathsf{E}\left[\xi^k\right] + \frac{(it)^n}{n!} \mathsf{E}[\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1)]$$

Осталось понять, что из себя представляет этот остаточный член. Введём обозначение

$$\mathcal{E}_n(t) = \mathsf{E}[\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1)].$$

Теперь заметим, что

$$|\mathcal{E}_n(t)| \leqslant \mathsf{E}[|\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1)|] \leqslant 3\,\mathsf{E}[|\xi|^n]$$

Далее, при $t \to 0$ выполнена поточечная сходимость:

$$\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1 \xrightarrow{\text{п.н.}} 0$$

И, конечно же, выражение внутри матожидания ограничено:

$$|\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1)| \leqslant 3|\xi|^n$$

Тогда, снова по теореме Лебега, можно перейти к пределу матожидания:

$$\lim_{t\to 0} \mathcal{E}_n(t) = 0. \qquad \square$$

Примечание. Заметим, что условие $\lim\limits_{t\to 0} \mathcal{E}_n(t) = 0$ равносильно тому, что

$$\mathsf{E}[\xi^n(\cos(\theta_1 t\xi) + i\sin(\theta_2 t\xi) - 1)] = o(1).$$

Тогда формулу разложения можно записать иным образом:

$$\varphi_\xi(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} \mathsf{E}\left[\xi^k\right] + o\left(\frac{t^n}{n!}\right)$$

Следствие. *Если $\mathsf{E}[\xi^2] \leqslant \infty$, то при $t \to 0$ верна следующая формула*

$$\varphi_\xi(t) = 1 + it\,\mathsf{E}[\xi] - \frac{t^2}{2}\mathsf{E}\left[\xi^2\right] + o(t^2)$$

1.17.2 Харатеристическая функция случайного вектора и её свойства

Ранее мы работали с характеристическими функциями случайных величин. Обобщим это понятие на случайные вектора:

Определение 75. Пусть $\xi \in \mathbb{R}^n$ — случайный вектор. Тогда *характеристической функцией* случайного вектора ξ будем называть функцию

$$\varphi_\xi(\mathbf{t}) = \mathsf{E}\big[e^{i\langle \xi, \mathbf{t}\rangle}\big],\ \mathbf{t} \in \mathbb{R}^n$$

В данном случае $\langle \xi, \mathbf{t}\rangle = \sum_{k=1}^n t_k \xi_k$ — скалярное произведение.

Свойства данной характеристической функции остаются теми же, что и в одномерном случае. Например:

Теорема 57 (единственности)**.** *Пусть ξ и η — некоторые случайные векторы в $\mathbb{R}^n$. Тогда*

$$\forall\, \mathbf{t} \in \mathbb{R}^n\ \varphi_\xi(\mathbf{t}) = \varphi_\eta(\mathbf{t}) \iff \xi \stackrel{d}{=} \eta.$$

С характеристическими функциями случайных векторов связан ещё один факт. Как вы помните, если случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности, то функция распределения случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$ разбивалось в произведение функций распределения для случайных величин $\xi_1, \ldots, \xi_n$. Оказывается, что похожее утверждение верно и для характеристических функций.

Теорема 58 (критерий независимости для характеристических функций)**.** *Случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности тогда и только тогда, когда характеристическая функция вектора $\xi = (\xi_1, \ldots, \xi_n)$ разбивается в произведение характеристических функций ξ_i:*

$$\forall \mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n\ \varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \varphi_{\xi_i}(t_i).$$

Доказательство. Как обычно, будем доказывать по очереди.

[⇒] Предположим, что случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности. Тогда рассмотрим характеристическую функцию вектора $\xi = (\xi_1, \ldots, \xi_n)$:

$$\varphi_\xi(\mathbf{t}) = \mathsf{E}\big[e^{i\langle \xi, \mathbf{t}\rangle}\big] = \mathsf{E}\Big[e^{i\sum_{k=1}^n t_k\xi_k}\Big] = \mathsf{E}\left[\prod_{k=1}^n e^{it_k\xi_k}\right]$$

Но, как известно, функции (борелевские) от незавимых случайных величин — это независимые случайные величины, а матожидание произведения независимых случайных величин разбивается в произведение матожиданий. Тогда

$$\varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \mathsf{E}\big[e^{it_k\xi_k}\big] = \prod_{k=1}^n \varphi_{\xi_i}(t_i).$$

[⇐] Теперь предположим, что

$$\forall \mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n\ \varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \varphi_{\xi_i}(t_i).$$

Покажем, что функция распределения $F_\xi(\mathbf{t})$ аналогичным образом разбивается в произвдение. Для этого рассмотрим случайный вектор $\eta = (\eta_1, \ldots, \eta_n)$, где случайные величины $\eta_1, \ldots, \eta_n$ независимы в совокупности и $\eta_i \overset{d}{=} \xi_i$ для $i \in \{1, \ldots, n\}$. Понятно, что его функция распределения равна

$$F_\eta(\mathbf{t}) = \prod_{k=1}^{n} F_{\eta_k}(t_k) = \prod_{k=1}^{n} F_{\xi_k}(t_k)$$

Осталось показать, что $\varphi_\eta(\mathbf{t}) = \varphi_\xi(\mathbf{t})$. Действительно, согласно теореме единственности и первому пункту получаем, что

$$\varphi_\eta(\mathbf{t}) = \prod_{k=1}^{n} \varphi_{\eta_k}(t_k) = \prod_{k=1}^{n} \varphi_{\xi_k}(t_k) = \varphi_\xi(\mathbf{t}).$$

Следовательно, по теореме единственности

$$\xi \overset{d}{=} \eta \implies F_\xi(\mathbf{t}) = F_\eta(\mathbf{t}) = \prod_{k=1}^{n} F_{\xi_k}(t_k).$$

А это означает, что $\xi_1, \ldots, \xi_n$ независимы в совокупности.

□

1.17.3 Центральная предельная теорема и примеры её применения

Наконец-то мы достигли «жемчужины» нашего курса — центральной предельной теоремы, на которой, можно сказать, держится математическая статистика. Но на самом деле нет. Перед этим сформулируем теорему, необходимую для её доказательства.

Теорема 59 (непрерывности)**.** *Пусть есть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные величины и $\varphi_n(t) = \varphi_{\xi_n}(t)$.*

1. *Пусть $\xi_n \xrightarrow{\mathrm{d}} \xi$. Тогда $\varphi_n(t) \to \varphi_\xi(t)$ для любого $t \in \mathbb{R}$.*
2. *Пусть существует непрерывная в нуле функция $\varphi(t)$ такая, что*

$$\forall t \in \mathbb{R} \lim_{n\to\infty} \varphi_n(t) = \varphi(t).$$

Тогда $\varphi_\xi(t) = \varphi(t)$ и $\xi_n \xrightarrow{\mathrm{d}} \xi$.

Доказательство. Без доказательства.[20] □

Впрочем, делать тут особо нечего: ЦПТ является простым следствием этой теоремы!

Теорема 60 (центральная предельная)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин, про которую известно, что $\mathsf{E}[\xi_1] < \infty$ и $0 < \mathsf{D}[\xi_1] < +\infty$. Далее, введём обозначения: $S_n = \xi_1 + \cdots + \xi_n$, $a = \mathsf{E}[\xi_1]$ и $\sigma^2 = \mathsf{D}[\xi_1]$. Тогда*

$$\frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1).$$

[20] Мне это уже надоело, ей богу. Самое вкусное идёт без доказательства.

Доказательство. Утверждение равносильно (по теореме непрерывности) тому, что характеристическая функция случайной величины

$$\mu_n = \frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}}$$

стремится к $e^{-t^2/2}$. Покажем это.

Для этого немного преобразуем μ_n, пользуясь независимостью ξ_i:

$$\mu_n = \frac{\sum_{k=1}^n (\xi_k - \mathsf{E}[\xi_k])}{\sqrt{\sum_{k=1}^n \mathsf{D}[\xi_k]}} = \sum_{k=1}^n \frac{\xi_k - a}{\sqrt{n}\sigma} = \left[\eta_k = \frac{\xi_k - a}{\sigma}\right] = \sum_{k=1}^n \frac{\eta_k}{\sqrt{n}}.$$

Заметим, что $\mathsf{E}[\eta_k] = 0$ и $\mathsf{D}[\eta_k] = 1$. Отсюда получаем, что $\mathsf{E}[\eta_k^2] = 1$ и

$$\varphi_{\eta_k}(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Теперь посмотрим на характеристическую функцию μ_n и воспользуемся тем, что η_k независимы в совокупности, как функции от независимых случайных величин:

$$\varphi_{\mu_n}(t) = \mathsf{E}\Big[e^{it\sum_{k=1}^n \frac{\eta_k}{\sqrt{n}}}\Big] = \prod_{k=1}^n \varphi_{\eta_k}\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n^2}\right)\right)^n$$

Но это, как известно, стремится к $e^{-t^2/2}$ при $n \to \infty$. Тогда $\mu_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$, что и требовалось доказать. □

Примечание. Утверждение ЦПТ равносильно следующему:

$$\sqrt{n}\left(\frac{S_n}{n} - \mathsf{E}[\xi_1]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathsf{D}[\xi_1]).$$

Следствие. *С помощью ЦПТ можно оценить скорость сходимости в законе больших чисел. Неформально,*

$$\frac{S_n}{n} - \mathsf{E}[\xi_1] = O\left(\frac{1}{\sqrt{n}}\right).$$

Теперь рассмотрим пример применения ЦПТ, более приближенный к реальности.

Пример. Допустим, что во время эксперимента мы получили 10^4 чисел, причём каждое из них измерено с точностью 10^{-6}. В каких пределах лежит суммарная ошибка, если все числа получены независимо друг от друга?

Решение. Пусть ξ_i — погрешность в вычислении i-го числа. Будем считать, что $\xi_i \sim U(-10^{-6}, 10^{-6})$. Тогда $\mathsf{E}[\xi_i] = 0$ и $\sigma^2 = \mathsf{D}[\xi_i] = \frac{10^{-12}}{3}$. Далее, введём обозначение $S_n = \xi_1 + \dots + \xi_n$ — суммарная ошибка первых n измерений. Тогда по центральной предельной теореме для достаточно больших n

$$\mathsf{P}\left(\frac{|S_n|}{\sigma\sqrt{n}} \leqslant x\right) \approx \Phi(x) - \Phi(-x).$$

Теперь возьмём x такое, чтобы $\Phi(x) - \Phi(-x) \geqslant 0.99$. Можно проверить, что $x \approx 2.58$ вполне подходит. Тогда с вероятностью 0.99 выполнено, что

$$|S_n| \leqslant x\sigma\sqrt{n} = 258 \cdot \frac{10^{-6}}{\sqrt{3}} \approx 1.5 \cdot 10^{-4}.$$ □

Хорошо, ЦПТ дала нам скорость сходимости для ЗБЧ. А насколько быстро сходится сама ЦПТ? Ответ на этот вопрос даёт следующая

Теорема 61 (Берри, Эссеен). *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин. Далее, пусть выполнены следующие ограничения: $\mathsf{E}[|\xi_1|^3] < +\infty$ и $0 < \mathsf{D}[\xi_1] < +\infty$. Введём следующие обозначения:*

$$S_n = \sum_{k=1}^{n} \xi_k$$
$$T_n = \frac{S_n - \mathsf{E}[S_n]}{\sqrt{\mathsf{D}[S_n]}}$$

Тогда

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leqslant \frac{C\,\mathsf{E}[|\xi_1 - \mathsf{E}[\xi_1]|^3]}{\mathsf{D}[\xi_1]^{3/2}\sqrt{n}},$$

где C — это некоторая постоянная, не зависящая от ξ_n.

Доказательство. Без доказательства. □

Что же мзвестно про C в теореме Берри-Эссеена? Вообще, первоначально было доказано, что $C \leqslant 100$ и $C \geqslant \frac{1}{\sqrt{2\pi}}$, причём оценка снизу неулучшаема. Верхнюю же оценку постепенно снижали. В 2013 году доказали, что $C \leqslant 0.48$.

Как показывает следующий пример, центральную предельную теорему можно использовать для оценки параметров распределения случайной величины.

Пример. Пусть нам известно, что $\xi_1, \ldots, \xi_n \sim \mathcal{N}(a, \sigma^2)$ и n достаточно большое, но не известны a и σ. Мы провели n независимых наблюдений и получили значения $\xi_1, \ldots, \xi_n$. Как оценить a и σ?

Решение. Для начала оценим a. Пусть $\overline{\xi}_n = \frac{\xi_1 + \cdots + \xi_n}{n}$. Вообще говоря, хотелось бы сразу сказать, что по ЗБЧ

$$\overline{\xi}_n \approx a.$$

Но это верно с какой-то неизвестной погрешностю — у нас же конечное число измерений. Придётся воспользоваться ЦПТ для оценки погрешности:

$$\sqrt{n}\left(\frac{\overline{\xi}_n - a}{\sigma}\right) \xrightarrow{\mathrm{d}} N(0,1) \implies \mathsf{P}\left(\sqrt{n}\left|\frac{\overline{\xi}_n - a}{\sigma}\right| \leqslant y\right) \approx \Phi(y) - \Phi(-y).$$

Возьмём $y \approx 2.58$. Тогда с вероятностью 0.99 верно, что

$$\overline{\xi}_n - \frac{\sigma y}{\sqrt{n}} \leqslant a \leqslant \overline{\xi}_n + \frac{\sigma y}{\sqrt{n}}$$

Проблема — нам всё ещё не известно значение σ. Как его найти? Это мы уже обсудим в следующий раз. □

1.18 Лекция от 27.01.2017

1.18.1 Сходимость случайных векторов

Ранее мы изучали сходимость для случайных величин. Обобщим это для случайных векторов, введя те же типы сходимости.

Определение 76. Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы из $\mathbb{R}^m$. Тогда

1. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *почти наверное* (или с вероятностью 1), если
$$\mathsf{P}\Big(\lim_{n\to\infty} \xi_n = \xi\Big) = 1.$$
2. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *по вероятности*, если для любого $\varepsilon > 0$
$$\lim_{n\to\infty} \mathsf{P}(\|\xi_n - \xi\|_2 \geqslant \varepsilon) = 0$$
В данном случае $\|x\|_2 = \sqrt{x_1^2 + \cdots + x_m^2}$.
3. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *по распределению*, если для любой непрерывной ограниченной функции $h : \mathbb{R}^m \mapsto \mathbb{R}$ выполнено, что
$$\lim_{n\to\infty} \mathsf{E}[h(\xi_n)] = \mathsf{E}[h(\xi)].$$

Так, стоп. Первые два похожи на определения в одномерном случае, но третье кардинально отличается от одномерного случая — ведь там была сходимость функций распределения. Сразу возникает вопрос: а совпадают ли эти определения для одномерного случая? Ведь случайная величина тоже является случайным вектором. Оказывается, что всё хорошо. Это утверждает следующая

Теорема 62 (об эквивалентности определений сходимости по распределению)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы из $\mathbb{R}^m$.*

1. *Пусть $m = 1$. Тогда $\xi_n \xrightarrow{d} \xi$ (в смысле случайных величин), тогда и только тогда, когда для любой непрерывной ограниченной функции $h : \mathbb{R}^m \mapsto \mathbb{R}$ выполнено, что $\lim\limits_{n\to\infty} \mathsf{E}[h(\xi_n)] = \mathsf{E}[h(\xi)]$.*
2. *Если $m > 1$, то $\xi_n \xrightarrow{d} \xi$ равносильно тому, что для любой точки $x \in \mathbb{R}^m$ непрерывности $F_\xi(x)$ следует, что $\lim\limits_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$.*

Доказательство. Как обычно, принимаем без доказательства. □

Следующиее упражнение может быть весьма полезно для доказательства некоторых фактов.

Упражнение. Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы в $\mathbb{R}^m$ и $\xi_n = (\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(m)})$. Тогда

1. $\xi_n \xrightarrow{\text{п.н.}} \xi \iff \forall i \in \{1, 2, \ldots, m\}\ \xi_n^{(i)} \xrightarrow{\text{п.н.}} \xi^{(i)}$
2. $\xi_n \xrightarrow{\mathsf{P}} \xi \iff \forall i \in \{1, 2, \ldots, m\}\ \xi_n^{(i)} \xrightarrow{\mathsf{P}} \xi^{(i)}$

3. $\xi_n \xrightarrow{\mathrm{d}} \xi \implies \forall i \in \{1, 2, \ldots, m\}\ \xi_n^{(i)} \xrightarrow{\mathrm{d}} \xi^{(i)}$, причём в обратную сторону это неверно.

Вспомним, что в одномерном случае существовала «иерархия» сходимостей: почти наверное $\Rightarrow$ по вероятности $\Rightarrow$ по распределению. Оказывается, что она остаётся верна и в многомерном случае.

Лемма (о взаимосвязи видов сходимости)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные векторы в $\mathbb{R}^m$. Тогда*

1. $\xi_n \xrightarrow{\text{п.н.}} \xi \implies \xi_n \xrightarrow{\mathsf{P}} \xi$

2. $\xi_n \xrightarrow{\mathsf{P}} \xi \implies \xi_n \xrightarrow{\mathrm{d}} \xi$

Доказательство. Вообще говоря, тут и доказывать нечего. Первый пункт автоматически следует из утверждения о равносильности общей и покоординатной сходимости для сходимостей почти наверное и по вероятности. Доказательство второго пункта аналогично одномерному случаю. □

1.18.2 Теорема о наследовании сходимости, лемма Слуцкого и их применения

Сформулируем и докажем две важные теоремы.

Теорема 63 (о наследовании сходимости)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные векторы в $\mathbb{R}^m$, а $h : \mathbb{R}^m \mapsto \mathbb{R}^n$ — это некоторая функция.*

1. *Если $\xi_n \xrightarrow{\text{п.н.}} \xi$ и h непрерывна почти всюду относительно распределения ξ (то есть существует $B \in \mathcal{B}(\mathbb{R}^m)$ такое, что h непрерывна на B и $\mathsf{P}(\xi \in B) = 1$), то $h(\xi_n) \xrightarrow{\text{п.н.}} h(\xi)$.*

2. *Аналогично, если $\xi_n \xrightarrow{\mathsf{P}} \xi$ и h непрерывна почти всюду относительно распределения ξ, то $h(\xi_n) \xrightarrow{\mathsf{P}} h(\xi)$.*

3. *Если $\xi_n \xrightarrow{\mathrm{d}} \xi$ и h непрерывна, то $h(\xi_n) \xrightarrow{\mathrm{d}} h(\xi)$.*

Доказательство. Вообще говоря, первый и третий пункт по сложности сравнимы с проверкой определений. Самое сложное — это второй пункт.

1. Для начала докажем утверждение для сходимости почти наверное. Рассмотрим вероятность
$$\mathsf{P}\Big(\lim_{n\to\infty} h(\xi_n) = h(\xi)\Big)$$
Заметим, что для любого $\xi \in B$ это равносильно тому, что $\xi_n \xrightarrow{\text{п.н.}} \xi$ (по непрерывности h). Тогда эту вероятность можно ограничить снизу:
$$\mathsf{P}\Big(\lim_{n\to\infty} h(\xi_n) = h(\xi)\Big) \geqslant \mathsf{P}\Big(\lim_{n\to\infty} \xi_n = \xi, \xi \in B\Big)$$
Вероятность каждого из событий в скобках равна единице, значит вероятность того, что событие не выполнится равна нулю, как объединение двух множеств меры ноль. Это и означает, что $h(\xi_n) \xrightarrow{\text{п.н.}} h(\xi)$.

2. Теперь докажем утверждение для сходимости по вероятности от противного. Допустим, что $h(\xi_n) \overset{\mathsf{P}}{\not\to} h(\xi)$. Тогда существуют положительные ε_0 и δ_0 и подпоследовательность $\{\xi_{n_k} \mid k \in \mathbb{N}\}$ такие, что
$$\mathsf{P}(\|h(\xi_{n_k}) - h(\xi))\| \geqslant \varepsilon_0) \geqslant \delta_0$$
Однако $\xi_{n_k} \xrightarrow{\mathsf{P}} \xi$. Тогда есть ещё одна подпоследовательность $\{\xi_{n_{k_s}} \mid s \in \mathbb{N}\}$ такая, что $\xi_{n_{k_s}} \xrightarrow{\text{п.н.}} \xi$. Тогда согласно первому пункту $h(\xi_{n_{k_s}}) \xrightarrow{\text{п.н.}} h(\xi)$. Но по лемме о взаимосвязи сходимостей $h(\xi_{n_{k_s}}) \xrightarrow{\mathsf{P}} h(\xi)$. Получаем противоречие. Следовательно, $h(\xi_n) \xrightarrow{\mathsf{P}} h(\xi)$.

3. Предположим, что $\xi_n \xrightarrow{\mathrm{d}} \xi$ и $h : \mathbb{R}^n \mapsto \mathbb{R}^k$ — непрерывная функция. Возьмём произвольную ограниченную и непрерывную функцию $f : \mathbb{R}^k \mapsto \mathbb{R}$. Тогда композиция $f \circ h : \mathbb{R}^n \mapsto \mathbb{R}$ — ограниченная и непрерывная функция. Следовательно, так как $\xi_n \xrightarrow{\mathrm{d}} \xi$, то
$$\mathsf{E}[f(h(\xi_n))] \to \mathsf{E}[f(h(\xi))].$$
Но это означает, что $h(\xi_n) \xrightarrow{\mathrm{d}} h(\xi)$.

□

Упражнение. Предположим, что некоторый случайный вектор ξ сходится к константе по распределению. Докажите, что он сходится к ней же по вероятности.

Лемма (Слуцкий)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ и $\{\eta_n \mid n \in \mathbb{N}\}$ — две последовательности случайных величин. Далее, пусть существуют константа и случайная величина ξ такие, что $\xi_n \xrightarrow{\mathrm{d}} \xi$ и $\eta_n \xrightarrow{\mathrm{d}} C$. Тогда*

1. $\xi_n + \eta_n \xrightarrow{\mathrm{d}} \xi + C$,

2. $\xi_n \eta_n \xrightarrow{\mathrm{d}} C\xi$

Доказательство. Докажем это утверждение только для суммы, ибо для произведения всё аналогично.

Пусть t — это точка непрерывности $F_{\xi+C}$. Выберем сколь угодно малый ε такой, что $t \pm \varepsilon$ тоже являются точками непрерывности $F_{\xi+C}$. Теперь рассмотрим $F_{\xi_n+\eta_n}(t) = \mathsf{P}(\xi_n + \eta_n \leqslant t)$. По формуле полной вероятности:
$$F_{\xi_n+\eta_n}(t) = \mathsf{P}(\xi_n + \eta_n \leqslant t, \eta_n > C - \varepsilon) + \mathsf{P}(\xi_n + \eta_n \leqslant t, \eta_n \leqslant C - \varepsilon).$$
Увеличим вторую вероятность следующим образом: забудем про первое условие, а второе заменим на $|\eta_n - C| \geqslant \varepsilon$. Тогда
$$F_{\xi_n+\eta_n}(t) \leqslant \mathsf{P}(\xi_n + \eta_n \leqslant t, \eta_n > C - \varepsilon) + \mathsf{P}(|\eta_n - C| \geqslant \varepsilon).$$
Теперь увеличим первую вероятность. Из условий следует, что $\xi_n + C \leqslant t + \varepsilon$. Тогда заменим оба условия на это. Следовательно,
$$F_{\xi_n+\eta_n}(t) \leqslant \mathsf{P}(\xi_n + C \leqslant t + \varepsilon) + \mathsf{P}(|\eta_n - C| \geqslant \varepsilon).$$
Осталось понять, куда стремятся полученные вероятности. Так как $\xi_n \xrightarrow{\mathrm{d}} \xi$, то $\xi_n + C \xrightarrow{\mathrm{d}} \xi + C$. Учитывая то, что $t + \varepsilon$ есть точка непрерывности, то первая вероятность стремится

к $F_{\xi+C}(t+\varepsilon)$. Вторая же вероятность стремится к нулю, так как $\eta_n \xrightarrow{\mathrm{d}} C$ (и, следовательно, $\eta_n \xrightarrow{\mathsf{P}} C$).

Отсюда получаем, что

$$\varlimsup_{n\to\infty} F_{\xi_n+\eta_n}(t) \leqslant F_{\xi+C}(t+\varepsilon).$$

Аналогичными рассуждениями получаем, что

$$\mathsf{P}(\xi_n+\eta_n \geqslant t) \geqslant \mathsf{P}(\xi_n + C \geqslant t+\varepsilon) + \mathsf{P}(|\eta_n - C| \geqslant \varepsilon).$$

Тогда

$$\varlimsup_{n\to\infty}(1 - F_{\xi_n+\eta_n}(t)) \geqslant 1 - F_{\xi+C}(t+\varepsilon) \text{ и } \varliminf_{n\to\infty} F_{\xi_n+\eta_n}(t) \geqslant F_{\xi+C}(t-\varepsilon).$$

Получаем следующую цепочку ограничений:

$$F_{\xi+C}(t-\varepsilon) \leqslant \varliminf_{n\to\infty} F_{\xi_n+\eta_n}(t) \leqslant \varlimsup_{n\to\infty} F_{\xi_n+\eta_n}(t) \leqslant F_{\xi+C}(t+\varepsilon).$$

Устремляя ε к нулю и пользуясь тем, что t — это точка непрерывности $F_{\xi+C}$, получаем, что

$$\lim_{n\to\infty} F_{\xi_n+\eta_n}(t) = F_{\xi+C}(t).$$ □

Теперь посмотрим, как применять их на практике.

Задача 1. *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин, про которые известно, что $a = \mathsf{E}\,\xi_1 < \infty$ и $0 < \sigma^2 = \mathsf{D}\,\xi_1 < +\infty$. Введём обозначение $S_n = \xi_1 + \dots + \xi_n$. Найдите предел по распределению*

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right)$$

Решение. Преобразуем это выражение следующим образом:

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) = -\frac{\sigma}{a}\cdot\frac{n}{S_n}\cdot\sqrt{n}\left(\frac{S_n}{n} - a\right)\frac{1}{\sigma}$$

Осталось заметить следующее:

$$\text{По теореме о наследовании сходимости и уЗБЧ } \frac{S_n}{n} \xrightarrow{\text{п.н.}} a \implies \frac{n}{S_n} \xrightarrow{\mathrm{d}} \frac{1}{a}$$

$$\text{По ЦПТ } \sqrt{n}\left(\frac{S_n}{n} - a\right)\frac{1}{\sigma} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

Следовательно, по лемме Слуцкого

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{\mathrm{d}} -\frac{\sigma}{a^2}\mathcal{N}(0,1) = \mathcal{N}\left(0, \frac{\sigma^2}{a^4}\right).$$ □

Теперь обобщим закон больших чисел:

Теорема 64 (усиленный ЗБЧ для векторов)**.** *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных m-мерных векторов. Далее, положим $a = \mathsf{E}\,\xi_1 \in \mathbb{R}^m$. Тогда*

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} a.$$

Доказательство. Автоматически следует из равносильности покомпонентной и общей сходимости и усиленного ЗБЧ для случайных величин. □

1.18.3 Многомерное нормальное распределение

Перед тем, как приступать к статистике, нужно ввести понятие многомерного нормального распределения.

Определение 77. Пусть есть случайный вектор $\xi \in \mathbb{R}^m$. Будем говорить, что он подчиняется *многомерному нормлаьному распределению*, если его характеристическая функция имеет вид

$$\varphi_\xi(\mathbf{t}) = \mathsf{E}\left[e^{i\langle \mathbf{t}, \xi\rangle}\right] = \exp\left\{i\left\langle \mathbf{a}, \mathbf{t}\right\rangle - \frac{1}{2}\left\langle \mathbf{\Sigma t}, \mathbf{t}\right\rangle\right\},$$

где $\mathbf{a} \in \mathbb{R}^m$ — некоторый фиксированный вектор, а $\mathbf{\Sigma}$ — некоторая симметрическая и неотрицательно определённая матрица $m \times m$. В таком случае пишут, что $\xi \sim \mathcal{N}(\mathbf{a}, \mathbf{\Sigma})$.

Есть одна весьма полезная теорема, которая даёт три эквивалентных определения.

Теорема 65 (о трёх эквивалентных определениях)**.** *Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор. Тогда следующие утверждения равносильны:*

1. *Случайный вектор ξ имеет нормальное многомерное распределение.*
2. *$\xi = \mathbf{A}\eta + \mathbf{b}$ почти наверное, где $\mathbf{A}$ — матрица размера $n \times m$, $\mathbf{b} \in \mathbb{R}^n$, а $\eta = (\eta_1, \ldots, \eta_m)$, где η_i независимы в совокупности и имеют нормальное стандартное распределение.*
3. *Для любого вектора $\mathbf{x} \in \mathbb{R}^n$ случайная величина $\langle \mathbf{x}, \xi\rangle$ имеет нормальное распределение.*

Доказательство. Докажем, что из (2) следует (3). Рассмотрим некоторый вектор $\mathbf{x} \in \mathbb{R}^n$ и случайную величину $\langle \mathbf{x}, \xi\rangle$:

$$\begin{aligned}\langle \mathbf{x}, \xi\rangle = \langle \mathbf{x}, \mathbf{\Sigma}\eta + \mathbf{a}\rangle &= \langle \mathbf{x}, \mathbf{a}\rangle + \langle \mathbf{x}, \mathbf{\Sigma}\eta\rangle \\ &= \langle \mathbf{x}, \mathbf{a}\rangle + \sum_{k=1}^{n}\left(\mathbf{\Sigma}^\top \mathbf{x}\right)_k \eta_k\end{aligned}$$

Несложно понять, что это сумма нормальных случайных величин, что есть нормальная случайная величина.

Теперь покажем, что из (3) следует (1). Пусть $\langle \mathbf{x}, \xi\rangle \sim \mathcal{N}(a_\mathbf{x}, \sigma^2_\mathbf{x})$. Тогда

$$\begin{gathered}a_\mathbf{x} = \mathsf{E}[\langle \mathbf{x}, \xi\rangle] = \sum_{k=1}^{n}\mathsf{E}[x_k\xi_k] = \sum_{k=1}^{n} x_k\,\mathsf{E}[\xi_k] = \langle \mathbf{x}, \mathsf{E}[\xi]\rangle \\ \sigma^2_\mathbf{x} = \mathsf{D}[\langle \mathbf{x}, \xi\rangle] = \mathsf{D}\left[\mathbf{x}^\top \xi\right] = \mathbf{x}^\top\,\mathsf{D}[\xi]\mathbf{x} = \langle \mathsf{D}[\xi]\mathbf{x}, \mathbf{x}\rangle\end{gathered}$$

Осталось посчитать характеристическую функцию ξ:

$$\begin{aligned}\varphi_\xi(\mathbf{x}) = \mathsf{E}\left[e^{i\langle \mathbf{x}, \xi\rangle}\right] = \varphi_{\langle \mathbf{x}, \xi\rangle}(1) &= \exp\left\{ia_\mathbf{x} - \frac{1}{2}\sigma^2_\mathbf{x}\right\} \\ &= \exp\left\{i\left\langle \mathsf{E}[\xi], \mathbf{x}\right\rangle - \frac{1}{2}\left\langle \mathsf{D}[\xi]\mathbf{x}, \mathbf{x}\right\rangle\right\}.\end{aligned}$$

А это и означает, что ξ имеет многомерное нормальное распределение.

Теперь докажем, что из (1) следует (2). Так как $\mathbf{\Sigma}$ — симметрическая и неотрицательно определённая матрица, то существует ортогональное преобразование $\mathbf{C}$ такое, что

$$\mathbf{C\Sigma C}^\top = \mathbf{D} = \mathrm{diag}(d_1, \ldots, d_k, 0, \ldots, 0),$$

где $d_1, \ldots, d_k > 0$. Тогда рассмотрим случайную величину $\xi' = \mathbf{C}(\xi - \mathbf{a})$. Найдём её характеристическую функцию:

$$\begin{aligned}\varphi_{\xi'}(\mathbf{t}) &= \mathsf{E}\Big[e^{i\langle \eta, \mathbf{t}\rangle}\Big] = \mathsf{E}\Big[e^{i\langle \mathbf{C}(\xi-\mathbf{a}), \mathbf{t}\rangle}\Big] = \mathsf{E}\Big[e^{i\big\langle (\xi-\mathbf{a}), \mathbf{C}^\top\mathbf{t}\big\rangle}\Big] \\ &= e^{-i\big\langle \mathbf{a}, \mathbf{C}^\top\mathbf{t}\big\rangle}\,\mathsf{E}\Big[e^{i\big\langle \xi, \mathbf{C}^\top\mathbf{t}\big\rangle}\Big]\end{aligned}$$

Заметим, что $\mathsf{E}\Big[e^{i\big\langle \xi, \mathbf{C}^\top\mathbf{t}\big\rangle}\Big] = \varphi_\xi(\mathbf{C}^\top\mathbf{t})$. Тогда

$$\begin{aligned}e^{-i\big\langle \mathbf{a}, \mathbf{C}^\top\mathbf{t}\big\rangle}\,\mathsf{E}\Big[e^{i\big\langle \xi, \mathbf{C}^\top\mathbf{t}\big\rangle}\Big] &= \exp\left\{-i\left\langle \mathbf{a}, \mathbf{C}^\top\mathbf{t}\right\rangle + i\left\langle \mathbf{a}, \mathbf{C}^\top\mathbf{t}\right\rangle - \frac{1}{2}\left\langle \mathbf{\Sigma}\mathbf{C}^\top\mathbf{t}, \mathbf{C}^\top\mathbf{t}\right\rangle\right\} \\ &= \exp\left\{-\frac{1}{2}\left\langle \mathbf{C\Sigma C}^\top\mathbf{t}, \mathbf{t}\right\rangle\right\} = \exp\left\{-\frac{1}{2}\left\langle \mathbf{D}\mathbf{t}, \mathbf{t}\right\rangle\right\} \\ &= \exp\left\{-\frac{1}{2}\sum_{i=1}^{k} d_i t_i^2\right\} = \prod_{i=1}^{k} e^{-\frac{d_i t_i^2}{2}}\end{aligned}$$

Получаем, что

$$\begin{cases}\xi_i' \sim \mathcal{N}(0, d_i), & i \in \{1, 2, \ldots, k\} \\ \xi_i' = 0 \text{ почти наверное}, & i \in \{k+1, \ldots, n\}\end{cases}$$

Введём случайные величины $\eta_i = \frac{\xi_i'}{\sqrt{d_i}}$ для всех $i \in \{1, 2, \ldots, k\}$ и построим по ним случайный вектор $\eta = (\eta_1, \ldots, \eta_k)^\top$. Далее, построим следующую матрицу:

$$\mathbf{B} = \begin{pmatrix} \sqrt{d_1} & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \sqrt{d_k} \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix}_{n\times k}$$

Тогда несложно понять, что $\mathbf{B}\eta = \xi'$ почти наверное. Отсюда получаем, что

$$\xi = \mathbf{C}^\top\mathbf{B}\eta + \mathbf{a}$$

□

1.19 Лекция от 03.02.2017

1.19.1 Гауссовские векторы. Продолжение

На предыдущей лекции мы доказали теорему о трёх эквивалентных определениях. Выпишем основные следствия из неё:

Следствие (Смысл параметров). *Если случайный вектор* $\xi \sim \mathcal{N}(\mathbf{a}, \mathbf{\Sigma})$, *то* $\mathbf{a} = \mathsf{E}[\xi]$, $\mathbf{\Sigma} = \mathsf{D}[\xi]$.

Доказательство. См. доказательство перехода от (2) к (3). □

Следствие (Корректность определения гауссовского вектора)**.** *Гауссовский вектор определён корректно, т.е. существует случайный вектор с таким распределением.*

Доказательство. Возьмём η, $\mathbf{C}$ и $\mathbf{B}$ из доказательства перехода от (1) к (2) (матрицы свободно восстанавливаются по $\mathbf{\Sigma}$). Тогда случайный вектор $\xi = \mathbf{C}^\top \mathbf{B}\eta + \mathbf{a}$ обладает нужным распределением. □

Следствие (Линейные преобразования)**.** *Если $\xi = (\xi_1, \ldots, \xi_n)^\top$ — гауссовский вектор, то его любое линейное преобразование тоже является гауссовским вектором.*

Доказательство. Пусть $\eta = \mathbf{B}\xi + \mathbf{b}$ — некоторое линейное преобразование ξ, где $\mathbf{B} \in \mathrm{Mat}(k \times n)$ и $\mathbf{b} = (b_1, \ldots, b_k)^\top$. Тогда рассмотрим произвольный вектор $\mathbf{t} \in \mathbb{R}^k$ и его скалярное произведение с η:

$$\langle \mathbf{t}, \eta \rangle = \langle \mathbf{t}, \mathbf{B}\xi + \mathbf{b} \rangle = \left\langle \mathbf{B}^\top \mathbf{t}, \xi \right\rangle + \langle \mathbf{t}, \mathbf{b} \rangle$$

Первое слагаемое есть нормальная случайная величина (по теореме о трёх определениях), а второе — просто константа. Тогда $\langle \mathbf{t}, \eta \rangle$ — это нормальная случайная величина. Следовательно, η — это гауссовский вектор. □

Теперь докажем одну простую, но достаточно полезную лемму:

Лемма (Критерий независимости компонент)**.** *Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — это гауссовский вектор. Тогда его компоненты независимы тогда и только тогда, когда они некоррелированы.*

Доказательство. Как известно, если случайные величины независимы, то они некоррелированы. Теперь докажем, что в данном случае верно и обратное.

Пусть $\mathsf{cov}(\xi_i, \xi_j) = 0$ при $i \neq j$. Тогда матрица ковариаций $\mathbf{\Sigma}$ диагональна. Теперь рассмотрим характеристическую функцию $\varphi_\xi(\mathbf{t})$:

$$\begin{aligned}\varphi_\xi(\mathbf{t}) &= \exp\left\{ i \langle \mathbf{a}, \mathbf{t} \rangle - \frac{1}{2} \langle \mathbf{\Sigma t}, \mathbf{t} \rangle \right\} = \exp\left\{ i \sum_{i=1}^n a_i t_i - \frac{1}{2} \sum_{i=1}^n \mathsf{cov}(\xi_i, \xi_i) t_i^2 \right\} \\ &= \exp\left\{ \sum_{i=1}^n \left(i a_i t_i - \frac{1}{2} \mathsf{D}[\xi_i] t_i^2 \right) \right\} = \prod_{i=1}^n \exp\left\{ i a_i t_i - \frac{1}{2} \mathsf{D}[\xi_i] t_i^2 \right\} = \prod_{i=1}^n \varphi_{\xi_i}(t_i)\end{aligned}$$

Отсюда по критерию независимости через характеристические функции получаем, что компоненты ξ независимы в совокупности. □

Порой хочется сказать, что любой вектор, состоящий из нормальных случайных величин, является гауссовским. Но, к сожалению, это не так.

Пример. Пусть $\xi \sim \mathcal{N}(0, 1)$, а η равновероятно принимает значения из $\{1, -1\}$. Введём случайную величину $\delta = \xi\eta$. Легко проверить, что $\delta \sim \mathcal{N}(0, 1)$ и $\mathsf{cov}(\xi, \delta) = 0$:

$$\mathsf{cov}(\xi, \delta) = \mathsf{E}\left[\xi^2 \eta\right] - \mathsf{E}[\xi]\,\mathsf{E}[\xi\eta] = 0.$$

Однако вектор (ξ, δ) не является гауссовским. (Почему?)

Как мы помним, у нормальных случайных величин была плотность. Возникает вопрос: а есть ли она у гауссовских векторов? Не всегда.

Теорема 66 (о плотности гауссовских векторов). *Пусть $\xi \sim \mathcal{N}(\mathbf{a}, \mathbf{\Sigma})$ — n-мерный гауссовский вектор. Тогда, если $\mathbf{\Sigma}$ положительно определена, то существует плотность $p_\xi(\mathbf{t})$ и она равна*

$$p_\xi(\mathbf{t}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{\Sigma}}} \exp\left\{-\frac{1}{2}\left\langle \mathbf{\Sigma}^{-1}(\mathbf{t}-\mathbf{a}), (\mathbf{t}-\mathbf{a})\right\rangle\right\}$$

Доказательство. Пусть $\mathbf{D} = \sqrt{\mathbf{\Sigma}}$, то есть матрица $\mathbf{D}$ такова, что $\mathbf{D}^2 = \mathbf{\Sigma}$. Если Σ положительно определённая, то $\mathbf{D}$ единственна. Тогда $\mathbf{D}$ тоже положительно определённая и симметрическая.

Рассмотрим $\eta = \mathbf{D}^{-1}(\xi - \mathbf{a})$. Тогда η — тоже гауссовский вектор (как линейное преобразование ξ) и $\eta \sim \mathcal{N}(0, \mathbf{D}^{-1}\mathbf{\Sigma}\mathbf{D}^{-1}) = \mathcal{N}(0, \mathbf{E}_n)$. Следовательно, $\eta = (\eta_1, \dots, \eta_n)$, где η_i независимы в совокупности и имеют стандартное нормальное распределение. Тогда

$$p_\eta(\mathbf{t}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} t_k^2\right\} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\langle \mathbf{t}, \mathbf{t}\rangle}$$

Теперь рассмотрим вероятностную меру ξ. Возьмём произвольное $B \in \mathcal{B}(\mathbb{R}^n)$

$$\mathsf{P}(\xi \in B) = \mathsf{P}\big(\eta \in \mathbf{D}^{-1}(B - \mathbf{a})\big) = \int\limits_{\mathbf{D}^{-1}(B-\mathbf{a})} p_\eta(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

Сделаем замену

$$\mathbf{y} = \mathbf{D}\mathbf{x} + a \iff \mathrm{d}\mathbf{y} = \det \mathbf{D}\,\mathrm{d}\mathbf{x}$$

Тогда

$$\mathsf{P}(\xi \in B) = \int\limits_B \frac{1}{(2\pi)^{n/2}\det \mathbf{D}} \exp\left\{-\frac{1}{2}\left\langle \mathbf{D}^{-1}(\mathbf{y}-\mathbf{a}), \mathbf{D}^{-1}(\mathbf{y}-\mathbf{a})\right\rangle\right\} \mathrm{d}y$$

А это в свою очередь равно

$$\int\limits_B \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{\Sigma}}} \exp\left\{-\frac{1}{2}\left\langle \mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{a}), (\mathbf{y}-\mathbf{a})\right\rangle\right\} \mathrm{d}y$$

Дифференцируя, получаем желаемое. □

1.19.2 Многомерная ЦПТ. Пример применения

Центральная предельная теорема верна и для случайных векторов:

Теорема 67 (многомерная ЦПТ). *Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных векторов из $\mathbb{R}^m$. Далее, пусть $\mathbf{a} = \mathsf{E}[\xi_1]$ и $\mathbf{\Sigma} = \mathsf{D}[\xi_1]$ конечны. Введём обозначение $S_n = \xi_1 + \dots + \xi_n$. Тогда*

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbf{a}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbf{\Sigma})$$

Доказательство. Доказательство идейно аналогично одномерному случаю: воспользуемся теоремой о непрерывности для характеристических функций и покажем, что последовательность характеристических функций сходится именно туда, куда нужно. □

Рассмотрим какой-нибудь пример применения многомерной ЦПТ.

Задача 1. *Пусть выполнены условия многомерной ЦПТ и $h : \mathbb{R}^m \mapsto \mathbb{R}$ — некоторая дифференцируемая в точке $\mathbf{a}$ функция. Докажите, что*

$$\sqrt{n}\left(h\left(\frac{S_n}{n}\right) - h(\mathbf{a})\right) \xrightarrow{\mathrm{d}} \langle h'(\mathbf{a}), \mathcal{N}(0, \mathbf{\Sigma})\rangle$$

где $h'(\mathbf{a})$ — вектор частных производных h в точке $\mathbf{a}$.

Доказательство. Рассмотрим разложение h рядом с точкой $\mathbf{a}$:

$$h(\mathbf{x} + \mathbf{a}) = h(\mathbf{a}) + \langle h'(\mathbf{a}), \mathbf{x}\rangle + \psi(\mathbf{x})$$

где $\frac{\psi(\mathbf{x})}{\|\mathbf{x}\|} \to 0$ при $\mathbf{x} \to \mathbf{0}$. Подставим $\mathbf{x} = S_n/n - \mathbf{a}$ и домножим обе части уравнения на $\sqrt{n}$. Это легально, так как по ЗБЧ $S_n/n \xrightarrow{\text{п.н.}} \mathbf{a}$. Тогда

$$\sqrt{n}h\left(\frac{S_n}{n}\right) = \sqrt{n}h(\mathbf{a}) + \left\langle h'(\mathbf{a}), \sqrt{n}\left(\frac{S_n}{n} - \mathbf{a}\right)\right\rangle + \sqrt{n}\psi\left(\frac{S_n}{n} - \mathbf{a}\right)$$

По многомерной ЦПТ

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbf{a}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathbf{\Sigma})$$

Рассмотрим последний член. Преобразуем его:

$$\sqrt{n}\psi\left(\frac{S_n}{n} - \mathbf{a}\right) = \frac{\psi\left(\frac{S_n}{n} - \mathbf{a}\right)}{\left\|\frac{S_n}{n} - \mathbf{a}\right\|}\left\|\sqrt{n}\left(\frac{S_n}{n} - \mathbf{a}\right)\right\|$$

Как мы сказали ранее, второй член произведения сходится к $\mathcal{N}(0, \mathbf{\Sigma})$ по распределению. Однако первый сходится к нулю почти наверное по теореме о наследовании сходимости. Тогда по лемме Слуцкого $\sqrt{n}\psi\left(\frac{S_n}{n} - \mathbf{a}\right) \xrightarrow{\mathrm{d}} 0$.

Следовательно,

$$\sqrt{n}\left(h\left(\frac{S_n}{n}\right) - h(\mathbf{a})\right) \xrightarrow{\mathrm{d}} \langle h'(\mathbf{a}), \mathcal{N}(0, \mathbf{\Sigma})\rangle = \mathcal{N}(0, (h'(\mathbf{a}))^\top \mathbf{\Sigma} h'(\mathbf{a})). \qquad \square$$

1.19.3 Условное математическое ожидание

> Будет больно, но мы попробуем сделать это немного помягче.
>
> Д.А. Шабанов

Как мы все помним, в теории вероятностей есть понятие условной вероятности. Однако можно поставить вопрос ребром: пусть ξ и η — две случайные величины. Что такое $\mathsf{E}[\xi \mid \eta = y]$ и $\mathsf{E}[\xi \mid \eta]$?

Давайте для начала посмотрим на самый простой случай — простые случайные величины. Как известно, для них

$$\mathsf{E}[\xi] = \sum_{i=1}^{n} x_i \,\mathsf{P}(\xi = x_i)$$

Было бы разумно ввести $\mathsf{E}[\xi \mid \eta = y]$ следующим образом:

$$\mathsf{E}[\xi \mid \eta = y] = \sum_{i=1}^{n} x_i \, \mathsf{P}(\xi = x_i \mid \eta = y)$$

Так и введём. Теперь посмотрим, что мы из этого имеем. Домножим на $\mathsf{P}(\eta = y)$. Тогда

$$\sum_{i=1}^{n} x_i \, \mathsf{P}(\xi = x_i, \eta = y) = \mathsf{E}\left[\sum_{i=1}^{n} x_i \mathbf{1}\{\xi = x_i\}\mathbf{1}\{\eta = y\}\right] = \mathsf{E}[\xi \mathbf{1}\{\eta = y\}]$$

Однако, если положить $\mathsf{E}[\xi \mid \eta = y] = \varphi(y)$, то

$$\varphi(y) \, \mathsf{P}(\eta = y) = \mathsf{E}[\varphi(y)\mathbf{1}\{\eta = y\}] = \begin{cases} \mathsf{E}[\xi \mathbf{1}\{\eta = y\}] \\ \mathsf{E}[\varphi(\eta)\mathbf{1}\{\eta = y\}] \end{cases}$$

Как видно, выполнено следующее свойство: $\mathsf{E}[\xi \mathbf{1}\{\eta = y\}] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}\{\eta = y\}]$. Теперь можно ввести определение для общего случая:

Определение 78. Пусть ξ — это случайная величина, а η — это случайный вектор из $\mathbb{R}^m$. Тогда *условным математическим ожиданием* ξ относительно η называется случайная величина $\mathsf{E}[\xi \mid \eta]$, удовлетворяющая двум условиям:

1. $\mathsf{E}[\xi \mid \eta] = \varphi(\eta)$, где $\varphi : \mathbb{R}^m \mapsto \mathbb{R}$ — некоторая борелевская функция (свойство измеримости).

2. Для любого $B \in \mathcal{B}(\mathbb{R}^m)$ $\mathsf{E}[\xi \mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}_B(\eta)]$ (интегральное свойство)

Теорема 68. *Если $\mathsf{E}[\xi] < \infty$, то $\mathsf{E}[\xi \mid \eta]$ существует и единственно с точностью до равенства почти наверное.*

Доказательство. Как обычно, нам не хватает знаний в теории меры для того, чтобы доказать эту теорему. □

Неформально говоря, $\mathsf{E}[\xi \mid \eta]$ — это усреднение значений ξ по значениям η.

1.20 Лекция от 10.02.2017

1.20.1 Вычисление условного математического ожидания

Ранее мы ввели условное математическое ожидание. Теперь хотелось бы научиться считать его. Для начала покажем, как это делается в дискретном случае.

Теорема 69. *Пусть η имеет дискретное распределение и $X = (x_1, \dots, x_n, \dots)$ — множество значений. Тогда для любой случайной величины ξ выполнено, что*

$$\mathsf{E}[\xi \mid \eta] = \sum_{k=1}^{\infty} \frac{\mathsf{E}[\xi \, \mathbf{I}\{\eta = x\}]}{\mathsf{P}(\eta = x)} \, \mathbf{I}\{\eta = x\}$$

Доказательство. Достаточно проверить измеримость и интегральное свойство. Как известно,
$I\{\eta = x\}$ является борелевской функцией от η для любого x. Тогда функция в правой части является борелевской функцией от η.

Теперь проверим интегральное свойство. Для начала возьмём $B = \{x_n\}$, где x_n — это одно из значений η, для которого $\mathsf{P}(\eta = x_n) > 0$.

$$\begin{aligned}\mathsf{E}[\mathsf{E}[\xi \mid \eta]\,\mathbf{I}\{\eta = x_n\}] &= \mathsf{E}\left[\frac{\mathsf{E}[\xi\,\mathbf{I}\{\eta = x_n\}]}{\mathsf{P}(\eta = x_n)}\,\mathbf{I}\,\{\eta = x_n\}\right] \\ &= \frac{\mathsf{E}[\xi\,\mathbf{I}\{\eta = x_n\}]}{\mathsf{P}(\eta = x_n)}\,\mathsf{E}[\mathbf{I}\,\{\eta = x_n\}] = \mathsf{E}[\xi\,\mathbf{I}\{\eta = x_n\}]\end{aligned}$$

Теперь обобщим это на произвольное B. Пусть в B есть элементы $y_1, \ldots, y_n, \ldots$, которые являются значениям η. Тогда по линености матожидания получаем, что

$$\begin{aligned}\mathsf{E}[\xi\,\mathbf{I}\{\eta \in B\}] &= \mathsf{E}\left[\xi\left(\sum_{i=1}^{\infty}\mathbf{I}\{\eta = y_i\}\right)\right] = \sum_{i=1}^{\infty}\mathsf{E}[\xi\,\mathbf{I}\{\eta = y_i\}] \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\mathsf{E}[\xi\,\mathbf{I}\{\eta = y_i\}]\end{aligned}$$

Заметим, что $|\xi\,\mathbf{I}\{\eta = y_i\}| \leqslant |\xi|$ и $\mathsf{E}[\xi\,\mathbf{I}\{\eta = y_i\}] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]\,\mathbf{I}\{\eta = y_i\}]$. Тогда по теореме Лебега о мажорируемой сходимости

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathsf{E}[\xi\,\mathbf{I}\{\eta = y_i\}] = \mathsf{E}\left[\mathsf{E}[\xi \mid \eta]\left(\sum_{i=1}^{\infty}\mathbf{I}\{\eta = y_i\}\right)\right] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]\,\mathbf{I}\{\eta \in B\}].$$ □

1.20.2 Свойства условного математического ожидания

У условного математического ожидания немало полезных свойств.

Свойство 1. *Если $\xi = f(\eta)$, где f — некоторая борелевская функция, то $\mathsf{E}[\xi \mid \eta] = \xi$.*

Доказательство. Несложно проверить, что для ξ выполняются свойство измеримости и интегральное свойство. □

Свойство 2 (Линейность)**.** *Для любых случайных величин (или векторов) ξ, η и δ выполнено, что*

$$\forall a, b \in \mathbb{R}\ \ \mathsf{E}[a\xi + b\delta \mid \eta] = a\,\mathsf{E}[\xi \mid \eta] + b\,\mathsf{E}[\delta \mid \eta].$$

Доказательство. Для начала заметим, что и $\mathsf{E}[\xi \mid \eta]$, и $\mathsf{E}[\delta \mid \eta]$ являются борелевскими функциями от η. Тогда $a\,\mathsf{E}[\xi \mid \eta] + b\,\mathsf{E}[\delta \mid \eta]$ тоже есть борелевская функция от η.

Осталось проверить интегральное свойство. По линейности матожидания

$$\mathsf{E}[(a\xi + b\delta)\mathbf{1}_B(\eta)] = a\,\mathsf{E}[\xi\mathbf{1}_B(\eta)] + b\,\mathsf{E}[\delta\mathbf{1}_B(\eta)].$$

Согласно интегральному свойству:

$$a\,\mathsf{E}[\xi\mathbf{1}_B(\eta)] + b\,\mathsf{E}[\delta\mathbf{1}_B(\eta)] = a\,\mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}_B(\eta)] + b\,\mathsf{E}[\mathsf{E}[\delta \mid \eta]\mathbf{1}_B(\eta)].$$

Свернём сумму назад:

$$a\,\mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}_B(\eta)] + b\,\mathsf{E}[\mathsf{E}[\delta \mid \eta]\mathbf{1}_B(\eta)] = \mathsf{E}[(a\,\mathsf{E}[\xi \mid \eta] + b\,\mathsf{E}[\delta \mid \eta])\mathbf{1}_B(\eta)].$$

Тем самым получаем желаемое. □

Свойство 3 (Формула полной вероятности)**.** *«Математическое ожидание убирает условие»:* $\mathsf{E}[\mathsf{E}[\xi \mid \eta]] = \mathsf{E}[\xi]$.

Доказательство. Пусть $\eta \in \mathbb{R}^n$. Воспользуемся интегральным свойством, положив $B = \mathbb{R}^n$:

$$\mathsf{E}[\xi] = \mathsf{E}[\xi \mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi \mid \eta] \mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]]. \qquad \square$$

Примечание. Название оно получило из-за того, что является обобщением формулы полной вероятности. Покажем это. Пусть $\{B_n \mid n \in \mathbb{N}\}$ — попарно не пересекающиеся события. Построим η следующим образом: она будет обладать дискретным распределением и $\mathsf{P}(\eta = n) = \mathsf{P}(B_n)$. Тогда, взяв $\xi = \mathbf{1}_B(x)$, получим, что

$$\mathsf{P}(B) = \mathsf{E}[\mathsf{E}[\mathbf{1}_B \mid \eta]] = \mathsf{E}[\mathsf{P}(B \mid \eta)].$$

А это, по сути, и есть формула полной вероятности.

Свойство 4. *Если ξ независимо с η, то* $\mathsf{E}[\xi \mid \eta] = \mathsf{E}[\xi]$.

Доказательство. Заметим, что $\mathsf{E}[\xi]$ — константа. Тогда можно считать, что $\mathsf{E}[\xi]$ есть борелевская функция от η. Проверим интегральное свойство:

$$\mathsf{E}[\xi \mathbf{1}_B(\eta)] = |\text{по независимости}| = \mathsf{E}[\xi]\, \mathsf{E}[\mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi] \mathbf{1}_B(\eta)]. \qquad \square$$

Свойство 5 (Сохранение отношения порядка)**.** *Если $\xi \leqslant \delta$ почти наверное, то* $\mathsf{E}[\xi \mid \eta] \leqslant \mathsf{E}[\delta \mid \eta]$ *почти наверное.*

Доказательство. Рассмотрим случайную величину $\zeta = \mathsf{E}[\xi \mid \eta] - \mathsf{E}[\delta \mid \eta]$. Несложно понять, что она является борелевской функцией от η. Теперь рассмотрим $\mathsf{E}\big[\chi \mathbf{1}_{(0,+\infty)}(\chi)\big]$. Заметим, что $\mathsf{E}\big[\chi \mathbf{1}_{(0,+\infty)}(\chi)\big] \geqslant 0$. Теперь скажем, что $\mathbf{1}_{(0,+\infty)}(\chi) = \mathbf{1}_B(\eta)$ для какого-то борелевского множества B.[21] Тогда

$$\mathsf{E}\big[\chi \mathbf{1}_{(0,+\infty)}(\chi)\big] = \mathsf{E}[(\mathsf{E}[\xi \mid \eta] - \mathsf{E}[\delta \mid \eta]) \mathbf{1}_B(\eta)].$$

Пользуясь линейностью матожидания и интегральным свойством, получаем, что

$$\mathsf{E}\big[\chi \mathbf{1}_{(0,+\infty)}(\chi)\big] = \mathsf{E}[\xi \mathbf{1}_B(\eta)] - \mathsf{E}[\delta \mathbf{1}_B(\eta)] \leqslant 0, \text{ так как } \xi \leqslant \delta.$$

Тогда получаем, что $\chi \mathbf{1}_{(0,+\infty)}(\chi) = 0$ почти наверное и $\xi \leqslant 0$. $\square$

Свойство 6 (Телескопическое)**.** *Пусть ξ, η и δ — случайные векторы. Тогда*

1. $\mathsf{E}[\mathsf{E}[\xi \mid \eta] \mid (\eta, \delta)] = \mathsf{E}[\xi \mid \eta]$,
2. $\mathsf{E}[\mathsf{E}[\xi \mid (\eta, \delta)] \mid \eta] = \mathsf{E}[\xi \mid \eta]$.

Доказательство. Первый пункт почти очевиден. Заметим, что $\mathsf{E}[\xi \mid \eta]$ — борелевская функция от η. Но тогда она является борелевской функцией от (η, δ). Следовательно, по свойству 1 получаем желаемое.

Теперь докажем второе утверждение. Заметим, что $\mathsf{E}[\xi \mid \eta]$ является борелевской функцией от η. Теперь проверим интегральное свойство. Пусть $\delta \in \mathbb{R}^m$. Тогда по интегральному свойству:

$$\begin{aligned}\mathsf{E}[\mathsf{E}[\xi \mid (\eta, \delta)] \mathbf{1}_B(\eta)] &= \mathsf{E}[\mathsf{E}[\xi \mid (\eta, \delta)] \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] = \mathsf{E}[\mathsf{E}[\xi \mid (\eta, \delta)] \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] \\ &= \mathsf{E}[\xi \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] = \mathsf{E}[\xi \mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi \mid \eta] \mathbf{1}_B(\eta)]. \qquad \square\end{aligned}$$

[21] На самом деле $B = \chi^{-1}(0, +\infty)$.

Свойство 7 (Предельный переход). *Пусть* $\{\xi_n \mid n \in \mathbb{N}\}$, ξ, η, δ *— случайные величины.*

1. *Если* $0 \leqslant \xi_n \uparrow \xi$ *почти наверное, то* $\mathsf{E}[\xi_n \mid \eta] \uparrow \mathsf{E}[\xi \mid \eta]$ *почти наверное.*
2. *Если* $\xi_n \xrightarrow{\text{п.н.}} \xi$ *и* $|\xi_n| \leqslant |\delta|$ *для всех натуральных* n, *то* $\mathsf{E}[\xi_n \mid \eta] \xrightarrow{\text{п.н.}} \mathsf{E}[\xi \mid \eta]$.

Доказательство. Так как ξ_n монотонно возрастают, то и $\mathsf{E}[\xi_n \mid \eta]$ монотонно возрастают (по свойству 5). Тогда существует (не обязательно конечный) предел $h = \lim\limits_{n\to\infty} \mathsf{E}[\xi_n \mid \eta]$. Так как для любого натурального n $\mathsf{E}[\xi_n \mid \eta]$ — это борелевская функция от η, то и h является борелевской функцией от η.

Теперь возьмём произвольное борелевское множество B и воспользуемся интегральным свойством:

$$\mathsf{E}[\xi_n \mathbf{1}_B(\eta)] = \mathsf{E}[\mathsf{E}[\xi_n \mid \eta]\mathbf{1}_B(\eta)].$$

Теперь заметим, что $\xi_n \mathbf{1}_B(\eta) \uparrow \xi \mathbf{1}_B(\eta)$ и $\mathsf{E}[\xi_n \mid \eta]\mathbf{1}_B(\eta) \uparrow h(\eta)\mathbf{1}_B(\eta)$. Тогда по теореме о монотонной сходимости

$$\mathsf{E}[\xi \mathbf{1}_B(\eta)] = \mathsf{E}[h(\eta)\mathbf{1}_B(\eta)].$$

А это и означает, что $\mathsf{E}[\xi_n \mid \eta] \uparrow \mathsf{E}[\xi \mid \eta]$.

Второй же пункт доказывается по аналогии с доказательством теоремы Лебега для обычного математического ожидания. □

Свойство 8. *Если* δ *— это борелевская функция от* η, *то* $\mathsf{E}[\xi\delta \mid \eta] = \delta\, \mathsf{E}[\xi \mid \eta]$.

Доказательство. Пусть $\delta = \mathbf{1}_A(\eta)$, где A — некоторое борелевское множество. Заметим, что $\delta\, \mathsf{E}[\xi \mid \eta]$ — это борелевская функция от η. Проверим интегральное свойство:

$$\begin{aligned}\mathsf{E}[\delta\xi\mathbf{1}_B(\eta)] &= \mathsf{E}[\xi\mathbf{1}_A(\eta)\mathbf{1}_B(\eta)] = \mathsf{E}[\xi\mathbf{1}_{A\cap B}(\eta)] = \mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}_{A\cap B}(\eta)]\\ &= \mathsf{E}[\mathsf{E}[\xi \mid \eta]\mathbf{1}_A(\eta)\mathbf{1}_B(\eta)] = \mathsf{E}[\delta\, \mathsf{E}[\xi \mid \eta]\mathbf{1}_B(\eta)].\end{aligned}$$

Пользуясь линейностью матожидания, интегральное свойство будет выполнено и для простых функций вида $\delta = \sum c_k \mathbf{1}_{A_k}(\eta)$.

В общем случае для любой $\delta = f(\eta)$ построим последовательность простых функций $\delta_n = f_n(\eta)$ такую, что $\delta_n \xrightarrow{\text{п.н.}} \delta$ и $|\delta_n| \leqslant |\delta|$. Тогда по свойству 7 получаем, что $\mathsf{E}[\delta_n\xi \mid \eta] \xrightarrow{\text{п.н.}} \mathsf{E}[\delta\xi \mid \eta]$. Однако

$$\mathsf{E}[\delta_n\xi \mid \eta] = \delta_n\, \mathsf{E}[\xi \mid \eta] \xrightarrow{\text{п.н.}} \delta\, \mathsf{E}[\xi \mid \eta].$$

Следовательно, $\mathsf{E}[\xi\delta \mid \eta] = \delta\, \mathsf{E}[\xi \mid \eta]$ (почти наверное, конечно). □

1.20.3 Условные распределения и плотности

Определение 79. *Условным математическим ожиданием при условии* $\eta = y$, $\eta, y \in \mathbb{R}^n$ называется $\mathsf{E}[\xi \mid \eta = y]$.

Вообще говоря, $\mathsf{E}[\xi \mid \eta = y]$ — это борелевская функция $\varphi(y)$ такая, что для любого $B \in \mathcal{B}(\mathbb{R}^n)$

$$\mathsf{E}[\xi \mathbf{1}_B(\eta)] = \mathsf{E}[\varphi(\eta)\mathbf{1}_B(\eta)].$$

Отсюда можно сделать простой вывод: $\mathsf{E}[\xi \mid \eta = y] = \varphi(y) \iff \mathsf{E}[\xi \mid \eta] = \varphi(\eta)$.

Свойства у такого условного математического ожидания похожи на свойства для обычного условного: например, выполнены линейность, сохранение относительного порядка и теорема Лебега.

Впрочем, мы не имеем представления о том, как его вычислять. Для этого введём условные распределение и плотность.

Определение 80. *Условным распределением* случайной величины ξ при условии, что $\eta = y$ назовём функцию $\mathsf{P}(\xi \in B \mid \eta = y) \equiv \mathsf{E}[\mathbf{1}_B(\xi) \mid \eta = y]$, рассматриваемую, как функцию от $B \in \mathcal{B}(\mathbb{R}^n)$ при фиксированном $y \in \mathbb{R}^k$.

Утверждение. При фиксированном $y \in \mathbb{R}^k$ с вероятностью 1 (по распределению η) условное распределение есть распределение вероятностей на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Определение 81. Если условное распределение имеет плотность $p_{\xi|\eta}(x \mid y)$, то назовём его *условной плотностью ξ относительно η*. То есть для любого $B \in \mathcal{B}(\mathbb{R}^n)$

$$\mathsf{P}(\xi \in \mathcal{B} \mid \eta = y) = \int\limits_B p_{\xi|\eta}(x \mid y)\,\mathrm{d}x.$$

Теперь сформулируем две теоремы (пока что без доказательства), которые и позволяют нам считать условное математическое ожидание.

Теорема 70 (о вычислении условного математического ожидания)**.** *Пусть ξ — случайная величина, а $f(x)$ — некоторая борелевская функция. Если $\mathsf{E}[|f(\xi)|] < +\infty$ и существует плотность $p_{\xi|\eta}(x \mid y)$, то*

$$\mathsf{E}[f(\xi) \mid \eta = y] = \int\limits_{\mathbb{R}} f(x) p_{\xi|\eta}(x \mid y)\,\mathrm{d}x.$$

Теорема 71. *Пусть ξ и η таковы, что есть совместная плотность $p_{\xi,\eta}(x, y)$. Тогда существует условная плотность $p_{\xi|\eta}(x \mid y)$ и она равна*

$$p_{\xi|\eta}(x \mid y) = \begin{cases} \dfrac{p_{\xi,\eta}(x, y)}{p_\eta(y)}, & p_\eta(y) > 0, \\ 0, & p_\eta(y) = 0. \end{cases}$$

Рассмотрим на примере, как их использовать.

Задача 1. *Пусть ξ и η — независимые случайные величины, причём $\xi, \eta \sim \mathrm{Exp}(1)$. Найти $\mathsf{E}[\xi^2 \mid \xi + \eta]$.*

Решение. Для начала найдём плотность $\xi + \eta$. По формуле свёртки:

$$\begin{aligned} p_{\xi+\eta}(x) &= \int_{-\infty}^{+\infty} e^{-(x-y)} e^{-y} \mathbf{1}_{(0,+\infty)}(x - y) \mathbf{1}_{(0,+\infty)}(y)\,\mathrm{d}y \\ &= \mathbf{1}_{(0,+\infty)}(x) \int_0^x e^{-x}\,\mathrm{d}y = x e^{-x} \mathbf{1}_{(0,+\infty)}(x). \end{aligned}$$

Теперь найдём совместную плотность распределения ξ и $\xi + \eta$. Для этого распишем вероятностную меру:

$$\mathsf{P}((\xi, \xi + \eta) \in B) = \iint\limits_{(x,x+y) \in B} e^{-x} e^{-y} \mathbf{1}_{(0,+\infty)}(x) \mathbf{1}_{(0,+\infty)}(y)\,\mathrm{d}x\,\mathrm{d}y$$

Сделаем замену переменной:

$$\begin{cases} u = x \\ v = x + y \end{cases} \implies \begin{cases} x = u \\ y = u - v \end{cases} \implies \mathbf{J}(u, v) = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}$$

Тогда $|\det \mathbf{J}(u,v)| = 1$ и

$$\mathsf{P}((\xi, \xi+\eta) \in B) = \iint\limits_{(u,v)\in B} e^{-v} \mathbf{1}_{\{0<u<v\}} \,\mathrm{d}u\,\mathrm{d}v.$$

Следовательно, $p_{\xi,\xi+\eta}(x,y) = e^{-y}\mathbf{1}_{\{0<x<y\}}$. Теперь посчитаем $p_{\xi|\eta}(x \mid y)$. Так как $\xi+\eta > 0$ с вероятностью 1, то:

$$p_{\xi|\eta}(x \mid y) = \frac{e^{-y}\mathbf{1}_{\{0<x<y\}}}{ye^{-y}\mathbf{1}_{(0,+\infty)}(y)} = \frac{1}{y}\mathbf{1}_{\{0<x<y\}}.$$

Недолго думая, считаем условное матожидание по формуле:

$$\mathsf{E}\big[\xi^2 \mid \xi+\eta = y\big] = \int_0^y \frac{x^2\,\mathrm{d}x}{y} = \frac{y^2}{3} \implies \mathsf{E}\big[\xi^2 \mid \xi+\eta\big] = \frac{(\xi+\eta)^2}{3}.$$

$\square$