

ЗАБАВНЫЕ ФАКТЫ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Источник: Чернова Н.И., Теория вероятностей

Авторы заметок: Хоружий Кирилл

От: 17 июня 2021 г.

Содержание

1	Основные понятия теории вероятностей	3
1.1	Элементы комбинаторики	3
1.2	События и операции над ними	3
1.3	Дискретное пространство элементарных исходов	3
1.4	Дискретное пространство элементарных исходов	4
1.5	Геометрическая вероятность	5
2	Аксиоматика теории вероятностей	5
2.1	Алгебра и σ -алгебра событий	5
2.2	Мера и вероятностная мера	5
3	Условная вероятность и независимость	6
3.1	Условная вероятность	6
3.2	(\mathfrak{B}) Операции с вероятностями	7
3.3	Независимость событий	7
3.4	Формула полной вероятности	7
3.5	Формула Байеса	7
4	Схема Бернулли	8
4.1	Распределение числа успехов в n испытаниях	8
4.2	Номер первого успешного испытания	8
4.3	Независимые испытания с несколькими исходами	8
4.4	Теорема Пуассона для схемы Бернулли	9
5	Случайные величины и их распределения	9
5.1	Случайные величины	9
5.2	Распределения случайных величин	9
5.3	Функция распределения	10
5.4	(3) Примеры дискретных распределений	10
5.5	(3) Примеры абсолютно непрерывных распределений	11
5.6	Свойства функций распределения	12
5.7	Свойства нормального распределения	12
6	Преобразования случайных величин	13
6.1	Измеримость функций от случайных величин	13
6.2	Распределения функций от случайных величин	13
7	Многомерные распределения	13
7.1	Совместное распределение	13
7.2	Типы многомерных распределений	14
7.3	Примеры многомерных распределений	14
7.4	Независимость случайных величин	14
7.5	Функции от двух случайных величин	15

8	Числовые характеристики распределений	15
8.1	Математическое ожидание случайной величины	15
8.2	Свойства математического ожидания	15
8.3	Дисперсия и моменты старших порядков	15
8.4	Свойства дисперсии	16
8.5	Математические ожидания и дисперсии стандартных распределений	16
8.6	Другие числовые характеристики распределений	16
8.7	Производящие функции	17
8.8	Вычисление моментов через производящие функции	17
9	Числовые характеристики зависимости	18
9.1	Ковариация двух случайных величин	18
9.2	Коэффициент корреляции	18
10	Характеристические функции	19
10.1	Определение и примеры	19
10.2	Свойства характеристических функций	19
11	Статистика	21

КОНТРОЛЬНАЯ РАБОТА №1 ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Автор работы: Хоружий Кирилл

От: 17 июня 2021 г.

1 Основные понятия теории вероятностей

1.1 Элементы комбинаторики

Для начала подружмся с комбинаторикой, взяв некоторую её проекцию на теорвер

Thr 1.1. Пусть множества $A = \{a_1, \dots, a_k\}$ состоит из k элементов, а множество $B = \{b_1, \dots, b_m\}$ – из m элементов. Тогда можно образовать равно $k \cdot m$ пар (a_i, b_j) .

Thr 1.2. Общее количество различных наборов при выборе k элементов из n **без** возвращения и с учётом порядка равняется

$$A_n^k = n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!},$$

где A_n^k называется числом размещений из n элементов по k элементов.

Thr 1.3. Общее количество различных наборов при выборе k элементов из n **без** возвращения и **без** учета порядка равняется

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

где число C_n^k называется числом сочетаний из n элементов по k элементов.

Thr 1.4. Общее количество различных наборов при выборе k элементов из n с возвращением и без учёта порядка равняется

$$C_{n+k-1}^k = C_{n+k-1}^{n-1}.$$

1.2 События и операции над ними

Def 1.5. Пространством элементарных исходов называют множество Ω , содержащее все возможные взаимоисключающие результаты данного случайного эксперимента. Элементы множества Ω называются *элементарными исходами* и обозначаются ω .

Def 1.6. Событиями называются подмножества Ω . Говорят, что *произошло событие* A , если эксперимент завершился одним из элементарных исходов, входящих в множество A .

Вообще в силу таких определений события и множества оказываются очень похожими, так что определены операции *объединения*, *пересечения*, *дополнения*, а также взятия *противоположного* $\bar{A} = \Omega \setminus A$. Также можно выделить достоверное событие Ω и невозможное \emptyset .

События A и B называются *несовместными*, если они не могут произойти одновременно: $A \cap B = \emptyset$. События A_1, \dots, A_n называются *попарно несовместными*, если несовместны любые два из них: $A_i \cap A_j = \emptyset$, $\forall i \neq j$. Говорят, что событие A *влечет* событие B ($A \subseteq B$), если $A \Rightarrow B$.

1.3 Дискретное пространство элементарных исходов

Пространство элементарных исходов назовём дискретным, если множество Ω конечно или счётно: $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$.

Def 1.7. Сопоставим каждому элементарному исходу ω_i число $p_i \in [0, 1]$ так, чтобы $\sum p_i = 1$. Вероятностью события A называют число

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

где в случае $A = \emptyset$ считаем $P(A) = 0$.

Def 1.8 (Классическое определение вероятности). Говорят, что эксперимент описывается *классической вероятностной моделью*, если пространство его элементарных исходов состоит из конечного числа равновероятных исходов. Для любого события верно, что

$$P(A) = \frac{\text{card } A}{\text{card } \Omega}. \quad (1.1)$$

Эту формулу называют *классическим определением вероятности*.

Тут стоит вспомнить три схемы из модели с урнами: схема выбора с возвращением и с учётом порядка (n^k), выбора без возвращения и с учётом порядка (A_n^k), а также выбора без возвращения и без учёта порядка (C_n^k), описываются классической вероятностной моделью. А вот схема выбора с возвращением и без учёта порядка уже не описывается классической вероятностью.

Пример с гипергеометрическим распределением

Из урны, в которой K белых и $N - K$ чёрных шаров, наудачу и без возвращения вынимают n шаров, где $n \leq N$. Термин «наудачу» означает, что появление любого набора из n шаров равновозможно. Найти вероятность того, что будет выбрано k белых и $n - k$ чёрных шаров.

Результат – набор из n шаров. Общее число $\text{card } \Omega = C_N^n$. Пусть A_k – событие, состоящее в том, что в наборе окажется k белых и $n - k$ чёрных. Есть ровно C_K^k способов выбрать k белых шаров из K , и C_{N-K}^{n-k} способов выбрать $n - k$ чёрных шаров из $N - K$. Тогда $\text{card } A_k = C_K^k C_{N-K}^{n-k}$,

$$P(A_k) = \frac{\text{card } A_k}{\text{card } \Omega} = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

Этот набор вероятностей называется *гипергеометрическим распределением* вероятностей.

1.4 Дискретное пространство элементарных исходов

Пространство элементарных исходов назовём дискретным, если множество Ω конечно или счётно: $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$.

Def 1.9. Сопоставим каждому элементарному исходу ω_i число $p_i \in [0, 1]$ так, чтобы $\sum p_i = 1$. Вероятностью события A называют число

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

где в случае $A = \emptyset$ считаем $P(A) = 0$.

Def 1.10 (Классическое определение вероятности). Говорят, что эксперимент описывается *классической вероятностной моделью*, если пространство его элементарных исходов состоит из конечного числа равновероятных исходов. Для любого события верно, что

$$P(A) = \frac{\text{card } A}{\text{card } \Omega}. \quad (1.2)$$

Эту формулу называют *классическим определением вероятности*.

Тут стоит вспомнить три схемы из модели с урнами: схема выбора с возвращением и с учётом порядка (n^k), выбора без возвращения и с учётом порядка (A_n^k), а также выбора без возвращения и без учёта порядка (C_n^k), описываются классической вероятностной моделью. А вот схема выбора с возвращением и без учёта порядка уже не описывается классической вероятностью.

Пример с гипергеометрическим распределением

Из урны, в которой K белых и $N - K$ чёрных шаров, наудачу и без возвращения вынимают n шаров, где $n \leq N$. Термин «наудачу» означает, что появление любого набора из n шаров равновозможно. Найти вероятность того, что будет выбрано k белых и $n - k$ чёрных шаров.

Результат – набор из n шаров. Общее число $\text{card } \Omega = C_N^n$. Пусть A_k – событие, состоящее в том, что в наборе окажется k белых и $n - k$ чёрных. Есть ровно C_K^k способов выбрать k белых шаров из K , и C_{N-K}^{n-k} способов выбрать $n - k$ чёрных шаров из $N - K$. Тогда $\text{card } A_k = C_K^k C_{N-K}^{n-k}$,

$$P(A_k) = \frac{\text{card } A_k}{\text{card } \Omega} = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

Этот набор вероятностей называется *гипергеометрическим распределением* вероятностей.

1.5 Геометрическая вероятность

Def 1.11. Пусть некоторая область $\Omega \subset \mathbb{R}^k$ такая, что $\mu(\Omega)$ конечна. Пусть эксперимент состоит из равновероятного выбора случайной точки в области Ω . *Геометрическое определение вероятности:*

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

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

2 Аксиоматика теории вероятностей

2.1 Алгебра и σ -алгебра событий

Def 2.1. Множество \mathcal{A} , элементами которого являются некоторые подмножества Ω называют *алгеброй*, если оно удовлетворяет следующим условиям:

- A1) $\Omega \in \mathcal{A}$ (алгебра содержит достоверные события);
- A2) если $A \in \mathcal{A}$, то $\bar{A} \in \mathcal{A}$ (вместе с любым множеством алгебра содержит противоположное к нему);
- A3) если $A \in \mathcal{A}$ и $B \in \mathcal{A}$, то $A \cup B \in \mathcal{A}$ (вместе с любыми двумя множествами алгебра содержит их объединение).

Вообще из A1 и A2 следует, что $\emptyset = \bar{\Omega} \in \mathcal{A}$. Пункт A3 экстраполируется на любой конечный набор. Кстати, объединение можно заменить (в силу закона де Моргана) на пересечение:

$$xy \in \mathcal{A} \quad \Leftrightarrow \quad \overline{xy} \in \mathcal{A} \quad \Leftrightarrow \quad \bar{x} + \bar{y} \in \mathcal{A}.$$

Thr 2.2 (закон де Моргана). *Для множеств x, y верно, что*

$$\overline{x + y} = \bar{x} \cdot \bar{y}, \quad \overline{xy} = \bar{x} + \bar{y},$$

где $xy = x \cap y$, $x + y = x \cup y$.

В случае счётного пространства элементарных исходов A3 алгебры оказывается недостаточно, так приходим к σ -алгебре:

Def 2.3. Множество \mathcal{F} , элементами которого являются некоторые подмножества Ω называется σ -алгеброй, если выполнены следующий условия:

- S1) $\Omega \in \mathcal{F}$ (алгебра содержит достоверные события);
- S2) если $A \in \mathcal{F}$, то $\bar{A} \in \mathcal{F}$ (вместе с любым множеством алгебра содержит противоположное к нему);
- S3) если $\{A_i\} \in \mathcal{F}$, то $\cup_i A_i \in \mathcal{F}$ (вместе с любым *счётным* набором событий σ -алгебра содержит их объединение).

Def 2.4. Минимальной σ -алгеброй, содержащей набор множеств \mathcal{U} , называется пересечение всех σ -алгебр, содержащих \mathcal{U} .

Def 2.5. Минимальная σ -алгебра, содержащая множество \mathcal{U} всех интервалов на вещественной прямой называется *борелевской сигма-алгеброй* в \mathbb{R} и обозначается $\mathfrak{B}(\mathbb{R})$.

Итак, оказался определен специальный класс \mathcal{F} подмножеств Ω , названный σ -алгеброй событий. Применение счетного числа любых операций к множествам из \mathcal{F} снова дает множество из \mathcal{F} . *Событиями* будем называть только множества $A \in \mathcal{F}$.

2.2 Мера и вероятностная мера

Def 2.6. Пусть Ω – некоторое непустое множество \mathcal{F} – σ -алгебра его подмножеств. Функция

$$\mu: \mathcal{F} \mapsto \mathbb{R} \cap [0, +\infty) \cup \{+\infty\}$$

называется *мерой* на (Ω, \mathcal{F}) , если она удовлетворяет условиям

- $\mu(1)$ $\mu(A) \geq 0$ для любого множества $A \in \mathcal{F}$;

$\mu_2) \forall$ счетного $\{A_i\} \in \mathcal{F}$ таких, что $A_i \cap A_j = \emptyset, \forall i \neq j$ мера их объединения равна сумме их мер:

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

Последнее свойство называют *счётное аддитивностью* или σ -аддитивностью меры.

Thr 2.7 (свойство непрерывности меры). Пусть дана убывающая последовательность $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ множеств из \mathcal{F} , причем $\mu(B_1) < \infty$. Пусть $B = \bigcap_{i=1}^{\infty} B_i$. Тогда $\mu(B) = \lim_{n \rightarrow \infty} \mu(B_n)$.

Def 2.8. Пусть Ω – непустое множество, \mathcal{F} – σ -алгебра его подмножеств. Мера $\mu: \mathcal{F} \mapsto \mathbb{R}$ называется *нормированной*, если $\mu(\Omega) = 1$. Другое название нормированной меры – *вероятность*.

Def 2.9. Пусть Ω – пространство элементарных исходов, \mathcal{F} – σ -алгебра его подмножеств (событий). *Вероятностью* или *вероятностной мерой* на (Ω, \mathcal{F}) называется функция

$$P: \mathcal{F} \mapsto \mathbb{R}$$

обладающая свойствами

P1) $P(A) \geq 0$ для любого события $A \in \mathcal{F}$;

P2) для любого счётного набора *попарно несовместных* событий $\{A_i\} \in \mathcal{F}$ имеет равенство

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

P3) вероятность достоверного события равна единице: $P(\Omega) = 1$.

Свойства (P1) – (P3) называют *аксиомами вероятности*.

Def 2.10. Тройка $\langle \Omega, \mathcal{F}, P \rangle$, в которой Ω – пространство элементарных исходов, \mathcal{F} – σ -алгебра его подмножеств и P – вероятная мера на \mathcal{F} , называется *вероятностным пространством*.

Вообще, для вероятности верны следующие свойства

1. $P(\emptyset) = 0$.
2. Для любого конечного набора попарно несовместных событий $A_1, \dots, A_n \in \mathcal{F}$ имеет место равенство $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$.
3. $P(\bar{A}) = 1 - P(A)$.
4. Если $A \subseteq B$, то $P(B \setminus A) = P(B) - P(A)$.
5. $A \subseteq B$, то $P(A) \leq P(B)$.
6. $P(A_1 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i)$.

И это всё, конечно, хорошо, но если мы хотим что-то посчитать, то

Thr 2.11 (Формула включения-исключения). Для вероятности, в частности для двух событий, верно, что

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

и, обобщая, для объединения n множеств

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

3 Условная вероятность и независимость

3.1 Условная вероятность

Def 3.1. Условной вероятностью события A при условии, что произошло событие B , называется число

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

которое само собой определено только при $P(B) > 0$.

Thr 3.2. Если $P(B) > 0$ и $P(A) > 0$, то

$$P(A \cap B) = P(B) P(A|B) = P(A) P(B|A).$$

Thr 3.3. Для любых событий A_1, \dots, A_n верно равенство:

$$P(A_1 \dots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 A_2) \cdot \dots \cdot P(A_n|A_1 \dots A_{n-1}),$$

если все участвующие в нём условные вероятности определены.

3.2 (B) Операции с вероятностями

Thr 3.4. Вероятность произведения (совмещения) двух зависимых событий равна произведению вероятностей одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило:

$$P(AB) = P(A) P_A(B) = P(A) P(B|A).$$

Con 3.5. Для конечного числа зависимых событий верна формула:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_k) = P(A_1) P_{A_1}(A_2) P_{A_1, A_2}(A_3) \dots P_{A_1 \dots A_{k-1}}(A_k).$$

3.3 Независимость событий

Def 3.6. События A и B называются *независимыми*, если $P(A \cap B) = P(A) P(B)$.

Из этого определения вытекают следующие леммы.

Lem 3.7. Пусть $P(B) > 0$. Тогда события A и B независимы тогда и только, когда $P(A|B) = P(A)$.

Lem 3.8. Пусть A и B несовместны. Тогда независимыми они будут только в том случае, если $P(A) = 0$ или $P(B) = 0$.

Другими словами несовместные события не могут быть независимыми. Зависимость между ними – просто причинно-следственная: если $A \cap B = \emptyset$, то $A \subseteq \bar{B}$, т.е. при выполнении A события B не происходит.

Lem 3.9. Если события A и B независимы, то независимы и события A и \bar{B} , \bar{A} и B , \bar{A} и \bar{B} .

Def 3.10. События A_1, \dots, A_n называются *независимыми в совокупности*, если для любого $1 \leq k \leq n$ и любого набора различных меж собой индексов $1 \leq i_1 < \dots < i_k \leq n$ имеет место равенство

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k}).$$

3.4 Формула полной вероятности

Def 3.11. Конечный или счётный набор попарно несовместных событий $\{H_i\}$ таких, что $P(H_i) > 0 \forall i$ и $\cup_i H_i = \Omega$, называется *полной группой событий* или разбиением пространства Ω . Также события, образующие полную группу событий, часто называют *гипотезами*.

При подходящем выборе гипотез для любого события A могут быть сравнительно просто вычислены $P(A|H_i)$ и, собственно, $P(H_i)$. Как посчитать вероятность события A ?

Thr 3.12 (формула полной вероятности). Пусть дана полная группа событий $\{H_i\}$. Тогда вероятность любого события A может быть вычислена по формуле

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

3.5 Формула Байеса

Thr 3.13 (формула Байеса). Пусть $\{H_i\}$ – полная группа событий, и A – некоторое событие, $P(A) > 0$. Тогда условная вероятность того, что имело место событие H_k , если в результате эксперимента наблюдалось событие A , может быть вычислена по формуле

$$P(H_k|A) = \frac{P(H_k) \cdot P(A|H_k)}{\sum_{i=1}^{\infty} P(H_i) \cdot P(A|H_i)}. \quad (3.1)$$

Def 3.14. Вероятности $P(H_i)$, вычисленные заранее, до проведения эксперимента, называют¹ *априорными* вероятностями. Условные вероятности $P(H_i|A)$ называют² *апостериорными* вероятностями.

Формула Байеса позволяет переоценить заранее известные вероятности после того, как получено знание о результате эксперимента. Эта формула находит многочисленные применения в экономике, статистике, социологии и т.п.

4 Схема Бернулли

4.1 Распределение числа успехов в n испытаниях

Def 4.1. *Схемой Бернулли* называется последовательность независимых в совокупности испытаний, в каждом из которых возможны лишь два исхода – «успех» и «неудача», при этом успех \checkmark в одном испытании происходит с вероятностью $p \in (0, 1)$, а неудача \times с вероятностью $q = 1 - p$.

В испытаниях схемы Бернулли независимость в совокупности испытаний означает, что при любом n независимы в совокупности события успехов в каждом событии.

Эти события принадлежат одному и тому же пространству элементарных исходов, полученному декартовым произведением бесконечного числа двухэлементных множеств $\{\checkmark, \times\}$:

$$\Omega = \{(a_1, a_2, \dots, a_n) \mid a_i \in \{\checkmark, \times\}, n \in \mathbb{Z}_+\}.$$

Далее количество успехов для n испытаний схемы Бернулли будем называть ν_n . Заметим, что $\nu_n \in \mathbb{Z}_+ \cap [0, n]$.

Thr 4.2 (формула Бенулли). *При любом $k = 0, 1, \dots, n$ имеет место равенство:*

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

Def 4.3 (\mathfrak{D}). Набор чисел $\{C_n^k p^k q^{n-k}, k = 0, 1, \dots, n\}$ называется *биномиальным* распределением.

4.2 Номер первого успешного испытания

Далее, для схемы Бернулли, введем величину $\tau \in \mathbb{Z}_+ \cap [1, +\infty)$ равную номеру первого успешного испытания.

Thr 4.4. *Вероятность того, что первый успех произойдёт в испытании с номером $k \in \mathbb{N} \cap [1, +\infty)$, равна*

$$P(\tau = k) = pq^{k-1}.$$

Def 4.5 (\mathfrak{D}). Набор чисел $\{pq^{k-1} \mid k = 1, 2, \dots\}$ называется *геометрическим* распределением вероятностей.

Thr 4.6 («Нестарение» геометрического распределения). *Пусть $P(\tau = k) = pq^{k-1} \forall k \in \mathbb{N}$. Тогда для любых неотрицательных целых n и k имеет место равенство:*

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

Другими название – свойство отсутствия последствия.

4.3 Независимые испытания с несколькими исходами

Теперь рассмотрим схему независимых испытаний независимых испытаний уже не с двумя, а с большим количеством возможных результатов в каждом испытании.

Пусть возможны m исходов, i -й исход в одном испытании случается с вероятностью p_i , где $\sum_i p_i = 1$. Через $P(n_1, \dots, n_m)$ обозначим вероятность того, что в n независимых испытаниях первый исход случится n_1 раз, \dots , m -исход – n_m раз.

Thr 4.7. *Для любого n и любых неотрицательных целых чисел $\{n_i\}$, сумма которых равна n , верна формула*

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

¹ *a priori* – «до опыта».

² *a posteriori* – «после опыта».

Def 4.8 (\mathfrak{D}). Набор чисел

$$\left\{ \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m} \mid n = 1, 2, \dots \right\}$$

называется *мультиномиальным (полиномиальным) распределением*.

4.4 Теорема Пуассона для схемы Бернулли

Сформулируем теорему о приближенном вычислении вероятности иметь k успехов в большом числе испытаний Бернулли с маленькой вероятностью успеха p .

Thr 4.9 (теорема Пуассона). Пусть $n \rightarrow \infty$ и $p_n \rightarrow 0$ так, что $np_n \rightarrow \lambda > 0$. Тогда для любого $k \geq 0$ вероятность получить k успехов в n испытаниях схемы Бернулли с вероятностью успеха p_n

$$P(\nu_n = k) = C_n^k p_n^k (1 - p_n)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}. \quad (4.1)$$

то есть стремится к величине $\lambda^k e^{-\lambda} / k!$.

Def 4.10 (\mathfrak{D}). Набор чисел

$$\left\{ \frac{\lambda^k}{k!} e^{-\lambda} \mid k = 0, 1, 2, \dots \right\}$$

называется *распределением Пуассона* с параметром $\lambda > 0$.

Для всех этих распределений можно посчитать вектора средних и матрицы ковариации.

5 Случайные величины и их распределения

5.1 Случайные величины

Пусть задано вероятностное пространство $\langle \Omega, \mathcal{F}, P \rangle$.

Def 5.1. Функция $\xi: \Omega \mapsto \mathbb{R}$ называется *случайной величиной*, если для любого борелевского множества $B \in \mathfrak{B}(\mathbb{R})$ множество $\xi^{-1}(B)$ является событием, т.е. принадлежит σ -алгебре \mathcal{F} .

Множество $\xi^{-1}(B) = \{\omega \mid \xi(\omega) \in B\}$, состоящее из элементарных исходов ω , называется *полным прообразом множества B* . Можно немного другим способом сформулировать требования к величине:

Def 5.2. Функция $\xi: \Omega \mapsto \mathbb{R}$ называется случайной величиной, если для любых вещественных $a < b$ множество

$$\{\omega: \xi(\omega) \in (a, b)\} \in \mathcal{F}$$

принадлежит σ -алгебре.

5.2 Распределения случайных величин

Def 5.3. *Распределением* случайной величины ξ называется *вероятностная мера* $\mu(B) = P(\xi \in B)$ на множестве борелевских подмножеств \mathbb{R} .

Можно представить себе распределение случайной величины ξ как соответствие между множествами $B \in \mathfrak{B}(\mathbb{R})$ и вероятностями $P(\xi \in B)$.

Def 5.4. Если две функции ξ и η отличаются на множестве меры нуль, при этом имеют одинаковое распределение, то говорят, что ξ и η совпадают *почти наверное*: $P(\xi = \eta) = 1$.

Def 5.5. Случайная величина ξ имеет *дискретное* распределение, если существует конечный, или счётный набор чисел $\{a_i\}$ такой, что

$$P(\xi = a_i) > 0 \quad \forall i, \quad \sum_{i=1}^{\infty} P(\xi = a_i) = 1.$$

Значения эти называют *атомами*: ξ имеет атом в точке x , если $P(\xi = x) > 0$.

Если случайная величина ξ имеет дискретное распределение, то для любого $B \subseteq \mathbb{R}$

$$P(\xi \in B) = \sum_{a_i \in B} P(\xi = a_i).$$

Вообще дискретные распределения удобно задавать вероятностной таблицей

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

Def 5.6. Случайная величина ξ имеет *абсолютно непрерывно* распределение, если существует неотрицательная функция $f_\xi(x)$ такая, что для любого борелевского множества B имеет место равенство:

$$P(\xi \in B) = \int_B f_\xi(x) dx.$$

Функцию $f_\xi(x)$ называют *плотностью распределения* величины ξ .

Thr 5.7. Плотность распределения обладает свойствами:

$$(f1) \quad f_\xi(x) \geq 0 \quad \forall x, \quad (f2) \quad \int_{-\infty}^{+\infty} f_\xi(t) dt = 1.$$

Thr 5.8. Если функция f обладает свойствами (f1) и (f2), то существует вероятностное пространство и случайная величина ξ на нём, для которой f является плотностью распределения.

Ещё бывает сингулярное распределение³, смешанные варианты, и всё (*Лебег approved*).

5.3 Функция распределения

Хотелось бы найти некоторый универсальный способ для описания распределения.

Def 5.9. Функцией распределения случайной величины ξ называется функция $F_\xi: \mathbb{R} \mapsto [0, 1]$, при каждом $x \in \mathbb{R}$ равная вероятности случайной величине ξ принимать значения, меньшие x :

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

Далее перечислены основные дискретные и абсолютно непрерывные распределения и найдены их функции распределения.

5.4 (3) Примеры дискретных распределений

Вырожденное распределение. Для удобства вводят *вырожденное распределение*, когда возможен единственный результат при $P(\xi = c) = 1$, тогда функция распределения имеет вид

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

В таком случае принято писать, что $\xi \in I_c$.

Распределение Бернулли. Говорят про *распределение Бернулли* с параметром p ($\xi \in B_p$), если ξ принимает значения 1 и 0 с вероятностью p и $1 - p$ соответственно. Случайная величина ξ с таким распределением равна *числу успехов* в одном испытании схемы Бернулли с вероятностью успеха p . Функция распределения случайной величины ξ тогда равна

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

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

Геометрическое распределение. Говорят, что случайная величина τ имеет геометрическое распределение с параметром $p \in (0, 1)$, и пишут $\tau \in G_p$, если τ принимает значения $k = 1, 2, 3, \dots$ с вероятностями $P(\tau = k) = p(1 - p)^{k-1}$. Случайная величина с таким распределением имеет смысл номера первого успешного испытания в схеме Бернулли с вероятностью успеха p .

³На континуальном множестве меры ноль.

Распределение Пуассона. Говорят, что случайная величина ξ имеет распределение Пуассона с параметром $\lambda > 0$, и пишут $\xi \in \Pi_\lambda$, если ξ принимает значения $k = 0, 1, \dots$ с вероятностью $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Иначе распределение Пуассона называют *распределением числа редких событий*.

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

5.5 (3) Примеры абсолютно непрерывных распределений

Равномерное распределение. Говорят, что ξ имеет равномерное распределение на отрезке $[a, b]$ ($\xi \in U_{a,b}$), если плотность распределения f_ξ постоянна на отрезке $[a, b]$ и равна нулю вне него:

$$f_\xi(x) = \begin{cases} (b-a)^{-1}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Площадь под графиком этой функции равна единице, $f_\xi \geq 0$, так что $f_\xi(x)$ действительно плотность.

Легко теперь посчитать функцию распределения величины ξ :

$$F_\xi(x) = P(\xi < x) = \int_{-\infty}^x f_\xi(t) dt = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b, \end{cases}$$

что вполне логично. График функции распределения и плотности распределения приведен ниже.

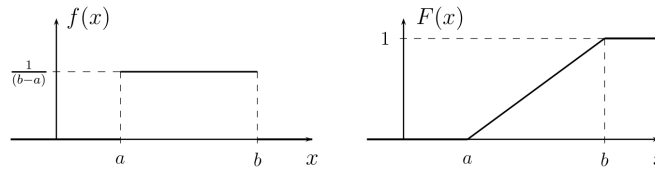


Рис. 1: Плотность и функция распределения $U_{a,b}$

Показательное распределение. Говорят, что ξ имеет показательное (экспоненциальное) распределение с параметром $\alpha > 0$ ($\xi \in E_\alpha$), если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \begin{cases} 0, & x < 0, \\ \alpha e^{-\alpha x}, & x \geq 0. \end{cases}$$

Функция распределения случайной величины ξ непрерывна:

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

Стоит заметить, что показательное распределение является единственным абсолютно непрерывным распределением, для которого выполнено свойство «нестарения» (а-ля геометрическое):

Thr 5.10. Пусть $\xi \in E_\alpha$. Тогда для любых $x, y > 0$ верно, что $P(\xi > x + y \mid \xi > x) = P(\xi > y)$.

Нормальное распределение. Говорят, что ξ имеет *нормальное (гауссовское)* распределение с параметрами a, σ^2 , где $a \in \mathbb{R}$, $\sigma > 0$ ($\xi \in N_{a,\sigma^2}$), если ξ имеет плотность распределения вида

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \quad (5.1)$$

Это действительно функция распределения, ведь вспоминая интеграл Пуассона

$$I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi},$$

нетрудно заменой переменных свести $\int f_\xi(x) dx$ к I .

Def 5.11. Нормальное распределение $N_{0,1}$ называется *стандартным нормальным* распределением.

Для функции распределения нормального закона N_{a,σ^2} далее будет использоваться $\Phi_{a,\sigma^2}(x)$ для функции распределения нормального закона N_{a,σ^2} .

Распределение Коши. Говорят, что ξ имеет распределение Коши с параметрами $a \in \mathbb{R}$, $\sigma > 0$ ($\xi \in C_{a,\sigma}$), если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x-a)^2}, \quad \forall x \in \mathbb{R}.$$

Плотность распределения Коши симметрична относительно $x = a$ и похожа на нормальное, но с более толстыми хвостами на $\pm\infty$. Функция распределения случайной величины ξ с распределением Коши равна

$$F_\xi(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \left(\frac{x-a}{\sigma} \right).$$

Гамма-распределение.

Распределение Парето.

5.6 Свойства функций распределения

Общие свойства функций распределения. Функцией распределения случайной величины ξ мы называли функцию $F_\xi(x) = P(\xi < x)$.

Thr 5.12. Любая функция распределения обладает свойствами

F1) она не убывает;

F2) в пределах $x \rightarrow -\infty$, и $x \rightarrow +\infty$ равна 0 и 1 соответственно;

F3) она в любой точке непрерывна слева.

Thr 5.13. Если функция $F: \mathbb{R} \mapsto [0, 1]$ удовлетворяет свойствам (F1)-(F3), то F есть функция распределения некоторой случайной величины ξ , т.е. найдётся вероятностное пространство $\langle \Omega, \mathcal{F}, P \rangle$ и случайная величина ξ на нём такая, что $F(x) \equiv F_\xi(x)$.

Lem 5.14. В любой точке x_0 разность $F_\xi(x_0 + 0) - F_\xi(x_0)$ равна $P(\xi = x_0)$:

$$F_\xi(x_0 + 0) = F_\xi(x_0) + P(\xi = x_0) = P(\xi \leq x_0).$$

Lem 5.15. Для любой случайной величины ξ

$$P(a \leq \xi < b) = F_\xi(b) - F_\xi(a).$$

Функция распределения дискретного распределения. Как мы помим, функция распределения может быть найдена по таблице распределения, как сумма $F_\xi(x) = P(\xi < x) = \sum_k P(\xi = a_k)$, где $a_k < x$.

Lem 5.16. Случайная величина ξ имеет дискретное распределение тогда и только тогда, когда функция распределения $F_\xi(x)$ имеет в точках a_i скачки с величиной $p_i = P(\xi = a_i) = F_\xi(a_i + 0) - F_\xi(a_i)$, и растёт только за счёт скачков.

Свойства абсолютно непрерывного распределения. Пусть случайная величина ξ имеет абсолютно непрерывное распределение с плотностью $f_\xi(t)$. Тогда функция распределения может быть найдена, как интеграл.

Lem 5.17. Если случайная величина ξ имеет абсолютно непрерывное распределение, то её функция распределения всюду непрерывна. Более того её функция распределения дифференцируема почти всюду: $f_\xi(x) = F'_\xi(x) = d_x F_\xi(x)$.

Функция распределения сингулярного распределения.

Функция распределения смешанного распределения.

5.7 Свойства нормального распределения

Lem 5.18. Для любого $x \in \mathbb{R}$ справедливо соотношение:

$$\Phi_{a,\sigma^2}(x) = \Phi_{0,1} \left(\frac{x-a}{\sigma} \right).$$

Аналогичное утверждение для случайных величин: если $\xi \in N_{a,\sigma^2}$, то $\eta = \frac{\xi-a}{\sigma} \in N_{0,1}$. Более того, если $\xi \in N_{a,\sigma^2}$, то

$$P(x_1 < \xi < x_2) = \Phi_{a,\sigma^2}(x_2) - \Phi_{a,\sigma^2}(x_1) = \Phi_{0,1}\left(\frac{x_2-a}{\sigma}\right) - \Phi_{0,1}\left(\frac{x_1-a}{\sigma}\right).$$

В общем вычисления любых вероятностей для нормального распределения сводятся к вычислению $\Phi_{0,1}(x)$, которое обладает следующими свойствами:

- $\Phi_{0,1}(0) = 0.5$, $\Phi_{0,1}(-x) = 1 - \Phi_{0,1}(x)$.
- Если $\xi \in N_{0,1}$, то для любого $x > 0$, верно что $P(|\xi| < x) = 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1$.

6 Преобразования случайных величин

6.1 Измеримость функций от случайных величин

Пусть на векторном пространстве $\langle \Omega, \mathcal{F}, P \rangle$ задана случайная величина ξ .

Thr 6.1. Пусть ξ – случайная величина, а $g: \mathbb{R} \mapsto \mathbb{R}$ – борелевская функция, т.е. такая, что для всякого борелевского множества B его прообраз $g^{-1}(B)$ есть снова борелевское множество. Тогда $g(\xi)$ – случайная величина.

6.2 Распределения функций от случайных величин

Линейные и монотонные преобразования. Если с дискретными распределениями всё понятно, то с абсолютно непрерывными чуть интереснее, о них дальше и поговорим. Пусть случайная величина ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$. Построим с помощью борелевской функции $g: \mathbb{R} \rightarrow \mathbb{R}$ случайную величину $\eta = g(\xi)$, и найдём плотность распределения (если она существует).

Thr 6.2. Пусть ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$, и постоянная a отлична от нуля. Тогда случайная величина $\eta = a\xi + b$ имеет плотность распределения

$$f_\eta(x) = \frac{1}{|a|} f_\xi\left(\frac{x-b}{a}\right).$$

Квантильное преобразование. Полезно уметь строить случайные величины с заданным распределением по равномерно распределённой случайной величине.

Thr 6.3. Пусть функция распределения $F(x) = F_\xi(x)$ непрерывна. Тогда случайная величина $\eta = F(\xi)$ имеет равномерное на отрезке $[0, 1]$ распределение.

Thr 6.4 (alarm). Пусть $\eta \in U_{0,1}$, а F – произвольная функция распределения. Тогда случайная величина $\xi = F^{-1}(\eta)$ («квантильное преобразование» над η) имеет функцию распределения F .

Как следствие, для $\eta \in U_{0,1}$ верны следующие утверждения:

$$-\frac{1}{\alpha} \ln(1 - \eta) \in E_\alpha, \quad a + \sigma \operatorname{tg}(\pi\eta - \pi/2) \in C_{\alpha,\sigma}, \quad \Phi_{0,1}^{-1}(\eta) \in N_{0,1}.$$

7 Многомерные распределения

7.1 Совместное распределение

Пусть случайные величины ξ_1, \dots, ξ_n заданы на одном вероятностном пространстве $\langle \Omega, \mathcal{F}, P \rangle$.

Def 7.1. Функция

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 < x_1, \dots, \xi_n < x_n) \quad (7.1)$$

называется функцией распределения вектора (ξ_1, \dots, ξ_n) или функцией совместного распределения случайных величин ξ_1, \dots, ξ_n .

7.2 Типы многомерных распределений

Далее рассмотрим два типичных случая, когда совместное распределение либо дискретно, либо непрерывно. Сингулярное распределение не является редкостью: стоит выбрать отрезок на плоскости.

Def 7.2. Случайные величины ξ_1, ξ_2 имеют *дискретное* совместное распределение, если существует конечный или счётный набор пар числе $\{a_i, b_j\}$ такой, что

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

Таблицу, на пересечении i -й строки и j -го столбца которых стоит $P(\xi_1 = a_i, \xi_2 = b_j)$, называют таблицей совместного распределения случайных величин ξ_1 и ξ_2 .

Def 7.3. Случайные величины ξ_1, ξ_2 имеют *абсолютно непрерывное* совместное распределение, если существует неотрицательная функция $f_{\xi_1, \xi_2}(x, y)$ такая, что для любого множества $B \in \mathfrak{B}(\mathbb{R}^2)$ имеет место равенство

$$P((\xi_1, \xi_2) \in B) = \iint_B f_{\xi_1, \xi_2}(x, y) dx dy.$$

Функция $f_{\xi_1, \xi_2}(x, y)$ называется *плотностью совместного распределения* случайных величин ξ_1 и ξ_2 .

Если случайные величины ξ_1 и ξ_2 имеют абсолютно непрерывное совместное распределение, то для любых x_1, x_2 имеет место равенство

$$F_{\xi_1, \xi_2}(x_1, x_2) = P(\xi_1 < x_1, \xi_2 < x_2) = \int_{-\infty}^{x_1} \left(\int_{-\infty}^{x_2} f_{\xi_1, \xi_2}(x, y) dy \right) dx.$$

По функции совместного распределения его плотность находится как смешанная частная производная:

$$f_{\xi_1, \xi_2}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{\xi_1, \xi_2}(x, y)$$

для почти всех (x, y) .

Thr 7.4. Если случайные величины ξ_1 и ξ_2 имеют абсолютно непрерывное совместное распределение с плотностью $f(x, y)$, то ξ_1 и ξ_2 в отдельности также имеют абсолютно непрерывное распределение с плотностями:

$$f_{\xi_1}(x) = \int_{-\infty}^{+\infty} f(x, y) dy; \quad f_{\xi_2}(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

7.3 Примеры многомерных распределений

Многомерное нормальное распределение. Пусть $\Sigma > 0$ – положительно определенная симметричная матрица. Говорят, что вектор (ξ_1, \dots, ξ_n) имеет многомерное нормально распределение $N_{\vec{a}, \Sigma}$ с вектором средних \vec{a} и матрицей ковариации Σ , если плотность совместного распределения $f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ равна

$$f_{\xi}(x) = \frac{1}{\sqrt{\det \Sigma} (\sqrt{2\pi})^n} \exp \left(-\frac{1}{2} (x - \vec{a})^T \cdot \Sigma^{-1} \cdot (x - \vec{a}) \right)$$

В частном случае, когда Σ – диагональная матрица с элементами $\sigma_1^2, \dots, \sigma_n^2$ на диагонали, совместная плотность превращается в произведение плотностей нормальных величин. Вообще это равенство означает независимость величин ξ_1, \dots, ξ_n .

7.4 Независимость случайных величин

Def 7.5. Случайные величины ξ_1, \dots, ξ_n называются *независимыми* (в совокупности), если для любого набора борелевских множеств $B_1, \dots, B_n \in \mathfrak{B}(\mathbb{R})$ имеет место равенство

$$P(\xi_1 \in B_1, \dots, \xi_n \in B_n) = P(\xi_1 \in B_1) \cdot \dots \cdot P(\xi_n \in B_n).$$

Определение независимости можно сформулировать в терминах функций распределения.

Def 7.6. Случайные величины ξ_1, \dots, ξ_n независимы (в совокупности), если для любых x_1, \dots, x_n имеет место равенство

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

Thr 7.7. Случайные величины ξ_1, \dots, ξ_n с абсолютно непрерывными распределениями независимы (в совокупности) тогда и только тогда, когда плотность их совместного распределения существует и равна произведению плотностей, т.е. для любых x_1, \dots, x_n имеет место равенство:

$$f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n).$$

7.5 Функции от двух случайных величин

Пусть ξ_1 и ξ_2 – случайные величины с плотностью совместного распределения $f_{\xi_1, \xi_2}(x_1, x_2)$, и задана борелевская функция $g: \mathbb{R}^2 \mapsto \mathbb{R}$. Требуется найти функцию (и плотность, если повезет) распределения случайной величины $\eta = g(\xi_1, \xi_2)$.

Thr 7.8. Пусть $x \in \mathbb{R}$, и область $D_x \subseteq \mathbb{R}^2$ состоит из точек (u, v) таких, что $g(u, v) < x$. Тогда случайная величина $\eta = g(\xi_1, \xi_2)$ имеет функцию распределения

$$F_\eta(x) = P(g(\xi_1, \xi_2) < x) = P((\xi_1, \xi_2) \in D_x) = \iint_{D_x} f_{\xi_1, \xi_2}(u, v) du dv.$$

Если ξ_1 и ξ_2 независимы, то распределение $g(\xi_1, \xi_2)$ полностью определяется частными распределениями величин ξ_1 и ξ_2 .

Thr 7.9 (формула свёртки). Если случайные величины ξ_1 и ξ_2 независимы и имеют абсолютно непрерывные распределения с плотностями $f_{\xi_1}(u)$ и $f_{\xi_2}(v)$, то плотность распределения суммы $\xi_1 + \xi_2$ существует и равна «свёртке» плотностей f_{ξ_1} и f_{ξ_2} :

$$f_{\xi_1 + \xi_2}(t) = \int_{-\infty}^{+\infty} f_{\xi_1}(u) f_{\xi_2}(t - u) du = \int_{-\infty}^{+\infty} f_{\xi_2}(u) f_{\xi_1}(t - u) du. \quad (7.2)$$

8 Числовые характеристики распределений

8.1 Математическое ожидание случайной величины

Далее будет использовать термин *математического ожидания*, и также можно встретить наименования: *среднее значение*, *первый момент*.

Def 8.1. Математическим ожиданием $E(\xi)$ случайной величины ξ с дискретным распределением называется число

$$E(\xi) = \sum_k a_k p_k = \sum_k a_k P(\xi = a_k),$$

если данный ряд абсолютно сходится, т.е. если $\sum_i |a_i| p_i < +\infty$. В противном случае говорят, что математическое ожидание не существует.

Def 8.2. Математическим ожиданием $E(\xi)$ случайной величины ξ с абсолютно непрерывным распределением с плотностью распределения $f_\xi(x)$ называется число

$$E(\xi) = \int_{-\infty}^{+\infty} x f_\xi(x) dx,$$

если этот интеграл абсолютно сходится, т.е. если $\int |x| f_\xi(x) dx < +\infty$.

8.2 Свойства математического ожидания

Далее всегда предполагается, что матожидание существует.

(E1) Для \forall борелевской $g: \mathbb{R} \mapsto \mathbb{R}$, для дискретного и непрерывного распределения, при существующем E :

$$E g(\xi) = \sum_k g(a_k) P(\xi = a_k), \quad E g(\xi) = \int_{-\infty}^{+\infty} g(x) f_\xi(x) dx.$$

(E3) Матожидание линейно по константам: $E(c\xi) = c E(\xi)$.

(E4) Матожидание суммы любых случайных величин равно сумме их матожиданий: $E(\xi + \eta) = E(\xi) + E(\eta)$.

(E7) Если ξ и η независимы и их матожидания существуют, то $E(\xi\eta) = E(\xi) \cdot E(\eta)$.

8.3 Дисперсия и моменты старших порядков

Def 8.3. Пусть $E|\xi|^k < +\infty$. Число $\nu_k = E\xi^k$ называется *моментом порядка k* , или *k -м моментом* случайной величины ξ , число $E|\xi|^k$ называется *абсолютным k -м моментом*. Число $E[\xi - E(\xi)]^k$ называется *центральный k -м моментом*, $E|\xi - E(\xi)|^2$ – *абсолютным центральным k -м моментом*.

Def 8.4. Число $D(\xi) = E(\xi - E\xi)^2$ (центральный момент второго порядка) называется *дисперсией* случайной величины ξ . Другими словами, это «среднее значение квадрата отклонения случайной величины ξ от своего среднего».

Def 8.5. Число $\sigma = \sqrt{D\xi}$ называют *среднеквадратичным отклонением* случайной величины ξ .

Thr 8.6 (неравенство Йенсена). Пусть вещественнозначная функция g выпукла. Тогда для любой случайной величины ξ с конечным первым моментом верно неравенство

$$Eg(\xi) \geq g(E\xi),$$

где для вогнутых функций знак неравенства меняется на противоположный.

Lem 8.7. Если $E|\xi|^t < \infty$, то для любого $0 < s < t$ верно, что

$$\sqrt[s]{E|\xi|^s} \leq \sqrt[t]{E|\xi|^t}.$$

Также из неравенства Йенсена вытекает ряд удобных неравенств:

$$Ee^\xi \geq e^{E\xi}, \quad \dots$$

8.4 Свойства дисперсии

Во всех свойствах ниже предполагается существование вторых моментов случайных величин.

(D1) Дисперсия может быть вычислена по формуле $D\xi = E\xi^2 - (E\xi)^2$.

$$D\xi = E(\xi - E\xi)^2 = \int a = E\xi \Big/ = E(\xi^2) - 2aE\xi + a^2 = E\xi^2 - (E\xi)^2.$$

(D2) Считая c константной: $D(c\xi) = c^2 D\xi$.

(D3) Дисперсия неотрицательна: $D\xi \geq 0$, более того обращается в ноль, только при $\xi = \text{const}$ почти наверное.

(D4) $D(\xi + c) = D\xi$.

(D5) Если ξ и η независимы, то $D(\xi + \eta) = D\xi + D\eta$. Вообще верна формула

$$D(\xi + \eta) = D\xi + D\eta + 2(E(\xi\eta) - E\xi E\eta). \quad (8.1)$$

(D6) Минимум среднеквадратичного отклонения ξ от точек числовой прямой есть $D\xi$:

$$D\xi = E(\xi - E\xi)^2 = \min_a E(\xi - a)^2.$$

8.5 Математические ожидания и дисперсии стандартных распределений

Посчитаем несколько характерных значений для различных распределений:

Имя	$E\xi$	$E\xi^2$	$D\xi$
$B_{1,p}$	p	p	pq
G_p	$1/p$		qp^{-2}
Π_λ	λ	$\lambda^2 + \lambda$	λ
$U_{a,b}$	$(a+b)/2$	$\frac{1}{3}(a^2 + ab + b^2)$	$\frac{1}{12}(b-a)^2$
$N_{0,1}$	0	1	1
$N_{a,\sigma}$	a		σ^2
E_α	$1/\alpha^1$	$2!/\alpha^2$	$1/\alpha^2$
$C_{0,1}$	\nexists	\nexists	\nexists

8.6 Другие числовые характеристики распределений

Далее кратко познакомимся с другими показателями из статистики.

Def 8.8. Медианой распределения случайной величины ξ называется любое из чисел μ таких, что

$$P(\xi \leq \mu) \geq \frac{1}{2}, \quad P(\xi \geq \mu) \geq \frac{1}{2}.$$

Обобщая, приходим к понятию *квантили* уровня $\delta \in (0, 1)$, так называется решение уравнения $P(x_\delta) = \delta$, где x_δ отрезает площадь δ слева от себя и $1 - \delta$ справа.

Вообще ещё есть такой зоопарк, что квантили уровней кратных 0.01 в прикладной статистике называют *процентилями*, кратных 0.1 – *децилями*, кратных 0.25 – *квартлями*.

Def 8.9. Модой абсолютно непрерывного распределения называют любую точку локального максимума плотности распределения. Для дискретных распределений модой считают любое значение a_i , вероятность которого больше соседних.

Для описания *унимодеальных* распределений используют следующие величины:

Def 8.10. Коэффициентом асимметрии распределения с конечным третьим моментом называют число

$$\beta_1 = E \left(\frac{x - a}{\sigma} \right)^3,$$

где $a = E\xi$, а $\sigma = \sqrt{D\xi}$.

Для симметричных распределений коэффициент асимметрии равен нулю, если $\beta_1 > 0$, то график плотности имеет более крутой наклон слева, и более пологий справа.

Def 8.11. Коэффициентом эксцесса распределения с конечным четвертым моментом называется число

$$\beta_2 = E \left(\frac{\xi - a}{\sigma} \right)^4 - 3,$$

где $a = E\xi$, а $\sigma = \sqrt{D\xi}$.

Для нормального распределения $\beta_2 = 0$, при $\beta_2 > 0$ плотность распределения имеет более острую вершину, чем у нормального распределения.

8.7 Производящие функции

Дискретные величины, рассмотренные ранее, принимают только целые значения $X = 0, 1, \dots$. Нахождение числовых характеристик упрощается, если рассмотреть *производящие функции*.

Def 8.12. Производящей функцией дискретной целочисленной случайной величины ξ с законом распределения $P(\xi = k) = p_k$, где $k = 0, 1, \dots$ называется функция, заданная степенным рядом

$$E(s^\xi) = P(s) = p_0 + p_1 s + p_2 s^2 + \dots, \quad (8.2)$$

который сходится по крайней мере для $|s| \leq 1$.

Thr 8.13. Производящая функция суммы независимых случайных величин ξ и η равна произведению производящих функций слагаемых

$$P_{\xi+\eta}(s) = P_\xi(s) \cdot P_\eta(s). \quad (8.3)$$

Так например для биномиального распределения производящая функция примет вид

$$P(s) = (q + ps)^n.$$

А для геометрического закона распределения

$$P(s) = ps + pqs^2 + pq^2s^3 + \dots = \frac{ps}{1 - qs}.$$

В случае же Пуассона

$$P(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}.$$

Thr 8.14. Сумма независимых случайных величин, распределенных по закону Пуассона, распределена по тому же закону.

Thr 8.15. Для дискретной случайной величины ξ с производящей функцией $P(s)$ выполняются следующие требования:

$$E(\xi) = P'_s(1), \quad D(\xi) = P''_{s,s}(1) + P'_s(1) - [P'_s(1)]^2. \quad (8.4)$$

8.8 Вычисление моментов через производящие функции

Def 8.16. Производящей функцией моментов случайной величины ξ называют математическое ожидание случайной величины $e^{s\xi}$, где s – действительный параметр:

$$\psi_\xi(s) = E(e^{s\xi}). \quad (8.5)$$

Thr 8.17. Если случайная величина ξ имеет начальный момент порядка n , то производящая функция $\psi_\xi(s)$ n раз дифференцируема по s , и для всех $k \leq n$ выполняется соотношение

$$\nu_k = \psi_\xi^{(k)}(0). \quad (8.6)$$

Действительно, разлагая функции моментов в ряд Маклорена, можно получить её разложение в ряд с начальными моментами

$$\psi_\xi(s) = 1 + \nu_1 s + \frac{\nu_2}{2!} s^2 + \dots$$

9 Числовые характеристики зависимости

9.1 Ковариация двух случайных величин

Дисперсия суммы двух случайных величин равна

$$D(\xi + \eta) = D\xi + D\eta + 2(E(\xi\eta) - E(\xi)E(\eta)).$$

Величина $E(\xi\eta) - E\xi E\eta = 0$, если ξ и η независимы, но это верно только в одну сторону, поэтому эту величину используют как «индикатор наличия зависимости» между двумя случайными величинами.

Def 9.1. Ковариацией $\text{cov}(\xi, \eta)$ случайных величин ξ и η называется число

$$\text{cov}(\xi, \eta) = E[(\xi - E\xi)(\eta - E\eta)]. \quad (9.1)$$

Для ковариации справедливы следующие равенства:

$$\text{cov}(\xi, \eta) = E(\xi\eta) - E(\xi)E(\eta); \quad \text{cov}(\xi, \xi) = D(\xi); \quad \text{cov}(\xi, \eta) = \text{cov}(\eta, \xi); \quad \text{cov}(c\xi, \eta) = c \text{cov}(\xi, \eta).$$

Lem 9.2. Дисперсия суммы нескольких случайных величин вычисляется по формуле:

$$D(\xi_1 + \dots + \xi_n) = \sum_{i=1}^n D(\xi_i) + \sum_{i \neq j} \text{cov}(\xi_i, \xi_j) = \sum_{i,j} \text{cov}(\xi_i, \xi_j). \quad (9.2)$$

Если ковариация $\text{cov}(\xi, \eta) \neq 0$, то ξ и η зависимы. Найти совместное распределение бывает сложнее, чем посчитать $E(\xi\eta)$, поэтому, если повезет, и $E(\xi\eta) \neq E(\xi)E(\eta)$, то, не находя совместное распределение, мы обнаружим зависимость ξ и η , не находя их совместного распределения. Это очень хорошо.

Однако есть проблема – ковариация не безразмерно, поэтому большие значения ковариации не говорят о более сильной зависимости. Хотелось бы как-то отнормировать $\text{cov}(\xi, \eta)$, получив «безразмерную» величину. Так мы приходим к коэффициенту корреляции.

9.2 Коэффициент корреляции

Def 9.3. Коэффициентом корреляции $\rho(\xi, \eta)$ случайных величин ξ и η , дисперсии которых существуют и отличны от нуля, называется число

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}}. \quad (9.3)$$

Можно наполнить это достаточно глубоким смыслом. На самом деле это «косинус угла» между двумя элементами $\xi - E\xi$ и $\eta - E\eta$ гильбертова пространства, образованного случайными величинами с нулевым матожиданием и конечным вторым моментом. Пространство нажжено скалярным произведением $\text{cov}(\xi, \eta)$ и «нормой», равной корню из дисперсии, или $\sqrt{\text{cov}(\xi, \xi)}$.

Thr 9.4. Коэффициент корреляции обладает свойствами:

- 1) если ξ и η независимы, то $\rho(\xi, \eta) = 0$;
- 2) всегда $|\rho(\xi, \eta)| \leq 1$;
- 3) $|\rho(\xi, \eta)| = 1$ тогда и только тогда, когда ξ и η почти наверное линейно связаны.

Def 9.5. Стандартизацией случайной величины называется преобразование

$$\hat{\xi} = \frac{\xi - E(\xi)}{\sqrt{D(\xi)}}. \quad (9.4)$$

В терминах стандартизации чуть проще записывается коэффициент корреляции:

$$\rho(\xi, \eta) = E(\hat{\xi} \cdot \hat{\eta}).$$

Def 9.6. Говорят, что ξ и η отрицательно коррелированы, если $\rho(\xi, \eta) < 0$; положительно коррелированы, если $\rho(\xi, \eta) > 0$; некоррелированы, если $\rho(\xi, \eta) = 0$.

Lem 9.7. Для любых случайных величин ξ и η с конечной и ненулевой дисперсией при любых постоянных $a \neq 0$ и b имеет место равенство

$$\rho(a\xi + b, \eta) = \text{sign}(a) \cdot \rho(\xi, \eta). \quad (9.5)$$

Разобрать пример 67 и далее.

10 Характеристические функции

10.1 Определение и примеры

Def 10.1. Функция $\varphi_\xi(t) = E(e^{it\xi})$ вещественной переменной t называется *характеристической функцией* случайной величины ξ .

Например, если характеристическая функция имеет стандартное нормальное распределение, то её характеристическая функция равна

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} e^{-(x-it)^2/2} d(x-it) = e^{-t^2/2}.$$

10.2 Свойства характеристических функций

(Ф1). Характеристическая функция всегда существует: $|\varphi_\xi(t)| = |E e^{it\xi}| \leq 1$.

(Ф2). По характеристической функции однозначно восстанавливается распределение. Например, если модуль характеристической функции интегрируем на всей прямой, то

$$f_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_\xi(t) dt.$$

(Ф3). Характеристическая функция случайной величины $a + b\xi$ связана с характеристической функцией случайной величины ξ равенством

$$\varphi_{a+b\xi}(t) = E e^{it(a+b\xi)} = e^{ita} E(e^{itb\xi}) = e^{ita} \varphi_\xi(tb).$$

(Ф4). Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых: если случайные величины ξ и η независимы, то

$$\varphi_{\xi+\eta}(t) = E(e^{it(\xi+\eta)}) = E(e^{it\xi}) E(e^{it\eta}) = \varphi_\xi(t) \varphi_\eta(t).$$

Собственно, это очень простой и приятный инструмент для доказательства *устойчивости* распределений. **Чем надо было бы и воспользоваться.**

(Ф5). Пусть существует момент порядка $k \in \mathbb{N}$ случайной величины ξ . Тогда характеристическая функция $\varphi_\xi(t)$ непрерывно дифференцируема k раз и её k -я производная в нуле связана с моментом порядка k равенством

$$\varphi_\xi^{(k)}(0) = \left(\frac{d^k}{dt^k} E e^{it\xi} \right) \Big|_{t=0} = (E i^k \xi^k e^{it\xi}) \Big|_{t=0} = i^k E(\xi^k).$$

Lem 10.2. Для случайной величины ξ со стандартным нормальным распределением момент чётного порядка $2k$ равен

$$E(\xi^{2k}) = (2k-1)!! = (2k-1) \cdot (2k-3) \cdot \dots \cdot 3 \cdot 1.$$

Все моменты нечётных порядков существуют и равны нулю.

Как только появились производные высших порядков, самое время разложить функцию в ряд Тейлора:

(Ф6). Пусть существует момент порядка $k \in \mathbb{N}$ случайной величины ξ , тогда характеристическая функция $\varphi_\xi(t)$ в окрестности точки $t = 0$ разлагается в ряд Тейлора

$$\varphi_\xi(t) = \varphi_\xi(0) + \sum_{j=1}^k \frac{t^j}{j!} \varphi_\xi^{(j)}(0) + o(|t^k|) = 1 + \sum_{j=1}^k \frac{i^j t^j}{j!} E(\xi^j) + o(|t^k|).$$

Thr 10.3 (теорема о непрерывно соответствии). *Случайные величины ξ_n слабо сходятся к случайной величине ξ тогда и только тогда, когда для любого t характеристические функции $\varphi_{\xi_n}(t)$ сходятся к характеристической функции $\varphi_\xi(t)$.*

11 Статистика

Гвоорим, что $T(\mathbf{X})$ – достаточная для параметра θ статистика, если $\forall D$

$$P(\mathbf{X} \in D \mid T(\mathbf{X})) \neq P(\theta).$$

Получается, что зная $T(\mathbf{X})$ мы всё узнаем о θ по выборке.

Есть критерий о достаточности статистики.

Thr 11.1 (Критерий факторизации). *Статистика $T(\mathbf{X})$ достаточна для θ тогда и только тогда, когда*

$$f_{\mathbf{X}}(\mathbf{x}, \theta) = g(T(\mathbf{x}), \theta) \cdot h(\mathbf{x}),$$

где $f_{\mathbf{x}}$ – плотность распределения.

\triangle . Пусть известно, что плотность факторизована, тогда вероятность

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} \mid T = t) &= \frac{P(\mathbf{X} = \mathbf{x}, T = t)}{P(T = t)} = I_{T=t} \frac{P(\mathbf{X} = \mathbf{x})}{P(T = t)} = I_{T(\mathbf{x})=t} \frac{g(T(\mathbf{x}), \theta)h(\mathbf{x})}{\sum_{\mathbf{x}: T(\mathbf{x})=t} P(\mathbf{X} = \mathbf{x})} = \\ &= I_{T(\mathbf{x})=t} \frac{g(T(\mathbf{x}), \theta)h(\mathbf{x})}{\sum_{\mathbf{x}: T(\mathbf{x})=t} g(T(\mathbf{x}), \theta)h(\mathbf{x})} = \end{aligned}$$

□