Предмет теории вероятностей и математической статистики

Теория вероятностей изучает случайные явления. Ее методы позволяют понимать и анализировать такие явления.

Основная идея состоит в том, что каждому случайному событию A приписывается некоторое число $\mathsf{P}(A) \in [0,1]$ — его *вероятность*. Это число показывает, насколько правдоподобным является наступление данного события.

Вероятность (вероятностная мера) P может быть введена на основе субъективных представлений или связана с частотой наступления события при многократном повторении эксперимента.

Пример

Какова вероятность выпадения орла при подбрасывании монеты? Если мы предполагаем, что монета симметрична, то $\mathsf{P}(A) = 1/2$. Вместо этого можно много раза подбросить монету и оценить вероятность выпадения орла как частоту наступления этого события.

Пример

Какова вероятность того, что сборная России станет следующим чемпионом мира по футболу? Эксперимент нельзя повторить.

Математическая статистика решает задачу построения вероятностной модели на основе наблюдений (на основе имеющихся данных). Задачи статистики:

- оценка параметров (оценить вероятность p выпадения орла)
- поверка гипотез (верно ли, что монета симметрична)
- анализ зависимостей: регрессия (оценка влияния цены нефти на индекс Доу-Джонса)

- Кардано "Книга об игре в кости", 1526 (опубликована 1663): идеи определения вероятности и закона больших чисел.
- Переписка между Паскалем и Ферма (1654) по поводу ряда задач, в том числе по поводу задачи о разделе ставки (приза), если игра прервана. Идея: делить ставку пропорционально вероятности ее выигрыша, если игра будет продолжена.
- Гюйгенс, "О расчетах в азартных играх", 1657. Идея понятия математического ожидания, использование формул сложения и умножения вероятностей, решение задач, связанных с бросанием костей и разделом ставок.
- Я. Бернулли "Искусство предположений", 1713. Анализ работы Гюйгенса, комбинаторика, применение комбинаторики к "случайным играм и играм в кости", доказательство закона больших чисел.
- Муавр "Учение о случаях" (1718, 1738, 1756), "Аналитическая смесь" (1730). Частный случай центральной предельной теоремы, понятия независимости, математического ожидания, условной вероятности, решение задачи о продолжительности азартной игры.

- Лаплас “Аналитическая теория вероятностей” (1812). Обобщение теоремы Муавра, применении вероятностных методов к теории ошибок наблюдений.
- Пуассон (1781—1840) и Гаусс (1777—1855): предельные теоремы.
- Чебышёв (1821—1894), Марков (1856—1922), Ляпунов (1857—1918): эффективные методы доказательства предельных теорем для сумм независимых произвольно распределенных случайных величин.
- Бернштейн (1880—1968), Мизес (1883—1953), Борель (1871—1956): аксиоматика теории вероятностей.
- **Колмогоров «Основные понятия теории вероятностей» (1933)**: предложена аксиоматика, получившая всеобщее признание и позволившая не только охватить все классические разделы теории вероятностей, но и дать строгую основу для развития ее новых разделов.

References

- Ширяев А.Н. Вероятность I. МНЦМО: М., 2004; II. МНЦМО: М., 2007.
- Bertsekas D.P., Tsitsiklis J.N. Introduction to Probability. Athena Scientific: Belmont, MIT, 2008.
- Tsitsiklis J. Introduction to probability
https://www.youtube.com/watch?v=j9WZyLZCBzs&list=PLmPcD-wiF4Ea_Doghiw3ya6XaLrmGrLUU (2012)
or
https://www.youtube.com/watch?v=1uW3qMFA9Ho&list=PLUl4u3cNGP60hI9ATjSFgLZpbNJ7myAg6 (2018)
- Чернова Н.И. Теория вероятностей. НГУ: 2007.
- Lipschutz S., Lipson M. Probability. 430 fully solved problems. McGraw-Hill: New York, 2011.
- Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Том I. МЦНМО: М., 2018 г.
- Рохлин Д.Б. Основы стохастического анализа. ЮФУ: Ростов-на-Дону, 2019, (гл. 1) (репозиторий ЮФУ).

Модель вероятностного эксперимента

В основе любых математических моделей, учитывающих случайные факторы, лежит модель вероятностного эксперимента. Согласно системе аксиом Колмогорова описание такой модели включает три объекта:

- множество Ω возможных исходов эксперимента (множество элементарных исходов),
- семейство $\mathscr{F}$ подмножеств Ω – семейство событий, которые могут наблюдаться в результате эксперимента,
- вероятность P всех событий из $\mathscr{F}$.

Множество элементарных исходов

- Эксперимент с однократным подбрасыванием монеты: $\Omega = \{H, T\}$. Возможные исходы: орел (Head) и решка (Tail).
- Эксперимент, состоящий в n-кратном подбрасывании монеты.

$$\Omega = \{\omega = (a_1, \ldots, a_n) : a_i = H \text{ или } a_i = T\}.$$

- Следующая цена акции: $\Omega = [0, \infty)$ или $\Omega = [a, b]$, если точно известно, что цена лежит на интервале $[a, b]$.
- Траектория цены акции на интервале времени $[0, T]$: $\Omega = C[0, T]$, если известно, что цена является непрерывной функцией времени.

σ-алгебра событий

Семейство событий $\mathscr{F}$ должно быть *σ-алгеброй*:

- $\emptyset \in \mathscr{F}$;
- из условия $A \in \mathscr{F}$ вытекает, что $A^c := \Omega \backslash A \in \mathscr{F}$;
- из условия $A_i \in \mathscr{F}$, $i \in \mathbb{Z}_+$ вытекает, что $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$.

Для любого семейства множеств $\mathscr{C}$ существует наименьшая σ-алгебра $\sigma(\mathscr{C})$, которая содержит $\mathscr{C}$. Данная σ-алгебра совпадает с пересечением всех σ-алгебр, содержащих $\mathscr{C}$.

Пример

$\Omega = \{H, T\}$, $\mathscr{C} = \{\{H\}, \{T\}\}$,

$$\sigma(\mathscr{C}) = \{\varnothing, \Omega, \{H\}, \{T\}\} = 2^{\Omega}.$$

Пример

$$\Omega = \{\omega = (a_1, \ldots, a_n) : a_i = H \text{ или } a_i = T\}.$$

$$\begin{aligned}
A_{HH} &= \{\omega = (a_1, \ldots, a_n) : a_1 = H, a_2 = H\},\\
A_{HT} &= \{\omega = (a_1, \ldots, a_n) : a_1 = H, a_2 = T\},\\
A_{TH} &= \{\omega = (a_1, \ldots, a_n) : a_1 = T, a_2 = H\},\\
A_{TT} &= \{\omega = (a_1, \ldots, a_n) : a_1 = T, a_2 = T\},
\end{aligned}$$

$$\mathscr{C} = \{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}.$$

$\mathscr{F}_2 = \sigma(\mathscr{C})$ содержит информацию о первых двух подбрасываниях монеты: любое событие, отражающее такую информацию содержится в $\mathscr{F}_2$.

Пример

$\Omega = [0, 1]$,
$$\mathscr{C} = \{(a, b) : a, b \in [0, 1], a < b\},$$
$\mathscr{B}([0, 1]) := \sigma(\mathscr{C})$ — борелевская σ-алгебра.

Пример

$\Omega = C[0, T]$,
$$C_i = \{\omega \in C[0, T] : \omega(t_i) \in (a_i, b_i), a_i < b_i\},$$
$\mathscr{C}$ — семейство C_i, $\sigma(\mathscr{C})$ — цилиндрическая σ-алгебра.

Вероятностная мера

Функция $\mathsf{P} : \mathscr{F} \mapsto [0, 1]$ называется *вероятностной мерой*, если

- $\mathsf{P}(\emptyset) = 0$, $\mathsf{P}(\Omega) = 1$;
- для любой последовательности попарно непересекающихся множеств $A_i \in \mathscr{F}$, $i \in \mathbb{N}$ верно равенство

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

  которое называется свойством счетной аддитивности (σ-аддитивности).

Пример

Ω — конечное или счетное множество, функция $\pi : \Omega \mapsto [0, 1]$ удовлетворяет условию

$$\sum_{\omega \in \Omega} \pi(\omega) = 1.$$

Для любой σ-алгебры $\mathscr{F}$ положим

$$\mathsf{P}(A) = \sum_{\omega \in A} \pi(\omega), \quad A \in \mathscr{F}.$$

Тогда P — вероятностная мера на σ-алгебре $\mathscr{F}$. В частности, если Ω конечно и $\pi(\omega)$ не зависит от ω, то $\pi(\omega) = 1/|\Omega|$,

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

Пример

Рассмотрим эксперимент с n кратным подбрасыванием монеты

$$\Omega = \{\omega = (a_1, \dots, a_n) : a_i = H \text{ или } a_i = T\}, \quad \mathscr{F} = 2^{\Omega}.$$

Положим $\pi(\omega) = p^{\nu(\omega)} q^{n-\nu(\omega)}$, где $\nu(\omega)$ – количество H в последовательности $\omega = (a_1, \dots, a_n)$, $p \in (0,1)$, $q = 1 - p$. Нетрудно проверить, что $\sum_{\omega \in \Omega} \pi(\omega) = 1$.

Данная модель подходит для описания многократного подбрасывания несимметричной монеты с вероятностью выпадения орла равной p.

Пример

$\Omega = \mathbb{N} = \{1, 2, \dots\}$. Положим $\pi(n) = 1/2^n$:

$$\sum_{n=1}^{\infty} \pi(n) = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1.$$

Пусть $\mathscr{F} = 2^{\Omega}$. Найдем вероятность множества четных чисел:

$$\mathsf{P}(\{2, 4, 6, \dots\}) = \mathsf{P}(\{2\}) + \mathsf{P}(\{4\}) + \mathsf{P}(\{6\}) + \dots = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \dots$$

$$= \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{4} \sum_{n=0}^{\infty} \frac{1}{4^n} = \frac{1}{4} \frac{1}{1 - \frac{1}{4}} = \frac{1}{3}.$$

Пример

Конечное семейство множеств $\mathscr{D} = \{D_1, \dots, D_m\}$ называется разбиением, если

$$\bigcup_{i=1}^{m} D_i = \Omega, \quad D_i \cap D_j = \emptyset, \quad i \neq j.$$

Элементы сигма-алгебры $\mathscr{F} = \sigma(\mathscr{D})$ являются конечными объединениями элементов $\mathscr{D}$. При этом D_i называются атомами $\mathscr{F}$. Они обладают тем свойством, что их нельзя раздробить: если $A \subset D_i$, $A \in \mathscr{F}$, то $A = D_i$.

Чтобы задать вероятностную меру на $\mathscr{F}$, достаточно задать ее на $\mathscr{D}$: $\mathsf{P}(D_i) = \pi_i$. Любой элемент $A \in \mathscr{F}$ однозначно представляется в виде

$$A = \bigcup_{i \in I} D_i,$$

и мера P продолжается на $\mathscr{F}$ по формуле

$$\mathsf{P}(A) = \sum_{i \in I} \mathsf{P}(D_i).$$

Пример

Допустим, что в модели многократного подбрасывания несимметричной монеты доступна информация о первых двух подбрасываниях. Данная информация содержится в сигма-алгебре $\mathscr{F}_2$, порожденной разбиением

$$\mathscr{D} = \{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}.$$

Для того, чтобы задать вероятностную меру на $\mathscr{F}_2$ достаточно положить

$$\mathsf{P}(A_{HH}) = p^2, \quad \mathsf{P}(A_{HT}) = \mathsf{P}(A_{TH}) = pq, \quad \mathsf{P}(A_{TT}) = q^2.$$

Теорема Каратеодори о продолжении меры

Пусть P' — вероятностная мера на алгебре $\mathscr{A}$. Имеется в виду, что свойство σ-аддитивности выполняется, если $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$. Тогда существует единственная вероятностная мера P на $\sigma(\mathscr{A})$, которая совпадает с P' на $\mathscr{A}$:

$$\mathsf{P}(A) = \mathsf{P}'(A), \quad A \in \mathscr{A}$$

Пример

Пусть $\Omega = (0, 1]$. Рассмотрим алгебру $\mathscr{A}$, элементами которой являются конечные объединения интервалов:

$$A = (a_1, b_1] \cup \dots \cup (a_k, b_k], \quad 0 \le a_1 \le b_1 \dots \le a_k \le b_k \le 1.$$

Данная алгебра порождает борелевскую сигма алгебру $\mathscr{B}((0, 1]) = \sigma(\mathscr{A})$. Для A указанного выше вида положим

$$\mathsf{P}'(A) = \sum_{i=1}^{k} (b_k - a_k).$$

Можно показать, что P' корректно определена на $\mathscr{A}$ (легкая часть) и является σ-аддитивной на $\mathscr{A}$ (трудная часть). По теореме Каратеодори о продолжении меры существует единственная вероятностная мера P на $\mathscr{B}((0, 1])$, которая совпадает с P' на $\mathscr{A}$. Данная мера называется мерой Лебега.

В данном примере вероятность события пропорциональна его размеру. В частности, вероятность любой точки равна 0.

$$1 = \mathsf{P}([0,1]) = \mathsf{P}\left(\bigcup_{\omega\in[0,1]} \{\omega\}\right) \neq \sum_{\omega\in[0,1]} \mathsf{P}(\omega) = 0.$$

Операции на множествами (закон дистрибутивности, законы де Моргана)

$$A \cap (B \cup C) = A \cap B \cup A \cap C,$$

$$\left(\bigcup_{i=1}^{m} A_i \right)^c = \bigcap_{i=1}^{m} A_i^c,$$

$$\left(\bigcap_{i=1}^{m} A_i \right)^c = \bigcup_{i=1}^{m} A_i^c.$$

Простые свойства вероятностной меры

- $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$

$$1 = \mathsf{P}(\Omega) = \mathsf{P}(A \cup A^c) = \mathsf{P}(A) + \mathsf{P}(A^c).$$

- $A \subset B \Longrightarrow \mathsf{P}(A) \leq \mathsf{P}(B)$

$$B = A \cup B \setminus A, \quad B \setminus A = A^c \cap B = (A \cup B^c)^c \in \mathscr{F},$$

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A) \geq \mathsf{P}(A).$$

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

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

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

$$A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)$$

$$\begin{aligned} \mathsf{P}(A \cup B) &= \mathsf{P}(A \backslash B) + \mathsf{P}(B \backslash A) + \mathsf{P}(A \cap B) \\ &= \mathsf{P}(A) - \mathsf{P}(A \cap B) + \mathsf{P}(B) - \mathsf{P}(A \cap B) + \mathsf{P}(A \cap B) \\ &= \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B). \quad \square \end{aligned}$$

$\mathsf{P}(A \cup B) \leq \mathsf{P}(A) + \mathsf{P}(B)$ (the union bound).

Равномерное распределение вероятностей

Если Ω конечно и $\mathsf{P}(A) = |A|/|\Omega|$, то чтобы подсчитать вероятность события нужно подсчитать число исходов, которые ему благоприятствуют.

Пример

Подбрасываются 2 симметричных кубика. Какова вероятность, того что сумма очков на гранях будет равна 10?

$\Omega = \{(a_1, a_2) : a_i \in \{1, \dots, 6\}\}$, $\mathscr{F} = 2^\Omega$. Легко видеть, что $|\Omega| = 36$. Событие A имеет вид

$$A = \{\omega = (a_1, a_2) : a_1 + a_2 = 10\} = \{(4, 6), (5, 5), (6, 4)\}.$$

Таким образом, $|A| = 3$,

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{36} = \frac{1}{12}.$$

Равномерное распределение вероятностей

Если Ω конечно и $\mathsf{P}(A) = |A|/|\Omega|$, то чтобы подсчитать вероятность события нужно подсчитать число исходов, которые ему благоприятствуют.

Пример

Подбрасываются 2 симметричных кубика. Какова вероятность, того что сумма очков на гранях будет равна 10?
$\Omega = \{(a_1, a_2) : a_i \in \{1, \dots, 6\}\}$, $\mathscr{F} = 2^{\Omega}$. Легко видеть, что $|\Omega| = 36$.
Событие A имеет вид

$$A = \{\omega = (a_1, a_2) : a_1 + a_2 = 10\} = \{(4, 6), (5, 5), (6, 4)\}.$$

Таким образом, $|A| = 3$,

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{36} = \frac{1}{12}.$$

Пример

$\Omega = [0,1]\times[0,1]$, $\mathscr{F} = \mathscr{B}([0,1]\times[0,1])$,

$$\mathsf{P}((a,b)\times(c,d)) = (b-a)(d-c)$$

и мера продолжена соответствующим образом на $\mathscr{B}([0,1]\times[0,1])$. Построенная мера также называется мерой Лебега. Для простых множеств она равна их площади.
Например,

$$A = \{\omega = (x,y) : x+y \le 1/2,\ x\ge 0, y\ge 0\}.$$

$\mathsf{P}(A) = 1/8$ — площадь прямоугольного треугольника.

Условная вероятность

- Определение условной вероятности
- Правило умножения
- Формула полной вероятности
- Формула Байеса

Известно, что произошло событие B. Как изменятся вероятности других событий при этом условии?

Пример

Подбрасываются два кубика. Какова вероятность того, что сумма очков будет равна 10, если известно, что на первом кубике выпало число 5 (или число 3)?

Ясно, что $1/6$ (соответственно, 0).

Условная вероятность

- Определение условной вероятности
- Правило умножения
- Формула полной вероятности
- Формула Байеса

Известно, что произошло событие B. Как изменятся вероятности других событий при этом условии?

Пример

Подбрасываются два кубика. Какова вероятность того, что сумма очков будет равна 10, если известно, что на первом кубике выпало число 5 (или число 3)?
Ясно, что $1/6$ (соответственно, 0).

Пусть $\mathsf{P}(B) > 0$. Условной вероятностью события A относительно B называется число

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

Данная формула показывает какая доля вероятности, приписанной B, относится к A.

Пример

В примере с последовательным подбрасыванием двух кубиков пусть

$$B = \{\omega : a_1 = 5\}, \quad A = \{\omega : a_1 + a_2 = 10\},$$

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(\{\omega : a_1 = 5, a_2 = 5\})}{\mathsf{P}(\{\omega : a_1 = 5\})} = \frac{1/36}{1/6} = 1/6.$$

Если $B = \{\omega : a_1 = 3\}$, то $A \cap B = \emptyset$ и $\mathsf{P}(A|B) = 0$.

Легко проверить, что функция

$$A \mapsto \mathsf{P}_B(A) := \mathsf{P}(A|B)$$

является вероятностной мерой на $\mathscr{F}$. Получаем новое вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P}_B)$. Данная модель учитывает информацию о наступлении события B. Можно рассмотреть также вероятностное пространство $(B, \mathscr{F}_B, \mathsf{P}_B)$.

Пример (Дело Симпсона)

В 1994 г. в своем доме Лос-Анджелесе была убита Николь Браун. Основным подозреваемым был ее (бывший) муж О. Джей Симпсон актер и в прошлом профессиональный футболист. До этого он неоднократно бил свою жену.
Аргумент адвоката: "Среди тех, кто избивает своих партнеров, лишь ничтожно малое число, менее 1 из 2500, совершают убийство".

$$A = \{\text{наличие домашнего насилия}\}, \quad K = \{\text{муж убил свою жену}\},$$

$$\mathsf{P}(K|A) = 1/2500 \quad (\text{по американской статистике}).$$

Пример (продолжение)

Правильный вопрос: если в семье было домашнее насилие и женщина убита, какова вероятность того, что ее убил муж?

- $M = \{\text{жена убита}\}$,

$$\mathsf{P}(K|M \cap A) = 8/9 \quad (\text{по американской статистике}).$$

Правило умножения

Пусть $\mathsf{P}(A_1 \cap \dots \cap A_m) > 0$. Равенство

$$\mathsf{P}(A_1 \cap \dots \cap A_m) = \mathsf{P}(A_1)\mathsf{P}(A_2|A_1)\dots\mathsf{P}(A_m|A_1 \cap \dots \cap A_{m-1})$$

называется *правилом умножения*.

Доказательство.

$(m = 3)$

$$\begin{aligned}\mathsf{P}(A_1 \cap A_2 \cap A_3) =& \mathsf{P}(A_3|A_1 \cap A_2)\mathsf{P}(A_1 \cap A_2)\\ =& \mathsf{P}(A_3|A_1 \cap A_2)\mathsf{P}(A_2|A_1)\mathsf{P}(A_1)\end{aligned}$$

Пример

Имеется колода из 36 карт. Из колоды последовательно наудачу вытаскиваются 3 карты. Какова вероятность того, что среди них не будет пики?

Пусть A_i — событие, состоящее в том, что карта, вытащенная на i-шаге не является пикой. Нас интересует вероятность события $A_1 \cap A_2 \cap A_3$. По правилу умножения,

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2).$$

Нетрудно понять, что

$$\mathrm{P}(A_1) = \frac{27}{36}, \quad \mathrm{P}(A_2|A_1) = \frac{26}{35}, \quad \mathrm{P}(A_3|A_1 \cap A_2) = \frac{25}{34}.$$

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) \approx 0.41.$$

Пример

Имеется колода из 36 карт. Из колоды последовательно наудачу вытаскиваются 3 карты. Какова вероятность того, что среди них не будет пики?
Пусть A_i — событие, состоящее в том, что карта, вытащенная на i-шаге не является пикой. Нас интересует вероятность события $A_1 \cap A_2 \cap A_3$. По правилу умножения,

$$\mathsf{P}(A_1 \cap A_2 \cap A_3) = \mathsf{P}(A_1)\mathsf{P}(A_2|A_1)\mathsf{P}(A_3|A_1 \cap A_2).$$

Нетрудно понять, что

$$\mathsf{P}(A_1) = \frac{27}{36}, \quad \mathsf{P}(A_2|A_1) = \frac{26}{35}, \quad \mathsf{P}(A_3|A_1 \cap A_2) = \frac{25}{34}.$$

$$\mathsf{P}(A_1 \cap A_2 \cap A_3) \approx 0.41.$$

Формула полной вероятности

Рассмотрим теперь разбиение $\mathscr{D} = \{A_1, \ldots, A_m\}$. Равенство

$$\mathsf{P}(B) = \sum_{i=1}^{m} \mathsf{P}(B|A_i)\mathsf{P}(A_i)$$

называется *формулой полной вероятности*.

Доказательство.

$$\sum_{i=1}^{m} \mathsf{P}(B|A_i)\mathsf{P}(A_i) = \sum_{i=1}^{m} \mathsf{P}(B \cap A_i) = \mathsf{P}\left(B \bigcap \bigcup_{i=1}^{m} A_i\right) = \mathsf{P}(B).$$

Пример

Имеются три урны, в каждой из которых могут лежать белые и черные шары:

- в первой урне 3 белых и 2 черных;
- во второй урне 0 белых и 4 черных;
- в третьей урне 5 белых и 2 черных.

На первом шаге эксперимента наудачу выбирается одна из урн. На втором шаге из нее наудачу извлекается один шар. Какова вероятность того, что он будет белым?

Пример (продолжение)

Пусть A_i — событие, состоящее в том, что выбрана урна с номером i, а B — событие состоящее в том, что вытащен белый шар. Вероятности того, что B произойдет при условии A_i легко посчитать, зная количество белых и черных шаров в соответствующей урне:

$$\mathsf{P}(B|A_1) = 3/5, \quad \mathsf{P}(B|A_2) = 0, \quad \mathsf{P}(B|A_3) = 5/7.$$

По формуле полной вероятности,

$$\mathsf{P}(B) = \sum_{i=1}^{3} \mathsf{P}(B|A_i)\mathsf{P}(A_i) = \frac{3}{5} \cdot \frac{1}{3} + \frac{5}{7} \cdot \frac{1}{3} = \frac{46}{105} \approx 0.44.$$

Пример (Monty Hall)

За одной из трех закрытых дверей находится приз, за двумя другими ничего нет.

- Игрок выбирает одну из дверей.
- Ведущий открывает одну из двух оставшихся дверей, за которой ничего нет.
- Игрок может сохранить или изменить свой первоначальный выбор.

Какая стратегия лучше?

Пример (продолжение)

Если игрок сохраняет свой первоначальный выбор, то вероятность того, что он получит приз равна $1/3$.
Обозначим через A событие, состоящее в том, что за дверью первоначально выбранной игроком, находится приз. Пусть B — событие, состоящее в том, что приз будет выигран. Если игрок изменяет свой первоначальный выбор, то

$$\mathsf{P}(B|A) = 0, \quad \mathsf{P}(B|A^c) = 1.$$

По формуле полной вероятности

$$\mathsf{P}(B) = \mathsf{P}(B|A)\mathsf{P}(A) + \mathsf{P}(B|A^c)\mathsf{P}(A^c) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Таким образом, вторая стратегия намного лучше.

Пример

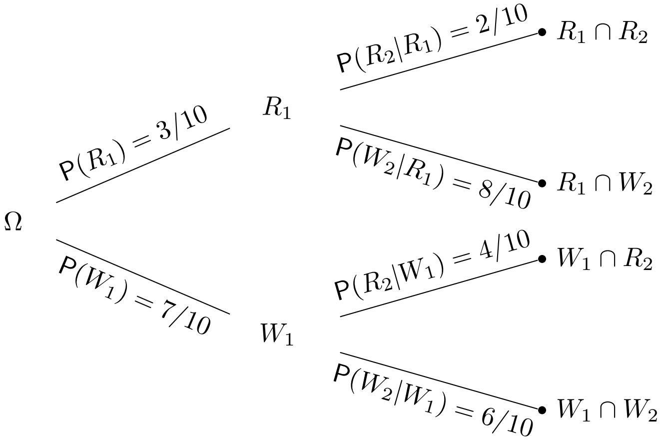
В коробке находятся 3 красных шара и 7 белых. Наудачу вытаскиваем шар из коробки и заменяем шаром другого цвета (белый меняем на красный, красный на белый). Кладем шар обратно и наудачу вытаскиваем шар еще раз.

(a) Какова вероятность в итоге вытащить красный шар?

(b) Если шары одного цвета, какова вероятность того, что они белые?

$$R_1 = \{\text{Первый шар красный}\}, \quad W_1 = \{\text{Первый шар белый}\},$$

$$R_2 = \{\text{Второй шар красный}\}, \quad W_2 = \{\text{Второй шар белый}\}.$$



Ω
$\mathsf{P}(R_1) = 3/10$
$\mathsf{P}(W_1) = 7/10$
R_1
W_1
$\mathsf{P}(R_2|R_1) = 2/10$
$\mathsf{P}(W_2|R_1) = 8/10$
$\mathsf{P}(R_2|W_1) = 4/10$
$\mathsf{P}(W_2|W_1) = 6/10$
$R_1 \cap R_2$
$R_1 \cap W_2$
$W_1 \cap R_2$
$W_1 \cap W_2$

Пример (продолжение)

(a) Вероятность того, что второй шар будет красным:

$$\begin{aligned} \mathsf{P}(R_2) &= \mathsf{P}(R_2|R_1)\mathsf{P}(R_1) + \mathsf{P}(R_2|W_1)\mathsf{P}(W_1) \\ &= \frac{2}{10}\frac{3}{10} + \frac{4}{10}\frac{7}{10} = \frac{17}{50} = 0.34 \end{aligned}$$

(b) Если шары одного цвета, то вероятность того, что они белые равна

$$\mathsf{P}(W_1 \cap W_2 | W_1 \cap W_2 \cup R_1 \cap R_2) = \frac{\mathsf{P}(W_1 \cap W_2)}{\mathsf{P}(W_1 \cap W_2 \cup R_1 \cap R_2)}$$

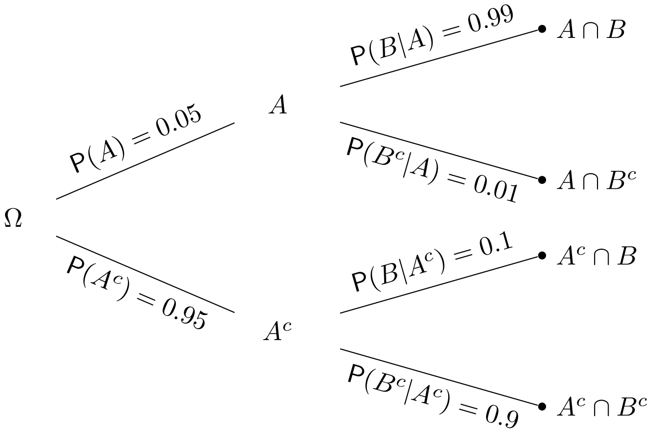
$$= \frac{6/10 \cdot 7/10}{6/10 \cdot 7/10 + 2/10 \cdot 3/10} = \frac{42}{48} = \frac{7}{8}$$

Пример (Радар)

- Если самолет находится в зоне ответственности радара, то радар его обнаруживает с вероятностью 0.99.
- Если самолета нет, то радар с создает ложную тревогу с вероятностью 0.1.
- Самолет присутствует с вероятностью 0.05.

Какова вероятность ложной тревоги? Какова вероятность того, что самолет есть, но радар его не обнаружил?

- $A = \{\text{самолет находится в зоне ответственности радара}\}$,
- $B = \{\text{радар создает тревогу}\}$.


Ω
$\mathsf{P}(A) = 0.05$
$\mathsf{P}(A^c) = 0.95$
A
A^c
$\mathsf{P}(B|A) = 0.99$
$\mathsf{P}(B^c|A) = 0.01$
$\mathsf{P}(B|A^c) = 0.1$
$\mathsf{P}(B^c|A^c) = 0.9$
$A \cap B$
$A \cap B^c$
$A^c \cap B$
$A^c \cap B^c$

- Вероятность ложной тревоги:

$$\mathsf{P}(A^c \cap B) = \mathsf{P}(A^c)\mathsf{P}(B|A^c) = 0.95 \cdot 0.1 = 0.095$$

- Вероятность того, что самолет есть, но радар его не обнаружил:

$$\mathsf{P}(A \cap B^c) = \mathsf{P}(A)\mathsf{P}(B^c|A) = 0.05 \cdot 0.01 = 0.0005$$

Вероятность разорения (Гюйгенс, 1657)

На каждом шаге игры игрок с вероятностью p выигрывает и с вероятностью $q = 1 - p$ проигрывает один рубль. Игрок с начальным капиталом x играет до тех пор, пока выиграет n рублей, либо разорится. Найти вероятность разорения (или выигрыша).
Пусть $v(x)$ — вероятность итогового выигрыша: событие W. Пусть события A_+, A_- соответствуют выигрышу и проигрышу на первом шаге. По формуле полной вероятности,

$$v(x) = \mathsf{P}(W|A_+)\mathsf{P}(A_+) + \mathsf{P}(W|A_-)\mathsf{P}(A_-) = \mathsf{P}(W|A_+)p + \mathsf{P}(W|A_-)q.$$

В случае выигрыша на первом шаге, игрок на следующем шаге начнет ту же большую игру с капиталом $x+1$. Поэтому ясно, что $\mathsf{P}(W|A_+) = v(x+1)$. Аналогично, $\mathsf{P}(W|A_-) = v(x-1)$.

Вероятность разорения (продолжение)

Итак, v удовлетворяет линейному однородному разностному уравнению

$$v(x) = pv(x+1) + qv(x-1)$$

и граничным условиям

$$v(0) = 0, \quad v(n) = 1.$$

Теория линейных разностных уравнений аналогична теории линейных дифференциальных уравнений. Будем искать решение в виде $v(x) = r^x$. Тогда r является корнем характеристического уравнения

$$1 = pr + q/r \iff pr^2 - r + q = 0.$$

Одним из корней является $r = 1$. Поэтому,

$$pr^2 - r + q = p(r-1)(r-\gamma) = pr^2 - p(1+\gamma)r + p\gamma.$$

Следовательно, второй корень $\gamma = q/p$.

Вероятность разорения (продолжение)

При $p \neq q$ корни различны и общее решение имеет вид

$$v(x) = C_1 + C_2 \left(\frac{q}{p}\right)^x .$$

Из граничных условий находим

$$v(0) = C_1 + C_2 = 0, \quad v(n) = C_1 + C_2 \left(\frac{q}{p}\right)^n = 1.$$

Таким образом,

$$v(x) = \frac{1 - (q/p)^x}{1 - (q/p)^n}$$

Вероятность разорения (продолжение)

При $p = q$ корень $r = 1$ является кратным и общее решение имеет вид

$$v(x) = C_1 + C_2 x.$$

Из граничных условий

$$v(0) = C_1 = 0, \quad v(n) = C_1 + C_2 n = 1$$

находим

$$v(x) = x/n.$$

Формула Байеса

Рассмотрим теперь разбиение $\mathscr{D} = \{A_1, \ldots, A_m\}$. Равенство

$$\mathsf{P}(A_i|B) = \frac{\mathsf{P}(B|A_i)\mathsf{P}(A_i)}{\sum_{j=1}^m \mathsf{P}(B|A_j)\mathsf{P}(A_j)}$$

называется *формулой Байеса*.

Доказательство.

$$\mathsf{P}(A_i|B) = \frac{\mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B|A_i)\mathsf{P}(A_i)}{\sum_{j=1}^m \mathsf{P}(B|A_j)\mathsf{P}(A_j)} \qquad \square$$

Формула Байеса позволяет учитывать информацию о наступлении события B для переоценки вероятностей других событий.

- $\mathsf{P}(A_i)$ — априорные вероятности (до опыта)
- $\mathsf{P}(A_i|B) = \mathsf{P}_B(A_i)$ — апостериорные вероятности (после опыта)

Пример

Вероятность того, что у пациента имеется некоторое редкое заболевание равна 0.001. Медицинский тест дает правильный результат в 95% случаев. Какова вероятность того, что конкретный пациент болен, если тест дал положительный результат?

- H — гипотеза, что пациент болен, $\mathscr{D} = \{H, H^c\}$
- B — событие, состоящее в том, что результат теста положителен.

Формула Байеса дает неожиданный результат:

$$\mathrm{P}(H|B) = \frac{\mathrm{P}(B|H)\mathrm{P}(H)}{\mathrm{P}(B|H)\mathrm{P}(H) + \mathrm{P}(B|H^c)\mathrm{P}(H^c)} = \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.05 \cdot 0.999} \approx 0.0187,$$

т.е. пациент действительно болен менее чем в 2% случаев. Значит для определения столь редкого заболевания требуется значительно более точный тест.

Пример

Вероятность того, что у пациента имеется некоторое редкое заболевание равна 0.001. Медицинский тест дает правильный результат в 95% случаев. Какова вероятность того, что конкретный пациент болен, если тест дал положительный результат?

- H — гипотеза, что пациент болен, $\mathscr{D} = \{H, H^c\}$
- B — событие, состоящее в том, что результат теста положителен.

Формула Байеса дает неожиданный результат:

$$\begin{aligned} \mathsf{P}(H|B) &= \frac{\mathsf{P}(B|H)\mathsf{P}(H)}{\mathsf{P}(B|H)\mathsf{P}(H) + \mathsf{P}(B|H^c)\mathsf{P}(H^c)} \\ &= \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.05 \cdot 0.999} \approx 0.0187, \end{aligned}$$

т.е. пациент действительно болен менее чем в 2% случаев. Значит для определения столь редкого заболевания требуется значительно более точный тест.

Пример

В коробке 10 монет:

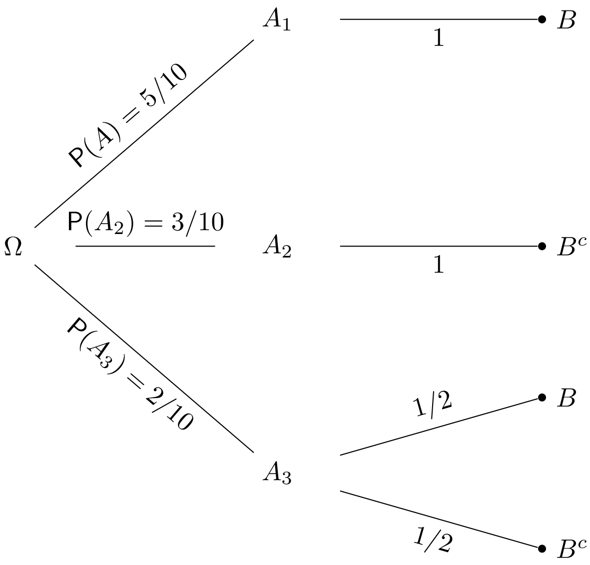
- 5 монет типа HH (орел с двух сторон)
- 3 монеты типа TT (решка с двух сторон)
- 2 монеты типа HT (обычные симметричные монеты)

Выбираем монету наудачу и подбрасываем.

(a) Найти вероятность выпадения орла.

(b) Если выпал орел, то какова вероятность того, что мы взяли монету типа HT?

- $A_1 = \{\text{выбрана монета } HH\}$
- $A_2 = \{\text{выбрана монета } TT\}$
- $A_3 = \{\text{выбрана монета } HT\}$
- $B = \{\text{выпал орел}\}$


Ω
$\mathsf{P}(A) = 5/10$
$\mathsf{P}(A_2) = 3/10$
$\mathsf{P}(A_3) = 2/10$
A_1
A_2
A_3
1
1
1/2
1/2
B
B^c
B
B^c

- Вероятность выпадения орла:

$$\mathsf{P}(B) = \mathsf{P}(A_1)\mathsf{P}(B|A_1) + \mathsf{P}(A_3)\mathsf{P}(B|A_3) = \frac{5}{10} \cdot 1 + \frac{2}{10} \cdot \frac{1}{2} = 0.6$$

- Вероятность того, что выбрана монета HT (обычная монета), если выпал орел:

$$\mathsf{P}(A_3|B) = \frac{\mathsf{P}(A_3 \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A_3)\mathsf{P}(B|A_3)}{\mathsf{P}(B)} = \frac{0.1}{0.6} = \frac{1}{6}.$$

Независимость

События A, B называются независимыми, если

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

Если $\mathsf{P}(B) > 0$, то независимость равносильна условию

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

Аналогично, если $\mathsf{P}(A) > 0$, то независимость равносильна условию

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

- Пусть $A \cap B = \emptyset$, $\mathsf{P}(A) > 0$, $\mathsf{P}(B) > 0$, тогда они зависимы:
$$0 = \mathsf{P}(A \cap B) \neq \mathsf{P}(A)\mathsf{P}(B) > 0$$
Если известно, что произошло одно из событий, то другое точно не произошло.
- Если A, B независимы, то A и B^c независимы:
$$\mathsf{P}(A) = \mathsf{P}(A \cap B) + \mathsf{P}(A \cap B^c) = \mathsf{P}(A)\mathsf{P}(B) + \mathsf{P}(A \cap B^c),$$
$$\mathsf{P}(A \cap B^c) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(B^c).$$

Пример

Из колоды 36 карт наугад вытаскивается одна карта. Независимы ли события $A = \{\text{дама}\}$, $B = \{\text{пика}\}$? Изменится ли ответ, если в колоде 52 карты?

- 36 карт

$$\mathrm{P}(A) = 1/9, \quad \mathrm{P}(B) = 1/4,$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(\text{дама пик}) = 1/36 = \mathrm{P}(A)\mathrm{P}(B).$$

- 52 карты

$$\mathrm{P}(A) = 1/13, \quad \mathrm{P}(B) = 1/4,$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(\text{дама пик}) = 1/52 = \mathrm{P}(A)\mathrm{P}(B).$$

Пример

Из колоды 36 карт наугад вытаскивается одна карта. Независимы ли события $A = \{\text{дама}\}$, $B = \{\text{пика}\}$? Изменится ли ответ, если в колоде 52 карты?

- 36 карт
$$\mathsf{P}(A) = 1/9, \quad \mathsf{P}(B) = 1/4,$$
$$\mathsf{P}(A \cap B) = \mathsf{P}(\text{дама пик}) = 1/36 = \mathsf{P}(A)\mathsf{P}(B).$$
- 52 карты
$$\mathsf{P}(A) = 1/13, \quad \mathsf{P}(B) = 1/4,$$
$$\mathsf{P}(A \cap B) = \mathsf{P}(\text{дама пик}) = 1/52 = \mathsf{P}(A)\mathsf{P}(B).$$

Пример

Построим модель двукратного подбрасывания несимметричной монеты, в которой подбрасывания независимы и вероятность выпадения орла равна $p \in (0,1)$.
Положим $\Omega = \{(a_1, a_2) : a_i = H \text{ или } a_i = T\}$, $\mathscr{F} = 2^\Omega$. Рассмотрим события

$$H_1 = \{\omega : a_1 = H\}, \quad H_2 = \{\omega : a_2 = H\},$$

$$T_1 = \{\omega : a_1 = T\}, \quad T_2 = \{\omega : a_2 = T\}.$$

Очевидно, что на P нужно наложить условия

$$\mathsf{P}(H_1) = \mathsf{P}(H_2) = p, \quad \mathsf{P}(T_1) = \mathsf{P}(T_2) = q = 1 - p,$$

чтобы вероятность выпадения орла была равна p

Пример (продолжение)

и условия

$$\mathsf{P}(\{HH\}) = \mathsf{P}(H_1 \cap H_2) = \mathsf{P}(H_1)\mathsf{P}(H_2) = p^2,$$

$$\mathsf{P}(\{HT\}) = \mathsf{P}(H_1 \cap T_2) = \mathsf{P}(H_1)\mathsf{P}(T_2) = pq,$$

$$\mathsf{P}(\{TH\}) = \mathsf{P}(T_1 \cap H_2) = \mathsf{P}(T_1)\mathsf{P}(H_2) = qp,$$

$$\mathsf{P}(\{TT\}) = \mathsf{P}(T_1 \cap T_2) = \mathsf{P}(T_1)\mathsf{P}(T_2) = q^2,$$

чтобы результаты подбрасываний были независимыми. Фактически, мера задается последними четырьмя условиями.

Пример

В условия предыдущего примера выясним, при каком условии события H_1 (орел при первом подбрасывании) и $B = \{HT, TH\}$ (один орел, одна решка) независимы.

$$\mathsf{P}(B|H_1) = \frac{\mathsf{P}(\{HT\})}{\mathsf{P}(H_1)} = \frac{pq}{p} = q,$$

$$\mathsf{P}(B) = \mathsf{P}(\{HT\}) + \mathsf{P}(\{TH\}) = 2pq.$$

Следовательно, данные события независимыми, если и только если $p = 1/2$.

Пример

В условия предыдущего примера выясним, при каком условии события H_1 (орел при первом подбрасывании) и $B = \{HT, TH\}$ (один орел, одна решка) независимы.

$$\mathsf{P}(B|H_1) = \frac{\mathsf{P}(\{HT\})}{\mathsf{P}(H_1)} = \frac{pq}{p} = q,$$

$$\mathsf{P}(B) = \mathsf{P}(\{HT\}) + \mathsf{P}(\{TH\}) = 2pq.$$

Следовательно, данные события независимыми, если и только если $p = 1/2$.

Прямое произведение вероятностных пространств

Пусть имеются вероятностные пространства

$$(\Omega_1, \mathscr{F}_1, \mathsf{P}_1), \dots, (\Omega_n, \mathscr{F}_n, \mathsf{P}_n).$$

Их прямым произведением называется вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$, где

$$\Omega = \Omega_1 \times \dots \times \Omega_n,$$

$$\mathscr{F} = \sigma(A_1 \times \dots \times A_n), \quad A_i \in \mathscr{F}_i,$$

$$\mathsf{P}(A_1 \times \dots \times A_n) = \prod_{i=1}^{n} \mathsf{P}_i(A_i).$$

и мера P продолжена на $\mathscr{F}$. Такое продолжение единственно.
Прямое произведение используется для описания n независимых экспериментов.
В прямом произведении события $H_i = \{(\omega_1, \dots, \omega_n) : \omega_i \in A_i\}$ при различных i независимы.

Пример

Для описания n-кратного подбрасывания (несимметричной) монеты введем вероятностные пространства $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$,

$$\Omega_i = \{H, T\}, \quad \mathscr{F}_i = \sigma(\{H\}, \{T\}), \quad \mathsf{P}_i(\{H\}) = p, \quad \mathsf{P}_i(\{T\}) = q.$$

В этом пространстве подбрасывания, т.е. события вида

$$H_i = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in A_i\}, \quad A_i \in \mathscr{F}_i$$

независимы и

$$\mathsf{P}(\{\omega\}) = \prod_{i=1}^{n} \mathsf{P}_i(\{\omega_i\}) = p^{\nu(\omega)}(1-p)^{n-\nu(\omega)},$$

где $\nu(\omega)$ – количество H в последовательности $\omega = (a_1, \dots, a_n)$.

Пример (задача о разделе ставки, решена Паскалем и Ферма)

Два игрока играют матч до 6 побед. В каждой партии вероятность того, что выиграет первый игрок равна p. Результаты партий независимы. Игра прервана при счете 5:3 в пользу первого игрока. Как справедливо разделить приз?

Идея: справедливо делить приз пропорционально вероятностям выиграть матч при продолжении игры. При этом может быть сыграно не более трех партий. Всего имеется 8 исходов: $\omega = (a_1, a_2, a_3)$, $a_i = \{1, 2\}$. Второй игрок побеждает только в при реализации исхода $(2, 2, 2)$. В силу независимости партий,

$$\mathrm{P}(2, 2, 2) = q^3, \quad q = 1 - p.$$

$$\frac{\text{вероятность победы в матче 1 игрока}}{\text{вероятность победы в матче 2 игрока}} = \frac{1 - q^3}{q^3}$$

При $p = 1/2$ справедливо делить приз в соотношении $7 : 1$.

Пример (задача о разделе ставки, решена Паскалем и Ферма)

Два игрока играют матч до 6 побед. В каждой партии вероятность того, что выиграет первый игрок равна p. Результаты партий независимы. Игра прервана при счете 5:3 в пользу первого игрока. Как справедливо разделить приз?

Идея: справедливо делить приз пропорционально вероятностям выиграть матч при продолжении игры. При этом может быть сыграно не более трех партий. Всего имеется 8 исходов: $\omega = (a_1, a_2, a_3)$, $a_i = \{1, 2\}$. Второй игрок побеждает только в при реализации исхода $(2, 2, 2)$. В силу независимости партий,

$$\mathsf{P}(2,2,2) = q^3, \quad q = 1 - p.$$

$$\frac{\text{вероятность победы в матче 1 игрока}}{\text{вероятность победы в матче 2 игрока}} = \frac{1 - q^3}{q^3}$$

При $p = 1/2$ справедливо делить приз в соотношении $7 : 1$.

Пример (де Мере, 17 век)

де Мере (Антуан Гомбо): французский писатель, игрок, математик-любитель.
Какова вероятность выпадения хотя бы одной 1 (событие A) при четырех подбрасываниях кубика?

$\Omega = \{a_1, \ldots, a_4\}$, $a_i \in \{1, \ldots, 6\}$. Все подбрасывания независимы.

$$\mathsf{P}(A^c) = \prod_{i=1}^{4} \mathsf{P}(a_i \neq 1) = \left(\frac{5}{6}\right)^4,$$

$$\mathsf{P}(A) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$$

Пример (де Мере, 17 век)

де Мере (Антуан Гомбо): французский писатель, игрок, математик-любитель.

Какова вероятность выпадения хотя бы одной 1 (событие A) при четырех подбрасываниях кубика?

$\Omega = \{a_1, \ldots, a_4\}$, $a_i \in \{1, \ldots, 6\}$. Все подбрасывания независимы.

$$\mathsf{P}(A^c) = \prod_{i=1}^{4} \mathsf{P}(a_i \neq 1) = \left(\frac{5}{6}\right)^4,$$

$$\mathsf{P}(A) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$$

Пример (де Мере)

Два кубика подбрасываются 24 раза. Какова вероятность того, что хотя бы один раз выпадет пара $(1,1)$ (событие A)?

$\Omega = \{(a_1,b_1),\ldots,(a_{24},b_{24})\}$, $a_i,b_i \in \{1,\ldots,6\}\}$. Все подбрасывания независимы.

$$\mathsf{P}(A^c) = \prod_{i=1}^{24} \mathsf{P}((a_i,b_i) \neq (1,1)) = \left(\frac{35}{36}\right)^{24},$$

$$\mathsf{P}(A) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914.$$

Пример (де Мере)

Два кубика подбрасываются 24 раза. Какова вероятность того, что хотя бы один раз выпадет пара $(1,1)$ (событие A)?
$\Omega = \{(a_1,b_1),\ldots,(a_{24},b_{24})\}$, $a_i, b_i \in \{1,\ldots,6\}\}$. Все подбрасывания независимы.

$$\mathsf{P}(A^c) = \prod_{i=1}^{24} \mathsf{P}((a_i,b_i) \neq (1,1)) = \left(\frac{35}{36}\right)^{24},$$

$$\mathsf{P}(A) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914.$$

Условная независимость

Пусть имеется событие C: $\mathsf{P}(C) > 0$. События A, B называются условно независимыми относительно C, если

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

Пример

Пусть A, B независимы и C обладает следующим свойством:

$$\mathsf{P}(A \cap C) > 0, \quad \mathsf{P}(B \cap C) > 0, \quad \mathsf{P}(A \cap B \cap C) = 0.$$

Тогда A, B условно зависимы относительно C.
Например, в модели двукратного подбрасывания монеты, $A = H_1$, $B = H_2$, $C = \{HT, TH\}$:

$$A \cap C = \{HT\}, \quad B \cap C = \{TH\}, \quad A \cap B \cap C = \emptyset.$$

Пример

Имеются две монеты: синяя и красная. Для синей монеты вероятность выпадения орла равна 0.99. Для красной монеты вероятность выпадения орла равна 0.01. Выбираем монету случайным образом и подбрасываем два раза. Будут ли результаты подбрасываний независимыми?

Интуитивно: нет, так как если при первом подбрасывании выпал орел, то есть серьезные основания полагать, что выбрана синяя монета и результатом следующего подбрасывания также будет орел.

С другой стороны, если известно, какая монета выбрана, то результаты подбрасываний независимы.

Пример

Имеются две монеты: синяя и красная. Для синей монеты вероятность выпадения орла равна 0.99. Для красной монеты вероятность выпадения орла равна 0.01. Выбираем монету случайным образом и подбрасываем два раза. Будут ли результаты подбрасываний независимыми?
Интуитивно: нет, так как если при первом подбрасывании выпал орел, то есть серьезные основания полагать, что выбрана синяя монета и результатом следующего подбрасывания также будет орел.
С другой стороны, если известно, какая монета выбрана, то результаты подбрасываний независимы.

Пример (продолжение)

- $B = \{\text{Выбрана синяя монета}\}, R = \{\text{Выбрана красная монета}\}$
- $H_1 = \{\text{Первый орел}\}, H_2 = \{\text{Второй орел}\}$

$$
\begin{aligned}
P(H_1 \cap H_2) &= \mathsf{P}(H_1 \cap H_2|B)\mathsf{P}(B) + \mathsf{P}(H_1 \cap H_2|R)\mathsf{P}(R) \\
&= \mathsf{P}(H_1|B)\mathsf{P}(H_2|B)\mathsf{P}(B) + \mathsf{P}(H_1|R)\mathsf{P}(H_2|R)\mathsf{P}(R) \\
&= 0.99 \cdot 0.99 \cdot 0.5 + 0.01 \cdot 0.01 \cdot 0.5 = 0.4901 \\
\mathsf{P}(H_1) &= \mathsf{P}(H_1|B)\mathsf{P}(B) + \mathsf{P}(H_1|R)\mathsf{P}(R) = 0.99 \cdot 0.5 + 0.01 \cdot 0.5 = 0.5 \\
\mathsf{P}(H_2) &= 0.5 \\
\mathsf{P}(H_1)\mathsf{P}(H_2) &= \frac{1}{4} \neq \mathsf{P}(H_1 \cap H_2) = 0.4901
\end{aligned}
$$

Пример (повторный тест)

Вероятность того, что у пациента имеется некоторое редкое заболевание равна 0.001. Медицинский тест дает правильный результат в 95% случаев. Какова вероятность того, что конкретный пациент болен, если два проведенных подряд теста дали положительные результаты?

- $H = \{\text{пациент болен}\}$
- $B_1 = \{\text{первый тест положителен}\}$
- $B_2 = \{\text{второй тест положителен}\}$

Пример с синей и красной монетами показывает, что результаты тестов скорее всего зависимы. Неявное предположение: результаты тестов независимы, если известно что пациент болен (или здоров).

Пример (продолжение)

Формула Байеса:

$$\begin{aligned}
\mathsf{P}(H|B_1 \cap B_2) &= \frac{\mathsf{P}(B_1 \cap B_2|H)\mathsf{P}(H)}{\mathsf{P}(B_1 \cap B_2|H)\mathsf{P}(H) + \mathsf{P}(B_1 \cap B_2|H^c)\mathsf{P}(H^c)} \\
&= \frac{\mathsf{P}(B_1|H)\mathsf{P}(B_2|H)\mathsf{P}(H)}{\mathsf{P}(B_1|H)\mathsf{P}(B_2|H)\mathsf{P}(H) + \mathsf{P}(B_1|H^c)\mathsf{P}(B_2|H^c)\mathsf{P}(H^c)} \\
&= \frac{0.95 \cdot 0.95 \cdot 0.001}{0.95 \cdot 0.95 \cdot 0.001 + 0.05 \cdot 0.05 \cdot 0.999} \approx 0.265
\end{aligned}$$

Ранее было получено, что $\mathsf{P}(H|B_1) \approx 0.0187$.

Пусть B_i, $i = 1, \ldots, n$ независимы относительно каждой гипотезы A_i и

$$\mathsf{P}(B_1|A_i) = \cdots = \mathsf{P}(B_n|A_i).$$

К чему стремится $\mathsf{P}(A_i|B_1 \cap \cdots \cap B_n)$ при $n \to \infty$?

Независимость в совокупности

Множества $A_1, \dots, A_m$ независимы в совокупности, если для любых $k = 2, \dots, m$, $1 \le i_1 < \dots \le i_k \le m$ верно равенство

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

Пример

Если события независимы в совокупности, то совместное наступление любой подсистемы событий не зависит от совестного наступления от другой подсистемы. Пусть $I, J \subset \{1, \dots, m\}$, $I \cap J = \emptyset$. Тогда

$$\mathsf{P}\left(\bigcap_{i\in I} A_i \middle| \bigcap_{j\in J} A_j\right) = \frac{\mathsf{P}\left(\bigcap_{k\in I\cup J} A_k\right)}{\mathsf{P}\left(\bigcap_{j\in J} A_j\right)} = \frac{\prod_{k\in I\cup J} \mathsf{P}(A_k)}{\prod_{j\in J} \mathsf{P}(A_j)}$$

$$= \prod_{k\in I} \mathsf{P}(A_k) = \mathsf{P}\left(\bigcap_{i\in I} A_i\right).$$

Пример

Независимость в совокупности трех событий A, B, C означает, что они попарно независимы:

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B), \quad \mathsf{P}(A \cap C) = \mathsf{P}(A)\mathsf{P}(C), \quad \mathsf{P}(B \cap C) = \mathsf{P}(B)\mathsf{P}(C),$$

и, кроме того,

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

Пример

Рассмотрим двукратное подбрасывание симметричной монеты. Пусть $A = H_1$ (орел при первом подбрасывании), $B = H_2$ (орел при втором подбрасывании), $C = \{HT, TH\}$ (один орел, одна решка). Данные события независимы попарно, но не являются независимыми в совокупности:

$$\begin{aligned}
\mathsf{P}(A \cap B) &= \mathsf{P}(\{HH\}) = \frac{1}{4} = \mathsf{P}(A)\mathsf{P}(B) = \frac{1}{2}\frac{1}{2},\\
\mathsf{P}(A \cap C) &= \mathsf{P}(\{HT\}) = \frac{1}{4} = \mathsf{P}(A)\mathsf{P}(C) = \frac{1}{2}\frac{1}{2},\\
\mathsf{P}(B \cap C) &= \mathsf{P}(\{TH\}) = \frac{1}{4} = \mathsf{P}(B)\mathsf{P}(C) = \frac{1}{2}\frac{1}{2},\\
\mathsf{P}(A \cap B \cap C) &= 0 \neq \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C).
\end{aligned}$$

Заметим, что если наступили A и B, то C наступить не может. В частности, $A \cap B$ и C зависимы.

Перечислительная комбинаторика

Перечислительная комбинаторика занимается подсчетом числа элементов конечных множеств, которые заданы при помощи некоторого неявного описания. Используется для подсчета числа исходов, благоприятствующих интересующему событию.

Основной принцип перечисления.

Пусть S *— множество последовательностей*

$$(a_1, \ldots, a_m) \in A_1 \times \cdots \times A_m,$$

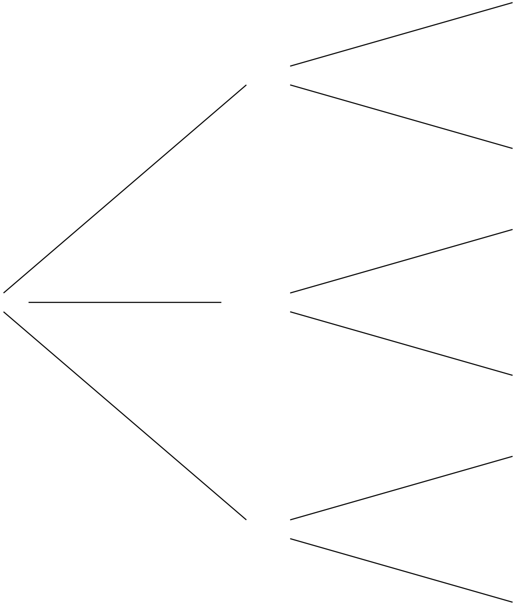
обладающих следующими свойствами:

- a_1 *может принимать одно из* n_1 *значений*
- *для любой последовательности* $(a_1, \ldots, a_{i-1})$ *элемент* a_i *может принимать одно из* n_i *значений, множество этих значений может зависеть от* $(a_1, \ldots, a_{i-1})$ *(но их количество равно* n_i*).*

Тогда $|S| = n_1 \ldots n_m$.

Действительно, на первом шаге элемент a_1 можно выбрать n_1 способами, на втором шаге при любом a_1 элемент a_2 можно выбрать n_2 способами и т.д.

$m = 2$, $n_1 = 3$, $n_2 = 2$, $|S| = 6$:



- Рассмотрим декартово произведение конечного числа конечных множеств:

$$A = A_1 \times \dots \times A_m = \{(a_1, \dots, a_m) : a_i \in A_i\}.$$

Справедливо равенство $|A| = |A_1| \dots |A_m|$.

- Количество выборок с возвращением = количество всех последовательностей $(a_1, \dots, a_m)$, $a_i \in A$, $|A| = n$:

$$|A^m| = n^m.$$

- Количество всех подмножеств множества множества A с $|A| = n$ = количество всех последовательностей $(a_1, \dots, a_n)$, $a_i \in \{0, 1\}$:

$$2^{|A|}.$$

.

- Количество выборок без возвращения = количество всех последовательностей $(a_1, \ldots, a_m)$, $a_i \in A$, элементы которых различны: $a_i \neq a_j, i \neq j$ = количество размещений. Если $|A| = n$, то

$$P_n^m = n(n-1)\ldots(n-m+1) = \frac{n!}{(n-m)!}, \quad n \geq m.$$

  На первом шаге есть n вариантов, на втором $n-1$ и т.д. При $m = n$ получаем число перестановок:

$$P_n^n = n! \quad (0! := 1).$$

- Число сочетаний=число неупорядоченных выборок без возвращения=количество подмножеств мощности m:

$$C_n^m := \binom{n}{m} = \frac{P_n^m}{m!} = \frac{n!}{m!(n-m)!}, \quad n \geq m.$$

  Количество всех подмножеств:

$$\sum_{m=0}^{n} C_n^m = (1+1)^n = 2^n.$$

Пример

Сколько существует возможных 10-значных телефонных номеров, если первая цифра не может быть 0:

$$9 \cdot 10^9.$$

Пример (Задача о днях рождения)

В комнате находится $m < 365$ человек. Какова вероятность того, все их дни рождения различны (предполагается, что никто не родился 29 февраля).

В качестве Ω возьмем множество всех последовательностей $(a_1, \ldots, a_m)$, $a_i \in \{1, \ldots, 365\}$ (a_i — день рождения i-го человека). Событию A благоприятствуют выборки без возвращения. Таким образом,

$$\mathrm{P}(A) = \frac{P_{365}^m}{365^m} = \frac{365(365-1)\ldots(365-m+1)}{365^m}.$$

Наименьшее m, при котором эта вероятность меньше $1/2$ равно 23.

Пример

Сколько существует возможных 10-значных телефонных номеров, если первая цифра не может быть 0:

$$9 \cdot 10^9.$$

Пример (Задача о днях рождения)

В комнате находится $m < 365$ человек. Какова вероятность того, все их дни рождения различны (предполагается, что никто не родился 29 февраля).

В качестве Ω возьмем множество всех последовательностей $(a_1, \dots, a_m)$, $a_i \in \{1, \dots, 365\}$ (a_i — день рождения i-го человека). Событию A благоприятствуют выборки без возвращения. Таким образом,

$$\mathsf{P}(A) = \frac{P_{365}^m}{365^m} = \frac{365(365-1)\dots(365-m+1)}{365^m}.$$

Наименьшее m, при котором эта вероятность меньше $1/2$ равно 23.

Вероятность совпадения дней рождения:

[2]:
```
import math
for m in (10,15,20,22,23,25,30,35,40,45,50,55,60):
    print('m=',m,'|',1-math.factorial(365)/(365**m*math.factorial(365-m)))
```

```
m= 10 | 0.11694817771107768
m= 15 | 0.25290131976368635
m= 20 | 0.41143838358057994
m= 22 | 0.4756953076625501
m= 23 | 0.5072972343239854
m= 25 | 0.5686997039694639
m= 30 | 0.7063162427192686
m= 35 | 0.8143832388747152
m= 40 | 0.891231809817949
m= 45 | 0.940975899465775
m= 50 | 0.9703735795779884
m= 55 | 0.9862622888164461
m= 60 | 0.994122660865348
```

Пример

21 июня 1995 в немецкой лотерее 6/49 выпала комбинация 15-25-27-30-42-48. Эта же комбинация выпала 20 декабря 1986. Такое произошло впервые за 3016 розыгрышей. Какова вероятность совпадения выпавших комбинаций в 3016 розыгрышах?

Всего имеется $N = C_{49}^{6} = 13983816 \approx 14 \cdot 10^6$ вариантов. Количество последовательностей (возможных) розыгрышей, все результаты которых различны:

$$N(N-1)\cdots(N-3015).$$

Количество всех последовательностей: N^{3016}. Вероятность отсутствия совпадений:

$$\frac{N(N-1)\cdots(N-3015)}{N^{3016}} \approx 0.7224$$

Вероятность наличия совпадений ≈ 0.2776.

Пример

21 июня 1995 в немецкой лотерее 6/49 выпала комбинация 15-25-27-30-42-48. Эта же комбинация выпала 20 декабря 1986. Такое произошло впервые за 3016 розыгрышей. Какова вероятность совпадения выпавших комбинаций в 3016 розыгрышах?

Всего имеется $N = C_{49}^{6} = 13983816 \approx 14 \cdot 10^6$ вариантов. Количество последовательностей (возможных) розыгрышей, все результаты которых различны:

$$N(N-1)\cdots(N-3015).$$

Количество всех последовательностей: N^{3016}. Вероятность отсутствия совпадений:

$$\frac{N(N-1)\cdots(N-3015)}{N^{3016}} \approx 0.7224$$

Вероятность наличия совпадений ≈ 0.2776.

Пример (Биномиальное распределение)

Рассмотрим модель m-кратного подбрасывания несимметричной монеты с вероятностью выпадения орла равной p. Найти вероятность того, что выпадет j орлов.

Событию A благоприятствуют последовательности, содержащие j орлов. Такие последовательности находятся в биективном отношении с подмножествами $\{i_1, \ldots, i_j\} \subset \{1, \ldots, m\}$, содержащими j элементов. Количество таких подмножеств равно C_m^j. Вероятность любого исхода с j орлами равна $p^j q^{m-j}$. Таким образом,

$$\mathsf{P}(j \text{ орлов}) = C_m^j p^j q^{m-j}.$$

Получаем вероятностную меру (распределение) на (самой богатой алгебре множества) $\{0, \ldots, m\}$.

Пример (Биномиальное распределение)

Рассмотрим модель m-кратного подбрасывания несимметричной монеты с вероятностью выпадения орла равной p. Найти вероятность того, что выпадет j орлов.

Событию A благоприятствуют последовательности, содержащие j орлов. Такие последовательности находятся в биективном отношении с подмножествами $\{i_1, \ldots, i_j\} \subset \{1, \ldots, m\}$, содержащими j элементов. Количество таких подмножеств равно C_m^j. Вероятность любого исхода с j орлами равна $p^j q^{m-j}$. Таким образом,

$$\mathsf{P}(j \text{ орлов}) = C_m^j p^j q^{m-j}.$$

Получаем вероятностную меру (распределение) на (самой богатой алгебре множества) $\{0, \ldots, m\}$.

Пример (Биномиальное распределение)

Рассмотрим модель m-кратного подбрасывания несимметричной монеты с вероятностью выпадения орла равной p. Найти вероятность того, что выпадет j орлов.

Событию A благоприятствуют последовательности, содержащие j орлов. Такие последовательности находятся в биективном отношении с подмножествами $\{i_1, \dots, i_j\} \subset \{1, \dots, m\}$, содержащими j элементов. Количество таких подмножеств равно C_m^j. Вероятность любого исхода с j орлами равна $p^j q^{m-j}$. Таким образом,

$$\mathsf{P}(j \text{ орлов}) = C_m^j p^j q^{m-j}.$$

Получаем вероятностную меру (распределение) на (самой богатой алгебре множества) $\{0, \dots, m\}$.

Пример

Самюэль Пипс (английский чиновник) задал следующий вопрос Исааку Ньютону. Какое событие имеет бо́льшую вероятность

- выпадение по крайней мере одной шестерки при 6 подбрасываниях кубика (A),
- выпадение по крайней мере двух шестерок при 12 подбрасываниях кубика (B),
- выпадение по крайней мере трех шестерок при 18 подбрасываниях кубика (C)?

Пример (продолжение)

$$\mathsf{P}(A^c) = \left(\frac{5}{6}\right)^6, \quad P(A) = 1 - \mathsf{P}(A^c) \approx 0.6651$$

B^c: выпадет одна шестерка или ни одной при 12 подбрасываниях:

$$\mathsf{P}(B^c) = \left(\frac{5}{6}\right)^{12} + C_{12}^1 \frac{1}{6}\left(\frac{5}{6}\right)^{11}, \quad \mathsf{P}(B) = 1 - \mathsf{P}(B^c) \approx 0.6187$$

C^c: две, одна или ни одной шестерки при 18 подбрасываниях:

$$\mathsf{P}(C^c) = \left(\frac{5}{6}\right)^{18} + C_{18}^1 \frac{1}{6}\left(\frac{5}{6}\right)^{17} + C_{18}^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{16},$$

$$\mathsf{P}(C) = 1 - \mathsf{P}(C^c) \approx 0.5973$$

Ньютон получил правильный ответ, используя похожие вычисления.

Пример (продолжение)

$$\mathsf{P}(A^c) = \left(\frac{5}{6}\right)^6, \quad P(A) = 1 - \mathsf{P}(A^c) \approx 0.6651$$

B^c: выпадет одна шестерка или ни одной при 12 подбрасываниях:

$$\mathsf{P}(B^c) = \left(\frac{5}{6}\right)^{12} + C_{12}^1 \frac{1}{6} \left(\frac{5}{6}\right)^{11}, \quad \mathsf{P}(B) = 1 - \mathsf{P}(B^c) \approx 0.6187$$

C^c: две, одна или ни одной шестерки при 18 подбрасываниях:

$$\mathsf{P}(C^c) = \left(\frac{5}{6}\right)^{18} + C_{18}^1 \frac{1}{6} \left(\frac{5}{6}\right)^{17} + C_{18}^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{16},$$

$$\mathsf{P}(C) = 1 - \mathsf{P}(C^c) \approx 0.5973$$

Ньютон получил правильный ответ, используя похожие вычисления.

Пример (продолжение)

$$\mathsf{P}(A^c) = \left(\frac{5}{6}\right)^6, \quad P(A) = 1 - \mathsf{P}(A^c) \approx 0.6651$$

B^c: выпадет одна шестерка или ни одной при 12 подбрасываниях:

$$\mathsf{P}(B^c) = \left(\frac{5}{6}\right)^{12} + C_{12}^1 \frac{1}{6}\left(\frac{5}{6}\right)^{11}, \quad \mathsf{P}(B) = 1 - \mathsf{P}(B^c) \approx 0.6187$$

C^c: две, одна или ни одной шестерки при 18 подбрасываниях:

$$\mathsf{P}(C^c) = \left(\frac{5}{6}\right)^{18} + C_{18}^1 \frac{1}{6}\left(\frac{5}{6}\right)^{17} + C_{18}^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{16},$$

$$\mathsf{P}(C) = 1 - \mathsf{P}(C^c) \approx 0.5973$$

Ньютон получил правильный ответ, используя похожие вычисления.

Пример

Сколько можно составить слов (различных последовательностей букв) переставляя буквы в слове STATISTICS?

- Рассмотрим множество всех перестановок, считая все буквы различными. Число таких перестановок $10!$.
- Отождествляем последовательности, в которые входят буквы S. Число последовательностей уменьшится в $3!$ раза.
- Затем отождествляем последовательности, которые содержат T. Число последовательностей уменьшится еще в $3!$ раза.
- Наконец, отождествляем последовательности, которые содержат I. Число последовательностей уменьшится еще в $2!$ раза.

Таким образом,

$$\frac{10!}{3!3!1!2!1!} = 50400.$$

Пример

Сколько можно составить слов (различных последовательностей букв) переставляя буквы в слове STATISTICS?

- Рассмотрим множество всех перестановок, считая все буквы различными. Число таких перестановок $10!$.
- Отождествляем последовательности, в которые входят буквы S. Число последовательностей уменьшится в $3!$ раза.
- Затем отождествляем последовательности, которые содержат T. Число последовательностей уменьшится еще в $3!$ раза.
- Наконец, отождествляем последовательности, которые содержат I. Число последовательностей уменьшится еще в $2!$ раза.

Таким образом,

$$\frac{10!}{3!3!1!2!1!} = 50400.$$

Пример

Сколько можно составить слов (различных последовательностей букв) переставляя буквы в слове STATISTICS?

- Рассмотрим множество всех перестановок, считая все буквы различными. Число таких перестановок $10!$.
- Отождествляем последовательности, в которые входят буквы S. Число последовательностей уменьшится в $3!$ раза.
- Затем отождествляем последовательности, которые содержат T. Число последовательностей уменьшится еще в $3!$ раза.
- Наконец, отождествляем последовательности, которые содержат I. Число последовательностей уменьшится еще в $2!$ раза.

Таким образом,

$$\frac{10!}{3!3!1!2!1!} = 50400.$$

В общем случае рассмотрим последовательности

$$(a_1, \ldots, a_m) : a_i \in \{b_1, \ldots, b_k\},$$

в каждой из который m_1 элементов 1-го типа, ..., m_k элементов k-го типа и $m_1 + \cdots + m_k = m$. Обозначим через

$$P_m^{m_1, \ldots, m_k} = \frac{m!}{m_1! \ldots m_k!}$$

число таких последовательностей.

Мультиномиальное распределение

Рассмотрим множество последовательностей

$$\Omega = \{\omega = (a_1, \dots, a_m) : a_i \in \{b_1, \dots, b_k\}\}.$$

Пусть $\nu_j(\omega)$ — количество элементов последовательности ω, равных b_j. Если на каждом из m шагов вероятность взять элемент $a_i = b_j$ равна p_j, $(p_1 + \dots + p_k = 1)$ и эти операции независимы, то

$$\mathsf{P}(\omega) = p_1^{\nu_1(\omega)} \dots p_k^{\nu_k(\omega)}.$$

Количество таких последовательностей равно $P_m^{\nu_1(\omega),\dots,\nu_k(\omega)}$. Таким образом,

$$\mathsf{P}(\{\omega : \nu_1(\omega) = m_1, \dots, \nu_k(\omega) = m_k\}) = \frac{m!}{m_1! \dots m_k!} p_1^{m_1} \dots p_k^{m_k}.$$

Получаем вероятностную меру на $\{1, \dots, m\}^k$: мультиномиальное распределение.

Пример (задача де Мере, вопрос был задан Паскалю)

Кубик подбрасывается 3 раза. Чему чаще равна сумма очков: 11 или 12?

Рассмотрим сочетания, приводящие к 11:

$$(6,4,1),\ (6,3,2),\ (5,5,1),\ (5,4,2),\ (5,3,3),\ (4,4,3)$$

и к 12:

$$(6,5,1),\ (6,4,2),\ (6,3,3),\ (5,5,2),\ (5,4,3),\ (4,4,4).$$

Их число одинаково, но они не равновероятны. Сочетания, содержащие 3 различных цифры порождаются $6 = 3!$ последовательностями. Сочетания, содержащие 2 различных цифры порождаются $3 = \frac{3!}{1!2!}$ последовательностями. Сочетания, не содержащие различных цифр — одной последовательностью.

Пример (задача де Мере, вопрос был задан Паскалю)

Кубик подбрасывается 3 раза. Чему чаще равна сумма очков: 11 или 12?
Рассмотрим сочетания, приводящие к 11:

$$(6,4,1),\ (6,3,2),\ (5,5,1),\ (5,4,2),\ (5,3,3),\ (4,4,3)$$

и к 12:

$$(6,5,1),\ (6,4,2),\ (6,3,3),\ (5,5,2),\ (5,4,3),\ (4,4,4).$$

Их число одинаково, но они не равновероятны. Сочетания, содержащие 3 различных цифры порождаются $6 = 3!$ последовательностями. Сочетания, содержащие 2 различных цифры порождаются $3 = \frac{3!}{1!2!}$ последовательностями. Сочетания, не содержащие различных цифр — одной последовательностью.

Пример (задача де Мере, вопрос был задан Паскалю)

Кубик подбрасывается 3 раза. Чему чаще равна сумма очков: 11 или 12?
Рассмотрим сочетания, приводящие к 11:

$$(6,4,1),\ (6,3,2),\ (5,5,1),\ (5,4,2),\ (5,3,3),\ (4,4,3)$$

и к 12:

$$(6,5,1),\ (6,4,2),\ (6,3,3),\ (5,5,2),\ (5,4,3),\ (4,4,4).$$

Их число одинаково, но они не равновероятны. Сочетания, содержащие 3 различных цифры порождаются $6 = 3!$ последовательностями. Сочетания, содержащие 2 различных цифры порождаются $3 = \frac{3!}{1!2!}$ последовательностями. Сочетания, не содержащие различных цифр — одной последовательностью.

Пример (продолжение)

Число исходов, благоприятствующих 11:

$$6 + 6 + 3 + 6 + 3 + 3 = 27.$$

Число исходов, благоприятствующих 12:

$$6 + 6 + 3 + 3 + 6 + 1 = 25.$$

Число всех исходов: $6^3 = 216$. Таким образом,

$$\mathsf{P}(11) = \frac{27}{216} = \frac{1}{8} = 0.125, \quad \mathsf{P}(12) = \frac{25}{216} = 0.1157.$$

[23]:
```
def dice_sum(r=3,generator='random.choices'):
    if generator=='random.choices':
        s=sum(random.choices([1,2,3,4,5,6],k=r))
    elif generator=='numpy.random':
        s=np.sum(np.random.randint(1,7,r))
    return s
```

[25]:
```
import random
import numpy as np
n=100000
for gen in ['random.choices','numpy.random']:
    n_11=n_12=0
    for i in range(n):
        s=dice_sum(r=3,generator=gen)
        if s==11:
            n_11+=1
        elif s==12:
            n_12+=1
    print(gen.ljust(14), '11:', n_11/n, '12:', n_12/n)
```

```
random.choices 11: 0.12514 12: 0.11715
numpy.random   11: 0.12503 12: 0.11739
```

Пример

На выборах в маленьком городке за кандидата A (de Martini) проголосовали 2656 человек, а за кандидата B (Power) – 2594 человека. Разница составляет 62 голоса. Расследование выборов, инициированное проигравшим, показало, что 136 человек, проголосовавших на выборах, не имели права этого делать. Следует ли отменить результаты выборов (других нарушений не было)?

Пусть голосам, отданным за кандидата A соответствуют белые шары, а кандидату B – черные. Удалим $m = 136$ шаров, выбирая их случайным образом. Гипотеза состоит в том, что каждый избиратель с одинаковой вероятностью был допущен к выборам ошибочно.

Пример

На выборах в маленьком городке за кандидата A (de Martini) проголосовали 2656 человек, а за кандидата B (Power) – 2594 человека. Разница составляет 62 голоса. Расследование выборов, инициированное проигравшим, показало, что 136 человек, проголосовавших на выборах, не имели права этого делать. Следует ли отменить результаты выборов (других нарушений не было)?

Пусть голосам, отданным за кандидата A соответствуют белые шары, а кандидату B – черные. Удалим $m = 136$ шаров, выбирая их случайным образом. Гипотеза состоит в том, что каждый избиратель с одинаковой вероятностью был допущен к выборам ошибочно.

Пример (продолжение)

- Количество всех наборов из 2656+2594=5250 по 136 равно C_{2520}^{136},
- k шаров из 2656 белых можно выбрать C_{2656}^{k} способами,
- $136-k$ шаров из 2594 черных можно выбрать C_{2594}^{136-k} способами.

Используя правило умножения, находим, что вероятность удалить ровно k белых шаров равна

$$\frac{C_{2656}^{k} C_{2594}^{136-k}}{C_{5250}^{136}}.$$

Результат выборов будет другим, если

$$2656-k \leq 2594-(136-k) \iff k \geq 99.$$

Пример (продолжение)

- Количество всех наборов из 2656+2594=5250 по 136 равно C_{2520}^{136},
- k шаров из 2656 белых можно выбрать C_{2656}^{k} способами,
- $136-k$ шаров из 2594 черных можно выбрать C_{2594}^{136-k} способами.

Используя правило умножения, находим, что вероятность удалить ровно k белых шаров равна

$$\frac{C_{2656}^{k} C_{2594}^{136-k}}{C_{5250}^{136}}.$$

Результат выборов будет другим, если

$$2656-k \leq 2594-(136-k) \iff k \geq 99.$$

Пример (продолжение)

Вероятность этого события равна

$$\sum_{k=99}^{136} \frac{C_{2656}^{k} C_{2594}^{136-k}}{C_{5250}^{136}} \approx 7.49 \cdot 10^{-8}$$

[16]:
```
import scipy.special
s=0
for i in range(99,137):
    s+=scipy.special.binom(2656,i)*scipy.special.binom(2594,136-i)/scipy.
 ↪special.binom(5250,136)
s
```

[16]:
```
7.492089252003911e-08
```

Эти расчеты подтверждают решение Апелляционного суда: результаты выборов оставлены в силе.

Комбинации в покере

52 карты, раздача (рука, hand) 5 карт.

(1) Стрит флеш (straight flush): 5 карт одной масти по порядку.

(2) Каре (four of a kind): 4 карты одного значения.

(3) Фулл хаус (full house): 3 карты одного значения и две карты другого значения.

(4) Флеш (flush): 5 карт одной масти (исключая стрит флеш).

(5) Стрит (straight): 5 карт по порядку (исключая стрит флеш).

(6) Тройка (three of a kind): 3 карты одного значения (исключая каре и фулл хаус).

(7) Две пары (two pairs): две пары карт одного значения (исключая тройку и каре).

(8) Пара (pair) — две карты одного значения, исключая старшие комбинации (две пары, тройка, фулл хаус, каре).

Вероятности комбинаций в покере

Количество комбинаций:

$$C_{52}^{5} = \frac{52!}{(52-5)!5!} = 2\,598\,960.$$

- Стрит флеш (straight flush): 5 карт одной масти по порядку.
  Начало от двойки до десятки (9 вариантов), 4 масти: 36 вариантов (рук). Вероятность:

$$36/2598960 \approx 1.385 \cdot 10^{-5} \approx 1/72193.3$$

- Каре (four of a kind): 4 карты одного значения.
  4 карты одного значения можно выбрать 13 способами. Пятую карту после этого 48 способами: $13 \cdot 48$ вариантов. Вероятность:

$$13 \cdot 48/2598960 \approx 2.401 \cdot 10^{-4} = 1/4165.$$

Вероятности комбинаций в покере

Количество комбинаций:

$$C_{52}^5 = \frac{52!}{(52-5)!5!} = 2\,598\,960.$$

- Стрит флеш (straight flush): 5 карт одной масти по порядку. Начало от двойки до десятки (9 вариантов), 4 масти: 36 вариантов (рук). Вероятность:

$$36/2598960 \approx 1.385 \cdot 10^{-5} \approx 1/72193.3$$

- Каре (four of a kind): 4 карты одного значения. 4 карты одного значения можно выбрать 13 способами. Пятую карту после этого 48 способами: $13 \cdot 48$ вариантов. Вероятность:

$$13 \cdot 48/2598960 \approx 2.401 \cdot 10^{-4} = 1/4165.$$

Вероятности комбинаций в покере

Количество комбинаций:

$$C_{52}^5 = \frac{52!}{(52-5)!5!} = 2\,598\,960.$$

- Стрит флеш (straight flush): 5 карт одной масти по порядку. Начало от двойки до десятки (9 вариантов), 4 масти: 36 вариантов (рук). Вероятность:

$$36/2598960 \approx 1.385 \cdot 10^{-5} \approx 1/72193.3$$

- Каре (four of a kind): 4 карты одного значения.
  4 карты одного значения можно выбрать 13 способами. Пятую карту после этого 48 способами: $13 \cdot 48$ вариантов. Вероятность:

$$13 \cdot 48/2598960 \approx 2.401 \cdot 10^{-4} = 1/4165.$$

Вероятности комбинаций в покере

Количество комбинаций:

$$C_{52}^5 = \frac{52!}{(52-5)!5!} = 2\,598\,960.$$

- Стрит флеш (straight flush): 5 карт одной масти по порядку. Начало от двойки до десятки (9 вариантов), 4 масти: 36 вариантов (рук). Вероятность:

$$36/2598960 \approx 1.385 \cdot 10^{-5} \approx 1/72193.3$$

- Каре (four of a kind): 4 карты одного значения. 4 карты одного значения можно выбрать 13 способами. Пятую карту после этого 48 способами: $13 \cdot 48$ вариантов. Вероятность:

$$13 \cdot 48/2598960 \approx 2.401 \cdot 10^{-4} = 1/4165.$$

- Фулл хаус (full house): 3 карты одного значения и две карты другого значения.

  $C_4^3 = 4$ способа выбрать 3 карты одного фиксированного значения. Всего 13 значений. Значит $4 \cdot 13 = 52$ варианта выбрать три одного значения.

  $C_4^2 = 6$ способов выбрать 2 карты одного значения. Из оставшихся 12 значений выбираем две карты: $6 \cdot 12 = 72$.

  Всего $52 \cdot 72 = 3744$ варианта (руки). Вероятность:

  $$3744/2598960 \approx 1.441 \cdot 10^{-3} \approx 1/694.2.$$

- Фулл хаус (full house): 3 карты одного значения и две карты другого значения.

  $C_4^3 = 4$ способа выбрать 3 карты одного фиксированного значения. Всего 13 значений. Значит $4 \cdot 13 = 52$ варианта выбрать три одного значения.

  $C_4^2 = 6$ способов выбрать 2 карты одного значения. Из оставшихся 12 значений выбираем две карты: $6 \cdot 12 = 72$.

  Всего $52 \cdot 72 = 3744$ варианта (руки). Вероятность:

  $$3744/2598960 \approx 1.441 \cdot 10^{-3} \approx 1/694.2.$$

- Флеш (flush): 5 карт одной масти (исключая стрит флеш).

  C_{13}^5 способов выбрать 5 карт фиксированной масти. Выбросить 9 стрит флешей. Всего $4 \cdot (C_{13}^5 - 9) = 5112$ вариантов.

  Вероятность:

$$5112/2598960 \approx 1.967 \cdot 10^{-2} \approx 1/508.4.$$

- Флеш (flush): 5 карт одной масти (исключая стрит флеш).

  C_{13}^5 способов выбрать 5 карт фиксированной масти. Выбросить 9 стрит флешей. Всего $4 \cdot (C_{13}^5 - 9) = 5112$ вариантов.

  Вероятность:

$$5112/2598960 \approx 1.967 \cdot 10^{-2} \approx 1/508.4.$$

- Стрит (straight): 5 карт по порядку (исключая стрит флеш).

Начиная с 10 имеется 4^5 вариантов. Перебираем все начала отсчета (их 9): $9 \cdot 4^5$ вариантов (рук). Выбрасываем стрит флеши: 36 вариантов. Вероятность:

$$(9 \cdot 4^5 - 36)/2598960 \approx 3.532 \cdot 10^{-2} \approx 1/283.1$$

- Стрит (straight): 5 карт по порядку (исключая стрит флеш).

  Начиная с 10 имеется 4^5 вариантов. Перебираем все начала отсчета (их 9): $9 \cdot 4^5$ вариантов (рук). Выбрасываем стрит флеши: 36 вариантов. Вероятность:

  $$(9 \cdot 4^5 - 36)/2598960 \approx 3.532 \cdot 10^{-2} \approx 1/283.1$$

- Тройка (three of a kind): 3 карты одного значения (исключая каре и фулл хаус).

Количество способов выбрать 3 карты одного значения: 52 (см. фулл хаус). Из оставшихся $12 \cdot 4 = 48$ значений (чтобы не получилось каре) выбираем любую карту. Все значения этого значения также будут запрещены. Из оставшихся 44 карт выбираем любую. Получаем количество троек и упорядоченных наборов из двух карт:

$$52 \cdot 48 \cdot 44$$

Убираем упорядочение:

$$52 \frac{48 \cdot 44}{2!} = 54192$$

вариантов. Вероятность:

$$54912/2598960 \approx 2.113 \cdot 10^{-2} \approx 1/47.3$$

- Тройка (three of a kind): 3 карты одного значения (исключая каре и фулл хаус).

  Количество способов выбрать 3 карты одного значения: 52 (см. фулл хаус). Из оставшихся $12 \cdot 4 = 48$ значений (чтобы не получилось каре) выбираем любую карту. Все значения этого значения также будут запрещены. Из оставшихся 44 карт выбираем любую. Получаем количество троек и упорядоченных наборов из двух карт:
$$52 \cdot 48 \cdot 44$$
  Убираем упорядочение:
$$52\frac{48 \cdot 44}{2!} = 54192$$
  вариантов. Вероятность:
$$54912/2598960 \approx 2.113 \cdot 10^{-2} \approx 1/47.3$$

- Две пары (two pairs): две пары карт одинакового значения (исключая тройку и каре):

$C_4^2 \cdot 13 \cdot C_4^2 \cdot 12$ — количество упорядоченных наборов из двух пар разного значения.
$C_4^2 \cdot 13 \cdot C_4^2 \cdot 12/2!$ — количество неупорядоченных наборов из двух пар разного значения.
Выбираем 5-ю карту из оставшихся $52 - 8 = 44$:

$$\frac{C_4^2 \cdot 13 \cdot C_4^2 \cdot 12}{2!} \cdot 44 = 123552$$

вариантов. Вероятность:

$$123552/2598960 \approx 4.754 \cdot 10^{-2} \approx 1/21.$$

- Две пары (two pairs): две пары карт одинакового значения (исключая тройку и каре):

  $C_4^2 \cdot 13 \cdot C_4^2 \cdot 12$ — количество упорядоченных наборов из двух пар разного значения.
  $C_4^2 \cdot 13 \cdot C_4^2 \cdot 12/2!$ — количество неупорядоченных наборов из двух пар разного значения.
  Выбираем 5-ю карту из оставшихся $52 - 8 = 44$:

  $$\frac{C_4^2 \cdot 13 \cdot C_4^2 \cdot 12}{2!} \cdot 44 = 123552$$

  вариантов. Вероятность:

  $$123552/2598960 \approx 4.754 \cdot 10^{-2} \approx 1/21.$$

- Пара (pair) — две карты одного значения, исключая старшие комбинации (две пары, тройка, фулл хаус, каре).

$13 \cdot C_4^2$ способов выбрать пару. Из оставшихся 48 карт выбираем любую. Все карты этого значения будут запрещены. Из оставшихся 44 карт выбираем любую. Все карты этого значения будут запрещены. Из оставшихся 40 карт выбираем любую. Получаем пару и упорядоченный набор из 3 карт:

$$13 \cdot C_4^2 \cdot 48 \cdot 44 \cdot 40.$$

Считаем количество неупорядоченных наборов:

$$13 \cdot C_4^2 \frac{48 \cdot 44 \cdot 40}{3!} = 1098240$$

Вероятность:

$$1098240/2598960 \approx 0.423 \approx 1/2.366.$$

- Пара (pair) — две карты одного значения, исключая старшие комбинации (две пары, тройка, фулл хаус, каре).

  $13 \cdot C_4^2$ способов выбрать пару. Из оставшихся 48 карт выбираем любую. Все карты этого значения будут запрещены. Из оставшихся 44 карт выбираем любую. Все карты этого значения будут запрещены. Из оставшихся 40 карт выбираем любую. Получаем пару и упорядоченный набор из 3 карт:

  $$13 \cdot C_4^2 \cdot 48 \cdot 44 \cdot 40.$$

  Считаем количество неупорядоченных наборов:

  $$13 \cdot C_4^2 \frac{48 \cdot 44 \cdot 40}{3!} = 1098240$$

  Вероятность:

  $$1098240/2598960 \approx 0.423 \approx 1/2.366.$$

Случайные величины

Измеримым пространством называется пара $(\Omega, \mathscr{F})$, где $\mathscr{F}$ — сигма алгебра подмножеств Ω.
Функция $\xi : \Omega \mapsto \mathbb{R}^d$ называется $\mathscr{F}$*-измеримой*, если множество $\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}$ принадлежит $\mathscr{F}$ для любого борелевского множества $B \subset \mathbb{R}^d$. Напомним, что борелевской называется наименьшая σ-алгебра $\mathscr{B}(\mathbb{R}^d)$, содержащая все открытые подмножества $\mathbb{R}^d$.
В теории вероятностей такие функции ξ называются *случайными величинами*. Считается, что после проведения вероятностного эксперимента для любого события из $\mathscr{F}$ становится известным, произошло оно или нет. Таким образом, $\mathscr{F}$ описывает доступную информацию. $\mathscr{F}$-измеримость ξ означает, что для любого борелевского множества $B \subset \mathbb{R}^d$ можно сказать, верно ли, что $\xi(\omega) \in B$ на основе информации, заключенной в $\mathscr{F}$.

Пример

$\mathscr{F} = \{\emptyset, \Omega\}$, ξ is $\mathscr{F}$-measurable $\iff$ $\xi = \mathrm{const}$.

Пример

$\Omega = \{(a_1, a_2) : a_i \in \{H, T\}\}$, $\mathscr{F}_1 = \sigma(A_H, A_T)$,

$$\xi_1 = I_{\{a_1 = H\}}, \quad \xi_2 = I_{\{a_2 = H\}},$$

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

ξ_1 является $\mathscr{F}_1$-измеримой, ξ_2 не является $\mathscr{F}_1$-измеримой.

Пример

$\mathscr{F} = \{\emptyset, \Omega\}$, ξ is $\mathscr{F}$-measurable $\iff \xi = \mathrm{const}.$

Пример

$\Omega = \{(a_1, a_2) : a_i \in \{H, T\}\}$, $\mathscr{F}_1 = \sigma(A_H, A_T)$,

$$\xi_1 = I_{\{a_1 = H\}}, \quad \xi_2 = I_{\{a_2 = H\}},$$

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

ξ_1 является $\mathscr{F}_1$-измеримой, ξ_2 не является $\mathscr{F}_1$-измеримой.

Пример

Рассмотрим разбиение с атомами $(A_i)_{i=1}^m$. При каком условии $\xi : \Omega \mapsto \mathbb{R}$ измерима относительно $\mathscr{F} = \sigma\left((A_i)_{i=1}^m\right)$?

- ξ — $\mathscr{F}$-измерима $\iff$ ξ постоянна на атомах.

Доказательство. Если ξ принимает хотя бы два различных значения на некотором атоме, то $\{\omega : \xi(\omega) = z\}$ не является объединением атомов для некоторого z. С другой стороны, любая функция, которая постоянна на атомах является $\mathscr{F}$-измеримой: достаточно рассмотреть прообразы ее значений.

Пример

Рассмотрим разбиение с атомами $(A_i)_{i=1}^m$. При каком условии $\xi : \Omega \mapsto \mathbb{R}$ измерима относительно $\mathscr{F} = \sigma\left((A_i)_{i=1}^m\right)$?

- ξ — $\mathscr{F}$-измерима $\iff$ ξ постоянна на атомах.

Доказательство. Если ξ принимает хотя бы два различных значения на некотором атоме, то $\{\omega : \xi(\omega) = z\}$ не является объединением атомов для некоторого z. С другой стороны, любая функция, которая постоянна на атомах является $\mathscr{F}$-измеримой: достаточно рассмотреть прообразы ее значений.

Пример

Рассмотрим измеримое пространство $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Случайными величинами $\xi : \mathbb{R} \mapsto \mathbb{R}$ на таком пространстве являются в точности борелевские вектор-функции, т.е. функции удовлетворяющие условию $\{\omega : \xi(\omega) \in B\} \in \mathscr{B}(\mathbb{R})$, $B \in \mathscr{B}(\mathbb{R})$.

Пример

Пусть $\mathscr{F}_\xi = \sigma(\xi) := \sigma(\xi^{-1}(B), B \in \mathscr{B}(\mathbb{R}))$ — σ-алгебра порожденная случайной величиной ξ. Это наименьшая σ-алгебра относительно которой ξ является измеримой.

Алгебраические операции и предельный переход сохраняют измеримость:

$$\xi \pm \eta, \quad \xi\eta, \quad \xi/\eta, \quad \lim_{n\to\infty} \xi_n, \quad \liminf_{n\to\infty} \xi_n, \quad \limsup_{n\to\infty} \xi_n.$$

Все определения переносятся на случай расширенных случайных величин:

$$\xi : \Omega \mapsto \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}.$$

Пример

Рассмотрим измеримое пространство $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Случайными величинами $\xi : \mathbb{R} \mapsto \mathbb{R}$ на таком пространстве являются в точности борелевские вектор-функции, т.е. функции удовлетворяющие условию $\{\omega : \xi(\omega) \in B\} \in \mathscr{B}(\mathbb{R})$, $B \in \mathscr{B}(\mathbb{R})$.

Пример

Пусть $\mathscr{F}_\xi = \sigma(\xi) := \sigma(\xi^{-1}(B), B \in \mathscr{B}(\mathbb{R}))$ — σ-алгебра порожденная случайной величиной ξ. Это наименьшая σ-алгебра относительно которой ξ является измеримой.

Алгебраические операции и предельный переход сохраняют измеримость:

$$\xi \pm \eta, \quad \xi\eta, \quad \xi/\eta, \quad \lim_{n\to\infty} \xi_n, \quad \liminf_{n\to\infty} \xi_n, \quad \limsup_{n\to\infty} \xi_n.$$

Все определения переносятся на случай расширенных случайных величин:

$$\xi : \Omega \mapsto \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}.$$

Пример

Рассмотрим измеримое пространство $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Случайными величинами $\xi : \mathbb{R} \mapsto \mathbb{R}$ на таком пространстве являются в точности борелевские вектор-функции, т.е. функции удовлетворяющие условию $\{\omega : \xi(\omega) \in B\} \in \mathscr{B}(\mathbb{R})$, $B \in \mathscr{B}(\mathbb{R})$.

Пример

Пусть $\mathscr{F}_\xi = \sigma(\xi) := \sigma(\xi^{-1}(B), B \in \mathscr{B}(\mathbb{R}))$ — σ-алгебра порожденная случайной величиной ξ. Это наименьшая σ-алгебра относительно которой ξ является измеримой.

Алгебраические операции и предельный переход сохраняют измеримость:

$$\xi \pm \eta, \quad \xi\eta, \quad \xi/\eta, \quad \lim_{n\to\infty} \xi_n, \quad \liminf_{n\to\infty} \xi_n, \quad \limsup_{n\to\infty} \xi_n.$$

Все определения переносятся на случай расширенных случайных величин:

$$\xi : \Omega \mapsto \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}.$$

Распределение

Распределением случайной величины ξ называется вероятностная мера P_ξ на $\mathscr{B}(\mathbb{R})$:

$$\mathsf{P}_\xi(B) := \mathsf{P}(\xi^{-1}(B)), \quad B \in \mathscr{B}(\mathbb{R}).$$

Таким образом, каждая случайная величина ξ порождает вероятностное пространство $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \mathsf{P}_\xi)$.

Дискретные случайные величины

ξ называется дискретной, если существует такое конечное или счетное множество Z, что

$$\mathsf{P}(\xi \in Z) = \mathsf{P}_\xi(Z) = 1.$$

Фактически ξ принимает только значения из Z: $\mathsf{P}(\xi \in Z^c) = 0$.
Распределение P_ξ определяется значениями функции

$$p_\xi(z) := \mathsf{P}(\xi = z) = \mathsf{P}_\xi(z),$$

которая называется вероятностной массой (probability mass function: pmf). Другое название: дискретное распределение.

Пример

Модель двукратного подбрасывания монеты с вероятностью выпадения орла равной p, $\Omega = \{(a_1, a_2) : a_i \in \{H, T\}\}$, $q = 1 - p$

- $\xi : \Omega \mapsto \{0, 1, 2\}$ — число орлов:
$$\xi(\omega) = I_{\{a_1 = H\}} + I_{\{a_2 = H\}}$$
- $\eta : \Omega \mapsto \{-2, 0, 2\}$ — разность между числом орлов и решек:
$$\eta(\omega) = \sum_{i=1}^{2} \left(I_{\{a_i = H\}} - I_{\{a_i = T\}}\right)$$
- $\zeta : \Omega \mapsto \{0, 1\}$ — число орлов при первом подбрасывании:
$$\zeta(\omega) = I_{\{a_1 = H\}}.$$

Пример (продолжение)

- $\xi(HH) = 2, \quad \xi(HT) = \xi(TH) = 1, \quad \xi(TT) = 0$

z	0	1	2
$p_\xi(z)$	q^2	$2pq$	p^2

- $\eta(HH) = 2, \quad \eta(HT) = \eta(TH) = 0, \quad \eta(TT) = -2$

z	-2	0	2
$p_\eta(z)$	q^2	$2pq$	p^2

- $\zeta(HH) = \zeta(HT) = 1, \quad \zeta(TH) = \zeta(TT) = 0$

z	0	1
$p_\zeta(z)$	q	p

Пример (продолжение)

- $\xi(HH) = 2, \quad \xi(HT) = \xi(TH) = 1, \quad \xi(TT) = 0$

z	0	1	2
$p_\xi(z)$	q^2	$2pq$	p^2

- $\eta(HH) = 2, \quad \eta(HT) = \eta(TH) = 0, \quad \eta(TT) = -2$

z	-2	0	2
$p_\eta(z)$	q^2	$2pq$	p^2

- $\zeta(HH) = \zeta(HT) = 1, \quad \zeta(TH) = \zeta(TT) = 0$

z	0	1
$p_\zeta(z)$	q	p

Пример (продолжение)

- $\xi(HH) = 2, \quad \xi(HT) = \xi(TH) = 1, \quad \xi(TT) = 0$

z	0	1	2
$p_\xi(z)$	q^2	$2pq$	p^2

- $\eta(HH) = 2, \quad \eta(HT) = \eta(TH) = 0, \quad \eta(TT) = -2$

z	-2	0	2
$p_\eta(z)$	q^2	$2pq$	p^2

- $\zeta(HH) = \zeta(HT) = 1, \quad \zeta(TH) = \zeta(TT) = 0$

z	0	1
$p_\zeta(z)$	q	p

Любая функция $p : Z \mapsto [0,1]$, удовлетворяющая условию

$$\sum_{z \in Z} p(z) = 1,$$

является вероятностной массой для некоторой случайной величины: $(\Omega, \mathscr{F}, \mathsf{P})$:

$$\Omega := Z, \quad \mathscr{F} := 2^Z, \quad \mathsf{P}(\omega) := p(\omega), \quad \xi(\omega) := \omega.$$

Поэтому часто случайные величины задаются с помощью вероятностной массы (распределения) без явного описания вероятностного пространства.

Пример

Распределение Бернулли: $\xi \sim \mathrm{Ber}(p)$,

z	0	1
p_ξ	$1-p$	p

Встречалось в предыдущем примере.

Пример

Биномиальное распределение: $\xi \sim \mathrm{B}(n,p)$, $q = 1-p$

z	0	$\dots$	k	$\dots$	n
p_ξ	q^n	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	p^n

Пусть в схеме Бернулли $\xi_i = I_{\{a_i = H\}}$,

$$\xi = \sum_{i=1}^{n} \xi_i$$

Пример

Распределение Бернулли: $\xi \sim \mathrm{Ber}(p)$,

z	0	1
p_ξ	$1-p$	p

Встречалось в предыдущем примере.

Пример

Биномиальное распределение: $\xi \sim \mathrm{B}(n,p)$, $q = 1-p$

z	0	$\dots$	k	$\dots$	n
p_ξ	q^n	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	p^n

Пусть в схеме Бернулли $\xi_i = I_{\{a_i = H\}}$,

$$\xi = \sum_{i=1}^{n} \xi_i$$

Пример (продолжение)

Тогда

$$p_\xi(k) = \mathsf{P}(\xi = k) = \mathsf{P}(\text{количество орлов равно } k) = C_n^k p^k q^{n-k}$$

Встречалось при подсчете вероятности выпадения k орлов.

Пример

Дискретное равномерное распределение: $\xi \sim DU(a, b)$, $a, b \in \mathbb{Z}_+$,

z	a	$\ldots$	b
p_ξ	$1/(b-a+1)$	$\ldots$	$1/(b-a+1)$

Например, ξ — число очков на грани симметричного кубика, который подбрасывается один раз ($a = 1$, $b = 6$).

Пример (продолжение)

Тогда

$$p_\xi(k) = \mathsf{P}(\xi = k) = \mathsf{P}(\text{количество орлов равно } k) = C_n^k p^k q^{n-k}$$

Встречалось при подсчете вероятности выпадения k орлов.

Пример

Дискретное равномерное распределение: $\xi \sim DU(a,b)$, $a, b \in \mathbb{Z}_+$,

z	a	$\ldots$	b
p_ξ	$1/(b-a+1)$	$\ldots$	$1/(b-a+1)$

Например, ξ — число очков на грани симметричного кубика, который подбрасывается один раз ($a = 1$, $b = 6$).

Пример

Геометрическое распределение: $\xi \sim G(p)$, $q = 1 - p$,

$$p_\xi(k) = pq^{k-1}, \quad k \in \mathbb{N}.$$

Корректность:

$$\sum_{k=1}^{\infty} pq^{k-1} = p(1 + q + q^2 + \dots) = \frac{p}{1-q} = 1.$$

Пример (продолжение)

Рассмотрим схему Бернулли с бесконечным числом испытаний и вероятностью успеха p: $\Omega = \{(a_i)_{i=1}^{\infty} : a_i \in \{H, T\}\}$,

$$\mathsf{P}(a_1, \dots a_n) = p^{\nu_n(\omega)} q^{n-\nu_n(\omega)}.$$

Пусть $\xi_i(\omega) = I_{\{a_i = H\}}$, $\xi = \min\{n \geq 1 : \xi_n = 1\}$ — время первого успеха $(\min \emptyset := +\infty)$. Тогда

$$p_\xi(k) = \mathsf{P}(\xi = k) = \mathsf{P}(\xi_1 = 0, \dots, \xi_{k-1} = 0, \xi_k = 1) = q^{k-1} p, \quad k \in \mathbb{N},$$

$$\mathsf{P}(\xi = \infty) \leq \mathsf{P}(\xi_1 = 0, \dots, \xi_k = 0) = q^k, \ \forall k \quad \Longrightarrow \quad \mathsf{P}(\xi = \infty) = 0.$$

Пример

Распределение Пуассона: $\xi \sim \Pi(\lambda)$, $\lambda > 0$,

$$p_\xi(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{Z}_+.$$

Данная вероятностная масса является пределом биномиального распределения при большом числе испытаний и малой вероятности успеха:

$$C_n^k p^k q^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!}, \quad n \to \infty, \quad p \to 0, \quad np \to \lambda.$$

Действительно,

$$C_n^k p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!}p^k(1-p)^{n-k}$$

Пример (продолжение)

$$\frac{n(n-1)\dots(n-k+1)}{k!}p^k = \frac{(1-1/n)\dots(1-(k-1)/n)}{k!}(np)^k \to \frac{\lambda^k}{k!}$$

$$(1-p)^{n-k} = e^{n(1-k/n)\ln(1-p)} = e^{n(1-k/n)(-p+o(p))} \to e^{-\lambda}.$$

$p_\xi(k)$ — вероятность того, что число успехов равно k, если число попыток велико и вероятность успеха в каждой попытке мала:

- число аварий автомобиля за год,
- число полисов, предъявленных к оплате,
- число опечаток в книге,
- число голов в футбольном матче.

Пример (продолжение)

$$\frac{n(n-1)\dots(n-k+1)}{k!}p^k = \frac{(1-1/n)\dots(1-(k-1)/n)}{k!}(np)^k \to \frac{\lambda^k}{k!}$$

$$(1-p)^{n-k} = e^{n(1-k/n)\ln(1-p)} = e^{n(1-k/n)(-p+o(p))} \to e^{-\lambda}.$$

$p_\xi(k)$ — вероятность того, что число успехов равно k, если число попыток велико и вероятность успеха в каждой попытке мала:

- число аварий автомобиля за год,
- число полисов, предъявленных к оплате,
- число опечаток в книге,
- число голов в футбольном матче.

Распределение преобразования дискретной случайной величины

Пусть ξ — дискретная случайная величина и $\eta = g(\xi)$. Тогда

$$p_\eta(y) = \mathsf{P}(g(\xi) = y) = \mathsf{P}(\xi \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} p_\xi(x).$$

Если g строго монотонна, то $p_\eta(y) = p_\xi(g^{-1}(y))$.

Пример

$\xi \sim DU(-1, 1)$:

x	-1	0	1
p_ξ	$1/3$	$1/3$	$1/3$

$\eta = |\xi|$, $\eta : \Omega \mapsto \{0, 1\}$, $p_\eta(1) = 2/3$, $p_\eta(0) = 1/3$, т.е.

$$\eta \sim \mathrm{Ber}(2/3).$$

Непрерывные случайные величины

Случайная величина ξ называется *непрерывной*, если существует борелевская функция $f_\xi : \mathbb{R} \mapsto \mathbb{R}_+$ такая, что

$$\mathsf{P}(\xi \in B) = \int_B f_\xi(x)dx, \quad B \in \mathscr{B}(\mathbb{R}).$$

Функция f_ξ называется *плотностью* распределения P_ξ (probability density function: pdf).
Заметим, что

$$\mathsf{P}(\xi \in [x, x+\delta]) \approx f_\xi(x)\delta,$$

$$f_\xi(x) = \lim_{\delta \to 0} \frac{\mathsf{P}(\xi \in [x, x+\delta])}{\delta},$$

$$\mathsf{P}(\xi \in [x, x]) = \mathsf{P}(\xi = x) = 0.$$

Пример

Равномерное распределение: $\xi \sim U(a, b)$, $a < b$,

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

Вероятность попадания на интервал $[c, d] \subset [a, b]$ пропорциональна длине интервала:

$$\mathsf{P}(\xi \in [c, d]) = \int_{[c,d]} f_\xi(x)\, dx = \frac{d - c}{b - a}.$$

Пример

Экспоненциальное распределение: $\xi \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$,

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

Корректность:

$$\int_{\mathbb{R}} f_\xi(x)\, dx = \int_0^\infty \lambda e^{-\lambda x}\, dx = 1.$$

Пример

Нормальное (гауссовское) распределение: $\xi \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$,

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

Стандартное нормальное распределение: $\xi \sim N(0, 1)$,

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

Корректность:

$$\int_{\mathbb{R}} f_\xi(x)\, dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1.$$

Пример (продолжение)

Действительно,

$$
\begin{aligned}
\left(\int_{\mathbb{R}} e^{-x^2/2}\, dx\right)^2 &= \int_{\mathbb{R}} e^{-x^2/2}\, dx \int_{\mathbb{R}} e^{-y^2/2}\, dy = \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy \\
&= \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r\, dr d\varphi = 2\pi \int_0^{\infty} e^{-r^2/2} r\, dr \\
&= 2\pi \int_0^{\infty} -d(e^{-r^2}) = 2\pi.
\end{aligned}
$$

Интеграл

Дано измеримое пространство $(\Omega, \mathscr{F})$. Пусть ξ — $\mathscr{F}$-измерима и μ — неотрицательная σ-аддитивная мера на $\mathscr{F}$: $\mu : \mathscr{F} \mapsto \mathbb{R} \cup \{+\infty\}$,

- $\mu(\emptyset) = 0$,
- $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$, $A_i \cap A_j = \emptyset$, $i \neq j$.

Типичным примером является мера Лебега λ на $\mathscr{B}(\mathbb{R}^d)$, которая однозначно определяется значениями на параллелепипедах:

$$\lambda([a_1, b_1] \times \cdots \times [a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i).$$

Вероятностная мера, естественно, также удовлетворяет данным условиям.

Простыми называются функции вида

$$\xi(\omega) = \sum_{j=1}^{n} c_j I_{A_j}(\omega), \quad A_j \in \mathscr{F}, \quad c_j \in \mathbb{R}.$$

Ясно, что такие функции являются $\mathscr{F}$-измеримыми.

Интеграл от простой функции по мере μ определяется следующим образом:

$$\int_\Omega \xi(\omega)\,\mu(d\omega) = \sum_{j=1}^{n} c_j \mu(A_j).$$

Используется также обозначение $\int_\Omega \xi\, d\mu$. Данное определение корректно: для различных представлений

$$\xi(\omega) = \sum_{j=1}^{n} c_j I_{A_j}(\omega) = \sum_{j=1}^{n} c'_j I_{A'_j}(\omega)$$

результат будет одним и тем же.
На любые неотрицательные измеримые функции ξ интеграл распространяется при помощи формулы

$$\int_\Omega \xi\, d\mu = \sup\left\{\int_\Omega \varphi\, d\mu,\ \text{где } \varphi \text{ простая и } 0 \le \varphi \le \xi\right\}.$$

При этом допускаются и бесконечные значения.

Далее, любая $\mathscr{F}$-измеримая функция может быть представлена в виде разности двух неотрицательных $\mathscr{F}$-измеримых функций: $\xi = \xi^+ - \xi^-$, где

$$\xi^+(\omega) := \xi(\omega) \vee 0 = \frac{|\xi| + \xi}{2}(\omega), \qquad \xi^-(\omega) := -(\xi(\omega) \wedge 0) = \frac{|\xi| - \xi}{2}(\omega)$$

и использованы обозначения

$$a \wedge b = \min\{a, b\}, \qquad a \vee b = \max\{a, b\}.$$

Если оба значения $\int_\Omega \xi^+ \, d\mathsf{P}$, $\int_\Omega \xi^- \, d\mathsf{P}$ конечны, то ξ называется *интегрируемой* и интеграл определяется формулой

$$\int_\Omega \xi \, d\mu = \int_\Omega \xi^+ \, d\mu - \int_\Omega \xi^- \, d\mu.$$

Для интеграла по мере Лебега принято обозначение

$$\int_{\mathbb{R}^d} f(x)\, dx := \int_{\mathbb{R}^d} f(x)\lambda(dx).$$

Интеграл по вероятностной мере P называется *математическим ожиданием* (expectation): $\mathsf{E}\xi = \int_\Omega \xi\, d\mathsf{P}$. Очевидно, что математическое ожидание обладает свойством линейности:

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

Формула замены переменных

Пусть g — (ограниченная) борелевская функция, тогда

$$\mathsf{E}g(\xi) = \int_{\mathbb{R}} g(x)\,\mathsf{P}_\xi(dx).$$

Схема доказательства.

- Пусть $g = I_B$, тогда

$$\begin{aligned}\mathsf{E}g = \mathsf{E}I_B(\xi) = \mathsf{E}I_{\{\xi\in B\}} = \mathsf{P}(\xi \in B) = \mathsf{P}_\xi(B) \\ = \int_{\mathbb{R}} I_B(x)\,\mathsf{P}_\xi(dx) = \int_{\mathbb{R}} g(x)\,\mathsf{P}_\xi(dx).\end{aligned}$$

- Для $g = \sum_{i=1}^m c_i I_{B_i}(\omega)$ — по линейности.
- Для $g \geq 0$ — предельный переход по монотонной последовательности простых функций: $0 \leq g_n \uparrow g$ (пока нет нужной теоремы).
- Для g произвольного знака — по линейности: $g = g^+ - g^-$.

Формула замены переменных

Пусть g — (ограниченная) борелевская функция, тогда

$$\mathsf{E}g(\xi) = \int_{\mathbb{R}} g(x)\, \mathsf{P}_\xi(dx).$$

Схема доказательства.

- Пусть $g = I_B$, тогда

$$\begin{aligned}\mathsf{E}g = \mathsf{E}I_B(\xi) = \mathsf{E}I_{\{\xi \in B\}} = \mathsf{P}(\xi \in B) = \mathsf{P}_\xi(B) \\ = \int_{\mathbb{R}} I_B(x)\, \mathsf{P}_\xi(dx) = \int_{\mathbb{R}} g(x)\, \mathsf{P}_\xi(dx).\end{aligned}$$

- Для $g = \sum_{i=1}^m c_i I_{B_i}(\omega)$ — по линейности.
- Для $g \geq 0$ — предельный переход по монотонной последовательности простых функций: $0 \leq g_n \uparrow g$ (пока нет нужной теоремы).
- Для g произвольного знака — по линейности: $g = g^+ - g^-$.

- ξ дискретна, $\mathsf{E}g(\xi) = \sum_{x \in Z} g(x) p_\xi(x)$,
- ξ непрерывна, $\mathsf{E}g(\xi) = \int_{\mathbb{R}} g(x) f_\xi(x)\, dx$.

Доказательство осуществляется по той же схеме, что и в общем случае. В частности,

- ξ дискретна, $\mathsf{E}\xi = \sum_{x \in Z} x p_\xi(x)$,
- ξ непрерывна, $\mathsf{E}\xi = \int_{\mathbb{R}} x f_\xi(x)\, dx$.

Стандартная техника

Описанная схема доказательства называется стандартной техникой (standard machine):

- для индикаторов (проверка равенства, определение),
- для линейных комбинаций индикаторов (линейность),
- для неотрицательных функций (монотонный предельный переход),
- для произвольных (линейность).

Дисперсия

Дисперсия (variance) случайной величины

$$\mathrm{Var}(\xi) := \mathsf{E}(\xi - \mathsf{E}\xi)^2 = \int (x - \mathsf{E}\xi)^2 \mathsf{P}_\xi(dx)$$

характеризует разброс ее значений вокруг среднего. Среднеквадратическое отклонение

$$\sigma(\xi) = \sqrt{\mathrm{Var}(\xi)}$$

играет ту же роль.
Заметим, что

$$\mathrm{Var}(\xi) := \mathsf{E}(\xi^2 - 2\xi\mathsf{E}\xi + (\mathsf{E}\xi)^2) = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2,$$

$$\mathrm{Var}(a\xi + b) = a^2\mathrm{Var}(\xi).$$

Математическое ожидание и дисперсия зависят только от распределения случайной величины. Таким образом, их значения для различных, но одинаково распределенных случайных величин будут одинаковыми.

$\xi \sim Ber(p)$

z	0	1
p_ξ	$1-p$	p

$$\mathsf{E}\xi = 0 \cdot p_\xi(0) + 1 \cdot p_\xi(1) = p,$$

$$\mathsf{E}(\xi^2) = 0 \cdot (1-p) + 1 \cdot p = p,$$

$$\mathrm{Var}(\xi) = \mathsf{E}(\xi^2) - (\mathsf{E}\xi)^2 = p - p^2 = p(1-p).$$

$\xi \sim B(n, p)$

$$p_\xi(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

Воспользуемся тем, что ξ имеет такое же распределение как $\eta = \sum_{i=1}^n \xi_i, \quad \xi_i = I_{\{a_i = H\}}$ в схеме Бернулли. Воспользуемся также линейностью математического ожидания:

$$\mathsf{E}\eta = \sum_{i=1}^n \mathsf{E}\xi_i = np.$$

Дисперсию вычислим позже.

$\xi \sim DU(a, b)$

Пусть $\eta = \xi - a$. Тогда $\eta \sim U(0, n)$, $n = b - a$,

$$\mathsf{E}\eta = \frac{1}{n+1}\sum_{k=0}^{n} k = \frac{1}{n+1}\frac{n(n+1)}{2} = \frac{n}{2},$$

$$\mathsf{E}\xi = \frac{n}{2} + a = \frac{b-a}{2} + a = \frac{a+b}{2}.$$

$$\mathsf{E}\eta^2 = \frac{1}{n+1}\sum_{k=0}^{n} k^2$$

По индукции докажем, что

$$\sum_{k=0}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

$\xi \sim DU(a,b)$ (продолжение)

При $n = 0$ верно. Пусть верно при $n-1$, тогда

$$\sum_{k=0}^{n} k^2 = \sum_{k=0}^{n-1} k^2 + n^2 = \frac{(n-1)n(2(n-1)+1)}{6} + n^2 = \frac{1}{6} n(n+1)(2n+1).$$

$$\begin{aligned} \mathrm{Var}(\eta) &= \mathsf{E}\eta^2 - (\mathsf{E}\eta)^2 = \frac{1}{n+1}\frac{1}{6} n(n+1)(2n+1) - \frac{n^2}{4} \\ &= \frac{2n^2+n}{6} - \frac{n^2}{4} = \frac{n(n+2)}{12} \end{aligned}$$

$$\mathrm{Var}(\xi) = \mathrm{Var}(\eta) = \frac{(b-a)(b-a+2)}{12}.$$

$\xi \sim G(p)$

$$p_\xi(k) = p(1-p)^{k-1},$$

$$\begin{aligned} \mathsf{E}\xi &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -p\frac{d}{dp}\sum_{k=1}^{\infty}(1-p)^k = -p\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k \\ &= -p\frac{d}{dp}\frac{1}{1-(1-p)} = -p\frac{d}{dp}\frac{1}{p} = \frac{1}{p}. \end{aligned}$$

Дисперсию вычислим позже.

$\xi \sim \Pi(\lambda)$

$$p_\xi(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{Z}_+.$$

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

$$\mathsf{E}\xi^2 = \sum_{k=0}^{\infty} k^2 p_\xi(k) = e^{-\lambda}\sum_{k=1}^{\infty}\frac{k\lambda^k}{(k-1)!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(k+1)\lambda^{k+1}}{k!}$$

$$= \lambda\left(e^{-\lambda}\sum_{k=0}^{\infty}\frac{k\lambda^k}{k!} + e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right) = \lambda(\lambda+1).$$

$$\mathrm{Var}(\xi) = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.$$

$\xi \sim \mathrm{Exp}(\lambda)$

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

$$\mathsf{E}\xi = \lambda \int_0^\infty x e^{-\lambda x}\, dx = -\int_0^\infty x\, d(e^{-\lambda x}) = \int_0^\infty e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

$$\mathsf{E}\xi^2 = \lambda \int_0^\infty x^2 e^{-\lambda x}\, dx = -\int_0^\infty x^2 d\left(e^{-\lambda x}\right) = \int_0^\infty 2x e^{-\lambda x}\, dx$$
$$= \frac{2}{\lambda}\mathsf{E}\xi = \frac{2}{\lambda^2}.$$

$$\mathrm{Var}(\xi) = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$\xi \sim N(0,1)$

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

$$\mathsf{E}\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\, dx = 0.$$

$$\begin{aligned} \mathrm{Var}(\xi) = \mathsf{E}\xi^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, de^{-x^2/2} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1. \end{aligned}$$

Формула для плотности строго монотонного преобразования непрерывной случайной величины

Пусть $\eta = g(\xi)$, где g — строго монотонная борелевская функция. Для любой ограниченной борелевской функции h имеем

$$\begin{aligned} \mathsf{E}h(\eta) = \mathsf{E}h(g(\xi)) &= \int_{\mathbb{R}} h(g(x))\,\mathsf{P}_\xi(dx) = \int_{\mathbb{R}} h(g(x)) f_\xi(x)\,dx \\ &= \int_{\mathbb{R}} h(y) f_\xi(g^{-1}(y))|(g^{-1})'(y)|\,dy = \int_{\mathbb{R}} h(y) f_\eta(y)\,dy \end{aligned}$$

В силу произвольности h,

$$f_\eta(y) = |(g^{-1})'(y)| f_\xi(g^{-1}(y)).$$

$\xi \sim U(0,1)$

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

Пусть

$$\eta = a + (b-a)\xi = g(\xi), \quad a < b.$$

В данном случае

$$g(x) = a + (b-a)x, \quad g^{-1}(y) = \frac{y-a}{b-a}, \quad (g^{-1})'(y) = \frac{1}{b-a}.$$

Следовательно,

$$f_\eta(y) = \frac{1}{b-a} f_\xi\left(\frac{y-a}{b-a}\right) = \begin{cases} \frac{1}{b-a}, & \frac{y-a}{b-a} \in [0,1] \\ 0, & \text{иначе} \end{cases} = \begin{cases} \frac{1}{b-a}, & y \in [a,b], \\ 0, & \text{иначе.} \end{cases}$$

Таким образом, $\eta \sim U(a,b)$.

Поскольку

$$\mathsf{E}\xi = \int_0^1 x\,dx = \frac{1}{2},$$

$$\mathsf{E}\xi^2 = \int_0^1 x^2\,dx = \frac{1}{3},$$

$$\mathrm{Var}(\xi) = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12},$$

то для $\eta \sim U(a,b)$ имеем

$$\mathsf{E}(\eta) = \mathsf{E}(a + (b-a)\xi) = a + \frac{b-a}{2} = \frac{a+b}{2},$$

$$\mathrm{Var}(\eta) = \mathrm{Var}(a + (b-a)\xi) = (b-a)^2 \cdot \mathrm{Var}(\xi) = \frac{(b-a)^2}{12}.$$

$\xi \sim N(0,1)$

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Пусть

$$\eta = \mu + \sigma\xi = g(\xi), \quad \sigma > 0.$$

В данном случае

$$g(x) = \mu + \sigma x, \quad g^{-1}(y) = \frac{y-\mu}{\sigma}, \quad (g^{-1})'(y) = \frac{1}{\sigma}.$$

Следовательно,

$$f_\eta(y) = \frac{1}{\sigma} f_\xi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

Таким образом, $\eta \sim N(\mu, \sigma^2)$ и для таких случайных величин

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

$$\mathrm{Var}(\eta) = \sigma^2 \mathrm{Var}(\xi) = \sigma^2.$$

Функции распределения

Функция $F_\xi(x) = \mathsf{P}(\xi \le x)$ называется *функцией распределения* (cdf: cumulative distribution function) случайной величины ξ.

$$F_\xi(x) = \begin{cases} \sum_{y \le x} p_\xi(y), & \xi \text{ дискретна}, \\ \int_{-\infty}^{x} f_\xi(y)\,dy, & \xi \text{ непрерывна}. \end{cases}$$

$$F_\xi(x) \le F_\xi(y), \qquad \text{если } x \le y,$$

$$\lim_{x \to -\infty} F_\xi(x) = 0, \qquad \lim_{x \to +\infty} F_\xi(x) = 1.$$

Для непрерывной случайной величины $F'_\xi(x) = f_\xi(x)$ в точках непрерывности функции f_ξ.
Кроме того, функции распределения непрерывны справа:

$$\lim_{y \downarrow x} F_\xi(y) = \lim_{Q \ni y \downarrow x} \mathsf{P}(\xi \le y) = \mathsf{P}(\cap_{y \ge x, y \in \mathbb{Q}} \{\xi \le y\}) = \mathsf{P}(\xi \le x) = F_\xi(x).$$

Во втором равенстве использована непрерывность вероятностной меры.

Непрерывность вероятностной меры

Пусть $A_k \subseteq A_{k+1}$, $k \geq 1$, $A = \cup_{k=1}^{\infty} A_k$. Тогда

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}(A).$$

Доказательство.

Пусть $A_0 = \varnothing$, тогда

$$\mathsf{P}(A) = \mathsf{P}\left(\cup_{k=1}^{\infty} A_k\right) = \mathsf{P}\left(\cup_{k=1}^{\infty} (A_k \backslash A_{k-1})\right) = \sum_{k=1}^{\infty} \mathsf{P}(A_k \backslash A_{k-1})$$

$$= \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathsf{P}(A_k \backslash A_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} (\mathsf{P}(A_k) - \mathsf{P}(A_{k-1}))$$

$$= \lim_{n\to\infty} \mathsf{P}(A_n). \qquad \square$$

Аналогичное утверждение для убывающих множеств:

$$A_k \supseteq A_{k+1}, k \geq 1 \Longrightarrow \mathsf{P}(\cap_{k=1}^{\infty} A_k) = \lim_{n\to\infty} \mathsf{P}(A_n)$$

можно вывести из закона де Моргана.

Пусть $\xi \sim G(p)$, тогда

$$F_\xi(x) = \mathsf{P}(\xi \le x) = \sum_{1 \le k \le x} p_\xi(k) = \sum_{1 \le k \le x} pq^{k-1},$$

$$F_\xi(n) = \sum_{k=1}^{n} pq^{k-1} = p\frac{1-q^n}{1-q} = 1 - q^n, \quad n \ge 1.$$

Пусть $\eta \sim \mathrm{Exp}(\lambda)$, тогда

$$F_\eta(x) = \int_0^x \lambda e^{-\lambda y}\, dy = -e^{-\lambda y}\Big|_{y=0}^{y=x} = 1 - e^{-\lambda x}, \quad x \ge 0.$$

Пусть $q = e^{-\lambda}$, тогда

$$F_\xi(n) = 1 - e^{-\lambda n} = F_\eta(n).$$

Пусть $\xi \sim N(0,1)$, тогда

$$\Phi(x) := F_\xi(x) = \mathsf{P}(\xi \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

Правило 3σ. $\Phi(3) \approx 0.9987$. Пусть $\eta \in N(\mu, \sigma^2)$. Тогда

$$\xi = \frac{\eta - \mu}{\sigma} \in N(0,1).$$

$$\begin{aligned}
\mathsf{P}(|\eta - \mu| \leq 3\sigma) &= \mathsf{P}(|\xi| \leq 3) = \mathsf{P}(\xi \leq 3) - \mathsf{P}(\xi \leq -3) \\
&= \mathsf{P}(\xi \leq 3) - (1 - \mathsf{P}(\xi \leq 3)) = 2\mathsf{P}(\xi \leq 3) - 1 \\
&= 2\Phi(3) - 1 \approx 0.997
\end{aligned}$$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983

Санкт-Петербургский парадокс

- Nicholas Bernoulli (1713)
- Daniel Bernoulli (1738), в издании Императорской Академии наук, Санкт-Петербург

Симметричная монета подбрасывается до тех пор, пока выпадет орел. Казино платит игроку 2^{n-1} рублей, если это произошло при n-м подбрасывании. Какова справедливая плата для участия в такой игре?

Момент τ окончания игры имеет геометрическое распределение, $\tau \sim G(1/2)$:

$$\mathsf{P}(\tau = n) = \frac{1}{2^n}, \quad n \in \mathbb{N}.$$

Распределение выигрыша ξ игрока:

$$p_\xi(2^{n-1}) := \mathsf{P}(\{\xi = 2^{n-1}\}) = \mathsf{P}(\{\omega : \tau(\omega) = n\}) = \frac{1}{2^n}, \quad n \in \mathbb{N}.$$

Санкт-Петербургский парадокс

- Nicholas Bernoulli (1713)
- Daniel Bernoulli (1738), в издании Императорской Академии наук, Санкт-Петербург

Симметричная монета подбрасывается до тех пор, пока выпадет орел. Казино платит игроку 2^{n-1} рублей, если это произошло при n-м подбрасывании. Какова справедливая плата для участия в такой игре?

Момент τ окончания игры имеет геометрическое распределение, $\tau \sim G(1/2)$:

$$\mathsf{P}(\tau = n) = \frac{1}{2^n}, \quad n \in \mathbb{N}.$$

Распределение выигрыша ξ игрока:

$$p_\xi(2^{n-1}) := \mathsf{P}(\{\xi = 2^{n-1}\}) = \mathsf{P}(\{\omega : \tau(\omega) = n\}) = \frac{1}{2^n}, \quad n \in \mathbb{N}.$$

Санкт-Петербургский парадокс (продолжение)

Математическое ожидание выигрыша (справедливая плата за участие в игре):

$$\mathsf{E}\xi = \sum_{n\in\mathbb{N}} 2^{n-1} p_\xi(n) = \sum_{n\in\mathbb{N}} \frac{1}{2} = \infty.$$

В то же время вероятность выиграть не более, скажем, $16 = 2^4$ рублей равна

$$\mathsf{P}(\tau \leq 5) = \sum_{n=1}^{5} \frac{1}{2^n} \approx 0.969$$

Один из подходов к разрешению парадокса (de Condorcet, Poisson, Catalan, Cramer). Казино не выплачивает выигрыши, превосходящие 2^N. В этом случае выигрыш игрока

$$\xi = \sum_{j=1}^{N+1} 2^{j-1} I_{\{\tau=j\}} + 2^N I_{\{\tau\geq N+2\}}.$$

Санкт-Петербургский парадокс (продолжение)

Математическое ожидание выигрыша (справедливая плата за участие в игре):

$$\mathsf{E}\xi = \sum_{n\in\mathbb{N}} 2^{n-1} p_\xi(n) = \sum_{n\in\mathbb{N}} \frac{1}{2} = \infty.$$

В то же время вероятность выиграть не более, скажем, $16 = 2^4$ рублей равна

$$\mathsf{P}(\tau \leq 5) = \sum_{n=1}^{5} \frac{1}{2^n} \approx 0.969$$

Один из подходов к разрешению парадокса (de Condorcet, Poisson, Catalan, Cramer). Казино не выплачивает выигрыши, превосходящие 2^N. В этом случае выигрыш игрока

$$\xi = \sum_{j=1}^{N+1} 2^{j-1} I_{\{\tau=j\}} + 2^N I_{\{\tau\geq N+2\}}.$$

Санкт-Петербургский парадокс (продолжение)

Поскольку

$$\mathsf{E} I_{\{\tau \ge N+2\}} = \mathsf{P}(\tau \ge N+2) = \sum_{j \ge N+2} \mathsf{P}(\tau = j),$$

то

$$\mathsf{E}\xi = \sum_{j=1}^{N+1} \frac{1}{2} + 2^N \sum_{j \ge N+2} \frac{1}{2^j} = \frac{1}{2}(N+1) + \frac{1}{2} = \frac{N}{2} + 1.$$

N	Максимальный выигрыш	Справедливая плата
4	16	3
5	32	3.5
10	1024	6
15	32768	8.5
20	1048576	11

Санкт-Петербургский парадокс (продолжение)

Поскольку

$$\mathsf{E} I_{\{\tau \geq N+2\}} = \mathsf{P}(\tau \geq N+2) = \sum_{j \geq N+2} \mathsf{P}(\tau = j),$$

то

$$\mathsf{E}\xi = \sum_{j=1}^{N+1} \frac{1}{2} + 2^N \sum_{j \geq N+2} \frac{1}{2^j} = \frac{1}{2}(N+1) + \frac{1}{2} = \frac{N}{2} + 1.$$

N	Максимальный выигрыш	Справедливая плата
4	16	3
5	32	3.5
10	1024	6
15	32768	8.5
20	1048576	11

Санкт-Петербургский парадокс (продолжение)

Другой подход к разрешению парадокса (Daniel Bernoulli). Для игрока ценность денег определяется функцией полезности: ценность x равна $U(x)$. Приращение полезности:

$$U(x + \Delta x) - U(x) \approx A \frac{\Delta x}{x}.$$

"выигрыш 1000 дукатов более значим для бедняка, чем для богатого человека" (Д. Бернулли)
Из дифференциального уравнения

$$\frac{dU}{dx} = \frac{A}{x}$$

находим

$$U(x) = A \ln x + B.$$

Для игрока с капиталом x справедливая плата $c(x)$ за участие в игре определяется из уравнения

$$U(x) = \mathsf{E} U(x + \xi - c).$$

$$\ln(x) = \mathsf{E}\ln(x + \xi - c) = \sum_{n=1}^{\infty} \frac{1}{2^n} \ln(2^{n-1} + x - c).$$

```
import numpy as np
from scipy import optimize
import math
def eq(c):
    s=0
    for n in range(1,101):
        s+=(math.log(2**(n-1)+x-c)-math.log(x))/2**n
    return s
for x in [1,2,5,10,1000,10000,100000,1000000]:
    print('x=',repr(x).ljust(7),'|','c(x)=',optimize.bisect(lambda c: eq(c), 0.
↪01, x+0.99))
```

```
x= 1       | c(x)= 1.6737849095727688
x= 2       | c(x)= 1.9999999999988085
x= 5       | c(x)= 2.485010013342313
x= 10      | c(x)= 2.8837618245817453
x= 1000    | c(x)= 5.968017344459698
x= 10000   | c(x)= 7.617581141106898
x= 100000  | c(x)= 9.276533622403035
x= 1000000 | c(x)= 10.937183977278348
```

Функции полезности и несклонность к риску

Будем использовать функцию полезности U для сравнения случайных капиталов:

$$\xi \preceq \eta \iff \mathsf{E}U(\xi) \le \mathsf{E}U(\eta).$$

Наложим следующие условия:

$$\xi = x \le \eta = y \Longrightarrow x \preceq y,$$

$$\xi \preceq \mathsf{E}\xi \quad \text{что означает} \quad \mathsf{E}U(\xi) \le \mathsf{E}U(\mathsf{E}\xi) = U(\mathsf{E}\xi).$$

Из первого условия вытекает, что функция U неубывающая

$$U(x) \le U(y), \quad x \le y,$$

из второго — что она вогнута:

$$\mathsf{P}(\xi = x_i) = p_i, \quad i = 1, 2,$$

$$\mathsf{E}U(\xi) = p_1 U(x_1) + p_2 U(x_2) \le U(\mathsf{E}\xi) = U(p_1 x_1 + p_2 x_2).$$

Пусть инвестор имеет капитал x и ему предстоит сделать выбор: подвергнуть свой капитал риску изменения на малую случайную величину $\varepsilon\xi$ с $\mathsf{E}\xi = 0$ или заплатить страховую сумму и уменьшить свой капитал на y_ε. Субъективная ожидаемая полезность этих альтернатив одинакова, если

$$U(x - y_\varepsilon) = \mathsf{E}U(x + \varepsilon\xi).$$

Чем бо́льшую страховую сумму y_ε готов платить инвестор при фиксированной величине ξ, тем менее он склонен к риску.
По формуле Тейлора

$$U(x - y_\varepsilon) = U(x) - U'(x)y_\varepsilon + o(y_\varepsilon)$$

$$\begin{aligned}\mathsf{E}U(X + \varepsilon\xi) &= U(x) + \varepsilon U'(x)\mathsf{E}\xi + \frac{\varepsilon^2}{2}U''(x)\mathsf{E}\xi^2 + o(\varepsilon^2)\\ &= U(x) + \frac{\varepsilon^2}{2}U''(x)\mathrm{Var}(\xi) + o(\varepsilon^2).\end{aligned}$$

Приравнивая эти выражения, заключаем, что

$$y_\varepsilon \approx \frac{\varepsilon^2}{2}\mathrm{Var}(\xi)\cdot r(x), \quad r(x) = -\frac{U''(x)}{U'(x)}.$$

Величина $r(x)$ называется *коэффициентом несклонности к риску*. Поскольку $U'' \leq 0$ для любой дважды дифференцируемой вогнутой функции U, то $r(x) \geq 0$.
Найдем коэффициенты несклонности к риску нескольких стандартных функций полезности

$$\begin{aligned} U(x) = \ln x; &\quad r(x) = 1/x; \\ U(x) = x^\gamma,\ \gamma \in (0,1); &\quad r(x) = (1-\gamma)/x; \\ U(x) = -e^{-\alpha x}; &\quad r(x) = \alpha. \end{aligned}$$

Парадокс Алле (Allais, 1953)

- В игре 1 вы заведомо выигрываете 500 000$.
- В игре 2 вы выигрываете

2 500 000$	с вероятностью 0.1
500 000$	с вероятностью 0.89
0$	с вероятностью 0.01

Какую игру вы предпочтете?

- В игре 3 вы выигрываете

500 000$	с вероятностью 0.11
0$	с вероятностью 0.89

- В игре 4 вы выигрываете

2 500 000$	с вероятностью 0.1
0$	с вероятностью 0.9

Какую игру вы предпочтете?

Парадокс Алле (Allais, 1953)

- В игре 1 вы заведомо выигрываете 500 000\$.
- В игре 2 вы выигрываете

2 500 000\$	с вероятностью 0.1
500 000\$	с вероятностью 0.89
0\$	с вероятностью 0.01

Какую игру вы предпочтете?

- В игре 3 вы выигрываете

500 000\$	с вероятностью 0.11
0\$	с вероятностью 0.89

- В игре 4 вы выигрываете

2 500 000\$	с вероятностью 0.1
0\$	с вероятностью 0.9

Какую игру вы предпочтете?

Для большинства людей $1 \succ 2$, $4 \succ 3$. Пусть предпочтения игрока определяются функцией полезности U:

$$X \succ Y \iff \mathsf{E}U(X) > \mathsf{E}U(Y).$$

Пусть $x_1 = 0$, $x_2 = 500\,000$, $x_3 = 2\,500\,000$, и предположим, что $U(0) = 0$. Тогда $1 \succ 2$ означает, что

$$U(x_2) > 0.1U(x_3) + 0.89U(x_2),$$

а $4 \succ 3$ означает, что

$$0.1U(x_3) > 0.11U(x_2).$$

Получаем противоречие:

$$U(x_2) > 0.11U(x_2) + 0.89U(x_2) = U(x_2).$$

Таким образом, не существует функции полезности, которая описывает такие предпочтения.

Для большинства людей $1 \succ 2$, $4 \succ 3$. Пусть предпочтения игрока определяются функцией полезности U:

$$X \succ Y \iff \mathsf{E}U(X) > \mathsf{E}U(Y).$$

Пусть $x_1 = 0$, $x_2 = 500\,000$, $x_3 = 2\,500\,000$, и предположим, что $U(0) = 0$. Тогда $1 \succ 2$ означает, что

$$U(x_2) > 0.1U(x_3) + 0.89U(x_2),$$

а $4 \succ 3$ означает, что

$$0.1U(x_3) > 0.11U(x_2).$$

Получаем противоречие:

$$U(x_2) > 0.11U(x_2) + 0.89U(x_2) = U(x_2).$$

Таким образом, не существует функции полезности, которая описывает такие предпочтения.

Оптимальная величина ставки в азартной игре

Пусть ставка игрока удваивается с вероятностью p и теряется с вероятностью $q = 1 - p$. Найти оптимальную величину ставки для заданной функции полезности U.

Пусть $\xi \sim \mathrm{Ber}(p)$. Через δ обозначим долю капитала x, которую ставит игрок. Цель состоит в максимизации ожидаемой полезности:

$$\varphi(\delta) = \mathsf{E}U(x + 2\xi\delta x - \delta x) = pU(x + \delta x) + qU(x - \delta x) \to \max_{\delta \in [0,1]} .$$

Если функция U выпукла, то функция φ также выпукла и ее максимум достигается на границе интервала $[0, 1]$. Таким образом $\delta^* \in \{0, 1\}$, т.е. оптимально ставить все или ничего в зависимости от величины p.

Оптимальная величина ставки в азартной игре

Пусть ставка игрока удваивается с вероятностью p и теряется с вероятностью $q = 1 - p$. Найти оптимальную величину ставки для заданной функции полезности U.

Пусть $\xi \sim \mathrm{Ber}(p)$. Через δ обозначим долю капитала x, которую ставит игрок. Цель состоит в максимизации ожидаемой полезности:

$$\varphi(\delta) = \mathsf{E}U(x + 2\xi\delta x - \delta x) = pU(x + \delta x) + qU(x - \delta x) \to \max_{\delta \in [0,1]}.$$

Если функция U выпукла, то функция φ также выпукла и ее максимум достигается на границе интервала $[0, 1]$. Таким образом $\delta^* \in \{0, 1\}$, т.е. оптимально ставить все или ничего в зависимости от величины p.

Если функция U вогнута и $p \le q$ (игра неблагоприятна), то

$$\begin{aligned} pU(x+\delta x) + qU(x-\delta x) &\le U(p(x+\delta x) + q(x-\delta x)) \\ &= U(x+\delta(p-q)x) \le U(x). \end{aligned}$$

Таким образом, $\delta^* = 0$ — одно из оптимальных решений. Единственный интересный случай, когда U вогнута и игра благоприятна: $p > q$. Пусть $U(x) = \ln x$. Тогда

$$\varphi(\delta) = \ln x + p\ln(1+\delta) + q\ln(1-\delta).$$

Оптимальное значение δ не зависит от x и определяется из уравнения

$$\frac{p}{1+\delta} - \frac{q}{1-\delta} = 0.$$

Таким образом, $\delta^* = p - q$.

Многомерные случайные величины

Пусть $(\xi_i)_{i=1}^d$ — $\mathscr{F}$-измеримые случайные величины на $(\Omega, \mathscr{F}, \mathsf{P})$. d-мерной случайной величиной называется отображение

$$\xi = (\xi_1, \dots, \xi_d) : \Omega \mapsto \mathbb{R}^d.$$

Распределением d-мерной случайной величины ξ (или *совместным распределением* $(\xi_1, \dots, \xi_d)$) называется вероятностная мера P_ξ на $\mathscr{B}(\mathbb{R}^d)$:

$$\mathsf{P}_\xi(B) := \mathsf{P}(\xi^{-1}(B)), \quad B \in \mathscr{B}(\mathbb{R}^d).$$

Дискретные и непрерывные многомерные случайные величины

ξ называется *дискретной*, если существует такое конечное или счетное множество $Z \subset \mathbb{R}^d$, что

$$\mathsf{P}(\xi \in Z) = \mathsf{P}_\xi(Z) = 1.$$

В этом случае P_ξ определяется значениями

$$p_\xi(x) = p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_d) = \mathsf{P}_\xi(x),$$

которые называются вероятностной массой (распределением вероятностей, дискретным распределением).
ξ называется *непрерывной*, если существует борелевская функция $f_\xi = f_{\xi_1,\ldots,\xi_d} : \mathbb{R}^d \mapsto \mathbb{R}_+$ такая, что

$$\mathsf{P}(\xi \in B) = \int_B f_\xi(x)dx, \quad B \in \mathscr{B}(\mathbb{R}^d).$$

Функция f_ξ называется *плотностью* распределения P_ξ (совместной плотностью $(\xi_1, \ldots, \xi_d)$).

Формула замены переменных

Как и в одномерном случае, справедлива формула замены переменных:

$$\mathsf{E}g(\xi_1, \ldots, \xi_d) = \int_{\mathbb{R}^d} g(x_1, \ldots, x_d) \mathsf{P}_\xi(dx_1 \ldots dx_d).$$

В частности,

$$\mathsf{E}\xi_i = \int_{\mathbb{R}^d} x_i \mathsf{P}_\xi(dx_1 \ldots dx_d).$$

Для дискретного совместного распределения:

$$\mathsf{E}g(\xi_1, \ldots, \xi_d) = \sum_{(x_1, \ldots, x_d) \in Z} g(x_1, \ldots, x_d) p_{\xi_1, \ldots, \xi_d}(x_1, \ldots, x_d).$$

Для непрерывного совместного распределения

$$\mathsf{E}g(\xi_1, \ldots, \xi_d) = \int_{\mathbb{R}^d} g(x_1, \ldots, x_d)\, f_{\xi_1, \ldots, \xi_d}(x_1, \ldots, x_d)\, dx_1 \ldots dx_d$$

Маргинальные распределения

Для упрощения обозначений будем рассматривать двумерные случайные величины (ξ, η). Зная совместное распределение (ξ, η), можно найти распределения ξ, η, которые называются маргинальными:

$$\mathsf{P}_\xi(B) = \mathsf{P}(\xi \in B) = \mathsf{P}(\xi \in B, \eta \in \mathbb{R}) = \mathsf{P}_{\xi,\eta}(B \times \mathbb{R}).$$

Для дискретного совместного распределения $(Z_x := \{y : (x, y) \in Z\})$:

$$p_\xi(x) = \mathsf{P}(\xi = x) = \mathsf{P}(\xi = x, \eta \in \mathbb{R}) = \mathsf{P}(\xi = x, \eta \in Z_x) = \sum_{y \in Z_x} p_{\xi,\eta}(x, y).$$

Для непрерывного совместного распределения:

$$\begin{aligned} \mathsf{P}(\xi \in B) &= \int_B f_\xi(x)\,dx = \mathsf{P}(\xi \in B, \eta \in \mathbb{R}) = \mathsf{P}_{\xi,\eta}(B \times \mathbb{R}) \\ &= \int_{B\times\mathbb{R}} f_{\xi,\eta}(x, y)\,dxdy = \int_B \int_{\mathbb{R}} f_{\xi,\eta}(x, y)\,dydx. \end{aligned}$$

Следовательно, $f_\xi(x) = \int_{\mathbb{R}} f_{\xi,\eta}(x, y)\,dy$.

Независимость

σ-алгебры $\mathscr{F}_1, \dots, \mathscr{F}_n$ называются независимыми, если

$$\mathsf{P}(A_1 \cap \dots \cap A_n) = \mathsf{P}(A_1) \dots \mathsf{P}(A_n).$$

ξ не зависит от $\mathscr{H}$, если $\mathscr{F}_\xi$ и $\mathscr{H}$ независимы

$$\mathsf{P}(\{\xi \in B\} \cap A) = \mathsf{P}(\xi \in B)\mathsf{P}(A), \quad B \in \mathscr{B}(\mathbb{R}), A \in \mathscr{H}.$$

$(\xi_i)_{i=1}^n$ независимы, если σ-алгебры $\mathscr{F}_{\xi_i}$ независимы:

$$\mathsf{P}(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \mathsf{P}(\xi_1 \in B_1) \dots \mathsf{P}(\xi_n \in B_n)$$

Другими словами,

$$\mathsf{P}_{\xi_1,\dots,\xi_n}(B_1, \dots, B_n) = \mathsf{P}_{\xi_1}(B_1) \dots \mathsf{P}_{\xi_n}(B_n).$$

Для дискретных ξ_i:

$$p_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \mathsf{P}_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \mathsf{P}_{\xi_1}(x_1)\dots\mathsf{P}_{\xi_n}(x_n)$$
$$= p_{\xi_1}(x_1)\dots p_{\xi_n}(x_n).$$

Для непрерывных ξ_i:

$$\int_{B_1}\cdots\int_{B_n} f_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)\,dx_1\dots dx_n = \mathsf{P}_{\xi_1,\dots,\xi_n}(B_1,\dots,B_n)$$
$$=\mathsf{P}_{\xi_1}(B_1)\dots\mathsf{P}_{\xi_n}(B_n) = \int_{B_1} f_{\xi_1}(x_1)\,dx_1\cdots\int_{B_n} f_{\xi_n}(x_n)\,dx_n$$
$$=\int_{B_1}\cdots\int_{B_n} f_{\xi_1}(x_1)\dots f_{\xi_n}(x_n)\,dx_1\dots dx_n.$$

Таким образом,

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

Пусть ξ, η независимы, $g, h : \mathbb{R} \mapsto \mathbb{R}$ — борелевские функции. Тогда $g(\xi)$, $h(\eta)$ независимы:

$$\begin{aligned}&\mathsf{P}(g(\xi) \in B, h(\eta) \in C) = \mathsf{P}(\xi \in g^{-1}(B), \eta \in h^{-1}(C)) \\ &= \mathsf{P}(\xi \in g^{-1}(B))\mathsf{P}(\eta \in h^{-1}(C)) = \mathsf{P}(g(\xi) \in B)\mathsf{P}(h(\eta) \in C).\end{aligned}$$

Пример: двумерная дискретная случайная величина

[Bertsekas, Tsitsiklis, p. 94]

Таблица: Совместное распределение $p_{\xi,\eta}(i,j)$ и маргинальные распределения $p_\xi(i)$, $p_\eta(j)$

					$p_\eta(j)$
4	0	1/20	1/20	1/20	3/20
3	1/20	2/20	3/20	1/20	7/20
2	1/20	2/20	3/20	1/20	7/20
1	1/20	1/20	1/20	0	3/20
j/i	1	2	3	4	
$p_\xi(i)$	3/20	6/20	8/20	3/20	

$$\mathsf{P}(\xi = \eta) = \frac{1}{20} + \frac{2}{20} + \frac{3}{20} + \frac{1}{20} = \frac{7}{20}.$$

$$\frac{p_{\xi,\eta}(i,j)}{p_\xi(i)} \neq p_\eta(j) \Longrightarrow \xi, \eta \text{ зависимы.}$$

(продолжение)

Пусть $\zeta = \xi + 2\eta$. Тогда

$$\begin{aligned}\mathsf{E}\zeta = \mathsf{E}\xi + 2\mathsf{E}\eta &= \sum_{i=1}^{4} i p_\xi(i) + 2\sum_{j=1}^{4} j p_\eta(j) \\ &= 1\cdot\frac{3}{20} + 2\cdot\frac{6}{20} + 3\cdot\frac{8}{20} + 4\cdot\frac{3}{20} \\ &+ 2\left(1\cdot\frac{3}{20} + 2\cdot\frac{7}{20} + 3\cdot\frac{7}{20} + 4\cdot\frac{3}{20}\right) \\ &= \frac{51}{20} + \frac{100}{20} = 7.55\end{aligned}$$

Распределение функции двумерной дискретной случайной величины

Пусть $\zeta = g(\xi, \eta)$. Тогда

$$p_\zeta(z) = \mathsf{P}(\zeta = z) = \mathsf{P}(g(\xi,\eta) = z) = \mathsf{P}((\xi,\eta) \in \{(x,y) : g(x,y) = z)\})$$
$$= \sum_{\{(x,y):g(x,y)=z\}} p_{\xi,\eta}(x,y)$$

В рассмотренном выше примере, $\zeta = \xi + 2\eta$,

$$p_\zeta(3) = p_{\xi,\eta}(1,1) = \frac{1}{20}, \quad p_\zeta(4) = p_{\xi,\eta}(2,1) = \frac{1}{20},$$
$$p_\zeta(5) = p_{\xi,\eta}(1,2) + p_{\xi,\eta}(3,1) = \frac{1}{20} + \frac{1}{20} = \frac{2}{20},$$
$$p_\zeta(6) = p_{\xi,\eta}(2,2) + p_{\xi,\eta}(4,1) = \frac{2}{20} + 0 = \frac{2}{20},$$

$$
\begin{gathered}
p_\zeta(7) = p_{\xi,\eta}(1,3) + p_{\xi,\eta}(3,2) = \frac{1}{20} + \frac{3}{20} = \frac{4}{20}, \\
p_\zeta(8) = p_{\xi,\eta}(2,3) + p_{\xi,\eta}(4,2) = \frac{2}{20} + \frac{1}{20} = \frac{3}{20}, \\
p_\zeta(9) = p_{\xi,\eta}(1,4) + p_{\xi,\eta}(3,2) = 0 + \frac{3}{20} = \frac{3}{20}, \\
p_\zeta(10) = p_{\xi,\eta}(2,4) + p_{\xi,\eta}(4,3) = \frac{1}{20} + \frac{1}{20} = \frac{2}{20}, \\
p_\zeta(11) = p_{\xi,\eta}(3,4) = \frac{1}{20}, \quad p_\zeta(12) = p_{\xi,\eta}(4,4) = \frac{1}{20}.
\end{gathered}
$$

$$
\begin{gathered}
\mathsf{E}\zeta = \sum_{i=3}^{12} i p_\zeta(i) = 3\cdot\frac{1}{20} + 4\cdot\frac{1}{20} + 5\cdot\frac{2}{20} + 6\cdot\frac{2}{20} + 7\cdot\frac{4}{20} \\
+ 8\cdot\frac{3}{20} + 9\cdot\frac{3}{20} + 10\cdot\frac{2}{20} + 11\cdot\frac{1}{20} + 12\cdot\frac{1}{20} = 7.55.
\end{gathered}
$$

Распределение максимума независимых случайных величин

Пусть $\xi_i \sim DU(1, 10)$ независимы,

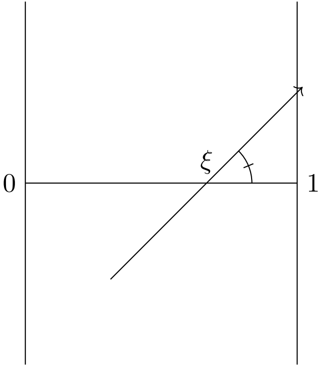
$$\eta = \max_{1 \le i \le n} \xi_i,$$

$$F_\eta(k) = \mathsf{P}(\xi_1 \le k, \ldots, \xi_n \le k) = \prod_{i=1}^{n} \mathsf{P}(\xi_i \le k) = \left(\frac{k}{10}\right)^n,$$

$$p_\eta(k) = F_\eta(k) - F_\eta(k-1) = \left(\frac{k}{10}\right)^n - \left(\frac{k-1}{10}\right)^n, \quad k = 1, \ldots, 10.$$

Игла Бюффона (1777 г.)

На полосу бесконечной длины и единичной ширины случайным образом бросается игла единичной длины. Какова вероятность того, что она пересечет границу полосы?



$\xi \in [0,1]$ — центр иглы, $\theta \in [-\pi/2, \pi/2]$ — угол с осью абсцисс.

Игла Бюффона (продолжение)

Игла пересекает границу полосы, если

- $\xi \in [0, 1/2]$ и $\xi - \frac{1}{2}\cos\theta \le 0$

или

- $\xi \in [1/2, 1]$ и $\xi + \frac{1}{2}\cos\theta \ge 1$

Другими словами, событие A, состоящее в том, что игла пересекает границу полосы имеет вид

$$\left\{\xi \in \left[0, \frac{1}{2}\cos\theta\right] \cup \left[1 - \frac{1}{2}\cos\theta, 1\right], \theta \in [-\pi/2, \pi/2]\right\}.$$

Предположим, что $\xi \sim U(0,1)$, $\theta \sim U(-\pi/2, \pi/2)$ и ξ, η независимы.

Игла Бюффона (продолжение)

Тогда

$$
\begin{aligned}
\mathsf{P}(A) &= \mathsf{E}\left(I_{\{\xi\in[0,\frac{1}{2}\cos\theta]\}} I_{\{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\}} + I_{\{\xi\in[1-\frac{1}{2}\cos\theta,1]\}} I_{\{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\}}\right)\\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\int_0^1 \left(I_{\{x\in[0,\frac{1}{2}\cos\varphi]\}} + I_{\{x\in[1-\frac{1}{2}\cos\varphi,1]\}}\right) dx d\varphi\\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\left(\int_0^{\frac{1}{2}\cos\varphi} dx + \int_{1-\frac{1}{2}\cos\varphi}^{1} dx\right) d\varphi\\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos\varphi\, d\varphi = \frac{2}{\pi}
\end{aligned}
$$

Считая, что при большом числе испытаний частота наступления события близка к его вероятности, можно использовать эту формулу для приближенного вычисления числа π (метод Монте-Карло). В 1850 г. Р. Вольф (астроном, Цюрих) подбросил иглу 5000 раз и получил для π приближенное значение 3.1596.

Игла Бюффона (продолжение)

Тогда

$$\begin{aligned}
\mathsf{P}(A) &= \mathsf{E}\left(I_{\{\xi\in[0,\frac{1}{2}\cos\theta]\}} I_{\{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\}} + I_{\{\xi\in[1-\frac{1}{2}\cos\theta,1]\}} I_{\{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\}}\right) \\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\int_0^1 \left(I_{\{x\in[0,\frac{1}{2}\cos\varphi]\}} + I_{\{x\in[1-\frac{1}{2}\cos\varphi,1]\}}\right) dx d\varphi \\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\left(\int_0^{\frac{1}{2}\cos\varphi} dx + \int_{1-\frac{1}{2}\cos\varphi}^1 dx\right) d\varphi \\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos\varphi\, d\varphi = \frac{2}{\pi}
\end{aligned}$$

Считая, что при большом числе испытаний частота наступления события близка к его вероятности, можно использовать эту формулу для приближенного вычисления числа π (метод Монте-Карло). В 1850 г. Р. Вольф (астроном, Цюрих) подбросил иглу 5000 раз и получил для π приближенное значение 3.1596.

Можно ли найти наибольшее из двух неизвестных чисел?

На двух листках бумаги написаны два различных вещественных числа. Игрок выбирает и переворачивает один из листков. Увидев написанное число, он должен сказать, больше ли оно, чем число на втором листке.

- *Существует метод, позволяющий угадать правильный ответ с вероятностью больше, чем 0.5.*

Пусть $a < b$ указанные неизвестные числа. Обозначим через X_1 число, которое видит игрок, и через X_2 — скрытое число. Выбор игрока случаен, поэтому

$$\mathsf{P}(X_1 = a) = \mathsf{P}(X_1 = b) = \frac{1}{2}.$$

Скрытое число определяется по формуле

$$X_2 = aI_{\{X_1=b\}} + bI_{\{X_1=a\}}.$$

Можно ли найти наибольшее из двух неизвестных чисел?

На двух листках бумаги написаны два различных вещественных числа. Игрок выбирает и переворачивает один из листков. Увидев написанное число, он должен сказать, больше ли оно, чем число на втором листке.

- *Существует метод, позволяющий угадать правильный ответ с вероятностью больше, чем 0.5.*

Пусть $a < b$ указанные неизвестные числа. Обозначим через X_1 число, которое видит игрок, и через X_2 — скрытое число. Выбор игрока случаен, поэтому

$$\mathsf{P}(X_1 = a) = \mathsf{P}(X_1 = b) = \frac{1}{2}.$$

Скрытое число определяется по формуле

$$X_2 = aI_{\{X_1=b\}} + bI_{\{X_1=a\}}.$$

Можно ли найти наибольшее из двух неизвестных чисел?

На двух листках бумаги написаны два различных вещественных числа. Игрок выбирает и переворачивает один из листков. Увидев написанное число, он должен сказать, больше ли оно, чем число на втором листке.

- *Существует метод, позволяющий угадать правильный ответ с вероятностью больше, чем 0.5.*

Пусть $a < b$ указанные неизвестные числа. Обозначим через X_1 число, которое видит игрок, и через X_2 — скрытое число. Выбор игрока случаен, поэтому

$$\mathsf{P}(X_1 = a) = \mathsf{P}(X_1 = b) = \frac{1}{2}.$$

Скрытое число определяется по формуле

$$X_2 = aI_{\{X_1=b\}} + bI_{\{X_1=a\}}.$$

Рассмотрим любую непрерывную случайную величину ξ с положительной плотностью $f_\xi(x) > 0$, $x \in \mathbb{R}$ (например, $\xi \sim N(0,1)$), *не зависящую* от X_1, X_2. Игрок может применить следующее правило выбора:

$$Y = X_1 I_{\{\xi \leq X_1\}} + X_2 I_{\{\xi > X_1\}}$$

т.е. выбрать первое число, если $\xi(\omega) \leq X_1(\omega)$ и второе в противном случае. Здесь $\xi(\omega)$ — реализация случайной величины (полученная с помощью соответствующего датчика случайных чисел).

Заметим, что

$$YI_{\{\xi<a\}} = X_1 I_{\{\xi\le X_1\}} I_{\{\xi<a\}} = X_1 I_{\{\xi<a\}}$$

$$YI_{\{\xi>b\}} = X_2 I_{\{\xi>X_1\}} I_{\{\xi>b\}} = X_2 I_{\{\xi>b\}}$$

$$\begin{aligned} YI_{\{a<\xi<b\}} &= X_1 I_{\{\xi\le X_1\}} I_{\{a<\xi<b\}} + X_2 I_{\{\xi>X_1\}} I_{\{a<\xi<b\}} \\ &= bI_{\{\xi\le X_1\}} I_{\{a<\xi<b\}} + bI_{\{\xi>X_1\}} I_{\{a<\xi<b\}} \\ &= bI_{\{a<\xi<b\}} \end{aligned}$$

Итак, игрок выбирает

- первое число, если $\xi < a$,
- второе число, если $\xi > b$,
- *наибольшое число*, т.е. правильный ответ, если $a < \xi < b$.

Заметим, что

$$YI_{\{\xi<a\}} = X_1 I_{\{\xi\le X_1\}} I_{\{\xi<a\}} = X_1 I_{\{\xi<a\}}$$

$$YI_{\{\xi>b\}} = X_2 I_{\{\xi> X_1\}} I_{\{\xi>b\}} = X_2 I_{\{\xi>b\}}$$

$$\begin{aligned} YI_{\{a<\xi<b\}} &= X_1 I_{\{\xi\le X_1\}} I_{\{a<\xi<b\}} + X_2 I_{\{\xi> X_1\}} I_{\{a<\xi<b\}} \\ &= bI_{\{\xi\le X_1\}} I_{\{a<\xi<b\}} + bI_{\{\xi> X_1\}} I_{\{a<\xi<b\}} \\ &= bI_{\{a<\xi<b\}} \end{aligned}$$

Итак, игрок выбирает

- первое число, если $\xi < a$,
- второе число, если $\xi > b$,
- *наибольшое число*, т.е. правильный ответ, если $a < \xi < b$.

$$\begin{aligned}
\mathsf{P}(Y=b|\xi<a) &= \frac{\mathsf{P}(Y=b,\xi<a)}{\mathsf{P}(\xi<a)} = \frac{\mathsf{P}(X_1=b,\xi<a)}{\mathsf{P}(\xi<a)} \\
&= \frac{\mathsf{P}(X_1=b)\mathsf{P}(\xi<a)}{\mathsf{P}(\xi<a)} = \mathsf{P}(X_1=b) = \frac{1}{2}, \\
\mathsf{P}(Y=b|\xi>b) &= \frac{\mathsf{P}(Y=b,\xi>b)}{\mathsf{P}(\xi>b)} = \frac{\mathsf{P}(X_2=b,\xi>b)}{\mathsf{P}(\xi>b)} \\
&= \frac{\mathsf{P}(X_2=b)\mathsf{P}(\xi>b)}{\mathsf{P}(\xi>b)} = \mathsf{P}(X_2=b) = \frac{1}{2}, \\
\mathsf{P}(Y=b|a<\xi<b) &= \frac{\mathsf{P}(Y=b,a<\xi<b)}{\mathsf{P}(a<\xi<b)} = \frac{\mathsf{P}(a<\xi<b)}{\mathsf{P}(a<\xi<b)} = 1.
\end{aligned}$$

По формуле полной вероятности, вероятность угадать правильный ответ равна

$$\begin{aligned}\mathsf{P}(Y = b) = {} & \mathsf{P}(Y = b \mid \xi < a)\mathsf{P}(\xi < a) \\ & + \mathsf{P}(Y = b \mid a < \xi < b)\mathsf{P}(a < \xi < b) \\ & + \mathsf{P}(Y = b \mid \xi > b)\mathsf{P}(\xi > b).\end{aligned}$$

Здесь учтено, что $\mathsf{P}(\xi \in \{a, b\}) = 0$. Итак,

$$\begin{aligned}\mathsf{P}(Y = b) &= \frac{1}{2}\mathsf{P}(\xi < a) + 1 \cdot \mathsf{P}(a < \xi < b) + \frac{1}{2}\mathsf{P}(\xi > b) \\ &= \frac{1}{2} + \frac{1}{2}\mathsf{P}(a < \xi < b) = \frac{1}{2} + \frac{1}{2}\int_a^b f_\xi(x)\, dx > \frac{1}{2}.\end{aligned}$$

Мат. ожидание произведения независимых с.в.

Пусть ξ, η независимы, тогда $\mathsf{E}(\xi\eta) = \mathsf{E}(\xi)\mathsf{E}(\eta)$ (при выполнении условий интегрируемости).

Схема доказательства.

- Пусть $\xi = I_{A_1}, \eta = I_{A_2}$, $A_1, A_2 \in \mathscr{F}$, тогда

$$\begin{aligned} \mathsf{E}(\xi\eta) &= \mathsf{E}(I_{A_1 \cap A_2}) = \mathsf{P}(A_1 \cap A_2) = \mathsf{P}(A_1)\mathsf{P}(A_2) \\ &= \mathsf{E}(I_{A_1})\mathsf{E}(I_{A_2}) = \mathsf{E}(\xi)\mathsf{E}(\eta) \end{aligned}$$

- ξ — простая функция, $\eta = I_{A_2}$: равенство вытекает из линейности по первому аргументу.
- ξ — неотрицательная функция, $\eta = I_{A_2}$: равенство выводится с помощью предельного перехода по монотонной последовательности простых функций.
- ξ — функция произвольного знака: равенство выводится из линейности по первому аргументу ($\xi = \xi^+ - \xi^-$). Приходим к равенству $\mathsf{E}(\xi I_{A_2}) = \mathsf{E}(\xi)\mathsf{E}(I_{A_2})$.
- повторяем рассуждения применительно ко второму аргументу (отдельно для ξ^+ и ξ^-).

Дисперсия суммы независимых с.в.

Пусть ξ, η независимы, тогда

$$\mathrm{Var}(\xi + \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta).$$

Действительно, пусть сначала $\mathsf{E}\xi = \mathsf{E}\eta = 0$. Тогда

$$\mathrm{Var}(\xi+\eta) = \mathsf{E}(\xi+\eta)^2 = \mathsf{E}\xi^2+2\mathsf{E}(\xi\eta)+\mathsf{E}\eta^2 = \mathsf{E}\xi^2+\mathsf{E}\eta^2 = \mathrm{Var}(\xi)+\mathrm{Var}(\eta),$$

так как

$$\mathsf{E}(\xi\eta) = \mathsf{E}(\xi)\mathsf{E}(\eta) = 0.$$

В общем случае

$$\begin{aligned}\mathrm{Var}(\xi + \eta) &= \mathrm{Var}(\xi - \mathsf{E}\xi + \eta - \mathsf{E}\eta) = \mathrm{Var}(\xi - \mathsf{E}\xi) + \mathrm{Var}(\eta - \mathsf{E}\eta)\\ &= \mathrm{Var}(\xi) + \mathrm{Var}(\eta).\end{aligned}$$

Пример: дисперсия биномиального распределения

$\xi \sim B(n, p)$: $p_\xi(k) = C_n^k p^k (1-p)^{n-k}$. Найти $\mathrm{Var}(\xi)$.

Сумма независимых бернуллиевских случайных величин

$$\eta = \sum_{i=1}^{n} \xi_i, \quad \xi_i = I_{\{a_i = H\}}, \quad \xi_i \sim \mathrm{Ber}(p)$$

имеет биномиальное распределение: $\eta \sim B(n, p)$. Следовательно,

$$\mathrm{Var}(\xi) = \mathrm{Var}(\eta) = \sum_{i=1}^{n} \mathrm{Var}(\xi_i) = np(1-p).$$

Ковариация

$$\mathrm{Cov}(\xi,\eta) := \mathsf{E}[(\xi-\mathsf{E}\xi)(\eta-\mathsf{E}\eta)] = \int_{\mathbb{R}^2}(x-\mathsf{E}\xi)(y-\mathsf{E}\eta)\mathsf{P}_{\xi,\eta}(dxdy)$$
$$= \mathsf{E}(\xi\eta)-\mathsf{E}\xi\cdot\mathsf{E}\eta.$$

Ковариация положительна (отрицательна), если отклонения ξ и η от из средних значений "чаще" имеют один знак (разные знаки).
Если ξ, η независимы, то

$$\mathrm{Cov}(\xi,\eta) := \mathsf{E}[\xi-\mathsf{E}\xi]\cdot\mathsf{E}[\eta-\mathsf{E}\eta] = 0.$$

Обратное неверно.

Пример: [Bertsekas, Tsitsiklis, p.219]

Рассмотрим дискретную двумерную случайную величину с распределением

$$p_{(\xi,\eta)}(1,0) = p_{(\xi,\eta)}(0,1) = p_{(\xi,\eta)}(0,-1) = p_{(\xi,\eta)}(-1,0) = 1/4$$

				$p_\eta(j)$
1	0	1/4	0	1/4
0	1/4	0	1/4	1/2
-1	0	1/4	0	1/4
j/i	-1	0	1	
$p_\xi(i)$	1/4	1/2	1/4	

$$\mathsf{E}\xi = \mathsf{E}\eta = 0,$$

$$\mathrm{Cov}(\xi,\eta) = \mathsf{E}(\xi\eta) = \sum_{x,y}(xy)p_{\xi,\eta}(x,y) = 0.$$

Но ξ, η зависимы:

$$p_{\xi,\eta}(0,0) = 0 \neq p_\xi(0)p_\eta(0) = \frac{1}{4}.$$

Пример

Пусть ξ непрерывна и имеет симметричную положительную на $\mathbb{R}$ плотность: $f_\xi(x) = f_\xi(-x)$. Случайные величины ξ и $\eta = \xi^2$ зависимы:

$$\begin{aligned} \mathsf{P}(\xi \in (a,b), \xi^2 \in (a^2,b^2)) &= \mathsf{P}(\xi \in (a,b)) \\ &\neq \mathsf{P}(\xi \in (a,b))\mathsf{P}(\xi^2 \in (a^2,b^2)), \quad 0 < a < b. \end{aligned}$$

Но

$$\mathsf{E}\xi = \int_{\mathbb{R}} x f_\xi(x)\,dx = 0$$

$$\operatorname{Cov}(\xi,\eta) = \mathsf{E}(\xi\eta) = \mathsf{E}(\xi^3) = \int_{\mathbb{R}} x^3 f_\xi(x)\,dx = 0$$

Свойства ковариации

$$\mathrm{Cov}(\xi, \xi) = \mathrm{Var}(\xi),$$

$$\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(\eta, \xi),$$

$$\mathrm{Cov}(\alpha\xi + \beta\eta, \zeta) = \alpha\mathrm{Cov}(\xi, \zeta) + \beta\mathrm{Cov}(\eta, \zeta).$$

$$\mathrm{Var}(\xi + \eta) = \mathrm{Cov}(\xi + \eta, \xi + \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta) + 2\mathrm{Cov}(\xi, \eta)$$

$$\mathrm{Var}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathrm{Var}(\xi_i) + \sum_{i \neq j} \mathrm{Cov}(\xi_i, \xi_j)$$

Неравенство Коши-Буняковского

$$|\mathsf{E}(\xi\eta)| \leq \sqrt{\mathsf{E}(\xi^2)}\sqrt{\mathsf{E}(\eta^2)}.$$

Доказательство.

$$\mathsf{E}(\xi + t\eta)^2 = \mathsf{E}\xi^2 + 2t\mathsf{E}(\xi\eta) + t^2\mathsf{E}\eta^2 \geq 0, \quad t \in \mathbb{R}.$$

Следовательно, дискриминант неположителен:

$$(\mathsf{E}(\xi\eta))^2 \leq \mathsf{E}\xi^2\mathsf{E}\eta^2,$$

Коэффициент корреляции

$\sigma(\xi) = \sqrt{\mathrm{Var}(\xi)}$.
Коэффициент корреляции

$$\rho(\xi, \eta) := \frac{\mathrm{Cov}(\xi, \eta)}{\sigma(\xi)\sigma(\eta)}, \quad \sigma(\xi), \sigma(\eta) > 0.$$

является безразмерной версией коэффициента ковариации. По неравенству Коши-Буняковского,

$$|\mathrm{Cov}(\xi, \eta)| \leq \sqrt{\mathsf{E}(\xi - \mathsf{E}\xi)^2}\sqrt{\mathsf{E}(\eta - \mathsf{E}\eta)^2} = \sigma(\xi)\sigma(\eta).$$

Следовательно, $\rho(\xi, \eta) \in [-1, 1]$.

Если ξ, η независимы, то $\rho(\xi, \eta) = 0$, так как $\mathrm{Cov}(\xi, \eta) = 0$.

Если $\xi = \alpha\eta + \beta$, то

$$\rho(\xi, \eta) := \frac{\mathrm{Cov}(\alpha\eta, \eta)}{\sigma(\alpha\eta)\sigma(\eta)} = \frac{\alpha}{|\alpha|}\frac{\mathrm{Cov}(\eta, \eta)}{\sigma(\eta)\sigma(\eta)} = \mathrm{sgn}\,(\alpha) = \pm 1.$$

Обратно, если $\rho(\xi, \eta) = \pm 1$, то

$$\xi - \mathsf{E}\xi = c(\eta - \mathsf{E}\eta).$$

Действительно, пусть

$$\overline{\xi} = \frac{\xi - \mathsf{E}\xi}{\sigma(\xi)}, \quad \overline{\eta} = \frac{\eta - \mathsf{E}\eta}{\sigma(\eta)}.$$

Тогда

$$\mathsf{E}\overline{\xi} = \mathsf{E}\overline{\eta} = 0, \quad \mathrm{Var}(\overline{\xi}) = \mathrm{Var}(\overline{\eta}) = 1,$$

$$\rho := \rho(\xi, \eta) = \mathsf{E}(\overline{\xi}\overline{\eta}),$$

$$\mathsf{E}(\overline{\xi} - \rho\overline{\eta})^2 = \mathsf{E}\overline{\xi}^2 - 2\rho\mathsf{E}(\overline{\xi}\overline{\eta}) + \rho^2\mathsf{E}\overline{\eta}^2 = 1 - 2\rho^2 + \rho^2 = 1 - \rho^2.$$

Если $|\rho| = 1$, то

$$\overline{\xi} = \rho\overline{\eta} \quad \text{п.н.}$$

Пример (Tsitsiklis, 2018, Lecture 12.11)

Пусть ξ, η, ζ независимы,

$$\mathsf{E}\xi = \mathsf{E}\eta = \mathsf{E}\zeta = 0, \quad \mathrm{Var}(\zeta) = 1.$$

Рассмотрим

$$X = \xi + \zeta, \quad Y = \eta + \zeta.$$

ζ — общий скрытый фактор, ξ, η — шум.

$$\rho(X, Y) = \frac{\mathrm{Cov}(\xi + \zeta, \eta + \zeta)}{\sigma(\xi + \zeta)\sigma(\eta + \zeta)} = \frac{1}{\sqrt{\mathrm{Var}(\xi) + 1}\sqrt{\mathrm{Var}(\eta) + 1}}$$

Корреляция тем меньше, чем больше дисперсия шума. Если $\mathrm{Var}(\xi) = \mathrm{Var}(\eta) = 1$, то $\rho(X, Y) = 1/2$.

Портфель ценных бумаг. Диверсификация Марковитца

Инвестор может инвестировать начальный капитал X_0 в d рисковых активов. Их цены в момент времени $t = 0$ известны: $S_0^1, \dots, S_0^d$. Пусть $\gamma_1, \dots, \gamma_d$ количество единиц каждого актива, в его начальном портфеле:

$$X_0 = \sum_{i=1}^{d} \gamma_i S_0^i.$$

В момент времени $t = 1$ цены случайны и капитал инвестора будет равен

$$X_1 = \sum_{i=1}^{d} \gamma_i S_1^i = X_0 \sum_{i=1}^{d} \delta_i R^i, \quad \delta^i = \frac{\gamma_i S_0^i}{X_0}, \quad R^i = \frac{S_1^i}{S_0^i}.$$

Г. Марковитц (1952) предложил оценивать качество портфеля по ожидаемому возврату (математическому ожиданию) и риску (дисперсии):

$$\mathsf{E}\left(\frac{X_1}{X_0}\right) = \sum_{i=1}^{d} \mu^i \delta_i, \quad \mu^i = \mathsf{E} R^i,$$

$$\operatorname{Var}\left(\frac{X_1}{X_0}\right) = \sum_{i=1}^{d} \sigma_i^2 \delta_i^2 + \sum_{i \neq j} \sigma_{ij} \delta_i \delta_j,$$

$$\sigma_i^2 = \operatorname{Var}(R^i), \quad \sigma_{ij} = \operatorname{Cov}(R^i, R^j).$$

Оптимизационная задача Марковитца:

$$\sum_{i=1}^{d} \sigma_i^2 \delta_i^2 + \sum_{i \neq j} \sigma_{ij} \delta_i \delta_j \to \min_{\delta},$$

$$\sum_{i=1}^{d} \mu^i \delta_i = m.$$

Если R^i некоррелированы и $\mathrm{Var}(R^i) \le C$, то

$$\mathrm{Var}\left(\frac{X_1}{X_0}\right) = \sum_{i=1}^{d} \sigma_i^2 \delta_i^2 \le C \sum_{i=1}^{d} \delta_i^2.$$

Полагая $\delta_i = 1/d$, находим

$$\mathrm{Var}\left(\frac{X_1}{X_0}\right) \le \frac{C}{d} \to 0, \quad d \to \infty.$$

В общем случае для такого портфеля имеем

$$\begin{aligned}\mathrm{Var}\left(\frac{X_1}{X_0}\right) &= \frac{1}{d^2} \sum_{i=1}^{d} \sigma_i^2 + \frac{1}{d^2} \sum_{i \ne j} \sigma_{ij} \\ &= \frac{1}{d} \overline{\sigma}_d^2 + \frac{d-1}{d} \overline{\mathrm{Cov}}_d.\end{aligned}$$

Средняя дисперсия и средняя ковариация:

$$\overline{\sigma}_d^2 = \frac{1}{d}\sum_{i=1}^{d}\sigma_i^2, \quad \overline{\mathrm{Cov}}_d = \frac{1}{d(d-1)}\sum_{i\neq j}\sigma_{ij}$$

определяют несистематический и систематический (рыночный) риски:

$$\frac{1}{d}\overline{\sigma}_d^2, \quad \frac{d-1}{d}\overline{\mathrm{Cov}}_d.$$

Несистематический риск может быть редуцирован диверсификацией (распределением капитала между большим числом рисковых активов). Систематический риск присущ рынку и не может быть редуцирован с помощью диверсификации:

$$\frac{1}{d}\overline{\sigma}_d^2 \to 0, \quad \frac{d-1}{d}\overline{\mathrm{Cov}}_d \to \overline{\mathrm{Cov}}, \quad d\to\infty.$$

Сумма независимых случайных величин, свертка

$\zeta = \xi + \eta$, ξ, η независимы.
ξ, η — дискретны:

$$p_\zeta(z) = \mathsf{P}(\xi + \eta = z) = \mathsf{E} I_{\{\xi+\eta=z\}} = \sum_{x,y} I_{\{x+y=z\}} p_\xi(x) p_\xi(y)$$

$$= \sum_{x+y=z} p_\xi(x) p_\xi(y) = \sum_x p_\xi(x) p_\eta(z-x)$$

ξ, η — непрерывны:

$$\mathsf{E} g(\zeta) = \int_{\mathbb{R}} h(z) f_\zeta(z)\, dz = \mathsf{E} h(\xi+\eta) = \int_{\mathbb{R}^2} h(x+y) f_\xi(x) f_\eta(y)\, dx dy$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} h(z) f_\xi(x) f_\eta(z-x)\, dx dz = \int_{\mathbb{R}} h(z) \left(\int_{\mathbb{R}} f_\xi(x) f_\eta(z-x)\, dx \right) dz,$$

$$f_\zeta(z) = \int_{\mathbb{R}} f_\xi(x) f_\eta(z-x)\, dx.$$

Пример (Bertsekas, Tsitsiklis, p.223-224)

$\xi, \eta \in U(0,1)$ и независимы, $\zeta = \xi + \eta$,

$$f_\zeta(z) = \int_{\mathbb{R}} f_\xi(x) f_\eta(z-x)\,dx = \int_0^1 f_\eta(z-x)\,dx = \int_{z-1}^{z} f_\eta(y)\,dy$$

При $z \in [0,1]$:

$$f_\zeta(z) = \int_0^z f_\eta(y)\,dy = z$$

При $z \in [1,2]$:

$$f_\zeta(z) = \int_{z-1}^{1} f_\eta(y)\,dy + \int_1^z f_\eta(y)\,dy = 2 - z + 0 = 2 - z$$

Пример (Tsitsiklis, Lecture 12.4)

$\xi_i \sim N(\mu_i, \sigma_i^2)$ и независимы, $\zeta = \xi_1 + \xi_2$,

$$f_\zeta(z) = \int_{\mathbb{R}} f_{\xi_1}(x) f_{\xi_2}(z-x)\, dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z-x-\mu_2)^2}{2\sigma_1^2}\right) dx$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2))}\right)$$

Таким образом, $\zeta \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Плотность преобразования многомерной случайной величины

Пусть $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ — взаимно-однозначное дифференцируемое отображение. Пусть $\eta = (\eta_1, \ldots, \eta_d) = g(\xi_1, \ldots, \xi_d) = g(\xi)$. Тогда

$$\mathsf{E}h(\eta) = \int_{\mathbb{R}^d} h(y) f_\eta(y)\, dy = \mathsf{E}h(g(\xi)) = \int_{\mathbb{R}^d} h(g(x)) f_\xi(x)\, dx$$
$$= \int_{\mathbb{R}^d} h(y) f_\xi(g^{-1}(y)) |\det \left[(g^{-1})'\right] |\, dy,$$

где $(g^{-1})'$ — якобиан отображения g^{-1}. Таким образом,

$$f_\eta(y) = f_\xi(g^{-1}(y)) \left|\det \left[(g^{-1})'\right]\right|.$$

В частности, если $\eta = A\xi + b$, где A — обратимая $d \times d$ матрица и $b \in \mathbb{R}^d$, то

$$f_\eta(y) = f_\xi(A^{-1}(y-b)) \left|\det \left[A^{-1}\right]\right| = \frac{1}{|\det A|} f_\xi(A^{-1}(y-b)).$$

Многомерное нормальное распределение

Пусть $\xi_i \sim N(0,1)$, $i = 1, \dots, d$ независимы. *Многомерным стандартным нормальным* распределением называется совместное распределение ξ_i:

$$f_{\xi_1,\dots,\xi_d}(x_1, \dots, x_d) = \prod_{i=1}^{d} f_{\xi_i}(x_i) = \frac{1}{(\sqrt{2\pi})^d} \prod_{i=1}^{d} e^{-x_i^2/2}$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\langle x,x\rangle/2}$$

Многомерным нормальным (гауссовским) распределением называется распределение вектора $\eta = A\xi + \mu$ (матрицу A пока будем считать обратимой). По формуле для преобразования плотности,

$$f_\eta(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{|\det A|} e^{-\langle A^{-1}(y-\mu), A^{-1}(y-\mu)\rangle/2}$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{|\det A|} e^{-\langle (A^{-1})^T A^{-1}(y-\mu), y-\mu\rangle/2}$$

$$(A^{-1})^T A^{-1} = (A^T)^{-1} A^{-1} = (AA^T)^{-1}.$$

Матрица AA^T симметрична и положительно определена:

$$\langle AA^T x, x \rangle = \|A^T x\|^2 > 0, \quad x \neq 0.$$

Положим $\Sigma = AA^T$. Поскольку

$$\det \Sigma = \det(AA^T) = (\det A)^2,$$

то

$$f_\eta(y) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\sqrt{\det \Sigma}} e^{-\langle \Sigma^{-1}(y-\mu), y-\mu \rangle / 2}$$

Для распределения гауссовского вектора η используется обозначение $\eta \sim N(\mu, \Sigma)$. Рассмотрим вектор математических ожиданий μ:

$$\mathsf{E}\eta = \mathsf{E}(\mu + A\xi) = \mu$$

и *ковариационную матрицу*

$$\begin{aligned}\mathrm{Cov}(\eta_i, \eta_j) &= \mathsf{E}(\langle a_i, \xi\rangle \cdot \langle a_j, \xi\rangle) = \mathsf{E}\sum_{kr} a_{ik}a_{jr}\xi_k\xi_r \\ &= \sum_k a_{ik}a_{jk} = (AA^T)_{ij} = \Sigma_{ij}.\end{aligned}$$

Данные формулы указывают на смысл параметров μ, Σ.

Если η_i некоррелированы, то матрица Σ диагональна и

$$f_\eta(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{\prod_{i=1}^d \Sigma_{ii}}} \exp\left(-\sum_{i=1}^d \frac{(y_i-\mu_i)^2}{2\Sigma_{ii}}\right)$$
$$= \prod_{i=1}^d \sqrt{\frac{1}{2\pi\Sigma_{ii}}} \exp\left(-\frac{(y_i-\mu_i)^2}{2\Sigma_{ii}}\right)$$

Таким образом, $f_\eta(y)$ распадается в произведение маргинальных плотностей одномерных нормальных случайных величин $\eta_i \sim N(\mu_i, \Sigma_{ii})$. Значит компоненты η независимы. Таким образом, для гауссовского вектора некоррелированность компонент равносильна их независимости.

Характеристические функции

Характеристической функцией случайной величины $\xi : \Omega \mapsto \mathbb{R}^d$ называется преобразование Фурье ее распределения:

$$\varphi_\xi(u) = \mathsf{E}e^{i\langle u,\xi\rangle} = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle}\,\mathsf{P}_\xi(dx).$$

Здесь $\langle u, x\rangle$ — обычное скалярное произведение в $\mathbb{R}^d$.

- Распределение P_ξ однозначно восстанавливается по своей характеристической функции (теорема единственности).

Поскольку характеристическая функция однозначно определяется распределением, можно называть ее также характеристической функцией распределения.

Характеристическую функцию можно использовать для вычисления математического ожидания и дисперсии, так как (в одномерном случае)

$$\varphi'_\xi(u) = i\mathsf{E}\left(\xi e^{iu\xi}\right), \quad \varphi''_\xi(0) = -\mathsf{E}\left(\xi^2 e^{iu\xi}\right),$$

$$\varphi'_\xi(0) = i\mathsf{E}\xi, \quad \varphi''_\xi(0) = -\mathsf{E}\xi^2.$$

Характеристическая функция распределения Пуассона

Найдем математическое ожидание и дисперсию распределения Пуассона, используя аппарат характеристических функций. Пусть $\xi \sim \Pi(\lambda)$. Тогда

$$\varphi_\xi(u) = \mathsf{E}e^{iu\xi} = \sum_{k=0}^{\infty} e^{iuk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{iu})^k}{k!} = \exp(-\lambda + \lambda e^{iu}),$$
$$\varphi'_\xi(u) = \lambda i e^{iu}\exp(\lambda(e^{iu}-1)),$$
$$\varphi''_\xi(u) = (-\lambda e^{iu} + (\lambda i e^{iu})^2)\exp(\lambda(e^{iu}-1)).$$

Следовательно,

$$\mathsf{E}\xi = -i\varphi'_\xi(0) = \lambda, \quad \mathsf{E}\xi^2 = -\varphi''_\xi(0) = \lambda + \lambda^2,$$
$$\mathrm{Var}(\xi) = \mathsf{E}(\xi^2) - (\mathsf{E}\xi)^2 = \lambda.$$

Характеристическая функция нормального распределения

Найдем характеристическую функцию нормального распределения. Пусть $\eta \sim N(0,1)$. Тогда

$$\varphi_\eta(u) = \mathsf{E} e^{iu\eta} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} e^{-x^2/2}\, dx,$$

$$\varphi'_\eta(u) = i\mathsf{E}(\eta e^{iu\eta}) = \frac{i}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{iux} e^{-x^2/2}\, dx = -\frac{i}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} d e^{-x^2/2}$$
$$= -\frac{u}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} e^{-x^2/2}\, dx = -u\varphi_\eta(u).$$

Решая дифференциальное уравнение $\varphi'_\eta(u) = -u\varphi_\eta(u)$ с начальным условием $\varphi_\eta(0) = 1$, находим

$$\varphi_\eta(u) = e^{-u^2/2}.$$

Чтобы найти характеристическую функцию произвольного нормального распределения, рассмотрим случайную величину $\xi = \mu + \sigma\eta \sim N(\mu, \sigma^2)$:

$$\varphi_\xi(u) = e^{iu\mu}\mathsf{E}e^{iu\sigma\eta} = e^{i\mu u - \sigma^2 u^2/2}.$$

$$\varphi'_\xi(u) = (i\mu - \sigma^2 u)e^{i\mu u - \sigma^2 u^2/2}, \quad \varphi''_\xi(u) = [(i\mu - \sigma^2 u)^2 - \sigma^2]e^{i\mu u - \sigma^2 u^2/2}.$$

Следовательно,

$$\mathsf{E}\xi = -i\varphi'_\xi(0) = \mu, \quad \mathsf{E}\xi^2 = -\varphi''_\xi(0) = \mu^2 + \sigma^2,$$
$$\mathrm{Var}(\xi) = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \sigma^2.$$

Характеристическая функция суммы независимых случайных величин

Характеристическая функция совместного распределения вектора, составленного из независимых случайных величин, равна произведению характеристических функций компонент:

$$\varphi_\xi(u) = \mathsf{E} e^{i\langle u, \xi\rangle} = \mathsf{E} \prod_{k=1}^{n} e^{iu_k\xi_k} = \prod_{k=1}^{n} \mathsf{E} e^{iu_k\xi_k} = \prod_{k=1}^{n} \varphi_{\xi_k}(u_k).$$

Из упомянутой теоремы единственности вытекает можно вывести, что верно и обратное утверждение, т.е. данное равенство является критерием независимости $(\xi_1, \ldots, \xi_n)$.

Характеристическая функция многомерного нормального распределения

$\eta = \mu + A\xi$:

$$\varphi_\eta(u) = e^{i\langle u,\mu\rangle}\mathsf{E}e^{i\langle u,A\xi\rangle} = e^{i\langle u,\mu\rangle}\mathsf{E}e^{i\langle A^T u,\xi\rangle} = e^{i\langle u,\mu\rangle}\prod_{k=1}^{d}\mathsf{E}e^{i(A^T u)_k\xi_k}$$

$$= e^{i\langle u,\mu\rangle}\prod_{i=1}^{d}e^{-(A^T u)_i^2/2} = e^{i\langle u,\mu\rangle}e^{-\|A^T u\|^2/2} = e^{i\langle u,\mu\rangle-\langle\Sigma u,u\rangle/2},$$

где $\Sigma = AA^T$. Таким образом, вектор η является *гауссовским*, если и только если

$$\varphi_\eta(u) = e^{i\langle u,\mu\rangle-\langle\Sigma u,u\rangle/2},$$

где $\mu \in \mathbb{R}^d$ и Σ — симметричная неотрицательно определенная матрица. Можно принять данное свойство за определение многомерной гауссовской случайной величины. При этом достаточно считать, что Σ неотрицательно определена: ее обратимости не требуется.

Распределение суммы одномерных независимых нормальных с.в.

$\xi_i \sim N(\mu_i, \sigma_i^2)$ и независимы, $\zeta = \xi_1 + \xi_2$.

$$\begin{aligned}\varphi_\zeta(u) = \varphi_{\xi_1}(u)\varphi_{\xi_2}(u) &= e^{iu\mu_1 - \sigma_1^2 u^2/2} e^{iu\mu_2 - \sigma_2^2 u^2/2} \\ &= e^{iu(\mu_1+\mu_2) - (\sigma_1^2+\sigma_2^2)u^2/2}.\end{aligned}$$

Из теоремы единственности вытекает, что $\zeta \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. С помощью вычисления свертки получить этот результат значительно сложнее. Выше он был сформулирован, но вычисления не проводились.

Сходимость случайных величин

Последовательность случайных величин ξ_n называется сходящейся к ξ

- *с вероятностью 1* (или *п.н.*) : $\xi_n \xrightarrow{\text{п.н.}} \xi$, если
$$\mathsf{P}\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\} = 1;$$
- *по вероятности*: $\xi_n \xrightarrow{\mathsf{P}} \xi$, если
$$\lim_{n\to\infty} \mathsf{P}\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\} = 0;$$
- *в среднем порядка* r (или в L^r): $\xi_n \xrightarrow{L^r} \xi$, $(0 < r < \infty)$, если
$$\lim_{n\to\infty} \mathsf{E}(|\xi_n - \xi|^r) = 0;$$
- *по распределению*: $\xi_n \xrightarrow{d} \xi$, если $\lim_{n\to\infty} \mathsf{E}g(\xi_n) = \mathsf{E}g(\xi)$ для любой ограниченной непрерывной функции $g : \mathbb{R}^d \mapsto \mathbb{R}$.

Соотношение между этими видами сходимости описывается следующим образом:

- сходимость с вероятностью $1 \Longrightarrow$ сходимость по вероятности $\Longrightarrow$ сходимость по распределению;
- сходимость в среднем порядка $r \Longrightarrow$ сходимость по вероятности;
- если $\xi_n \to \xi$ по вероятности, то существует подпоследовательность ξ_{n_k}, сходящаяся к ξ с вероятностью 1.

Приведем примеры, показывающие, что “стрелки нельзя обратить”.

Пример

Пример последовательности ξ_n, сходящейся к 0 по вероятности (и даже в L^r), но не с вероятностью 1.
На пространстве $((0,1), \mathscr{B}((0,1)), \lambda)$, где λ — мера Лебега, рассмотрим последовательность

$$\xi_1 = I_{(0,1/2)}, \quad \xi_2 = I_{[1/2,1)},$$

$$\xi_3 = I_{(0,1/3)}, \quad \xi_4 = I_{[1/3,2/3)}, \quad \xi_5 = I_{[2/3,1)}, \quad \ldots$$

Легко видеть, что $\mathsf{E}|\xi_n|^r \to 0$ (и значит $\xi_n \xrightarrow{\mathsf{P}} 0$), но

$$\limsup_{n\to\infty} \xi_n(\omega) = 1, \quad \liminf_{n\to\infty} \xi_n(\omega) = 0 \quad \text{для всех } \omega.$$

Пример

Пример последовательности, сходящейся по распределению, но не по вероятности.
Пусть $\mathsf{P}(\xi = 1) = \mathsf{P}(\xi = 0) = 1/2$. Рассмотрим последовательность $\xi_n = \xi$, $n \geq 1$ и с.в. $\eta = 1 - \xi$. Ясно, что $\xi_n \xrightarrow{d} \eta$, т.к. $\xi \sim \eta$. Но $|\xi_n - \eta| = 1$. Значит сходимости по вероятности нет.

Пример

Пример последовательности, сходящейся по вероятности, но не в L^1.
Рассмотрим вероятностное пространство $([0,1], \mathscr{B}([0,1]), \lambda)$, где λ — мера Лебега. Последовательность $\xi_n = nI_{[0,1/n]}$ сходится к 0 по вероятности:

$$\mathsf{P}(|\xi_n| \geq \varepsilon) \leq 1/n \to 0, \quad n \to \infty,$$

но не в L^1: $\mathsf{E}|\xi_n| = 1$.

Отметим, что первый пример показывает, что из сходимости в L^r, $r > 0$ не вытекает сходимость п.н., а последний — что из сходимости п.н. не вытекает сходимость в L^r.

Теоремы о предельном переходе под знаком математического ожидания

Известно три основных результата о предельном переходе под знаком интеграла (математического ожидания):

- *Теорема о монотонной сходимости*: пусть $0 \le \xi_n \nearrow \xi$ п.н., тогда $\mathsf{E}\xi = \lim_{n\to\infty} \mathsf{E}\xi_n$.
- *Теорема о мажорируемой сходимости*: пусть $|\xi_n| \le \eta$, $\mathsf{E}\eta < \infty$ и $\xi_n \xrightarrow{\mathsf{P}} \xi$, тогда $\mathsf{E}\xi = \lim_{n\to\infty} \mathsf{E}\xi_n$.
- *Лемма Фату*: пусть $\xi_n \ge 0$, тогда

$$\mathsf{E} \liminf_{n\to\infty} \xi_n \le \liminf_{n\to\infty} \mathsf{E}\xi_n.$$

Стандартная техника

Если требуется доказать, что какое-то свойство выполняется для всех интегрируемых $\mathscr{F}$-измеримых случайных величин, то можно использовать *стандартную технику* (standard machine):

- доказать, что данное свойство верно для индикаторов I_A, $A \in \mathscr{F}$,
- по линейности распространить его на неотрицательные простые функции,
- по теореме о *монотонной сходимости* доказать, что оно верно для неотрицательных $\mathscr{F}$-измеримых функций,
- используя равенство $\xi = \xi^+ - \xi^-$, распространить его на все $\mathscr{F}$-измеримые функции.

Неравенства Маркова и Чебышёва

- *Неравенство Маркова*. Пусть $\xi \geq 0$, тогда

$$\mathsf{P}(\xi \geq \varepsilon) \leq \frac{\mathsf{E}\xi}{\varepsilon}.$$

Действительно,

$$\mathsf{E}\xi \geq \mathsf{E}(\xi I_{\{\xi \geq \varepsilon\}}) \geq \varepsilon \mathsf{E}(I_{\{\xi \geq \varepsilon\}}) = \varepsilon \mathsf{P}(\xi \geq \varepsilon).$$

- *Неравенство Чебышёва*

$$\mathsf{P}(|\xi - \mu| \geq \varepsilon) = \mathsf{P}(|\xi - \mu|^2 \geq \varepsilon^2) \leq \frac{\mathrm{Var}(\xi)}{\varepsilon^2}, \quad \mu = \mathsf{E}\xi$$

вытекает из неравенства Маркова.

Используя теорему о мажорируемой сходимости, можно доказать, что

- сходимость с вероятностью 1 влечет сходимость по вероятности: если $\xi_n \xrightarrow{\text{п.н.}} \xi$, то

$$\mathsf{P}(|\xi_n - \xi| \geq \varepsilon) = \mathsf{E} I_{\{|\xi_n - \xi| \geq \varepsilon\}} \to 0, \quad n \to \infty$$

- сходимость по вероятности влечет сходимость по распределению: если $\xi_n \xrightarrow{\mathsf{P}} \xi$, то, как можно доказать, $g(\xi_n) \xrightarrow{\mathsf{P}} g(\xi)$ для непрерывной функции g (теорема о непрерывном преобразовании). Следовательно,

$$\mathsf{E} g(\xi_n) \to \mathsf{E} g(\xi), \quad n \to \infty$$

  для ограниченной непрерывной функции g.
- сходимость в L^r влечет сходимость по вероятности:

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

  в силу неравенства Маркова.

Первая лемма Бореля-Кантелли

Для последовательности множеств A_n, $n \in \mathbb{N}$ введем обозначение

$$\{A_n \text{ б.ч.}\} := \{A_n \text{ бесконечно часто}\} = \bigcap_m \bigcup_{n \geq m} A_n = \left\{ \sum_{n=1}^{\infty} I_{A_n} = \infty \right\}$$

- Если $\sum_{n=1}^{\infty} \mathsf{P}(A_n) < \infty$, то

$$\mathsf{P}(A_n \text{ б.ч.}) = 0.$$

Доказательство. По теореме о монотонной сходимости,

$$\mathsf{E} \sum_{n=1}^{\infty} I_{A_n} = \sum_{n=1}^{\infty} \mathsf{E} I_{A_n} = \sum_{n=1}^{\infty} \mathsf{P}(A_n) < \infty.$$

Следовательно, $\sum_{n=1}^{\infty} I_{A_n}$ не может быть бесконечной на множестве положительной вероятности. □

Существование подпоследовательности, сходящейся п.н.

- Пусть $\xi_n \xrightarrow{\mathsf{P}} \xi$. Тогда существует подпоследовательность $\xi_{n_k} \overset{\text{п.н.}}{\to} \xi$.

Доказательство. Пусть $0 < \varepsilon_k \to 0$. Поскольку

$$\mathsf{P}(|\xi_n - \xi| \geq \varepsilon_k) \to 0, \quad n \to \infty,$$

то для любого k существует n_k:

$$\mathsf{P}(|\xi_{n_k} - \xi| \geq \varepsilon_k) \leq 1/2^k,$$

и последовательность n_k может быть выбрана возрастающей. Отсюда следует, что

$$\sum_{k=1}^{\infty} \mathsf{P}(|\xi_{n_k} - \xi| \geq \varepsilon_k) < \infty.$$

По лемме Бореля-Кантелли, $\mathsf{P}(\{|\xi_{n_k} - \xi| \geq \varepsilon_k\} \text{ б.ч.}) = 0$. Это означает, существует множество Ω' с $\mathsf{P}(\Omega') = 1$ такое, что

$$|\xi_{n_k} - \xi|(\omega) < \varepsilon_k, \quad \omega \in \Omega'$$

при достаточно больших k. Следовательно, $\xi_{n_k} \to \xi$ поточечно на Ω'.

Вторая лемма Бореля-Кантелли

- Если события $(A_n, n \in \mathbb{N})$ независимы и $\sum_{n=1}^{\infty} \mathsf{P}(A_n) = \infty$, то

$$\mathsf{P}(A_n \text{ б.ч.}) = \mathsf{P}\left(\bigcap_m \bigcup_{n \geq m} A_n\right) = 1.$$

Доказательство.

$$\mathsf{P}((A_n \text{ б.ч.})^c) = \mathsf{P}\left(\bigcup_m \bigcap_{n \geq m} A_n^c\right)$$

$$\mathsf{P}\left(\bigcap_{n \geq m} A_n^c\right) = \lim_{r \to \infty} \mathsf{P}\left(\bigcap_{n=m}^{r} A_n^c\right) = \lim_{r \to \infty} \prod_{n=m}^{r} \mathsf{P}(A_n^c) = \prod_{n=m}^{\infty} (1 - \mathsf{P}(A_n))$$

Корректность первого предельного перехода вытекает из непрерывности вероятностной меры. Из элементарного неравенства $1 - x \leq e^{-x}$ следует, что

$$\prod_{n=m}^{\infty}(1-\mathsf{P}(A_n)) \le \prod_{n=m}^{\infty} e^{-\mathsf{P}(A_n)} = \exp\left(-\sum_{n=m}^{\infty}\mathsf{P}(A_n)\right) = 0. \quad \Box$$

Независимость важна: если $A_n = A$, $\mathsf{P}(A) < 1$, то $\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \infty$, но $\mathsf{P}(A_n \text{ б.ч}) = \mathsf{P}(A) < 1$.

Пример

Какова вероятность получить k орлов подряд при бесконечном подбрасывании симметричной монеты? Здесь k — любое число, например, $k = 100^{100}$.

Пусть один эксперимент состоит в k подбрасываниях. Вероятность “успеха” в n-м эксперименте: $\mathsf{P}(A_n) = 1/2^k$. Вероятность того, что будет бесконечное число успехов: $\mathsf{P}(A_n \text{ б.ч}) = 1$, так как

$$\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \sum_{n=1}^{\infty}\frac{1}{2^k} = \infty.$$

Аналогично, вероятность того, что (бессмертная) обезьяна, случайно стуча по клавиатуре, в конце концов напечатает роман “Война и мир”, равна 1. Более того, она это сделает бесконечное число раз.

$$\prod_{n=m}^{\infty}(1-\mathsf{P}(A_n)) \le \prod_{n=m}^{\infty} e^{-\mathsf{P}(A_n)} = \exp\left(-\sum_{n=m}^{\infty}\mathsf{P}(A_n)\right) = 0. \quad \square$$

Независимость важна: если $A_n = A$, $\mathsf{P}(A) < 1$, то $\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \infty$, но $\mathsf{P}(A_n \text{ б.ч}) = \mathsf{P}(A) < 1$.

Пример

Какова вероятность получить k орлов подряд при бесконечном подбрасывании симметричной монеты? Здесь k — любое число, например, $k = 100^{100}$.

Пусть один эксперимент состоит в k подбрасываниях. Вероятность "успеха" в n-м эксперименте: $\mathsf{P}(A_n) = 1/2^k$. Вероятность того, что будет бесконечное число успехов: $\mathsf{P}(A_n \text{ б.ч}) = 1$, так как

$$\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \sum_{n=1}^{\infty}\frac{1}{2^k} = \infty.$$

Аналогично, вероятность того, что (бессмертная) обезьяна, случайно стуча по клавиатуре, в конце концов напечатает роман "Война и мир", равна 1. Более того, она это сделает бесконечное число раз.

$$\prod_{n=m}^{\infty}(1-\mathsf{P}(A_n)) \le \prod_{n=m}^{\infty} e^{-\mathsf{P}(A_n)} = \exp\left(-\sum_{n=m}^{\infty}\mathsf{P}(A_n)\right) = 0. \quad \square$$

Независимость важна: если $A_n = A$, $\mathsf{P}(A) < 1$, то $\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \infty$, но $\mathsf{P}(A_n \text{ б.ч}) = \mathsf{P}(A) < 1$.

Пример

Какова вероятность получить k орлов подряд при бесконечном подбрасывании симметричной монеты? Здесь k — любое число, например, $k = 100^{100}$.

Пусть один эксперимент состоит в k подбрасываниях. Вероятность "успеха" в n-м эксперименте: $\mathsf{P}(A_n) = 1/2^k$. Вероятность того, что будет бесконечное число успехов: $\mathsf{P}(A_n \text{ б.ч}) = 1$, так как

$$\sum_{n=1}^{\infty}\mathsf{P}(A_n) = \sum_{n=1}^{\infty}\frac{1}{2^k} = \infty.$$

Аналогично, вероятность того, что (бессмертная) обезьяна, случайно стуча по клавиатуре, в конце концов напечатает роман "Война и мир", равна 1. Более того, она это сделает бесконечное число раз.

Пример

Пусть $(\xi_n)_{n\in\mathbb{N}}$ независимы, и $\xi_n \sim \mathrm{Ber}(1/n)$:

$$\mathsf{P}(\xi_n = 1) = \frac{1}{n}, \quad \mathsf{P}(\xi_n = 0) = 1 - \frac{1}{n},$$

тогда

$$\mathsf{P}(|\xi_n| \geq \varepsilon) = \mathsf{P}(\xi_n = 1) = \frac{1}{n}$$

Следовательно, $\xi_n \xrightarrow{\mathsf{P}} 0$. В то же время, $\limsup_{n\to\infty} \xi_n = 1$ п.н., так как $\mathsf{P}(\xi_n = 1 \text{ б.ч.}) = 1$:

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

Слабый закон больших чисел

$(\xi_i)_{i=1}^{\infty}$ — независимые одинаково распределенные случайные величины, $\mathsf{E}\xi_i = \mu$, $\mathrm{Var}(\xi_i) = \sigma^2$.

- *Слабый закон больших чисел:*

$$M_n = \frac{1}{n}\sum_{i=1}^{n} \xi_i \xrightarrow{\mathsf{P}} \mu, \quad n \to \infty.$$

Доказательство.

$$\mathsf{E}M_n = \mu, \quad \mathrm{Var}(M_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(\xi_i) = \frac{\sigma^2}{n}.$$

$$\mathsf{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad n \to \infty. \quad \square$$

В частности, частота появления события в серии независимых опытов приближается к его вероятности: A_i независимы, $\mathsf{P}(A_i) = p$,

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} I_{A_i} \xrightarrow{\mathsf{P}} p$$

Теорема непрерывности

Сходимость $\xi_n \xrightarrow{d} \xi$ влечет сходимость характеристических функций: $\varphi_{\xi_n}(u) \to \varphi(u)$, $n \to \infty$ при всех u. Обратное утверждение играет ключевую роль при доказательстве предельных теорем.

- *Теорема непрерывности*: если $\varphi_{\xi_n}(u) \to \varphi(u)$, $n \to \infty$ при всех u, то $\xi_n \xrightarrow{d} \xi$.

Центральная предельная теорема

Пусть $(\xi_k)_{k=0}^{\infty}$ — независимые одинаково распределенные случайные величины, $\mathsf{E}\xi_n = 0$, $\mathsf{D}\xi_n = 1$.

- *Центральная предельная теорема*:
$$\frac{S_n}{\sqrt{n}} := \frac{1}{\sqrt{n}}(\xi_1 + \cdots + \xi_n) \xrightarrow{d} \eta \sim N(0,1).$$

Доказательство. Пусть $\xi_j \sim \xi$. Тогда

$$\begin{aligned}
\varphi_{S_n/\sqrt{n}}(u) &= \mathsf{E}\prod_{j=1}^{n} e^{iu\xi_j/\sqrt{n}} = \prod_{j=1}^{n} \mathsf{E}e^{iu\xi_j/\sqrt{n}} = (\mathsf{E}e^{iu\xi/\sqrt{n}})^n \\
&= \left(\mathsf{E}\left(1 + \frac{iu\xi}{\sqrt{n}} - \frac{1}{2}\frac{u^2\xi^2}{n} + o\left(\frac{1}{n}\right)\right)\right)^n = \left(1 - \frac{1}{2}\frac{u^2}{n} + o\left(\frac{1}{n}\right)\right)^n \\
&= \exp\left(n\ln\left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)\right) \to e^{-u^2/2}, \quad n \to \infty \quad \square
\end{aligned}$$

- $\xi_n \xrightarrow{d} \xi$, если и только если
$$F_{\xi_n}(x) \to F_{\xi}(x)$$
во всех точках непрерывности F_{ξ}.

В частности, в центральной предельной теореме
$$\mathsf{P}\left(\frac{\eta_1 + \cdots + \eta_n}{\sqrt{n}} \le u\right) \to \Phi(u) := \mathsf{P}(\zeta \le u),$$
$$\zeta \in N(0,1), \quad \mathsf{E}\eta_i = 0, \quad \mathrm{Var}(\eta_i) = 1.$$

Galton board

Схема серий и теорема Пуассона

Рассмотрим *схему серий*: $S_n = \xi_{1,n} + \dots + \xi_{n,n}$, $n \geq 1$, где $\xi_{1,n}, \dots, \xi_{n,n}$ — независимые одинаково распределенные бернуллиевские случайные величины:

$$\mathsf{P}(\xi_{k,n} = 1) = p_n, \quad \mathsf{P}(\xi_{k,n} = 0) = q_n, \quad 1 \leq k \leq n, \quad p_n + q_n = 1.$$

Пусть $p_n \to 0$, $np_n \to \lambda > 0$, $n \to \infty$.

- *Теорема Пуассона*:
$$S_n \xrightarrow{d} \eta \sim \Pi(\lambda).$$

Доказательство. Найдем предел характеристической функции S_n:

$$\varphi_{S_n}(u) = \prod_{k=1}^{n} \mathsf{E} e^{iu\xi_{k,n}} = (p_n e^{iu} + q_n)^n = (1 + p_n(e^{iu} - 1))^n$$
$$= \exp\left(n \ln\left(1 + \frac{\lambda}{n}(e^{iu} - 1) + o(1)\right)\right) \to e^{\lambda(e^{iu}-1)}, \quad n \to \infty.$$

Таким образом, $\varphi_{S_n}(u) \to \varphi_\eta(u)$. $\square$

Применение центральной предельной теоремы

$$S_n = \xi_1 + \dots + \xi_n, \quad \mu = \mathsf{E}\xi_i, \quad \sigma = \sqrt{\mathrm{Var}(\xi_i)}$$

$$Z_n = \frac{S_n - \mathsf{E}S_n}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - \mu n}{\sqrt{n}\sigma}, \quad \mathsf{E}Z_n = 0, \quad \mathrm{Var}(Z_n) = 1$$

$$\lim_{n\to\infty} \mathsf{P}(Z_n \le z) = \mathsf{P}(Z \le z), \quad Z \sim N(0,1).$$

На практике Z_n считается стандартной нормальной: $Z_n \approx N(0,1)$. Следовательно,

$$S_n = \mu n + \sqrt{n}\sigma Z_n \approx N(\mu n, \sigma^2 n).$$

Часто в приближенном равенстве

$$\mathsf{P}(S_n \ge a) \approx b$$

известны два из трех параметров a, b, n и требуется найти третий.

Пример (Tsitsiklis, 2018, Lecture 19.5)

В контейнер загружаются пакеты, которые имеют случайный вес $\xi_i \sim \mathrm{Exp}(\lambda)$, $\lambda = 1/2$. В данном случае $\mu = \sigma = 1/\lambda = 2$. Пусть имеется $n = 100$ пакетов. Найти вероятность того, что общий вес пакетов не превзойдет емкости контейнера $a = 210$.

$$\begin{aligned}
\mathsf{P}(S_n \geq 210) &= \mathsf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \geq \frac{210 - n\mu}{\sigma\sqrt{n}}\right) \\
&= \mathsf{P}\left(\frac{S_n - 200}{20} \geq \frac{210 - 200}{20}\right) = \mathsf{P}(Z_n \geq 1/2) \\
&\approx 1 - \Phi(0.5) \approx 1 - 0.6915 = 0.3085.
\end{aligned}$$

В этом примере a, n были фиксированы и нужно было найти b.

Пример (Tsitsiklis, 2018, Lecture 19.5)

При тех же условиях найти емкость контейнера, при которой общий вес пакетов не превзойдет ее с вероятностью $b = 0.05$. Найдем a из неравенства

$$\mathsf{P}(S_n \geq a) \leq 0.05$$

$$\mathsf{P}(S_n \geq a) = \mathsf{P}\left(\frac{S_n - 200}{20} \geq \frac{a - 200}{20}\right) = \mathsf{P}\left(Z_n \geq \frac{a - 200}{20}\right)$$

$$\approx 1 - \Phi\left(\frac{a - 200}{20}\right) = 0.05$$

В таблице нормального распределения ищем число 0.95 и получаем решение уравнения

$$\Phi(z) = 0.95.$$

Это решение $z^* \approx 1.645$. Следовательно,

$$a = 200 + 20z^* \approx 232.9.$$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983

Пример (Tsitsiklis, 2018, Lecture 19.5)

Пусть $a = 210$. Сколько пакетов можно загружать, чтобы емкость контейнера a не была превышена с вероятностью 0.95? Найдем n из неравенства $\mathsf{P}(S_n \geq 210) \leq 0.05$:

$$\begin{aligned}
\mathsf{P}(S_n \geq a) &= \mathsf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \geq \frac{210 - n\mu}{\sigma\sqrt{n}}\right) \\
&= \mathsf{P}\left(\frac{S_n - 2n}{2\sqrt{n}} \geq \frac{210 - 2n}{2\sqrt{n}}\right) \approx 1 - \Phi\left(\frac{210 - 2n}{2\sqrt{n}}\right) = 0.05.
\end{aligned}$$

Снова рассматриваем решение уравнения

$$\Phi(z) = 0.95, \quad z^* \approx 1.645$$

и ищем n, для которого

$$\frac{210 - 2n}{2\sqrt{n}} = z^*.$$

Решая квадратное уравнение, находим, что $n \in (89, 90)$. Выбираем $n = 89$.

Нормальная нормальная аппроксимация биномиального распределения (Tsitsiklis, 2018, Lecture 19.6)

Пусть $\xi_i \sim \mathrm{Ber}(p)$ независимы. Тогда

$$S_n = \sum_{i=1}^{n} \xi_i \sim B(n,p), \quad \mathsf{E} S_n = np, \quad \mathrm{Var}(S_n) = np(1-p).$$

Пусть $n = 36$, $p = 1/2$. Найдем

$$\mathsf{P}(S_n \le 21) = \sum_{k=0}^{21} \mathsf{P}(S_n = k) = \sum_{k=0}^{21} C_n^k \frac{1}{2^n} \approx 0.8785.$$

Воспользуемся нормальной аппроксимацией:

$$Z_n = \frac{S_n - \mathsf{E} S_n}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{S_n - 18}{3} \approx N(0,1).$$

$$\mathsf{P}(S_n \le 21) = \mathsf{P}\left(\frac{S_n - 18}{3} \le \frac{21-18}{3}\right) = \mathsf{P}(Z_n \le 1) \approx \Phi(1) \approx 0.8413$$

Вместо этого можно было рассмотреть

$$\mathsf{P}(S_n < 22) = \mathsf{P}\left(\frac{S_n - 18}{3} < \frac{22-18}{3}\right) = \mathsf{P}(Z_n < 4/3) \approx \Phi(1.33) \approx 0.9082$$

Точное значение: $0.8785 \in (0.8413, 0.9082)$. Естественно попробовать еще один вариант:

$$\mathsf{P}(S_n \le 21.5) = \mathsf{P}(Z_n \le 3.5/3) \approx \Phi(1.17) \approx 0.8790$$

Аппроксимируем также индивидуальные вероятности (вероятностную массу)

$$\mathsf{P}(S_n = 19) = \mathsf{P}(18.5 \le S_n \le 19.5) \approx \mathsf{P}\left(\frac{18.5-18}{3} \le Z_n \le \frac{19.5-18}{3}\right)$$

$$= \mathsf{P}\left(0.17 \le Z_n \le 0.5\right) = \Phi(0.5) - \Phi(0.17) = 0.6915 - 0.5675 = 0.124$$

Точное значение:

$$P(S_n = 19) = C_{36}^{19}\frac{1}{2^{36}} = 0.1251.$$

$$P(S_n \le 21) = P\left(\frac{S_n - 18}{3} \le \frac{21-18}{3}\right) = P(Z_n \le 1) \approx \Phi(1) \approx 0.8413$$

Вместо этого можно было рассмотреть

$$P(S_n < 22) = P\left(\frac{S_n - 18}{3} < \frac{22-18}{3}\right) = P(Z_n < 4/3) \approx \Phi(1.33) \approx 0.9082$$

Точное значение: $0.8785 \in (0.8413, 0.9082)$. Естественно попробовать еще один вариант:

$$P(S_n \le 21.5) = P(Z_n \le 3.5/3) \approx \Phi(1.17) \approx 0.8790$$

Аппроксимируем также индивидуальные вероятности (вероятностную массу)

$$P(S_n = 19) = P(18.5 \le S_n \le 19.5) \approx P\left(\frac{18.5-18}{3} \le Z_n \le \frac{19.5-18}{3}\right)$$
$$= P\left(0.17 \le Z_n \le 0.5\right) = \Phi(0.5) - \Phi(0.17) = 0.6915 - 0.5675 = 0.124$$

Точное значение:

$$P(S_n = 19) = C_{36}^{19}\frac{1}{2^{36}} = 0.1251.$$

Задача об опросе, нормальная аппроксимация

Перед голосованием (референдумом) проводится опрос с целью выяснить, какова доля p населения, которая будет голосовать из “за”. Наудачу выбранный человек будет голосовать “за” с вероятностью p. Последовательно опрашивая таких людей, получаем последовательность независимых бернуллиевских случайных величин ξ_i: $\mathsf{P}(\xi_i = 1) = p$ (за), $\mathsf{P}(\xi_i = 1) = 1 - p$ (против).
Хотим найти для p *доверительный интервал* — случайный интервал, который накроет p с вероятностью $1-\alpha$. В качестве оценки p возьмем частоту появления ответа “за”:

$$M_n = \frac{1}{n}\sum_{i=1}^{n} \xi_i.$$

Воспользуемся сначала неравенством Чебышёва:

$$\mathsf{P}\left(|M_n - p| > \varepsilon\right) \le \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{\mathrm{Var}(\xi)}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2} = \alpha.$$

Здесь использовано неравенство $p(1-p) \le 1/4$.

Полагая $\varepsilon = 1/(2\sqrt{\alpha n})$, находим

$$\mathsf{P}\left(|M_n - p| > \frac{1}{2\sqrt{\alpha n}}\right) \le \alpha$$

и значит

$$p \in \left(M_n - \frac{1}{2\sqrt{\alpha n}}, M_n + \frac{1}{2\sqrt{\alpha n}}\right)$$

с вероятностью не менее $1 - \alpha$.
Потребуем, что погрешность в определении p не превосходила 0.03 с вероятностью не менее 0.95. Полагая $\alpha = 0.05$, заключаем что должно выполняться неравенство

$$\frac{1}{2\sqrt{0.05n}} \le 0.03 \Longrightarrow n \ge 5555.5(5)$$

Рассмотрим теперь нормальную аппроксимацию. Положим

$$\eta_i = \frac{\xi_i - \mathsf{E}\xi_i}{\sigma(\xi_i)} = \frac{\xi_i - p}{\sqrt{p(1-p)}}$$

Тогда

$$\begin{aligned}
\mathsf{P}\left(|M_n - p| > \varepsilon\right) &= \mathsf{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}(\xi_i - p)\right| > \varepsilon\right) \\
&= \mathsf{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\eta_i\right| > \frac{\sqrt{n}}{\sqrt{p(1-p)}}\varepsilon\right) \\
&\leq \mathsf{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\eta_i\right| > 2\sqrt{n}\varepsilon\right) \\
&\approx \mathsf{P}(|\zeta| > 2\sqrt{n}\varepsilon) \leq \alpha.
\end{aligned}$$

Снова потребуем, чтобы ошибка не превзошла $\varepsilon = 0.03$ с вероятностью не менее $1 - \alpha = 0.95$.

Пример (продолжение)

$$\mathsf{P}(|\zeta| > 0.06\sqrt{n}) = 2\mathsf{P}(\zeta > 0.06\sqrt{n}) = 2(1 - \Phi(0.06\sqrt{n})) \le 0.05$$

$$\Phi(0.06\sqrt{n}) \ge 0.975$$

По таблице

$$\Phi(1.96) \approx 0.975$$

Следовательно,

$$\sqrt{n} \approx \frac{1.96}{0.06} = 32.6(6) \Longrightarrow n \approx 1067.$$

Для $\varepsilon = 0.01$ получилось бы

$$\sqrt{n} \approx \frac{1.96}{0.02} = 98 \Longrightarrow n \approx 9604.$$

Условное математическое ожидание относительно σ-алгебры

Рассмотрим σ-алгебру $\mathscr{H} \subset \mathscr{F}$. Пусть $\xi \in L^1(\mathscr{F}) := L^1(\Omega, \mathscr{F}, \mathsf{P}, \mathbb{R})$.

- Случайная величина $\eta \in L^1(\mathscr{H})$ называется *условным математическим ожиданием* ξ относительно $\mathscr{H}$, если
$$\mathsf{E}(\xi I_A) = \mathsf{E}(\eta I_A), \quad A \in \mathscr{H}.$$
Принято обозначение $\eta = \mathsf{E}(\xi|\mathscr{H})$.

- Из свойств интеграла вытекает, что если $\eta' = \eta$ п.н., т.е. $\mathsf{P}(\eta = \eta') = 1$, то
$$\mathsf{E}(\eta I_A) = \mathsf{E}(\eta' I_A).$$
Следовательно, η' также будет у.м.о. Таким образом, у.м.о. определено с точностью до множества нулевой меры.

Теорема

Для любого $\xi \in L^1(\mathscr{F})$ *условное математическое ожидание* $\eta = \mathsf{E}(\xi|\mathscr{H}) \in L^1(\mathscr{H})$ *существует и единственно (как класс эквивалентности).*

Очевидные свойства:

- Пусть $\xi = C$, тогда $\mathsf{E}(\xi|\mathscr{H}) = \xi$ (по определению).
- Пусть $\xi \in L^1(\mathscr{H})$, тогда $\mathsf{E}(\xi|\mathscr{H}) = \xi$ (по определению).
- Пусть $\mathscr{H} = \{\emptyset, \Omega\}$. Тогда $\mathsf{E}(\xi|\mathscr{H}) = \mathsf{E}\xi$.

Пример

Пусть имеется разбиение $(A_i)_{i=1}^m$, причем вероятностная мера нагружает все его атомы: $\mathsf{P}(A_i) > 0$, $i = 1, \ldots, m$, и $\mathscr{H}$ порождается разбиением $(A_i)_{i=1}^m$. Тогда

$$\mathsf{E}(\xi|\mathscr{H}) = \sum_{i=1}^m \frac{\mathsf{E}(\xi I_{A_i})}{\mathsf{P}(A_i)} I_{A_i}, \qquad \xi \in L^1(\mathscr{F}).$$

Действительно, обозначим правую часть последнего равенства через η. Для любого атома A_j имеем

$$\mathsf{E}(\eta I_{A_j}) = \mathsf{E}\left(\sum_{i=1}^m \frac{\mathsf{E}(\xi I_{A_i})}{\mathsf{P}(A_i)} I_{A_i} I_{A_j}\right) = \mathsf{E}\left(\frac{\mathsf{E}(\xi I_{A_j})}{\mathsf{P}(A_j)} I_{A_j}\right) = \mathsf{E}(\xi I_{A_j}).$$

Утверждение вытекает из того, что любое множество $A \in \mathscr{H}$ является объединением атомов A_j.
Полученная формула показывает, что вычисление условного математического ожидания сводится к усреднению ξ по атомам разбиения.

Пример

Модель n-кратного подбрасывания монеты.

- $\Omega = \{(a_1, \dots, a_n) : a_i \in \{H, T\}\}$,
- $\mathscr{F}_t = \sigma(\{\omega : (a_1, \dots, a_t) \text{ фиксированы}\})$,
- $\mathsf{P}(\omega) = p^{\nu(\omega)} q^{n-\nu(\omega)}$, $\nu(\omega)$ — количество H в последовательности $\omega = (a_1, \dots, a_n)$, $p \in (0,1)$, $q = 1 - p$.
- $A_H := H_1 = \{\omega : a_1 = H\}$,
- $A_{HH} := H_1 \cap H_2 = \{\omega : a_1 = a_2 = H\}$,
- $A_{HHH} := H_1 \cap H_2 \cap H_3 = \{\omega : a_1 = a_2 = a_3 = H\}$.

$$
\begin{aligned}
\mathsf{E} I_{A_{HHH}} &= \mathsf{P}(A_{HHH}) = p^3, \\
\mathsf{E}(I_{A_{HHH}} | \mathscr{F}_1) &= \frac{\mathsf{E}(I_{A_{HHH}} I_{A_H})}{\mathsf{P}(A_H)} I_{A_H} + \frac{\mathsf{E}(I_{A_{HHH}} I_{A_T})}{\mathsf{P}(A_T)} I_{A_T} \\
&= \frac{\mathsf{P}(A_{HHH})}{\mathsf{P}(A_H)} I_{A_H} = \frac{p^3}{p} I_{A_H} = p^2 I_{A_H}.
\end{aligned}
$$

Пример (продолжение)

$$
\begin{aligned}
\mathsf{E}(I_{A_{HHH}}|\mathscr{F}_2) &= \frac{\mathsf{E}(I_{A_{HHH}} I_{A_{HH}})}{\mathsf{P}(A_{HH})} I_{A_{HH}} + \frac{\mathsf{E}(I_{A_{HHH}} I_{A_{HT}})}{\mathsf{P}(A_{HT})} I_{A_{HT}} \\
&+ \frac{\mathsf{E}(I_{A_{HHH}} I_{A_{TH}})}{\mathsf{P}(A_{TH})} I_{A_{TH}} + \frac{\mathsf{E}(I_{A_{HHH}} I_{A_{TT}})}{\mathsf{P}(A_{TT})} I_{A_{TT}} \\
&= \frac{\mathsf{E}(I_{A_{HHH} \cap A_{HH}})}{\mathsf{P}(A_{HH})} I_{A_{HH}} = \frac{\mathsf{P}(A_{HHH})}{\mathsf{P}(A_{HH})} I_{A_{HH}} = \frac{p^3}{p^2} I_{A_{HH}} \\
&= p I_{A_{HH}}, \\
\mathsf{E}(I_{A_{HHH}}|\mathscr{F}_3) &= I_{A_{HHH}}.
\end{aligned}
$$

Пример

На том же вероятностном пространстве определим последовательность случайных S_t величин (случайный процесс) по рекуррентной формуле

$$S_{t+1} = S_t(uI_{\{a_{t+1}=H\}} + dI_{\{a_{t+1}=T\}}),$$

где S_0 — некоторая константа, $u > d$. Данный процесс является простейшей моделью цены рискового актива (биномиальная модель, модель Кокса-Росса-Рубинштейна). Найдем $\mathsf{E}(S_2|\mathscr{F}_1)$.
По общей формуле

$$\mathsf{E}(S_2|\mathscr{F}_1) = \frac{\mathsf{E}(S_2 I_{A_H})}{\mathsf{P}(A_H)} I_{A_H} + \frac{\mathsf{E}(S_2 I_{A_T})}{\mathsf{P}(A_T)} I_{A_T}.$$

Поскольку $S_2 = S_0(u^2 I_{A_{HH}} + udI_{A_{HT}} + duI_{A_{TH}} + d^2 I_{A_{TT}})$, то

$$\mathsf{E}(S_2 I_{A_H}) = S_0 \mathsf{E}(u^2 I_{A_{HH}} + udI_{A_{HT}}) = S_0(u^2p^2 + udpq),$$

$$\mathsf{E}(S_2 I_{A_T}) = S_0 \mathsf{E}(duI_{A_{TH}} + d^2 I_{A_{TT}}) = S_0(udpq + d^2q^2).$$

Пример (продолжение)

Таким образом,

$$
\begin{aligned}
\mathsf{E}(S_2|\mathscr{F}_1) &= \frac{S_0(u^2p^2 + udpq)}{p} I_{A_H} + \frac{S_0(udpq + d^2q^2)}{q} I_{A_T} \\
&= S_0u(up + dq)I_{A_H} + S_0d(up + dq)I_{A_T} \\
&= (up + dq)S_1
\end{aligned}
$$

Пример

- $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}(0,1)$, $\mathsf{P} = \lambda$ — мера Лебега.
- Пусть $\xi(\omega) = \cos(\pi\omega)$, $\mathscr{H} = \{A \subset (0,1) : A \text{ или } A^c \text{ счетно}\}$.

Если A счетно, то

$$\mathsf{E}(\xi I_A) = \int_A \cos(\pi\omega)\, d\omega = 0.$$

Если A^c счетно, то

$$\mathsf{E}(\xi I_A) = \int_0^1 \cos(\pi\omega)\, d\omega - \int_{A^c} \cos(\pi\omega)\, d\omega = 0.$$

Следовательно, $\mathsf{E}(\xi|\mathscr{H}) = 0$.

Свойства условного математического ожидания

(1) Условное математическое ожидание является линейным оператором из $L^1(\mathscr{F})$ в $L^1(\mathscr{H})$:

$$\mathsf{E}(a_1\xi_1 + a_2\xi_2|\mathscr{H}) = a_1\mathsf{E}(\xi_1|\mathscr{H}) + a_2\mathsf{E}(\xi_2|\mathscr{H}).$$

Вытекает из линейности математического ожидания.

(2) Пусть $\xi \in L^1(\mathscr{F})$ и η — $\mathscr{H}$-измерима и ограничена. Тогда

$$\mathsf{E}(\xi\eta|\mathscr{H}) = \eta\mathsf{E}(\xi|\mathscr{H}).$$

$\mathscr{H}$-измеримая случайная величина η ведет себя как константа при вычислении у.м.о. относительно $\mathscr{H}$.

Свойства условного математического ожидания

Доказательство. Нужно проверить, что

$$\mathsf{E}(\xi\eta I_A) = \mathsf{E}(\eta\mathsf{E}(\xi|\mathscr{H})I_A), \quad A \in \mathscr{H}.$$

Из представления $\xi = \xi^+ - \xi^-$ вытекает, что можно считать ξ неотрицательной.
Воспользуемся стандартной техникой. 1) Пусть $\eta = I_B$, $B \in \mathscr{H}$. Тогда левая и правая части равны соответственно

$$\mathsf{E}(\xi\eta I_A) = \mathsf{E}(\xi I_B I_A) = \mathsf{E}(\xi I_{A\cap B})$$

$$\mathsf{E}(\eta\mathsf{E}(\xi|\mathscr{H})I_A) = \mathsf{E}(\mathsf{E}(\xi|\mathscr{H})I_{A\cap B}) = \mathsf{E}(\xi I_{A\cap B}).$$

2) По линейности равенство верно для линейных комбинаций индикаторов.
3) На неотрицательные η равенство распространяется с помощью монотонного предельного перехода (теорема о монотонной сходимости).
4) На произвольные (ограниченные) функции η формула распространяется по линейности: $\eta = \eta^+ - \eta^-$.

Свойства условного математического ожидания

(3) Пусть $\xi \in L^1(\mathscr{F})$ не зависит от $\mathscr{H}$. Тогда

$$\mathsf{E}(\xi|\mathscr{H}) = \mathsf{E}\xi.$$

Таким образом, информация содержащаяся в $\mathscr{H}$ бесполезна, если ξ не зависит от $\mathscr{H}$.
Доказательство:

$$\mathsf{E}(\xi I_A) = \mathsf{E}\xi \cdot \mathsf{E}I_A = \mathsf{E}(I_A \mathsf{E}\xi).$$

(4) Пусть $\xi \in L^1(\mathscr{F})$. Тогда

$$\mathsf{E}\mathsf{E}(\xi|\mathscr{H}) = \mathsf{E}\xi$$

(формула полного математического ожидания, law of total expectation).
Доказательство. В определении у.м.о. положим $A = \Omega$:

$$\mathsf{E}\xi = \mathsf{E}(\xi I_\Omega) = \mathsf{E}(\mathsf{E}(\xi|\mathscr{H}) I_\Omega) = \mathsf{E}\mathsf{E}(\xi|\mathscr{H}).$$

Свойства условного математического ожидания

(5) Пусть $\xi \in L^1(\mathscr{F})$ и $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$. Тогда

$$\mathsf{E}(\mathsf{E}(\xi|\mathscr{G})|\mathscr{H}) = \mathsf{E}(\xi|\mathscr{H}).$$

Данное свойство имеет много названий: свойство последовательного усреднения, телескопическое свойство (tower property, law of iterated expectations). Оно обобщает свойство (4).

Доказательство. Нужно доказать, что

$$\mathsf{E}(\xi I_A) = \mathsf{E}(\mathsf{E}(\mathsf{E}(\xi|\mathscr{G})|\mathscr{H})I_A), \quad A \in \mathscr{H}.$$

Для преобразования правой части воспользуемся свойствами (2) и (4):

$$\mathsf{E}(\mathsf{E}(\mathsf{E}(\xi|\mathscr{G})|\mathscr{H})I_A) \overset{(2)}{=} \mathsf{E}(\mathsf{E}(\mathsf{E}(\xi|\mathscr{G})I_A|\mathscr{H})) \overset{(4)}{=} \mathsf{E}(\mathsf{E}(\xi|\mathscr{G})I_A)$$
$$\overset{(2)}{=} \mathsf{E}(\mathsf{E}(\xi I_A|\mathscr{G})) \overset{(4)}{=} \mathsf{E}(\xi I_A).$$

Свойства условного математического ожидания

(6) Пусть $\xi \in L^2(\mathscr{F})$. Тогда $\eta = \mathsf{E}(\xi|\mathscr{H}) \in L^2(\mathscr{H})$ является наилучшим L^2-приближением ξ (на основе информации, заключенной в $\mathscr{H}$):

$$\mathsf{E}(\xi - \eta)^2 = \min_{\zeta \in L^2(\mathscr{H})} \mathsf{E}(\xi - \zeta)^2.$$

Другими словами, $\xi \mapsto \mathsf{E}(\xi|\mathscr{H})$ является ортопроектором на $L^2(\mathscr{H})$.

Доказательство.

$$\begin{aligned} \mathsf{E}((\xi-\zeta)^2|\mathscr{H}) &= \mathsf{E}(\xi^2|\mathscr{H}) - 2\zeta\mathsf{E}(\xi|\mathscr{H}) + \zeta^2 \\ &= \mathsf{E}(\xi^2|\mathscr{H}) + (\mathsf{E}(\xi|\mathscr{H}) - \zeta)^2 - (\mathsf{E}(\xi|\mathscr{H}))^2, \\ \mathsf{E}\mathsf{E}((\xi-\zeta)^2|\mathscr{H}) &= \mathsf{E}(\mathsf{E}(\xi|\mathscr{H}) - \zeta)^2 + \mathsf{E}\xi^2 - \mathsf{E}(\mathsf{E}(\xi|\mathscr{H}))^2. \end{aligned}$$

Очевидно, что наименьшее значение достигается при $\zeta^* = \mathsf{E}(\xi|\mathscr{H})$.

Пример

В биномиальной модели найдем $\mathsf{E}(S_{t+1}|\mathscr{F}_t)$, $\mathsf{E}S_{t+1}$.
В формуле

$$S_{t+1} = S_t(uI_{\{a_{t+1}=H\}} + dI_{\{a_{t+1}=T\}}),$$

первый множитель является $\mathscr{F}_t$-измеримым, а второй не зависит от $\mathscr{F}_t$. Следовательно,

$$\begin{aligned}
\mathsf{E}(S_{t+1}|\mathscr{F}_t) &\overset{(2)}{=} S_t\mathsf{E}(uI_{\{a_{t+1}=H\}} + dI_{\{a_{t+1}=T\}}|\mathscr{F}_t) \\
&\overset{(3)}{=} S_t\mathsf{E}(uI_{\{a_{t+1}=H\}} + dI_{\{a_{t+1}=T\}}) \\
&= S_t(u\mathsf{E}I_{\{a_{t+1}=H\}} + d\mathsf{E}I_{\{a_{t+1}=T\}}) = (up+dq)S_t. \\
\mathsf{E}S_{t+1} &\overset{(4)}{=} \mathsf{E}\mathsf{E}(S_{t+1}|\mathscr{F}_t) = (up+dq)\mathsf{E}S_t = (up+dq)^2\mathsf{E}S_{t-1} \\
&= (up+dq)^{t+1}S_0.
\end{aligned}$$

Условное математическое ожидание относительно случайной величины

Пусть $\xi \in L^1(\mathscr{F})$. По определению,

$$\mathsf{E}(\xi|\eta) = \mathsf{E}(\xi|\mathscr{F}_\eta),$$

где $\mathscr{F}_\eta := \sigma(\eta) = \{\eta^{-1}(B), B \in \mathscr{B}(\mathbb{R})\}$ — σ-алгебра, порожденная случайной величиной η.

Пример

Пусть $\eta : \Omega \mapsto \{z_1, \ldots, z_m\}$. Тогда

$$\mathscr{F}_\eta = \sigma(A_1, \ldots, A_m), \quad A_i = \eta^{-1}(\{z_i\}).$$

$$\mathsf{E}(\xi|\eta) = \sum_{i=1}^{m} \frac{\mathsf{E}(\xi I_{A_i})}{\mathsf{P}(A_i)} I_{A_i} = \sum_{i=1}^{m} \frac{\mathsf{E}(\xi I_{\{\eta = z_i\}})}{\mathsf{P}(\eta = z_i)} I_{\{\eta = z_i\}}.$$

- Для случайной величины η, принимающей конечное число значений, в предыдущем примере была получена формула

$$\mathsf{E}(\xi|\eta) = \sum_{i=1}^{m} c_i I_{\{\eta=z_i\}} = g(\eta).$$

- Известно, что если ζ является $\sigma(\eta)$ измеримой, то $\zeta = g(\eta)$ для некоторой борелевской функции g (*лемма Дуба-Дынкина*).
- Отсюда следует, что для любой случайной величины η существует борелевская функция g такая, что $\mathsf{E}(\xi|\eta) = g(\eta)$. В связи с этим используется обозначение

$$\mathsf{E}(\xi|\eta = x) = g(x).$$

  В том числе появляется возможность вычислять условное математическое ожидание относительно событий $\{\eta = x\}$ нулевой вероятности.
- Заметим, что функция g определяется неединственным образом. Например, в приведенном выше примере можно было определить ее произвольным образом вне множества $\{z_1, \ldots, z_m\}$.

Пример

Пусть $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные интегрируемые случайные величины. Положим $\xi = \xi_1 + \cdots + \xi_n$ и найдем $\mathsf{E}(\xi_1|\xi)$, $\mathsf{E}(\xi|\xi_1)$.

Из соображений симметрии,

$$\mathsf{E}(\xi_1|\xi) = \cdots = \mathsf{E}(\xi_n|\xi).$$

Из определения условного математического ожидания и свойства линейности вытекает, что

$$\mathsf{E}(\xi_1|\xi) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(\xi_i|\xi) = \frac{1}{n}\mathsf{E}(\xi_1 + \cdots + \xi_n|\xi) = \frac{1}{n}\mathsf{E}(\xi|\xi) = \frac{\xi}{n}.$$

Далее,

$$\begin{aligned}\mathsf{E}(\xi|\xi_1) &= \mathsf{E}(\xi_2 + \cdots + \xi_n|\xi_1) + \mathsf{E}(\xi_1|\xi_1) = \mathsf{E}(\xi_2 + \cdots + \xi_n) + \xi_1 \\ &= (n-1)\mathsf{E}\xi_1 + \xi_1.\end{aligned}$$

Пример

Пусть $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные интегрируемые случайные величины. Положим $\xi = \xi_1 + \cdots + \xi_n$ и найдем $\mathsf{E}(\xi_1|\xi)$, $\mathsf{E}(\xi|\xi_1)$.

Из соображений симметрии,

$$\mathsf{E}(\xi_1|\xi) = \cdots = \mathsf{E}(\xi_n|\xi).$$

Из определения условного математического ожидания и свойства линейности вытекает, что

$$\mathsf{E}(\xi_1|\xi) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(\xi_i|\xi) = \frac{1}{n}\mathsf{E}(\xi_1 + \cdots + \xi_n|\xi) = \frac{1}{n}\mathsf{E}(\xi|\xi) = \frac{\xi}{n}.$$

Далее,

$$\begin{aligned}\mathsf{E}(\xi|\xi_1) &= \mathsf{E}(\xi_2 + \cdots + \xi_n|\xi_1) + \mathsf{E}(\xi_1|\xi_1) = \mathsf{E}(\xi_2 + \cdots + \xi_n) + \xi_1 \\ &= (n-1)\mathsf{E}\xi_1 + \xi_1.\end{aligned}$$

Пример

Пусть $\xi_1, \dots, \xi_n$ — независимые одинаково распределенные интегрируемые случайные величины. Положим $\xi = \xi_1 + \dots + \xi_n$ и найдем $\mathsf{E}(\xi_1|\xi)$, $\mathsf{E}(\xi|\xi_1)$.

Из соображений симметрии,

$$\mathsf{E}(\xi_1|\xi) = \dots = \mathsf{E}(\xi_n|\xi).$$

Из определения условного математического ожидания и свойства линейности вытекает, что

$$\mathsf{E}(\xi_1|\xi) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(\xi_i|\xi) = \frac{1}{n}\mathsf{E}(\xi_1 + \dots + \xi_n|\xi) = \frac{1}{n}\mathsf{E}(\xi|\xi) = \frac{\xi}{n}.$$

Далее,

$$\begin{aligned}\mathsf{E}(\xi|\xi_1) &= \mathsf{E}(\xi_2 + \dots + \xi_n|\xi_1) + \mathsf{E}(\xi_1|\xi_1) = \mathsf{E}(\xi_2 + \dots + \xi_n) + \xi_1 \\ &= (n-1)\mathsf{E}\xi_1 + \xi_1.\end{aligned}$$

Лемма о независимости (лемма о заморóзке)

Пусть ξ не зависит от η и $g : \mathbb{R}^2 \mapsto \mathbb{R}$ — (ограниченная) борелевская функция. Тогда

$$\mathsf{E}(g(\xi,\eta)|\eta) = h(\eta), \quad h(y) = \mathsf{E}g(\xi,y).$$

Доказательство. Достаточно доказать утверждения для индикаторов $g = I_{B_1\times B_2}$:

$$\begin{aligned}
\mathsf{E}(g(\xi,\eta)|\eta) &= \mathsf{E}(I_{B_1\times B_2}(\xi,\eta)|\eta) = \mathsf{E}(I_{B_1}(\xi)I_{B_2}(\eta)|\eta) \\
&= I_{B_2}(\eta)\mathsf{E}(I_{B_1}(\xi)|\eta) = I_{B_2}(\eta)\mathsf{E}(I_{B_1}(\xi)) \\
&= \mathsf{E}(I_{B_1}(\xi)I_{B_2}(y))|_{y=\eta} = h(\eta), \\
h(y) &= \mathsf{E}(I_{B_1}(\xi)I_{B_2}(y)) = \mathsf{E}g(\xi,y)
\end{aligned}$$

и воспользоваться стандартной техникой.

Пример

Пусть ξ и η независимые экспоненциально распределенные случайные величины с параметрами λ и μ соответственно. Найдем $\mathsf{P}(\xi \le \eta)$.

$$\mathsf{P}(\xi \le \eta) = \mathsf{E}\mathsf{E}(I_{\{\xi \le \eta\}}|\eta) = \mathsf{E}h(\eta),$$

$$h(y) = \mathsf{E}I_{\{\xi \le y\}} = \mathsf{P}(\xi \le y) = \int_0^y \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_{x=0}^{x=y} = 1 - e^{-\lambda y},$$

$$\mathsf{P}(\xi \le \eta) = 1 - \mathsf{E}e^{-\lambda\eta} = 1 - \int_0^\infty e^{-\lambda y}\mu e^{-\mu y}\,dy = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}.$$

Пример

Пусть $(\xi_i)_{i=1}^{\infty}$ — независимые одинаково распределенные случайные величины, $\mathsf{E}\xi_i^2 < \infty$, и пусть $N : \Omega \mapsto \mathbb{Z}_+$ не зависит от $(\xi_i)_{i=1}^{\infty}$. Рассмотрим случайную сумму

$$S_N = \sum_{i=1}^{N} \xi_i.$$

В теории страхования ξ_i — размер выплаты по страховому полису, N — количество полисов, предъявленных к оплате. Найдем $\mathsf{E}S_N$:

$$\mathsf{E}(S_N|N) = \mathsf{E}\left(\sum_{i=1}^{N} \xi_i \middle| N\right) = \mathsf{E}\left(\sum_{i=1}^{n} \xi_i\right)\Bigg|_{n=N} = (n\mathsf{E}\xi_1)\Big|_{n=N} = N \cdot \mathsf{E}\xi_1,$$

$$\mathsf{E}S_N = \mathsf{E}\mathsf{E}(S_N|N) = \mathsf{E}N \cdot \mathsf{E}\xi_1.$$

По определению условного математического ожидания:

$$\mathsf{E}((\xi - \mathsf{E}(\xi|\mathscr{H}))I_A) = 0, \quad A \in \mathscr{H}.$$

Стандартная техника:

$$\mathsf{E}((\xi - \mathsf{E}(\xi|\mathscr{H}))\eta) = 0, \quad \eta — \mathscr{H}\text{-измерима и ограничена.}$$

Если $\xi \in L^2(\mathscr{F})$, то

$$\mathsf{E}((\xi - \mathsf{E}(\xi|\mathscr{H}))\eta) = 0, \quad \eta \in L^2(\mathscr{H}),$$

т.е. $\xi - \mathsf{E}(\xi|\mathscr{H})$ ортогональна $L^2(\mathscr{H})$ в $L^2(\mathscr{F})$.

Пример

Введем *условную дисперсию*

$$\mathrm{Var}(\xi|\mathscr{H}) = \mathsf{E}((\xi - \mathsf{E}(\xi|\mathscr{H}))^2|\mathscr{H}) = \mathsf{E}(\xi^2|\mathscr{H}) - (\mathsf{E}(\xi|\mathscr{H}))^2.$$

По указанному свойству случайные величины

$$\xi - \mathsf{E}(\xi|\mathscr{H}) \in L^2(\mathscr{F}), \quad \mathsf{E}(\xi|\mathscr{H}) - \mathsf{E}\xi \in L^2(\mathscr{H})$$

ортогональны в $L^2(\mathscr{F})$. Следовательно,

$$\begin{aligned}
\mathrm{Var}(\xi) &= \mathsf{E}(\xi - \mathsf{E}(\xi|\mathscr{H}) + \mathsf{E}(\xi|\mathscr{H}) - \mathsf{E}\xi)^2 \\
&= \mathsf{E}(\xi - \mathsf{E}(\xi|\mathscr{H}))^2 + \mathsf{E}(\mathsf{E}(\xi|\mathscr{H}) - \mathsf{E}\xi)^2 \\
&= \mathsf{E}\mathsf{E}((\xi - \mathsf{E}(\xi|\mathscr{H}))^2|\mathscr{H}) + \mathsf{E}(\mathsf{E}(\xi|\mathscr{H}) - \mathsf{E}\mathsf{E}(\xi|\mathscr{H}))^2 \\
&= \mathsf{E}\mathrm{Var}(\xi|\mathscr{H}) + \mathrm{Var}(\mathsf{E}(\xi|\mathscr{H})).
\end{aligned}$$

Таким образом, дисперсия случайной величины равна сумме математического ожидания условной дисперсии и дисперсии условного математического ожидания.

Пример

Для рассмотренной выше случайной суммы

$$S_N = \sum_{i=1}^{N} \xi_i,$$

где $(\xi_i)_{i=1}^{\infty}$ не зависят от N, имеем

$$\begin{aligned}
\mathsf{E}(S_N|N) &= \mathsf{E}S_n\Big|_{n=N} = N\mathsf{E}\xi_1,\\
\mathrm{Var}(S_N|N) &= \mathrm{Var}(S_n)\Big|_{n=N} = N\mathrm{Var}(\xi_1),\\
\mathrm{Var}(S_N) &= \mathsf{E}\mathrm{Var}(S_N|N) + \mathrm{Var}(\mathsf{E}(S_N|N))\\
&= \mathsf{E}N \cdot \mathrm{Var}(\xi_1) + \mathrm{Var}(N) \cdot (\mathsf{E}\xi_1)^2.
\end{aligned}$$

Переходная вероятность

Функция $Q : \mathbb{R}^m \times \mathscr{B}(\mathbb{R}^d) \mapsto [0,1]$ называется *переходной вероятностью* из $\mathbb{R}^m$ в $\mathbb{R}^d$, если

- при любом $B \in \mathscr{B}(\mathbb{R}^d)$ функция $y \mapsto Q(y, B)$ является $\mathscr{B}(\mathbb{R}^m)$-измеримой;
- при любом $y \in \mathbb{R}^m$ функция $B \mapsto Q(y, B)$ является вероятностной мерой на $\mathscr{B}(\mathbb{R}^d)$.

Условное распределение

Рассмотрим случайные величины $\xi : \Omega \mapsto \mathbb{R}^d$, $\eta : \Omega \mapsto \mathbb{R}^m$. Переходная вероятность $\mathsf{P}_{\xi|\eta}(B|y) = Q(y,B)$ из $\mathbb{R}^m$ в $\mathbb{R}^d$ называется *(регулярным) условным распределением* ξ *относительно* η, если

$$\mathsf{E}(g(\xi)|\eta)(\omega) = \int_{\mathbb{R}^d} g(x)\, \mathsf{P}_{\xi|\eta}(dx|\eta) \quad \mathsf{P}\text{-п.н.} \tag{1}$$

для любой ограниченной борелевской функции на $\mathbb{R}^d$.
Известно, что регулярное условное распределение всегда существует. Заметим, что по упомянутой лемме Дуба-Дынкина существует функция $(y,B) \mapsto \mathsf{P}_{\xi|\eta}(B|y) : \mathbb{R}^m \times \mathscr{B}(\mathbb{R}^d) \mapsto [0,1]$ такая, что

$$\mathsf{E}(I_B(\xi)|\eta) = \mathsf{P}_{\xi|\eta}(B|\eta), \quad A \in \mathscr{B}(\mathbb{R}^d).$$

Более того, функция $y \mapsto \mathsf{P}_{\xi|\eta}(B|y)$ является $\mathscr{B}(\mathbb{R}^m)$-измеримой. Однако, доказательство того, что Q может быть выбрана таким образом, что $B \mapsto \mathsf{P}_{\xi|\eta}(B|y)$ является вероятностной мерой на $\mathscr{B}(\mathbb{R}^d)$, требует дополнительных рассуждений.

Формула в определении условного распределения обобщается следующим образом:

$$\mathsf{E}(g(\xi,\eta)|\eta) = \int_{\mathbb{R}^d} g(x,\eta)\mathsf{P}_{\xi|\eta}(dx|\eta).$$

Доказательство можно провести с помощью стандартной техники.

- Лемма о независимости (о заморо́зке) для условных распределений.
  Пусть ξ, η независимы, тогда условное распределение $\zeta = \psi(\xi,\eta)$ относительно $\eta = y$ совпадает с распределением случайной величины $\psi(\xi, y)$:

$$\mathsf{E}(g(\psi(\xi,\eta))|\eta) = \mathsf{E}g(\psi(\xi,y))\Big|_{y=\eta} = \int_{\mathbb{R}} g(z)\mathsf{P}_{\psi(\xi,y)}(dz)\Big|_{y=\eta},$$

$$\mathsf{E}(g(\zeta)|\eta) = \int_{\mathbb{R}} g(z)\mathsf{P}_{\zeta|\eta}(dz|\eta).$$

Значит $\mathsf{P}_{g(\zeta)|\eta}(dz|\eta) = \mathsf{P}_{\psi(\xi,y)}(dz)\Big|_{y=\eta}$.

Пример: анализ геометрического распределения

Пусть $\xi \sim G(p)$: $\mathsf{P}(\xi = n) = q^{n-1}p$, $n \in \mathbb{N}$. Найдем $\mathsf{P}_{\xi|\eta}$, $\eta = I_{\{\xi>1\}}$.
Искомое условное распределение удовлетворяет равенству

$$\begin{aligned}
\mathsf{E}(g(\xi)h(\eta)) = \mathsf{E}(\mathsf{E}(g(\xi)|\eta)h(\eta)) &= \mathsf{E}\left(\int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|\eta)h(\eta)\right) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|y)h(y)\mathsf{P}_{\eta}(dy)\right) \\
&= \int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|0)h(0)p + \int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|1)h(1)q,
\end{aligned}$$

так как $\mathsf{P}(\eta = 0) = \mathsf{P}(\xi = 1) = p$, $\mathsf{P}(\eta = 1) = \mathsf{P}(\xi > 1) = q$.
Рассмотрим левую часть:

$$\begin{aligned}
\mathsf{E}(g(\xi)h(\eta)) = \mathsf{E}(g(\xi)h(I_{\{\xi>1\}})) &= \sum_{k=1}^{\infty} g(k)h(I_{\{k>1\}})q^{k-1}p \\
&= g(1)h(0)p + \sum_{k=2}^{\infty} g(k)q^{k-1}ph(1)
\end{aligned}$$

Пример: анализ геометрического распределения

$$
\begin{aligned}
\mathsf{E}(g(\xi)h(\eta)) &= g(1)h(0)p + \sum_{j=1}^{\infty} g(j+1)q^j p h(1) \\
&= g(1)\cdot h(0)p + \sum_{j=1}^{\infty} g(j+1)q^{j-1}p \cdot h(1)q \\
&= g(1)\cdot h(0)p + \mathsf{E}g(\xi+1)\cdot h(1)q.
\end{aligned}
$$

В силу произвольности h,

$$
\int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|0) = g(1), \quad \int_{\mathbb{R}} g(x)\,\mathsf{P}_{\xi|\eta}(dx|1) = \mathsf{E}g(\xi+1).
$$

Это означает, что

$$
\mathsf{P}_{\xi|\eta}(dx|0) = \delta_1(dx), \quad 1+\xi \sim \mathsf{P}_{\xi|\eta}(dx|1).
$$

Математическое ожидание геометрического распределения

$$
\begin{aligned}
\mathsf{E}\xi &= \mathsf{E}\mathsf{E}(\xi|\eta) = \mathsf{E}\int_{\mathbb{R}} x\mathsf{P}_{\xi|\eta}(dx|\eta) \\
&= \int_{\mathbb{R}} x\mathsf{P}_{\xi|\eta}(dx|0)\cdot p + \int_{\mathbb{R}} x\mathsf{P}_{\xi|\eta}(dx|1)\cdot q \\
&= p + q\mathsf{E}(1+\xi) = 1 + q\mathsf{E}\xi, \\
\mathsf{E}\xi &= \frac{1}{1-q} = \frac{1}{p}
\end{aligned}
$$

Дисперсия геометрического распределения

$$
\begin{aligned}
\mathsf{E}\xi^2 &= \mathsf{E}\mathsf{E}(\xi^2|\eta) = \mathsf{E}\int_{\mathbb{R}} x^2 \mathsf{P}_{\xi|\eta}(dx|\eta) \\
&= \int_{\mathbb{R}} x^2 \mathsf{P}_{\xi|\eta}(dx|0)\cdot p + \int_{\mathbb{R}} x^2 \mathsf{P}_{\xi|\eta}(dx|1)\cdot q \\
&= p + q\mathsf{E}(1+\xi)^2 = p + q(\mathsf{E}\xi^2 + 2\mathsf{E}\xi + 1) = 1 + q\mathsf{E}\xi^2 + 2\frac{q}{p} \\
\mathsf{E}\xi^2 &= \frac{1}{1-q}\left(1 + 2\frac{q}{p}\right) = \frac{1}{p} + 2\frac{1-p}{p^2} = \frac{2}{p^2} - \frac{1}{p} \\
\mathrm{Var}(\xi) &= \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

Анализ двумерных дискретных случайных величин

Рассмотрим двумерную случайную величину $(\xi, \eta) : \Omega \mapsto \mathbb{R}^2$ с распределением

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

Предположим, что существует такое конечное или счетное множество $Z \subset \mathbb{R}^2$, что

$$\mathsf{P}((\xi, \eta) \in Z) = \mathsf{P}_{\xi,\eta}(Z) = 1.$$

Пусть $p_{\xi,\eta}(x, y) = \mathsf{P}_{\xi,\eta}(x, y)$ — соответствующая вероятностная масса. По формуле замены переменных

$$\mathsf{E}g(\xi, \eta) = \sum_{(x,y)\in Z} g(x, y) p_{\xi,\eta}(x, y).$$

Найдем $\mathsf{E}(g(\xi)|\eta)$ и условное распределение ξ относительно η.

Согласно общей теории,

$$\mathsf{E}(g(\xi)|\eta) = r(\eta).$$

По определению у.м.о. должно выполняться равенство

$$\mathsf{E}(g(\xi)h(\eta)) = \mathsf{E}(r(\eta)h(\eta))$$

для любой ограниченной борелевской функции h (индикатор $I_A = h(\eta)$, $A \in \mathscr{F}_\eta$ является частным случаем, общий устанавливается с помощью стандартной техники). По формуле замены переменных,

$$\mathsf{E}(g(\xi)h(\eta)) = \sum_{x,y} g(x)h(y)p_{\xi,\eta}(x,y)$$

$$\mathsf{E}(r(\eta)h(\eta)) = \sum_{y} r(y)h(y)p_\eta(y),$$

где $p_\eta(y)$ — маргинальное распределение η. В силу произвольности h,

$$r(y)p_\eta(y) = \sum_{x} g(x)p_{\xi,\eta}(x,y).$$

Следовательно,

$$\mathsf{E}(g(\xi)|\eta) = r(\eta) = \sum_x g(x)\frac{p_{\xi,\eta}(x,\eta)}{p_\eta(\eta)}.$$

Введем *условную вероятностную массу*

$$p_{\xi|\eta}(x|y) = \frac{p_{\xi,\eta}(x,y)}{p_\eta(y)} \left(= \frac{\mathsf{P}(\xi = x, \eta = y)}{\mathsf{P}(\eta = y)}\right).$$

Тогда

$$\mathsf{E}(g(\xi)|\eta) = \sum_x g(x) p_{\xi|\eta}(x|\eta).$$

Следовательно,

$$\mathsf{P}_{\xi|\eta}(A|y) = \sum_{x \in A} p_{\xi|\eta}(x|y)$$

является условным распределением ξ относительно η.

Формула полной вероятности

$$p_\xi(x) = \mathsf{P}_\xi(x) = \mathsf{E} I_{\{\xi = x\}} = \mathsf{E}\mathsf{E}(I_{\{\xi=x\}}|\eta) = \mathsf{E} p_{\xi|\eta}(x|\eta)$$
$$= \sum_y p_{\xi|\eta}(x|y) p_\eta(y).$$

Формула полного математического ожидания

$$\mathsf{E} g(\xi) = \mathsf{E}\mathsf{E}(g(\xi)|\eta) = \mathsf{E} \sum_x g(x) p_{\xi|\eta}(x|\eta) = \sum_y \sum_x g(x) p_{\xi|\eta}(x|y) p_\eta(y).$$

Формула Байеса

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

$$p_{\eta|\xi}(y|x) = \frac{p_{\xi|\eta}(x|y) p_\eta(y)}{p_\xi(x)} = \frac{p_{\xi|\eta}(x|y) p_\eta(y)}{\sum_{y'} p_{\xi|\eta}(x|y') p_\eta(y')}.$$

Анализ двумерных непрерывных случайных величин

Рассмотрим двумерную непрерывную случайную величину $(\xi, \eta) : \Omega \mapsto \mathbb{R}^2$ с распределением

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

Предположим, что существует борелевская функция (плотность) $f_{\xi,\eta} : \mathbb{R}^2 \mapsto \mathbb{R}$ такая, что

$$\mathsf{P}((\xi, \eta) \in B) = \int_B f_{\xi,\eta}(x, y)\, dxdy.$$

По формуле замены переменных

$$\mathsf{E} g(\xi, \eta) = \int_{\mathbb{R}^2} g(x, y) f_{\xi,\eta}(x, y)\, dxdy.$$

Найдем $\mathsf{E}(g(\xi)|\eta)$ и условное распределение ξ относительно η.

Согласно общей теории, $\mathsf{E}(g(\xi)|\eta) = r(\eta)$. По определению у.м.о. должно выполняться равенство

$$\mathsf{E}(g(\xi)h(\eta)) = \mathsf{E}(r(\eta)h(\eta))$$

для любой ограниченной борелевской функции h. По формуле замены переменных,

$$\begin{aligned}\mathsf{E}(g(\xi)h(\eta)) &= \int_{\mathbb{R}^2} g(x)h(y)f_{\xi,\eta}(x,y)\,dxdy \\ &= \int_{\mathbb{R}} h(y)\left(\int_{\mathbb{R}} g(x)f_{\xi,\eta}(x,y)\,dx\right)\,dy\end{aligned}$$

$$\mathsf{E}(r(\eta)h(\eta)) = \int_{\mathbb{R}} r(y)h(y)f_{\eta}(y)\,dy,$$

где $f_\eta(y)$ — маргинальная плотность η. В силу произвольности h,

$$r(y)f_\eta(y) = \int_{\mathbb{R}} g(x)f_{\xi,\eta}(x,y)\,dx.$$

Следовательно,

$$\mathsf{E}(g(\xi)|\eta) = r(\eta) = \int_{\mathbb{R}} g(x)\frac{f_{\xi,\eta}(x,\eta)}{f_\eta(\eta)}\,dx.$$

Введем *условную плотность*

$$f_{\xi|\eta}(x|y) = \frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}.$$

Тогда

$$\mathsf{E}(g(\xi)|\eta) = \int_{\mathbb{R}} g(x) f_{\xi|\eta}(x|\eta)\,dx.$$

Следовательно,

$$\mathsf{P}_{\xi|\eta}(A|y) = \int_{x\in A} f_{\xi|\eta}(x|y)\,dx$$

является условным распределением ξ относительно η.

Формула полной вероятности

$$f_\xi(x) = \int_{\mathbb{R}} f_{\xi,\eta}(x,y)\,dy = \int_{\mathbb{R}} f_{\xi|\eta}(x|y) f_\eta(y)\,dy$$

Формула полного математического ожидания

$$\begin{aligned}
\mathsf{E} g(\xi) &= \mathsf{E}\mathsf{E}(g(\xi)|\eta) = \mathsf{E}\int_{\mathbb{R}} g(x) f_{\xi|\eta}(x|\eta)\,dx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x) f_{\xi|\eta}(x|y)\,dx \right) f_\eta(y)\,dy \\
&= \int_{\mathbb{R}} \mathsf{E}(g(\xi)|\eta = y) f_\eta(y)\,dy
\end{aligned}$$

Формула Байеса

$$f_{\xi,\eta}(x,y) = f_{\xi|\eta}(x|y) f_\eta(y) = f_{\eta|\xi}(y|x) f_\xi(x),$$

$$f_{\eta|\xi}(y|x) = \frac{f_{\xi|\eta}(x|y) f_\eta(y)}{f_\xi(x)} = \frac{f_{\xi|\eta}(x|y) f_\eta(y)}{\int_{\mathbb{R}} f_{\xi|\eta}(x|y') f_\eta(y')\, dy'}.$$

Пример

Имеется трость длины l (интервал $(0,l)$). Мы ломаем ее в точке $\xi \sim U(0,l)$, а затем в точке $\eta \sim U(0,\xi)$. Найдем $\mathsf{E}(\eta|\xi)$, $\mathsf{E}\eta$, совместное распределение (ξ,η) и маргинальное распределение η.

$$f_\xi(x) = \frac{1}{l} I_{[0,l]}(x), \quad f_{\eta|\xi}(y|x) = \frac{1}{x} I_{[0,x]}(y),$$

$$\mathsf{E}(\eta|\xi = x) = \int y f_{\eta|\xi}(y|x)\, dy = \frac{1}{x}\int_0^x y\, dy = \frac{x}{2},$$

$$\mathsf{E}\eta = \mathsf{E}\mathsf{E}(\eta|\xi) = \mathsf{E}\frac{\xi}{2} = \frac{l}{4},$$

$$f_{\xi,\eta}(x,y) = f_\xi(x) f_{\eta|\xi}(y|x) = \frac{1}{l} I_{[0,l]}(x) \frac{1}{x} I_{[0,x]}(y) = \begin{cases} \frac{1}{lx}, & 0 \le y \le x \le l, \\ 0, & \text{иначе}, \end{cases}$$

$$f_\eta(y) = \int f_{\xi,\eta}(x,y)\, dx = \int_y^l \frac{1}{lx}\, dx = \frac{1}{l} \ln x \Big|_{x=y}^{x=l} = \frac{1}{l} \ln \frac{l}{y}, \quad y \in [0,l].$$

Пример: обнаружение сигнала

Пусть одномерный сигнал η распределен по стандартному нормальному закону $N(0,1)$. Шум $w \sim N(0,\sigma^2)$ не зависит от сигнала. Наблюдению доступна сумма $\xi = \eta + w$. Требуется найти условное распределение η относительно ξ.
Решение. В данном случае

$$f_\eta(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad f_{\xi|\eta}(x|y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-y)^2/(2\sigma^2)},$$

так как $f_{\xi|\eta}(x|y)$ является плотностью случайной величины

$$y + w \sim N(y,\sigma^2)$$

(см. лемму о независимости для условных распределений).

По формуле Байеса,

$$f_{\eta|\xi}(y|x) = c_1(x) f_{\xi|\eta}(x|y) f_\eta(y) = c_2(x) \exp\left(-\frac{1}{2}\left(y^2 + \frac{(x-y)^2}{\sigma^2}\right)\right)$$
$$= c_3(x) \exp\left(-\frac{1+\sigma^2}{2\sigma^2}\left(y - \frac{x}{1+\sigma^2}\right)^2\right),$$

где $c_i(x)$ — некоторые нормирующие множители. Отсюда следует, что при фиксированном значении $\xi = x$ случайная величина η имеет нормальное распределение

$$N\left(\frac{x}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right).$$

Случай когда ξ непрерывна, η дискретна

Пусть ξ непрерывна и имеет плотность $f_\xi(x)$, а η дискретна и имеет вероятностную массу $p_\eta(y)$. *Введем* условную плотность $f_{\xi|\eta}(x|y)$,

$$f_{\xi|\eta}(x|y) \geq 0, \quad \int_{\mathbb{R}} f_{\xi|\eta}(x|y)\,dx = 1.$$

Другими словами, будем считать, что условное распределение ξ относительно η задается формулой

$$\mathsf{E}(g(\xi,\eta)|\eta) = \int_{\mathbb{R}} g(x,\eta) f_{\xi|\eta}(x|\eta)\,dx.$$

Теперь можно найти совместное распределение:

$$\begin{aligned}\mathsf{E}g(\xi,\eta) &= \mathsf{E}\mathsf{E}(g(\xi,\eta)|\eta) = \mathsf{E}\int_{\mathbb{R}} g(x,\eta) f_{\xi|\eta}(x|\eta)\,dx \\ &= \sum_y \int_{\mathbb{R}} g(x,y) f_{\xi|\eta}(x|y)\,dx \cdot p_\eta(y).\end{aligned}$$

Частным случаем является формула полного математического ожидания:

$$\mathsf{E}g(\xi) = \sum_y \int_{\mathbb{R}} g(x) f_{\xi|\eta}(x|y)\, dx \cdot p_\eta(y) = \int_{\mathbb{R}} g(x) \sum_y f_{\xi|\eta}(x|y) p_\eta(y)\, dx.$$

Отсюда вытекает формула полной вероятности:

$$f_\xi(x) = \sum_y f_{\xi|\eta}(x|y) p_\eta(y).$$

Найдем $\mathsf{E}(g(\eta)|\xi) = r(\xi)$. По определению,

$$\mathsf{E}(g(\eta)h(\xi)) = \mathsf{E}(r(\xi)h(\xi))$$

для любой ограниченной борелевской функции h.

$$
\begin{aligned}
\mathsf{E}(g(\eta)h(\xi)) &= \sum_y \int_{\mathbb{R}} g(y)h(x)f_{\xi|\eta}(x|y)\,dx \cdot p_\eta(y) \\
&= \int_{\mathbb{R}} \left(\sum_y g(y) f_{\xi|\eta}(x|y) p_\eta(y) \right) h(x)\,dx, \\
\mathsf{E}(r(\xi)h(\xi)) &= \int_{\mathbb{R}} r(x)h(x)f_\xi(x)\,dx.
\end{aligned}
$$

Следовательно,

$$
r(x) = \sum_y g(y) \frac{f_{\xi|\eta}(x|y)p_\eta(y)}{f_\xi(x)},
$$

$$
\mathsf{E}(g(\eta)|\xi) = r(\xi) = \sum_y g(y) \frac{f_{\xi|\eta}(\xi|y)p_\eta(y)}{f_\xi(x)} = \sum_y g(y) p_{\eta|\xi}(y|\xi),
$$

где условная вероятностная масса $p_{\eta|\xi}(y|x)$ определяется формулой Байеса:

$$
p_{\eta|\xi}(y|x) = \frac{f_{\xi|\eta}(x|y)p_\eta(y)}{f_\xi(x)} = \frac{f_{\xi|\eta}(x|y)p_\eta(y)}{\sum_{y'} f_{\xi|\eta}(x|y')p_\eta(y')}.
$$

Пример: обнаружение сигнала

Бинарный сигнал описывается случайной величиной η:

$$p_\eta(1) = p, \quad p_\eta(-1) = q = 1 - p.$$

На сигнал накладывается стандартный гауссовский шум $w \sim N(0,1)$, который не зависит от η. Полученный сигнал имеет вид

$$\xi = \eta + w.$$

Найти вероятность того, что $\eta = 1$ при условии, что наблюдаемое значение ξ равно x.

Решение. В данном случае $f_{\xi|\eta}(x|y)$ совпадает с плотностью случайной величины $y + w \sim N(y,1)$:

$$f_{\xi|\eta}(x|y) = \frac{1}{\sqrt{2\pi}} e^{-(x-y)^2/2}.$$

По формуле Байеса,

$$p_{\eta|\xi}(y|x) = \frac{f_{\xi|\eta}(x|y)p_\eta(y)}{\sum_{y'} f_{\xi|\eta}(x|y')p_\eta(y')}.$$

В частности,

$$p_{\eta|\xi}(1|x) = \frac{e^{-(x-1)^2/2}p}{e^{-(x-1)^2/2}p + e^{-(x+1)^2/2}q} = \frac{p}{p+qe^{-2x}}.$$

Заметим, что функция $x \mapsto p_{\eta|\xi}(1|x)$ является монотонно возрастающей и

$$\lim_{x\to-\infty} p_{\eta|\xi}(1|x) = 0, \quad \lim_{x\to\infty} p_{\eta|\xi}(1|x) = 1.$$

Случай, когда ξ дискретна, η непрерывна

Пусть ξ дискретна и имеет вероятностную массу $p_\xi(x)$, а η непрерывна и имеет плотность $f_\eta(y)$. *Введем* условную вероятностную массу $p_{\xi|\eta}(x|y)$:

$$p_{\xi|\eta}(x|y) \geq 0, \quad \sum_x p_{\xi|\eta}(x|y) = 1.$$

Другими словами, будем считать, что условное распределение ξ относительно η задается формулой

$$\mathsf{E}(g(\xi,\eta)|\eta) = \sum_x g(x,\eta) p_{\xi|\eta}(x|\eta).$$

Найдем совместное распределение:

$$\mathsf{E} g(\xi,\eta) = \mathsf{E}\mathsf{E}(g(\xi,\eta)|\eta) = \mathsf{E} \sum_x g(x,\eta) p_{\xi|\eta}(x|\eta)$$
$$= \int_{\mathbb{R}} \sum_x g(x,y) p_{\xi|\eta}(x|y) f_\eta(y)\, dy.$$

Частным случаем является формула полного математического ожидания:

$$\mathsf{E}g(\xi) = \sum_x g(x) \int_{\mathbb{R}} p_{\xi|\eta}(x|y) f_\eta(y)\, dy$$

из которой вытекает формула полной вероятности

$$p_\xi(x) = \int_{\mathbb{R}} p_{\xi|\eta}(x|y) f_\eta(y)\, dy.$$

Найдем $\mathsf{E}(g(\eta)|\xi) = r(\xi)$. По определению,

$$\mathsf{E}(g(\eta)h(\xi)) = \mathsf{E}(r(\xi)h(\xi))$$

для любой ограниченной борелевской функции h.

$$\begin{aligned}
\mathsf{E}(g(\eta)h(\xi)) &= \int_{\mathbb{R}} \sum_x g(y)h(x)p_{\xi|\eta}(x|y)f_\eta(y)\,dy \\
&= \sum_x \left(\int_{\mathbb{R}} g(y)p_{\xi|\eta}(x|y)f_\eta(y)\,dy \right) h(x), \\
\mathsf{E}(r(\xi)h(\xi)) &= \sum_x r(x)h(x)p_\xi(x).
\end{aligned}$$

Следовательно,

$$r(x) = \int_{\mathbb{R}} g(y)\frac{p_{\xi|\eta}(x|y)f_\eta(y)}{p_\xi(x)}\,dy,$$

$$\mathsf{E}(g(\eta)|\xi) = r(\xi) = \int_{\mathbb{R}} g(y)\frac{p_{\xi|\eta}(\xi|y)f_\eta(y)}{p_\xi(\xi)}\,dy = \int_{\mathbb{R}} g(y)f_{\eta|\xi}(y|\xi)\,dy,$$

где условная плотность $f_{\eta|\xi}(y|x)$ определяется формулой Байеса:

$$f_{\eta|\xi}(y|x) = \frac{p_{\xi|\eta}(x|y)f_\eta(y)}{p_\xi(x)} = \frac{p_{\xi|\eta}(x|y)f_\eta(y)}{\int_{\mathbb{R}} p_{\xi|\eta}(x|y')f_\eta(y')\,dy'}.$$

Пример: оценка несимметричности монеты

Результат подбрасывания монеты описывается бернуллиевской случайной величиной ξ с вероятностной массой

$$p_\xi(1) = \eta, \quad p_\xi(0) = 1 - \eta,$$

1 — орел, 0 — решка. Степень несимметричности монеты неизвестна. Пусть $\eta \sim U(0,1)$ — априорная (случайная!) вероятность выпадения орла. Монета подброшена 1 раз и выпал орел. Найти соответствующую условную плотность $f_{\eta|\xi}(y|x)$, $x = 1$ вероятности выпадения орла.

Решение. По формуле Байеса

$$f_{\eta|\xi}(y|x) = \frac{p_{\xi|\eta}(x|y) f_\eta(y)}{\int_{y'} p_{\xi|\eta}(x|y') f_\eta(y')\, dy'}.$$

В данном случае

$$p_{\xi|\eta}(1|y) = y, \quad p_{\xi|\eta}(0|y) = 1 - y, \quad f_\eta(y) = \begin{cases} 1, & y \in [0,1] \\ 0, & y \notin [0,1] \end{cases}.$$

Таким образом,

$$f_{\eta|\xi}(y|1) = \frac{p_{\xi|\eta}(1|y) f_\eta(y)}{\int_{y'} p_{\xi|\eta}(1|y') f_\eta(y')\, dy'} = \frac{y}{\int_0^1 y'\, dy'} = 2y, \quad y \in [0, 1].$$

ИЗБРАННЫЕ ВОПРОСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

II. Математическая статистика

Рохлин Д.Б.

ИММиКН, ЮФУ, 2022

References

- Боровков А.А. Математическая статистика. Лань: СПб, 2010.
- Чернова Н.И. Математическая статистика: учебное пособие. РИЦ НГУ: Новосибирск, 2014.
- Коршунов Д.А., Чернова Н.И. Сборник задач и упражнений по математической статистике: учебное пособие. Изд-во Института математики: Новосибирск, 2004.
- Wasserman L. All of statistics: a concise course in statistical inference. Springer: New York, 2004.
- Sahoo P. Probability and mathematical statistics. University of Louisville, USA, 2013.
- Bertsekas D.P., Tsitsiklis J.N. Introduction to Probability. Athena Scientific: Belmont, MIT, 2008.

  - Rigollet P. (Massachusetts Institute of Technology) Statistics for Applications, 2016, youtube.

Статистика: искусство извлекать информацию из данных и делать выводы на ее основе.

- Нидерланды, 10 век, строительство плотин и дамб на основе данных об уровнях воды. Сооружение должно быть достаточно высоким для большинства наводнений, но не должно быть слишком дорогим (высоким).
- Страхование автомобиля. Определить “справедливую” на основе данных о водительском стаже, возрасте, информации об автомобиле.
- Клинические испытания лекарства. Препарат протестирован на 100 пациентах; 56 из них были излечены, а в 44 случаях улучшения не было. Эффективен ли препарат?

Общие черты: в данных присутствует случайность.

- Использование средних характеристик (справедливая премия)
- Количественная характеристика шансов наступления определенных событий (превышения уровня)
- Значимость (достаточно ли 56 из 100 для вывода об эффективности)

Для описания данных используется некоторая вероятностная модель. Однако, эта модель определена лишь с точностью до некоторых параметров, которые нужно оценить на основе имеющихся данных.

Пример: Задача об опросе, нормальная аппроксимация

Перед голосованием (референдумом) проводится опрос с целью выяснить, какова доля p населения, которая будет голосовать из "за". Наудачу выбранный человек будет голосовать "за" с вероятностью p. Последовательно опрашивая таких людей, получаем последовательность независимых бернуллиевских случайных величин X_i: $\mathsf{P}(X_i = 1) = p$ (за), $\mathsf{P}(X_i = 1) = 1 - p$ (против).

Хотим найти для p *доверительный интервал* — случайный интервал, который накроет p с вероятностью $1-\alpha$. В качестве оценки p возьмем частоту появления ответа "за":

$$p = \mathsf{E}X_1 \approx \overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- *Слабый закон больших чисел.* Пусть $(\xi_i)_{i=1}^{\infty}$ — независимые одинаково распределенные случайные величины, $\mathsf{E}\xi_i = \mu$, $\mathrm{Var}(\xi_i) < \infty$. Тогда
$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow{\mathsf{P}} \mu, \quad n \to \infty.$$
- *Центральная теорема*. Пусть $(\xi_k)_{k=0}^{\infty}$ — независимые одинаково распределенные случайные величины, $\mathsf{E}\xi_n = 0$, $\mathrm{Var}(\xi_n) = 1$. Тогда
$$\frac{1}{\sqrt{n}}(\xi_1 + \cdots + \xi_n) \xrightarrow{d} Z \sim N(0,1).$$

X_i независимы, одинаково распределены, и $\mathsf{E}X_i = p$, $\mathrm{Var}(X_i) = p(1-p)$. Поскольку

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{p(1-p)}{n},$$

то, согласно центральной предельной теореме,

$$\frac{\overline{X}-p}{\sqrt{\mathrm{Var}(\overline{X})}} = \frac{\sum_{i=1}^{n}(X_i-p)/n}{\sqrt{p(1-p)/n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i-p}{\sqrt{p(1-p)}} \xrightarrow{d} Z \sim N(0,1).$$

Другими словами,

$$\sqrt{n}\frac{\overline{X}-p}{\sqrt{p(1-p)}} \xrightarrow{d} Z \sim N(0,1).$$

Далее мы покажем, что

$$\sqrt{n}\frac{\overline{X}-p}{\sqrt{\overline{X}(1-\overline{X})}} \xrightarrow{d} Z \sim N(0,1).$$

Таким образом,

$$\mathsf{P}(|\overline{X}-p|>\gamma)=\mathsf{P}\left(\sqrt{n}\frac{|\overline{X}-p|}{\sqrt{\overline{X}(1-\overline{X})}}>\gamma\frac{\sqrt{n}}{\sqrt{\overline{X}(1-\overline{X})}}\right)$$

$$\approx \mathsf{P}\left(|Z|>\gamma\frac{\sqrt{n}}{\sqrt{\overline{X}(1-\overline{X})}}\right)=\mathsf{P}\left(|Z|>z_{\alpha/2}\right)=\alpha.$$

Доверительный интервал для p:

$$p\in(\overline{X}-\gamma,\overline{X}+\gamma),\quad \gamma=z_{\alpha/2}\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}.$$

Уравнение для $z_{\alpha/2}$ можно переписать так:

$$\alpha=2\mathsf{P}(Z>z_{\alpha/2})=2(1-\mathsf{P}(Z\le z_{\alpha/2}))\Longrightarrow \mathsf{P}(Z\le z_{\alpha/2})=1-\alpha/2.$$

Например, для $\alpha=0.05$ имеем $z_{0.025}=1.96$ по таблице для стандартного нормального распределения.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983

Допустим, что длина интервала не превосходит 0.06 (т.е. $|\overline{X} - p| \leq 0.03$), и мы хотим, чтобы он накрывал p с вероятностью 0.95. Сколько людей тогда требуется опросить?
Для $\alpha = 0.05$, $\gamma \leq 0.03$ имеем

$$\gamma = 1.96 \frac{\sqrt{\overline{X}(1 - \overline{X})}}{\sqrt{n}} \leq 0.03.$$

Данное неравенство будет выполнено, если заменить $\overline{X}(1-\overline{X})$ верхней оценкой $1/4$:

$$\sqrt{n} \geq \frac{1.96}{0.03}\frac{1}{2} = 32.6(6) \Longrightarrow n \approx 1067.$$

Для $\gamma = 0.01$ мы получили бы

$$\sqrt{n} \geq \frac{1.96}{0.02} = 98 \Longrightarrow n \approx 9604.$$

Статистические модели, выборочный метод

Имеется семейство вероятностных вероятностных мер $\mathcal{P}$. Данные

$$x_1, \ldots, x_n$$

являются реализациями (значениями) независимых случайных величин

$$X_1, \ldots, X_n \sim \mathsf{P},$$

где $\mathsf{P} \in \mathcal{P}$ — неизвестная мера. Последовательность $X = (X_i)$ называется *выборкой*. Цель математической статистики состоит в том, чтобы сделать выводы относительно P на основе полученных значений $x_1, \ldots, x_n$.

- Любая (борелевская) функция $\widehat{\theta}_n(X_1, \ldots, X_n)$, зависящая от данных, называется *статистикой*.

Например,

$$\widehat{\theta}_n(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Заметим, что функция $\widehat{\theta}_n$ зависит от n.
Эмпирическим распределением, соответствующим выборке X, называется случайная вероятностная мера

$$\widehat{\mathsf{P}}(B) = \frac{1}{n} \sum_{i=1}^{n} I_B(X_i), \quad B \in \mathscr{B}(\mathbb{R}).$$

Таким образом, $\widehat{\mathsf{P}}(B)$ совпадает с долей элементов выборки, попавших в B.

Из закона больших чисел вытекает, что

$$\widehat{\mathsf{P}}(B) \xrightarrow{\mathsf{P}} \mathsf{E} I_B(X_1) = \mathsf{P}(X_1 \in B).$$

Более общим образом, пусть $g : \mathbb{R} \mapsto \mathbb{R}$ — борелевская функция. Тогда по закону больших чисел

$$\int_{\mathbb{R}} g(x)\,\widehat{\mathsf{P}}(dx) = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{\mathsf{P}} \mathsf{E}_{\mathsf{P}} g(X_1) = \int_{\mathbb{R}} g(x)\,\mathsf{P}(dx).$$

Таким образом, знание выборки позволяет аппроксимировать (оценивать) числовые характеристики случайных величин на основе выборочных характеристик. Например,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m \xrightarrow{\mathsf{P}} \mathsf{E}_{\mathsf{P}} X^m.$$

Эмпирической функцией распределения называется (случайная) функция

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}} = \widehat{\mathsf{P}}((-\infty, x]).$$

Пусть F — функция распределения X_1. Тогда

$$\mathsf{E}\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathsf{P}(X_i \le x) = \mathsf{P}(X_1 \le x) = F(x),$$

$$\mathrm{Var}(\widehat{F}_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(I_{\{X_i \le x\}})\right) = \frac{1}{n}F(x)(1 - F(x)),$$

так как $I_{\{X_i \le x\}} \sim \mathrm{Ber}(F(x))$.

Упорядочим случайные величины $X_1, \dots, X_n$ по возрастанию:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

$$X_{(1)} = \min_{1 \leq i \leq n} X_i, \quad X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

Случайная величина $X_{(j)}$ называется *j-й порядковой статистикой*. *Выборочной медианой* называется

$$\widehat{\zeta} = \begin{cases} X_{(m)}, & n = 2m-1, \\ (X_{(m)} + X_{(m+1)})/2, & n = 2m. \end{cases}$$

Найдем функцию распределения $X_{(k)}$. Поскольку

$$\{X_{(k)} \leq x\} = \bigcup_{j=k}^{n} \left\{ \sum_{s=1}^{n} I_{\{X_s \leq x\}} = j \right\}$$

и $I_{\{X_s \leq x\}} \sim \mathrm{Ber}(F(x))$, то

$$\mathsf{P}(X_{(k)} \leq x\} = \sum_{j=k}^{n} \mathsf{P}\left(\sum_{s=1}^{n} I_{\{X_s \leq x\}} = j \right) = \sum_{j=k}^{n} C_n^j F(x)^j (1 - F(x)^{n-j}.$$

Разобьем множество значений X_1 на интервалы A_j, $j = 1, \ldots, m$. Пусть l_j — длина A_j и

$$\nu_j = \sum_{i=1}^{n} I_{A_j}(X_i)$$

— число элементов выборки попавших на интервал A_j.
Гистограммой называется (случайная) функция

$$\widehat{f}(x) = \frac{\nu_j}{n l_j}, \quad x \in A_j.$$

Заметим, что

$$\int \widehat{f}(x)\,dx = \sum_{j=1}^{m} \frac{\nu_j}{n} = 1,$$

$$\mathsf{E}\widehat{f}(x) = \frac{\mathsf{P}(X_1 \in A_j)}{l_j}, \quad x \in A_j.$$

Если X_1 имеет плотность f, то

$$\mathsf{E}\widehat{f}(x) = \frac{1}{l_j}\int_{A_j} f(y)\,dy, \quad x \in A_j.$$

Для случайной величины с функцией распределения F *квантилью* (quantile) порядка (или уровня) p называется число

$$\zeta_p = F^{-1}(p) := \inf\{x : F(x) \geq p\}.$$

(Quantile function = percent point function = inverse cumulative distribution function.)
Квантиль порядка 0.5 называется медианой. Квантили порядков, кратных 0.01 — процентилями, кратных 0.25 — квартилями.

Выборочная квантиль порядка p по определению равна

$$\widehat{\zeta}_p = \inf\{x : \widehat{F}_n(x) \geq p\}.$$

В каждой точке $X_{(k)}$ сосредоточена масса $1/n$ эмпирического распределения. Для кратных точек массы складываются. Из анализа неравенств

$$\frac{j}{n} \leq p \leq \frac{j+1}{n}$$

вытекает, что если np является целым, то $\widehat{\zeta}_p = X_{(np)}$. Иначе $\widehat{\zeta}_p = X_{([np]+1)}$.
Например,

$$\widehat{\zeta}_{1/2} = \begin{cases} X_{n/2}, & n \text{ четно}, \\ X_{(n+1)/2}, & n \text{ нечетно}. \end{cases}$$

При нечетном n выборочная квантиль $\widehat{\zeta}_{1/2}$ совпадает с выборочной медианой.

Сходимость случайных величин

Напомним, что последовательность случайных величин ξ_n называется сходящейся к ξ

- *по вероятности*: $\xi_n \xrightarrow{\mathsf{P}} \xi$, если

$$\lim_{n\to\infty} \mathsf{P}\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\} = 0;$$

- *в среднем порядка* r (или в L^r): $\xi_n \xrightarrow{L^r} \xi$, $(0 < r < \infty)$, если

$$\lim_{n\to\infty} \mathsf{E}(|\xi_n - \xi|^r) = 0;$$

- *по распределению*: $\xi_n \xrightarrow{d} \xi$, если $\lim_{n\to\infty} \mathsf{E}g(\xi_n) = \mathsf{E}g(\xi)$ для любой ограниченной непрерывной функции $g : \mathbb{R}^d \mapsto \mathbb{R}$.

Соотношения между этими видами сходимости:

- сходимость в среднем порядка $r \Longrightarrow$ сходимость по вероятности $\Longrightarrow$ сходимость по распределению.

Дополнительные свойства

- $\xi_n \xrightarrow{d} \xi$, если и только если
$$F_{\xi_n}(x) \to F_\xi(x)$$
во всех точках непрерывности F_ξ.
- Если $\xi_n \xrightarrow{d} c$, то $\xi_n \xrightarrow{\mathsf{P}} c$, где c — константа.
- Если $\xi_n \xrightarrow{\mathsf{P}} \xi$, $\eta_n \xrightarrow{\mathsf{P}} \eta$, то
$$\xi_n + \eta_n \xrightarrow{\mathsf{P}} \xi + \eta, \quad \xi_n \eta_n \xrightarrow{\mathsf{P}} \xi\eta.$$
- Если $\xi_n \xrightarrow{L^r} \xi$, $\eta_n \xrightarrow{L^r} \eta$, то $\xi_n + \eta_n \xrightarrow{L^r} \xi + \eta$.

Дополнительные свойства

- (Теорема Слуцкого). Если $\xi_n \xrightarrow{d} c$, $\eta_n \xrightarrow{d} \eta$, то
$$\xi_n + \eta_n \xrightarrow{d} c + \eta, \quad \xi_n \eta_n \xrightarrow{d} c\eta.$$
- (Теорема о непрерывном преобразовании). Пусть g — непрерывна во всех точках множества C такого, что $\mathsf{P}(\xi \in C) = 1$. Тогда
$$\text{если} \quad \xi_n \xrightarrow{\mathsf{P}} \xi, \quad \text{то} \quad g(\xi_n) \xrightarrow{\mathsf{P}} g(\xi),$$
$$\text{если} \quad \xi_n \xrightarrow{d} \xi, \quad \text{то} \quad g(\xi_n) \xrightarrow{d} g(\xi).$$

Из сходимости $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$ в общем случае не следует, что $(\xi_n, \eta_n) \xrightarrow{d} (\xi, \eta)$. Например, если $\xi_n, \eta_n, \xi, \eta \sim N(0, 1)$, $\xi_n = \xi = -\eta_n = \eta$, то

$$\xi_n + \eta_n = 0 \overset{d}{\not\to} \xi + \eta = 2\xi \sim N(0, 4).$$

Если $(\xi_n, \eta_n) \xrightarrow{d} (\xi, \eta)$, то это было бы невозможно по теореме о непрерывном преобразовании, примененной к векторной последовательности (ξ_n, η_n).

Пример. Пусть $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ и $\mathsf{E}X_1 = \mu$, $\mathrm{Var}(X_1) = \sigma^2$. Тогда

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} Z \sim N(0, \sigma^2),$$

$$\overline{X} - \mu \approx N(0, \sigma^2/n).$$

Пример. Пусть $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. Тогда

$$\sqrt{n}(\overline{X^k} - \mathsf{E}X_1^k) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i^k - \mathsf{E}X_1^k) \xrightarrow{d} Z \sim N(0, \mathrm{Var}(X_1^k)).$$

Пример. Пусть $\mathsf{E}X_1 = 0$, $\mathrm{Var}(X_1) = \sigma^2$ и

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)$$

$$= \frac{1}{n}\sum_{i=1}^{N} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{N} X_i^2 - \overline{X}^2.$$

Тогда

$$\sqrt{n}(\widehat{\sigma}^2 - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - \sigma^2) - \overline{X}^2\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i^2 - \sigma^2) - \sqrt{n}\cdot\overline{X}^2 \xrightarrow{d} Z \sim N(0, \mathrm{Var}(X_1^2))$$

по теореме Слуцкого, так как

$$\sqrt{n}\cdot\overline{X}^2 = \overline{X}(\sqrt{n}\cdot\overline{X}) \xrightarrow{d} 0.$$

Действительно,

$$\overline{X} \xrightarrow{\mathsf{P}} \mathsf{E}X_1 = 0, \quad \sqrt{n}\cdot\overline{X} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i \xrightarrow{d} N(0, \sigma^2).$$

Итак, $\sqrt{n}(\widehat{\sigma}^2 - \sigma^2) \approx N(0, \mathrm{Var}(X_1^2))$.

Пример. Пусть X_1 имеет функцию распределения F, тогда

$$\sqrt{n}(\widehat{F}_n(x)-F(x)) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(I_{\{X_i\le x\}}-F(x)) \xrightarrow{d} N(0, F(x)(1-F(x)),$$

так как $\mathrm{Var}(I_{\{X_i\le x\}}) = F(x)(1-F(x))$. Таким образом,

$$\widehat{F}_n(x) - F(x) \approx N\left(0, \frac{F(x)(1-F(x))}{n}\right).$$

Интегрирование по методу Монте-Карло

Пусть $X_i \sim U(0,1)$. Тогда

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{\mathsf{P}} \mathsf{E} g(X_1) = \int_0^1 g(x)\,dx,$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_0^1 g(x)\,dx\right) \xrightarrow{d} N(0, \mathrm{Var}(g(X_1))).$$

Пусть X_i имеют некоторую плотность f на $\mathbb{R}$. Тогда

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{\mathsf{P}} \mathsf{E} g(X_1) = \int_{\mathbb{R}} g(x) f(x)\,dx.$$

Таким образом, если компьютер может генерировать реализацию выборки с заданным распределением, то получаем очень простой метод приближенного вычисления интегралов.

Покажем, что фактически достаточно иметь генератор для равномерного распределения $\zeta \sim U(0,1)$. Пусть $F_\xi(x) = \mathsf{P}(\xi \le x)$ — функция распределения ξ. Положим

$$F_\xi^{-1}(u) = \min\{x : F_\xi(x) \ge u\}, \quad u \in (0,1)$$

(минимум достигается, так как F_ξ непрерывна справа).

- Метод обратного преобразования (преобразования Смирнова):

$$F_\xi^{-1}(\zeta) \sim \mathsf{P}_\xi.$$

Доказательство. Если F_ξ непрерывна и является строго возрастающей, то уравнение

$$F_\xi(x) = u$$

имеет единственное решение при любом $u \in (0,1)$. В этом случае F_ξ^{-1} совпадает с обычной обратной функцией и

$$\mathsf{P}(F_\xi^{-1}(\zeta) \le x) = \mathsf{P}(\zeta \le F_\xi(x)) = F_\xi(x).$$

В общем случае из определения вытекает, что

$$F_\xi(x) > u \Longrightarrow x \geq F_\xi^{-1}(u),$$

$$F_\xi(F_\xi^{-1}(u)) \geq u, \text{ так как } F_\xi \text{ непрерывна справа.}$$

Отсюда вытекает, что

$$\{\zeta < F_\xi(x)\} \subseteq \{F_\xi^{-1}(\zeta) \leq x\} \subseteq \{F_\xi(F_\xi^{-1}(\zeta)) \leq F_\xi(x)\} \subseteq \{\zeta \leq F_\xi(x)\},$$

Следовательно, $\mathsf{P}(\zeta < F_\xi(x)) = \mathsf{P}(\zeta \leq F_\xi(x)) = F_\xi(x)$. Таким образом,

$$F_\xi(x) = \mathsf{P}(F_\xi^{-1}(\zeta) \leq x). \qquad \square$$

Пример. $\xi \sim \mathrm{Ber}(p)$,

$$F_\xi(x) = \begin{cases} 0, & x < 0, \\ p, & x \in [0,1), \\ 1, & x \geq 1, \end{cases} \qquad F_\xi^{-1}(u) = \begin{cases} 0, & u \in (0,p], \\ 1, & u \in (p,1). \end{cases}$$

$$\mathsf{P}(F_\xi^{-1}(\zeta) = 0) = \mathsf{P}(\zeta \in (0,p]) = p, \quad \mathsf{P}(F_\xi^{-1}(\zeta) = 1) = 1 - p.$$

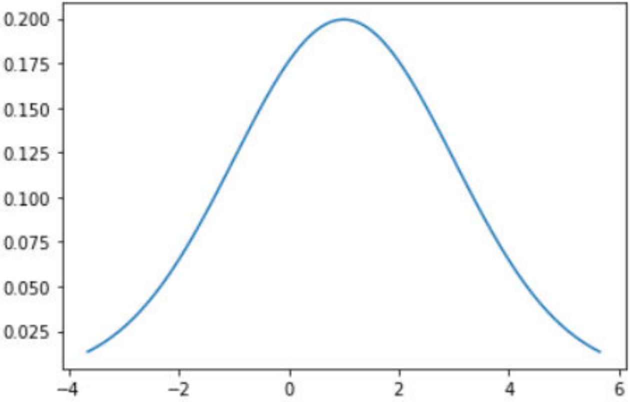
In [10]:
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
rv = stats.norm(1,2)
```

In [11]:
```python
print('mean:', rv.mean())
print('variance:', rv.var())
print('standard deviation:', rv.std())
print('median:', rv.median())
print('0.01 percentile:', rv.ppf(0.01))
print('0.99 percentile:', rv.ppf(0.99))
```

```
mean: 1.0
variance: 4.0
standard deviation: 2.0
median: 1.0
0.01 percentile: -3.6526957480816815
0.99 percentile: 5.6526957480816815
```

In [12]:
```python
x=np.linspace(rv.ppf(0.01),rv.ppf(0.99),100)
plt.plot(x,rv.pdf(x))
```

Out[12]:
```
[<matplotlib.lines.Line2D at 0x1d0295ca128>]
```



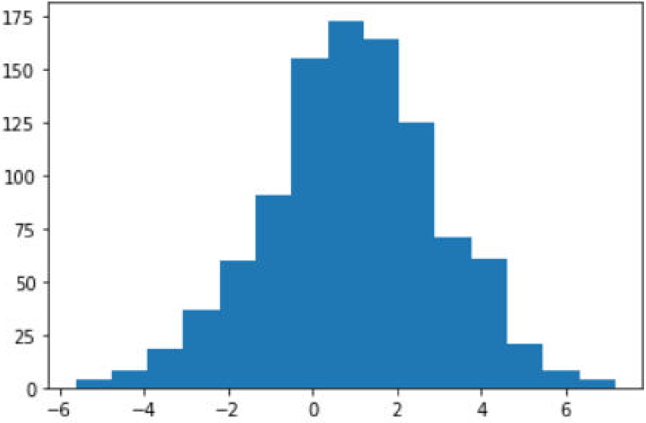
In [14]:
```
X_samples=rv.rvs(1000)
print(X_samples[:4])
```

```
[-1.26321354 -1.12273014  0.94553942  1.94721138]
```

In [6]:
```
print('sample mean:', X_samples.mean())
print('sample variance:', X_samples.var())
print('sample standard deviation:', X_samples.std())
print('sample median:', np.median(X_samples))
print('0.01 percentile:', np.percentile(X_samples,1.))
print('0.99 percentile:', np.percentile(X_samples,99.))
```

```
sample mean: 1.0234708179992147
sample variance: 4.20289766675017
sample standard deviation: 2.0500969895958994
sample median: 1.0317126998060728
0.01 percentile: -3.7197014673708044
0.99 percentile: 5.67401027625593
```

In [15]:
```
plt.hist(X_samples, bins=15)
plt.show()
```



Приближенное вычисление интеграла $\int_0^1 x^2\,dx = 1/3$

по методу Монте-Карло:

In [17]:
```
X_samples=stats.uniform.rvs(size=1000)
np.mean(X_samples**2)
```

Out[17]: 0.3434236797277684

Приближенное вычисление интеграла $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx = 1$

по методу Монте-Карло:

In [18]:
```
X_samples=stats.norm.rvs(size=1000)
np.mean(X_samples**2)
```

Out[18]: 0.975283463854397

Точечные оценки параметров

Параметрическая модель имеет вид $\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$. Обычно рассматриваются семейства плотностей или вероятностных масс

$$\{x \mapsto f(x;\theta) : \theta \in \Theta\},$$

которые будем обозначать той же буквой $\mathcal{P}$. Например, семейство плотностей нормальных распределений:

$$f(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad \theta = (\mu,\sigma) \in \mathbb{R} \times (0,\infty).$$

На основе имеющихся данных:

$$\theta \approx \widehat{\theta}_n = \widehat{\theta}_n(X_1,\ldots,X_n)$$

требуется построить статистику, которая является оценкой неизвестного параметра θ.

Оценка $\widehat{\theta}_n$ называется

- несмещенной, если $\mathsf{E}\widehat{\theta}_n = \theta$,
- состоятельной, если $\widehat{\theta}_n \xrightarrow{\mathsf{P}} \theta$,
- асимптотически нормальной с коэффициентом $\sigma^2(\theta)$, если

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \sigma^2(\theta)).$$

Покажем, что асимптотически нормальная оценка является состоятельной. Положим

$$\xi_n = \frac{1}{\sqrt{n}}, \quad \eta_n = \sqrt{n}(\widehat{\theta}_n - \theta).$$

Поскольку $\xi_n \xrightarrow{d} 0$, $\eta_n \xrightarrow{d} Z$, то

$$\widehat{\theta}_n - \theta = \xi_n \eta_n \xrightarrow{d} 0 \cdot Z = 0 \Longrightarrow \widehat{\theta}_n \xrightarrow{d} \theta$$

по теореме Слуцкого. Но тогда $\widehat{\theta}_n \xrightarrow{\mathsf{P}} \theta$, так как θ является константой.

Метод моментов

Пусть имеется k неизвестных параметров $\theta = (\theta_1, \ldots, \theta_k)$. Оценим первые k моментов по закону больших чисел:

$$\mu_j(\theta) := \mathsf{E}_\theta(X_1^j) \approx \widehat{\mu}_j := \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

и определим $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ как решение системы уравнений

$$\mu_j(\theta) = \widehat{\mu}_j, \quad j = 1, \ldots k.$$

Пример. $X_i \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma)$,

$$\mu_1(\theta) = \mu, \quad \mu_2(\theta) = \mathsf{E}_\theta X^2 = \mathrm{Var}_\theta(X) + (\mathsf{E}_\theta X)^2 = \sigma^2 + \mu^2,$$

$$\mu = \widehat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{N} X_i,$$

$$\sigma^2 + \mu^2 = \widehat{\mu}_2 := \frac{1}{n}\sum_{i=1}^{N} X_i^2.$$

Таким образом,

$$\widehat{\mu} = \overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

поскольку

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2) \\ &= \frac{1}{n}\sum_{i=1}^{N} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{N} X_i^2 - \overline{X}^2.\end{aligned}$$

Пример. $X_i \sim U(0, \theta)$:

$$f_\theta(x) = \begin{cases} 1/\theta, & x \in (0, \theta) \\ 0, & \text{иначе}, \end{cases}$$

$$\mu_1(\theta) = \mathsf{E}_\theta X = \frac{\theta}{2} = \widehat{\mu}_1 = \overline{X},$$

$$\widehat{\theta} = 2\overline{X}.$$

Вместо моментов можно брать более общие оценочные (пробные) функции $g_1, \ldots, g_k$ такие, что система уравнений

$$\mu_j(\theta) = \frac{1}{n} \sum_{i=}^{n} g_j(X_i), \quad j = 1, \ldots, k,$$
$$\mu_j(\theta) := \mathsf{E}_\theta g_j(X_1)$$

однозначно разрешима относительно θ.

Пример. $X_i \sim \Pi(\lambda)$. Оценим параметр $\theta = \lambda e^{-\lambda}$, используя функцию

$$g(y) = I_{\{y=1\}}.$$

Имеем,

$$\mathsf{E}_\lambda g(X_1) = \mathsf{P}_\lambda(X_1 = 1) = \lambda e^{-\lambda} = \theta,$$
$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i=1\}}.$$

Таким образом, оценкой θ является

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i=1\}}.$$

Заметим, что λ не определяется однозначно из уравнения $\lambda e^{-\lambda} = \theta$.

Состоятельность оценок метода моментов

- Пусть $\widehat{\theta}_n = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^n g(X_i)\right)$, где
$$\mu(\theta) = \mathsf{E}_\theta g(X_1)$$
и функция μ^{-1} непрерывна. Тогда оценка $\widehat{\theta}_n$ состоятельна.

Доказательство. По закону больших чисел

$$\frac{1}{n}\sum_{i=1}^n g(X_i) \xrightarrow{\mathsf{P}} \mathsf{E}_\theta g(X_1) = \mu(\theta).$$

По теореме о непрерывном преобразовании,

$$\mu^{-1}\left(\frac{1}{n}\sum_{i=1}^n g(X_i)\right) \xrightarrow{\mathsf{P}} \theta. \qquad \square$$

Метод максимального правдоподобия

Функция правдоподобия

$$L_n(x;\theta) = \begin{cases} \prod_{i=1}^n f(x_i;\theta), & \text{для семейства плотностей;} \\ \prod_{i=1}^n p(x_i;\theta), & \text{для семейства вероятностных масс.} \end{cases}$$

Во втором случае $L_n(x;\theta) = \mathsf{P}_\theta(X_1 = x_1, \ldots, X_n = x_n)$.
Логарифмическая функция правдоподобия:

$$l_n(x;\theta) = \ln L_n(x;\theta).$$

Оценка максимального правдоподобия (MLE: maximum likelihood estimate):

$$\widehat{\theta} \in \arg\max_\theta L_n(x;\theta) = \arg\max_\theta l_n(x;\theta).$$

В связи с данной оптимизационной задачей можно считать, что функция правдоподобия определена с точностью до положительного множителя $C(x)$, не зависящего от θ. Используется запись

$$L_n(x;\theta) \propto C(x) L_n(x;\theta).$$

Пример. $X_i \sim \mathrm{Ber}(\theta)$,

$$p_\theta(x) = \begin{cases} \theta, & x = 1, \\ 1-\theta, & x = 0. \end{cases}$$

$$L_n(x;\theta) = \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^S(1-\theta)^{n-S}, \quad S = \sum_{i=1}^n x_i,$$

$$l_n(x;\theta) = S\ln\theta + (n-S)\ln(1-\theta),$$

$$\frac{\partial l_n(x;\theta)}{\partial\theta} = \frac{S}{\theta} - \frac{n-S}{1-\theta} = 0, \quad S = n\theta,$$

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}.$$

Пример. $X \sim N(\mu, 1)$, $\theta = \mu$,

$$f_\theta(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2},$$

$$L_n(x;\mu) \propto \prod_{i=1}^{n} e^{-(x_i-\mu)^2/2} \propto e^{-n(\overline{x}-\mu)^2/2}, \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Действительно,

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n} x_i^2 - 2n\mu\overline{x} + n\mu^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 + n(\overline{x}-\mu)^2.$$

Следовательно (пренебрегая слагаемыми, не зависящими от μ),

$$l_n(x;\mu) = -n(\overline{x}-\mu)^2/2,$$

$$\widehat{\mu} = \overline{X}.$$

Пример. $X_i \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma)$,

$$L_n(x;\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu)^2/(2\sigma^2)} \propto \frac{1}{\sigma^n} \prod_{i=1}^{n} e^{-(x_i-\mu)^2/(2\sigma^2)},$$

$$l_n(x;\theta) = \ln L_n(x;\theta) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$(l_n)_\mu = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} x_i - n\mu \right) = 0,$$

$$(l_n)_\sigma = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0.$$

Отсюда,

$$\widehat{\mu} = \overline{X}, \quad \widehat{\sigma} = \left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right)^{1/2}.$$

Такие же оценки были получены методом моментов.

Пример. $X_i \sim U(0, \theta)$,

$$f(y; \theta) = \begin{cases} 1/\theta, & y \in (0, \theta), \\ 0, & \text{иначе}, \end{cases}$$

$$L_n(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \begin{cases} \frac{1}{\theta^n}, & x_i < \theta, i = 1, \dots, n, \\ 0, & \text{иначе}. \end{cases}.$$

Оценкой максимального правдоподобия будет наименьшее θ, которое превосходит все x_i:

$$\widehat{\theta} = X_{(n)} := \max_{1 \le i \le n} X_i.$$

Методом моментов была получена другая оценка:

$$\widehat{\theta} = 2\overline{X}.$$

Пример. $X_i \sim \Pi(\lambda)$,

$$f(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x \in \mathbb{Z}_+,$$

$$L_n(x;\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\frac{1}{x_i!},$$

$$l_n(x;\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda + \mathrm{Const},$$

$$\frac{\partial l_n}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0.$$

Оценка максимального правдоподобия:

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

Пример. $X_i \sim \mathrm{Exp}(\theta)$,

$$f(x;\theta) = \begin{cases} \theta e^{-\theta x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

$$L_n(x;\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta(x_1 + \ldots x_n)},$$

$$l_n(x;\theta) = n \ln \theta - \theta \sum_{i=1}^{n} x_i,$$

$$\frac{\partial l_n}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0.$$

Оценка максимального правдоподобия:

$$\widehat{\theta} = \frac{1}{\sum_{i=1}^{n} X_i / n} = \frac{1}{\overline{X}}.$$

Смещение

Смещением оценки $\widehat{\theta}$ называется

$$\mathrm{bias}(\widehat{\theta}) = \mathsf{E}_\theta \widehat{\theta} - \theta.$$

Оценка θ называется *несмещенной*, если $\mathrm{bias}(\widehat{\theta}) = 0$.

Пример. Несмещенная оценка дисперсии определяется формулой

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Доказательство.

$$\begin{aligned}
\frac{n-1}{n}S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n}X_i + \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2,
\end{aligned}$$

Пусть $\mathsf{E}X_i = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. Тогда

$$\frac{1}{n}\mathsf{E}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n}\big(\mathrm{Var}(X_i) + (\mathsf{E}X_i)^2\big) = \sigma^2 + \mu^2,$$

$$\mathsf{E}\overline{X}^2 = \mathrm{Var}(\overline{X}) + (\mathsf{E}\overline{X})^2 = \frac{1}{n}\sigma^2 + \mu^2.$$

Таким образом,

$$\frac{n-1}{n}\mathsf{E}S^2 = \sigma^2 + \mu^2 - \frac{1}{n}\sigma^2 - \mu^2 = \frac{n-1}{n}\sigma^2,$$

$$\mathsf{E}S^2 = \sigma^2. \qquad \square$$

Отметим, что оценки дисперсии нормального распределения, полученные методом моментов и методом наибольшего правдоподобия, являются смещенными:

$$\mathsf{E}\widehat{\sigma}^2 = \mathsf{E}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n-1}{n}\sigma^2.$$

Среднеквадратическая ошибка

Одной из основных мер качества статистической оценки $\widehat{\theta}_n$ среднеквадратическая ошибка (mean squared error: MSE):

$$\mathsf{MSE}(\widehat{\theta}) = \mathsf{E}_\theta(\widehat{\theta} - \theta)^2.$$

Разложение смещения–дисперсии (bias-variance decomposition):

$$\mathsf{MSE}(\widehat{\theta}) = (\mathsf{bias}(\widehat{\theta}))^2 + \mathrm{Var}_\theta(\widehat{\theta}).$$

Доказательство. Пусть $\overline{\theta} = \mathsf{E}_\theta\widehat{\theta}$. Тогда

$$\begin{aligned}
\mathrm{MSE}(\widehat{\theta}) &= \mathsf{E}_\theta(\widehat{\theta} - \theta)^2 = \mathsf{E}_\theta(\widehat{\theta} - \overline{\theta} + \overline{\theta} - \theta)^2 \\
&= \mathsf{E}_\theta(\widehat{\theta} - \overline{\theta})^2 + 2(\overline{\theta} - \theta)\mathsf{E}_\theta(\widehat{\theta} - \overline{\theta}) + \mathsf{E}_\theta(\overline{\theta} - \theta)^2 \\
&= \mathsf{E}_\theta(\widehat{\theta} - \overline{\theta})^2 + (\overline{\theta} - \theta)^2 = \mathrm{Var}_\theta(\widehat{\theta}) + (\mathsf{bias}(\widehat{\theta}))^2.
\end{aligned}$$

- X_1 — несмещенная оценка $\mu = \mathsf{E}X_1$ с большой дисперсией,
- $\widehat{\sigma}^2$ — оценка σ^2 с малым смещением и малой дисперсией (при больших n).

Сравнение оценок в среднеквадратическом

- Оценка $\widehat{\theta}$ лучше оценки $\widetilde{\theta}$ в среднеквадратическом, если

$$\mathsf{E}_\theta(\widehat{\theta}-\theta)^2 \le \mathsf{E}_\theta(\widetilde{\theta}-\theta)^2, \quad \theta \in \Theta.$$

В классе всех возможных оценок наилучшей в среднеквадратическом обычно не существует. Действительно, пусть $\widehat{\theta}$ — наилучшая оценка. Положим $\widetilde{\theta}=\theta'$, где θ' — любая фиксированная точка из Θ. Тогда

$$\mathsf{E}_{\theta'}(\widehat{\theta}-\theta')^2 \le \mathsf{E}_{\theta'}(\widetilde{\theta}-\theta')^2 = 0.$$

Следовательно, $\mathsf{E}_{\theta'}(\widehat{\theta}-\theta')^2 = 0$ для всех $\theta' \in \Theta$. Это возможно только в вырожденном случае, когда значение выборки однозначно определяется значением параметра θ.

Пример. $X_i \in U(\theta, \theta+1)$, $\theta \in \mathbb{Z}$; $\widehat{\theta} = [X_1]$.

Эффективные оценки

Рассмотрим класс K_0 несмещенных оценок $\widehat{\theta}$:

$$\mathsf{E}_\theta \widehat{\theta} = \theta, \quad \widehat{\theta} \in K_0.$$

Оценка $\theta^* \in K_0$ называется *эффективной*, если она является наилучшей в среднеквадратическом в K_0:

$$\mathsf{E}_\theta (\theta^* - \theta)^2 \leq \mathsf{E}_\theta (\widetilde{\theta} - \theta)^2, \quad \theta \in \Theta$$

для любой $\widetilde{\theta} \in K_0$.
Сравнение несмещенных оценок в среднеквадратическом сводится к сравнению их дисперсий, так как

$$\mathsf{E}_\theta (\widehat{\theta} - \theta)^2 = \mathsf{E}_\theta (\widehat{\theta} - \mathsf{E}_\theta \widehat{\theta})^2 = \mathrm{Var}_\theta (\widehat{\theta}), \quad \widehat{\theta} \in K_0.$$

Таким образом, эффективная оценка — это несмещенная оценка с наименьшей (при всех θ) дисперсией.

Единственность эффективной оценки

- Любые две эффективных оценки совпадают п.н. относительно любой P_θ, $\theta \in \Theta$.

Пусть θ^*, $\widehat{\theta}$ — эффективные оценки. Для любого $\lambda \in \mathbb{R}$ имеем

$$\theta^* + \lambda(\widehat{\theta} - \theta^*) \in K_0.$$

Следовательно,

$$\mathrm{Var}_\theta(\theta^* + \lambda(\widehat{\theta} - \theta^*)) - \mathrm{Var}(\theta^*) = \lambda^2 \mathrm{Var}_\theta(\widehat{\theta} - \theta^*) + 2\lambda \mathrm{Cov}_\theta(\theta^*, \widehat{\theta} - \theta^*) \geq 0.$$

В силу произвольности λ отсюда следует, что

$$\mathrm{Cov}_\theta(\theta^*, \widehat{\theta} - \theta^*) = \mathsf{E}_\theta(\theta^*(\widehat{\theta} - \theta^*)) = 0.$$

Аналогичным образом,

$$\mathsf{E}_\theta(\widehat{\theta}(\theta^* - \widehat{\theta})) = 0.$$

Комбинируя эти два равенства, заключаем, что $\mathsf{E}_\theta(\theta^* - \widehat{\theta})^2 = 0$.

Пример сравнения оценок по среднеквадратическому критерию

Пусть $X_i \sim U(0,\theta)$, $f(y;\theta) = I_{(0,\theta)}(y)/\theta$. Рассмотрим следующие оценки параметра θ:

$$\widehat{\theta}_n = 2\overline{X}, \quad \widetilde{\theta}_n = \frac{n+1}{n} X_{(n)} := \frac{n+1}{n} \max_{1 \le i \le n} X_i.$$

Первая из них была получена методом моментов, вторая отличается поправочным множителем от той, которая была получена методом максимального правдоподобия.
Ясно, что $\widehat{\theta}$ является несмещенной:

$$\mathsf{E}_\theta \widehat{\theta}_n = 2\mathsf{E}_\theta X_1 = 2\frac{\theta}{2} = \theta.$$

Найдем функцию распределения $X_{(n)}$:

$$\mathsf{P}_\theta(X_{(n)} \le y) = \prod_{i=1}^{n} \mathsf{P}_\theta(X_i \le y) = (\mathsf{P}_\theta(X_1 \le y))^n = \left(\int_0^y \frac{1}{\theta}\,dx\right)^n$$
$$= \left(\frac{y}{\theta}\right)^n, \quad y \in (0, \theta).$$

Дифференцируя данное выражение по y, находим плотность:

$$f_{X_{(n)}}(y;\theta) = n\frac{y^{n-1}}{\theta^n}, \quad y \in (0, \theta).$$

Теперь легко убедиться, что оценка $\widetilde{\theta}_n$ также является несмещенной:

$$\mathsf{E}_\theta \widetilde{\theta}_n = \frac{n+1}{n}\int_0^\theta y f_{X_{(n)}}(y;\theta)\,dy = \frac{n+1}{\theta^n}\int_0^\theta y^n\,dy = \theta.$$

Найдем дисперсии оценок $\widehat{\theta}_n$, $\widetilde{\theta}_n$:

$$\mathrm{Var}(\widehat{\theta}_n) = 4\mathrm{Var}(\overline{X}) = \frac{4}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\mathrm{Var}(X_1) = \frac{\theta^2}{3n}.$$

Здесь использована формула для дисперсии равномерного распределения: $\mathrm{Var}(X_1) = \theta^2/12$. Далее,

$$\mathsf{E}_\theta X_{(n)}^2 = \int_0^\theta y^2 f_{X_{(n)}}(y;\theta)\,dy = \frac{n}{\theta^n}\int_0^\theta y^{n+1}\,dy = \frac{n}{n+2}\theta^2$$

$$\begin{aligned}\mathrm{Var}(\widetilde{\theta}_n) &= \mathsf{E}\widetilde{\theta}_n^2 - (\mathsf{E}\widetilde{\theta}_n)^2 = \frac{(n+1)^2}{n^2}\mathsf{E}_\theta X_{(n)}^2 - \theta^2 = \left(\frac{(n+1)^2}{n(n+2)} - 1\right)\theta^2 \\ &= \frac{n^2+2n+1-n^2-2n}{n(n+2)}\theta^2 = \frac{\theta^2}{n(n+2)}.\end{aligned}$$

$\widetilde{\theta}_n$ значительно лучше $\widehat{\theta}_n$ при больших n. Однако, что для отыскания эффективных оценок попарного сравнения оценок недостаточно.

Информация Фишера

Напомним, определение логарифмической функции правдоподобия

$$l_n(\boldsymbol{X};\theta) = l_n(X_1,\ldots,X_n;\theta) = \sum_{i=1}^{n} \ln f(X_i;\theta),$$

и положим $l(y;\theta) := l_1(y;\theta) = \ln f(y;\theta)$.
Определим *информацию Фишера* следующим образом:

$$I_n(\theta) = \mathrm{Var}_\theta\left(\frac{\partial l_n(\boldsymbol{X};\theta)}{\partial\theta}\right) = \mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln f(X_i,\theta)\right)$$
$$= \sum_{i=1}^{n}\mathrm{Var}_\theta(l_\theta(X_i;\theta)) = n\mathrm{Var}_\theta(l_\theta(X_1;\theta)) = nI(\theta),$$

где $I(\theta) := I_1(\theta) = \mathrm{Var}_\theta(l_\theta(X_1;\theta))$.

Пример. $X_i \sim N(\mu, \sigma^2)$, σ известно, $\theta = \mu$. Тогда

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)},$$

$$l(x;\mu) = -\frac{(x-\mu)^2}{2\sigma^2} + \mathrm{Const},$$

$$l_\mu(x;\mu) = \frac{\partial}{\partial\mu}\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{x-\mu}{\sigma^2},$$

$$I(\mu) = \mathrm{Var}_\mu(l_\mu(X_1;\mu)) = \mathrm{Var}_\mu\left(\frac{X_1-\mu}{\sigma^2}\right) = \frac{1}{\sigma^2}.$$

Условие регулярности

В дальнейших вычислениях существенным образом используются формулы вида

$$\frac{\partial}{\partial\theta}\int_{\mathbb{R}} h(x)f(x;\theta)\,dx = \int_{\mathbb{R}} h(x)\frac{\partial}{\partial\theta}f(x;\theta)\,dx,$$

$$\frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}} f(x;\theta)\,dx = \int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2}f(x;\theta)\,dx,$$

которые мы назовем *условием регулярности*.
Пример. Условие регулярности выполнено для семейства экспоненциальных распределений $X_i \sim \mathrm{Exp}(\theta)$:

$$f_\theta(x) = \theta e^{-\theta x} I_{[0,\infty)}(x), \quad \theta > 0,$$

но не выполнено для семейства равномерных распределений $X_i \sim U(0,\theta)$:

$$f_\theta(x) = \frac{1}{\theta} I_{[0,\theta]}(x), \quad \theta > 0.$$

- Пусть выполнено условие регулярности, тогда

$$I(\theta) = -\mathsf{E}_\theta \left(l_{\theta\theta}(X_1; \theta) \right).$$

Заметим, что

$$\mathsf{E}_\theta l_\theta(X_1, \theta) = \mathsf{E}_\theta \frac{f_\theta(X_1; \theta)}{f(X_1; \theta)} = \int_{\mathbb{R}} \frac{f_\theta(y; \theta)}{f(y; \theta)} f(y; \theta)\, dy$$
$$= \int_{\mathbb{R}} f_\theta(y; \theta)\, dy = \frac{d}{d\theta} \int_{\mathbb{R}} f(y; \theta)\, dy = 0$$

Аналогичным образом,

$$\mathsf{E}_\theta \left(\frac{f_{\theta\theta}(X_1, \theta)}{f(X_1; \theta)} \right) = \int_{\mathbb{R}} \frac{f_{\theta\theta}(y; \theta)}{f(y; \theta)} f(y; \theta)\, dy = \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}} f(y; \theta)\, dy = 0.$$

Далее,

$$l_{\theta\theta}(y; \theta) = \frac{\partial}{\partial \theta} \frac{f_\theta(y; \theta)}{f(y; \theta)} = \frac{f_{\theta\theta}(y; \theta)}{f(y; \theta)} - \frac{f_\theta^2(y; \theta)}{f^2(y; \theta)}.$$

Следовательно,

$$I(\theta) = \mathrm{Var}_\theta(l_\theta(X_1,\theta)) = \mathsf{E}_\theta\left(l_\theta^2(X_1,\theta)\right) = \mathsf{E}_\theta\left(\frac{f_\theta(X_1;\theta)}{f(X_1;\theta)}\right)^2$$

$$= \mathsf{E}_\theta\frac{f_{\theta\theta}(X_1;\theta)}{f(X_1;\theta)} - \mathsf{E}_\theta(l_{\theta\theta}(X_1;\theta)) = -\mathsf{E}_\theta(l_{\theta\theta}(X_1;\theta)). \quad \square$$

Пример. $X_i \sim N(\mu,\sigma^2)$, σ известно, $\theta=\mu$. Тогда

$$l_\mu(x;\mu) = \frac{x-\mu}{\sigma^2}, \quad I(\mu) = -\mathsf{E}_\mu l_{\mu\mu}(X_1;\mu) = \frac{1}{\sigma^2}.$$

Пример. $X_i \sim \Pi(\lambda)$:

$$f(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x\in\mathbb{Z}_+,$$

$$l(x;\lambda) = -\lambda + x\ln\lambda + \mathrm{Const}, \quad l_{\lambda\lambda}(x;\lambda) = -\frac{x}{\lambda^2},$$

$$I(\lambda) = -\mathsf{E}_\lambda\left(l_{\lambda\lambda}(X_1;\lambda)\right) = \frac{\mathsf{E}_\lambda X_1}{\lambda^2} = \frac{1}{\lambda}.$$

Пример. $X_i \sim \mathrm{Ber}(\theta)$:

$$\begin{aligned}
f(x;\theta) &= \theta^x (1-\theta)^{1-x}, \quad x \in \{0,1\}, \\
l(x;\theta) &= \ln f(x;\theta) = x \ln\theta + (1-x)\ln(1-\theta) \\
l_{\theta\theta}(x;\theta) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}, \\
I(\theta) &= -\mathsf{E}_\theta\left(l_{\theta\theta}(X_1;\theta)\right) = \frac{\mathsf{E}_\theta X_1}{\theta^2} + \frac{1-\mathsf{E}_\theta X_1}{(1-\theta)^2} \\
&= \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}.
\end{aligned}$$

Неравенство Рао-Крамера

- Пусть $\widehat{\theta}_n$ — несмещенная оценка θ, выполнено условие регулярности и $I(\theta) \neq 0$. Тогда

$$\mathrm{Var}(\widehat{\theta}_n) \geq \frac{1}{nI(\theta)}.$$

Доказательство. Применим неравенство Коши-Буняковского:

$$(\mathsf{E}(\xi\eta))^2 \leq \mathsf{E}\xi^2\mathsf{E}\eta^2$$

с $\xi = (l_n)_\theta(X;\theta)$, $\eta = \widehat{\theta}_n(X;\theta) - \theta$:

$$\left[\mathsf{E}_\theta\left((l_n)_\theta(\widehat{\theta}_n - \theta)\right)\right]^2 \leq \mathsf{E}_\theta\left((l_n)^2_\theta\right)\mathsf{E}_\theta\left((\widehat{\theta}_n - \theta)^2\right)$$
$$= \mathrm{Var}((l_n)_\theta)\mathrm{Var}(\widehat{\theta}_n) = nI(\theta)\mathrm{Var}(\widehat{\theta}_n).$$

Кроме того,

$$\mathsf{E}_\theta\left((l_n)_\theta(\widehat{\theta}_n-\theta)\right)=\mathsf{E}_\theta\left((l_n)_\theta\widehat{\theta}_n\right)=\int_{\mathbb{R}^n}\frac{\partial\ln f(\boldsymbol{y};\theta)}{\partial\theta}\widehat{\theta}_n(\boldsymbol{y})f(\boldsymbol{y};\theta)\,d\boldsymbol{y},$$

где $f(\boldsymbol{y};\theta)$ — плотность совместного распределения $X_1,\ldots,X_n$ (для определенности X_i считаются одномерными). Отсюда следует, что

$$\mathsf{E}_\theta\left((l_n)_\theta(\widehat{\theta}_n-\theta)\right)=\int_{\mathbb{R}^n}f_\theta(\boldsymbol{y};\theta)\widehat{\theta}_n(\boldsymbol{y})\,d\boldsymbol{y}=\frac{d}{d\theta}\int_{\mathbb{R}^n}f(\boldsymbol{y};\theta)\widehat{\theta}_n(\boldsymbol{y})\,d\boldsymbol{y}$$
$$=\frac{d}{d\theta}\mathsf{E}_\theta\widehat{\theta}_n=\frac{d}{d\theta}\theta=1.\qquad\square$$

Оценка $\theta^*\in K_0$ называется R-эффективной, если

$$\mathrm{Var}(\theta^*)=\frac{1}{nI(\theta)}.$$

Естественно, такая оценка является эффективной.

Пример. $X_i \sim N(\mu, \sigma^2)$, σ известно, $\theta = \mu$. Тогда $I(\mu) = 1/\sigma^2$ и

$$\mathrm{Var}(\widehat{\mu}_n) \geq \frac{\sigma^2}{n}$$

для любой несмещенной оценки $\widehat{\mu}_n$ параметра μ. Оценка максимального правдоподобия $\widehat{\mu}_n = \overline{X}$ является R-эффективной:

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}.$$

Пример. $X_i \sim U(0, \theta)$, $f(y; \theta) = I_{(0,\theta)}/\theta$. Ранее было показано, что оценка

$$\widetilde{\theta}_n = \frac{n+1}{n} X_{(n)} := \frac{n+1}{n} \max_{1 \le i \le n} X_i$$

является несмещенной. Парадоксальным образом дисперсия оценки $\widetilde{\theta}_n$ (она также была вычислена выше) при больших n убывает быстрее, чем позволяет неравенство Рао-Крамера:

$$\mathrm{Var}(\widetilde{\theta}_n) = \frac{\theta^2}{n(n+2)} = O\left(\frac{1}{n^2}\right), \quad n \to \infty.$$

Причина состоит в том, что семейство равномерных распределений не удовлетворяет условию регулярности.

Асимптотическая нормальность

Напомним, что

- оценка $\widehat{\theta}_n$ параметра θ называется *состоятельной*, если

$$\widehat{\theta}_n \xrightarrow{\mathsf{P}} \theta.$$

асимптотически нормальной с коэффициентом $\sigma^2(\theta)$, если

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \sigma^2(\theta)).$$

Было доказано, что асимптотически нормальная оценка является состоятельной.

Пример. $X_i \sim U(0, \theta)$. Проверим, являются ли оценки

$$\widehat{\theta}_n = 2\overline{X}, \quad \widetilde{\theta}_n = X_{(n)} := \max_{1 \le i \le n} X_i$$

параметра θ асимптотически нормальными.

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}\left(\frac{2}{n}\sum_{i=1}^n X_i - \theta\right) = 2\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \theta/2)$$

$$= 2(\mathrm{Var}(X_1))^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \theta/2}{(\mathrm{Var}(X_i))^{1/2}} \xrightarrow{d} 2(\mathrm{Var}(X_1))^{1/2} Z,$$

где $Z \sim N(0, 1)$. Напомним, что $\mathrm{Var}(X_1) = \theta^2/12$. Следовательно,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \theta^2/3).$$

Покажем, что $\widetilde{\theta}_n$ не является асимптотически нормальной. Если $\widetilde{\theta}_n$ является *асимптотически нормальной* с коэффициентом $\sigma^2(\theta)$, то

$$\mathsf{P}_\theta(\sqrt{n}(\widetilde{\theta}_n - \theta) \le x) \to \mathsf{P}(Z \le x), \quad n \to \infty, \quad Z \sim N(0, \sigma^2(\theta))$$

для всех $x \in \mathbb{R}$. Но

$$1 = \mathsf{P}_\theta(\sqrt{n}(\widetilde{\theta}_n - \theta) \le 0) \not\to \mathsf{P}(Z \le 0) = 1/2, \quad n \to \infty.$$

Пример. Пусть $\mathsf{E}g^2(X_1) < \infty$. Доказать, что статистика $\overline{g(X)}$ является асимптотически нормальной оценкой параметра $\theta = \mathsf{E}_\theta g(X_1)$. Найти коэффициент асимптотической нормальности.
Решение.

$$\sqrt{n}(\overline{g(X)} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g(X_i) - \theta)$$

$$= \frac{\sqrt{\mathrm{Var}(g(X_1))}}{\sqrt{n}} \sum_{i=1}^{n} \frac{g(X_i) - \theta}{\sqrt{\mathrm{Var}(g(X_i))}} \xrightarrow{d} \sqrt{\mathrm{Var}(g(X_1))} Z \sim N(0, \mathrm{Var}(g(X_1))).$$

Дельта метод

Дельта метод позволяет находить предельное распределение функции от асимптотически нормальной случайной величины.

- Пусть
$$\sqrt{n}\frac{Y_n - \mu}{\sigma} \xrightarrow{d} Z \sim N(0,1),$$
g непрерывно дифференцируема в точке μ, и $g'(\mu) \neq 0$. Тогда
$$\sqrt{n}\frac{g(Y_n) - g(\mu)}{|g'(\mu)|\sigma} \xrightarrow{d} \widetilde{Z} \sim N(0,1).$$

Доказательство. По формуле Тейлора,
$$\sqrt{n}\frac{g(Y_n) - g(\mu)}{|g'(\mu)|\sigma} = \sqrt{n}\frac{Y_n - \mu}{\sigma}\frac{g'(\widetilde{\mu}_n)}{|g'(\mu)|},$$
где $\widetilde{\mu}_n$ лежит между Y_n и μ. Поскольку $Y_n \xrightarrow{\mathsf{P}} \mu$, то $\widetilde{\mu}_n \xrightarrow{\mathsf{P}} \mu$ и
$$g'(\widetilde{\mu}_n) \xrightarrow{\mathsf{P}} g'(\mu)$$
по теореме о непрерывном преобразовании. По теореме Слуцкого,
$$\sqrt{n}\frac{Y_n - \mu}{\sigma}\frac{g'(\widetilde{\mu}_n)}{|g'(\mu)|} \xrightarrow{d} \pm Z \sim N(0,1). \qquad \square$$

Пример. Пусть X_i — независимые одинаково распределенные случайные величины с $\mathsf{E}X_i = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. Согласно ЦПТ,

$$\sqrt{n}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma} = \sqrt{n}(\overline{X}_n-\mu)/\sigma \xrightarrow{d} Z \sim N(0,1).$$

Рассмотрим $W_n = g(\overline{X}_n) = e^{\overline{X}_n}$. Согласно дельта-методу,

$$\sqrt{n}\frac{W_n - e^{\mu}}{e^{\mu}\sigma} \xrightarrow{d} Z \sim N(0,1).$$

Пример. При тех же условиях и $\mu \neq 0$ найдем асимптотическое распределение $W_n = g(\overline{X}_n) = \overline{X}_n^2$:

$$\sqrt{n}\frac{\overline{X}_n^2 - \mu^2}{2\mu\sigma} \xrightarrow{d} Z \sim N(0,1).$$

Асимптотическая нормальность оценок метода моментов

- Пусть $\mu(\theta) = \mathsf{E}_\theta g(X_1)$ и

$$\widehat{\theta}_n = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right).$$

Если функция $H = \mu^{-1}$ дифференцируема и $H'(\mathsf{E}_\theta g(X_1)) \neq 0$, то $\widehat{\theta}_n$ параметра θ является асимптотически нормальной с коэффициентом

$$\sigma^2(\theta) = (H'(\mathsf{E}_\theta g(X_1)))^2 \cdot \mathrm{Var}(g(X_1)).$$

Доказательство. Было показано, что $\overline{g(X)}$ является асимптотически нормальной оценкой параметра $\mathsf{E}_\theta g(X_1)$:

$$\sqrt{n}(\overline{g(X)} - \mathsf{E}_\theta g(X_1)) \xrightarrow{d} N(0, \mathrm{Var}(g(X_1))),$$
$$\sqrt{n}\frac{\overline{g(X)} - \mathsf{E}_\theta g(X_1)}{(\mathrm{Var}(g(X_1)))^{1/2}} \xrightarrow{d} N(0, 1).$$

Согласно дельта-методу,

$$\sqrt{n}\frac{H(\overline{g(X)}) - H(\mathsf{E}_\theta g(X_1))}{|H'(\mathsf{E}_\theta g(X_1))| \cdot (\mathrm{Var}(g(X_1)))^{1/2}} \xrightarrow{d} N(0, 1).$$

Но $H(\mathsf{E}_\theta g(X_1)) = \theta$, $H(\overline{g(X)}) = \widehat{\theta}_n$. $\square$

Пример. $X_i \sim U(0, \theta)$:

$$f_\theta(x) = \begin{cases} 1/\theta, & x \in (0, \theta) \\ 0, & \text{иначе}, \end{cases}$$

$$\mu_1(\theta) = \mathsf{E}_\theta X = \frac{\theta}{2},$$

$$\mu_1(\widehat{\theta}) = \overline{X},$$

$$\widehat{\theta} = H(\overline{X}) = 2\overline{X},$$

$$\sqrt{n}(2\overline{X} - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)),$$

$$\sigma^2(\theta) = (H'(\mathsf{E}_\theta X_1))^2 \cdot \mathrm{Var}(X_1) = 4 \cdot \mathrm{Var}(X_1) = \frac{\theta^2}{3}.$$

Асимптотическая нормальность оценок метода максимального правдоподобия

- Пусть $\widehat{\theta}_n$ — оценка максимального правдоподобия и выполнено условие регулярности. Тогда

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N\left(0, \frac{1}{I(\theta)}\right).$$

Нестрого говоря, $\widehat{\theta}_n$ является асимптотически несмещенной: $\mathsf{E}(\theta_n - \theta) \approx 0$, а ее дисперсия асимптотически достигает нижней границы, указанной в неравенстве Рао-Крамера:

$$\mathrm{Var}(\widehat{\theta}_n) \approx \frac{1}{nI(\theta)}.$$

Доказательство. По формуле Тейлора,

$$0 = \frac{\partial l_n}{\partial \theta}(X; \widehat{\theta}_n) = \frac{\partial l_n}{\partial \theta}(X; \theta) + (\widehat{\theta}_n - \theta)\frac{\partial^2 l_n}{\partial \theta^2}(X; \widetilde{\theta}_n),$$

где $\widetilde{\theta}_n$ лежит между $\widehat{\theta}_n$ и θ. Отсюда следует, что

$$\sqrt{n}(\widehat{\theta}_n - \theta) = -\frac{(l_n)_\theta/\sqrt{n}}{(l_n)_{\theta\theta}/n},$$

$$\xi_n := \frac{(l_n)_\theta(X, \theta)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_\theta(X_i; \theta) \xrightarrow{d} Y \sim N(0, I(\theta))$$

в силу центральной предельной теоремы, так как

$$\mathsf{E}_\theta l_\theta(X_i; \theta) = 0, \quad \mathrm{Var}_\theta l_\theta(X_i; \theta) = I(\theta).$$

$$-\frac{(l_n)_{\theta\theta}(X,\theta)}{n} = -\frac{1}{n}\sum_{i=1}^{n} l_{\theta\theta}(X_i;\theta) \xrightarrow{\mathsf{P}} I(\theta)$$

по закону больших чисел, так как $-\mathsf{E}_\theta l_{\theta\theta}(X_i;\theta) = I(\theta)$. Естественно ожидать, что данный результат сохраняет силу, если заменить θ на $\widetilde{\theta}_n$:

$$\eta_n = -\frac{(l_n)_{\theta\theta}(X,\widetilde{\theta}_n)}{n} \xrightarrow{\mathsf{P}} I(\theta).$$

Итак,

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{\xi_n}{\eta_n}, \quad \xi_n \xrightarrow{d} Y \sim N(0, I(\theta)); \quad \eta_n \xrightarrow{\mathsf{P}} I(\theta).$$

По теореме Слуцкого,

$$\frac{\xi_n}{\eta_n} \xrightarrow{d} \frac{Y}{I(\theta)} \sim N\left(0, \frac{1}{I(\theta)}\right). \quad \square$$

Все результаты работают как для плотностей, так и для вероятностных масс.

Пример. $X_i \sim \mathrm{Ber}(\theta)$,

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0, 1\}.$$

Оценка максимального правдоподобия:

$$L_n(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^S (1-\theta)^{n-S}, \quad S = \sum_{i=1}^{n} x_i,$$

$$l_n(x;\theta) = S \ln\theta + (n-S)\ln(1-\theta),$$

$$\frac{\partial l_n(x;\theta)}{\partial\theta} = \frac{S}{\theta} - \frac{n-S}{1-\theta} = 0, \quad S = n\theta,$$

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

Информация Фишера:

$$\begin{aligned} l(x;\theta) &= \ln f(x;\theta) = x \ln\theta + (1-x)\ln(1-\theta) \\ l_{\theta\theta}(x;\theta) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}, \\ I(\theta) &= -\mathsf{E}_\theta\left(l_{\theta\theta}(X_1;\theta)\right) = \frac{\mathsf{E}_\theta X_1}{\theta^2} + \frac{1-\mathsf{E}_\theta X_1}{(1-\theta)^2} \\ &= \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}. \end{aligned}$$

Асимптотическая нормальность:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, 1/I(\theta)) = N(0, \theta(1-\theta)).$$

Заметим, что

$$\sqrt{n} I^{1/2}(\theta)(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0,1).$$

Кроме того,

$$\frac{I^{1/2}(\widehat{\theta}_n)}{I^{1/2}(\theta)} \xrightarrow{\mathsf{P}} 1$$

по теореме о непрерывном преобразовании. Следовательно,

$$\frac{\sqrt{n}}{\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)}}(\widehat{\theta}_n - \theta) = \sqrt{n} I^{1/2}(\widehat{\theta}_n)(\widehat{\theta}_n - \theta)$$

$$= \sqrt{n} I^{1/2}(\widehat{\theta})(\widehat{\theta}_n - \theta)\frac{I^{1/2}(\widehat{\theta}_n)}{I^{1/2}(\theta)} \xrightarrow{d} N(0,1)$$

по теореме Слуцкого.

Дельта метод для оценок максимального правдоподобия

Пусть $\widehat{\theta}_n$ — оценка максимального правдоподобия для θ. Тогда

$$\sqrt{n} I^{1/2}(\theta)(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0,1).$$

Согласно дельта-методу,

$$\sqrt{n} I^{1/2}(\theta) \frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\theta)|} \xrightarrow{d} N(0,1).$$

Используя теорему о непрерывном преобразовании и теорему Слуцкого, находим

$$\sqrt{n} I^{1/2}(\widehat{\theta}_n) \frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\widehat{\theta}_n)|} \xrightarrow{d} N(0,1).$$

Пример. $X_i \sim \mathrm{Ber}(\theta)$. Построить оценку для

$$\tau = g(\theta) = \frac{\theta}{1-\theta}.$$

Оценкой максимального правдоподобия является $\widehat{\theta}_n = \overline{X}_n$. Информация Фишера:

$$I(\theta) = \frac{1}{\theta(1-\theta)}.$$

Поскольку $g'(\theta) = 1/(1-\theta)^2$, то

$$\begin{aligned}\sqrt{n} I^{1/2}(\widehat{\theta}_n) \frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\widehat{\theta}_n)|} &= \frac{\sqrt{n}}{\sqrt{\widehat{\theta}_n(1-\widehat{\theta}_n)}}(1-\widehat{\theta}_n)^2(g(\widehat{\theta}_n) - \tau) \\ &= \sqrt{n}\sqrt{\frac{(1-\overline{X}_n)^3}{\overline{X}_n}}\big(g(\overline{X}_n) - \tau\big) \xrightarrow{d} N(0,1).\end{aligned}$$

Доверительные интервалы

Пусть $a_n(X_1, \ldots, X_n)$, $b_n(X_1, \ldots, X_n)$ — статистики. Случайный интервал (a_n, b_n) называется *доверительным интервалом* для θ уровня доверия $1 - \alpha$ (или $1 - \alpha$-доверительным интервалом), если

$$\mathsf{P}_\theta(\theta \in (a_n, b_n)) \geq 1 - \alpha.$$

Случайный интервал (a_n, b_n) называется *асимптотическим доверительным* интервалом уровня доверия $1 - \alpha$, если

$$\liminf_{n \to \infty} \mathsf{P}_\theta(\theta \in (a_n, b_n)) \geq 1 - \alpha.$$

Принцип построения (асимптотических) доверительных интервалов

- Найти последовательность функций $G_n(X_1, \ldots, X_n, \theta)$, сходящуюся по распределению к распределению μ, не зависящему от параметра θ. Функции G_n должны быть обратимы по θ.
- Найти числа γ_1, γ_2 такие, что

$$\mu(\gamma_1, \gamma_2) = \mathsf{P}(\eta \in (\gamma_1, \gamma_2)) = (\text{или } \geq) 1 - \alpha, \quad \eta \sim \mu.$$

- Разрешив неравенства $\gamma_1 < G_n(\boldsymbol{X}; \theta) < \gamma_2$ относительно θ, получить асимптотический доверительный интервал (a_n, b_n):

$$\lim_{n\to\infty} \mathsf{P}_\theta(\gamma_1 < G_n(\boldsymbol{X}; \theta) < \gamma_2) = \lim_{n\to\infty} \mathsf{P}_\theta(\theta \in (a_n, b_n))$$
$$= \mathsf{P}(\eta \in (\gamma_1, \gamma_2)) = 1 - \alpha$$

Для неасимптотических доверительных интервалов требуется, чтобы $G_n(X_1, \ldots, X_n, \theta) \sim \mu$ (без предельного перехода).

Пример. $X_i \sim N(\mu, \sigma^2)$, σ известно. Построить для μ точный двусторонний доверительный интервал и односторонние доверительные интервалы уровня доверия $1-\alpha$.
Решение. $\overline{X} \sim N(\mu, \sigma^2/n)$,

$$G_n(\boldsymbol{X}; \mu) = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

$$\gamma_1 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \gamma_2 \iff \mu \in \left(\overline{X} - \frac{\sigma}{\sqrt{n}}\gamma_2, \overline{X} - \frac{\sigma}{\sqrt{n}}\gamma_1 \right),$$

$$\mathsf{P}_\mu \left(\mu \in \left(\overline{X} - \frac{\sigma}{\sqrt{n}}\gamma_2, \overline{X} - \frac{\sigma}{\sqrt{n}}\gamma_1 \right) \right) = \mathsf{P}(\gamma_1 < \eta < \gamma_2) = 1 - \alpha,$$

где $\eta \sim N(0,1)$.

Положим $\gamma_1 = -\gamma$, $\gamma_2 = \gamma$, $\Phi(\gamma) = \mathsf{P}(\eta \le \gamma) = F_\eta(\gamma)$:

$$\mathsf{P}(-\gamma < \eta < \gamma) = \Phi(\gamma) - \Phi(-\gamma) = \Phi(\gamma) - (1 - \Phi(\gamma)) = 2\Phi(\gamma) - 1 = 1 - \alpha,$$

$$\Phi(\gamma) = 1 - \alpha/2.$$

Тогда $\gamma = z_{\alpha/2}$ — квантиль уровня $1 - \alpha/2$:

$$\mathsf{P}(\eta \ge z_{\alpha/2}) = \alpha/2 \iff \Phi(z_{\alpha/2}) = 1 - \alpha/2$$

и

$$\mathsf{P}_\mu\left(\mu \in \left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)\right)$$
$$= \mathsf{P}(-z_{\alpha/2} < \eta < z_{\alpha/2}) = 1 - \alpha.$$

α	0.1	0.05	0.01
$\alpha/2$	0.05	0.025	0.005
$z_{\alpha/2}$	1.65	1.96	2.58

Для построения односторонних доверительных интервалов рассмотрим

$$\mathsf{P}_\mu\left(\frac{\sqrt{n}}{\sigma}(\overline{X}-\mu)<\gamma\right)=\mathsf{P}(\eta<\gamma)=1-\alpha.$$

Отсюда $\gamma = z_\alpha$,

$$\mathsf{P}_\mu\left(\mu\in\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_\alpha,\infty\right)\right)=1-\alpha.$$

Аналогично,

$$\mathsf{P}_\mu\left(\frac{\sqrt{n}}{\sigma}(\overline{X}-\mu)>-\gamma\right)=\mathsf{P}(\eta>-\gamma)=1-\alpha,$$

$$\mathsf{P}_\mu\left(\mu\in\left(-\infty,\overline{X}+\frac{\sigma}{\sqrt{n}}z_\alpha\right)\right)=1-\alpha.$$

α	0.1	0.05	0.01
z_α	1.28	1.645	2.33

Пример. Пусть $X_i \sim N(\mu, 4)$, $i = 1, \dots, 9$ описывают передаваемый сигнал. Получены значения

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5.$$

Построить двусторонний и односторонние 95% доверительные интервалы для μ.
Решение. $\overline{x} = 81/9 = 9$, $\sigma = 2$, $\sqrt{n} = 3$, $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, $z_\alpha = 1.645$,

$$\begin{aligned}
\mu \in &\left(\overline{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) \\
&= \left(9 - 2 \cdot \frac{1.96}{3}, 9 + 2\frac{1.96}{3}\right) = (7.69, 10.31), \\
\mu > &\ \overline{x} - \frac{\sigma}{\sqrt{n}} z_\alpha = 9 - 2 \cdot \frac{1.645}{3} = 7.903, \\
\mu < &\ \overline{x} + \frac{\sigma}{\sqrt{n}} z_\alpha = 9 + 2 \cdot \frac{1.645}{3} = 10.097.
\end{aligned}$$

Гамма-распределение

$\xi \sim \Gamma(\alpha, \lambda)$:

$$f_\xi(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & \text{иначе}, \end{cases}$$

где $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$ — гамма-функция, $\alpha, \lambda > 0$.

$$\Gamma(\alpha+1) = \int_0^\infty y^\alpha e^{-y}\, dy = -\int_0^\infty y^\alpha\, de^{-y} = \alpha \int_0^\infty y^{\alpha-1} e^{-y}\, dy = \alpha\Gamma(\alpha).$$

В частности, $\Gamma(n+1) = n!$, $n \in \mathbb{Z}_+$. Отметим также, что $\Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$.

$$\mathsf{E}\xi = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x}\, dx = \frac{1}{\lambda} \int_0^\infty \frac{1}{\Gamma(\alpha)} x^\alpha e^{-x}\, dx = \frac{1}{\lambda} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

Производящая функция моментов

Функция

$$\psi_X(t) = \mathsf{E}e^{tX}$$

называется производящей функцией моментов (moment generating function). Название связано с тем, что

$$\psi_X^{(n)}(0) = \mathsf{E}X^n.$$

- Если $\psi_X(t) = \psi_Y(t)$, $t \in (-\varepsilon, \varepsilon)$, то X и Y одинаково распределены.

Пусть $\xi \sim \Gamma(\alpha, \lambda)$. Тогда

$$\psi_\xi(t) = \mathsf{E} e^{t\xi} = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x}\, dx$$
$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{y^{\alpha-1}}{(\lambda-t)^{\alpha-1}} e^{-y} \frac{dy}{\lambda-t} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{(\lambda-t)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y}\, dy$$
$$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha}.$$

Пусть $\xi_i \sim \Gamma(\alpha_i, \lambda)$ независимы. Тогда

$$\psi_{\xi_1+\cdots+\xi_n}(t) = \prod_{i=1}^n \psi_{\xi_i}(t) = \frac{\lambda^{\alpha_1+\cdots+\alpha_n}}{(\lambda-t)^{\alpha_1+\cdots+\alpha_n}}.$$

Следовательно, $\xi_1 + \cdots + \xi_n \sim \Gamma(\alpha_1 + \cdots + \alpha_n, \lambda)$.

χ^2-распределение

χ^2*-распределением* с m степенями свободы называется распределение случайной величины

$$\zeta = \sum_{j=1}^{n} \xi_j^2,$$

где $\xi_j \sim N(0,1)$ — независимые случайные величины. Используется обозначение $\zeta \sim \chi^2(n)$.

- Покажем, что $\chi^2(n) = \Gamma(n/2, 1/2)$.

Пусть $\xi \sim N(0,1)$. Тогда $\xi^2 \sim \Gamma(1/2, 1/2)$:

$$F_{\xi^2}(y) = \mathsf{P}(\xi^2 \le y) = \mathsf{P}(-\sqrt{y} \le \xi \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}),$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz.$$

$$f_{\xi^2}(y) = F'_{\xi^2}(y) = \Phi'(\sqrt{y})\frac{1}{2\sqrt{y}} + \Phi'(-\sqrt{y})\frac{1}{2\sqrt{y}} = \Phi'(\sqrt{y})\frac{1}{\sqrt{y}}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{1/2-1} e^{-y/2},$$

так как $\Gamma(1/2) = \sqrt{\pi}$. С учетом независимости ξ_i,

$$\xi_1^2 + \cdots + \xi_n^2 \sim \Gamma(n/2, 1/2),$$

$$f_{\xi_1^2+\cdots+\xi_n^2}(x) = \begin{cases} \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0, \\ 0, & \text{иначе.} \end{cases}$$

Пример. $X_i \sim N(\mu, \sigma^2)$, μ известно. Построить доверительный интервал для σ^2, используя статистику $\overline{(X-\mu)^2}$.
Решение.

$$G_n(\boldsymbol{X}, \sigma^2) = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = Z, \quad Z \sim \chi^2(n).$$

Пусть $\chi^2_{\alpha,n}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(n)$: $\mathsf{P}(Z \geq \chi^2_{\alpha,n}) = \alpha$. Тогда

$$\begin{aligned}
1-\alpha &= \mathsf{P}(\chi^2_{1-\alpha/2,n} < Z < \chi^2_{\alpha/2,n}) \\
&= \mathsf{P}\left(\frac{1}{\chi^2_{\alpha/2,n}} \sum_{i=1}^{n} (X_i-\mu)^2 < \sigma^2 < \frac{1}{\chi^2_{1-\alpha/2,n}} \sum_{i=1}^{n} (X_i-\mu)^2 \right) \\
&= \mathsf{P}\left(\frac{n}{\chi^2_{\alpha/2,n}} \overline{(X-\mu)^2} < \sigma^2 < \frac{n}{\chi^2_{1-\alpha/2,n}} \overline{(X-\mu)^2} \right).
\end{aligned}$$

Многомерное нормальное (гауссовское) распределение

Напомним, что по определению, $\boldsymbol{Y} = (Y^1, \ldots, Y^k) \sim N(\boldsymbol{\mu}, \Sigma)$, если

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{u}) = \mathsf{E} e^{i\langle \boldsymbol{u}, \boldsymbol{Y} \rangle} = e^{i\langle \boldsymbol{u}, \boldsymbol{\mu} \rangle - \langle \Sigma \boldsymbol{u}, \boldsymbol{u} \rangle / 2}.$$

Здесь $\boldsymbol{\mu}, \boldsymbol{u} \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$ — симметричная неотрицательно определенная матрица. Параметры $\boldsymbol{\mu}$ и Σ определяют вектор математических ожиданий и ковариационную матрицу $\boldsymbol{Y}$:

$$\mathsf{E}\boldsymbol{Y} = \mu, \quad \mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = (\mathrm{Cov}(Y_i, Y_j))_{i,j=1}^k = \Sigma.$$

Заметим, что если $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \Sigma)$ и $A \in \mathbb{R}^{m \times k}$, то

$$A\boldsymbol{Y} \sim N(A\boldsymbol{\mu}, A\Sigma A^T).$$

Действительно,

$$\begin{aligned} \varphi_{A\boldsymbol{Y}}(u) &= \mathsf{E} e^{i\langle \boldsymbol{u}, A\boldsymbol{Y} \rangle} = \mathsf{E} e^{i\langle A^T \boldsymbol{u}, \boldsymbol{Y} \rangle} = e^{i\langle A^T \boldsymbol{u}, \boldsymbol{\mu} \rangle - \langle \Sigma A^T \boldsymbol{u}, A^T \boldsymbol{u} \rangle / 2} \\ &= e^{i\langle \boldsymbol{u}, A\boldsymbol{\mu} \rangle - \langle A\Sigma A^T \boldsymbol{u}, \boldsymbol{u} \rangle / 2}. \end{aligned}$$

- (Теорема Фишера) Пусть $X_i \sim N(\mu, \sigma^2)$ независимы. Тогда

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

независимы, и $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$.

Доказательство. $Z_i := (X_i - \mu)/\sigma \sim N(0,1)$. l^2-норма вектора

$$(1/\sqrt{n}, \dots, 1/\sqrt{n})$$

равна 1, поэтому его можно дополнить до ортонормированного базиса в $\mathbb{R}^n$. Из векторов(-строк) этого базиса составим матрицу C и положим $\boldsymbol{Y} = C\boldsymbol{Z}$. Поскольку $Y \sim N(0, CIC^T) = N(0, I)$, то компоненты Y независимы. При этом

$$Y_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Z_i = \sqrt{n} \cdot \overline{Z} = \sqrt{n}(\overline{X} - \mu)/\sigma.$$

$$
\begin{aligned}
\frac{n-1}{\sigma^2}S^2 &= \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} - \frac{\overline{X} - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} \left(Z_i - \overline{Z}\right)^2 \\
&= \sum_{i=1}^{n} \left(Z_i - \overline{Z}\right)^2 = \sum_{i=1}^{n} Z_i^2 - 2\overline{Z} \sum_{i=1}^{n} Z_i + n\overline{Z}^2 = \sum_{i=1}^{n} Z_i^2 - n\overline{Z}^2 \\
&= \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2.
\end{aligned}
$$

Следовательно, S^2 не зависит от $\overline{X}$ и $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$. $\square$

Пример. $X_i \sim N(\mu, \sigma^2)$, μ неизвестно. Построить доверительный интервал для σ^2, используя статистику

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Решение.

$$G_n(\boldsymbol{X}, \sigma^2) = \frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1).$$

Пусть $\chi^2_{\alpha,n}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(n)$: $\mathsf{P}(V \geq \chi^2_{\alpha,n-1}) = \alpha$, $V \sim \chi^2(n-1)$. Тогда

$$\begin{aligned}
1-\alpha &= \mathsf{P}(\chi^2_{1-\alpha/2,n-1} < V < \chi^2_{\alpha/2,n-1}) \\
&= \mathsf{P}\left(\chi^2_{1-\alpha/2,n-1} < \frac{n-1}{\sigma^2}S^2 < \chi^2_{\alpha/2,n-1}\right) \\
&= \mathsf{P}\left(\frac{n-1}{\chi^2_{\alpha/2,n-1}}S^2 < \sigma^2 < \frac{n-1}{\chi^2_{1-\alpha/2,n-1}}S^2\right).
\end{aligned}$$

Пример. Построить 90% доверительный интервал для дисперсии нормальной случайной величины, если дана выборка из 10 ее значений

$$0.123,\ 0.133,\ 0.124,\ 0.125,\ 0.126,\ 0.128,\ 0.120,\ 0.124,\ 0.130,\ 0.126.$$

Решение. $\alpha = 0.1$, $\chi^2_{0.05,9} = 16.92$, $\chi^2_{0.95,9} = 3.325$,

$$s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 = 1.366 \cdot 10^{-5},$$

$$\sigma^2 \in \left(\frac{9}{\chi^2_{0.05,9}}s^2, \frac{9}{\chi^2_{0.95,9}}s^2\right) = (7.26 \cdot 10^{-6}, 3.70 \cdot 10^{-5}).$$

```
import numpy as np
x=np.zeros(10)
x[0]=0.123
x[1]=0.133
x[2]=0.124
x[3]=0.125
x[4]=0.126
x[5]=0.128
x[6]=0.120
x[7]=0.124
x[8]=0.130
x[9]=0.126
```

```
from scipy.stats import chi2
from IPython.display import Latex
n=10
```

```
Latex("$\chi^2_{{0.05},9}$=%0.3e" % chi2.ppf(0.95,n-1))
```

$\chi^2_{0.05,9}$=1.692e+01

```
Latex("$\chi^2_{{0.95},9}$=%0.3e" % chi2.ppf(0.05,n-1))
```

$\chi^2_{0.95,9}$=3.325e+00

```
s_squared=np.sum((x-np.mean(x))**2)/(n-1)
Latex("$s^2$=%0.3e" %s_squared)
```

s^2=1.366e-05

```
lower=(n-1)*s_squared/chi2.ppf(0.95,n-1)
upper=(n-1)*s_squared/chi2.ppf(0.05,n-1)
print("lower=%0.2e,upper=%0.2e" % (lower, upper))
```

```
lower=7.26e-06,upper=3.70e-05
```

- Пусть $X \sim N(0, I_p)$ и Π — ортопроектор на подпространство размерности $r \le p$. Тогда

$$\langle \Pi X, X \rangle = \|\Pi X\|^2 \sim \chi^2(r).$$

Доказательство. Характеристическая функция $\xi := \langle \Pi X, X \rangle$:

$$\varphi_\xi(u) = \mathsf{E} e^{iu\langle \Pi X, X\rangle} = \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} e^{iu\langle \Pi x, x\rangle} e^{-(x_1^2+\dots+x_p^2)/2}\, dx.$$

Составим базис из ортонормированных собственных векторов v_i матрицы Π. Пусть $Q = (v_1, \dots, v_p)$ — матрица, столбцами которой являются v_i. Тогда

$$Q^T \Pi Q = \operatorname{diag}(\lambda_1, \dots, \lambda_p).$$

Собственными значениями Π могут быть только 0 и 1:

$$\Pi v = \lambda v \Longrightarrow \Pi^2 v = \Pi v = \lambda \Pi v = \lambda^2 v \Longrightarrow \lambda = \lambda^2.$$

Пусть для определенности $v_1, \ldots, v_r$ соответствуют собственные числа 1, а остальным: 0. Тогда

$$\langle \Pi Q y, Q y \rangle = \langle Q^T \Pi Q y, y \rangle = \sum_{i=1}^{r} y_i^2.$$

Сделаем в интеграле замену переменных $x = Qy$. В силу ортогональности Q,

$$|\det Q| = 1, \quad \|x\|^2 = \|Qy\|^2 = \|y\|^2,$$

$$\begin{aligned}
\varphi_\xi(u) &= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} e^{iu(y_1^2 + \cdots + y_r^2)} e^{-(y_1^2 + \cdots + y_p^2)/2} \, dy \\
&= \frac{1}{(2\pi)^{r/2}} \int_{\mathbb{R}^r} e^{iu(y_1^2 + \cdots + y_r^2)} e^{-(y_1^2 + \cdots + y_r^2)/2} \, dy \\
&= \varphi_\eta(u), \quad \eta \sim \chi^2(r).
\end{aligned}$$

Значит, $\xi \sim \chi^2(r)$. $\Box$

t-распределение Стьюдента

Пусть $Z \sim N(0,1)$, $V \sim \chi^2(n)$ — независимые случайные величины. t-распределением Стьюдента с n степенями свободы называется распределение случайной величины

$$T = \frac{Z}{\sqrt{V/n}},$$

обозначение: $T \sim \mathcal{T}(n)$.
Чтобы найти плотность T рассмотрим двумерную случайную величину

$$(S,T) = \boldsymbol{h}(V,Z), \quad \boldsymbol{h}(x,y) = \left(x, \frac{y}{\sqrt{x/n}}\right),$$

$$(V,Z) = \boldsymbol{h}^{-1}(S,T), \quad h^{-1}(s,t) = (s, t\sqrt{s/n}).$$

Пусть $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ — взаимно-однозначное дифференцируемое отображение. Пусть $\eta = (\eta_1, \ldots, \eta_d) = g(\xi_1, \ldots, \xi_d) = g(\xi)$. Тогда

$$\mathsf{E}h(\eta) = \int_{\mathbb{R}^d} h(y) f_\eta(y)\, dy = \mathsf{E}h(g(\xi)) = \int_{\mathbb{R}^d} h(g(x)) f_\xi(x)\, dx$$
$$= \int_{\mathbb{R}^d} h(y) f_\xi(g^{-1}(y)) |\det\left[(g^{-1})'\right]|\, dy,$$

где $(g^{-1})'$ — якобиан отображения g^{-1}. Таким образом,

$$f_\eta(y) = f_\xi(g^{-1}(y)) \left|\det\left[(g^{-1})'\right]\right|.$$

В нашем случае

$$f_{S,T}(s,t) = f_{V,Z}(h^{-1}(s,t)) |\det\left[(h^{-1})'(s,t)\right]|,$$

$$f_{V,Z}(x,y) = f_V(x) f_Z(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

$x > 0, y \in \mathbb{R}$.

Подставляя $x = s$, $y = t\sqrt{s/n}$, находим

$$f_{V,Z}(h^{-1}(s,t)) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} s^{n/2-1} e^{-s/2} \frac{1}{\sqrt{2\pi}} e^{-t^2 s/(2n)},$$

$$(h^{-1})'(s,t) = \begin{pmatrix} 1 & 0 \\ t/(2\sqrt{sn}) & \sqrt{s/n} \end{pmatrix}, \quad \det\left[(h^{-1})'(s,t)\right] = \sqrt{\frac{s}{n}}.$$

Таким образом,

$$\begin{aligned} f_{S,T}(s,t) &= \frac{(1/2)^{n/2}}{\sqrt{n}\Gamma(n/2)} s^{n/2-1/2} e^{-s/2} \frac{1}{\sqrt{2\pi}} e^{-t^2 s/(2n)} \\ &= \frac{s^{(n-1)/2}}{2^{(n+1)/2}\sqrt{n\pi}\Gamma(n/2)} e^{-s(1+t^2/n)/2}, \quad s > 0, t \in \mathbb{R}, \end{aligned}$$

$$
\begin{aligned}
f_T(t) &= \int_0^\infty f_{S,T}(s,t)\,ds \quad // \quad u = s(1+t^2/n)/2 \\
&= \frac{1}{\sqrt{n\pi}\Gamma(n/2)} \frac{1}{(1+t^2/n)^{(n+1)/2}} \int_0^\infty u^{(n+1)/2-1} e^{-u} du \\
&= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \frac{1}{(1+t^2/n)^{(n+1)/2}}.
\end{aligned}
$$

При $n = 1$ получаем плотность распределения Коши:

$$
f_T(t) = \frac{1}{\pi}\frac{1}{1+t^2},
$$

так как

$$
\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(1) = 1.
$$

Пример. $X_i \sim N(\mu, \sigma^2)$, σ неизвестно. Построить доверительный интервал для μ.

Решение. Рассмотрим

$$\begin{aligned} G_n(\boldsymbol{X}, \sigma^2) &= \sqrt{n}\frac{\overline{X} - \mu}{S} \\ &= \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \frac{1}{\sqrt{\frac{1}{n-1}(n-1)S^2/\sigma^2}} \sim \mathcal{T}(n-1), \end{aligned}$$

так как

$$\sqrt{n}\frac{\overline{X} - \mu}{\sigma} \sim N(0,1), \quad (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$$

независимы.

Распределение Стьюдента симметрично. Пусть $t_{\alpha,n}$ — квантиль уровня $1-\alpha$ распределения $\mathcal{T}(n)$: $\mathsf{P}(T \geq \chi^2_{\alpha,n-1}) = \alpha$, $T \sim \mathcal{T}(n-1)$. Тогда

$$\begin{aligned} 1-\alpha &= \mathsf{P}(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) \\ &= \mathsf{P}\left(-t_{\alpha/2,n-1} < \sqrt{n}\frac{\overline{X}-\mu}{S} < t_{\alpha/2,n-1}\right) \\ &= \mathsf{P}\left(\overline{X} - \frac{S}{\sqrt{n}}t_{\alpha/2,n-1} < \mu < \overline{X} + \frac{S}{\sqrt{n}}t_{\alpha/2,n-1}\right). \end{aligned}$$

Пример. Пусть $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, 9$ описывают передаваемый сигнал. Получены значения

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5.$$

Построить двусторонний 95% доверительный интервал для μ, если значение σ неизвестно.
Решение. $\overline{x} = 9$,

$$s^2 = \frac{1}{8}\sum_{i=1}^{9}(x_i - \overline{x}) = 9.5, \quad s = 3.082,$$

$\alpha = 0.05$, $t_{0.025,8} = 2.306$,

$$\mu \in \left(\overline{x} - \frac{s}{\sqrt{n}} t_{0.025,8} < \mu < \overline{x} + \frac{s}{\sqrt{n}} t_{0.025,8}\right)$$
$$= \left(9 - \frac{3.082}{3} 2.306, 9 + \frac{3.082}{3} 2.306\right) = (6.63, 11.37).$$

Асимптотические доверительные интервалы для асимптотически нормальных оценок

Напомним, что оценка $\widehat{\theta}_n$ параметра θ называется *асимптотически нормальной* с коэффициентом $\sigma^2(\theta)$, если

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \sigma^2(\theta)).$$

В частности, для оценок максимального правдоподобия $\sigma^2(\theta) = 1/I(\theta)$ при выполнении условия регулярности.

- Пусть $\widehat{\theta}_n$ является асимптотически нормальной оценкой параметра θ с коэффициентом нормальности $\sigma^2(\theta)$, и функция $\sigma^2(\theta)$ непрерывна. Тогда интервал

$$\left(\widehat{\theta}_n - \frac{z_{\alpha/2}\sigma(\widehat{\theta}_n)}{\sqrt{n}}, \widehat{\theta}_n + \frac{z_{\alpha/2}\sigma(\widehat{\theta}_n)}{\sqrt{n}}\right)$$

  является асимптотическим доверительным интервалом для θ уровня доверия $1-\alpha$. Здесь $z_{\alpha/2}$ — $(1-\alpha/2)$-квантиль стандартного нормального распределения.

Доказательство. По определению асимптотической нормальности,

$$\sqrt{n}\frac{\widehat{\theta}_n - \theta}{\sigma(\theta)} \xrightarrow{d} Z \sim N(0,1).$$

По теореме о непрерывном преобразовании,

$$\frac{\sigma(\theta)}{\sigma(\widehat{\theta}_n)} \xrightarrow{\mathsf{P}} 1,$$

так как $\widehat{\theta}_n$ является состоятельной. Следовательно,

$$G(\boldsymbol{X};\theta) = \sqrt{n}\frac{\widehat{\theta}_n - \theta}{\sigma(\widehat{\theta}_n)} = \frac{\sigma(\theta)}{\sigma(\widehat{\theta}_n)}\sqrt{n}\frac{\widehat{\theta}_n - \theta}{\sigma(\theta)} \xrightarrow{d} Z \sim N(0,1)$$

по теореме Слуцкого. В силу симметрии стандартного нормального распределения,

$$\mathsf{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Разрешим неравенство

$$-z_{\alpha/2} < G(\boldsymbol{X};\theta) = \sqrt{n}\frac{\widehat{\theta}_n - \theta}{\sigma(\widehat{\theta}_n)} < z_{\alpha/2}$$

относительно θ:

$$\lim_{n\to\infty} \mathsf{P}_\theta\left(\widehat{\theta}_n - \frac{z_{\alpha/2}\sigma(\widehat{\theta}_n)}{\sqrt{n}} < \theta < \widehat{\theta}_n + \frac{z_{\alpha/2}\sigma(\widehat{\theta}_n)}{\sqrt{n}}\right)$$

$$= \mathsf{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha. \quad \square$$

Пример. $X_i \sim \Pi(\theta)$. Построить для θ асимптотический доверительный интервал уровня доверия $1-\alpha$.
Решение. Напомним, что $\mathsf{E}X_1 = \theta$, $\mathrm{Var}(X_1) = \theta$. Согласно ЦПТ,

$$G_n(\boldsymbol{X};\theta) = \sqrt{n}\frac{\overline{X} - \mathsf{E}X_1}{\sqrt{\mathrm{Var}(X_1)}} = \sqrt{n}\frac{\overline{X} - \theta}{\sqrt{\theta}} \xrightarrow{d} \eta \sim N(0,1),$$

Следовательно, $\sigma(\theta) = \sqrt{\theta}$,

$$\mathsf{P}_\theta\left(\overline{X} - \frac{z_{\alpha/2}\sqrt{\overline{X}}}{\sqrt{n}} < \theta < \overline{X} + \frac{z_{\alpha/2}\sqrt{\overline{X}}}{\sqrt{n}}\right) \to 1-\alpha.$$

Пример. $X_i \sim \mathrm{Ber}(\theta)$,

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0,1\}.$$

Для оценки максимального правдоподобия $\widehat{\theta} = \overline{X}$ был получен следующий результат:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, 1/I(\theta)) = N(0, \theta(1-\theta)).$$

Следовательно,

$$\mathsf{P}_\theta \left(\overline{X} - \frac{z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}} < \theta < \overline{X} + \frac{z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}} \right) \to 1-\alpha.$$

Ross (2021), Example 7.5.b

В августе 2013 года газета New York Times сообщила, что согласно опросу 52 процента населения одобряют работу президента Обамы с погрешностью ± 4 процента. Что это значит? Можем ли мы сделать вывод, сколько людей было опрошено?

Решение. Обычно в новостях сообщают 95% доверительные интервалы. В данном случае 95% доверительный интервал для доли θ людей, одобряющих работу Обамы, имеет вид

$$\left(0.52 - \frac{z_{0.025}\sqrt{0.52 \cdot 0.48}}{\sqrt{n}}, 0.52 + \frac{z_{0.025}\sqrt{0.52 \cdot 0.48}}{\sqrt{n}}\right), \quad z_{0.025} = 1.96.$$

По условию,

$$\frac{1.96\sqrt{0.52 \cdot 0.48}}{\sqrt{n}} = 0.04, \quad n = 599.29.$$

Таким образом, было опрошено примерно 599 человек, и 52% из них ответили, что одобряют работу Обамы.

Пример. $X_i \sim \mathrm{Exp}(\theta)$. Построить для θ асимптотический доверительный интервал уровня доверия $1-\alpha$.
Решение. Для θ была получена оценка максимального правдоподобия: $\widehat{\theta} = 1/\overline{X}$, которая является асимптотически нормальной:

$$l(x;\theta) = \ln f(x;\theta) = \ln(\theta e^{-\theta x}) = \ln\theta - \theta x,$$

$$I(\theta) = -\mathsf{E}_\theta\left(l_{\theta\theta}(X_1;\theta)\right) = \frac{1}{\theta^2},$$

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}\left(\frac{1}{\overline{X}} - \theta\right) \xrightarrow{d} N(0, 1/I(\theta)) = N(0,\theta^2).$$

Следовательно,

$$\mathsf{P}_\theta\left(\frac{1}{\overline{X}} - \frac{z_{\alpha/2}}{\overline{X}\sqrt{n}} < \theta < \frac{1}{\overline{X}} + \frac{z_{\alpha/2}}{\overline{X}\sqrt{n}}\right) \to 1-\alpha.$$

Пример. Пусть X_i — независимые одинаково распределенные случайные величины с конечной дисперсией. Построить доверительный интервал для $\mu = \mathsf{E}X_1$.
Согласно центральной предельной теореме,

$$\sqrt{n}(\overline{X} - \mu)/\sigma \xrightarrow{d} Z \sim N(0,1).$$

S^2 является состоятельной оценкой дисперсии σ^2. По теореме Слуцкого,

$$\sqrt{n}(\overline{X} - \mu)/S \xrightarrow{d} Z \sim N(0,1).$$

Асимптотический доверительный интервал:

$$\mathsf{P}\left(-z_{\alpha/2} < \sqrt{n}(\overline{X} - \mu)/S < z_{\alpha/2}\right)$$
$$= \mathsf{P}\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right) \to \mathsf{P}\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha.$$

Пример. Пусть X_i, $i = 1, \dots, 9$ описывают передаваемый сигнал. Получены значения

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5.$$

Построить асимптотический 95% доверительный интервал для μ. В отличие от примера, рассмотренного ранее, нормальность X_i не предполагается.

Решение. $\overline{x} = 9$, $s = 3.082$, $\alpha = 0.05$, $z_{0.025} = 1.96$, $n = 9$,

$$\mu \in \left(\overline{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right) = (6.99, 11.01).$$

Проверка гипотез

Пусть множество Θ значений параметра θ разбито на два подмножества:

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset.$$

Имеются две гипотезы о параметре θ (виде распределения):

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1.$$

Гипотеза H_0 называется *нулевой*, а гипотеза H_1 — *альтернативной*.
На основе анализа выборки $X_1, \ldots, X_n$ мы хотим принять H_0 или отвергнуть ее в пользу H_1.
Гипотезы неравноправны: для того, чтобы отвергнуть H_0 нужны веские основания. Это соответствует презумпции невиновности: H_0 — гипотеза о невиновности.
Вопрос ставится так: дает ли выборка достаточные основания, чтобы отвергнуть H_0?

Для проверки гипотезы H_0 вводится *критическое множество* $R_n \subset \mathbb{R}^n$. Гипотеза H_0 отвергается в пользу H_1, если $(X_1, \ldots, X_n) \in R_n$ и принимается (не отвергается) в противном случае. Обычно

$$R_n = \{\boldsymbol{x} = (x_1, \ldots, x_n) : T_n(\boldsymbol{x}) > c\},$$

где T_n — тестовая статистика (статистика критерия) и c — критическое значение.

Говорят, что произошла *ошибка первого рода*, если критерий отверг *верную* гипотезу H_0, и *ошибка второго рода*, если критерий принял (не отверг) *неверную* гипотезу H_0.

	Принять H_0	Отвергнуть H_0
Верна H_0	Верное решение	Ошибка I рода
Верна H_1	Ошибка II рода	Верное решение

Функция

$$\beta(\theta) = \mathsf{P}_\theta(\boldsymbol{X} \in R)$$

называется *функцией мощности* критерия. Мы хотим, чтобы эта величина была малой, если H_0 верна $(\theta \in \Theta_0)$ и большой, если не верна $(\theta \in \Theta_1)$:

$$\beta(\theta) = \begin{cases} \text{вероятность отвергнуть нулевую гипотезу,} & \\ \text{когда она верна,} & \text{при } \theta \in \Theta_0 \\ \text{вероятность отвергнуть нулевую гипотезу,} & \\ \text{когда она неверна,} & \text{при } \theta \in \Theta_1 \end{cases}$$

Размером критерия называется

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta).$$

Говорят, что критерий имеет *уровень значимости* α, если его размер не превосходит α:

$$\beta(\theta) \le \alpha, \quad \theta \in \Theta_0.$$

Это означает, что вероятность ошибки первого рода не превосходит α.

Зафиксировав уровень значимости, стремятся найти критерий, функция мощности которого при $\theta \in \Theta_1$ является большой. Критерий называется *состоятельным*, если

$$\beta(\theta) \to 1, \quad n \to \infty, \quad \theta \in \Theta_1.$$

Пример. $X_i \sim N(\theta, \sigma^2)$, σ^2 известно.

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$$

Отклоняем H_0, если разность $\overline{X}_n - \theta_0$ велика по модулю:

$$T_n = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \sim N(0,1),$$

$$R_n = \{\boldsymbol{x} : |T_n| > c\}.$$

Контроль вероятности ошибки первого рода (выбираем критерий размера α): $c = z_{\alpha/2}$,

$$\mathsf{P}_{\theta_0}(|T_n| > z_{\alpha/2}) = \alpha.$$

Функция мощности:

$$\begin{aligned}
\beta(\theta) &= \mathsf{P}_\theta(|T_n| > z_{\alpha/2}) = 1 - \mathsf{P}_\theta\left(\frac{|\overline{X}_n - \theta_0|}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) \\
&= 1 - \mathsf{P}_\theta\left(-z_{\alpha/2} \le \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) \\
&= 1 - \mathsf{P}_\theta\left(-z_{\alpha/2} + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \le \frac{\overline{X}_n - \theta}{\sigma/\sqrt{n}} \le z_{\alpha/2} + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\
&= 1 - \left(\Phi\left(z_{\alpha/2} + \sqrt{n}\frac{\theta_0 - \theta}{\sigma}\right) - \Phi\left(-z_{\alpha/2} + \sqrt{n}\frac{\theta_0 - \theta}{\sigma}\right)\right),
\end{aligned}$$

где $\Phi(x) = \mathsf{P}(Z \le x)$, $Z \sim N(0,1)$. Поскольку

$$\lim_{x\to-\infty} \Phi(x) = 0, \quad \lim_{x\to+\infty} \Phi(x) = 1,$$

то $\lim_{n\to\infty} \beta(\theta) = 1$ при $\theta \ne \theta_0$. Естественно, $\beta(\theta_0) = \alpha$.

Пример. $X_i \sim N(\theta, \sigma^2)$, σ^2 известно.

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0.$$

Отклоняем H_0, если разность $\overline{X}_n - \theta_0$ принимает большие положительные значения:

$$T_n = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \sim N(0,1),$$

$$R_n = \{\boldsymbol{x} : T_n > c\}.$$

Контроль вероятности ошибки первого рода (выбираем критерий размера α): $c = z_\alpha$,

$$\mathsf{P}_{\theta_0}(T_n > z_\alpha) = \alpha.$$

Функция мощности:

$$\begin{aligned}\beta(\theta) = \mathsf{P}_\theta(T_n > z_\alpha) &= 1 - \mathsf{P}_\theta\left(\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \le z_\alpha\right) \\ &= 1 - \mathsf{P}_\theta\left(\frac{\overline{X}_n - \theta}{\sigma/\sqrt{n}} \le z_\alpha + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(z_\alpha + \sqrt{n}\frac{\theta_0 - \theta}{\sigma}\right).\end{aligned}$$

Функция $\theta \mapsto \beta(\theta)$ является неубывающей,

$$\lim_{n\to\infty} \beta(\theta) = 1, \quad \theta > \theta_0; \quad \lim_{n\to\infty} \beta(\theta) = 0, \quad \theta < \theta_0; \quad \beta(\theta_0) = \alpha.$$

Пример. $X_i \sim N(\theta, \sigma^2)$, σ^2 неизвестно.

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$$

Отклоняем H_0, если разность $\overline{X}_n - \theta_0$ принимает большие по модулю значения. Поскольку σ^2 неизвестно, будем использовать его несмещенную оценку $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Получаем распределение Стьюдента:

$$T_n = \frac{\overline{X}_n - \theta_0}{S/\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \theta_0)/\sigma}{\sqrt{\frac{1}{n-1}(n-1)S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}} \sim \mathcal{T}(n-1),$$

так как $Z \sim N(0,1)$, $V \sim \chi^2(n-1)$ независимы.

$$R_n = \{\boldsymbol{x} : |T_n| > c\}.$$

Контроль вероятности ошибки первого рода: $c = t_{n-1,\alpha/2}$,

$$\mathsf{P}_{\theta_0}(|T_n| > t_{n-1,\alpha/2}) = \alpha.$$

Пример. Пусть имеется асимптотически нормальная оценка параметра θ_0:

$$\sqrt{n}\frac{\widehat{\theta}_n - \theta_0}{\sigma} \xrightarrow{d} N(0,1).$$

Если $\widehat{\sigma}_n \xrightarrow{\mathsf{P}} \sigma$, то

$$\sqrt{n}\frac{\widehat{\theta}_n - \theta_0}{\widehat{\sigma}_n} \xrightarrow{d} N(0,1)$$

по теореме Слуцкого. Для проверки гипотезы

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

можно использовать статистику

$$T_n = \sqrt{n}\frac{\widehat{\theta}_n - \theta_0}{\widehat{\sigma}_n}.$$

Пусть $R_n = \{\mathbf{x} : |T_n| > c\}$: критерий Вальда. Выберем c так, чтобы получить критерий асимптотического размера α:

$$\mathsf{P}_{\theta_0}(|T_n| > z_{\alpha/2}) \to \alpha.$$

Условие

$$|T_n| \le z_{\alpha/2} \iff \theta_0 \in \left(\widehat{\theta}_n - z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}, \widehat{\theta}_n + z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}\right)$$

означает, что θ_0 накрывается соответствующим асимптотическим доверительным интервалом. Таким образом, гипотеза H_0 отвергается на уровне значимости α тогда и только тогда, когда θ_0 не накрывается данным интервалом.

p-значения

Не будем фиксировать уровень значимости и рассмотрим семейство критических областей $R_n(\alpha) = \{\mathbf{x} : T_n(\mathbf{x}) \geq c_\alpha\}$, соответствующих уровням значимости α:

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(T_n(X) \geq c_\alpha) = \alpha.$$

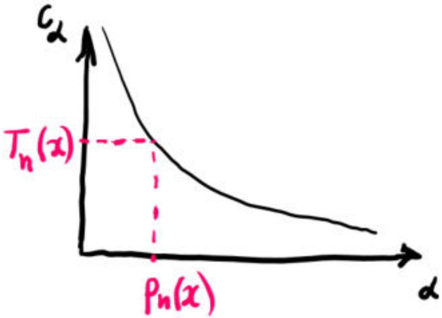
p-значение (наблюдаемый уровень значимости) определим следующим образом:

$$p_n(x) = \inf\{\alpha : x \in R_n(\alpha)\} = \inf\{\alpha : T_n(x) \geq c_\alpha\}.$$

$p_n(x)$ — наименьший уровень значимости, при котором x попадает в критическую область, т.е. нулевая гипотеза будет отвергнута.

$\alpha \mapsto c_\alpha$ — невозрастающая непрерывная функция, то

$$T_n(x) = c_{p_n(x)}.$$



Таким образом,

$$p_n(x) = \sup_{\theta \in \Theta_0} \mathsf{P}_\theta(T_n(X) \geq c_{p_n(x)}) = \sup_{\theta \in \Theta_0} \mathsf{P}_\theta(T_n(X) \geq T_n(x)),$$

т.е. $p_n(x)$ — максимальная вероятность увидеть $T_n(x)$ или большее значение при выполнении гипотезы H_0.

- p-значение $p(X)$ имеет равномерное распределение $U(0,1)$ при выполнении гипотезы $H_0 : \theta = \theta_0$.

Доказательство. Если

$$p_n(x) = \inf\{\alpha : T_n(x) \geq c_\alpha\} \leq u < v,$$

то $T_n(x) \geq c_v$ и

$$\mathsf{P}_{\theta_0}(p_n(X) \leq u) \leq \mathsf{P}_{\theta_0}(T_n(X) \geq c_v) = v, \quad v > u.$$

Следовательно, $\mathsf{P}_{\theta_0}(p_n(X) \leq u) \leq u$. С другой стороны, если $T_n(x) \geq c_u$, то $p_n(x) \leq u$. Таким образом,

$$\mathsf{P}_{\theta_0}(p_n(X) \leq u) \geq \mathsf{P}_{\theta_0}(T_n(X) \geq c_u) = u. \quad \square$$

p-значение	вывод
(0,0.01)	очень сильное свидетельство против H_0
(0.01,0.05)	сильное свидетельство против H_0
(0.05,0.1)	слабое свидетельство против H_0
(0.1,1)	нет свидетельства против H_0

Ross (2021), Example 8.3.h

Представитель общественного здравоохранения утверждает, что среднее потребление воды в доме 350 галлонов (1 галлон = 3.79 литра) в день. Чтобы проверить это утверждение, было проведено исследование 20 случайно выбранных домов, в результате чего среднесуточное потребление воды в этих 20 домах было следующим:

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Не противоречат ли данные заявлению чиновника?

Решение. Предположим, что данные подчиняются нормальному закону: $X_i \sim N(\mu, \sigma^2)$.

$$H_0 : \mu = \mu_0 = 350, \quad H_1 : \mu \neq 350.$$

$$\overline{x} = 353.8, \quad s = 21.8478.$$

Гипотеза H_0 отвергается на уровне значимости α, если

$$|T| = \frac{\sqrt{n}|\overline{x} - \mu_0|}{s} > t_{\alpha/2,n-1}.$$

В нашем случае

$$|T| = \frac{\sqrt{20} \cdot 3.8}{21.8478} = 0.7778.$$

Поскольку это меньше чем $t_{19,0.05} = 1.7$, то нулевая гипотеза принимается (не отвергается) на 10-процентном уровне значимости ($\alpha = 0.1$). p-значение

$$p\text{-value} = \mathsf{P}_{\mu_0}(|T| > 0.7778) = 2\mathsf{P}_{\mu_0}(T > 0.7778) = 0.4462$$

показывает, что гипотеза не будет отвергнута на любом разумном уровне значимости.

Для сравнения применим критерий Вальда

$$|T| = \sqrt{n}\frac{|\overline{x} - \mu_0|}{\widehat{\sigma}} > z_{\alpha/2}$$

с $\widehat{\sigma} = s$. Эта та же самая тестовая статистика, но ее распределение приближенно считается нормальным. Для $\alpha = 0.1$ имеем

$$|T| = 0.7778 < z_{\alpha/2} = z_{0.05} = 1.65.$$

Значит гипотеза не отвергается на 10-процентном уровнен значимости. Приближенное p-значение:

$$\begin{aligned} p\text{-value} &= \mathsf{P}_{\mu_0}(|T| > 0.7778) = 2\mathsf{P}_{\mu_0}(T > 0.7778) \\ &\approx 2(1 - \Phi(0.7778)) \approx 2(1 - 0.78) \approx 0.44, \quad \text{если } T \approx N(0,1). \end{aligned}$$

Вывод тот же самый: гипотеза не будет отвергнута на любом разумном уровне значимости.

Пример. Проверка гипотезы о симметричности монеты: $X_i \sim \mathrm{Ber}(p)$,

$$H_0 : p = p_0 := 0.5, \quad H_1 : p \neq p_0.$$

$$T = \frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} \to N(0,1), \quad T' = \frac{\overline{X} - p_0}{\sqrt{\overline{X}(1-\overline{X})/n}} \to N(0,1).$$

Рассмотрим критерий Вальда с критическим множеством:

$$R = \{\mathbf{x} : |T(x)| > z_{\alpha/2}\} \quad \text{или} \quad R' = \{\mathbf{x} : |T'(x)| > z_{\alpha/2}\}.$$

Для $n = 25$ найдем (приближенные) p-значения для всех возможных значений тестовой статистики T:

$$\mathsf{P}_{p_0}(|T(X)| > |T(x)|) \approx 2(1 - \Phi(|T(x)|)), \quad x = 0, \ldots, 25.$$

```
from scipy.stats import norm
import numpy as np
import pandas as pd
p_0=0.5
n=25
T=np.zeros(n+1)
p=np.zeros(n+1)
h=np.arange(0,n+1)
for x in h:
    T[x]=np.sqrt(n)*(x/n-p_0)/np.sqrt(p_0*(1-p_0))
    p[x]=2*(1-norm.cdf(np.abs(T[x])))
df=pd.DataFrame({'number of heads':h,'p-value':p})
print(df.to_string(index=False))
```

number of heads	p-value
0	5.733031e-07
1	4.224909e-06
2	2.669150e-05
3	1.446961e-04
4	6.738585e-04
5	2.699796e-03
6	9.322376e-03
7	2.780690e-02
8	7.186064e-02
9	1.615133e-01
10	3.173105e-01
11	5.485062e-01
12	8.414806e-01
13	8.414806e-01
14	5.485062e-01
15	3.173105e-01
16	1.615133e-01
17	7.186064e-02
18	2.780690e-02
19	9.322376e-03
20	2.699796e-03
21	6.738585e-04
22	1.446961e-04
23	2.669150e-05
24	4.224909e-06
25	5.733031e-07

p-значение	число орлов	вывод
(0,0.01)	0-6, 19-25	очень сильное свидетельство против H_0
(0.01,0.05)	7, 18	сильное свидетельство против H_0
(0.05,0.1)	8, 17	слабое свидетельство против H_0
(0.1,1)	9-16	нет свидетельства против H_0

Рандомизированным критерием называется (борелевская) функция

$$\psi : \mathbb{R}^n \mapsto [0,1].$$

Гипотеза H_0 отвергается с вероятностью $\psi(X_1, \ldots, X_n)$. "Средняя" вероятность ошибки первого рода равна

$$\mathsf{E}_\theta \psi(\boldsymbol{X}), \quad \theta \in \Theta_0.$$

Функция мощности определяется формулой

$$\beta_\psi(\theta) = \mathsf{E}_\theta \psi(\boldsymbol{X}), \quad \theta \in \Theta.$$

При $\psi(\boldsymbol{x}) = I_R(\boldsymbol{x})$ получаем нерандомизированный критерий.

Случай простых гипотез

Если нулевая и альтернативная гипотезы являются простыми:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1,$$

то задача построения наиболее мощного критерия при заданном уровне значимости $\alpha \in (0, 1)$ имеет вид:

$$\beta_\psi(\theta_1) = \mathsf{E}_{\theta_1} \psi(\boldsymbol{X}) \to \max_\psi$$

$$\beta_\psi(\theta_0) = \mathsf{E}_{\theta_0} \psi(\boldsymbol{X}) \le \alpha, \quad 0 \le \psi \le 1.$$

Введем отношение правдоподобия

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)}$$

и рассмотрим рандомизированный критерий отношения правдоподобия:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & f(\boldsymbol{x};\theta_1) > Kf(\boldsymbol{x};\theta_0), \\ 0, & f(\boldsymbol{x};\theta_1) < Kf(\boldsymbol{x};\theta_0), \\ \rho(\boldsymbol{x}), & f(\boldsymbol{x};\theta_1) = Kf(\boldsymbol{x};\theta_0), \end{cases}$$

$K > 0$, $0 \le \rho(\boldsymbol{x}) \le 1$.

Лемма Неймана-Пирсона

(a) (Оптимальность) Для любых K и $\rho(\mathbf{x})$ критерий ψ имеет наибольшую мощность среди всех рандомизированных критериев φ, размер которых не превосходит размера ψ:

$$\mathsf{E}_{\theta_0}\varphi(\boldsymbol{X}) \leq \mathsf{E}_{\theta_0}\psi(\boldsymbol{X}).$$

(b) (Существование) Для любого $\alpha \in (0,1)$ существуют константы K и ρ_0 такие, что критерий отношения правдоподобия ψ с параметрами K и $\rho(x) = \rho_0$ имеет размер α:

$$\mathsf{E}_{\theta_0}\psi(\boldsymbol{X}) = \alpha.$$

Доказательство. Young G.A., Smith R.L. Essentials of statistical inference (CUP, 2005).
(a) Пусть φ — любой критерий, удовлетворяющий условию

$$\mathsf{E}_{\theta_0}\varphi(\boldsymbol{X}) \leq \mathsf{E}_{\theta_0}\psi(\boldsymbol{X}).$$

Функция

$$U(\boldsymbol{x}) = (\psi(\boldsymbol{x}) - \varphi(\boldsymbol{x}))(f(\boldsymbol{x};\theta_1) - Kf(\boldsymbol{x};\theta_0))$$

удовлетворяет неравенству $U(\boldsymbol{x}) \geq 0$. Действительно,

$$\begin{aligned}
f(\boldsymbol{x};\theta_1) > Kf(\boldsymbol{x};\theta_0) &\Longrightarrow \psi(\boldsymbol{x}) = 1 \Longrightarrow U(\boldsymbol{x}) \geq 0,\\
f(\boldsymbol{x};\theta_1) < Kf(\boldsymbol{x};\theta_0) &\Longrightarrow \psi(\boldsymbol{x}) = 0 \Longrightarrow U(\boldsymbol{x}) \geq 0,\\
f(\boldsymbol{x};\theta_1) = Kf(\boldsymbol{x};\theta_0) &\Longrightarrow U(\boldsymbol{x}) = 0.
\end{aligned}$$

Если $f(\boldsymbol{x};\theta_0)$, $f(\boldsymbol{x};\theta_1)$ являются плотностями, то из неравенства

$$
\begin{aligned}
0 &\le \int_{\mathbb{R}^n} U(\boldsymbol{x})\,d\boldsymbol{x} = \int_{\mathbb{R}^n} (\psi(\boldsymbol{x}) - \varphi(\boldsymbol{x}))(f(\boldsymbol{x};\theta_1) - Kf(\boldsymbol{x};\theta_0))\,d\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} (\psi(\boldsymbol{x}) - \varphi(\boldsymbol{x}))f(\boldsymbol{x};\theta_1)\,d\boldsymbol{x} - K\int_{\mathbb{R}^n} (\psi(\boldsymbol{x}) - \varphi(\boldsymbol{x}))f(\boldsymbol{x};\theta_0)\,d\boldsymbol{x} \\
&= \mathsf{E}_{\theta_1}\psi(\boldsymbol{X}) - \mathsf{E}_{\theta_1}\varphi(\boldsymbol{X}) - K(\mathsf{E}_{\theta_0}\psi(\boldsymbol{X}) - \mathsf{E}_{\theta_0}\varphi(\boldsymbol{X})) \\
&\le \mathsf{E}_{\theta_1}\psi(\boldsymbol{X}) - \mathsf{E}_{\theta_1}\varphi(\boldsymbol{X})
\end{aligned}
$$

вытекает, что мощность φ не превосходит мощности ψ.
Если $f(\boldsymbol{x};\theta_0)$, $f(\boldsymbol{x};\theta_1)$ являются вероятностными массами, нужно рассмотреть соответствующие суммы.

(b) При $\rho(\boldsymbol{x}) = \rho_0$ уравнение $\mathsf{E}_{\theta_0}\psi(\boldsymbol{X}) = \alpha$ принимает вид (заметим, что $\mathsf{P}_{\theta_0}(f(\boldsymbol{X};\theta_0) = 0) = 0$)

$$\begin{aligned}\mathsf{E}_{\theta_0}\psi(\boldsymbol{X}) &= \mathsf{E}_{\theta_0} I_{\{f(\boldsymbol{X};\theta_1) > K f(\boldsymbol{X};\theta_0)\}} + \rho_0 \mathsf{E}_{\theta_0} I_{\{f(\boldsymbol{X};\theta_1) = K f(\boldsymbol{X};\theta_0)\}} \\ &= \mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) > K) + \rho_0 \mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) = K) = \alpha.\end{aligned}$$

Перепишем полученное уравнение в виде

$$G(K) = \mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) \le K) = 1 - \alpha + \rho_0 \mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) = K).$$

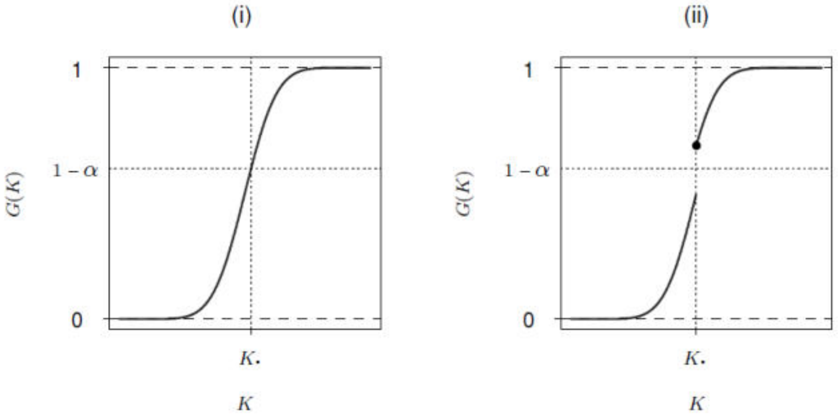
Функция $G(K)$ является неубывающей и непрерывной справа. (i) Если уравнение $G(K) = 1 - \alpha$ имеет решение K_0, то положим $K = K_0$, $\rho_0 = 0$. (ii) Иначе существует K_0:

$$G_-(K_0) = \mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) < K_0) \le 1 - \alpha < G(K_0).$$

В этом случае искомое значение ρ_0:

$$\rho_0 = \frac{G(K_0) - (1 - \alpha)}{G(K_0) - G_-(K_0)},$$

так как $\mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) = K_0) = G(K_0) - G_-(K_0)$.



(i)
(ii)
1
$1-\alpha$
0
$G(K)$
$K.$
K

Пример. $X_i \sim N(\theta, 1)$,

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1.$$

В данном случае

$$\begin{aligned} f(\boldsymbol{x}; \theta) &= \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\sum_{i=1}^n (x_i - \theta)^2/2} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\sum_{i=1}^n x_i^2/2 + \theta \sum_{i=1}^n x_i + n\theta^2/2}. \end{aligned}$$

Отношение правдоподобия:

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}; \theta_1)}{f(\boldsymbol{x}; \theta_0)} = e^{n(\theta_1^2 - \theta_0^2)} e^{n\overline{x}(\theta_1 - \theta_0)}.$$

Пусть $\theta_1 > \theta_0$. Тогда условие $\Lambda(x) > K$ эквивалентным образом переписывается так:

$$\overline{x} > \gamma.$$

В данном случае

$$\mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) = K) = \mathsf{P}_{\theta_0}(\overline{X} = \gamma) = 0.$$

Поэтому критерий отношения правдоподобия можно выбрать нерандомизированным:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \Lambda(\boldsymbol{x}) > K, \\ 0, & \Lambda(\boldsymbol{x}) \leq K. \end{cases}$$

Для любого $\alpha \in (0, 1)$ существует K такое, что соответствующий критерий имеет размер α (пункт (b) леммы Неймана-Пирсона).

$$\mathsf{P}_{\theta_0}(\Lambda(\mathbf{X}) > \lambda) = \mathsf{P}_{\theta_0}(\overline{X} > \gamma) = \alpha.$$

Такой критерий будет оптимальным среди всех критериев уровня значимости α (согласно пункту (a) леммы Неймана-Пирсона). Остается найти γ.

Поскольку $\sqrt{n}(\overline{X} - \theta_0) \sim N(0,1)$ относительно P_{θ_0}, то

$$\mathsf{P}_{\theta_0}(\overline{X} > \gamma) = \mathsf{P}_{\theta_0}(\sqrt{n}(\overline{X} - \theta_0) > \sqrt{n}(\gamma - \theta_0)) = 1 - \Phi(\sqrt{n}(\gamma - \theta_0)) = \alpha,$$

$$\Phi(\sqrt{n}(\gamma - \theta_0)) = 1 - \alpha, \quad \sqrt{n}(\gamma - \theta_0) = z_\alpha, \quad \gamma = \theta_0 + \frac{z_\alpha}{\sqrt{n}}.$$

Окончательная форма критерия уровня α:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \overline{x} > \theta_0 + \dfrac{z_\alpha}{\sqrt{n}}, \\ 0, & \text{иначе.} \end{cases}$$

Аналогичный критерий был получен ранее для проверки гипотезы

$$H_0 : \theta \le \theta_0, \quad H_1 : \theta > \theta_0.$$

Пример. $X_i \sim N(0, \sigma^2)$,

$$H_0 : \sigma = \sigma_0, \quad H_1 : \sigma = \sigma_1,$$

$$f(\boldsymbol{x}; \sigma^2) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\sum_{i=1}^n x_i^2/(2\sigma^2)}.$$

Отношение правдоподобия:

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}; \sigma_1^2)}{f(\boldsymbol{x}; \sigma_0^2)} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n \frac{x_i^2}{2}\right].$$

Пусть $\sigma_0 < \sigma_1$. Тогда условие $\Lambda(\boldsymbol{x}) > K$ принимает вид

$$\sum_{i=1}^n x_i^2 > \gamma.$$

Поскольку

$$\mathsf{P}_{\sigma_0}(\Lambda(\boldsymbol{X}) = K) = \mathsf{P}_{\sigma_0}\left(\sum_{i=1}^{n} X_i^2 = \gamma\right) = 0,$$

то критерий отношения правдоподобия можно выбрать нерандомизированным:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \Lambda(\boldsymbol{x}) > K, \\ 0, & \Lambda(\boldsymbol{x}) \le K. \end{cases}$$

Из леммы Неймана-Пирсона вытекает, что оптимальное значение параметра определяется из условия

$$\mathsf{P}_{\sigma_0}(\Lambda(\mathbf{X}) > K) = \mathsf{P}_{\sigma_0}\left(\sum_{i=1}^{n} X_i^2 > \gamma\right) = \alpha.$$

Другими словами,

$$\alpha = \mathsf{P}_{\sigma_0}\left(\sum_{i=1}^{n} X_i^2 > \gamma\right) = \mathsf{P}_{\sigma_0}\left(\sum_{i=1}^{n} X_i^2/\sigma_0^2 > \gamma/\sigma_0^2\right) = \mathsf{P}_{\sigma_0}\left(V > \frac{\gamma}{\sigma_0^2}\right)$$

где $V \sim \chi^2(n)$. Следовательно,

$$\mathsf{P}_{\sigma_0}\left(V \leq \frac{\gamma}{\sigma_0^2}\right) = \mathsf{P}_{\sigma_0}\left(V \leq \chi^2_{\alpha,n}\right) = 1 - \alpha, \quad \gamma = \chi^2_{\alpha,n}\sigma_0^2.$$

Окончательная форма критерия уровня α:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \displaystyle\sum_{i=1}^{n} x_i^2 > \chi^2_{\alpha,n}\sigma_0^2, \\ 0, & \text{иначе.} \end{cases}$$

Пусть теперь $\sigma_0^2 > \sigma_1^2$. Тогда условие $\Lambda(\boldsymbol{x}) > K$ принимает вид

$$\sum_{i=1}^{n} x_i^2 < \gamma.$$

Повторяя те же рассуждения находим, что оптимальный критерий имеет вид

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \sum_{i=1}^{n} x_i^2 < \chi^2_{1-\alpha,n}\sigma_0^2, \\ 0, & \text{иначе.} \end{cases}$$

Пример. $X_i \sim \mathrm{Exp}(\theta)$,

$$f(\boldsymbol{x};\theta) = \theta^n e^{-\theta(x_1+\dots+x_n)},$$

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1.$$

Отношение правдоподобия:

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} = \frac{\theta_1^n}{\theta_0^n} e^{(\theta_0-\theta_1)(x_1+\dots+x_n)}.$$

Пусть $\theta_1 < \theta_0$, тогда условие $\Lambda(\boldsymbol{x}) > K$ принимает вид

$$\sum_{i=1}^n x_i > \gamma.$$

Поскольку $\mathsf{P}_{\theta_0}(\Lambda(\boldsymbol{X}) = K) = \mathsf{P}_{\theta_0}\left(\sum_{i=1}^n X_i = \gamma\right) = 0$, то критерий отношения правдоподобия можно выбрать нерандомизированным:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \Lambda(\boldsymbol{x}) > K, \\ 0, & \Lambda(\boldsymbol{x}) \le K. \end{cases}$$

Из леммы Неймана-Пирсона вытекает, что оптимальное значение параметра определяется из условия

$$\mathsf{P}_{\theta_0}(\Lambda(\mathbf{X}) > K) = \mathsf{P}_{\theta_0}\left(\sum_{i=1}^{n} X_i > \gamma\right) = \alpha.$$

Напомним, что если $\xi \sim \Gamma(\alpha, \lambda)$, то

$$\psi_\xi(t) = \frac{\lambda^\alpha}{(\lambda - t)^\alpha}.$$

В частности, поскольку $X_i \sim \Gamma(1, \theta_0)$ относительно P_{θ_0}, то

$$\psi_{X_1}(t) = \mathsf{E}_{\theta_0} e^{tX_1} = \frac{\theta_0}{\theta_0 - t}.$$

Отсюда следует, что

$$\psi_{2\theta_0 X_1}(t) = \mathsf{E}_{\theta_0} e^{t2\theta_0 X_1} = \psi_{X_1}(t2\theta_0) = \frac{1}{1 - 2t} = \frac{1/2}{1/2 - t}.$$

Таким образом, $2\theta_0 X_1 \sim \Gamma(1, 1/2)$,

$$2\theta_0 \sum_{i=1}^{n} X_i \sim \Gamma(n, 1/2) = \chi^2(2n),$$

так как $\chi^2(n) = \Gamma(n/2, 1/2)$. Из условия

$$\mathsf{P}_{\theta_0}\left(\sum_{i=1}^{n} X_i > \gamma\right) = 1 - \mathsf{P}_{\theta_0}\left(2\theta_0 \sum_{i=1}^{n} X_i \le 2\theta_0\gamma\right) = \alpha$$

находим, что

$$2\theta_0\gamma = \chi^2_{\alpha,2n},$$

где $\mathsf{P}(V \le \chi^2_{\alpha,2n}) = 1 - \alpha$, $V \sim \chi^2(2n)$.
Окончательный вид оптимального критерия уровня α:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \displaystyle\sum_{i=1}^{n} x_i > \frac{\chi^2_{\alpha,2n}}{2\theta_0}, \\ 0, & \text{иначе.} \end{cases}$$

Если $\theta_1 > \theta_0$, то условие $\Lambda(\boldsymbol{x}) > K$ принимает вид

$$\sum_{i=1}^{n} x_i < \gamma.$$

Повторяя те же рассуждения, получаем следующий вид оптимального критерия:

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \displaystyle\sum_{i=1}^{n} x_i < \frac{\chi^2_{1-\alpha,n}}{2\theta_0}, \\ 0, & \text{иначе.} \end{cases}$$

Пример. $X_i \sim \mathrm{Ber}(\theta)$,

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1,$$

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0,1\},$$

$$f_n(\boldsymbol{x};\theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} = \theta^S (1-\theta)^{n-S}, \quad \boldsymbol{x} \in \{0,1\}^n,$$

где $S := \sum_{i=1}^n x_i$. Отношение правдоподобия:

$$\Lambda(\boldsymbol{x}) = \frac{f_n(\boldsymbol{x};\theta_1)}{f_n(\boldsymbol{x};\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^S \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-S}$$
$$= \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^S.$$

Пусть $\theta_1 > \theta_0$. Функция $y \mapsto y/(1-y)$ является возрастающей на $(0,1)$. Следовательно,

$$\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} = \frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)} > 1.$$

Отношение правдоподобия является возрастающей функцией S. Условие $\Lambda(\boldsymbol{x}) > K$ принимает вид $S > \gamma$. Для заданного уровня значимости α по лемме Неймана-Пирсона оптимальный (наиболее мощный) критерий имеет вид

$$\psi^*(\boldsymbol{x}) = \begin{cases} 1, & \Lambda(\boldsymbol{x}) > K \\ 0, & \Lambda(\boldsymbol{x}) < K \\ \rho, & \Lambda(\boldsymbol{x}) = K \end{cases} = \begin{cases} 1, & S > \gamma \\ 0, & S < \gamma \\ \rho, & S = \gamma \end{cases}$$

где γ, ρ удовлетворяют уравнению

$$F(\gamma) := \mathsf{P}_{\theta_0}(S \le \gamma) = 1 - \alpha + \rho \mathsf{P}_{\theta_0}(S = \gamma).$$

Здесь мы перешли от Λ, K к S, γ в уравнении из доказательства леммы Неймана-Пирсона.
Если уравнение $F(\gamma) = 1 - \alpha$ имеет решение γ_0, то полагаем $\gamma = \gamma_0$, $\rho = 0$. Иначе найдем γ_0:

$$F_-(\gamma_0) \le 1 - \alpha < F(\gamma_0)$$

и положим

$$\rho_0 = \frac{F(\gamma_0) - (1 - \alpha)}{F(\gamma_0) - F_-(\gamma_0)} = \frac{F(\gamma_0) - (1 - \alpha)}{\mathsf{P}_{\theta_0}(S = \gamma_0)}.$$

При выполнении гипотезы H_0 случайная величина S имеет биномиальное распределение $B(n, \theta_0)$. Найдем ее функцию распределения F при $\theta_0 = 0.5$, $n = 25$.

```
from scipy.stats import binom
import numpy as np
import pandas as pd
theta_0=0.5
n=25
h=np.arange(0,n+1)
F=np.zeros(n+1)
for x in h:
    F[x]=binom.cdf(x,n,theta_0)
df=pd.DataFrame({'x':h,'F(x)':F})
print(df.to_string(index=False))
```

x	F(x)
0	2.980232e-08
1	7.748604e-07
2	9.715557e-06
3	7.826090e-05
4	4.552603e-04
5	2.038658e-03
6	7.316649e-03
7	2.164263e-02
8	5.387607e-02
9	1.147615e-01
10	2.121781e-01
11	3.450190e-01
12	5.000000e-01
13	6.549810e-01
14	7.878219e-01
15	8.852385e-01
16	9.461239e-01
17	9.783574e-01
18	9.926834e-01
19	9.979613e-01
20	9.995447e-01
21	9.999217e-01
22	9.999903e-01
23	9.999992e-01
24	1.000000e+00
25	1.000000e+00

При $\alpha = 0.05$ уравнение $F(\gamma) = 1 - \alpha = 0.95$ неразрешимо. Условию

$$F_{-}(\gamma_0) \leq 1 - \alpha = 0.95 < F(\gamma_0)$$

удовлетворяет $\gamma_0 = 17$. Поскольку

```
binom.pmf(17,n,theta_0)
```

```
0.03223344683647148
```

то

$$\rho_0 = \frac{F(\gamma_0) - (1-\alpha)}{\mathsf{P}_{\theta_0}(S = \gamma_0)} = \frac{0.978 - 0.95}{0.032} = 0.875.$$

Таким образом, критерий Неймана-Пирсона уровня $\alpha = 0.05$ имеет вид

$$\psi(\boldsymbol{x}) = \begin{cases} 1, & \sum_{i=1}^{25} x_i > 17 \\ 0, & \sum_{i=1}^{25} x_i < 17 \\ 0.875, & \sum_{i=1}^{25} x_i = 17 \end{cases}$$

Для сравнения применим критерий Вальда. Согласно ЦПТ,

$$T_n = \sqrt{n}\frac{S/n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} \xrightarrow{d} Z \sim N(0,1).$$

Гипотеза H_0 отклоняется, если T_n принимает большие положительные значения:

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & S > \gamma \\ 0, & S \le \gamma \end{cases}$$

Найдем γ, соответствующее критерию уровня α:

$$\begin{aligned} &\mathsf{P}_{\theta_0}(S > \gamma) = \alpha \\ &= \mathsf{P}_{\theta_0}\left(\sqrt{n}\frac{S/n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} > \sqrt{n}\frac{\gamma/n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}}\right) = \alpha \\ &\approx \mathsf{P}\left(Z > \sqrt{n}\frac{\gamma/n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}}\right) = \alpha. \end{aligned}$$

$$\sqrt{n}\frac{\gamma/n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} = z_\alpha, \quad \gamma = n\left(\theta_0 + z_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right),$$

где z_α — квантиль уровня $1-\alpha$ стандартного нормального распределения.

```
stats.norm(0,1).ppf(0.95)
```

```
1.6448536269514722
```

При $\theta_0 = 0.5$, $n = 25$, $\alpha = 0.05$ имеем

$$\gamma = n\left(\theta_0 + z_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right) = 25 \cdot (0.5 + 1.645 \cdot 0.1) = 16.613.$$

Критерии отношения правдоподобия для сложных гипотез

Статистика отношения правдоподобия:

$$\Lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta} f_n(\boldsymbol{x}; \theta)}{\sup_{\theta \in \Theta_0} f_n(\boldsymbol{x}; \theta)}.$$

Отклонить H_0, если $\Lambda(x)$ велико. Если оптимальные значения функций правдоподобия в ограниченной модели Θ_0 и полной модели Θ близки, то нет оснований отказываться от ограниченной модели в пользу полной.
Размер критерия не должен превосходить уровня значимости α:

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(\Lambda(\boldsymbol{X}) > c) \leq \alpha.$$

Пример. $X_i \sim N(\mu, \sigma^2)$ (оба параметра неизвестны),

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0.$$

$$\Theta = \{\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}, \quad \Theta_0 = \{\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\},$$

$$f_n(x; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} e^{-(x_i-\mu)^2/(2\sigma^2)}.$$

Максимизация по Θ дает оценки максимального правдоподобия, полученные ранее:

$$\widehat{\mu} = \overline{x}, \quad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

$$f_n(x; \widehat{\mu}, \widehat{\sigma}^2) = \frac{1}{(2\pi\widehat{\sigma}^2)^{n/2}} \exp\left(-\frac{1}{2\widehat{\sigma}^2} \sum_{i=1}^{n} (x_i - \overline{x})^2\right) = \frac{1}{(2\pi\widehat{\sigma}^2)^{n/2}} e^{-n/2}.$$

Максимизация по Θ_0:

$$\ln f_n(x;\mu_0,\sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2 + \mathrm{Const},$$

$$\frac{\partial}{\partial\sigma^2}\ln f_n(x;\mu_0,\sigma^2) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu_0)^2 = 0,$$

$$\widehat{\sigma}_0^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2,$$

$$f_n(x;\mu_0,\widehat{\sigma}_0^2) = \frac{1}{(2\pi\widehat{\sigma}_0^2)^{n/2}}\exp\left(-\frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right) = \frac{1}{(2\pi\widehat{\sigma}_0^2)^{n/2}}e^{-n/2}.$$

Отношение правдоподобия:

$$\Lambda(x) = \frac{f_n(x;\widehat{\mu},\widehat{\sigma}^2)}{f_n(x;\mu_0,\widehat{\sigma}_0^2)} = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{n/2}.$$

Поскольку $\sum_{i=1}^{n}(x_i - \overline{x})(\overline{x} - \mu_0) = 0$, то

$$\begin{aligned}
\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2} &= \frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x} + \overline{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= 1 + \frac{n(\overline{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1 + \frac{n(\overline{x} - \mu_0)^2}{(n-1)s^2} = 1 + \frac{T^2}{n-1}, \\
T &= \frac{\sqrt{n}(\overline{x} - \mu_0)}{s} \sim \mathcal{T}(n-1),
\end{aligned}$$

при выполнении гипотезы H_0. Критическая область:

$$\{x : \Lambda(x) > c\} = \{x : |T| > \gamma\}.$$

Поскольку

$$\mathsf{P}_{\mu_0}(|T| > t_{\alpha/2, n-1}) = \alpha,$$

то гипотеза H_0 отвергается на уровне значимости α, если

$$\frac{\sqrt{n}|\overline{x} - \mu_0|}{s} > t_{\alpha/2, n-1}.$$

Аналогичный критерий был получен ранее из других соображений.

Пример. $X_i \sim N(\mu, \sigma^2)$ (оба параметра неизвестны),

$$H_0 : \mu \le \mu_0, \quad H_1 : \mu > \mu_0.$$

$$\Theta = \{\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}, \quad \Theta_0 = \{\mu, \sigma^2) : \mu \le \mu_0, \sigma^2 > 0\},$$

$$f_n(x; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} e^{-(x_i-\mu)^2/(2\sigma^2)}.$$

Как и в предыдущем примере, максимизация по Θ дает оценки максимального правдоподобия:

$$\widehat{\mu} = \overline{x}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

$$f_n(x; \widehat{\mu}, \widehat{\sigma}^2) = \frac{1}{(2\pi\widehat{\sigma}^2)^{n/2}} \exp\left(-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) = \frac{1}{(2\pi\widehat{\sigma}^2)^{n/2}} e^{-n/2}.$$

Максимизация по $\Theta_0 = \{(\mu, \sigma^2) \in (-\infty, \mu_0] \times (0, \infty)\}$:

$$\ln f_n(x; \mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 + \mathrm{Const},$$

$$\frac{\partial}{\partial \mu} \ln f_n(x; \mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = \frac{n}{\sigma^2} (\overline{x} - \mu),$$

$$\widehat{\mu}_0 := \begin{cases} \overline{x}, & \overline{x} < \mu_0 \\ \mu_0, & \overline{x} \geq \mu_0 \end{cases} = \min\{\overline{x}, \mu_0\}.$$

при любом σ^2. Остается максимизировать

$$\ln f_n(x; \widehat{\mu}_0, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \widehat{\mu}_0)^2 + \mathrm{Const}.$$

Как и в предыдущем примере, получаем

$$\widehat{\sigma}_0^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i - \widehat{\mu}_0)^2,$$

$$f_n(x;\widehat{\mu}_0,\widehat{\sigma}_0^2) = \frac{1}{(2\pi\widehat{\sigma}_0^2)^{n/2}} \exp\left(-\frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^n (x_i-\widehat{\mu}_0)^2\right) = \frac{1}{(2\pi\widehat{\sigma}_0^2)^{n/2}} e^{-n/2}.$$

Отношение правдоподобия внешне выглядит так же, как и в предыдущем примере:

$$\Lambda(x) = \frac{f_n(x;\widehat{\mu},\widehat{\sigma}^2)}{f_n(x;\widehat{\mu}_0,\widehat{\sigma}_0^2)} = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{n/2}.$$

Условие $\Lambda(x) > c$ можно переписать в виде

$$\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2} = \frac{\sum_{i=1}^n (x_i - \min\{\overline{x},\mu_0\})^2}{\sum_{i=1}^n (x_i-\overline{x})^2} > c_1.$$

При $\overline{x} \le \mu_0$ имеем $\Lambda(x) = 1$. Но $\Lambda(x) \ge 1$ при всех x. Следовательно, гипотеза H_0 не может отвергаться при этом условии. В противном случае она отвергалась бы всегда.

Остается рассмотреть случай $\overline{x} > \mu_0$. При этом мы получаем ту же статистику, что и в предыдущем примере:

$$\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2} = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = 1 + \frac{T^2}{n-1}, \quad T = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s},$$

но теперь нужно рассматривать $\mu \in (-\infty, \mu_0]$, а $T \sim \mathcal{T}(n-1)$ только при $\mu = \mu_0$. Критическая область:

$$\{|T| I_{\{\overline{X} > \mu_0\}} > \gamma\} = \left\{ \frac{\sqrt{n}|\overline{X} - \mu_0|}{S} I_{\{\overline{X} > \mu_0\}} > \gamma \right\} = \left\{ \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} > \gamma \right\}$$
$$= \left\{ \frac{\sqrt{n}(\overline{X} - \mu)}{S} > \gamma + \frac{\sqrt{n}(\mu_0 - \mu)}{S} \right\}$$

Выражение

$$\mathsf{P}_\mu(\{|T| I_{\{\overline{X} > \mu_0\}}) = \mathsf{P}_\mu\left(T > \gamma + \frac{\sqrt{n}(\mu_0 - \mu)}{S} \right), \quad T \sim \mathcal{T}(n-1)$$

возрастает по μ.

Следовательно,

$$\mathsf{P}_{\mu}(\{|T|I_{\{\overline{X}>\mu_0\}}) \leq \mathsf{P}_{\mu_0}(\{|T|I_{\{\overline{X}>\mu_0\}}) = \mathsf{P}_{\mu_0}(T > \gamma).$$

При $\gamma = t_{\alpha,n-1}$ получаем критерий размера α. Гипотеза H_0 отвергается на уровне значимости α, если

$$T = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s} > t_{\alpha,n-1}.$$

Асимптотическое распределение отношения правдоподобия

- Пусть оценка максимального правдоподобия является асимптотически нормальной:

  $$\sqrt{n}I^{1/2}(\theta)(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0,1).$$

  Тогда при выполнении некоторых технических условий отношение правдоподобия Λ_n обладает следующим свойством:

  $$2\ln\Lambda_n \xrightarrow{d} V \sim \chi^2(1),$$

  если гипотеза $H_0 : \theta = \theta_0$ верна.

Доказательство. Пусть $l(x;\theta) = \ln f(x;\theta)$. Тогда

$$\begin{aligned}
\ln \Lambda_n &= -\sum_{i=1}^{n} (l(X_i;\theta_0) - l(X_i;\widehat{\theta}_n)) \\
&= -(\theta_0 - \widehat{\theta}_n) \sum_{i=1}^{n} l_\theta(X_i;\widehat{\theta}_n) - \frac{1}{2}(\theta_0 - \widehat{\theta}_n)^2 \sum_{i=1}^{n} l_{\theta\theta}(X_i;\widetilde{\theta}) \\
&= -\frac{n}{2}(\theta_0 - \widehat{\theta}_n)^2 \frac{1}{n} \sum_{i=1}^{n} l_{\theta\theta}(X_i;\widetilde{\theta}),
\end{aligned}$$

где $\widetilde{\theta}_n$ лежит между θ_0 и $\widehat{\theta}_n$. Пусть верна гипотеза H_0. Тогда

$$\frac{1}{n} \sum_{i=1}^{n} l_{\theta\theta}(X_i;\widetilde{\theta}) \xrightarrow{d} \mathsf{E}_{\theta_0} l_{\theta\theta}(X_1;\theta_0) = -I(\theta_0)$$

по закону больших чисел, и

$$n(\theta_0 - \widehat{\theta}_n)^2 I(\theta_0) \to Z^2 \sim \chi^2(1)$$

по теореме о непрерывном преобразовании. Надо сослаться также на теорему Слуцкого. $\square$

Рассмотрим случай многомерного параметра $\theta = (\theta_1, \dots, \theta_d)$. Пусть

$$H_0 : \theta_1 = \theta_{10}, \dots, \theta_r = \theta_{r0}.$$

- Пусть оценка максимального правдоподобия является асимптотически нормальной:

  $$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \sim N(0, I^{-1}(\theta)).$$

  Тогда при выполнении некоторых технических условий отношение правдоподобия Λ_n обладает следующим свойством:

  $$2 \ln \Lambda_n \xrightarrow{d} V \sim \chi^2(r),$$

  если гипотеза H_0 верна.

Заметим, что

$$\dim \Theta_0 = d - r, \quad \dim \Theta - \dim \Theta_0 = d - (d - r) = r,$$

т.е. r равно разности между размерностью модели и размерностью подмодели. При $r = d$ доказательство предыдущей теоремы легко адаптируется.

Закон Менделя

Семена гороха имеют два признака: цвет и форму. Цвет может быть желтым (Y: yellow) и зеленым (g: green). Форма может быть гладкой (R: round) и морщинистой (w: wrinkled). Множество генотипов описывается декартовым произведением

$$\{YY, Yg, gg\} \times \{RR, Rw, ww\}.$$

Доминантному признаку соответствует большая буква. Например, для генотипа (Yg, Rw) мы увидим фенотип YR.
При скрещивании двух растений случайно (с одинаковой вероятностью) выбирается один элемент из каждой пары, описывающей признаки (независимо для каждого признака). Скрещивая (YY, RR) и (gg, ww), Мендель получил (Yg, Rw).

Результат скрещивания полученных растений (Yg, Rw) (гибридов) между собой можно описать двумерной случайной величиной (ξ, η) с независимыми компонентами, имеющими следующие распределения:

x	YY	Yg	gg
$p_\xi(x)$	$1/4$	$1/2$	$1/4$

y	RR	Rw	ww
$p_\eta(y)$	$1/4$	$1/2$	$1/4$

Найдем вероятности получения каждого фенотипа:

$$p_1 = \mathsf{P}(YR) = \mathsf{P}(\xi \in \{YY, Yg\}, \eta \in \{RR, Rw\}\} = \frac{3}{4}\frac{3}{4} = \frac{9}{16},$$
$$p_2 = \mathsf{P}(Yw) = \mathsf{P}(\xi \in \{YY, Yg\}, \eta = ww\} = \frac{3}{4}\frac{1}{4} = \frac{3}{16},$$
$$p_3 = \mathsf{P}(gR) = \mathsf{P}(\xi = g, \eta \in \{RR, Rw\}\} = \frac{1}{4}\frac{3}{4} = \frac{3}{16},$$
$$p_4 = \mathsf{P}(gw) = \mathsf{P}(\xi = gg, \eta = ww\} = \frac{1}{4}\frac{1}{4} = \frac{1}{16}.$$

В 556 опытах были получены следующие количества растений соответствующих фенотипов: $(315, 101, 108, 32)$. Возьмем закон Менделя в качестве нулевой гипотезы:

$$H_0 : p = p^0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right), \quad H_1 : p \neq p_0.$$

Найдем оценку максимального правдоподобия для полной модели $\mathsf{P}(X = z_j) = p_j$, $j = 1, \dots, k$ $(k = 4)$. Пусть дана выборка $x_i \in \{z_1, \dots, z_k\}$, $i = 1, \dots, n$. Функция правдоподобия:

$$L_n(x; p) = \prod_{i=1}^{n} p_1^{\nu_1} \dots p_k^{\nu_k}, \quad \nu_j = \sum_{i=1}^{n} I_{\{x_i = z_j\}}.$$

$$\ln L_n(x; p) = \nu_1 \ln p_1 + \dots + \nu_d \ln p_d \to \max,$$

$$p_1 + \dots + p_d = 1, \quad p \geq 0.$$

$$L(p, \lambda) = \sum_{j=1}^{k} \nu_j \ln p_j + \lambda \left(1 - \sum_{j=1}^{k} p_j\right),$$

$$\frac{\partial L}{\partial p_j} = \frac{\nu_j}{p_j} - \lambda = 0, \quad p_j = \frac{\nu_j}{\lambda},$$

$$\sum_{j=1}^{k} p_j = \frac{1}{\lambda} \sum_{j=1}^{k} \nu_j = \frac{n}{\lambda} = 1, \quad \lambda = n.$$

Оценка максимального правдоподобия: $\widehat{p}_j = \nu_j/n$. Отношение правдоподобия:

$$\Lambda_n(x) = \frac{\sup_{\theta\in\Theta} L_n(x;\theta)}{\sup_{\theta\in\Theta_0} L_n(x;\theta)} = \frac{L_n(x;\widehat{p})}{L_n(x;p_0)},$$

$$\begin{aligned}
\ln\Lambda_n(x) &= \ln L_n(x;\widehat{p}) - \ln L_n(x;p_0) \\
&= \nu_1 \ln\frac{\nu_1}{n} + \dots + \nu_k \ln\frac{\nu_k}{n} - \nu_1 \ln p_1^0 - \dots - \nu_d \ln p_1^0 \\
&= \nu_1 \ln\frac{\nu_1}{np_1^0} + \dots + \nu_k \ln\frac{\nu_k}{np_k^0}.
\end{aligned}$$

В нашем случае

$$\nu = (315, 101, 108, 32), \quad p^0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right), \quad n = 556,$$

$$2\Lambda_n(x) = 0.48.$$

Поскольку имеется 4 параметра (p_1, p_2, p_3, p_4) и одно ограничение $p_1 + \cdots + p_4 = 1$, то остается 3 степени свободы в полной модели. В ограниченной модели 0 степеней свободы. Следовательно,

$$2\Lambda_n(x) \xrightarrow{d} V \sim \chi^2(3)$$

при выполнении гипотезы H_0. Найдем приближенное p-значение:

$$p\text{-значение} = \mathsf{P}(V \geq 0.48) = 0.92.$$

Таким образом, нет оснований отвергать гипотезу Менделя.

Критерий согласия χ^2 (Пирсона)

Хотим проверить гипотезу о принадлежности распределения к заданному семейству $\mathcal{P} = (\mathsf{P}_\theta)_{\theta\in\Theta_0}$. Альтернативой является любое другое распределение:

$$H_0 : \mathsf{P}_{X_i} \in \mathcal{P}, \quad H_1 : \mathsf{P}_{X_i} \notin \mathcal{P}.$$

Разобьем область значений S рассматриваемой случайной величины на непересекающиеся подмножества:

$$S = \bigcup_{j=1}^{k} A_j, \quad A_i \cap A_j = \emptyset, i \neq j,$$

и заменим основную и альтернативную гипотезы следующим образом:

$$\begin{aligned} H_0 &: \mathsf{P}(X_i \in A_j) = p_j(\theta),\ j = 1, \dots, k,\ \theta \in \Theta_0, \\ H_1 &: \mathsf{P}(X_i \in A_j) = p_j(\theta),\ j = 1, \dots, k,\ \theta \in \Theta_1 := \Delta_k \backslash \Theta_0, \end{aligned}$$

Здесь $\Delta_k = \{p \geq 0 : \sum_{j=1}^k p_j = 1\}$.

Для простой гипотезы H_0 эта задача, фактически, анализировалась в предыдущем примере. Критерий отношения правдоподобия:

$$\Lambda_n(x) = \frac{\sup_{\theta\in\Delta_k} L_n(x;\theta)}{\sup_{\theta\in\Theta_0} L_n(x;\theta)} = \sum_{j=1}^{k} \nu_j \ln\frac{\nu_j}{np_j^0}, \quad \nu_j = \sum_{i=1}^{n} I_{\{X_i\in A_j\}},$$

$$2\ln\Lambda_n \xrightarrow{d} V \sim \chi^2(k-1).$$

По формуле Тейлора

$$f(z) = z\ln\frac{z}{z_0} = z - z_0 + \frac{1}{2z_0}(z-z_0)^2 + o((z-z_0)^2),$$

так как $f(z_0) = 0$,

$$f'(z) = \ln\frac{z}{z_0} + z\frac{1}{z/z_0}\frac{1}{z_0} = \ln\frac{z}{z_0} + 1, \quad f'(z_0) = 1,$$

$$f''(z) = 1/z.$$

Полагая $z = \nu_j/n$, $z_0 = p_j^0$, находим

$$\Lambda_n = n\sum_{j=1}^{k}\frac{\nu_j}{n}\ln\frac{\nu_j/n}{p_j^0} \approx n\sum_{j=1}^{k}\left(\frac{\nu_j}{n} - p_j^0 + \frac{1}{2p_j^0}\left(\frac{\nu_j}{n} - p_j^0\right)^2\right)$$

$$= \frac{1}{2}Q_n, \quad Q_n = \sum_{j=1}^{k}\frac{(\nu_j - np_j^0)^2}{np_j^0}.$$

Можно показать, что $2\Lambda_n - Q_n \xrightarrow{\mathsf{P}} 0$. Следовательно,

$$Q_n \xrightarrow{d} V \sim \chi^2(k-1).$$

Пример. Для рассмотренного опыта Менделя $Q_n = 0.47$,

$$p\text{-значение} = \mathsf{P}(V \geq 0.47) = 0.93.$$

Эти значения очень близки к тем, которые получены для критерия отношения правдоподобия.

Ross (2021), Example 11.2.a

Верно ли, что известные люди с меньшей вероятностью умирают незадолго до своего рождения, чем после него? (Ожидание дня рождения поднимает настроение и улучшает здоровье (?))
Данные об умерших людях случайным образом выбраны из Who Was Who in America. D. Phillips, "Death Day and Birthday: An Unexpected Connection", Statistics: A Guide to the Unknown, Holden-Day, 1972.

Месяц до дня рождения

-6	-5	-4	-3	-2	-1	0	1	2	3	4	5
90	100	87	96	101	86	119	118	121	114	113	106

$$H_0 : p_0 = \frac{1}{12}, \; i = 1, \ldots, 12.$$

$n = 1251$, $k = 12$, $Q_n = 17.192$, $V \sim \chi^2(11)$,

$$p\text{-значение} = \mathsf{P}(V \geq 17.192) = 0.102.$$

Нет свидетельства (или очень слабое свидетельство) против нулевой гипотезы об отсутствии влияния даты рождения на дату смерти.

Уменьшим количество категорий до 4-х:

“1“:	-6,-5,-4
“2“:	-3,-2,-1
“3“:	0, 1, 2
“4“:	3, 4, 5

С этой классификацией

“1“:	277
“2“:	283
“3“:	358
“4“:	333

$$H_0 : p_0 = 1/4,\ i = 1, \ldots, 4.$$

$n = 1251$, $k = 4$, $Q_n = 14.775$, $V \sim \chi^2(3)$,

$$p\text{-значение} = \mathsf{P}(V \geq 14.775) = 0.002.$$

Очень сильное свидетельство против H_0. Но нулевая гипотеза была выбрана после наблюдения данных (присутствует элемент манипуляции с данными).

Knight (2001), Example 9.2

Количество голов забитых в матчах английского первого дивизиона, 1978-1979.

Кол-во голов	0	1	2	3	4	≥ 5
Кол-во матчей	252	344	180	104	28	16

Подчиняется ли количество голов распределению Пуассона?

$$H_0 : p_j(\lambda) = e^{-\lambda}\frac{\lambda^j}{j!},\ j = 0, \dots, 4; \quad p_5(\lambda) = 1 - \sum_{j=0}^{4} p_j(\lambda).$$

Оценка максимального правдоподобия для полной модели с шестью возможными значениями: $\widehat{p}_j = \nu_j/n$, где ν_j берутся из таблицы, и $n = 924$ — общее количество матчей.

Оценка максимального правдоподобия для подмодели может быть найдена численно:

$$\prod_{i=1}^{n}\prod_{j=0}^{5} p_j(\lambda)^{\nu_j} \to \max_{\lambda>0} \Longrightarrow \widehat{\lambda} = 1.3118.$$

Статистики отношения правдоподобия и χ^2:

$$2\Lambda_n = 2\sum_{j=0}^{5}\nu_j \ln\left(\frac{\nu_j}{np_j(\widehat{\lambda}))}\right) = 10.87,$$

$$Q_n = \sum_{j=0}^{5}\frac{(\nu_j - np_j(\widehat{\lambda}))^2}{np_j(\widehat{\lambda})} = 11.09.$$

Размерность модели: 5, размерность подмодели: 1. Пусть $V \sim \chi^2(4)$, тогда

$$p\text{-value} = \mathsf{P}(V \geq 10.87) = 0.028$$

для критерия отношения правдоподобия,

$$p\text{-value} = \mathsf{P}(V \geq 11.09) = 0.026$$

для критерия χ^2 (Пирсона). Сильное свидетельство против модели Пуассона.

Критерий согласия Колмогорова-Смирнова

Пусть X_i имеют неизвестную функцию распределения F и пусть F_0 — непрерывная функция распределения.

$$H_0 : F = F_0, \quad H_1 : F \neq F_0.$$

Сильный закон больших чисел:

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq t\}} \overset{\text{п.н.}}{\to} F(t) = \mathsf{P}(X_1 \leq t)$$

Центральная предельная теорема:

$$\sqrt{n}(\widehat{F}_n(t) - F(t)) \overset{d}{\to} N(0, F(t)(1 - F(t)),$$

Теорема Гливенко-Кантелли:

$$\sup_{t\in\mathbb{R}} |\widehat{F}_n(t) - F(t)| \overset{\text{п.н.}}{\to} 0.$$

Теорема Колмогорова:

$$\sqrt{n}\sup_{t\in\mathbb{R}}|\widehat{F}_n(t)-F(t)|\xrightarrow{d}\eta,$$

$$\mathsf{P}(\eta>x)=2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2x^2}.$$

Статистика Колмогорова-Смирнова:

$$T_n:=\sqrt{n}\sup_{t\in\mathbb{R}}|\widehat{F}_n(t)-F_0(t)|.$$

Критерий Колмогорова-Смирнова определяется критической областью

$$R_\alpha=\{x:T_n(x)>q_\alpha\}.$$

Он имеет асимптотический размер α, если q_α является $(1-\alpha)$-квантилью распределения η:

$$\mathsf{P}_F(\eta\geq q_\alpha)=\alpha.$$

- Функция мощности критерия Колмогорова-Смирнова (определенная на множестве всех функций распределения) стремится к 1 при $F \neq F_0$, т.е. если верна H_1.

Действительно, пусть X_i имеют функцию распределения $F_1 \neq F_0$, и пусть $F_1(t_0) \neq F_0(t_0)$. По усиленному закону больших чисел, $\widehat{F}_n(t_0) \to F_1(t_0)$ п.н. Следовательно,

$$T_n = \sqrt{n} \sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F_0(t)| \geq \sqrt{n} |\widehat{F}_n(t_0) - F_0(t_0)| \stackrel{\text{п.н.}}{\to} \infty,$$

$$\mathsf{P}_{F_1}(T_n \geq q_\alpha) \to 1,$$

а вероятность ошибки второго рода стремится к нулю:

$$\mathsf{P}_{F_1}(T_n < q_\alpha) \to 0.$$

- Распределение T_n (а не только предельное распределение η) не зависит от F.

Доказательство.

$$\frac{T_n}{\sqrt{n}} = \sup_{t\in\mathbb{R}} \left| \sum_{i=1}^{n} \frac{1}{n} I_{\{X_i \le t\}} - F(t) \right| = \sup_{t\in\mathbb{R}} \left| \sum_{i=1}^{n} \frac{1}{n} I_{\{F(X_i) \le F(t)\}} - F(t) \right|$$
$$= \sup_{u\in[0,1]} \left| \sum_{i=1}^{n} \frac{1}{n} I_{\{F(X_i) \le u\}} - u \right|.$$

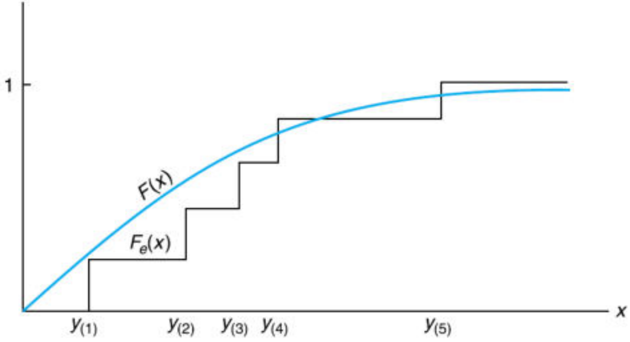
Но $F(X_i) \sim U(0,1)$:

$$\mathsf{P}_F(F(X_i) \le v) = \mathsf{P}_F(X_i \le F^{-1}(v)) = F(F^{-1}(v)) = v. \quad \square$$

Вычисление статистики Колмогорова-Смирнова:

$$T_n = \sqrt{n} \max_{1 \le i \le n} \max \left\{ \frac{i-1}{n} - F_0(X_{(i)}), \frac{i}{n} - F_0(X_{(i)}) \right\},$$

поскольку F_0 является неубывающей, а $\widehat{F}_n$ — кусочно-постоянной со скачками величины $1/n$ в точках X_i.



```python
from scipy.stats import kstwo
import pandas as pd
import numpy as np
ks_table=pd.DataFrame(columns=['n','alpha=0.1','alpha=0.05','alpha=0.01'])
n=np.arange(1,30)
ks_table['n']=n
ks_table['alpha=0.1']=kstwo.ppf(0.9,n)*np.sqrt(n)
ks_table['alpha=0.05']=kstwo.ppf(0.95,n)*np.sqrt(n)
ks_table['alpha=0.01']=kstwo.ppf(0.99,n)*np.sqrt(n)
print(ks_table.to_string(index=False))
```

Квантили распределения Колмогорова-Смирнова

n	alpha=0.1	alpha=0.05	alpha=0.01
1	0.950000	0.975000	0.995000
2	1.097986	1.190607	1.314214
3	1.101662	1.225596	1.435874
4	1.130432	1.247877	1.468476
5	1.139163	1.259522	1.494881
6	1.146345	1.271927	1.510373
7	1.153729	1.279020	1.523456
8	1.158586	1.284860	1.532421
9	1.162389	1.290033	1.539952
10	1.165811	1.294150	1.546138
11	1.168840	1.297542	1.551193
12	1.171383	1.300527	1.555539
13	1.173561	1.303163	1.559304
14	1.175508	1.305469	1.562575
15	1.177273	1.307504	1.565458
16	1.178867	1.309334	1.568029
17	1.180305	1.310994	1.570336
18	1.181610	1.312504	1.572416
19	1.182806	1.313881	1.574304
20	1.183911	1.315145	1.576029
21	1.184935	1.316310	1.577612
22	1.185886	1.317391	1.579070
23	1.186771	1.318395	1.580418
24	1.187597	1.319331	1.581670
25	1.188372	1.320207	1.582835
26	1.189099	1.321028	1.583924
27	1.189785	1.321800	1.584943
28	1.190433	1.322527	1.585900
29	1.191046	1.323214	1.586801

Квантили предельного распределения Колмогорова-Смирнова

```
#kstwobign: Kolmogorov-Smirnov two-sided big n
from scipy.stats import kstwobign
print(f'alpha=0.1: {kstwobign.ppf(0.9):.4f}\nalpha=0.05: {kstwobign.ppf(0.95):.4f}
      \nalpha=0.01: {kstwobign.ppf(0.99):.4f}')
```

```
alpha=0.1: 1.2238
alpha=0.05: 1.3581
alpha=0.01: 1.6276
```

Пример использования статистики Колмогорова-Смирнова

Testing that given distribution is standard normal

```
import numpy as np
from scipy import stats
#standard normal
np.random.seed(987654321)
stats.kstest(stats.norm.rvs(size=100), stats.norm.cdf)
```

```
KstestResult(statistic=0.058352892479417884, pvalue=0.8853119094415125)
```

```
# Student with 3 degrees of reedom
np.random.seed(42)
stats.kstest(stats.t.rvs(3, size=1000), stats.norm.cdf)
```

```
KstestResult(statistic=0.0510284487475544, pvalue=0.010561290731722296)
```

```
# Student with 1000 degrees of reedom
np.random.seed(42)
stats.kstest(stats.t.rvs(100, size=1000), stats.norm.cdf)
```

```
KstestResult(statistic=0.0398366578940299, pvalue=0.08145099694781682)
```

Пример использования статистики Колмогорова-Смирнова

Testing that given distribution is uniform

```
np.random.seed(987654321)
stats.kstest(stats.uniform.rvs(size=1000), stats.uniform.cdf)
```

```
KstestResult(statistic=0.03180749522953352, pvalue=0.2588514927573616)
```

Линейная регрессия

$$Y_i = x_i^T \beta + \varepsilon_i,$$

$x_i = (1, x_{i1}, \dots, x_{id})$: известные детерминированные величины, $\beta = (\beta_0, \dots, \beta)^T$: неизвестные параметры регрессии, $\varepsilon_1, \dots, \varepsilon_n \sim N(0, \sigma^2)$: независимые случайные величины. Дисперсия σ^2 неизвестна. Другими словами,

$$Y_i \sim N(x_i^T \beta, \sigma^2)$$

— независимые случайные величины.
Функция правдоподобия (совместная плотность):

$$L(\beta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2\sigma^2}\right).$$

Метод максимального правдоподобия:

$$\ln L(\beta, \sigma) = n \ln \frac{1}{\sqrt{2\pi}} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \to \max_{\beta \in \mathbb{R}^{d+1}, \sigma > 0}.$$

$$\mathbb{X} = \begin{pmatrix} x_{1,0} & \cdots & x_{1,d} \\ \vdots & \vdots & \vdots \\ x_{n,0} & \cdots & x_{n,d} \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

$$\begin{aligned} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 &= \|y - \mathbb{X}\beta\|^2 = \|y\|^2 - 2\langle y, \mathbb{X}\beta \rangle + \langle \mathbb{X}\beta, \mathbb{X}\beta \rangle \\ &= \|y\|^2 - 2\langle \mathbb{X}^T y, \beta \rangle + \langle \mathbb{X}^T \mathbb{X}\beta, \beta \rangle \end{aligned}$$

Пусть $A = \mathbb{X}^T \mathbb{X}$. Тогда

$$\begin{aligned} \frac{\partial}{\partial \beta_k} \langle A\beta, \beta \rangle &= \frac{\partial}{\partial \beta_k} \sum_{i,j=0}^{d} a_{ij} \beta_i \beta_j = \sum_{i,j=0}^{d} a_{ij} \frac{\partial \beta_i}{\partial \beta_k} \beta_j + \sum_{i,j=0}^{d} a_{ij} \beta_i \frac{\partial \beta_j}{\partial \beta_k} \\ &= \sum_{j=0}^{d} a_{kj} \beta_j + \sum_{i=0}^{d} a_{ik} \beta_i = \sum_{j=0}^{d} a_{kj} \beta_j + \sum_{i=0}^{d} a_{ki} \beta_i = 2(A\beta)_k. \end{aligned}$$

Кроме того, $\frac{\partial}{\partial \beta_k} \langle \mathbb{X}^T y, \beta \rangle = (\mathbb{X}^T y)_k$.

Таким образом,

$$\nabla_\beta \|y - \mathbb{X}\beta\|^2 = -2\mathbb{X}^T y + 2\mathbb{X}^T\mathbb{X}\beta,$$

$$\nabla_\beta \ln L(\beta, \sigma) = -\frac{1}{\sigma^2}\left(-\mathbb{X}^T y + \mathbb{X}^T\mathbb{X}\beta\right).$$

Условие оптимальности $\nabla_\beta \ln L = 0$ приводит к *нормальным уравнениям*:

$$\mathbb{X}^T\mathbb{X}\beta = \mathbb{X}^T y.$$

Обозначим решение этой системы через $\widehat{\beta}$. Второе условие оптимальности

$$\frac{\partial}{\partial\sigma} \ln L(\beta, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\|y - \mathbb{X}\beta\|^2 = 0$$

дает оценку σ^2:

$$\widehat{\sigma}^2 = \frac{1}{n}\|y - \mathbb{X}\widehat{\beta}\|^2.$$

$\widehat{\beta}$ является решением системы линейных уравнений $\mathbb{X}\beta = y$ по методу наименьших квадратов: $\|y - \mathbb{X}\widehat{\beta}\|^2 \to \min_\beta$. В подпространстве

$$\operatorname{im} X = \{\mathbb{X}\beta : \beta \in \mathbb{R}^{d+1}\} \subset \mathbb{R}^n$$

мы хотим найти элемент, ближайший к y. Такой элемент $\mathbb{H}$ существует, единственен и называется проекцией y на $\operatorname{im} X$. Вектор $\widehat{\beta}$, удовлетворяющий условию $\mathbb{H}y = \mathbb{X}\widehat{\beta}$ всегда существует, но не обязательно единственен.
Решение $\widehat{\beta}$ единственно при выполнении любого из эквивалентных условий

- $\ker(\mathbb{X}^T\mathbb{X}) = \{0\}$
- $\ker(\mathbb{X}) = \{0\}$
- столбцы $x^{(0)}, \ldots, x^{(d)}$ матрицы $\mathbb{X}$ независимы
- $\operatorname{rank}(\mathbb{X}) = d + 1$ (в предположении $n \geq d + 1$)

В любом из этих случаев матрица $\mathbb{X}^T\mathbb{X}$ положительно определена и

$$\widehat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T y.$$

Пусть $\mathbb{X}^T\mathbb{X}$ обратима. Тогда

$$\widehat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T y.$$

Предсказанные значения:

$$\widehat{y}_i = x_i^T \widehat{\beta}.$$

В матричной форме:

$$\widehat{y} = \mathbb{X}\widehat{\beta} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T y = \mathbb{H}y.$$

Оценка максимального правдоподобия для σ^2:

$$\widehat{\sigma}^2 = \frac{1}{n}\|y - \mathbb{X}\widehat{\beta}\|^2 = \frac{1}{n}\|(I - \mathbb{H})y\|^2.$$

- $\widehat{\beta} \sim N_{d+1}(\beta, \sigma^2(\mathbb{X}^T\mathbb{X})^{-1})$

Поскольку $Y \sim N_{d+1}(\mathbb{X}\beta, \sigma^2 I)$, то

$$\begin{aligned}
\widehat{\beta} &= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T Y = AY \sim N_{d+1}(A\mathbb{X}\beta, \sigma^2 AA^T),\\
A\mathbb{X}\beta &= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\beta = \beta,\\
AA^T &= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1} = (\mathbb{X}^T\mathbb{X})^{-1}.
\end{aligned}$$

- $n\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n-d-1)$

$$n\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\|(I-\mathbb{H})Y\|^2,$$

$$Y - \mathbb{H}Y = \mathbb{X}\beta + \varepsilon - \mathbb{H}(\mathbb{X}\beta + \varepsilon) = (I-\mathbb{H})\varepsilon,$$

так как $\mathbb{X}$ является проекцией на $\operatorname{im}\mathbb{X}$. Поскольку $I-\mathbb{H}$ является проектором на подпространство $(\operatorname{im}\mathbb{H})^\perp$ размерности $n-d-1$, то

$$n\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\|(I-\mathbb{H})\varepsilon\|^2 = \frac{1}{\sigma^2}\langle (I-\mathbb{H})\varepsilon, \varepsilon\rangle \sim \chi^2(n-d-1).$$

- $\widehat{\beta}$ и $\widehat{\sigma}^2$ независимы.

Покажем, что

$$\widehat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T Y = AY,$$

$$Y - \mathbb{X}\widehat{\beta} = (I - \mathbb{H})Y = BY$$

независимы:

$$\begin{pmatrix} A \\ B \end{pmatrix} Y = \begin{pmatrix} A \\ B \end{pmatrix} \mathbb{X}\beta + \begin{pmatrix} A \\ B \end{pmatrix} \varepsilon,$$

$$\begin{pmatrix} A \\ B \end{pmatrix} \varepsilon \sim N\left(0, \sigma^2 \begin{pmatrix} A \\ B \end{pmatrix} \begin{pmatrix} A^T B^T \end{pmatrix}\right) = N\left(0, \sigma^2 \begin{pmatrix} AA^T & AB^T \\ BA^T & BB^T \end{pmatrix}\right),$$

$$AB^T = AB = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(I - \mathbb{H}) = 0,$$

так как $\mathbb{X}^T\mathbb{H} = (\mathbb{H}^T\mathbb{X})^T = (\mathbb{H}\mathbb{X})^T = \mathbb{X}^T$.

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0.$$

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2 \gamma_j), \quad \gamma_j = (\mathbb{X}^T \mathbb{X})^{-1}_{jj}.$$

Положим

$$SS_R = \|Y - \mathbb{X}\widehat{\beta}\|^2.$$

При выполнении H_0,

$$\frac{\widehat{\beta}_j}{\sigma\sqrt{\gamma_j}} \sim N(0,1),$$

$$T_n = \sqrt{n-d-1}\frac{\widehat{\beta}_j}{\sqrt{SS_R\gamma_j}} = \frac{\widehat{\beta}_j}{\sigma\sqrt{\gamma_j}}\frac{1}{\sqrt{\dfrac{SS_R}{\sigma^2(n-d-1)}}} \sim \mathcal{T}(n-d-1),$$

так как

$$\frac{SS_R}{\sigma^2} = n\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-d-1),$$

и SS_R не зависит от $\widehat{\beta}_j$.

H_0 отвергается на уровне α, если

$$\mathsf{P}(|T_n| > t_{\alpha/2,n-d-1}) = 2\mathsf{P}(T_n > t_{\alpha/2,n-d-1}),$$

где $t_{\alpha/2,n-d-1}$ — квантиль уровня $1-\alpha/2$ распределения $\mathcal{T}(n-d-1)$,

$$p\text{-значение} = \mathsf{P}(|T_n| > v_j) = 2\mathsf{P}(T_n > v_j),$$

где v_j — значение T_n на заданном наборе данных.
Найдем $(1-\alpha)$-доверительный интервал для β_j:

$$\sqrt{n-d-1}\,\frac{\widehat{\beta}_j - \beta_j}{\sqrt{SS_R\gamma_j}} \sim \mathcal{T}(n-d-1),$$

$$1-\alpha = \mathsf{P}_\beta\left(-t_{\alpha/2,n-d-1} < \sqrt{n-d-1}\,\frac{\widehat{\beta}_j - \beta_j}{\sqrt{SS_R\gamma_j}} < t_{\alpha/2,n-d-1}\right)$$

$$= \mathsf{P}_\beta\left(\widehat{\beta}_j - t_{\alpha/2,n-d-1}\frac{\sqrt{SS_R\gamma_j}}{\sqrt{n-d-1}} < \beta_j < \widehat{\beta}_j + t_{\alpha/2,n-d-1}\frac{\sqrt{SS_R\gamma_j}}{\sqrt{n-d-1}}\right).$$

Пусть $d = 2$:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad x_i \in \mathbb{R}.$$

$$SS_R = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Нормальные уравнения:

$$\frac{\partial SS_R}{\partial \beta_0} = 0 \Longrightarrow \beta_0 + \beta_1 \overline{x} = \overline{y},$$

$$\frac{\partial SS_R}{\partial \beta_1} = 0 \Longrightarrow \sum_{i=1}^{n} x_i \beta_0 + \sum_{i=1}^{n} x_i^2 \beta_1 = \sum_{i=1}^{n} x_i y_i.$$

Решение:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \quad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}.$$

$$\mathbb{X}^T\mathbb{X} = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

$$= n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}$$

$$(\mathbb{X}^T\mathbb{X})^{-1} = \frac{1}{n} \frac{1}{\overline{x^2} - (\overline{x})^2} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

$$\gamma_0 = (\mathbb{X}^T\mathbb{X})^{-1}_{00} = \frac{\overline{x^2}}{S_{xx}}, \quad \gamma_1 = (\mathbb{X}^T\mathbb{X})^{-1}_{11} = \frac{1}{S_{xx}},$$

$$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2 = n(\overline{x^2} - (\overline{x})^2).$$

Bertsekas, Tsitsiklis, Example 9.9

Отклонение (в метрах) фиксированной точки Пизанской башни от положения, которое соответствует прямой башне.

1975	1976	1977	1978	1979	1980	1981
2.9642	2.9644	2.9656	2.9667	2.9673	2.9688	2.9696

1982	1983	1984	1985	1986	1987	
2.9698	2.9713	2.9717	2.9725	2.9742	2.9757	

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x,$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0.00093, \quad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 1.1233.$$

Проверим гипотезу об отсутствии изменения наклона с течением времени:

$$H_0 : \beta_1 = 0, \quad H_1 : \beta_1 \neq 0.$$

$$SS_R = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta} x_i)^2 = 1.9229, \quad \gamma_1 = \frac{1}{S_{xx}} = 0.0055,$$

$$T_n = \sqrt{n-2} \frac{\widehat{\beta}_1}{\sqrt{SS_R \gamma_1}} = 30.0685,$$

$$p\text{-значение} = 2\mathsf{P}(T_n > v_j) = 2(1 - \mathsf{P}(T_n \leq v_j)) = 6.5 \cdot 10^{-12},$$

где $T_n \sim \mathcal{T}(n-2)$. Гипотеза H_0 отвергается на любом разумном уровне значимости.

Найдем 95% доверительные интервалы для β_0, β_1:

$$\beta_1 \in \left(\widehat{\beta}_1 - t_{\alpha/2, n-2} \frac{\sqrt{SS_R \gamma_1}}{\sqrt{n-2}}, \widehat{\beta}_1 + t_{\alpha/2, n-2} \frac{\sqrt{SS_R \gamma_1}}{\sqrt{n-2}} \right) = (0.00086, 0.00100),$$

$$\beta_0 \in (0.9882, 1.2585).$$

Здесь $\alpha = 0.05$.

Рассмотрим задачу о проверке сложной гипотезы

$$H_0 : \beta_{r+1} = \cdots = \beta_d = 0, \quad H_1 : H_0 \text{ неверна.}$$

Функция правдоподобия:

$$\begin{aligned} L_n(x;\beta) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \prod_{i=1}^n \exp\left(-\frac{(y_i - \langle\beta, x_i\rangle)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\|y - \mathbb{X}\beta\|^2\right), \\ \ln L_n(x;\beta) &= -n\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\|y - \mathbb{X}\beta\|^2. \end{aligned}$$

Метод максимального правдоподобия: $\widehat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T y$ (метод наименьших квадратов),

$$\frac{\partial L_n(x;\beta)}{\partial\sigma} = 0 \Longrightarrow \widehat{\sigma}^2 = \frac{\mathrm{RSS}}{n} := \frac{1}{n}\|y - \mathbb{X}\widehat{\beta}\|^2.$$

Ограниченной модели соответствует матрица $\mathbb{X}_r$, содержащая первые $r+1$ столбец матрицы $\mathbb{X}$. Метод максимального правдоподобия дает следующие оценки:

$$\widehat{\beta}_r = (\mathbb{X}_r^T \mathbb{X}_r)^{-1} \mathbb{X}_r^T y, \quad \widehat{\sigma}_r^2 = \frac{\mathrm{RSS}_r}{n} := \frac{1}{n} \|y - \mathbb{X}\widehat{\beta}_r\|^2.$$

Поскольку

$$L_n(x, \widehat{\beta}) = \left(\frac{1}{\sqrt{2\pi}\widehat{\sigma}}\right)^n \exp\left(-\frac{1}{2n}\right),$$

$$L_n(x, \widehat{\beta}_r) = \left(\frac{1}{\sqrt{2\pi}\widehat{\sigma}_r}\right)^n \exp\left(-\frac{1}{2n}\right),$$

то статистика отношения правдоподобия

$$\Lambda_n(x) = \frac{L_n(x, \widehat{\beta})}{L_n(x, \widehat{\beta}_r)} = \left(\frac{\widehat{\sigma}_r}{\widehat{\sigma}}\right)^n = \left(\frac{\mathrm{RSS}_r}{\mathrm{RSS}}\right)^{n/2}.$$

Условие $\Lambda_n(x) > K$ эквивалентным образом можно переписать так:

$$F_n = \frac{(\mathrm{RSS}_r - \mathrm{RSS})/(d-r)}{\mathrm{RSS}/(n-d-1)} > \gamma.$$

Пусть $V \sim \chi^2(n)$, $W \sim \chi^2(m)$ — независимые случайные величины. По определению, случайная величина

$$F = \frac{V/n}{W/m}$$

имеет F-распределение (распределение Фишера) с n и m степенями свободы. Обозначение: $F \sim \mathcal{F}(n,m)$. Плотность:

$$f(x) = \frac{n^{n/2}\Gamma((n+m)/2)}{m^{m/2}\Gamma(m/2)\Gamma(n/2)} \frac{x^{(n-2)/2}}{(1+nx/m)^{(n+m)/2}}, \quad x > 0.$$

- $F_n \sim \mathcal{F}(d-r, n-d-1)$ при выполнении H_0.

Доказательство. Нужно проверить, что

$$(\mathrm{RSS}_r - \mathrm{RSS})/\sigma^2 \sim \chi^2(d-r), \quad \mathrm{RSS}/\sigma^2 \sim \chi^2(n-d-1)$$

и эти случайные величины независимы. Свойство

$$\mathrm{RSS}/\sigma^2 = n\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n-d-1)$$

установлено выше.

$$\mathrm{RSS}_r = \|Y - \mathbb{X}_r\widehat{\beta}_r\|^2 = \|Y - \mathbb{X}_r(\mathbb{X}_r^T\mathbb{X}_r)^{-1}\mathbb{X}_r^T Y\|^2 = \|(I-\mathbb{H}_r)Y\|^2$$

$$(I-\mathbb{H}_r)Y = (I-\mathbb{H}_r)(\mathbb{X}_r\widehat{\beta} + \varepsilon) = (I-\mathbb{H}_r)\varepsilon,$$

так как $\mathbb{H}_r$ является проектором на $\operatorname{im}\mathbb{X}_r$. Следовательно,

$$\mathrm{RSS}_r = \|(I-\mathbb{H}_r)\varepsilon\|^2 = \langle (I-\mathbb{H}_r)\varepsilon, \varepsilon\rangle.$$

Аналогичным образом,

$$\mathrm{RSS} = \langle (I - \mathbb{H})\varepsilon, \varepsilon \rangle.$$

Тогда

$$\frac{\mathrm{RSS}_r - \mathrm{RSS}}{\sigma^2} = \frac{1}{\sigma^2}(\langle (I - \mathbb{H}_r)\varepsilon, \varepsilon \rangle - \langle (I - \mathbb{H})\varepsilon, \varepsilon \rangle) = \frac{1}{\sigma^2}\langle (\mathbb{H} - \mathbb{H}_r)\varepsilon, \varepsilon \rangle.$$

$\mathbb{H} - \mathbb{H}_r$ является ортопроектором на подпространство, порожденное последними $d - r$ столбцами матрицы $\mathbb{X}$. Поскольку $\varepsilon/\sigma \sim N(0, I_n)$, то отсюда следует, что

$$\frac{\mathrm{RSS}_r - \mathrm{RSS}}{\sigma^2} \sim \chi^2(d - r).$$

Для завершения доказательства остается проверить, что

$$(I - \mathbb{H})\varepsilon, \quad (\mathbb{H} - \mathbb{H}_r)\varepsilon$$

независимы.

Пусть $A = I - \mathbb{H}$, $B = \mathbb{H} - \mathbb{H}_r$. Тогда

$$\begin{pmatrix} A \\ B \end{pmatrix} \varepsilon \sim N\left(0, \sigma^2 \begin{pmatrix} A \\ B \end{pmatrix} \left(A^T B^T\right)\right) = N\left(0, \sigma^2 \begin{pmatrix} AA^T & AB^T \\ BA^T & BB^T \end{pmatrix}\right),$$

$$AB^T = AB = (I - \mathbb{H})(\mathbb{H} - \mathbb{H}_r) = -\mathbb{H}_r + \mathbb{H}\mathbb{H}_r = 0,$$

так как $\mathbb{H}_r$ является ортопроектором на подпространство $\operatorname{im} \mathbb{X}_r \subset \operatorname{im} \mathbb{X}$. $\square$

В дисперсионном анализе (ANOVA: Analysis of variance) рассматривается модель

$$Y_j = \mu + \alpha_i + \varepsilon_j, \quad j \in J_i = \{n_{i-1}+1, \dots, n_i\}, \quad i = 1, \dots, k,$$

$n_0 = 0 < n_1 < \dots < n_k$, $\alpha_1 = 0$, в которой имеется k групп наблюдений. Проверим гипотезу об отсутствии влияния некоторого воздействия:

$$H_0 : \alpha_2 = \dots = \alpha_k = 0, \quad H_1 : H_0 \text{ неверна.}$$

Условие $\alpha_1 = 0$ связано с идентифицируемостью параметров. Оно гарантирует обратимость матрицы $\mathbb{X}^T\mathbb{X}$.
Воспользуемся критерием

$$F_n = \frac{(\mathrm{RSS}_r - \mathrm{RSS})/(d-r)}{\mathrm{RSS}/(n-d-1)} > \gamma.$$

В данном случае $d = k-1$, $r = 0$.

В полной модели вектор параметров $\beta = (\mu, \alpha_2, \ldots, \alpha_k)$ является решением оптимизационной задачи

$$\sum_{i=1}^{n}(y_i - \langle \beta, x_i \rangle)^2 = \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \mu - \alpha_i)^2$$

$$= \sum_{j=1}^{n_1}(y_j - \mu)^2 + \sum_{i=2}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \mu - \alpha_i)^2 \to \min_{\mu, \alpha_2, \ldots, \alpha_k}.$$

Минимизируем по α_i:

$$\alpha_i = \overline{y}_i - \mu, \quad \overline{y}_i := \frac{1}{n_i - n_{i-1}} \sum_{j=n_{i-1}+1}^{n_i} y_j.$$

При таком α_i двойная сумма не будет зависеть от μ. Минимизируем первую сумму по μ:

$$\widehat{\mu} = \overline{y}_1 := \frac{1}{n_1} \sum_{j=1}^{n_1} y_j.$$

При этом $\widehat{\alpha}_i = \overline{y}_i - \widehat{\mu} = \overline{y}_i - \overline{y}_1$.

Таким образом,

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \langle \widehat{\beta}, x_i \rangle)^2 = \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \overline{y}_i)^2.$$

В ограниченной модели единственный параметр μ является решением оптимизационной задачи:

$$\sum_{i=1}^{n} (y_i - \mu)^2 \to \min_{\mu}.$$

$$\widehat{\mu} = \overline{y} := \frac{1}{n} \sum_{j=1}^{n} y_j, \quad \mathrm{RSS}_r = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Используются следующие обозначения:

$$SS_T = \sum_{j=1}^{n} (y_j - \overline{y})^2, \quad SS_W = \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \overline{y}_i)^2,$$

$$SS_B = \sum_{i=1}^{k} \sum_{j=N_{i-1}+1}^{N_i} (\overline{y}_i - \overline{y})^2 = \sum_{i=1}^{k} n_i (\overline{y}_i - \overline{y})^2.$$

SS_T: соответствует общей выборочной дисперсии (total), SS_W: соответствует сумме внутригрупповых дисперсий (within); SS_B: соответствует межгрупповой дисперсии (between).

$$SS_T = \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \overline{y})^2 = \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \overline{y}_i + \overline{y}_i - \overline{y})^2$$

$$= \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (y_j - \overline{y}_i)^2 + \sum_{i=1}^{k} \sum_{j=n_{i-1}+1}^{n_i} (\overline{y}_i - \overline{y})^2 = SS_W + SS_B,$$

так как

$$\sum_{i=1}^{k}\sum_{j=n_{i-1}+1}^{n_i}(y_j-\overline{y}_i)(\overline{y}_i-\overline{y})=\sum_{i=1}^{k}(\overline{y}_i-\overline{y})\sum_{j=n_{i-1}+1}^{n_i}(y_j-\overline{y}_i)$$

$$=\sum_{i=1}^{k}(\overline{y}_i-\overline{y})\left(\sum_{j=n_{i-1}+1}^{n_i}y_j-n_i\overline{y}_i\right)=0.$$

Таким образом,

$$F_n=\frac{(\mathrm{RSS}_r-\mathrm{RSS})/(k-1)}{\mathrm{RSS}/(n-k)}=\frac{(SS_T-SS_W)/(k-1)}{SS_W/(n-k)}$$

$$=\frac{SS_B/(k-1)}{SS_W/(n-k)}\sim\mathcal{F}(k-1,n-k).$$

Sahoo, Example 20.3

Проверить нулевую гипотезу о том, что среднее число баллов, полученное тремя группами студентов одинаково.

75	90	17
91	80	81
83	50	55
45	93	70
82	53	61
75	87	43
68	76	89
47	82	73
38	78	58
	80	70
	33	
	79	

В данном случае $k = 3$, $n_1 = 9$, $n_2 = 12$, $n_3 = 10$, $n = 31$,

$$\overline{y}_1 = 67.11, \quad \overline{y}_2 = 73.42, \quad \overline{y}_3 = 61.70, \quad \overline{y} = 67.81.$$

$$SS_B = n_1(\overline{y_1} - \overline{y})^2 + n_2(\overline{y_2} - \overline{y})^2 + n_3(\overline{y_3} - \overline{y})^2 = 754.93,$$

$$SS_W = \sum_{i=1}^{k} \sum_{j=N_{i-1}+1}^{N_i} (y_j - \overline{y}_i)^2 = 10361.91,$$

$$F_n = \frac{SS_B/(k-1)}{SS_W/(n-k)} = \frac{SS_B/2}{SS_W/28} \sim \mathcal{F}(2, 28).$$

Полученное значение: $F_n = 1.02$,

$$p\text{-значение} = \mathsf{P}(F > 1.02) = 0.37, \quad F \sim \mathcal{F}(2, 28).$$

Гипотеза о равенстве средних не отвергается.

Sahoo, Example 20.2

Проверить нулевую гипотезу о равенстве средних для следующих двух наборов данных.

Data set 1			Data set 2		
8.1	8.0	14.8	9.2	9.5	9.4
4.2	15.1	5.3	9.1	9.5	9.3
14.7	4.7	11.1	9.2	9.5	9.3
9.9	10.4	7.9	9.2	9.6	9.3
12.1	9.0	9.3	9.3	9.5	9.2
6.2	9.8	7.4	9.2	9.4	9.3

Для каждого набора данных $k = 3$, $n_1 = n_2 = n_3 = 6$, $n = 18$,

$$\overline{y}_1 = 9.20, \quad \overline{y}_2 = 9.50, \quad \overline{y}_3 = 9.30, \quad \overline{y} = 9.33,$$

$$SS_B = n_1(\overline{y_1} - \overline{y})^2 + n_2(\overline{y_2} - \overline{y})^2 + n_3(\overline{y_3} - \overline{y})^2 = 0.28,$$

$$F_n = \frac{SS_B/(k-1)}{SS_W/(n-k)} = \frac{SS_B/2}{SS_W/15} \sim \mathcal{F}(2, 15).$$

Для первого набора:

$$SS_W = \sum_{i=1}^{k} \sum_{j=N_{i-1}+1}^{N_i} (y_j - \overline{y}_i)^2 = 187.22.$$

Полученное значение $F_n = 0.01$,

$$p\text{-значение} = \mathsf{P}(F > 0.01) = 0.99, \quad F \sim \mathcal{F}(2, 15).$$

Гипотеза о равенстве средних не отвергается.
Для второго набора: $SS_W = 0.06$, полученное значение $F_n = 35$,

$$p\text{-значение} = \mathsf{P}(F > 0.01) = 2.2 \cdot 10^{-6}, \quad F \sim \mathcal{F}(2, 15).$$

Гипотеза о равенстве средних отвергается.
ANOVA обнаруживает более значительные различия между средними значениями во втором наборе данных: при одинаковой межгрупповой дисперсии, внутригрупповые дисперсии во втором случае намного меньше.

Ошибки обучения и обобщения

Ошибка обучения:

$$\widehat{T} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \langle\widehat{\beta}, x_i\rangle)^2.$$

Ожидаемая ошибка обобщения:

$$G = \mathsf{E}(Y - \langle\widehat{\beta}, X\rangle)^2,$$

где (x, Y) — новый пример: случайная величина с тем же распределением, не зависящая от $(x_1, Y_1), \ldots, (x_n, Y_n)$.
Наряду с исходной выборкой $Y_i = \langle\beta, x_i\rangle + \varepsilon_i$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ рассмотрим еще одну с теми же x_i: $Y_i' = \langle\beta, x_i\rangle + \varepsilon_i'$. Случайные векторы $\varepsilon, \varepsilon'$, описывающие шум, независимы.
Пусть $\widehat{Y}_i = \langle\widehat{\beta}, x_i\rangle$ — предсказанные значения. Тогда ошибки обучения и обобщения имеют вид

$$\widehat{T} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2, \quad \widehat{G} = \frac{1}{n}\sum_{i=1}^{n}(Y_i' - \widehat{Y}_i)^2.$$

Заметим, что

$$\mathsf{E}\widehat{G} = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(Y_i' - \langle\widehat{\beta}, x_i\rangle)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\mathsf{E}((Y_i' - \langle\widehat{\beta}, x_i\rangle)^2|\varepsilon)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\mathsf{E}((Y - \langle\widehat{\beta}, x_i\rangle)^2|\varepsilon) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}((Y - \langle\widehat{\beta}, x_i\rangle)^2) = G.$$

- Ожидаемая ошибка обучения меньше ожидаемой ошибки обобщения:
$$\mathsf{E}\widehat{G} = \mathsf{E}\widehat{T} + \frac{2}{n}\sigma^2(d+1).$$

Доказательство. Имеем,

$$\mathsf{E}(Y_i - \widehat{Y}_i)^2 = \mathrm{Var}(Y_i - \widehat{Y}_i) + (\mathsf{E}Y_i - \mathsf{E}\widehat{Y}_i)^2$$
$$= \mathrm{Var}(Y_i) + \mathrm{Var}(\widehat{Y}_i) - 2\mathrm{Cov}(Y_i, \widehat{Y}_i) + (\mathsf{E}Y_i - \mathsf{E}\widehat{Y}_i)^2.$$

$$
\begin{aligned}
\mathsf{E}(Y_i' - \widehat{Y}_i)^2 &= \mathrm{Var}(Y_i' - \widehat{Y}_i) + (\mathsf{E}Y_i' - \mathsf{E}\widehat{Y}_i)^2 \\
&= \mathrm{Var}(Y_i') + \mathrm{Var}(\widehat{Y}_i) - 2\mathrm{Cov}(Y_i', \widehat{Y}_i) + (\mathsf{E}Y_i' - \mathsf{E}\widehat{Y}_i)^2 \\
&= \mathrm{Var}(Y_i) + \mathrm{Var}(\widehat{Y}_i) + (\mathsf{E}Y_i - \mathsf{E}\widehat{Y}_i)^2
\end{aligned}
$$

так как $\mathrm{Cov}(Y_i', \widehat{Y}_i) = 0$, $\mathsf{E}Y_i' = \mathsf{E}Y_i$, $\mathrm{Var}(Y_i') = \mathrm{Var}(Y_i)$.
Следовательно,

$$
\mathsf{E}(Y_i' - \widehat{Y}_i)^2 = \mathsf{E}(Y_i - \widehat{Y}_i)^2 + 2\mathrm{Cov}(Y_i, \widehat{Y}_i),
$$

$$
\mathsf{E}\frac{1}{n}\sum_{i=1}^{n}(Y_i' - \widehat{Y}_i)^2 = \mathsf{E}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \frac{2}{n}\sum_{i=1}^{n}\mathrm{Cov}(Y_i, \widehat{Y}_i).
$$

Но

$$
\mathrm{Cov}(Y_i, \widehat{Y}_i) = \sum_{j=1}^{n}\mathbb{H}_{ij}\mathrm{Cov}(Y_i, Y_j) = \mathbb{H}_{ii}\mathrm{Var}(Y_i) = \sigma^2\mathbb{H}_{ii}.
$$

Следовательно,

$$\mathsf{E}\frac{1}{n}\sum_{i=1}^{n}(Y_i' - \widehat{Y}_i)^2 = \mathsf{E}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \frac{2}{n}\sigma^2 \mathrm{Tr}(\mathbb{H}).$$

Напомним, что собственными числами проектора могут быть только 0 и 1. Количество единиц равно размерности образа, значит $\mathrm{Tr}(\mathbb{H}) = d + 1$. $\square$

Величину $2\sigma^2(d+1)/n$ можно назвать оптимизмом модели регрессии. Она равна смещению оценки ошибки обучения, если рассматривать ее как оценку ожидаемой ошибки обобщения.

Кросс-валидация (скользящий контроль)

Рассмотрим случайное разбиение $\{I_1, \ldots, I_K\}$ множества индексов $I = \{1, \ldots, n\}$. Обозначим через $\beta^{(-k)}$ коэффициенты модели линейной регрессии, для определения которых использованы только данные (x_i, y_i) с индексами из $I \backslash I_k$. Оценка ошибки обобщения "полной" модели $\widehat{f}(x) = \langle \widehat{\beta}, x \rangle$ строится следующим образом:

$$\mathrm{CV}(\widehat{f}) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{|I_k|} \sum_{i \in I_k} (y_i - \langle \widehat{\beta}^{-k}, x_i \rangle)^2.$$

Для больших K все модели $\widehat{f}^{(-i)}$ близки друг к другу и к $\widehat{f}$ (сильно коррелированы). Для малых K они могут сильно отличаться (слабо коррелитрованы). Оптимальное значение K зависит от данных. Иногда рекомендуют $K = 5$ или $K = 10$.

Leave-one-out CV (LOOCV): контроль по отдельным объектам

LOOCV-оценка ошибки обобщения:

$$CV(\widehat{f}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\widehat{\beta}_{(-i)})^2.$$

Здесь $\widehat{\beta}_{(-i)}$ — оценка метода наименьших квадратов для данных с индексами $I\backslash\{i\}$:

$$\widehat{\beta}_{(-i)} = (\mathbb{X}_{(-i)}^T\mathbb{X}_{(-i)})^{-1}\mathbb{X}_{(-i)}^T y_{(-i)},$$

$$\mathbb{X}_{(-i)} = \begin{pmatrix} x_1^T \\ \vdots \\ x_{i-1}^T \\ x_{i+1}^T \\ \vdots \\ x_n^T \end{pmatrix} \in \mathbb{R}^{(n-1)\times(d+1)}, \quad y_{(-i)} = \begin{pmatrix} y_1^T \\ \vdots \\ y_{i-1}^T \\ y_{i+1}^T \\ \vdots \\ y_n^T \end{pmatrix} \in \mathbb{R}^{(n-1)\times 1}.$$

Формула Шермана-Моррисона (Википедия)

Теорема

Пусть $A \in \mathbb{R}^{n\times n}$ — обратимая матрица и $u, v \in \mathbb{R}^n$ — векторы столбцы. Матрица $A + uv^T$ обратима, если и только если $1 + v^T A^{-1} u \neq 0$. В этом случае,

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}.$$

Доказательство. "$\Longleftarrow$" Предположим, что $1 + v^T A^{-1} u \neq 0$ и положим

$$Y = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}.$$

Имеем

$$
\begin{aligned}
(A+uv^T)Y &= \left(A+uv^T\right)\left(A^{-1}-\frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\right)\\
&= AA^{-1}+uv^TA^{-1}-\frac{AA^{-1}uv^TA^{-1}+uv^TA^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\\
&= I+uv^TA^{-1}-\frac{uv^TA^{-1}+uv^TA^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\\
&= I+uv^TA^{-1}-\frac{u\left(1+v^TA^{-1}u\right)v^TA^{-1}}{1+v^TA^{-1}u}\\
&= I+uv^TA^{-1}-uv^TA^{-1}=I
\end{aligned}
$$

$\Longrightarrow \operatorname{im}(A+uv^T)=\mathbb{R}^n \Longrightarrow A+uv^T$ обратима и обратной является Y.
"$\Longrightarrow$" Если $1+v^TA^{-1}u=0$, то $A+uv^T$ не является обратимой:

$$\left(A+uv^T\right)A^{-1}u=u+uv^TA^{-1}u=u(1+v^TA^{-1}u)=0. \quad \square$$

Теорема

$$CV(\widehat{f}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\widehat{\beta}_{(-i)})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - x_i^T\widehat{\beta}}{1-H_{ii}}\right)^2,$$

где $\widehat{\beta}$ — оценка метода наименьших квадратов, построенная по полному набору данных,

$$\mathbb{H} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T.$$

Доказательство. Покажем, что $y_i - x_i^T\widehat{\beta}_{(-i)} = \frac{y_i - x_i^T\widehat{\beta}}{1-H_{ii}}$. Для этого выразим

$$\widehat{\beta}_{(-i)} = (\mathbb{X}_{(-i)}^T\mathbb{X}_{(-i)})^{-1}\mathbb{X}_{(-i)}^T y_{(-i)}$$

в терминах $\mathbb{X}$ и y:

$$\left(\mathbb{X}_{(-i)}^T y_{(-i)}\right)_k = \sum_{s=1}^{n}\mathbb{X}_{ks}^T y_s - \mathbb{X}_{ki}^T y_i = (\mathbb{X}^T y)_k - \mathbb{X}_{ik} y_i = (\mathbb{X}^T y)_k - (x_i y_i)_k,$$

$$\mathbb{X}_{(-i)}^T y_{(-i)} = \mathbb{X}^T y - x_i y_i.$$

$$\left(\mathbb{X}^T_{(-i)}\mathbb{X}_{(-i)}\right)_{jk} = \sum_{s=1}^{n-1}\left(\mathbb{X}^T_{(-i)}\right)_{js}\left(\mathbb{X}_{(-i)}\right)_{sk} = \sum_{s=1}^{n-1}\left(\mathbb{X}_{(-i)}\right)_{sj}\left(\mathbb{X}_{(-i)}\right)_{sk}$$

$$= \sum_{s=1}^{n} x_{sj}x_{sk} - x_{ij}x_{ik} = (\mathbb{X}^T\mathbb{X})_{jk} - (x_i x_i^T)_{jk},$$

$$\mathbb{X}^T_{(-i)}\mathbb{X}_{(-i)} = \mathbb{X}^T\mathbb{X} - x_i x_i^T.$$

Положим $A = \mathbb{X}^T\mathbb{X}$, $u = x_i$, $v = -x_i$ в формуле Шермана-Моррисона:

$$\left(\mathbb{X}^T_{(-i)}\mathbb{X}_{(-i)}\right)^{-1} = \left(\mathbb{X}^T\mathbb{X} - x_i x_i^T\right)^{-1}$$

$$= \left(\mathbb{X}^T\mathbb{X}\right)^{-1} + \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i x_i^T \left(\mathbb{X}^T\mathbb{X}\right)^{-1}}{1 - x_i^T \left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}$$

$$= \left(\mathbb{X}^T\mathbb{X}\right)^{-1} + \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i x_i^T \left(\mathbb{X}^T\mathbb{X}\right)^{-1}}{1 - H_{ii}},$$

так как $x_i^T \left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i = (\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T)_{ii} = H_{ii}$.

Для применения этой формулы требуется, чтобы $H_{ii} \neq 1$. Из равенств $H = H^2$, $H = H^T$ вытекает, что

$$H_{ii} = \sum_{k=1}^{n} H_{is} H_{si} = H_{ii}^2 + \sum_{s \neq i} H_{is}^2.$$

Следовательно, $0 \leq H_{ii} \leq 1$. If $H_{ii} = 1$, then $H_{is} = 0$, $s \neq i$. В этом экзотическом случае $\widehat{y}_i = (Hy)_i = y_i$.

$$\begin{aligned}
\widehat{\beta}_{(-i)} &= (\mathbb{X}_{(-i)}^T \mathbb{X}_{(-i)})^{-1} \mathbb{X}_{(-i)}^T y_{(-i)} \\
&= \left(\left(\mathbb{X}^T \mathbb{X}\right)^{-1} + \frac{\left(\mathbb{X}^T \mathbb{X}\right)^{-1} x_i x_i^T \left(\mathbb{X}^T \mathbb{X}\right)^{-1}}{1 - H_{ii}} \right) (\mathbb{X}^T y - x_i y_i) \\
&= \left(\mathbb{X}^T \mathbb{X}\right)^{-1} \mathbb{X}^T y - \left(\mathbb{X}^T \mathbb{X}\right)^{-1} x_i y_i \\
&+ \frac{\left(\mathbb{X}^T \mathbb{X}\right)^{-1} x_i}{1 - H_{ii}} \left(x_i^T \left(\mathbb{X}^T \mathbb{X}\right)^{-1} \mathbb{X}^T y - x_i^T \left(\mathbb{X}^T \mathbb{X}\right)^{-1} x_i y_i \right)
\end{aligned}$$

$$
\begin{aligned}
\widehat{\beta}_{(-i)} &= \widehat{\beta} - \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}{1 - H_{ii}}(1 - H_{ii})y_i + \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}{1 - H_{ii}} \left(x_i^T \widehat{\beta} - H_{ii} y_i\right) \\
&= \widehat{\beta} - \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}{1 - H_{ii}} \left(y_i - x_i^T \widehat{\beta}\right)
\end{aligned}
$$

$$
\begin{aligned}
y_i - x_i^T \widehat{\beta}_{(-i)} &= y_i - x_i^T \left(\widehat{\beta} - \frac{\left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}{1 - H_{ii}} \left(y_i - x_i^T \widehat{\beta}\right)\right) \\
&= y_i - x_i^T \widehat{\beta} + \frac{x_i^T \left(\mathbb{X}^T\mathbb{X}\right)^{-1} x_i}{1 - H_{ii}} \left(y_i - x_i^T \widehat{\beta}\right) \\
&= (y_i - x_i^T \widehat{\beta}) \left(1 + \frac{H_{ii}}{1 - H_{ii}}\right) = \frac{y_i - x_i^T \widehat{\beta}}{1 - H_{ii}}. \quad \square
\end{aligned}
$$

Выбор модели с помощью кросс-валидации

Для семейства моделей $f_\theta = f(x;\theta)$, индексированных гиперпараметром θ (в случае линейной регресси $\theta = d$), разыскивается $\widehat{\theta}$ с наилучшей CV-оценкой ошибки обобщения:

$$\widehat{\theta} \in \arg\min_{\theta} \mathrm{CV}(\widehat{f}_\theta).$$

Результат зависит от случайного разбиения, за исключением случая LOOCV. Для малых K значение $\widehat{\theta}$ может существенно зависеть от случайного разбиения: см. примеры в James, Witten, Hastie, Tibshirani "An Introduction to Statistical Learning with Applications in R" (2014).

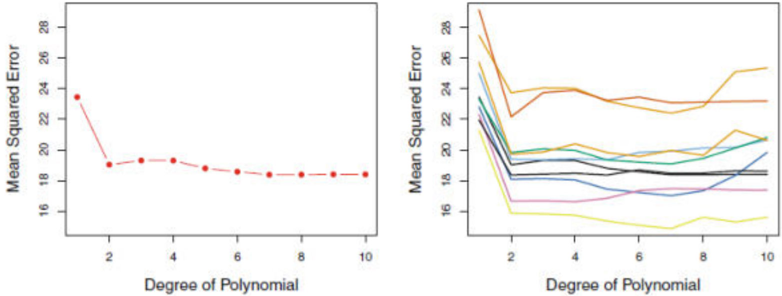


FIGURE 5.2. *The validation set approach was used on the* `Auto` *data set in order to estimate the test error that results from predicting* `mpg` *using polynomial functions of* `horsepower`. Left: *Validation error estimates for a single split into training and validation data sets.* Right: *The validation method was repeated ten times, each time using a different random split of the observations into a training set and a validation set. This illustrates the variability in the estimated test MSE that results from this approach.*

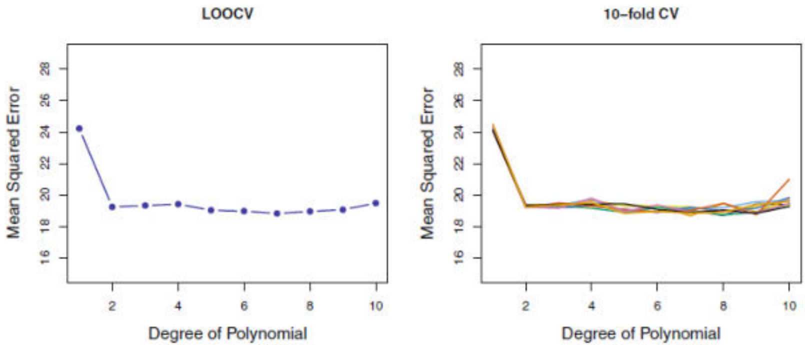


FIGURE 5.4. *Cross-validation was used on the* `Auto` *data set in order to estimate the test error that results from predicting* `mpg` *using polynomial functions of* `horsepower`. Left: *The LOOCV error curve.* Right: 10*-fold CV was run nine separate times, each with a different random split of the data into ten parts. The figure shows the nine slightly different CV error curves.*

ИЗБРАННЫЕ ВОПРОСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ
III. Элементы теории случайных процессов

Рохлин Д.Б.

ИММиКН, ЮФУ, 2022

References

- *Булинский А.В., Ширяев А.Н.* Теория случайных процессов. М.: Физматлит, 2005.
- *Рохлин Д.Б.* Основы стохастического анализа. Ростов-на-Дону – Таганрог, Издательство ЮФУ, 2020.
- *Baldi P.* Stochastic calculus: an introduction through theory and exercises, Cham: Springer, 2017.
- *Bass R.F.* Stochastic processes. Cambridge: Cambridge University Press, 2011.
- *Williams D.* Probability with martingales. Cambridge: Cambridge University Press, 1991.
- *Zitkovic G.* Lecture notes for "Introduction to Stochastic Processes", University of Texas at Austin, 2022.

Расширяющаяся последовательность σ-алгебр $\mathscr{F}_0 \subset \dots \mathscr{F}_n \subset \dots \subset \mathscr{F}$ называется *фильтрацией*, а четверка $(\Omega, \mathscr{F}, \mathsf{P}, (\mathscr{F}_n)_{n=0}^{\infty})$ — фильтрованным вероятностным пространством.

Случайным процессом называется последовательность случайных величин $(X_n)_{n=0}^{\infty}$. Говорят, что *случайный процесс* $X = (X_n)_{n=0}^{\infty}$ *согласован* с фильтрацией $\mathbb{F} = (\mathscr{F}_n)_{n=0}^{\infty}$, если случайные величины X_n — $\mathscr{F}_n$-измеримы.

Согласованный с фильтрацией $\mathbb{F}$ случайный процесс $X = (X_n)_{n=0}^{\infty}$ с $\mathsf{E}|X_n| < \infty$ называется

- *мартингалом*, если $\mathsf{E}(X_{n+1}|\mathscr{F}_n) = X_n$,
- *субмартингалом*, если $\mathsf{E}(X_{n+1}|\mathscr{F}_n) \geq X_n$,
- *супермартингалом*, если $\mathsf{E}(X_{n+1}|\mathscr{F}_n) \leq X_n$.

Пример 1

Рассмотрим последовательность независимых случайных величин ξ_n:

$$\mathsf{P}(\xi_n = 1) = p, \quad \mathsf{P}(\xi_n = -1) = q, \quad p + q = 1, \quad p, q > 0.$$

Относительно естественной фильтрации процесса ξ: $\mathscr{F}_n = \sigma(\xi_0, \dots, \xi_n)$ процесс $X_n = \sum_{k=1}^{n} \xi_k$ является мартингалом, если $p = 1/2$; субмартингалом, если $p > 1/2$; супермартингалом, если $p < 1/2$. Действительно,

$$\mathsf{E}(X_{n+1}|\mathscr{F}_n) = \mathsf{E}(\xi_{n+1}|\mathscr{F}_n) + \mathsf{E}(X_n|\mathscr{F}_n) = \mathsf{E}\xi_{n+1} + X_n = p - q + X_n.$$

Пример 2

Пусть $Z \in L^1(\mathscr{F})$. Тогда процесс $X_n = \mathsf{E}(Z|\mathscr{F}_n)$ является мартингалом:

$$\mathsf{E}(X_{n+1}|\mathscr{F}_n) = \mathsf{E}(\mathsf{E}(Z|\mathscr{F}_{n+1})|\mathscr{F}_n) = \mathsf{E}(Z|\mathscr{F}_n) = X_n.$$

Процесс A называется *предсказуемым*, если случайные величины A_n являются $\mathscr{F}_{n-1}$-измеримыми ($\mathscr{F}_{-1} := \mathscr{F}_0$).

Теорема 3 (Разложение Дуба)

Пусть X — интегрируемый процесс: $X_n \in L^1(\mathscr{F}_n)$. Тогда существуют предсказуемый процесс A с $A_0 = 0$ и мартингал M такие, что

$$X = A + M. \tag{1}$$

Процессы A, M определяются однозначно с точностью до множества нулевой вероятности. Процесс X является субмартингалом, если и только если процесс A — неубывающий: $A_n \leq A_{n+1}$.

Доказательство. Пусть $\Delta X_k = X_k - X_{k-1}$. Тогда

$$X_n = X_0 + \sum_{k=1}^{n} \Delta X_k = X_0 + \sum_{k=1}^{n} (\Delta X_k - \mathsf{E}(\Delta X_k | \mathscr{F}_{k-1})) + \sum_{k=1}^{n} \mathsf{E}(\Delta X_k | \mathscr{F}_{k-1}).$$

Таким образом, X представим в виде суммы мартингала

$$M_n = X_0 + \sum_{k=1}^{n} (\Delta X_k - \mathsf{E}(\Delta X_k | \mathscr{F}_{k-1})), \quad n \geq 0$$

и предсказуемого процесса

$$A_n = \sum_{k=1}^{n} \mathsf{E}(\Delta X_k | \mathscr{F}_{k-1}), \quad n \geq 1, \quad A_0 = 0.$$

Если имеется другое разложение $X = A' + M'$ вида (1), то $A - A' = M' - M$ и

$$\begin{aligned} A_{n+1} - A'_{n+1} &= \mathsf{E}(A_{n+1} - A'_{n+1} | \mathscr{F}_n) = \mathsf{E}(M'_{n+1} - M_{n+1} | \mathscr{F}_n) \\ &= M'_n - M_n = A_n - A'_n \end{aligned}$$

при почти всех ω. Поскольку $A_0 = A'_0$, то отсюда следует, что $A = A'$. Но тогда и $M = M'$.
Ясно также, что X является субмартингалом, если и только если

$$\mathsf{E}(X_{n+1} - X_n | \mathscr{F}_n) = \mathsf{E}(A_{n+1} - A_n | \mathscr{F}_n) = A_{n+1} - A_n \geq 0. \quad \square$$

(B,S)-рынок и портфель ценных бумаг

Рынок состоит из банковского счета (безрискового актива) и акции (рискового актива). Динамика банковского счета описывается предсказуемой стохастической последовательностью $(B_k, \mathcal{F}_{k-1})$, а динамика акции — стохастической последовательностью $(S_k, \mathcal{F}_k)$. Введем процентные ставки (“на завтра”):

$$r_k = \frac{B_k - B_{k-1}}{B_{k-1}}, \quad \rho_k = \frac{S_k - S_{k-1}}{S_{k-1}}.$$

Очевидно, что величины r_k — $\mathcal{F}_{k-1}$-измеримы, величины ρ_k — $\mathcal{F}_k$-измеримы и

$$B_k = (1 + r_k)B_{k-1} = \Pi_{j=1}^{k}(1 + r_j)B_0,$$

$$S_k = (1 + \rho_k)S_{k-1} = \Pi_{j=1}^{k}(1 + \rho_j)S_0.$$

На рынке можно

- размещать деньги на банковском счете и брать с него в долг
- покупать и продавать акции

Обозначим количество единиц банковского счета, которыми владеет инвестор через β, а количество акций — через γ. Двумерная предсказуемая стохастическая последовательность $((\beta_k, \gamma_k), \mathcal{F}_{k-1})$ называется портфелем ценных бумаг или стратегией инвестора.
Капиталом портфеля ценных бумаг называется стохастическая последовательность

$$X_k = \beta_k B_k + \gamma_k S_k.$$

Условие самофинансируемости:

$$X_{k-1} = \beta_k B_{k-1} + \gamma_k S_{k-1}.$$

Пусть $\Delta a_k = a_k - a_{k-1}$. Воспользовавшись формулой

$$\begin{aligned} \Delta(a_k b_k) &= a_k b_k - a_{k-1} b_{k-1} = a_k(b_k - b_{k-1}) + b_{k-1}(a_k - a_{k-1}) \\ &= a_k \Delta b_k + b_{k-1} \Delta a_k, \end{aligned}$$

находим

$$\begin{aligned} \Delta X_k &= \beta_k \Delta B_k + \gamma_k \Delta S_k + (B_{k-1} \Delta \beta_k + S_{k-1} \Delta \gamma_k) \\ &= \beta_k \Delta B_k + \gamma_k \Delta S_k. \end{aligned}$$

Часто полезно измерять все величины в единицах банковского счета (произвести дисконтирование). Пусть $B_k(\omega) > 0$ при всех ω и k. Положим

$$\widetilde{S}_k = S_k/B_k, \quad \widetilde{X}_k = X_k/B_k = \beta_k + \gamma_k \widetilde{S}_k.$$

Заметим, что при постоянной банковской процентной ставке $r_k = r$ и $B_0 = 1$ дисконтированная цена акции принимает вид $\widetilde{S}_k = S_k/(1+r)^k$. Вычислим приращение дисконтированного капитала, используя условие самофинансируемости:

$$\Delta \widetilde{X}_k = \Delta(\beta_k + \gamma_k \widetilde{S}_k) = \gamma_k \Delta \widetilde{S}_k + \Delta \beta_k + \widetilde{S}_{k-1} \Delta \gamma_k = \gamma_k \Delta \widetilde{S}_k.$$

Полезно отметить, что в эту формулу явно не входит банковский счет.

Арбитраж

Арбитраж — это возможность получения прибыли без риска и при отсутствии начального капитала. Самофинансируемая стратегия (β_k, γ_k) реализует арбитражную возможность, если соответствующий капитал удовлетворяет условиям

- $X_0 = 0, \quad X_N(\omega) \geq 0, \quad \forall\omega$
- $\exists\,\omega : X_N(\omega) > 0$

Полнота

Пусть инвестор взял на себя обязательство выплатить в момент времени N некоторую сумму, величина которой зависит от цены акции в момент N или более ранние моменты времени. Такое обязательство описывается $\mathscr{F}_N$-измеримой случайной величиной $f_N(\omega)$.
Платежное обязательство f_N при этом называется достижимым или воспроизводимым, если существует такой портфель, что

$$X_0 = x, \quad X_N(\omega) = f_N(\omega).$$

Справедливой ценой платежного обязательства является x, иначе у продавца или покупателя возникает арбитражная возможность.
Рынок ценных бумаг называется полным, если всякое $\mathscr{F}_N$-измеримое платежное обязательство f_N воспроизводимо.

Биномиальная модель рынка

В N-шаговой биномиальной модели предполагается, что цена акции на каждом шаге умножается на u с вероятностью p или на d с вероятностью q $(u > d > 0)$. Банковская процентная ставка считается постоянной и равной r.

Для вероятностного описания цены акции используется схема Бернулли. Каждый элементарный исход представляет собой последовательность: $\omega = (a_1, \dots a_N)$, где a_i может принимать два значения: H и T. В качестве алгебры $\mathscr{F}$ возьмем множество всех подмножеств Ω. Вероятности элементарных исходов определяются по формуле

$$\mathsf{P}(\omega) = p^{\nu(\omega)} q^{N-\nu(\omega)},$$

где $\nu(\omega)$ — число H в последовательности ω. Обозначим через $\mathscr{F}_k$ σ-алгебру, содержащую информацию о первых k элементах последовательности ω.

Цена акции описывается стохастической последовательностью

$$S_k(\omega) = u^{\nu_k(\omega)} d^{k-\nu_k(\omega)} S_0, \quad \nu_k(\omega) = \#\{i : a_i = H,\ i = 1, \dots k\},$$

где $S_0 = \mathrm{const}$ — начальная цена. Часто более удобна эквивалентная рекуррентная форма записи:

$$S_k(\omega) = (uI_{a_k=H}(\omega) + dI_{a_k=T}(\omega))S_{k-1}(\omega).$$

Банковский счет является детерминированной последовательностью:

$$B_k = (1+r)B_{k-1} = (1+r)^k B_0.$$

Арбитраж в одношаговой модели

В случае одношаговой модели имеем $\Omega = \{H, T\}$, $B_1 = (1+r)B_0$,

$$S_1(\omega) = \begin{cases} uS_0, & \text{если } \omega = H, \\ dS_0, & \text{если } \omega = T. \end{cases}$$

Величины β_1, γ_1 измеримы относительно тривиальной алгебры $\mathcal{F}_0 = \{\emptyset, \Omega\}$, т.е. постоянны. При нулевом начальном капитале условие самофинансируемости принимает вид

$$B_0\beta_1 + S_0\gamma_1 = 0.$$

Условие неотрицательности капитала $X_1 = \beta_1 B_1 + \gamma_1 S_1$ распадается на два:

$$\begin{aligned} X_1(H) &= \beta_1(1+r)B_0 + \gamma_1 uS_0 \geq 0, \\ X_1(T) &= \beta_1(1+r)B_0 + \gamma_1 dS_0 \geq 0. \end{aligned}$$

Последнее условие арбитражности портфеля состоит в том, что $X_1(H) > 0$ или $X_1(T) > 0$.
Из условия самофинансируемости следует $B_0\beta_1 = -S_0\gamma_1$, поэтому

$$X_1(H) = (u - 1 - r)\gamma_1 S_0, \quad X_1(T) = (d - 1 - r)\gamma_1 S_0.$$

Таким образом, арбитраж возможен, если

- акция заведомо лучше банковского счета: $u > d \geq 1 + r$
- банковский счет заведомо лучше акции: $1 + r \geq u > d$

В первом случае следует занять деньги в банке и купить акции: $\gamma_1 S_0 = -\beta_1 B_0 > 0$. Во втором — осуществить короткую продажу акций и деньги положить в банк: $\gamma_1 S_0 = -\beta_1 B_0 < 0$. Безрисковая прибыль в обоих случаях ничем не ограничена.

Полнота одношаговой модели

Наиболее общее платежное обязательство в одношаговой модели выглядит так:

$$f_1(\omega) = aI_{A_H} + bI_{A_T},$$

поскольку данным соотношением описывается любая $\mathcal{F}_1$-измеримая с.в.

При начальном капитале $X_0 = x$ условие самофинансируемости имеет вид

$$B_0\beta_1 + S_0\gamma_1 = x.$$

Условие воспроизводимости f_1: $X_1(\omega) = f_1(\omega)$ распадается на два

$$\begin{aligned} X_1(H) &= \beta_1(1+r)B_0 + \gamma_1 u S_0 = a, \\ X_1(T) &= \beta_1(1+r)B_0 + \gamma_1 d S_0 = b. \end{aligned}$$

Выражая $\beta_1 B_0$ из условия самофинансируемости, получаем два уравнения

$$\begin{aligned} X_1(H) &= (1+r)x + (u-1-r)\gamma_1 S_0 = a, \\ X_1(T) &= (1+r)x + (d-1-r)\gamma_1 S_0 = b \end{aligned}$$

с двумя неизвестными: x, γ_1. Решая эту систему, находим

$$\gamma_1 = \frac{a-b}{(u-d)S_0}, \quad x = \frac{\widetilde{p}a + \widetilde{p}b}{1+r},$$

$$\widetilde{p} = \frac{1+r-d}{u-d}, \quad \widetilde{q} = \frac{u-1-r}{u-d}.$$

Таким образом, данная модель является полной.

Пусть рынок является безарбитражным, тогда $\widetilde{p} > 0$, $\widetilde{q} > 0$. Поскольку $\widetilde{p} + \widetilde{q} = 1$, то можно ввести вероятностную меру $\widetilde{\mathsf{P}}$:

$$\widetilde{\mathsf{P}}(A_H) = \widetilde{p}, \quad \widetilde{\mathsf{P}}(A_T) = \widetilde{p},$$

а полученную формулу для начального капитала (справедливой цены платежного обязательства) переписать так:

$$x = (1 + r)^{-1} \widetilde{\mathsf{E}} f_1,$$

где $\widetilde{\mathsf{E}}$ — математическое ожидание по мере $\widetilde{\mathsf{P}}$. Мера P называется рыночной, а мера $\widetilde{\mathsf{P}}$ — мартингальной.

Мартингальная мера в N-шаговой биномиальной модели

В рамках N-шаговой биномиальной модели введем вероятностную меру по формуле

$$\widetilde{\mathsf{P}}(\omega) = \widetilde{p}^{\nu(\omega)}\widetilde{q}^{N-\nu(\omega)}.$$

Обозначения $\nu(\omega)$, $\widetilde{p}$, $\widetilde{q}$ введены выше. Оказывается, что относительно $\widetilde{\mathsf{P}}$ дисконтированная цена акции $\{S_k/B_k, \mathscr{F}_k\}$ и дисконтированный капитал $\{X_k/B_k, \mathscr{F}_k\}$ являются мартингалами. Действительно,

$$\widetilde{\mathsf{E}}(S_{k+1}/B_{k+1}|\mathscr{F}_k) = (u\widetilde{p} + d\widetilde{q})S_k/B_{k+1} = S_k/B_k$$

так как

$$\widetilde{\mathsf{E}}(S_{k+1}|\mathscr{F}_k) = \widetilde{\mathsf{E}}(S_k S_{k+1}/S_k|\mathscr{F}_k) = S_k\widetilde{\mathsf{E}}(S_{k+1}/S_k) = (u\widetilde{p} + d\widetilde{q})S_k.$$

и

$$\frac{u\widetilde{p} + d\widetilde{q}}{1+r} = \frac{u(1+r-d) + d(u-1-r)}{(1+r)(u-d)} = 1.$$

Чтобы установить аналогичное равенство для капитала воспользуемся соотношением

$$\frac{X_{k+1}}{B_{k+1}} = \frac{X_k}{B_k} + \gamma_{k+1}\left(\frac{S_{k+1}}{B_{k+1}} - \frac{S_k}{B_k}\right)$$

Вычислим у.м.о. и воспользуемся свойством 5:

$$\widetilde{\mathsf{E}}\left(\frac{X_{k+1}}{B_{k+1}}\middle|\mathcal{F}_k\right) = \widetilde{\mathsf{E}}\left(\frac{X_k}{B_k}\middle|\mathcal{F}_k\right) + \gamma_{k+1}\widetilde{\mathsf{E}}\left(\frac{S_{k+1}}{B_{k+1}} - \frac{S_k}{B_k}\middle|\mathscr{F}_k\right).$$

Поскольку уже доказано, что дисконтированная цена акции является $\widetilde{\mathsf{P}}$-мартингалом, то второе слагаемое в правой части последнего соотношения обращается в нуль и получаем требуемое равенство:

$$\widetilde{\mathsf{E}}(X_{k+1}/B_{k+1}|\mathscr{F}_k) = X_k/B_k.$$

Теорема о полноте многошаговой биномиальной модели

Любое $\mathscr{F}_N$-измеримое платежное обязательство f_N воспроизводимо. Начальный капитал и портфель вычисляются по формулам

$$x = \frac{\widetilde{\mathsf{E}} f_N}{(1+r)^N}, \tag{2}$$

$$\gamma_k(a_1, \dots a_{k-1}) = \frac{V_k(a_1, \dots, a_{k-1}, H) - V_k(a_1, \dots, a_{k-1}, T)}{S_k(a_1, \dots, a_{k-1}, H) - S_k(a_1, \dots, a_{k-1}, T)}, \tag{3}$$

$$V_N = f_N, \tag{4}$$

$$V_{k-1}(a_1, \dots a_{k-1}) = \frac{\widetilde{p} V_k(a_1, \dots, a_{k-1}, H) + \widetilde{q} V_k(a_1, \dots a_{k-1}, T)}{1+r}. \tag{5}$$

Доказательство

Предположим сначала, что воспроизводящая стратегия построена. Тогда $X_N = f_N$ и начальный капитал должен быть равен

$$X_0 = \widetilde{\mathsf{E}}(X_N/(1+r)^N) = \widetilde{\mathsf{E}} f_N/(1+r)^N,$$

поскольку $(X_k/B_k, \mathscr{F}_k)$ является $\widetilde{\mathsf{P}}$-мартингалом. Итак, формула (2) получена.

Построим $\widetilde{\mathsf{P}}$-мартингал $(V_k/B_k, \mathscr{F}_k)$ по формулам

$$V_N = f_N, \quad V_k/B_k = \widetilde{\mathsf{E}}(V_N/B_N|\mathscr{F}_k).$$

Мартингалы $(X_k/B_k, \mathscr{F}_k)$ и $(V_k/B_k, \mathscr{F}_k)$ совпадают при $k = N$, поэтому они совпадают при всех k.

Воспользовавшись равенством

$$X_k = (1+r)(X_{k-1} - \gamma_k S_{k-1}) + \gamma_k S_k, \tag{6}$$

находим

$$\begin{aligned} V_k(a_1, \dots a_{k-1}, H) &= (1+r)(V_{k-1} - \gamma_k S_{k-1})(a_1, \dots a_{k-1}) \\ &+ \gamma_k(a_1, \dots a_{k-1}) S_k(a_1, \dots a_{k-1}, H), \end{aligned}$$

$$\begin{aligned} V_k(a_1, \dots a_{k-1}, T) &= (1+r)(V_{k-1} - \gamma_k S_{k-1})(a_1, \dots a_{k-1}) \\ &+ \gamma_k(a_1, \dots a_{k-1}) S_k(a_1, \dots a_{k-1}, T), \end{aligned}$$

После вычитания получаем формулу (3).

Чтобы получить формулу (5) заметим, что

$$\begin{aligned} & (1+r)V_{k-1} = \widetilde{\mathsf{E}}(V_k|\mathscr{F}_{k-1}) \\ = \quad & \widetilde{\mathsf{E}}(V_k(a_1,\dots a_{k-1},H)I_{a_k=H} + V_k(a_1,\dots a_{k-1},T)I_{a_k=T}) \\ = \quad & V_k(a_1,\dots a_{k-1},H)\widetilde{p} + V_k(a_1,\dots a_{k-1},T)\widetilde{q}. \end{aligned}$$

Теперь докажем, что если начальный капитал и портфель определяются по формулам (2) – (5), то $X_k = V_k$ при всех k.
При $k = 0$ это соотношение выполнено: $X_0 = x = \widetilde{\mathsf{E}}(V_N/B_N) = V_0$.
Предполагая, что $X_{k-1} = V_{k-1}$, находим

$$X_k = (1+r)V_{k-1} + \gamma_k(S_k - (1+r)S_{k-1}).$$

Фиксируя $a_k = H$ и подставляя V_{k-1} и γ_k, получаем

$$
\begin{aligned}
& X_k(a_1, \dots a_{k-1}, H) = V_k(a_1, \dots a_{k-1}, H)\widetilde{p} + V_k(a_1, \dots a_{k-1}, T)\widetilde{q} \\
+ \quad & \frac{V_k(a_1, \dots, a_{k-1}, H) - V_k(a_1, \dots, a_{k-1}, T)}{S_k(a_1, \dots, a_{k-1}, H) - S_k(a_1, \dots, a_{k-1}, T)}(u - 1 - r)S_{k-1} \\
= \quad & V_k(a_1, \dots a_{k-1}, H)\widetilde{p} + V_k(a_1, \dots a_{k-1}, T)\widetilde{q} \\
+ \quad & \widetilde{q}(V_k(a_1, \dots a_{k-1}, H) - V_k(a_1, \dots a_{k-1}, T)) = V_k(a_1, \dots a_{k-1}, H)
\end{aligned}
$$

Аналогично, $X_k(a_1, \dots a_{k-1}, T) = V_k(a_1, \dots a_{k-1}, T)$. $\square$

Опцион Call

Опцион Call дает право купить акцию в момент времени $t = N$ по заранее оговариваемой цене K, которая называется страйком (strike). При $S_N > K$ продавец опциона должен купить акцию на рынке по цене S_N и продать ее покупателю опциона по цене K. При $S_N \leq K$ покупатель опциона не воспользуется своим правом.
Следовательно, платежное обязательство продавца опциона Call имеет вид

$$f_N = (S_N - K)^+, \quad (y - K)^+ = \max\{y - K, 0\}.$$

Пример. В случае двухшаговой биномиальной модели с параметрами $S_0 = 4$, $u = 2$, $d = 1/2$, $r = 1/4$ найти рациональную цену и хеджирующую стратегию для опциона Call со страйком $K = 5$.

Решение

Найдем параметры мартингальной меры:

$$\widetilde{p} = (1 + r - d)/(u - d) = 1/2 = \widetilde{q}$$

и платежную функцию $V_2 = f_2$:

$$\begin{aligned} V_2(HH) &= (16-5)^+ = 11, \quad V_2(HT) = V_2(TH) = (4-5)^+ = 0, \\ V_2(TT) &= (1-5)^+ = 0. \end{aligned}$$

Рациональная цена вычисляется по формуле (2):

$$x = \frac{\widetilde{\mathsf{E}} V_2}{(1+r)^2} = \frac{16}{25} \widetilde{\mathsf{P}}(HH) V_2(HH) = 44/25.$$

Найдем V_1, V_0:

$$
\begin{aligned}
V_1(H) &= (\widetilde{p}V_2(HH) + \widetilde{q}V_2(HT))/(1+r) = 22/5, \\
V_1(T) &= (\widetilde{p}V_2(TH) + \widetilde{q}V_2(TT))/(1+r) = 0, \\
V_0 &= (\widetilde{p}V_1(H) + \widetilde{q}V_1(T))/(1+r) = 44/25.
\end{aligned}
$$

Естественно, $V_0 = x$ и можно пользоваться двумя способами расчета рациональной цены.
Количество акций в портфеле вычисляется по формулам (3):

$$
\begin{aligned}
\gamma_1 &= (V_1(H) - V_1(T))/(S_1(H) - S_1(T)) = 11/25, \\
\gamma_2(H) &= (V_2(HH) - V_2(HT))/(S_2(HH) - S_2(HT)) = 11/12, \\
\gamma_2(T) &= (V_2(TH) - V_2(TT))/(S_2(HH) - S_2(HT)) = 0.
\end{aligned}
$$

Оптимальное инвестирование: мартингальный подход

Пусть инвестор имеет начальный капитал $X_0 = x$ и хочет построить инвестиционную стратегию, максимизирующую $\mathsf{E} \ln X_N$. Поскольку последовательность

$$(X_k/B_k, \mathscr{F}_k)_{k=0}^N$$

является $\widetilde{\mathsf{P}}$-мартингалом, то при любой стратегии выполнено соотношение

$$\widetilde{\mathsf{E}}(X_N/B_N) = x, \tag{7}$$

которое естественно назвать бюджетным ограничением.
С другой стороны, согласно теореме о полноте биномиальной модели рынка, для любой $\mathscr{F}_N$-измеримой случайной величины ξ, удовлетворяющей бюджетному ограничению: $\widetilde{\mathsf{E}}(\xi/B_N) = x$, существует такой портфель, что его капитал стартует с x и совпадает с ξ в момент времени $t = N$.

Таким образом, все возможные значения капитала X_N в точности описываются бюджетным ограничением (7), и получаем следующую задачу условной оптимизации:

$$\mathsf{E}\ln X_N \to \max, \quad \widetilde{\mathsf{E}} X_N = xB_N. \tag{8}$$

С учетом обозначения $Z = \widetilde{\mathsf{P}}(\omega)/\mathsf{P}(\omega)$ имеем

$$\widetilde{\mathsf{E}}(X_N) = \sum_{\omega\in\Omega} X_N(\omega)\widetilde{\mathsf{P}}(\omega) = \sum_{\omega\in\Omega} X_N(\omega)Z(\omega)\mathsf{P}(\omega) = \mathsf{E}(X_N Z).$$

Случайная величина Z называется плотностью или производной Радона-Никодима меры $\widetilde{\mathsf{P}}$ относительно P.

Для решения задачи (8) перепишем ее в следующей форме

$$\sum_{\omega \in \Omega} \ln(X_N(\omega))\mathsf{P}(\omega) \to \max_{X_N(\omega)}, \tag{9}$$

$$\sum_{\omega \in \Omega} X_N(\omega) Z(\omega) \mathsf{P}(\omega) = x B_N \tag{10}$$

и составим функцию Лагранжа

$$L(X_N, \lambda) = \sum_{\omega \in \Omega} (\ln X_N(\omega) - \lambda X_N(\omega) Z(\omega)) \mathsf{P}(\omega).$$

Приравнивая нулю производные по $X_N(\omega)$, находим

$$X_N(\omega) = \frac{1}{(\lambda Z(\omega))}.$$

Значение множителя Лагранжа λ определяется из ограничения (10):

$$\sum_{\omega\in\Omega}\frac{Z(\omega)\mathsf{P}(\omega)}{\lambda Z(\omega)}=\frac{1}{\lambda}=xB_N.$$

Итак, капитал, максимизирующий критерий (9), выглядит следующим образом

$$X_N(\omega)=xB_N/Z(\omega). \tag{11}$$

Стратегия, воспроизводящая этот капитал, состоит в воспроизведении искусственного “платежного обязательства” $xB_N/Z(\omega)$ и определяется по формулам (3).

Пример. Решить поставленую задачу оптимального инвестирования для двухшаговой биномиальной модели с параметрами $u=2$, $d=1/2$, $r=1/4$, $p=1/3$, $q=2/3$, $S_0=4$, $X_0=2$.

Решение примера

Значения $\widetilde{p} = \widetilde{q} = 1/2$ уже были найдены в примере 9.1. Найдем с.в. Z:

$$Z(HH) = \frac{\widetilde{\mathsf{P}}(HH)}{\mathsf{P}(HH)} = \frac{1/4}{1/9} = 9/4, \quad Z(TT) = \frac{\widetilde{\mathsf{P}}(TT)}{\mathsf{P}(TT)} = \frac{1/4}{4/9} = 9/16,$$

$$Z(HT) = Z(TH) = \frac{\widetilde{\mathsf{P}}(HT)}{\mathsf{P}(HT)} = \frac{1/4}{2/9} = 9/8.$$

Оптимальный конечный капитал $X_2 = V_2$ определяется по формуле (11):

$$V_2(HH) = \frac{x(1+r)^2}{Z(HH)} = 25/18, \quad V_2(TT) = \frac{x(1+r)^2}{Z(TT)} = 50/9,$$

$$V_2(HT) = V_2(TH) = \frac{x(1+r)^2}{Z(HT)} = 25/9.$$

Для определения оптимальной стратегии найдем V_1 — капитал в момент времени $t = 1$:

$$
\begin{aligned}
V_1(H) &= (\widetilde{p}V_2(HH) + \widetilde{q}V_2(HT))/(1+r) = 5/3, \\
V_1(T) &= (\widetilde{p}V_2(TH) + \widetilde{q}V_2(TT))/(1+r) = 10/3.
\end{aligned}
$$

Количество акций в портфеле вычисляется по формулам (3):

$$
\begin{aligned}
\gamma_1 &= (V_1(H) - V_1(T))/(S_1(H) - S_1(T)) = -5/18, \\
\gamma_2(H) &= (V_2(HH) - V_2(HT))/(S_2(HH) - S_2(HT)) = -25/216, \\
\gamma_2(T) &= (V_2(TH) - V_2(TT))/(S_2(HH) - S_2(HT)) = -25/27.
\end{aligned}
$$

Процесс X с $\mathsf{E}X_n^2 < \infty$ называется *квадратично интегрируемым*.

Теорема 4

Пусть M — квадратично интегрируемый мартингал, ограниченный в L^2: $\sup_{n\geq 1} \mathsf{E}M_n^2 < \infty$. Тогда M сходится в L^2.

Доказательство. Для квадратично интегрируемого мартингала верно равенство

$$\begin{aligned} \mathsf{E}(M_{n+k} - M_n)^2 &= \mathsf{E}M_{n+k}^2 - 2\mathsf{E}\mathsf{E}(M_n M_{n+k} | \mathscr{F}_n) + \mathsf{E}M_n^2 \\ &= \mathsf{E}M_{n+k}^2 - \mathsf{E}M_n^2, \quad k > 0. \end{aligned} \tag{12}$$

Отсюда следует, что последовательность $\mathsf{E}M_n^2$ является неубывающей. Если мартингал M ограничен в L^2, то последовательность $\mathsf{E}M_n^2$ сходится. Из (12) следует тогда, что последовательность M_n фундаментальна в L^2, так как

$$\sup_{k\geq 0} \|M_{n+k} - M_n\|_{L^2}^2 = \sup_{k\geq 0} \mathsf{E}(M_{n+k} - M_n)^2 = \sup_{k\geq 0} (\mathsf{E}M_{n+k}^2 - \mathsf{E}M_n^2) < \varepsilon$$

при достаточно больших n. В силу полноты L^2 процесс M сходится к некоторому элементу $M_\infty \in L^2$.

Случайная величина $\tau : \Omega \mapsto \overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{+\infty\}$ называется моментом остановки, если $\{\tau \leq n\} \in \mathscr{F}_n$, $n \in \mathbb{Z}_+$. Смысл данного условия состоит в том, что в момент любой фиксированный времени n должно быть известно, наступил ли момент τ ранее.

Пример 5

Для любого борелевского множества $G \in \mathbb{R}$ введем момент первого попадания в него процесса X:

$$\tau_G = \inf\{n \geq 0 : X_n \in G\}$$

с учетом соглашения $\inf \emptyset := +\infty$. Случайная величина τ_G является моментом остановки:

$$\{\tau_G = n\} = \{X_0 \notin G\} \cap \dots \cap \{X_{n-1} \notin G\} \cap \{X_n \in G\} \in \mathscr{F}_n,$$

$$\{\tau_G \leq n\} = \bigcup_{k=0}^{n} \{\tau_G = k\} \in \mathscr{F}_n.$$

Другой пример момента остановки — постоянная случайная величина $\tau \equiv n$.

Пример 6

В модели эксперимента с n-кратным подбрасыванием монеты положим

$$\xi_i(\omega) = \xi_i((a_1, \ldots, a_n)) = I_{\{a_i = H\}}(\omega).$$

Пусть σ-алгебра $\mathscr{F}_j$ содержит информации о первых j подбрасываниях: $\mathscr{F}_j = \sigma(\xi_1, \ldots, \xi_j)$. Тогда случайная величина

$$\tau(\omega) = \begin{cases} k, & \omega = (H, \ldots, H), \\ n, & \text{иначе}, \end{cases}$$

где $k < n$, не является моментом остановки относительно фильтрации $(\mathscr{F}_j)_{j=1}^n$, так как $\{\tau \leq k\} = \{(H, \ldots, H)\} \notin \mathscr{F}_k$. Ясно, что в момент времени k еще неизвестно, выпадет ли монета орлом при всех n подбрасываниях.

Чтобы иметь возможность определить случайную величину

$$X_\tau = \sum_{n\in\mathbb{Z}_+} X_n I_{\{\tau=n\}} + X_\infty I_{\{\tau=+\infty\}},$$

введем σ-алгебру $\mathscr{F}_\infty = \bigvee_{n=1}^\infty \mathscr{F}_n := \sigma\left(\bigcup_{n=1}^\infty \mathscr{F}_n\right)$ и $\mathscr{F}_\infty$-измеримую случайную величину X_∞.

Теорема 7 (критерий мартингальности)

Пусть $X_n \in L^1(\mathscr{F}_n)$, $n\in\mathbb{Z}_+$. Процесс X является мартингалом, если и только если

$$\mathsf{E}X_\tau = \mathsf{E}X_0 \tag{13}$$

для любого ограниченного момента остановки τ.

Доказательство. Необходимость. Пусть $\tau \le K \in \mathbb{N}$. Тогда

$$\mathsf{E}X_\tau = \sum_{j=0}^K \mathsf{E}\left(I_{\{\tau=j\}}X_j\right) = \sum_{j=0}^K \mathsf{E}\left(I_{\{\tau=j\}}X_K\right) = \mathsf{E}\left(X_K \sum_{j=0}^K I_{\{\tau=j\}}\right)$$
$$= \mathsf{E}X_K = \mathsf{E}X_0.$$

Достаточность. Рассмотрим множество $A \in \mathscr{F}_k$ и положим

$$\sigma = kI_A + (k+1)I_{A^c}.$$

Легко видеть, что σ — момент остановки:

$$\{\sigma \le n\} = \begin{cases} \varnothing, & n < k \\ A, & n = k \\ \Omega, & n > k. \end{cases}$$

Применяя (13) при $\tau = k+1$ и $\tau = \sigma$, находим

$$\mathsf{E}X_0 = \mathsf{E}X_{k+1} = \mathsf{E}X_\sigma = \mathsf{E}(X_k I_A) + \mathsf{E}(X_{k+1} I_{A^c}).$$

Следовательно, $\mathsf{E}(X_k I_A) = \mathsf{E}(X_{k+1} I_A)$, $A \in \mathscr{F}_k$. Это и означает, что X является мартингалом:

$$X_k = \mathsf{E}(X_{k+1} | \mathscr{F}_k). \quad \square$$

Лемма 8

Если τ, σ — моменты остановки, то $\tau \vee \sigma$, $\tau \wedge \sigma$, $\tau + \sigma$ будут моментами остановки.

С процессом X и моментом остановки τ свяжем «остановленный» процесс $X_n^\tau = X_{n \wedge \tau}$. Остановленный мартингал является мартингалом.

Теорема 9

Пусть X — мартингал и τ — момент остановки, тогда остановленный процесс X^τ является мартингалом.

Доказательство. Для любого ограниченного момента остановки σ верно равенство

$$\mathsf{E} X_\sigma^\tau = \mathsf{E} X_{\tau \wedge \sigma} = \mathsf{E} X_0$$

по предыдущей теореме. Из этой же теоремы следует, что X^τ является мартингалом.

Пример 10 (Задача о разорении)

Рассмотрим случайное блуждание $X_n = \sum_{k=1}^n \xi_k$, $n \geq 1$, $X_0 = 0$, где ξ_k — независимые радемахеровские величины: $\mathsf{P}(\xi_k = 1) = \mathsf{P}(\xi_k = -1) = 1/2$. Процесс X является мартингалом относительно фильтрации $\mathscr{F}_n = \sigma(\xi_0, \ldots, \xi_n)$. Положим

$$\tau = \inf\{n \in \mathbb{Z}_+ : X_n \notin (-a, b)\}, \quad a, b \in \mathbb{N}.$$

Процесс X можно рассматривать как капитал игрока, участвующего в азартной игре, где на каждом шаге он выигрывает или проигрывает один рубль с одинаковой вероятностью. Величины $\mathsf{P}(X_\tau = -a)$, $\mathsf{E}\tau$ определяют вероятность разорения и ожидаемую продолжительность игры.

Справедливы равенства $\mathsf{E}\tau = ab$,

$$\mathsf{P}(X_\tau = -a) = \frac{b}{b+a}, \qquad \mathsf{P}(X_\tau = b) = \frac{a}{b+a}. \tag{14}$$

Доказательство

Процесс $X_n^2 - n$ является мартингалом:

$$\begin{aligned}\mathsf{E}(X_{n+1}^2 - (n+1)|\mathscr{F}_n) &= \mathsf{E}(X_{n+1}^2 - X_n^2|\mathscr{F}_n) + X_n^2 - (n+1)\\ &= \mathsf{E}((X_{n+1} - X_n)^2|\mathscr{F}_n) + X_n^2 - (n+1)\\ &= X_n^2 - n.\end{aligned}$$

Следовательно, $(X_{n\wedge m}^2 - n\wedge m)_{n\geq 0}$ — мартингал при любом $m > 0$.
По критерию мартингальности,

$$\mathsf{E}X_{\tau\wedge m}^2 = \mathsf{E}(\tau\wedge m). \tag{15}$$

Здесь использовано, что момент остановки $\tau\wedge m$ ограничен.

Процесс $(X^2_{\tau\wedge m})_{m\in\mathbb{Z}_+}$ ограничен, поэтому

$$\mathsf{E}X^2_{\tau\wedge m} \to X^2_\tau, \quad m\to\infty$$

по теореме о мажорируемой сходимости. С другой стороны,

$$\mathsf{E}(\tau\wedge m) \nearrow \mathsf{E}\tau, \quad m\to\infty$$

по теореме о монотонной сходимости. Переходя в (15) к пределу при $m\to\infty$, находим

$$\mathsf{E}\tau = \mathsf{E}X^2_\tau < \infty.$$

В частности, $\tau<\infty$ п.н., и значит,

$$\mathsf{P}(X_\tau \in \{-a,b\}) = \mathsf{P}(X_\tau=-a) + \mathsf{P}(X_\tau=b) = 1. \tag{16}$$

Далее, поскольку процесс $(X_n)_{n\in\mathbb{Z}_+}$ является мартингалом, и момент остановки $\tau\wedge m$ ограничен, то $\mathsf{E}X_{\tau\wedge m}=0$. Из ограниченности процесса $(X_{\tau\wedge m})_{m\in\mathbb{Z}_+}$ вытекает, что

$$0=\lim_{m\to\infty}\mathsf{E}X_{\tau\wedge m}=\mathsf{E}X_\tau$$

по теореме о мажорируемой сходимости. В силу (16) имеем

$$0=\mathsf{E}X_\tau=-a\mathsf{P}(X_\tau=-a)+b\mathsf{P}(X_\tau=b). \tag{17}$$

Из уравнений (16), (17) вытекают равенства (14), используя которые, находим

$$\mathsf{E}\tau=\mathsf{E}X_\tau^2=a^2\mathsf{P}(X_\tau=-a)+b^2\mathsf{P}(X_\tau=b)=a^2\frac{b}{a+b}+b^2\frac{a}{a+b}=ab. \quad \square$$

Введем σ-алгебру $\mathscr{F}_\tau = \{A \in \mathscr{F}_\infty : A \cap \{\tau \le n\} \in \mathscr{F}_n, n \in \mathbb{Z}_+\}$. Смысл определения $\mathscr{F}_\tau$ можно понимать следующим образом: в каждый момент времени n известно, наступило ли $A \in \mathscr{F}_\tau$, если наступило τ. Семейство $\mathscr{F}_\tau$ называется также σ-алгеброй событий, наблюдаемых до момента времени τ. Некоторые свойства этой σ-алгебры описаны в следующей лемме.

Лемма 11

(i) *τ является $\mathscr{F}_\tau$-измеримой,*

(ii) *X_τ является $\mathscr{F}_\tau$-измеримой,*

(iii) *Если $\tau \le \sigma$, то $\mathscr{F}_\tau \subseteq \mathscr{F}_\sigma$,*

(iv) *$\mathscr{F}_\tau \cap \mathscr{F}_\sigma = \mathscr{F}_{\tau\wedge\sigma}$,*

(v) *$\{\tau \le \sigma\} \in \mathscr{F}_{\tau\wedge\sigma}$ для любых моментов остановки τ, σ.*

Теорема о свободном выборе

Пусть τ_1, τ_2 — моменты остановки, $0 \le \tau_1 \le \tau_2 \le K \in \mathbb{N}$ и M — мартингал. Тогда

$$\mathsf{E}(M_{\tau_2}|\mathscr{F}_{\tau_1}) = M_{\tau_1}. \tag{18}$$

Доказательство. Нужно доказать, что для $A \in \mathscr{F}_{\tau_1}$ верно равенство $\mathsf{E}(M_{\tau_2} I_A) = \mathsf{E}(M_{\tau_1} I_A)$, т.е.

$$\sum_{j=0}^{K} \mathsf{E}(M_j I_{A\cap\{\tau_2=j\}}) = \sum_{j=0}^{K} \mathsf{E}(M_j I_{A\cap\{\tau_1=j\}}).$$

Поскольку $A \in \mathscr{F}_{\tau_1} \subset \mathscr{F}_{\tau_2}$, то

$$A \cap \{\tau_1 = j\} \in \mathscr{F}_j, \qquad A \cap \{\tau_2 = j\} \in \mathscr{F}_j.$$

Следовательно, $\mathsf{E}(M_j I_{A\cap\{\tau_r=j\}}) = \mathsf{E}(M_K I_{A\cap\{\tau_r=j\}})$, $r = 1, 2$. Отсюда вытекает, что

$$\sum_{j=0}^{K} \mathsf{E}(M_j I_{A\cap\{\tau_r=j\}}) = \sum_{j=0}^{K} \mathsf{E}(M_K I_{A\cap\{\tau_r=j\}}) = \mathsf{E}(M_K I_A), \quad r = 1, 2.$$

Теорему о свободном выборе можно несколько усилить следующим образом.

Теорема 12

Пусть τ_1, τ_2 *— моменты остановки,* $0 \le \tau_1 \le \tau_2 \le K \in \mathbb{N}$ *и* X *— субмартингал (соотв., супермартингал). Тогда*

$$\mathsf{E}(X_{\tau_2}|\mathscr{F}_{\tau_1}) \geq X_{\tau_1}, \quad (\text{соотв.,} \ \leq X_{\tau_1}).$$

Рассматриваемые ниже *неравенства Дуба* позволяют оценивать процесс $X_n^* = \max_{i=0,\dots,n} X_i$ максимума неотрицательного субмартингала.

Теорема 13

Пусть X *— неотрицательный субмартингал. Тогда*

$$\mathsf{P}(X_n^* \geq a) \leq \frac{1}{a}\mathsf{E}(X_n I_{\{X_n^* \geq a\}}) \leq \frac{1}{a}\mathsf{E}X_n, , \quad a > 0. \tag{19}$$

Доказательство. Положим $\tau_a = \inf\{k \geq 0 : \; X_k \geq a\}$. Тогда $\tau_a \leq n$, если и только если $X_n^* \geq a$. Далее,

$$\begin{aligned}\mathsf{E}(X_n I_{\{\tau_a \leq n\}}) &= \mathsf{E}(\mathsf{E}(X_n|\mathscr{F}_{\tau_a \wedge n}) I_{\{\tau_a \leq n\}}) \geq \mathsf{E}(X_{n\wedge\tau_a} I_{\{\tau_a \leq n\}}) \\ &= \mathsf{E}(X_{\tau_a} I_{\{\tau_a \leq n\}}) \geq a\mathsf{P}(\tau_a \geq n).\end{aligned}$$

Таким образом,

$$\mathsf{P}(X_n^* \geq a) = \mathsf{P}(\tau_a \leq n) \leq \frac{\mathsf{E}(X_n I_{\{\tau_a \leq n\}})}{a} \leq \frac{\mathsf{E}X_n}{a}. \quad \square$$

Теорема 14

Пусть X — неотрицательный квадратично интегрируемый субмартингал. Тогда

$$\mathsf{E}(X_n^*)^2 \le 4\mathsf{E}X_n^2. \tag{20}$$

Заметим, что

$$\mathsf{E}(X_n^*)^2 = \mathsf{E}\int_0^{X_n^*} 2a\,da = \mathsf{E}\int_0^{\infty} I_{\{a\le X_n^*\}} 2a\,da = \int_0^{\infty} \mathsf{E}I_{\{a\le X_n^*\}} 2a\,da$$

по теореме Фубини. С учетом неравенства (19) отсюда следует, что

$$\begin{aligned}
\mathsf{E}(X_n^*)^2 &= 2\int_0^{\infty} \mathsf{P}(X_n^* \ge a)a\,da \le 2\int_0^{\infty} \mathsf{E}(X_n I_{\{X_n^*\ge a\}})\,da \\
&= 2\mathsf{E}\left(X_n \int_0^{X_n^*} da\right) \\
&= 2\mathsf{E}(X_n X_n^*) \le 2(\mathsf{E}X_n^2)^{1/2}(\mathsf{E}(X_n^*)^2)^{1/2}
\end{aligned}$$

по неравенству Коши-Буняковского. □

Для вещественных чисел $a < b$ введем последовательность моментов остановки

$$\tau_1 = \inf\{n \geq 0 : X_n \leq a\}, \qquad \tau_2 = \inf\{n > \tau_1 : X_n \geq b\},$$

$$\tau_{2m-1} = \inf\{n > \tau_{2m-2} : X_n \leq a\}, \qquad \tau_{2m} = \inf\{n > \tau_{2m-1} : X_n \geq b\}.$$

Случайная величина

$$U_N[a,b](X) = \begin{cases} 0, & \tau_2 > N, \\ \max\{m : \tau_{2m} \leq N\}, & \tau_2 \leq N \end{cases}$$

называется *числом пересечений вверх* $U_N[a,b](X)$ интервала $[a,b]$ последовательностью X_n, $n = 0, \ldots, N$. Очевидно, что $U_N[a,b](X)$ совпадает с наибольшим k, для которого существуют

$$q_1 < r_1 < q_2 < r_2 < \cdots < q_k < r_k \leq N$$

со следующим свойством:

$$X_{q_i} \leq a, \quad X_{r_i} \geq b, \quad i = 1, \ldots, k. \tag{21}$$

Лемма 15 (лемма Дуба о числе пересечений)

Если X — супермартингал, то

$$\mathsf{E}U_N[a,b](X) \le \frac{\mathsf{E}(X_N - a)^-}{b-a}, \qquad (22)$$

Доказательство. Будем рассматривать X_n как дисконтированную цену рискового актива. Пусть γ — предсказуемый процесс портфеля γ_n. Тогда капитал после k шагов равен

$$Y_k = \sum_{n=1}^{k} \gamma_n (X_n - X_{n-1}).$$

Пусть $a < b$. Рассмотрим следующую стратегию

(1) дождаться пока X станет меньше a,

(2) купить одну единицу актива и держать ее пока $X \le b$,

(3) продать как только X станет больше b,

(4) перейти на шаг (1).

Формально,

$$\gamma_1 = I_{\{X_0<a\}},$$

$$\gamma_n = I_{\{\gamma_{n-1}=1\}} I_{\{X_{n-1}\le b\}} + I_{\{\gamma_{n-1}=0\}} I_{\{X_{n-1}<a\}}, \quad n \ge 2.$$

Ясно, что каждое пересечение вверх интервала $[a,b]$ увеличивает значение выигрыша Y по крайней мере на $b-a$. Строя нижнюю оценку Y_N, нужно учесть также возможное падение X на последнем временно́м интервале. Таким образом,

$$Y_N \ge (b-a)U_N[a,b](X) - (X_N - a)^-.$$

Легко видеть, что процесс Y является супермартингалом, поэтому

$$0 = Y_0 \ge \mathsf{E}Y_N \ge (b-a)\mathsf{E}U_N[a,b](X) - \mathsf{E}(X_N - a)^-,$$

т.е. выполнено неравенство (22). $\square$

Лемма 16

Числовая последовательность X_n сходится в $\overline{\mathbb{R}}$, если и только если

$$\sup_N U_N[a,b](X) < \infty$$

для всех $a, b \in \mathbb{Q}$, $a < b$.

Доказательство. Если для некоторой пары $a < b$, $a, b \in \mathbb{Q}$ верно, что

$$\sup_N U_N[a,b](X) = \lim_{N\to\infty} U_N[a,b](X) = \infty,$$

то существуют подпоследовательности

$$X_{q_i} < a < b < X_{r_i}, \quad i \in \mathbb{N}.$$

Следовательно, $\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n$.
Наоборот, если существуют $a, b \in \mathbb{Q}$:

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n,$$

то легко построить последовательности $q_1 < r_1 < \dots < q_n < r_n < \dots$ такие, что $X_{q_i} \to \liminf_{n\to\infty} X_n$, $X_{r_i} \to \limsup_{n\to\infty} X_n$ и выполнены неравенства (21). $\square$

Теорема о сходимости мартингалов п.н.

Теорема 17

Пусть M — мартингал, ограниченный в L^1: $\sup_{n\geq 0} \mathsf{E}|M_n| < \infty$.
Тогда M сходится $\mathbb{R}$ п.н., и $M_\infty = \lim_{n\to\infty} M_n \in L^1$.

Доказательство. Из ограниченности M в L^1 и леммы 15 вытекает, что

$$\sup_N \mathsf{E} U_N[a,b](M) = \lim_{N\to\infty} \mathsf{E} U_N[a,b](M) < \infty.$$

По теореме о монотонной сходимости отсюда следует, что

$$\lim_{N\to\infty} U_N[a,b](M) = \sup_N U_N[a,b](M) < \infty \quad \text{п.н.}$$

для любых фиксированных $a < b$. В силу счетности множества $\{a, b \in \mathbb{Q} : a < b\}$, существует $\Omega' \in \mathscr{F}$ с $\mathsf{P}(\Omega') = 1$ такое, что

$$\sup_N U_N[a,b](M) < \infty \quad \text{для всех } a, b \in \mathbb{Q},\ \ a < b \text{ на } \Omega'.$$

По предыдущей лемме отсюда следует, что M сходится к расширенной случайной величине M_∞ на Ω'. Но по лемме Фату

$$\mathsf{E}|M_\infty| \le \liminf_{n\to\infty} \mathsf{E}|M_n| < \infty.$$

Следовательно, $M_\infty \in L^1$. $\square$

Пример 18

Приведем другое доказательство того, что момент τ первого выхода случайного блуждания X из интервала $(-a, b)$ конечен п.н.
Заметим, что процесс X^τ является ограниченным мартингалом и значит сходится п.н. На множестве $\{\tau = \infty\}$ данный процесс совпадает с X. Но X не может сходиться, так как $|X_{n+1} - X_n| = 1$. Следовательно, $\mathsf{P}(\tau = \infty) = 0$.

Семейство случайных величин $(\xi_\alpha)_{\alpha\in J}$, где J — любое множество индексов, называется *равномерно интегрируемым*, если

$$\sup_{\alpha\in J} \mathsf{E}\left(|\xi_\alpha| I_{\{|\xi_\alpha|\geq c\}}\right) \to 0, \quad c\to\infty.$$

Легко видеть, что если семейство $(\xi_\alpha)_{\alpha\in I}$ равномерно интегрируемо, то оно ограничено в L^1.

Лемма 19

Любое из следующих свойств влечет равномерную интегрируемость:

- ξ_α *равномерно ограничены:* $|\xi_\alpha|\leq C$,
- ξ_α *ограничены в* L^p*:* $\sup_{\alpha\in I}\mathsf{E}|\xi_\alpha|^p<\infty$ *с* $p>1$,
- $\xi_\alpha=\mathsf{E}(\xi|\mathscr{H}_\alpha)$, *где* $\xi\in L^1$ *и* $\mathscr{H}_\alpha$ *— некоторое семейство* σ*-алгебр.*

Для равномерно интегрируемой последовательности сходимость по вероятности равносильна сходимости в L^1.

Теорема 20

Пусть $X_n, X \in L^1(\mathscr{F})$. Последовательность X_n сходится X в L^1, если и только если $X_n \xrightarrow{\mathsf{P}} X$ и $(X_n)_{n\in\mathbb{Z}_+}$ равномерно интегрируема.

Теорема о сходимости мартингалов в L^1

Теорема 21

Для мартингала M следующие условия эквивалентны:

(i) *M сходится в L^1 к некоторой случайной величине M_∞,*

(ii) *M допускает представление $M_n = \mathsf{E}(M_\infty | \mathscr{F}_n)$ с $M_\infty \in L^1$,*

(iii) *M равномерно интегрируем.*

Доказательство. (i) $\Longrightarrow$ (ii). Поскольку M является мартингалом, то

$$\mathsf{E}(M_n I_A) = \mathsf{E}(M_k I_A), \quad A \in \mathscr{F}_n, \quad n < k.$$

Из сходимости M_k к M_∞ в L^1 следует, что

$$\mathsf{E}(M_n I_A) = \mathsf{E}(M_\infty I_A),$$

т.е. выполнено (ii).
(ii) $\Longrightarrow$ (iii). Вытекает из леммы 19.
(iii) $\Longrightarrow$ (i). Из равномерной интегрируемости вытекает ограниченность в L^1. Следовательно M_n сходится п.н. Но тогда, по предыдущей теореме M_n сходится в L^1. $\square$

Доопределяя равномерно интегрируемый мартингал $(M_n)_{n\in\mathbb{Z}_+}$ значением $M_\infty = \lim_{n\to\infty} M_n$, получаем мартингал на $\overline{\mathbb{Z}}_+$:

$$M_k = \mathsf{E}(M_n|\mathscr{F}_k), \quad k, n \in \overline{\mathbb{Z}}_+.$$

Для равномерно интегрируемых мартингалов теорема о свободном выборе верна без предположения об ограниченности моментов остановки.

Теорема 22

Пусть M — равномерно интегрируемый мартингал на $\overline{\mathbb{Z}}_+$ и $\tau_1 \le \tau_2$ — моменты остановки. Тогда

$$\mathsf{E}(M_{\tau_2}|\mathscr{F}_{\tau_1}) = M_{\tau_1}.$$

Оказывается, что время достижения случайным блужданием любого фиксированного уровня конечно п.н. Но его ожидаемое значение бесконечно.

Пример 23

Для случайного блуждания X положим

$$\tau_b = \inf\{n \geq 0 : X_n \geq b\}, \quad b \in \mathbb{N}.$$

Имеем, $\mathsf{P}(\tau_b < \infty) = 1$, $\mathsf{E}\tau_b = \infty$.

Доказательство. Момент τ достижения границы интервала $(-a, b)$ можно представить в виде $\tau_b \wedge \tau_{-a}$. Ясно, что если $X_{\tau_b \wedge \tau_{-a}}(\omega) = b$, то $\tau_b(\omega) < \infty$. Используя соответствующее равенство (14), находим

$$\mathsf{P}(\tau_b(\omega) < \infty) \geq \mathsf{P}(X_{\tau_b \wedge \tau_{-a}}(\omega) = b) = \frac{a}{b+a}.$$

Переходя к пределу при $a \to +\infty$, заключаем, что $\mathsf{P}(\tau_b < \infty) = 1$.

Далее, если $\mathsf{E}\tau_b < \infty$, то из равенства (15) вытекает, что

$$\mathsf{E}X^2_{\tau_b \wedge m} \le \mathsf{E}\tau_b < \infty.$$

Но тогда остановленный мартингал $X^{\tau_b} = (X_{\tau_b \wedge m})_{m=0}^{\infty}$ равномерно интегрируем и, по теореме 22,

$$\mathsf{E}X_{\tau_b} = \mathsf{E}X_0 = 0. \tag{23}$$

Однако, это невозможно, так как $X_{\tau_b} = 1$ п.н. $\square$
Противоречие в равенстве (23) показывает также, что условие равномерной интегрируемости в теореме 22 существенно.

Предельный переход по расширяющейся последовательности σ-алгебр вычисляется в следующей теореме Леви.

Теорема 24

Пусть $\xi \in L^1$. Тогда равномерно интегрируемый мартингал $\mathsf{E}(\xi|\mathscr{F}_n)$ *сходится к* $\mathsf{E}(\xi|\mathscr{F}_\infty)$ *в* L^1 *при* $n \to \infty$.

Доказательство.

В силу теоремы 21 $M_n = \mathsf{E}(\xi|\mathscr{F}_n)$ сходится к M_∞ в L^1. Нужно лишь доказать, что $M_\infty = \mathsf{E}(\xi|\mathscr{F}_\infty)$.

Очевидно, что достаточно рассмотреть случай $\xi \geq 0$. Меры

$$\mu(A) := \mathsf{E}(\mathsf{E}(\xi|\mathscr{F}_\infty)I_A), \quad \nu(A) := \mathsf{E}(M_\infty I_A)$$

совпадают на алгебре $\cup_{n=0}^{\infty}\mathscr{F}_n$, так как для $A \in \mathscr{F}_n$ имеют место равенства

$$\mu(A) = \mathsf{E}(\xi I_A), \quad \nu(A) = \mathsf{E}(M_n I_A) = \mathsf{E}(\mathsf{E}(\xi|\mathscr{F}_n)I_A) = \mathsf{E}(\xi I_A).$$

По теореме о монотонном классе, примененной к π-системе $\mathcal{I} = \{A : A \in \cup_{n=0}^{\infty}\mathscr{F}_n\}$ и семейству $\mathcal{H}$ ограниченных $\mathscr{F}_\infty$-измеримых функций, отсюда следует, что μ и ν совпадают на $\mathscr{F}_\infty = \sigma\left(\cup_{n=0}^{\infty}\mathscr{F}_n\right)$, т.е.

$$\mathsf{E}(\mathsf{E}(\xi|\mathscr{F}_\infty)I_A) = \mathsf{E}(M_\infty I_A), \quad A \in \mathscr{F}_\infty.$$

Таким образом, $\mathsf{E}(\xi|\mathscr{F}_\infty) = M_\infty$.

Пример 25

Закон 0 или 1 Колмогорова. Пусть $(\xi_n)_{n=1}^{\infty}$ — последовательность независимых случайных величин. Рассмотрим «хвостовую» σ-алгебру

$$\mathscr{G} = \cap_{n=1}^{\infty} \sigma(\xi_n, \xi_{n+1}, \dots).$$

Если $F \in \mathscr{G}$, то $\mathsf{P}(F) = 0$ или $\mathsf{P}(F) = 1$.

Доказательство.

Заметим, что $F \in \mathscr{G}$ не зависит от $\mathscr{F}_n = \sigma(\xi_1, \dots, \xi_n)$. Рассмотрим равномерно интегрируемый мартингал

$$M_n = \mathsf{E}(I_F | \mathscr{F}_n) = \mathsf{E}(I_F) = \mathsf{P}(F).$$

По теореме Леви,

$$M_\infty = \mathsf{E}(I_F | \mathscr{F}_\infty) = \mathsf{P}(F).$$

Но $\mathscr{G} \subset \mathscr{F}_\infty$. Следовательно, I_F — $\mathscr{F}_\infty$-измерима и $I_F = \mathsf{P}(F)$ п.н. Отсюда вытекает, что $I_F = 0$ или $I_F = 1$ п.н.

Например, если ξ_n независимы, то событие

$$A = \left\{ \sum_{k=1}^{\infty} \xi_k \text{ сходится} \right\} = \left\{ \sum_{k=n}^{\infty} \xi_k \text{ сходится} \right\} \in \sigma(\xi_n, \xi_{n+1}, \dots)$$

принадлежит $\mathscr{G} = \cap_{n=1}^{\infty} \sigma(\xi_n, \xi_{n+1}, \dots)$. Поэтому, согласно закону 0 или 1 Колмогорова, ряд $\sum_{k=1}^{\infty} \xi_k$ сходится или расходится с вероятностью 1.

Обращенным мартингалом называется согласованный с фильтрацией $(\mathscr{F}_n)_{n\in\mathbb{Z}_-}$ процесс $(M_n)_{n\in\mathbb{Z}_-}$, $\mathbb{Z}_- := -\mathbb{Z}_+$ со следующими свойствами: $M_n \in L^1$,

$$M_{n-1} = \mathsf{E}(M_n|\mathscr{F}_{n-1}), \quad n \in \mathbb{Z}_-.$$

Теорема 26

Для обращенного мартингала M сходимость

$$M_{-\infty} = \lim_{n\to-\infty} M_n$$

имеет место п.н. и в L^1.

Доказательство.

Обозначим через $U_{-N}[a,b](X)$ число пересечений вверх интервала $[a,b]$ последовательностью $(M_n)_{n=-N}^{0}$. По лемме 15 и теореме о монотонной сходимости справедлива оценка

$$\mathsf{E}\sup_N U_{-N}[a,b](X) = \lim_{N\to\infty} \mathsf{E}U_{-N}[a,b](X) \leq \frac{\mathsf{E}(X_0-a)^-}{b-a}.$$

По лемме 16 отсюда следует, что X_n сходится п.н. в $\overline{\mathbb{R}}$ при $n \to -\infty$. Из равномерной интегрируемости $M_n = \mathsf{E}(M_0|\mathscr{F}_n)$, $n \in \mathbb{Z}_-$ вытекает конечность $M_{-\infty}$ и тот факт, что сходимость имеет место в L^1 (см. теорему 20).

Пример 27

Усиленный закон больших чисел. Пусть $(\xi_n)_{n=1}^{\infty}$ — последовательность независимых одинаково распределенных интегрируемых случайных величин. Тогда

$$\frac{S_n}{n} \to \mathsf{E}\xi_1 \quad \text{п.н. и в } L^1, \quad \text{где} \;\; S_n = \sum_{i=1}^{n} \xi_i.$$

Доказательство.

Рассмотрим последовательность σ-алгебр

$$\mathscr{G}_{-n} = \sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \dots).$$

Равенство

$$\mathsf{E}(\xi_k | \mathscr{G}_{-n}) = \mathsf{E}(\xi_1 | \mathscr{G}_{-n}), \quad k = 2, \dots, n$$

верно из соображений симметрии. Суммируя эти равенства, находим

$$S_n = \mathsf{E}(S_n | \mathscr{G}_{-n}) = \sum_{k=1}^{n} \mathsf{E}(\xi_k | \mathscr{G}_{-n}) = n\mathsf{E}(\xi_1 | \mathscr{G}_{-n}).$$

Таким образом,

$$\frac{S_n}{n} = \mathsf{E}(\xi_1 | \mathscr{G}_{-n})$$

сходится п.н. и в L^1 по теореме 26. По закону 0 или 1 Колмогорова $\lim_{n\to\infty} S_n/n = c$ п.н. для некоторой константы c. При этом

$$c = \mathsf{E} \lim_{n\to\infty} S_n/n = \lim_{n\to\infty} \mathsf{E}(S_n/n) = \mathsf{E}\xi_1.$$

Мартингалы с непрерывным временем

Пусть имеется фильтрованное вероятностное пространство

$$(\Omega, \mathscr{F}, \mathsf{P}, (\mathscr{F}_t)_{t\in\mathbb{R}_+}).$$

Согласованный с фильтрацией $\mathbb{F} = (\mathscr{F}_t)_{t\in\mathbb{R}_+}$ случайный процесс $X = (X_t)_{t\in\mathbb{R}_+}$, $\mathsf{E}|X_t| < \infty$ с непрерывными траекториями $t \mapsto X_t(\omega)$, называется

- *мартингалом*, если $\mathsf{E}(X_t|\mathscr{F}_s) = X_s$,
- *субмартингалом*, если $\mathsf{E}(X_t|\mathscr{F}_s) \geq X_s$,
- *супермартингалом*, если $\mathsf{E}(X_t|\mathscr{F}_s) \leq X_s$

при $s < t$.

Обозначим через $X_t^* = \sup_{0 \le s \le t} X_s$ процесс максимума. Максимальные неравенства Дуба (19), (20) обобщаются на случай непрерывного времени.

Лемма 28

Пусть X — неотрицательный непрерывный субмартингал. Тогда

$$\mathsf{P}\left(\sup_{0 \le t \le T} X_t \ge a\right) \le \frac{\mathsf{E} X_T}{a}, \quad a > 0. \tag{24}$$

Если, кроме того, $\mathsf{E}|X_T|^2 < \infty$, *то*

$$\mathsf{E}(X_T^*)^2 \le 4\mathsf{E}X_T^2. \tag{25}$$

Теоремы о сходимости мартингалов в п.н. и в L^1 также сохраняют силу.

Броуновское движение

Случайный процесс $(W_t)_{t\geq 0}$ называется *броуновским движением* (или *винеровским процессом*) относительно фильтрации $\mathbb{F}$, если

(a) $W_0 = 0$ п.н.,

(b) $W_t - W_s$ не зависит от $\mathscr{F}_s$, $t > s$,

(c) $W_t - W_s$ имеет нормальное распределение $N(0, t)$ со средним 0 и дисперсией t:

$$\mathsf{P}(W_t - W_s \in (a, b)) = \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-x^2/(2t)}\, dx,$$

(d) W непрерывен.

Отметим, что независимость приращений броуновского движения плохо согласуется с регулярностью траекторий. Действительно, в каждый момент времени приращение траектории не зависит от предыстории. Следовательно, следует ожидать, что в каждой точки траектории имеется излом. Строгий результат состоит в том, что почти все траектории броуновского движения недифференцируемы ни в одной точке. Для стохастического анализа более важно, что почти все траектории W имеет неограниченную вариацию. Более подробно этот вопрос будет рассмотрен далее.

Пример

Пусть W — броуновское движение. Тогда процессы

$$W, \quad X = W^2 - t, \quad Y = e^{\theta W_t - \frac{\theta^2}{2} t}$$

являются мартингалами:

$$\begin{aligned}
\mathsf{E}(W_t | \mathscr{F}_s) &= \mathsf{E}(W_t - W_s + W_s | \mathscr{F}_s) = \mathsf{E}(W_t - W_s | \mathscr{F}_s) + \mathsf{E}(W_s | \mathscr{F}_s) \\
&= \mathsf{E}(W_t - W_s) + W_s = W_s, \\
\mathsf{E}(X_t | \mathscr{F}_s) &= \mathsf{E}((W_t - W_s + W_s)^2 - t | \mathscr{F}_s) = \mathsf{E}((W_t - W_s)^2 | \mathscr{F}_s) \\
&+ \mathsf{E}(2W_s(W_t - W_s) | \mathscr{F}_s) + W_s^2 - t = \mathsf{E}(W_t - W_s)^2 \\
&+ 2W_s \mathsf{E}(W_t - W_s | \mathscr{F}_s) + W_s^2 - t = W_s^2 - s = X_s, \\
\mathsf{E}(Y_t | \mathscr{F}_s) &= e^{-\frac{\theta^2}{2} t} \mathsf{E}\left(e^{\theta(W_t - W_s + W_s)} | \mathscr{F}_s \right) = e^{-\frac{\theta^2}{2} t} e^{\theta W_s} \mathsf{E}\left(e^{\theta(W_t - W_s)} \right) = Y_s.
\end{aligned}$$

В последнем равенстве использована формула

$$
\begin{aligned}
\mathsf{E}\left(e^{\theta(W_t-W_s)}\right) &= \frac{1}{\sqrt{2\pi(t-s)}}\int_{\mathbb{R}} e^{\theta x} e^{-\frac{x^2}{2(t-s)}}\,dx \\
&= \frac{e^{\frac{1}{2}(t-s)\theta^2}}{\sqrt{2\pi(t-s)}}\int_{\mathbb{R}} e^{-\frac{1}{2(t-s)}(x-(t-s)\theta)^2} dx = e^{\frac{1}{2}(t-s)\theta^2}.
\end{aligned}
$$

Пример 29

Для броуновского движения справедлив закон больших чисел:

$$\lim_{t\to\infty} \frac{W_t}{t} = 0 \quad \text{п.н.}$$

По неравенству Дуба

$$\mathsf{E}\left(\sup_{2^n\le t\le 2^{n+1}} \left(\frac{W_t}{t}\right)^2\right) \le \frac{1}{2^{2n}}\mathsf{E}\left(\sup_{2^n\le t\le 2^{n+1}} W_t^2\right) \le \frac{4}{2^{2n}}\mathsf{E}W_{2^{n+1}}^2 = \frac{4}{2^{2n}}2^{n+1} =$$

По неравенству Чебышева

$$\mathsf{P}\left(\sup_{2^n\le t\le 2^{n+1}} \left(\frac{W_t}{t}\right)^2 \ge \varepsilon^2\right) \le \frac{8}{n\varepsilon^2}.$$

По лемме Бореля-Кантелли

$$\mathsf{P}\left(\sup_{2^n\le t\le 2^{n+1}}\frac{|W_t|}{t}\ge\varepsilon\quad\text{б.ч.}\right)=0.$$

Таким образом, для почти всех ω существует $k(\omega)$:

$$\sup_{2^n\le t\le 2^{n+1}}\frac{|W_t|}{t}<\varepsilon,\quad n\ge k.$$

Теорема 17 о сходимости супермартингалов переносится на случай непрерывного времени.

Теорема 30

Пусть X — мартингал, ограниченный в L^1: $\sup_{t\geq 0} \mathsf{E}|X_t| < \infty$. Тогда $X_\infty := \lim_{t\to\infty} X_t$ существует п.н. в $\mathbb{R}$.

Теорема 21 о сходимости в L^1 также сохраняет силу. Ее доказательство не отличается от случая дискретного времени.

Теорема 31

Для непрерывного мартингала M следующие условия эквивалентны:

(i) *M сходится в L^1 к некоторой случайной величине M_∞,*

(ii) *M допускает представление $M_t = \mathsf{E}(M_\infty | \mathscr{F}_t)$ с $M_\infty \in L^1$,*

(iii) *M равномерно интегрируем.*

Ясно, что равномерно интегрируемый мартингал порождает мартингал $(M_t)_{t\in\overline{\mathbb{R}}_+}$ на $\overline{\mathbb{R}}_+$.

Моменты остановки

Случайная величина $\tau : \Omega \mapsto \overline{\mathbb{R}}_+$ называется *моментом остановки*, если $\{\tau \le t\} \in \mathscr{F}_t$, $t \in \mathbb{R}_+$. Говорят, что фильтрация $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ *непрерывна справа*, если $\mathscr{F}_t = \cap_{\varepsilon>0}\mathscr{F}_{t+\varepsilon}$.

Лемма 32

Пусть τ — момент остановки, тогда $\{\tau < t\} \in \mathscr{F}_t$. Если $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ непрерывна справа, то верно и обратное.

Доказательство. Поскольку $\{\tau \le s\} \in \mathscr{F}_s \subset \mathscr{F}_t$, $s < t$, то

$$\{\tau < t\} = \bigcup_{s<t}\{\tau \le s\} \subset \mathscr{F}_t.$$

Пусть теперь $\{\tau < t\} \in \mathscr{F}_t$. Второе утверждение вытекает из соотношения

$$\{\tau \le t\} = \bigcap_{s>t}\{\tau < s\} = \bigcap_{s\in(t,t+\varepsilon]\cap\mathbb{Q}}\{\tau < s\} \in \mathscr{F}_{t+\varepsilon},$$

которое справедливо при любом $\varepsilon > 0$. $\square$

Введем момент первого попадания процесса X в множество $B \subset \mathbb{R}^d$:

$$\tau_B = \inf\{t \geq 0 : X_t \in B\}.$$

Теорема 33

Пусть X — согласованный с $\mathbb{F}$ процесс, траектории которого непрерывны справа или слева. Если G открыто и фильтрация $\mathbb{F}$ непрерывна справа, то τ_G — момент остановки.

Следующий пример показывает, что условие непрерывности справа существенно.

Пример. Пусть ξ — стандартная нормальная величина: $\xi \sim N(0,1)$. Рассмотрим естественную фильтрацию $\mathscr{F}_t = \sigma(X_s, s \le t)$ случайного процесса $X_t = t\xi$. Рассмотрим момент первого попадания X в множество $(0,\infty)$:

$$\tau_{(0,+\infty)} = \inf\{t \ge 0 : X_t \in (0,+\infty)\} = \begin{cases} 0, & \xi(\omega) > 0, \\ +\infty, & \xi(\omega) \le 0. \end{cases}$$

Ясно, что

$$\{\tau_{(0,+\infty)} \le 0\} = \{\xi > 0\} \notin \mathscr{F}_0 = \sigma(X_0) = \{\emptyset, \Omega\},$$

и $\tau_{(0,+\infty)}$ не является моментом остановки относительно естественной фильтрации процесса X, которая разрывна в нуле: $\{\xi > 0\}$ принадлежит $\mathscr{F}_t$, $t > 0$, но не принадлежит $\mathscr{F}_0$.

Говорят, что фильтрация $\mathscr{F}_t)_{t\geq 0}$ *удовлетворяет обычным условиям*, если она непрерывна справа и каждая из σ-алгебр $\mathscr{F}_t$ содержит все *нулевые множества* N, для которых

$$\inf\{\mathsf{P}(A) : N \subset A, A \in \mathscr{F}\} = 0.$$

Естественная фильтрация броуновского движения, как и фильтрация, рассмотренная в предыдущем примере, разрывна при $t = 0$. Однако известно, что *пополненная фильтрация* броуновского движения, полученная присоединением к $\mathscr{F}_t$ всех нулевых множеств, удовлетворяет обычным условиям.

Лемма 34

(i) *Пусть* τ, σ *— моменты остановки, тогда* $\tau \wedge \sigma$, $\tau \vee \sigma$ *— моменты остановки.*

(ii) *Пусть* $\tau_1 \leq \tau_2 \leq \ldots$ *— неубывающая последовательность моментов остановки, тогда* $\sup_{n\geq 1} \tau_n$ *— момент остановки.*

(iii) *Пусть* $\tau_1 \geq \tau_2 \geq \ldots$ *— невозрастающая последовательность моментов остановки, тогда* $\inf_{n\geq 1} \tau_n$ *— момент остановки.*

Теорема. *Пусть* X *— согласованный с* $\mathbb{F}$ *процесс с непрерывными траекториями. Если* G *замкнуто и фильтрация* $\mathbb{F}$ *непрерывна справа, то* τ_G *— момент остановки.*

Любой момент остановки может быть аппроксимирован убывающей последовательностью моментов остановки со счетным набором значений. Этот важный факт позволяет сводить многие вопросы к случаю дискретного времени.

Лемма. *Для любого момента остановки* $\tau : \Omega \mapsto \overline{\mathbb{R}}_+$ *последовательность моментов остановки*

$$\tau_n = \sum_{k \in \mathbb{Z}_+} \frac{k+1}{2^n} I_{[k/2^n,(k+1)/2^n)}(\tau) + (+\infty) I_{\{\tau=+\infty\}} : \Omega \mapsto \overline{\mathbb{Z}}_+$$

монотонно убывает и равномерно сходится к τ*:*

$$\tau_n \searrow \tau, \quad 0 \le \tau_n - \tau \le 1/2^n.$$

Как и в случае дискретного времени, введем σ-алгебру событий

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_\infty : A \cap \{\tau \le t\} \in \mathscr{F}_t, t \in \mathbb{R}_+\},$$

наблюдаемых до момента времени τ. Здесь

$$\mathscr{F}_\infty = \bigvee_{t \ge 0} F_t = \sigma\left(\bigcup_{t \ge 0} \mathscr{F}_t\right).$$

Лемма. *Пусть τ, σ, τ_n — моменты остановки.*

(i) *Случайная величина τ измерима относительно $\mathscr{F}_\tau$.*

(ii) *Если $\tau \le \sigma$, то $\mathscr{F}_\tau \subset \mathscr{F}_\sigma$.*

(iii) $\mathscr{F}_\tau \cap \mathscr{F}_\sigma = \mathscr{F}_{\tau \wedge \sigma}$.

(iv) *Если $\tau_n \searrow \tau$ и $\mathbb{F}$ непрерывна справа, то $\cap_{n=1}^\infty \mathscr{F}_{\tau_n} = \mathscr{F}_\tau$.*

Для $\mathbb{F}$-согласованного случайного процесса X и момента остановки τ введем остановленный процесс $X_t^\tau = X_{t\wedge\tau}$ и случайную величину $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$. Последний случай требует доопределения процесса $(X_t)_{t\in\mathbb{R}_+}$ $\mathscr{F}_\infty$-измеримой случайной величиной X_∞.

Лемма. *Пусть τ — момент остановки, X — непрерывный согласованный процесс и фильтрация $\mathbb{F}$ непрерывна справа. Тогда*

(i) *случайная величина X_τ является $\mathscr{F}_\tau$-измеримой,*

(ii) *остановленный процесс X^τ является непрерывным и согласованным, а фильтрация $\mathbb{F}^\tau = (\mathscr{F}_{t\wedge\tau})_{t\geq 0}$ — непрерывной справа.*

Сформулируем теорему о свободном выборе для случая непрерывного времени.

Теорема 35

(i) Для непрерывного равномерно интегрируемого мартингала $(M_t)_{t\in\overline{\mathbb{R}}_+}$ *и любых моментов остановки* $\sigma \le \tau$ *верно равенство*

$$M_\sigma = \mathsf{E}(M_\tau | \mathscr{F}_\sigma). \tag{26}$$

(ii) Для непрерывного мартингала M *равенство (26) верно при условии ограниченности* τ*.*

В терминах моментов остановки формулируется следующий удобный критерий мартингальности процесса аналогичный теореме 7.

Теорема 36

Пусть $\mathbb{F}$ *непрерывна справа. Непрерывный процесс* M *является мартингалом, если и только если* $\mathsf{E}M_\tau = \mathsf{E}M_0$ *для любого ограниченного момента остановки.*

Остановленный непрерывный мартингал является мартингалом.

Теорема 37

Пусть M *— непрерывный мартингал и* τ*-момент остановки. Если* $\mathbb{F}$ *непрерывна справа, то* M^τ *является мартингалом.*

Пример

Пусть W — броуновское движение, относительно некоторой фильтрации, непрерывной справа. Положим

$$\tau = \inf\{t \geq 0 : W_t \notin (-a, b)\}, \quad a, b > 0.$$

Тогда $\mathsf{E}\tau = ab$,

$$\mathsf{P}(W_\tau = -a) = \frac{b}{b+a}, \qquad \mathsf{P}(W_\tau = b) = \frac{a}{b+a}. \tag{27}$$

Доказательство. Процесс $W_t^2 - t$. При любом $u > 0$ процесс $(W_{t\wedge u}^2 - t \wedge u)_{t \geq 0}$ является мартингалом и

$$\mathsf{E}W_{\tau\wedge u}^2 = \mathsf{E}(\tau \wedge u), \tag{28}$$

так как момент остановки $\tau \wedge u$ ограничен. Далее, процесс $(W_{\tau\wedge u}^2)_{u \geq 0}$ ограничен, поэтому

$$\mathsf{E}W_{\tau\wedge u}^2 \to \mathsf{E}W_\tau^2, \quad u \to \infty$$

по теореме о мажорируемой сходимости.

С другой стороны, $\mathsf{E}(\tau \wedge u) \nearrow \mathsf{E}\tau, \quad u \to \infty$ по теореме о монотонной сходимости. Переходя в (28) к пределу при $u \to \infty$, находим

$$\mathsf{E}\tau = \mathsf{E}W_\tau^2 < \infty.$$

В частности, $\tau < \infty$ п.н., и значит,

$$\mathsf{P}(W_\tau \in \{-a, b\}) = \mathsf{P}(W_\tau = -a) + \mathsf{P}(W_\tau = b) = 1. \tag{29}$$

Из критерия мартингальности и теоремы о свободном выборе вытекает, что

$$0 = \lim_{u \to \infty} \mathsf{E}W_{\tau \wedge u} = \mathsf{E}W_\tau.$$

В силу (29) имеем

$$0 = \mathsf{E}W_\tau = -a\mathsf{P}(W_\tau = -a) + b\mathsf{P}(W_\tau = b). \tag{30}$$

Из уравнений (29), (30) вытекает, что имеют место равенства (27), из которых следует, что

$$\mathsf{E}\tau = \mathsf{E}W_\tau^2 = a^2\mathsf{P}(W_\tau = -a) + b^2\mathsf{P}(W_\tau = b) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab. \quad \square$$

Пример

Момент остановки

$$\tau_b = \inf\{t \geq 0 : W_t \geq b\}, \quad b \in \mathbb{R}_+.$$

обладает следующими свойствами: $\mathsf{P}(\tau_b < \infty) = 1$, $\mathsf{E}\tau_b = \infty$.
Доказательство. Момент τ выхода из интервала $(-a, b)$ можно представить в виде $\tau_b \wedge \tau_{-a}$. Ясно, $\tau = \tau_b$ на множестве $\{W_\tau = b\}$. Следовательно,

$$\mathsf{P}(\tau_b < \infty) \geq \mathsf{P}(\tau_b < \infty, W_\tau = b) = \mathsf{P}(W_\tau = b) = \frac{a}{a+b}.$$

Переходя к пределу при $a \to +\infty$, заключаем, что $\mathsf{P}(\tau_b < \infty) = 1$.
Далее, если $\mathsf{E}\tau_b < \infty$, то из равенства (28) вытекает, что

$$\mathsf{E}W^2_{\tau_b \wedge u} \leq \mathsf{E}\tau_b < \infty.$$

Но тогда остановленный мартингал $W^{\tau_b} = (W_{\tau_b \wedge u})_{u \geq 0}$ равномерно интегрируем и $\mathsf{E}W_{\tau_b} = W_0 = 0$ по теореме о свободном выборе. Но это невозможно, так как $W_{\tau_b} = 1$ п.н. $\square$

Марковские процессы: общие определения

Пусть $\mathbb{T} = \mathbb{Z}_+$ или $\mathbb{T} = [0, \infty)$. Рассмотрим семейство фильтрованных вероятностных пространств $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathsf{P}_x)$, $x \in \mathbb{R}^d$ и случайный процесс $X = (X_t)_{t\in\mathbb{T}}$ со значениями в $\mathbb{R}^d$, согласованный с фильтрацией $(\mathscr{F}_t)_{t\in\mathbb{T}}$. Вместо $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ можно рассматривать другое измеримое пространство $(E, \mathscr{E})$. В частности, в примерах будет рассматриваться конечное или счетное множество E с самой богатой σ-алгеброй $\mathscr{E}$.

Семейство $(Q_t)_{t\in\mathbb{T}}$ переходных вероятностей из $\mathbb{R}^d$ в $\mathbb{R}^d$ называется *переходной функцией*, если

$$Q_0(x, B) = \delta_x(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B, \end{cases} \tag{31}$$

$x \in \mathbb{R}^d$, и выполняются *уравнения Колмогорова-Чепмена*:

$$Q_{t+u}(x, B) = \int_{\mathbb{R}^d} Q_t(x, dy) Q_u(y, B), \quad t, u \geq 0,$$

для всех $x \in \mathbb{R}^d$, $B \in \mathscr{B}(\mathbb{R}^d)$.

Пусть имеется переходная функция такая, что

$$Q_t(x, B) = \mathsf{P}_x(X_t \in B). \tag{32}$$

Процесс X называется *(однородным) марковским процессом* с переходной функцией $(Q_t)_{t\in\mathbb{T}}$, если

$$\mathsf{E}_x(g(X_{t+s})|\mathscr{F}_t) = \int g(y)Q_s(X_t, dy), \quad \mathsf{P}_x\text{-п.н.} \tag{33}$$

для любой ограниченной борелевской функции g. Марковским процессом удобно назвать также пятерку $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, X, \mathsf{P}_x)$.
Равенство (33) означает, что при определении будущего поведения процесса X существенна только информация о настоящем, а всю информацию о прошлом, которая содержится в $\mathscr{F}_t$, можно отбросить.
Из (33) следует также, что

$$\mathsf{E}_x(g(X_{t+s})|X_t) = \int g(y)Q_s(X_t, dy), \quad \mathsf{P}_x\text{-п.н.}$$

Таким образом, при фиксированном s, $Q_s(x, B)$ является условным распределением X_{t+s} относительно X_t для всех мер P_x.

Равенство (32) показывает, что $Q_s(x, B)$ следует интерпретировать как вероятность того, что процесс X, стартующий из состояния x попадает в множество B в момент времени t. Отметим также, что уравнение Колмогорова-Чепмена являются следствием условий (31) – (33). Действительно, из (33) следует, что

$$
\begin{aligned}
Q_{t+u}(X_0, B) &= \int I_B(y) Q_{t+u}(X_0, dy) = \mathsf{E}_x(I_B(X_{t+u})|\mathscr{F}_0) \\
&= \mathsf{E}_x(\mathsf{E}_x(I_B(X_{t+u})|\mathscr{F}_t)|\mathscr{F}_0) = \mathsf{E}_x\left(\int I_B(y) Q_u(X_t, dy)|\mathscr{F}_0\right) \\
&= \mathsf{E}_x(Q_u(X_t, B)|\mathscr{F}_0) = \int Q_u(y, B) Q_t(X_0, dy), \quad \mathsf{P}_x\text{-п.н.}
\end{aligned} \tag{34}
$$

Условия (2.3.1), (2.3.2) можно переписать так:

$$
\int g(y) Q_0(x, dy) = g(x),
$$

$$
\int g(y) Q_t(x, dy) = \mathsf{E}_x g(X_t), \tag{35}
$$

где g — любая ограниченная борелевская функция на $\mathbb{R}^d$.

В частности,

$$\mathsf{E}_x g(X_0) = \int g(y) Q_0(x, dy) = g(x).$$

Используя эту формулу, найдем математическое ожидание левой и правой частей (34):

$$\mathsf{E}_x Q_{t+u}(X_0, B) = Q_{t+u}(x, B),$$

$$\mathsf{E}_x \int Q_u(y, B) Q_t(X_0, dy) = \int Q_u(y, B) Q_t(x, dy).$$

Таким образом, Q удовлетворяет уравнениям Колмогорова-Чепмена. Сопоставляя формулы (2.3.3), (35), получаем соотношение

$$\mathsf{E}_x(g(X_{t+u})|\mathscr{F}_t) = \int Q_u(X_t, dy) g(y) = \mathsf{E}_{X_t} g(X_u), \quad \mathsf{P}_x\text{-п.н.},$$

где $\mathsf{E}_{X_t} g(X_u) := \mathsf{E}_x g(X_u)|_{x=X_t}$.

Сформулируем общий результат о существовании марковского процесса с заданной переходной функцией.

Теорема 38

Для любой переходной функции $(Q_t)_{t\in\mathbb{T}}$ *существует семейство вероятностных пространств* $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathsf{P}_x)$, $x \in \mathbb{R}^d$ *и согласованный с* $(\mathscr{F}_t)_{t\in\mathbb{T}}$ *марковский случайный процесс с переходной функцией* $(Q_t)_{t\in\mathbb{T}}$.

Пример 39

Рассмотрим семейство переходных ядер из $\mathbb{R}$ в $\mathbb{R}$:

$$Q_0(x,B)=\delta_x(B);\quad Q_t(x,B)=\frac{1}{\sqrt{2\pi t}}\int_B \exp\left(-\frac{(y-x)^2}{2t}\right)dy,\quad t>0.$$

Отметим, что $Q_t(x,\cdot)\sim N(x,t)$, $t>0$.
Покажем, что данное семейство является переходной функцией, т.е. удовлетворяет уравнениям Колмогорова-Чепмена. Заметим, что

$$\int_{\mathbb{R}} Q_t(x,dy)Q_u(y,B)=\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-\frac{(y-x)^2}{2t}}dy\frac{1}{\sqrt{2\pi u}}\int_B e^{-\frac{(z-y)^2}{2u}}dz$$
$$=\int_B\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi u}}\,e^{-\frac{(z-y)^2}{2u}}\frac{1}{\sqrt{2\pi t}}e^{-\frac{(y-x)^2}{2t}}dydz=\int_B\int_{\mathbb{R}} f_\xi(z-y)f_\eta(y)\,dydz,$$

где f_ξ, f_η — плотности случайных величин $\xi\sim N(0,u)$, $\eta\sim N(x,t)$, заданные на некотором вероятностном пространстве $(\Omega',\mathscr{F}',\mathbb{P}')$.

Следовательно

$$f_\zeta(z) = \int_{\mathbb{R}} f_\xi(z-y) f_\eta(y)\, dy$$

является плотностью суммы *независимых* случайных величин ξ и η. Но тогда

$$\mathsf{E} e^{i\theta(\xi+\eta)} = \mathsf{E} e^{i\theta\xi} \mathsf{E} e^{i\theta\eta} = e^{-\theta^2 u/2} e^{ix\theta - \theta^2 t/2} = e^{ix\theta - \theta^2(t+u)/2},$$

где математическое ожидание вычисляется относительно P' и использована формула для характеристической функции нормального распределения. Таким образом, $\zeta \sim N(x, t+u)$ и

$$\int_{\mathbb{R}} Q_t(x, dy) Q_u(y, B) = \int_B f_\zeta(z) dz = \int_B Q_{t+u}(x, dz) = Q_{t+u}(x, B).$$

По теореме 38 существует марковский процесс $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, X_t, \mathsf{P}_x)$ с переходной функцией $(Q_t)_{t\geq 0}$. Покажем, что относительно P_0 процесс X удовлетворяет условиям (a) – (c) из определения броуновского движения.

По формулам (32), (31),

$$\mathsf{P}_0(X_0 = 0) = Q_0(0, \{0\}) = 1.$$

Далее, покажем, что $X_t - X_s$ не зависит от $\mathscr{F}_s$ при $t > s$:

$$\mathsf{E}_0(e^{i\theta(X_t - X_s)} | \mathscr{F}_s) = e^{-i\theta X_s} \mathsf{E}_0(e^{i\theta X_t} | \mathscr{F}_s) = e^{-i\theta X_s} \int e^{i\theta y} Q_{t-s}(X_s, dy)$$
$$= e^{-i\theta X_s} e^{i\theta X_s - \theta^2 (t-s)/2} = e^{-\theta^2 (t-s)/2} = \mathsf{E}_0(e^{i\theta(X_t - X_s)}). \quad (36)$$

Здесь использовано, что $Q_{t-s}(x, \cdot) \sim N(x, t-s)$. Из полученного равенства следует, что $X_t - X_s$ не зависит от $\mathscr{F}_s$. Наконец, последнее равенство в (36) означает, что $X_t - X_s \sim N(0, t-s)$.

Построенный в примере 39 процесс X называется *предброуновским движением*: он удовлетворяет всем условиям из определения броуновского движения кроме непрерывности траекторий. Существование броуновского движения вытекает из следующей *теоремы Колмогорова о непрерывной модификации*.

Теорема 40

Пусть $(X_t)_{t\geq 0}$ *— случайный процесс для которого существуют константы* $\alpha, \beta, C > 0$ *такие, что*

$$\mathsf{E}|X_t - X_s|^\beta \leq C|t-s|^{1+\alpha}. \tag{37}$$

Тогда у процесса X *существует непрерывная модификация.*

Для построенного предброуновского движения X имеем $X_t - X_s = \sqrt{t-s}\cdot Z$, $Z \sim N(0,1)$, и неравенство (37) выполняется с $\beta = 4$, $\alpha = 1$:

$$\mathsf{E}_0|X_t - X_s|^4 = (t-s)^2\mathsf{E}_0|Z|^4.$$

Пусть $\tau : \Omega \mapsto \mathbb{T}$ — конечный момент остановки. Марковский процесс X называется *строго марковским*, если

$$\mathsf{E}_x(g(X_{\tau+s})|\mathscr{F}_\tau) = \int Q_s(X_\tau, dy)g(y), \quad \mathsf{P}_x\text{-п.н.} \tag{38}$$

для любой ограниченной борелевской функции g.

Теорема 41

В случае дискретного времени: $\mathbb{T} = \mathbb{Z}_+$, любой марковский процесс является строго марковским.

Доказательство.

Покажем, что

$$\mathsf{E}_x(g(X_{\tau+s})|\mathscr{F}_\tau) = \mathsf{E}_{X_\tau} g(X_s) \quad \mathsf{P}_x\text{-п.н.} \tag{39}$$

для ограниченной борелевской функции g и конечного момента остановки τ. Пусть $A \in \mathscr{F}_\tau$, тогда $A \cap \{\tau = k\} \in \mathscr{F}_k$. Суммируя равенства

$$\begin{aligned} \mathsf{E}_x(g(X_{k+s})I_{\{\tau=k\}}I_A) &= \mathsf{E}_x(\mathsf{E}_x(g(X_{k+s})|\mathscr{F}_k)I_{\{\tau=k\}\cap A}) \\ &= \mathsf{E}_x(\mathsf{E}_{X_k} g(X_s)I_{\{\tau=k\}\cap A}) \end{aligned}$$

по $k \in \mathbb{Z}_+$ и используя теорему о мажорируемой сходимости, находим

$$\mathsf{E}_x(g(X_{\tau+s})I_A) = \mathsf{E}_x(\mathsf{E}_{X_\tau} g(X_s)I_A),$$

что равносильно (39).

Фундаментальное уравнение для марковских процессов с дискретным временем

Пусть $(X_t)_{t\in\mathbb{Z}_+}$ — марковский процесс в $\mathbb{R}^d$ и $G \subset \mathbb{R}^d$ — борелевское подмножество. Введем неотрицательные ограниченные борелевские функции

$$f : G \mapsto \mathbb{R}, \quad g : G^c \mapsto \mathbb{R}$$

и обозначим через τ момент выхода процесса X из множества G:

$$\tau = \inf\{k \geq 0 : X_k \notin G\}.$$

Предположим, что $\tau < \infty$ P_x-п.н. и рассмотрим семейство функционалов

$$v(x) = \mathsf{E}_x \left(\sum_{k=0}^{\tau-1} f(X_k) + g(X_\tau) \right), \quad x \in G.$$

Можно считать, что под знаком математического ожидания стоит выигрыш, который подсчитывается для каждой траектории процесса X до ее выхода из множества G. Тогда v — средний (ожидаемый) выигрыш. Интуиция подсказывает, что ожидаемый выигрыш $v(x)$ равен сумме выигрыша $f(x)$ на первом шаге и ожидаемого выигрыша, подсчитанного для состояния системы после первого шага: $Tv(x) := \mathsf{E}_x v(X_1)$. Таким образом,

$$v(x) = f(x) + \mathsf{E}_x v(X_1) = f(x) + Tv(x), \quad x \in G. \tag{40}$$

Данное рассуждение называется *методом первого шага*, а равенство (40) — *фундаментальным уравнением*. Граничное условие

$$v(x) = g(x), \quad x \in G^c \tag{41}$$

выполняется по определению.

В следующей теореме рассматривается более общий случай, в котором допускаются бесконечные значения τ.

Теорема 42

Пусть $f \geq 0$, $g \geq 0$, тогда функция

$$u(x) = \mathsf{E}_x \left(\sum_{k=0}^{\tau-1} f(X_k) + I_{\{\tau<\infty\}} g(X_\tau) \right), \quad x \in \mathbb{R}^d$$

является наименьшим неотрицательным решением задачи (40), (41). Случай бесконечных значений $u(x)$ не исключается.

Доказательство

Запишем задачу (40), (41) в следующей форме:

$$\begin{aligned} v(x) &= I_G(x)(f(x) + Tv(x)) + I_{G^c}(x)g(x) \\ &= (I_G f + I_{G^c} g)(x) + I_G T v(x). \end{aligned} \tag{42}$$

В силу неотрицательности v,

$$v \geq I_G f + I_{G^c} g$$

Подставляя данную оценку в правую часть (42), выводим неравенство

$$v \geq I_G f + I_{G^c} g + I_G T(I_G f + I_{G^c} g) = (I + I_G T)(I_G f + I_{G^c} g).$$

Используя данную оценку и применяя тот же прием, находим

$$\begin{aligned} v &\geq I_G f + I_{G^c} g + I_G T(I + I_G T)(I_G f + I_{G^c} g) \\ &= (I + I_G T + (I_G T)^2)(I_G f + I_{G^c} g). \end{aligned}$$

Продолжая этот процесс, приходим к оценке

$$v \geq \sum_{n \geq 0} (I_G T)^n (I_G f + I_{G^c} g). \tag{43}$$

Покажем, что

$$(I_G T)^n \varphi(x) = \mathsf{E}_x(I_G(X_0) I_G(X_1) \ldots I_G(X_{n-1}) \varphi(X_n)), \quad n \geq 1 \quad (44)$$

для неотрицательной (или ограниченной) борелевской функции $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$. Действительно,

$$\begin{aligned} I_G T \varphi(x) &= I_G(x) \mathsf{E}_x \varphi(X_1) = \mathsf{E}_x(I_G(X_0)\varphi(X_1)), \\ (I_G T)^2 \varphi(x) &= (I_G T)(I_G T \varphi)(x) = \mathsf{E}_x(I_G(X_0)(I_G T\varphi)(X_1)) \\ &= \mathsf{E}_x(I_G(X_0) I_G(X_1)(T\varphi)(X_1)) = \mathsf{E}_x(I_G(X_0) I_G(X_1) \mathsf{E}_{X_1} \varphi(X_2)) \\ &= \mathsf{E}_x(I_G(X_0) I_G(X_1) \mathsf{E}_x(\varphi(X_2)|\mathscr{F}_1)) = \mathsf{E}_x(I_G(X_0) I_G(X_1) \varphi(X_2)). \end{aligned}$$

В случае произвольного n равенство (44) можно доказать по индукции.

Используя определение τ и формулу (44), находим

$$\begin{aligned} \mathsf{E}_x(I_{\{\tau=n\}}g(X_n)) =& \mathsf{E}_x(I_G(X_0)\dots I_G(X_{n-1})I_{G^c}(X_n)g(X_n)) \\ =& (I_G T)^n(I_{G^c}g)(x). \end{aligned}$$

Следовательно,

$$\sum_{n\geq 0}(I_G T)^n(I_{G^c}g)(x) = \sum_{n\geq 0}\mathsf{E}_x(I_{\{\tau=n\}}g(X_n)) = \mathsf{E}_x(I_{\{\tau<\infty\}}g(X_\tau)). \tag{45}$$

Аналогичным образом,

$$\begin{aligned} \mathsf{E}_x(I_{\{\tau>n\}}f(X_n)) =& \mathsf{E}_x(I_G(X_0)\dots I_G(X_{n-1})I_G(X_n)f(X_n)) \\ =& (I_G T)^n(I_G f)(x). \end{aligned}$$

Следовательно,

$$\sum_{n\geq 0}(I_G T)^n(I_G f)(x) = \sum_{n\geq 0}\mathsf{E}_x(I_{\{\tau>n\}}f(X_n)) = \mathsf{E}_x\sum_{n=0}^{\tau-1}f(X_n). \tag{46}$$

Формулы (43), (45), (46) показывают, что для любого решения v задачи (40), (41) справедлива оценка

$$v(x) \geq u(x) := \sum_{n\geq 0} (I_G T)^n (I_G f + I_{G^c} g)(x) = \mathsf{E}_x \left(\sum_{n=0}^{\tau-1} f(X_n) + I_{\{\tau<\infty\}} g(X_\tau) \right)$$

Остается проверить, что u является решением задачи (40), (41):

$$\begin{aligned} u &= I_G f + I_{G^c} g + \sum_{n\geq 1} (I_G T)^n (I_G f + I_{G^c} g) \\ &= I_G f + I_{G^c} g + I_G T \sum_{n\geq 0} (I_G T)^n (I_G f + I_{G^c} g) \\ &= I_G f + I_{G^c} g + I_G T u. \end{aligned}$$

Пример 43

Пусть ξ_k — независимые радемахеровские случайные величины: $\mathsf{P}(\xi_k = 1) = \mathsf{P}(\xi_k = -1) = 1/2$, заданные на некотором вероятностном пространстве $(\Omega', \mathscr{F}', \mathsf{P}')$. Рассмотрим марковский процесс X случайного блуждания с независимыми приращениями ξ_k. Процесс X принимает значения в $\mathbb{Z}$ и

$$Tv(x) = \mathsf{E}_x v(X_1) = \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1), \quad x \in \mathbb{Z}.$$

Рассмотрим задачу о разорении из примера 10. Пусть

$$\tau = \inf\{n \in \mathbb{Z}_+ : X_n \notin (-a, b)\}, \quad a, b \in \mathbb{N}.$$

В области $G = (-a, b)$, $a, b \in \mathbb{N}$ рассмотрим краевую задачу (40), (41):

$$v(x) = \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1) + f(x), \quad x \in (-a, b) \cap \mathbb{Z}, \tag{47}$$

$$v(a) = g(-a), \quad v(b) = g(b). \tag{48}$$

В данной ситуации, которая является типичной, достаточно задать значения g лишь на части G^c.

Нетрудно показать, что решение задачи (47), (48) единственно (упр. 2.27). Конечность τ была установлена в примере 10. Следовательно, по теореме 42,

$$v(x) = \mathsf{E}_x \left(\sum_{k=0}^{\tau-1} f(X_k) + g(X_\tau) \right).$$

Полагая $f = 0$, $g(-a) = 1$, $g(b) = 0$ заключаем, что v представляет собой вероятность выхода из интервала $(-a, b)$ через левую границу:

$$v(x) = \mathsf{E}_x I_{\{X_\tau = -a\}} = \mathsf{P}_x(X_\tau = -a).$$

Полагая $f = 1$, $g = 0$, заключаем, что v совпадает с ожидаемой продолжительностью игры:

$$v(x) = \mathsf{E}_x \sum_{k=0}^{\tau-1} 1 = \mathsf{E}_x \tau.$$

Итак, пусть $f = 0$, $g(-a) = 1$, $g(b) = 0$:

$$v(x) - \frac{1}{2}v(x+1) - \frac{1}{2}v(x-1) = 0, \quad x \in (a,b) \cap \mathbb{Z}; \quad v(-a) = 1, \ v(b) = 0.$$

Решением данной задачи является функция $v(x) = (b-x)/(b+a)$. В частности, вероятность того, что процесс, стартующий в начале координат, выйдет из интервала $(-a, b)$ через левую границу (вероятность разорения), равна

$$v(0) = b/(b+a).$$

Данная формула совпадает с (14).
Далее, пусть $f = 1$, $g = 0$:

$$v(x) - \frac{1}{2}v(x+1) - \frac{1}{2}v(x-1) = 1, \quad x \in (a,b) \cap \mathbb{Z}; \quad v(-a) = 0, \ v(b) = 0.$$

Будем искать решение в виде

$$v(x) = A + Bx + Cx^2.$$

Подставляя данную функцию в уравнение, находим

$$C\left(x^2 - \frac{1}{2}(x+1)^2 - \frac{1}{2}(x-1)^2\right) = -C = 1.$$

Граничные условия дают уравнения

$$v(-a) = A - Ba - a^2 = 0, \quad A + Bb - b^2 = 0.$$

Отсюда $A = ab$, $B = b - a$. Таким образом, ожидаемая продолжительность игры равна

$$v(x) = ab + (b-a)x - x^2.$$

Формула $v(0) = ab$ совпадает с полученной в примере 10.

Пример 44

В обозначениях примера 43 покажем, что ожидаемое число посещений начала координат симметричным случайным блужданием X бесконечно:

$$u(x) = \mathsf{E}_x \sum_{n=0}^{\infty} I_{\{0\}}(X_n) = \infty, \quad x \in \mathbb{Z}.$$

Пусть $G = \mathbb{Z}$. Тогда процесс X не может покинуть G: $\tau = \infty$, и значит u является наименьшим неотрицательным решением уравнения

$$v(x) = I_{\{0\}}(x) + \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1), \quad x \in \mathbb{Z}. \tag{49}$$

Если $v(x) < \infty$ для некоторого x, то $v(x \pm 1) < \infty$. Отсюда следует, что $v(x) < \infty$ для всех x. Таким образом, достаточно доказать, что уравнение (49) не имеет решений, которые конечны при всех $x \in \mathbb{Z}$.

Перепишем уравнение (49) в виде

$$w(x) = 2I_{\{0\}}(x) + w(x+1), \quad w(x) = v(x) - v(x-1). \tag{50}$$

Отсюда следует, что

$$w(x) = A, \quad x < 0; \qquad w(x) = B, \quad x > 0,$$

где A, B — некоторые константы. Из условия неотрицательности v вытекает, что $A \le 0$, $B \ge 0$, так как

$$v(-x-1) = v(-x) - A = v(-x+1) - 2A = v(0) - (x+1)A,$$
$$v(x) = v(x-1) + B = v(x-2) + 2B = v(0) + Bx,$$

при $x \ge 2$. Полагая $x = -1$ и $x = 0$, из (50) находим

$$A = w(-1) = w(0) = 2 + w(1) = 2 + B,$$

что невозможно.

Пример 45

В примере 23 были найдены вероятность достижения симметричным случайным блужданием фиксированного уровня $b \in \mathbb{N}$ и математическое ожидания времени достижения указанного уровня. Получим эти результаты с помощью теоремы 42.

Положим $G = (-\infty, b) \cap \mathbb{Z}$. Для симметричного случайного блуждания на $\mathbb{Z}$ задача (40), (41) принимает вид

$$v(x) = f(x) + \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1), \quad x \in (-\infty, b) \cap \mathbb{Z}, \quad v(b) = g(b).$$

Пусть сначала $f = 0$, $g = 1$. Найдем наименьшее неотрицательное решение задачи

$$v(x) = \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1), \quad x \in (-\infty, b) \cap \mathbb{Z}, \quad v(b) = 1. \tag{51}$$

Рассуждения примера 44 показывают, что функция $w(x) = v(x) - v(x-1)$ является постоянной при $x < b$. Следовательно, функция v линейна:

$$v(x) = Ax + B, \quad x \in (-\infty, b) \cap \mathbb{Z}. \tag{52}$$

Константа $u(x) = 1$ является решением краевой задачи (51), и других постоянных решений нет. Из вида семейства решений (52) ясно, что u — наименьшее неотрицательное решение. По теореме 42,

$$u(x) = \mathsf{E}_x I_{\{\tau<\infty\}} g(X_\tau) = \mathsf{P}_x(\tau < \infty), \quad \tau = \inf\{n \geq 0 : X_\tau = b\}.$$

Таким образом, вероятность достижения симметричным случайным блужданием любого уровня b равна 1.

Рассмотрим теперь случай $f = 1$, $g = 0$ и покажем, что единственным неотрицательным решением задачи

$$v(x) = 1 + \frac{1}{2}v(x+1) + \frac{1}{2}v(x-1), \quad x \in (-\infty, b) \cap \mathbb{Z}, \quad v(b) = 0 \quad (53)$$

является функция $u(x) = +\infty$, $x \in (-\infty, b) \cap \mathbb{Z}$, $u(b) = 0$. Как указано в примере 44, достаточно установить, что уравнение не имеет решений, которые конечны при всех x. Как уже было отмечено, общим решением однородной задачи (51) является линейная функция:

$$\hat{u}(x) = Ax + b, \quad x \in (-\infty, b) \cap \mathbb{Z}.$$

Легко проверить, что функция

$$\hat{v}(x) = -x^2, \quad x \in (-\infty, b) \cap \mathbb{Z}$$

является частным решением неоднородной задачи (53). Таким образом,

$$v(x) = Ax + b - x^2, \quad x \in (-\infty, b) \cap \mathbb{Z}.$$

Но эта функция не является неотрицательной.
Итак, по теореме 42,

$$u(x) = \mathsf{E}_x \sum_{k=0}^{\tau-1} 1 = \mathsf{E}_x \tau = \infty, \quad x \in (-\infty, b) \cap \mathbb{Z},$$

т.е. ожидаемое время достижения уровня b бесконечно.

Гейм в теннисе

В начале гейма оба игрока имеют 0 очков. Каждый раз, когда игрок 1 получает очко, его счет поднимается на ступень вверх в следующей иерархии

$$0 \mapsto 15 \mapsto 30 \mapsto 40.$$

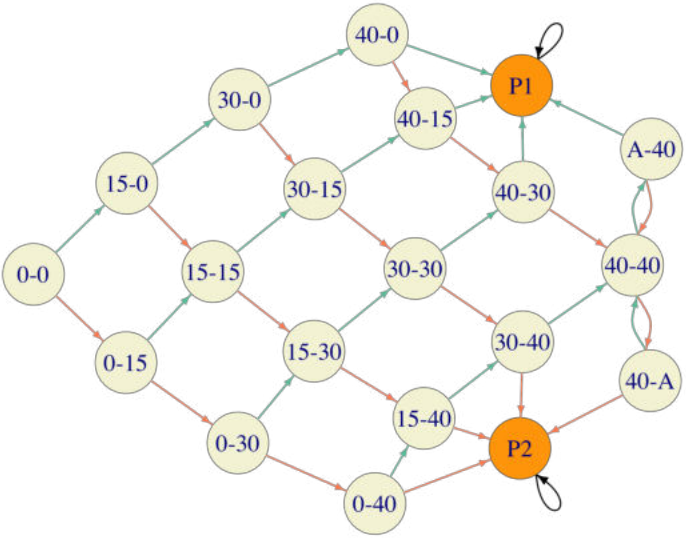
Когда один из игроков имеет 40 и выигрывает очко, могут произойти три вещи:

- если игрок 2 имеем 30 или меньше, игрок 1 выигрывает гейм,
- если игрок 2 имеет 40, счет игрока 1 переходит в состояние "больше",
- если счет игрока 2 "больше", счет игрока 1 не меняется, но счет игрока 2 понижается до 40.

Наконец, если игрок 1 имеет «больше» и выигрывает очко, то он выигрывает гейм. Ситуация полностью симметрична для игрока 2. Мы предполагаем, что вероятность того, что игрок 1 выиграет очко, равна $p \in (0,1)$, независимо от текущего счета и предшествующих событий.
Введем следующее пространство состояний:

$$S = \Big\{(0,0),(0,15),(0,30),(0,40),(15,0),(15,15),(15,30),(15,40),(30,0),\\ (30,15),(30,30),(30,40),(40,0),(40,15),(40,30),(40,40),(40,A),\\ (A,40),P_1,P_2\Big\}.$$

Здесь A (“advantage”) соответствует состоянию “больше” и P_1 (соотв., P_2) соответствует тому, что выиграл игрок 1 (соотв., игрок 2). Начальным состоянием всегда является $(0,0)$.


40-0
P1
30-0
40-15
A-40
15-0
30-15
40-30
0-0
15-15
30-30
40-40
0-15
15-30
30-40
40-A
15-40
0-30
P2
0-40

Достижимость

Пусть $(X_n)_{n\in\mathbb{Z}_+}$ — марковская цепь с конечным пространством состояний S. Моментом первого попадания в множество $B \subset S$ называется момент остановки

$$\tau_B = \min\{n \in \mathbb{Z}_+ : X_n \in B\}.$$

Если B состоит только из одного элемента, то будем писать τ_i вместо $\tau_{\{i\}}$.
Для рассмотренного примера вероятность выигрыша игрока 1 может быть выражена следующим образом:

$$\mathsf{P}[\text{выигрыш игрока 1}] = \mathsf{P}[\tau_{i_1} < \tau_{i_2}],$$

где $i_1 = P_1$, $i_2 = P_2$.

Удобно рассматривать одну и ту же цепь Маркова с разными начальными распределениями. Положим

$$\mathsf{P}_i[A] := \mathsf{P}[A|X_0 = i], \quad \mathsf{E}_i[Y] := \mathsf{E}[Y|X_0 = i]$$

для любого события A и любой случайной величины Y. Заметим, что $\mathsf{P}_i[X_1 = j] = p_{ij}$ и $\mathsf{P}_i[X_n = j] = p_{ij}^{(n)}$ для любого n.
Говорят, что состояние $i \in S$ достижимо из $j \in S$ (обозначение $i \to j$), если

$$\mathsf{P}_i[\tau_j < \infty] > 0.$$

Как нетрудно проверить, это равносильно тому, что $p_{ij}^{(n)} > 0$ для некоторого $n \in \mathbb{Z}_+$. Отношение достижимости является рефлексивным и транзитивным:

- $i \to i$,
- если $i \to j, j \to k$, то $i \to k$.

Периодом $d(i)$ состояния $i \in S$ называется наибольший общий делитель множества моментов возврата

$$R(i) = \{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$$

в состояние i. Если $R(i) = \emptyset$, то полагаем $d(i) = 1$. Состояние $i \in S$ называется апериодическим, если $d(i) = 1$.

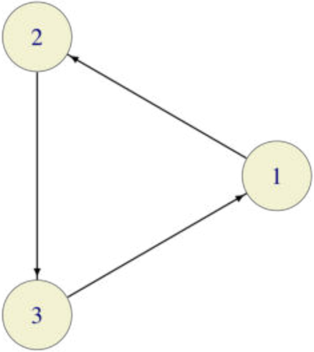
Пример 46

Для двух марковских цепей с тремя состояниями и переходными матрицами

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

найти множества моментов возврата и периоды для всех состояний.

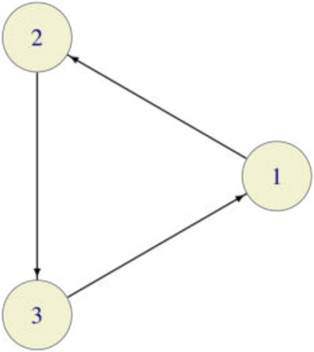
Для первой цепи множества моментов возврата одинаковы для всех состояний: $R(i) = \{3, 6, 9, 12, \dots\}$. Следовательно, $d(i) = 3$ для всех $i \in \{1, 2, 3\}$.



Для второй цепи, с внешне похожим переходным графом, ситуация кардинально другая:

$$
\begin{aligned}
R(1) &= \{3, 4, 5, 6, \dots\}, \\
R(2) &= \{2, 3, 4, 5, 6, \dots\}, \\
R(3) &= \{1, 2, 3, 4, 5, 6, \dots\}.
\end{aligned}
$$

Таким образом, $d(i) = 1$ для $i \in \{1, 2, 3\}$.



Классы

Будем говорить, что состояния i и j являются сообщающимися (обозначение: $i \leftrightarrow j$), если $i \to j$ и $j \to i$. Множество состояний $B \subseteq S$ называется неразложимым, если $i \leftrightarrow j$ для всех $i, j \in D$. Ясно, что $\leftrightarrow$ является отношением эквивалентности. По этому отношению множество состояний распадается на классы:

$$S = S_1 \cup S_2 \cup S_3 \cup \ldots .$$

Алгоритм идентификации классов состоит в следующем:

- Начать с произвольного состояния (назовем его 1).
- Найти все сообщающиеся с ним состояния (1 всегда среди них).
- Полученное множество является первым классом. Если других состояний нет, имеется всего один класс $S_1 = S$. Если существует элемент $S \setminus S_1$, то повторить описанную процедуру, начиная с этого элемента.

Подмножество $B \subseteq S$ называется замкнутым, если $i \not\to j$ для всех $i \in B$ и всех $j \in S \setminus B$. Другими словами, B замкнуто, если из него невозможно выйти. Состояние $i \in S$, для которого множество $\{i\}$ замкнуто, называется поглощающим. Нетрудно проверить, что B является замкнутым, если и только если $p_{ij} = 0$ для всех $i \in B$, $j \in B^c = S \setminus B$.
Из того, что B замкнуто в общем случае не следует, что $S \setminus B$ замкнуто. Кроме того, класс не обязан быть замкнутым, а замкнутое подмножество не обязано быть классом. В примере с теннисным геймом:

- множество $B = \{P_1 \text{ выиграл}\}$: замкнуто и является классом, но $S \setminus B$ незамкнуто,
- множество $B = S \setminus \{(0,0)\}$ замкнуто, но не является классом,
- множество $B = \{(0,0)\}$ является классом, но незамкнуто.

Лемма 47

Любое замкнутое множество B является классом или объединением классов.

Доказательство.

Обозначим через $\widehat{B}$ объединение всех классов C для которых $C \cap B \neq \emptyset$. Другими словами, $\widehat{B}$ состоит из состояний, которые сообщаются хотя бы с одним элементом B. Ясно, что $B \subset \widehat{B}$. Пусть существует $j \in \widehat{B} \setminus B$. По предположению j сообщается с некоторым $i \in B$. В частности, $i \to j$. В силу замкнутости B отсюда следует, что $j \in B$. Это противоречит предположению, что $j \in \widehat{B} \setminus B$.

Обратное утверждение, конечно, неверно. В "теннисном примере" множество $B = \{(0,0),(0,15)\}$ является объединением двух классов, но не является замкнутым.

Возвратность и невозвратность

Моментом первого попадания в состояние j называется

$$T_j(1) = \min\{n \in \mathbb{N} : X_n = j\}.$$

Второй, третий и последующие моменты попадания определяются следующим образом:

$$T_j(2) = \min\{n > T_j(1) : X_n = j\},$$
$$T_j(3) = \min\{n > T_j(2) : X_n = j\}, \ldots$$

Здесь подразумевается, что если $T_j(n) = \infty$, то $T_j(m) = \infty$ для всех $m > n$.

Состояние $i \in S$ называется

- возвратным, если $\mathsf{P}_i[T_i(1) < \infty] = 1$,
- положительно возвратным, если $\mathsf{E}_i[T_i(1)] < \infty$
- возвратным нулевым, если оно возвратно, но не является положительно возвратным,
- невозвратным, если оно не является возвратным.

Состояние является возвратным, если цепь возвращается в него с вероятностью 1. Оно положительно возвратно, если ожидаемое время между двумя последовательными его посещениями конечно. Нулевая возвратность состояния означает, что цепь вернется в него, но время ожидания может быть очень долгим. Состояние является невозвратным, если существует положительная вероятность (пусть и малая) того, что цепь в него не вернется. Пусть

$$f_i = \mathsf{P}_i[T_i(1) < \infty].$$

Ясно, что i возвратно, если и только если $f_i = 1$.

Интересно то, что каждый раз, когда цепь посещает состояние i, ее будущая эволюция не зависит от прошлого (за исключением названия текущего состояния), и она ведет себя точно так же, как новая и независимая цепь, начинающаяся с i. Это частный случай так называемого сильного марковского свойства, которое утверждает, что (обычное) марковское свойство выполняется по отношению к моментам остановки (а не только по отношению к фиксированным моментам времени). Данное свойство использовано в доказательстве следующей теоремы.

Теорема 48 (теорема о возвратности)

Пусть X — марковская цепь со счетным пространством состояний S и детерминированным начальным состоянием $X_0 = i$. Тогда верно в точности одно из следующих двух утверждений верно с вероятностью 1:

- *цепь возвращается в состояние i бесконечно много раз,*
- *цепь возвращается в состояние i конечное число раз N_i, где N_i имеет геометрическое распределение с параметром*

$$f_i = \mathsf{P}_i[T_i(1) < \infty].$$

В первом случае i возвратно, во втором — невозвратно.

Доказательство.

Если $f_i = 1$, то X гарантированно вернется в i хотя бы один раз. Когда это происходит, сильное марковское свойство "удаляет" прошлое, и процесс "обновляется". Это возвращает нас к исходной ситуации, когда мы имеем цепь, начинается в i и гарантированно возвращается туда хотя бы один раз. Продолжая в том же духе, мы получим бесконечную последовательность моментов остановки

$$T_i(1) < T_i(2) < \dots,$$

когда X оказывается в точке i. Все элементы этой последовательности конечны с вероятностью 1.

Если $f_i < 1$, то каждый раз когда X возвращается в i существует положительная вероятность $1 - f_i$ того, что процесс не вернется в i независимо от предшествующего поведения. Если считать возврат в i успехом, то число успехов до первой неудачи, т.е. общее число возвратов в i есть ни что иное как геометрически распределенная случайная величина с параметром f_i.

Теорема 49

Пусть S конечно. Тогда существует хотя бы одно возвратное состояние.

Доказательство.

Пусть все состояния невозвратны. Покажем, что в этом случае общее число посещений N_i каждого состояния i конечно для любого начального состояния i_0. Действительно, если $i = i_0$, то это вытекает из теоремы о возвратности. Если $i \neq i_0$, то число посещений i либо равно 0, если цепь не попадает в i, либо $1 + N_i$, если попадает. В любом случае, оно конечно. Поскольку S конечно, то сумма

$$\sum_{i \in S} N_i$$

конечна, что невозможно, так как хотя бы одно состояние должно посещаться бесконечное число раз.

Подобное утверждение неверно для счетного пространства состояний. Например, для детерминированного сдвига вправо на $\mathbb{Z}$.

Критерий возвратности

Важным следствием теоремы о возвратности является следующий критерий возвратности. Он работает как для конечного, так и для счетного пространства состояний

Теорема 50 (критерий возвратности)

Состояние $i \in S$ является возвратным, если и только если

$$\sum_{n \in \mathbb{N}} p_{ii}^{(n)} = \infty.$$

Доказательство.

Пусть N_i — общее число (конечное или бесконечное) посещений состояния i без учета его посещения в начальный момент времени:

$$N_i = \sum_{n=1}^{\infty} I_{\{X_n = i\}}.$$

Вычисляя математическое ожидание, находим

$$\mathsf{E}[N_i] = \mathsf{E}_i \left[\sum_{n=1}^{\infty} I_{\{X_n = i\}} \right] = \sum_{n=1}^{\infty} \mathsf{P}_i[X_n = i] = \sum_{n=1}^{\infty} p_{ii}^{(n)}.$$

Если i невозвратно, т.е. если $f_i < 1$, то теорема о возвратности и формула для математического ожидания геометрического распределения показывают, что

$$\mathsf{E}_i[N_i] = \frac{f_i}{1 - f_i} < \infty, \text{ и значит } \sum_{n=1}^{\infty} p_{ii}^{(n)} = \mathsf{E}_i[N_i] < \infty.$$

Если же i возвратно, то по теореме о возвратности $N_i = \infty$. Следовательно,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \mathsf{E}_i[N_i] = \infty,$$

что и требовалось доказать.

В данном доказательстве удалось проверить конечность N_i только исходя из ее математического ожидания. Это оказалось возможным благодаря тому, что каждый раз, когда цепь покидает состояние i, она возвращается в него с одинаковой вероятностью, независимо от прошлого. Это дает дополнительную информацию о случайной величине N_i: будет ли она бесконечной с вероятностью 1 или иметь геометрическое распределение.

Свойства класса

Некоторое свойство называется свойством класса, если из того, что оно выполнено для некоторого состояния данного класса вытекает, что оно выполнено для всех его состояний.

Теорема 51

Невозвратность и возвратность являются свойствами класса.

Доказательство.

Воспользуемся критерием возвратности. Пусть i является возвратным и j лежит в том же классе: $i \leftrightarrow j$. Тогда существуют натуральные числа m и k такие, что $p_{ij}^{(m)} > 0$ и $p_{ji}^{(n)} > 0$. Из уравнений Колмогорова-Чепмена вытекает, что

$$p_{jj}^{(n+m+k)} = \sum_{l_1 \in S} \sum_{l_2 \in S} p_{jl_1}^{(k)} p_{l_1 l_2}^{(n)} p_{l_2 m}^{(m)} \geq p_{ji}^{(k)} p_{ii}^{(n)} p_{ij}^{(m)}, \quad n \in \mathbb{N}.$$

Другими словами, существует положительная константа $c = p_{ji}^{(k)} p_{ij}^{(m)}$, не зависящая от n, такая, что

$$p_{jj}^{(n+m+k)} \geq c p_{ii}^{(n)}.$$

Из возвратности i вытекает, что $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, и значит

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \geq \sum_{n=m+k+1}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{jj}^{(n+m+k)} \geq c \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty,$$

откуда следует, что j является возвратным. Таким образом, возвратность является свойством класса. Невозвратность — это противоположное свойство, и значит оно также является свойством класса.

Теорема 52

Период является свойством класса, т.е. все элементы класса имеют один и тот же период.

Доказательство.

Пусть $d = d(i)$ — период состояния i, и $j \leftrightarrow i$. Тогда существуют натуральные числа m и k такие, что $p_{ij}^{(m)} > 0$ и $p_{ji}^{(k)} > 0$. В силу уравнений Колмогорова-Чепмена,

$$p_{ii}^{(m+k)} \geq p_{ij}^{(m)} p_{ji}^{(k)} > 0,$$

и значит $m + k \in R(i)$. Аналогично, для любого $n \in R(j)$,

$$p_{ii}^{(m+k+n)} \geq p_{ij}^{(m)} p_{jj}^{(n)} p_{ji}^{(k)} > 0.$$

Следовательно, $m + k + n \in R(i)$. Поскольку $d(i)$ делит $m + k$ и $m + k + n$, то он делит n. Это верно для всех $n \in R(j)$. Значит $d(i)$ является общим делителем всех элементов $R(j)$. Отсюда в свою очередь следует, что $d(i) \leq d(j)$. Меняя ролями i и j, заключаем что $d(j) \leq d(i)$. Следовательно, $d(i) = d(j)$.

Каноническое разложение

Поскольку число состояний не более чем счетно, то не более чем счетно и число классов. Пусть $C_1, C_2, \ldots$ — возвратные классы, $T_1, T_2, \ldots$ — невозвратные классы, и $T = T_1 \cup T_2 \cup \ldots$. Разложение

$$S = T \cup C_1 \cup C_2 \cup C_3 \cup \ldots,$$

называется каноническим разложением (пространства состояний) марковской цепи X.

Следующий результат показывает, что возвратные классы могут рассматриваться как самостоятельные марковские цепи.

Теорема 53

Возвратные классы замкнуты.

Доказательство.

Если возвратный класс C не является замкнутым, то существуют состояния $i \in C$ и $j \in C^c$ такие, что $i \to j$. При этом $j \not\to i$, так как C является классом. Стартуя в i, цепь достигнет j с положительной вероятностью и не вернется. Следовательно, число посещений i будет конечным с положительной вероятностью, что противоречит возвратности i по теореме о возвратности.

Для конечной цепи данный результат можно усилить.

Теорема 54

Для марковской цепи с конечным числом состояний возвратность класса равносильна его замкнутости.

Доказательство.

Мы знаем, что возвратные классы замкнуты. Чтобы доказать обратное, достаточно проверить, что невозвратные классы незамкнуты. Пусть T — замкнутый невозвратный класс. Поскольку T замкнут, мы можем рассматривать его как самостоятельную марковскую цепь. Но тогда она имеет возвратное состояние, что противоречит сделанному предположению.

Условие конечности пространства состояний существенно. Например, несимметричное случайное блуждание на $\mathbb{Z}$ представляет собой марковскую цепь, у которой только один (замкнутый) класс $\mathbb{Z}$ и все состояния невозвратны.

Если пронумеровать по порядку сначала состояния класса C_1, затем C_2 и т.д, и, наконец, состояния T, то переходная матрица примет вид

$$P = \begin{bmatrix} P_1 & 0 & 0 & \dots & 0 \\ 0 & P_2 & 0 & \dots & 0 \\ 0 & 0 & P_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_1 & Q_2 & Q_3 & \dots & \dots \end{bmatrix},$$

где P_1 — переходная матрица первого класса: $P_1 = (p_{ij}, i \in C_1, j \in C_1)$, и т.д. Q_k содержит вероятности перехода из невозвратных состояний в состояния класса C_k.

Стационарность и стационарные распределения

Переходы между различными состояниями цепи Маркова описывают кратковременное поведение цепи. Важно, однако, понимание поведения цепи Маркова на очень больших временных интервалах.

Случайный процесс $(X)_{n\in\mathbb{Z}_+}$ называется стационарным, если случайные векторы

$$(X_0, X_1, X_2, \ldots, X_k) \quad \text{и} \quad (X_m, X_{m+1}, X_{m+2}, \ldots, X_{m+k})$$

имеют одинаковые совместные распределения для $m, k \in \mathbb{Z}_+$.

Пример 55

Простое случайное блуждание не является стационарным. Действительно, X_0 является константой, а X_1 принимает два значений, следовательно их распределения различны. Ясно, что распределение X_n с течением времени все более расплывается.

Пример 56

В качестве примера стационарного процесса рассмотрим марковскую цепт с переключением режимов с $p_{01} = p_{10} = 1$, и начальным распределением $\mathsf{P}[X_0 = 0] = \mathsf{P}[X_0 = 1] = \frac{1}{2}$. Тогда $X_n = X_0$ если n четно, и $X_n = 1 - X_0$ если n нечетно. Кроме того, X_0 and $1 - X_0$, а $X_0, X_1, \ldots$ имеют одинаковые распределения.

Проверим, что распределения

$$(X_0, X_1, \ldots, X_k) \quad \text{и} \quad (X_m, X_{m+1}, \ldots, X_{m+k})$$

одинаковы.

Для $i_0, i_1, \dots, i_k \in \{0,1\}$ согласно марковскому свойству,

$$\begin{aligned}\mathsf{P}[X_0 = i_0, X_1 = i_i, \dots, X_k = i_k] &= \mathsf{P}[X_0 = i_0] p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k} \\ &= \frac{1}{2} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}.\end{aligned}$$

Аналогичным образом,

$$\begin{aligned}\mathsf{P}[X_m = i_0, X_1 = i_i, \dots, X_{m+k} = i_k] &= \mathsf{P}[X_m = i_0] p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k} \\ &= \frac{1}{2} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}.\end{aligned}$$

Следовательно, распределения совпадают.
В данном примере из того, что распределение X_0 и X_m одинаковы, было выведено, что процесс X является стационарным.

Распределение $\pi = (\pi_i)_{i \in S}$ на пространстве состояний S марковской цепи с переходной матрицей P называется стационарным, если

$$\mathsf{P}[X_1 = i] = \pi_i \text{ для всех } i \in S, \text{ если } \mathsf{P}[X_0 = i] = \pi_i, \text{ для всех } i \in S.$$

Лемма 57

Неотрицательный вектор $\pi = (\pi_i, i \in S)$ с $\sum_{i \in S} \pi_i = 1$ является стационарным распределением, если и только если

$$\pi P = \pi,$$

где π — вектор-строка. В этом случае марковская цепь с начальным распределением π и переходной матрицей P является стационарной и распределение X_m совпадает с π для всех m.

Доказательство.

Пусть $a^{(i)}$ — распределение X_i. В силу марковского свойства,

$$a^{(1)} = a^{(0)} P.$$

Условие стационарности π, очевидно, равносильно условию $\pi = \pi P$. Покажем, что марковская цепь с начальным распределением π является стационарной. Ясно, что все X_i имеют одинаковые распределения:

$$a^{(m)} = a^{(0)} P^m = \pi P^m = (\pi P) P^{m-1} = \pi P^{m-1} = \cdots = \pi.$$

Стационарность X вытекает из того, что выражение

$$\begin{aligned} \mathsf{P}[X_m = i_0, X_{m+1} = i_1, \ldots, X_{m+k} = i_k] &= \mathsf{P}[X_m = i_0] p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k} \\ &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k} \end{aligned}$$

не зависит от m.

Теорема 58 (Кац)

Пусть X *— неразложимая марковская цепь с конечным числом состояний. Тогда существует единственное стационарное распределение* $\pi = (\pi_j)_{j \in S}$, *которое определяется явной формулой*

$$\pi_j = \frac{\nu_{ij}}{m_i}, \ j \in S,$$

где i *— произвольное фиксированное состояние и*

$$\nu_{ij} = \mathsf{E}_i \left[\sum_{n=0}^{T_i(1)-1} I_{\{X_n = j\}} \right], \quad m_i = \mathsf{E}_i[T_i(1)] < \infty$$

— ожидаемое число посещений состояния j *между двумя последовательными посещениями состояния* i, *и ожидаемое время возврата в состояние* i, *соответственно.*

Заметим, что

- $\pi_i > 0$ для всех $i \in S$.
- Отношение $\frac{\nu_{ij}}{m_i}$ не зависит от состояния i.
- Поскольку $\nu_{ii} = 1$, то полагая $j = i$, заключаем, что

$$\pi_i = \frac{1}{m_i} \quad \text{и} \quad \nu_{ij} = \frac{\pi_j}{\pi_i}.$$

На практике сначала находят стационарное распределения, решая уравнение $\pi = \pi P$, а затем используют его для определения m_i или ν_{ij}.

Пример 59

Предположим, что статистика движения по данной дороге такова: в среднем за тремя из каждых четырех грузовиков следует легковой автомобиль, но только за одним из каждых пяти легковых автомобилей следует грузовик. Мимо вас проезжает грузовик. Сколько легковых автомобилей вы ожидаете увидеть, прежде чем проедет еще один грузовик?

Тип проезжающего мимо автомобиля (автомобиль или грузовик) можно моделировать с помощью цепи Маркова с двумя состояниями и матрицей переходов:

$$P = \begin{pmatrix} 4/5 & 1/5 \\ 3/4 & 1/4 \end{pmatrix}$$

где первая строка (столбец) соответствует состоянию “легковой автомобиль”, а вторая строка (столбец) — состоянию “грузовик”.

Нас интересует число ν_{21} посещений состояния “легковой автомобиль” ($j = 1$) между двумя посещениями состояния “грузовик” ($i = 2$). Найдем стационарное распределение:

$$\pi_1 = \frac{4}{5}\pi_1 + \frac{3}{4}\pi_2$$
$$\pi_2 = \frac{1}{5}\pi_1 + \frac{1}{4}\pi_2.$$

С учетом условия $\pi_1 + \pi_2 = 1$, находим

$$\pi_1 = \frac{15}{19}, \pi_2 = \frac{4}{19}.$$

Поскольку $v_{ij} = m_i \pi_j$, то

$$v_{21} = m_2 \pi_1 = \frac{15}{19} m_2.$$

Кроме того, $m_2 = 1/\pi_2 = 19/4$. Следовательно мы ожидаем увидеть $15/4 = 3.75$ легковых автомобилей между двумя грузовиками.

Долгосрочные средние значения

Одним из наиболее важных свойств стационарных распределений является то, что они описывают долгосрочное поведение цепи Маркова. Для анализа асимптотического поведения средних значений мы не можем воспользоваться классическим законом больших чисел для независимых одинаково распределенных случайных величин, так как случайные величины $X_0, X_1, \ldots$ таковыми не являются. Кроме того, значения X_k могут не быть числами, поэтому выражения $X_0 + X_1$ или $\mathsf{E}[X_0]$ могут не иметь смысла. Поэтому рассматривают некоторую функцию "дохода" $f : S \to \mathbb{R}$ и суммы вида $f(X_0) + f(X_1) + \cdots + f(X_{n-1})$. Свойство независимости в некоторой степени можно заменить марковским свойством и неразложимостью.

Теорема 60 (эргодическая теорема для марковских цепей)

Пусть X — конечная неразложимая марковская цепь. Для любой функции $f : S \to \mathbb{R}$ имеем

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \mathsf{E}_\pi[f(X_0)] := \sum_{j \in S} f(j)\pi_j,$$

где π — единственное стационарное распределение X.

Если в качестве f взять индикатор некоторого состояния, то получим следующий результат. Пусть N_n^i — число посещений состояния i на первых n шагах. Тогда

$$\pi_i = \lim_n \frac{1}{n} N_n^i,$$

где π — единственное стационарное распределение. Другими словами, для конечной неразложимой цепи Маркова компоненту π_i стационарного распределения π можно интерпретировать как долю времени, проводимого цепью в состоянии i.

Предельные распределения

Распределение $\pi = (\pi_i, i \in S)$ на пространстве состояний S марковской цепи с переходной матрицей P называется предельным, если

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j,$$

для всех $i, j \in S$.
Ясно, что может существовать только одно предельное распределение.
Если S конечно, то π является вероятностным распределением:

$$\sum_{j\in S} \pi_j = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n)} = 1.$$

Связь со стационарным распределением указана в следующем утверждении:

Теорема 61

Любое предельное распределение π является стационарным.

Доказательство.

Чтобы установить равенство $\pi = \pi P$, т.е.

$$\pi_j = \sum_{i \in S} \pi_i p_{ij},$$

воспользуемся уравнениями Колмогорова-Чепмена

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}.$$

Поскольку $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$, то

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k \in S} p_{ik}^{(n)} p_{kj} = \sum_{k \in S} \lim_{n\to\infty} p_{ik}^{(n)} p_{kj} = \sum_{k \in S} \pi_k p_{kj}.$$

Предельные распределения необязательно существуют даже при наличии стационарных распределений.

Пример 62

Пусть X осуществляет детерминированные переходы между состояниями 0 и 1. Ее переходная матрица имеет вид

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Легко видеть, что

$$P^{2n-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{и} \quad P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad n \in \mathbb{N}.$$

Это означает, что значения $p_{ij}^{(n)}$ не могут сходиться при $n \to \infty$. С другой стороны, существует стационарное распределение $\pi = (1/2, 1/2)$.

Пример 63

В предыдущем примере пределы последовательностей $p_{ij}^{(n)}$ не существовали. Возможна также ситуация когда они существуют, но зависят от начальных условий. Простейшим примером является марковская цепь с двумя состояниями $i = 1, 2$, которая не движется, т.е. $p_{11} = p_{22} = 1$. It follows that for each n, we have $p_{ij}^{(n)} = 1$ if $i = j$ and $p_{ij}^{(n)} = 0$ if $i \neq j$. Следовательно, пределы $\lim_{n\to\infty} p_{ij}^{(n)}$ существуют для каждой пары i, j, но их значения зависят от начального состояния i:

$$\lim_{n\to\infty} p_{12}^{(n)} = 0 \quad \text{и} \quad \lim_{n\to\infty} p_{22}^{(n)} = 1.$$

В первом примере цепь была периодической, во втором — разложимой.

Теорема 64 (о сходимости)

Пусть X — неразложимая, апериодическая конечная марковская цепь. Тогда она обладает предельным распределением.