1. Дискретное вероятностное пространство.

Предположим, что мы провели некоторый эксперимент (например бросание монеты или бросание игрального кубика), в результате которого мы могли получить значения из множества $\Omega = \{\omega_1, \dots, \omega_N\}$. Предположим, что мы можем повторить этот эксперимент неограниченное число раз, причем так, что эти эксперименты ни как не связаны друг с другом. В этом случае естественно считать вероятностью исхода ω_j число $p_{\omega_j} = \lim\limits_{n\to\infty} \frac{\#(\omega_j)}{n}$, где $\#(\omega_j)$ — количество экспериментов с результатом ω_j среди первых n проведенных.

Из такого подхода к вероятноти, как к частоте определенного результата, видны следующие важные свойсива

(i) $p_{\omega_j} \in [0,1]$

(ii) $\sum\limits_j p_{\omega_j} = 1$.

Абстрагируясь от нашего изначального "экспереминтального" подхода перейдем теперь к формальному определению вероятности.

Определение 1. Пусть задано некоторое множество возможных исходов (эксперимента) $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. Это множество Ω называют **множеством элементарных исходов**. Всякое подмножество $A \subset \Omega$ называют **событием**. Функцию $P\colon 2^\Omega \to [0,1]$, удовлетворяющую следующим свойствам:

(i) $P(\Omega) = 1$,

(ii) $A \cap B = \emptyset \Rightarrow P(A \bigcup B) = P(A) + P(B)$ (правило суммы или аддитивность),

называют **вероятностной мерой**, а значение $P(A)$ **вероятностью** события A.

Отметим, что свойства (i), (ii) из определения выше равносильны свойствам (i), (ii) из подхода к вероятности, как к частоте результата эксперимента. Действительно, вероятностная мера P полностью определяется значениями $P(\{\omega_j\}) = p_{\omega_j}$. Из определения вероятностной меры следует, что

$$p_{\omega_j} \geq 0 \quad \text{и} \quad \sum_j p_{\omega_j} = 1.$$

Вероятность произвольного события A вычисляется по формуле

$$P(A) = \sum_{j\,:\, \omega_j \in A} p_{\omega_j}.$$

Пример 2. Пусть мы бросаем правильную монетку. Тогда $\Omega = \{O, P\}$, где O — выпал орел, P — выпала решка. Естественно считать, что $P(\{O\}) = P(\{P\}) = 1/2$.

Вопрос: какое множество элементарных исходов и вероятность для броска двух правильных монеток?

Если все элементарные исходы равновозможны, т.е. $p_{\omega_1} = \ldots = p_{\omega_n} = 1/n$, то $P(A) = \frac{\#A}{\#\Omega}$, где $\#A$ — количество элементов (исходов) в множестве A. Так определенная вероятность называется **классической** моделью. Конечно предположение о равновозможности не всегда корректно. Если эксперимент состоит в бросании правильного кубика, у которого три грани зеленые, две красные и одна синяя, то выпадение зеленого, красного или синего цвета нельзя считать равновозможными исходами. В данном случае выпадение любой из шести граней равновозможно, а зеленый, красный и синий цвета выпадают с вероятностями 1/2, 1/3 и 1/6 соответственно.

Дни рождения. Какова вероятность того, что в группе из 30 человек у каких-то студентов дни рождения совпадают?

Посчитаем обратную вероятность, что у всех студентов дни рождения в разные дни. Такая веротяноть равна $\frac{365\cdot 364\cdot\ldots\cdot(365-30+1)}{365^{30}}$, а искомая веротяность равна

$$1-\Big(1-\frac{1}{365}\Big)\cdot\ldots\cdot\Big(1-\frac{29}{365}\Big)\approx 1-\exp(-\frac{1+2+\ldots+29}{365})=1-e^{-1.2}\approx 0.7.$$

2. Формула включений и исключений.

Из свойства (ii) вероятностной меры по индукции выводится следующее утверждение.

Предложение 3. *Для произвольных событий $A_1,\ldots A_n$ верно неравенство*

$$P(A_1\cup\ldots\cup A_n)\le\sum_{k=1}^{n}P(A_k).$$

Предложение 4 (формула включений и исключений)**.** *Для произвольных событий $A_1,\ldots A_n$ верно равенство*

$$P(A_1\cup\ldots\cup A_n)=\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<\ldots<i_k}P(A_{i_1}\cap\ldots\cap A_{i_k}).$$

Доказательство. Докажим утверждение по индукции. База:

$$\begin{aligned}P(A_1\cup A_2)&=P((A_1\setminus A_2)\cup(A_1\cap A_2)\cup(A_2\setminus A_1))\\&=P(A_1\setminus A_2)+P(A_1\cap A_2)+P(A_2\setminus A_1)=P(A_1)+P(A_2)-P(A_1\cap A_2).\end{aligned}$$

Предположим, что утверждение выполнено для n произвольных множеств. Проверим, что оно выполнено для $n+1$-го множества.

$$P(A_1\cup\ldots\cup A_n\cup A_{n+1})=P(A_1\cup\ldots\cup A_n)+P(A_{n+1})-P((A_1\cap A_{n+1})\cup\ldots\cup(A_n\cap A_{n+1}))$$

$$=\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<\ldots<i_k}P(A_{i_1}\cap\ldots\cap A_{i_k})+P(A_{n+1})-\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<\ldots<i_k}P(A_{i_1}\cap\ldots\cap A_{i_k}\cap A_{n+1})$$

$$=\sum_{k=1}^{n+1}(-1)^{k-1}\sum_{i_1<\ldots<i_k}P(A_{i_1}\cap\ldots\cap A_{i_k}).$$

□

Парадокс распределения подарков

n человек принесли подарки друг для друга. Затем эти подарки сложили в мешок и наугад каждый вынул из мешка себе подарок. Какова вероятность того, что конкретный человек вынул подарок, который он принес? Какова вероятность того, что никто не вытащил подарок, который сам принес?

Пространство исходов Ω состоит из всех возможных перестановок чисел $1,2,\ldots,n$, причем все перестановки являются равновозможными. Значит вероятность конкретной перестановки равна $1/n!$. Событие, состоящее в том, что конкретный человек вытащил подарок, который сам принес, состоит из $(n-1)!$ исходов. Следовательно, вероятность такого события равна $1/n$. При больших n эта вероятность стремится к нулю.

Можно было бы думать, что вероятность события: ни один человек не вытащил подарок, который сам принес, стремится к единице, но это ошибочное мнение.

Пусть A_k – событие состоящее в том, что k-й человек вытащил свой подарок. Тогда $A_1\cup\ldots\cup A_n$ — событие, состоящее в том, что хотя бы один вытащил свой подарок. По формуле включения и исключения

$$P(A_1\cup\ldots\cup A_n)=\sum_{k=1}^{n}(-1)^{k-1}C_n^k\frac{(n-k)!}{n!}=\sum_{k=1}^{n}\frac{(-1)^{k-1}}{k!}.$$

Таким образом, вероятность того, что ни один человек не вытащил подарок, который сам принес, равна

$$1 - P(A_1 \cup \ldots \cup A_N) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots$$

и стремится к $1/e$.

3. Бесконечное множество элементарных исходов.

Конечного множества исходов Ω не всегда хватает для описания результатов эксперимента. Например, будем подбрасывать монету до тех пор, пока не выпадет орел, после чего будем прекращать эксперимент. В данной ситуации элементарным исходом будет последовательность $(pp\ldots ppo)$. Таких последовательностей уже счетное число. С другой стороны, в случае бесконечного множества исходов, ситуация с определением вероятностной меры значительно сложнее. Так на $\Omega = \mathbb{N}$ существует удовлетворяющая свойствам (i) и (ii) вероятности функция P, которая равна нулю на всяком конечном множестве и единице на всем $\mathbb{N}$. Для такой функции

$$\sum_k P(\{k\}) = 0 \neq 1 = P(\mathbb{N}).$$

Т.к. нам хотелось бы, чтобы определение вероятности в случае счетного Ω было бы согласовано с уже данным выше определением в случае конечного множества исходов, то попробуем задать вероятностную меру набором $\{p_{\omega_j}, \omega_j \in \Omega\}$ вероятностей каждого элементарного исхода ($p_{\omega_j} \geq 0, \quad \sum\limits_j p_{\omega_j} = 1$). В таком случае вероятность каждого события A считается по уже знакомой формуле

$$P(A) = \sum_{j\colon \omega_j \in A} p_{\omega_j},$$

причем, как известно из курса анализа, эта сумма корректно определена, в том смысле, что результат не зависит от порядка суммирования. Заметим, что в данном случае вместо свойства (ii) вероятностной меры выполнено более сильное свойство

(ii)' $P(\cup_n A_n) = \sum\limits_n P(A_n)$

для произвольного не более чем счетного набора попарно непересекающихся событий A_n. Чтобы это проверить докажем следующую лемму.

Лемма 5. *Пусть $a_{n,m}$ — двупараметрическая последовательность неотрицательных чисел. Тогда*

$$\sum_{j=1}^{\infty} a_{n(j),m(j)} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{n,m},$$

где в левой части равенства стоит сумма ряда в произвольном порядке (т.е. $j \mapsto (n(j), m(j))$ — произвольная биекция).

Доказательство. Действительно, для произвольных M, N выполнено

$$\sum_{n=1}^{N}\sum_{m=1}^{M} a_{n,m} \leq \sum_{j=1}^{\infty} a_{n(j),m(j)}$$

и для произвольного J выполнено

$$\sum_{j=1}^{J} a_{n(j),m(j)} \leq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{n,m}.$$

□

Выполнение свойств (i) и $(ii)'$ для вероятностной меры, в случае не более чем счетного множества исходов, равносильно заданию этой меры с помощью набора элементарных вероятностей. Правило суммы (ii) называют аддитивностью меры P, а свойство (ii)' называют σ-аддитивностью или счетной аддитивностью.

4. Условная вероятность. Формула полной вероятности. Формула Байеса.

Рассмотрим следующий пример. Пусть эксперимент состоит в трехкратном подбрасывании симметричной монеты и пусть событие A заключается в том, что решка выпала ровно один раз. Вероятность данного события равна $3/8$, т.е. подходит всего три исхода из 8. Предположим теперь об исходе эксперимента дополнительно известно, что при втором бросании выпал орел. Обозначим это событие буквой B. Какова вероятность события A при этой дополнительной информации? Событие B состоит из 4-х исходов, причем ровно два из этих исходов являются благоприятными для события A, т.е. в событие A попадают два из 4-х возможных (принадлежащих B) исходов. Таким образом, естественно принять новую вероятность A (при условии выполнения события B) равной $2/4 = 1/2$.

Этот простой пример приводит нас к следующему общему определению.

Определение 6. Пусть $P(B) > 0$. **Условной вероятностью** события A при условии B называется число

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

Если фиксировать событие B, то функция $P(\,\cdot\,|B)$ является новой вероятностной мерой, т.е. удовлетворяет свойствам (i) и (ii) (или (ii)', если таковой была исходная вероятностная мера P). Равенство из определения часто переписывают в виде

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

и называют **правилом произведения**.

Теорема 7. (Формула полной вероятности) *Пусть* $\Omega = A_1 \cup A_2 \ldots \cup A_n$ *и* $A_i \cap A_j = \emptyset$ *для всех* $i \neq j$. *Предположим, что* $P(A_i) > 0$. *Тогда для каждого события* B *имеет место равенство*

$$P(B) = \sum_i P(B|A_i)P(A_i).$$

Доказательство. Имеем $P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i)$. □

На практике часто возникает ситуация, когда заданы или легко вычисляются именно условные вероятности. Предположим, что задано разбиение Ω на попарно непересекающиеся, непустые события A_i, их вероятности p_i и условные вероятности $P_i(\,\cdot\,) = P(\,\cdot\,|A_i)$ (произвольная вероятностная мера со свойством $P_i(B) = P_i(B \cap A_i)$). Тогда формула полной вероятности позволяет **определить** вероятностную меру P, для которой числа p_i и $P_i(B)$ действительно являются вероятностями A_i и условными вероятностями B при условии A_i. Более того, формула полной вероятности гарантирует, что такая вероятностная мера только одна.

Действительно, положим

$$P(B) = \sum_i P_i(B)p_i.$$

Свойства (i) и (ii) очевидно верны, кроме того $P(A_i) = p_i$ и $P(B \cap A_i) = P_i(B)p_i$.

Задача о сумасшедшей старушке

На посадку в самолет стоят $N \geq 2$ пассажиров, среди которых сумасшедшая старушка. Старушка расталкивает всех пассажиров и садится в самолет на произвольное место. Затем пассажиры, когда заходят в самолет, садятся на свое место, если оно свободно, и на произвольное свободное место в противном случае. Какова вероятность того, что последний пассажир сядет на свое место?

Пусть эта вероятность равна P_N. Если $N = 2$, то $P_N = 1/2$. Предположим, что уже для всех $k \leq N$ доказано, что $P_k = 1/2$. Докажем равенство $P_{N+1} = 1/2$. Событие B состоит из тех исходов, когда последний пассажир садится на свое место. Событие A_m состоит из тех исходов, когда старушка села на место m-го пассажира. По формуле полной вероятности

$$P_{N+1} = P(B) = \sum_m P(B|A_m)P(A_m).$$

Заметим, что $P(A_m) = 1/(N+1)$ и все кроме двух (когда старушка села на свое место или на место последнего пассажира) вероятности $P(B|A_m) = 1/2$. Следовательно, имеем

$$P_{N+1} = \frac{N-1}{2(N+1)} + \frac{1}{N+1} = \frac{1}{2}.$$

5. Формула Байеса

Следующее утверждение позволяет пересчитывать априорные вероятности на основе новых данных.

Теорема 8. (формула Байеса) *Пусть* $P(A) > 0$ *и* $P(B) > 0$. *Тогда имеет место равенство*

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

Доказательство. Достаточно заметить, что

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

□

Парадокс Байеса

Пусть имеется тест, используемый для диагностики некоторого редкого заболевания. Известно, что доля больных этим заболеванием равна $0,001$. Если человек болен, то тест дает положительный результат с вероятностью $0,99$. Если человек здоров, то тест дает положительный результат с вероятностью $0,01$. Известно, что тест оказался положительным. Какова вероятность того, что человек на самом деле здоров (т.е. тест ложноположительный)?

Пусть T_+ и T_- – события состоящие в том, что тест дал положительный результат и тест дал отрицательный результат. Пусть также Z_+ и Z_- – события состоящие в том, что человек болен (иммет заболевание) и человек здоров (не имеет) соответственно. По формуле полной вероятности

$$P(T_+) = P(T_+|Z_+)P(Z_+) + P(T_+|Z_-)P(Z_-) = 0,99 \cdot 0,001 + 0,01 \cdot 0,999 = 0,01098.$$

По формуле Байеса

$$P(Z_-|T_+) = \frac{P(T_+|Z_-)P(Z_-)}{P(T_+)} = \frac{0,01 \cdot 0,999}{0,01098} \geq 0,9.$$

Таким образом, при положительном тесте вероятность того, что человек на самом деле здоров, составляет более 90%. Это связано с тем, что точность теста снижается из-за редкости заболевания. В этом случае надо обязательно повторять тест.

Парадокс Монти Холла

Вы участвуете в игре, в которой вам нужно выбрать одну из трёх дверей. За одной из дверей спрятан автомобиль, за двумя другими — козы. Вы выбираете одну из дверей (например 1-ую) после этого ведущий, который знает, где находится автомобиль, открывает одну из оставшихся дверей (например 3-ю) за которой находится коза. После этого ведущий предлагает вам изменить свой выбор и выбрать дверь номер 2. Увеличатся ли ваши шансы выиграть автомобиль, если вы примете предложение ведущего и измените свой выбор?

Заметим, что в рамках описанной ситуации мы не можем дать однозначного ответа, т.к. мы не знаем, по какому правилу ведущий открывает дверь. Для определенности предположим, что ведущий равновероятно открывает одну из невыбранных игроком дверей, за которой находится коза.

Пусть событие A_i заключается в том, что автомобиль находится за i-й дверью, событие B — в том, что ведущий открыл 3-ю дверь. Если автомобиль изначально находился за 1-й дверью (которую мы выбрали), то ведущий откроет любую из оставшихся дверей случайным образом, т.е. $P(B|A_1) = 1/2$. В других же случаях условные вероятности равны: $P(B|A_2) = 1$ и $P(B|A_3) = 0$. Теперь применим формулу Байеса:

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} = \frac{1/3}{1/6 + 1/3} = \frac{2}{3}.$$

6. Независимые события

С точки зрения вычисления вероятностей независимость события A от события B означает, что вероятность A не меняется от того произошло событие B или нет. Формализовать эту идею позволяет условная вероятность. Событие A не зависит от события B (мы пока предполагаем, что $P(B) > 0$), если

$$P(A|B) = P(A),$$

что по определению условной вероятности можно переписать в виде

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

Это равенство и принимают в качестве определения независимости.

Определение 9. События A и B называются **независимыми**, если

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

Если события не являются независимыми, то говорят, что они зависимые.

Вопрос: Пусть в городе есть умные жители и есть богатые, причем ум и богатство независимы. Верно ли, что есть богатый и умный горожанин?

Пример 10. Если $A \cap B = \emptyset$ и $P(A) > 0$, $P(B) > 0$, то события A и B зависимые. Действительно, если одно из событий произошло, то другое событие уже произойти не может и значит его вероятность изменилась (стала равна нулю).

Пример 11. Если $P(A) = P(B) = 1$, то события независимые (например, $A = B = \Omega$). Действительно,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 = P(A) \cdot P(B).$$

Пример 12. Выбираем наугад натуральное число от 1 до 100. Событие A состоит из тех чисел, которые делятся на 2, а событие B состоит из тех чисел, которые делятся на 5. Эти события независимые. Действительно, $P(A) = 1/2$, $P(B) = 1/5$ и

$$P(A \cap B) = \frac{1}{10} = \frac{1}{2} \cdot \frac{1}{5} = P(A) \cdot P(B).$$

Пусть теперь выбирается натуральное число от 1 до 101. Тогда уже события A и B окажутся зависимыми:

$$P(A \cap B) = \frac{10}{101} \neq \frac{50}{101} \cdot \frac{20}{101} = P(A) \cdot P(B).$$

Отметим, что из независимости A и B следует независимость $\overline{A} = \Omega \setminus A$ и B:

$$P(\overline{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(\overline{A})P(B).$$

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

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

для произвольного $k \in \{2, \ldots, n\}$ и произвольных $1 \leq i_1 < i_2 < \ldots < i_k \leq n$.

Отметим, что независимость в совокупности не совпадает с попарной независимостью.

Это можно проиллюстрировать **парадоксом независимости**.

Два раза бросаем правильную монету. Событие A – при первом бросании выпал герб. Событие B – при втором бросании выпал герб. Событие C – ровно на одном бросании выпал герб. Эти события попарно независимы, но не являются независимыми в совокупности. Действительно, любые два однозначно определяют третье, в частности пересечение A и B исключает C, т. е. $P(A \cap B \cap C) = 0 \neq P(A) \cdot P(B) \cdot P(C)$.

7. Случайные величины на дискретном вероятностном пространстве

Определение 13. Случайной величиной на дискретном вероятностном пространстве $\Omega = \{\omega_1, \omega_2, \ldots\}$ называют произвольную функцию $X \colon \Omega \to \mathbb{R}$.

Если X — случайная величина на дискретном вероятностном пространстве, то множество ее значений не более чем счетно.

Определение 14. Пусть X — случайная величина на дискретном вероятностном пространстве и $x_1, \ldots, x_n, \ldots$ — все различные значения X. **Распределением** случайной величины X называется новая вероятностная мера μ_X на пространстве $\{x_1, \ldots, x_n, \ldots\}$, для которой $\mu_X(\{x_j\}) = P(\omega \colon X(\omega) = x_j)$.

В силу того, что события $A_j := \{\omega \colon X(\omega) = x_j\}$ попарно не пересекаются и $\cup_j A_j = \Omega$, μ действительно является вероятностной мерой. Пусть $p_j = P(\omega \colon X(\omega) = x_j)$, тогда распределение дискретной случайной можно задать таблицей:

x_1	x_2	$\ldots$	x_n	$\ldots$
p_1	p_2	$\ldots$	p_n	$\ldots$

Пример 15. *Пусть эксперимент состоит в бросании правильного кубика, у которого три грани зеленые, две красные и одна синяя. В случае выпадения зеленой грани будем говорить, что* $X = 0$, *в случае выпадения красной —* $X = 1$, *в случае синей —* $X = 2$. *Тогда* X *— случайная величина и ее распределение задано таблицей*

0	1	2
1/2	1/3	1/6

8. Случайные величины на дискретном вероятностном пространстве. Основные примеры распределений.

Приведем несколько примеров дискретных случайных величин.

(I) **Бернуллиевская случайная величина.**

Таблица распределения бернуллиевской случайной величины имеет вид

0	1
q	p

Эта случайная величина моделирует однократное бросание монеты с вероятностью орла p. Такая случайная величина обычно появляется, как индикатор какого-то события A

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

Тогда $p = P(A)$ и $q = 1 - P(A)$.

(II) **Схема Бернулли (биномиальное распределение)**

Ω – все возможные наборы последовательностей длины N из 0 и 1. Вероятностная мера P задана на каждом элементарном исходе следующим правилом: если исход содержит k единиц, то вероятность этого исхода $p^k q^{N-k}$, где $p, q \geq 0$ и $p + q = 1$.

Случайная величина $X(\omega)$ – число единиц в исходе ω. Таблица ее распределения имеет вид

0	1	$\ldots$	k	$\ldots$	N
q^N	Npq^{N-1}	$\ldots$	$C_N^k p^k q^{N-k}$	$\ldots$	p^N

(III) **Геометрическое распределение**

Распределение случайной величины X задается следующей таблицей:

1	2	$\ldots$	k	$\ldots$
p	qp	$\ldots$	$q^{k-1}p$	$\ldots$

Эта случайная величина моделирует подбрасывание монеты до первого успеха.

(IV) **Распределение Пуассона**

Распределение случайной величины X задается таблицей:

0	1	$\ldots$	k	$\ldots$
$e^{-\lambda}$	$\frac{\lambda}{1!}e^{-\lambda}$	$\ldots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	$\ldots$

9. Совместное распределение случайных величин.

Определение 16. Пусть X, Y — две случайные величины на дискретном вероятностном пространстве с множествами (различных) значений $\{x_1, \ldots, x_k, \ldots\}$ и $\{y_1, \ldots, y_k, \ldots\}$ соответственно. Их **совместным распределением** называется вероятностная мера $\mu_{(X,Y)}$ на вероятностном пространстве всех пар (x_j, y_k), для которой

$$\mu_{(X,Y)}(\{(x_j, y_k)\}) = P(\omega\colon X(\omega) = x_j,\ Y(\omega) = y_k) = P(\{\omega\colon X(\omega) = x_j\} \cap \{\omega\colon Y(\omega) = y_k\}).$$

Аналогично определяется совместное распределение трех и более случайных величин.

Определение 17. Случайные величины X, Y с множествами значений $\{x_1, \ldots, x_k, \ldots\}$ и $\{y_1, \ldots, y_k, \ldots\}$ соответственно называются **независимыми**, если

$$\mu_{(X,Y)}(\{(x_j, y_k)\}) = \mu_X(\{x_j\}) \cdot \mu_Y(\{y_k\}) \quad \forall k, j$$

или другими словами

$$P(\omega\colon X(\omega) = x_j,\ Y(\omega) = y_k) = P(\omega\colon X(\omega) = x_j) \cdot P(\omega\colon Y(\omega) = y_k) \quad \forall k, j.$$

Аналогично определяется независимость трех и более случайных величин.

Предложение 18. *Пусть X и Y две случайные величины. Тогда независимость X и Y равносильна тому, что*

$$P(\omega\colon X(\omega) \in A, Y(\omega) \in B) = P(\omega\colon X(\omega) \in A) \cdot P(\omega\colon Y(\omega) \in B)$$

для произвольных множесвт $A, B \subset \mathbb{R}$.

Доказательство. Пусть $\{x_1, x_2, \ldots\}$ и $\{y_1, y_2, \ldots\}$ множества значений случайных величин X и Y соответственно. Ясно, что достаточно рассмотреть множества $A \subset \{x_1, x_2, \ldots\}$ и $B \subset \{y_1, y_2, \ldots\}$. Тогда

$$P(\omega\colon X(\omega) \in A) \cdot P(\omega\colon Y(\omega) \in B) = \sum_{i,j: x_i \in A, y_j \in B} P(\omega\colon X(\omega) = x_i) \cdot P(\omega\colon Y(\omega) = y_j)$$

$$= \sum_{i,j: x_i \in A, y_j \in B} P(\omega\colon X(\omega) = x_i, Y(\omega) = y_j) = P(\omega\colon X(\omega) \in A, Y(\omega) \in B),$$

где была использована абсолютная сходимость рядов $\sum\limits_{i:x_i\in A} P(X=x_i)$ и $\sum\limits_{j:y_j\in B} P(Y=y_j)$.

$\square$

10. Математическое ожидание.

Пусть случайная величина X принимает значения $\{x_1,\dots,x_n,\dots\}$ (предполагаем, что значения перечислены без повторений).

Определение 19. Математическим ожиданием случайной величины X называют число

$$\mathbb{E}X=\sum_j x_jP(\omega\colon X(\omega)=x_j),$$

в предположении абсолютной сходимости ряда. Если ряд не сходится абсолютно, то говорят, что случайная величина не имеет конечного математического ожидания.

Пример 20. Пусть X — бернуллиевская случайная величина, т.е.

$$P(X=1)=p, P(X=0)=1-p.$$

Тогда $\mathbb{E}X=1\cdot p+0\cdot(1-p)=p$.

Пример 21. Пусть случайная величина S_n имеет биномиальное распределение, т.е.

$$P(S_n=k)=C_n^kp^k(1-p)^{n-k},\ k\in\{0,\dots,n\}.$$

Вычислим ожидание случайной величины S_n:

$$\mathbb{E}S_n=\sum_{k=1}^n kC_n^kp^k(1-p)^{n-k}=np\sum_{k=1}^n C_{n-1}^{k-1}p^{k-1}(1-p)^{n-1-(k-1)}=np.$$

Напомним, что слагаемые абсолютно сходящегося ряда можно группировать и переставлять, а произведение двух абсолютно сходящихся рядов равно ряду из всех возможных попарных произведений слагаемых этих рядов, взятых в произвольном порядке.

Лемма 22. *Пусть случайная величина X с конечным математическим ожиданием принимает значения y_k (не обязательно различные) на множествах B_k, причем события B_k попарно не пересекаются, тогда*

$$\mathbb{E}X=\sum_k y_kP(B_k).$$

Доказательство. Заметим, что

$$\sum_{k=1}^{k_0}|y_k|P(B_k)=\sum_j\sum_{\substack{k:y_k=x_j\\k\le k_0}}|y_k|P(B_k)=\sum_j|x_j|\sum_{\substack{k:y_k=x_j\\k\le k_0}}P(B_k)\le\sum_j|x_j|P(X=x_j).$$

Значит ряд $\sum_k y_kP(B_k)$ абсолютно сходится. Поэтому

$$\sum_k y_kP(B_k)=\sum_j\sum_{k:y_k=x_j}y_kP(B_k)=\sum_j x_jP(X=x_j)=\mathbb{E}X.$$

$\square$

Замечание 23. Заметим, что если в предыдущей лемме ряд $\sum_k y_kP(B_k)$ абсолютно сходится, то математическое ожидание определено. Действительно,

$$\sum_{j=1}^{j_0}|x_j|P(X=x_j)=\sum_{j=1}^{j_0}\sum_{k:y_k=x_j}|y_k|P(B_k)\le\sum_k|y_k|P(B_k).$$

Теорема 24. *Пусть X — случайная величина, принимающая значения $\{x_1, x_2, \dots\}$. Рассмотрим случайную величину $Y = \varphi(X)$, тогда*

$$\mathbb{E}Y = \mathbb{E}\varphi(X) = \sum_k \varphi(x_k) P(X = x_k),$$

при условии абсолютной сходимости последнего ряда.

Доказательство. Действительно, случайная величина Y принимает значения $\varphi(x_k)$ на множествах $\{\omega\colon X(\omega) = x_k\}$. □

Аналогичная формула справедлива и для функции от нескольких случайных величин $\varphi(X_1, \dots, X_k)$.

11. Свойства математического ожидания.

Если некоторое свойство имеет место с вероятностью единица, то говорят, что оно имеет место **почти наверное**.

Теорема 25. (i) $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y$ *(математическое ожидание линейно),*

(ii) *если $X \geq 0$ почти наверное, то $\mathbb{E}X \geq 0$ (математическое ожидание монотонно).*

(iii) *если $X \geq 0$ почти наверное и $\mathbb{E}X = 0$, то $X = 0$ почти наверное.*

Доказательство. Второе и третье утверждения следуют из определения. Докажем первое утверждение. Пусть $\{x_1, x_2, \dots\}$ и $\{y_1, y_2, \dots\}$ множества значений случайных величин X и Y соответственно. Пусть $A_i = \{X = x_i\}$ и $B_j = \{Y = y_j\}$. Случайная величина X принимает значения x_i на множествах $A_i \cap B_j$, случайная величина Y принимает значения y_j на множествах $A_i \cap B_j$. Поэтому

$$\alpha \mathbb{E}X + \beta \mathbb{E}Y = \alpha \sum_{i,j} x_i P(A_i \cap B_j) + \beta \sum_{i,j} y_j P(A_i \cap B_j) = \sum_{i,j} (\alpha x_i + \beta y_j) P(A_i \cap B_j).$$

Последний ряд абсолютно сходится, как сумма двух абсолютно сходящихся рядов. Случайная величина $\alpha X + \beta Y$ принимает значения $\alpha x_i + \beta y_j$ на множествах $A_i \cap B_j$. Таким образом, математическое ожидание $\alpha X + \beta Y$ существует и равно

$$\sum_{i,j} (\alpha x_i + \beta y_j) P(A_i \cap B_j),$$

что и требовалось доказать. □

Следствие 26. *Имеет место оценка*

$$|\mathbb{E}X| \leq \mathbb{E}|X|.$$

Доказательство. Достаточно заметить, что $-|X| \leq X \leq |X|$ и воспользоваться свойствами математического ожидания, полученными на предыдущей лекции. □

Следствие 27. (Неравенство Чебышева) *Если $X \geq 0$ п.н., то для всякого $t > 0$ верно неравенство*

$$P(\omega\colon X(\omega) \geq t) \leq \frac{\mathbb{E}X}{t}.$$

Доказательство. Пусть $A = \{\omega\colon X(\omega) \geq t\}$. Тогда $X \geq t \cdot I_A$ и значит

$$\mathbb{E}X \geq tP(A).$$

□

Теорема 28. *Если случайные величины X и Y независимы и существуют математические ожидания $\mathbb{E}X$ и $\mathbb{E}Y$, то*

$$\mathbb{E}[X \cdot Y] = [\mathbb{E}X] \cdot [\mathbb{E}Y].$$

Доказательство. Пусть $\{x_1, x_2, \ldots\}$ и $\{y_1, y_2, \ldots\}$ множества значений случайных величин X и Y соответственно и пусть $A_i = \{\omega\colon X(\omega) = x_i\}$, $B_j = \{\omega\colon Y(\omega) = y_j\}$. Тогда по теорема о произведении абсолютно сходящихся рядов

$$[\mathbb{E}X] \cdot [\mathbb{E}Y] = \Big(\sum_i x_i P(A_i)\Big)\Big(\sum_j y_j P(B_j)\Big) = \sum_{i,j} x_i y_j P(A_i)P(B_j) = \sum_{i,j} x_i y_j P(A_i \cap B_j),$$

где в последнем равенстве была использована независимость X и Y. Последняя сумма равна $\mathbb{E}[X \cdot Y]$ $\square$

12. Дисперсия и ковариация.

Математическое ожидание отвечает за среднее значение случайной величины. Теперь обсудим величину, которая отвечает за разброс значений случайной величины.

Дисперсией случайной величины X называется число

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

Ковариацией пары случайных величин называется число

$$\operatorname{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

Заметим, что ковариация является неотрицательно определенной билинейной формой,

$$\operatorname{cov}(X, Y) = \mathbb{E}[X \cdot Y] - [\mathbb{E}X] \cdot [\mathbb{E}Y],$$

в частности $\mathbb{D}X = \operatorname{cov}(X, X) = \mathbb{E}[X^2] - [\mathbb{E}X]^2$ и

$$\mathbb{D}[X + Y] = \operatorname{cov}(X + Y, X + Y) = \mathbb{D}X + 2\operatorname{cov}(X, Y) + \mathbb{D}Y.$$

Величину

$$\varrho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathbb{D}X}\sqrt{\mathbb{D}Y}}$$

называют **коэффициентом корреляции**.

Предложение 29. (i) *Если $\mathbb{D}X = 0$, то $X = \mathbb{E}X$ почти наверное.*

(ii) *Для произвольных чисел α и β верно равенство $\mathbb{D}[\alpha X + \beta] = \alpha^2 \mathbb{D}X$.*

(iii) *Если X и Y независимы, то $\operatorname{cov}(X, Y) = 0$ и $\mathbb{D}[X + Y] = \mathbb{D}X + \mathbb{D}Y$.*

Доказательство. Эти свойства немедленно следуют из свойств математического ожидания. $\square$

Следствие 30. (Неравенство Коши–Буняковского) *При условии $\mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty$ имеет место неравенство*

$$\Big|\mathbb{E}[XY]\Big| \le \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2},$$

причем равенство возможно тогда и только тогда, когда найдутся числа α и β такие, что $\alpha X + \beta Y = 0$ почти наверное.

Доказательство. Если $\mathbb{E}X^2 = 0$, то утверждение очевидно верно. Пусть $\mathbb{E}X^2 > 0$. Рассмотрим квадратичную функцию

$$p(t) = \mathbb{E}(Y - tX)^2 = \mathbb{E}Y^2 - 2t\mathbb{E}[XY] + t^2\mathbb{E}X^2.$$

Из неотрицательности $p(t)$ для всех t следует, что

$$D = 4\big(\mathbb{E}[XY]\big)^2 - 4[\mathbb{E}Y^2] \cdot [\mathbb{E}X^2] \le 0.$$

Последнее неравенство равносильно требуемому. $\square$

Рассмотрим линейное пространство случайных величин с нулевым математическим ожиданием. На этом пространстве введем скалярное произведение

$$\langle X, Y\rangle = \mathrm{cov}(X, Y).$$

Соответственно появляется норма $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{D}X}$. Коэффициент корреляции на случайных величинах с нулевым математическим ожиданием имеет вид

$$\varrho(X, Y) = \frac{\langle X, Y\rangle}{\|X\|\|Y\|}.$$

Таким образом, коэффициент корреляции играет роль косинуса угла между векторами X и Y.

Пример 31. Пусть случайная величина S_n имеет биномиальное распределение, т.е. $P(S_n = k) = C_n^k p^k (1-p)^{n-k},\ k \in \{0, \ldots, n\}$. Вычислим ожидание случайной величины S_n, используя свойства математического ожидания. Заметим, что математическое ожидание зависит только от распределения случайной величины. Пусть $X_1, \ldots, X_n$ — независимые бернуллиевские случайные величины с вероятностью успеха p. Тогда распрделение случайной величины $X_1 + \ldots + X_n$ совпадает с биномиальным, т.е. с распределением случайной величины S_n. Поэтому

$$\mathbb{E}S_n = \mathbb{E}X_1 + \ldots + \mathbb{E}X_n = np.$$

Теперь вычислим дисперсию:

$$\mathbb{D}S_n = \mathbb{D}X_1 + \ldots + \mathbb{D}X_n = n\mathbb{D}X_1 = n((1-p)^2 p + p^2(1-p)) = np(1-p).$$

Пример 32 (Балансировка векторов)**.** Пусть $v_1, \ldots, v_n \in \mathbb{R}^n$, причем $\|v_j\| = 1$. Всегда существует такая расстановка знаков $\varepsilon_1, \ldots, \varepsilon_n \in \{1, -1\}$, что $\|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n\| \le \sqrt{n}$.

Действительно, будем расставлять знаки случайно. Тогда

$$\mathbb{E}\|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n\|^2 = \sum_{i,j} (v_i, v_j)\mathbb{E}\varepsilon_i\varepsilon_j = \sum_{j=1}^n \|v_j\|^2 = n.$$

Если бы $\|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n\|^2 > n$ при каждом выборе знаков ε_j, то $\mathbb{E}\|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n\|^2 > n$.

13. Закон больших чисел.

Предложение 33. *Пусть $\mathbb{E}X^2 < \infty$. Тогда для каждого $\varepsilon > 0$ выполнено*

$$P(|X - \mathbb{E}X| \ge \varepsilon) \le \frac{\mathbb{D}X}{\varepsilon^2}.$$

Доказательство. Применяем неравенство Чебышева к случайной величине $|X - \mathbb{E}X|^2$. □

Следующее утверждение показывает, что при большом числе испытаний наблюдаемое эмпирическое среднее значение результатов экспериментов (например средняя частота успеха в схеме Бернулли) с большой долей вероятности близка к некоторому конкретному теоретическому среднему значению.

Следствие 34. (Закон больших чисел в слабой форме) *Пусть $\{X_n\}_n$ – последовательность таких независимых одинаково распределенных случайных величин (т.е. таблички, задающие их распределения, совпадают), что $\mathbb{E}X_n^2 < \infty$. Пусть $\mathbb{E}X_1 = m$ (а значит и $\mathbb{E}X_n = m$ при каждом n) тогда для каждого $\varepsilon > 0$*

$$\lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - m\Big| \ge \varepsilon\Big) = 0.$$

Доказательство. Заметим, что $\mathbb{E}\frac{X_1+\ldots+X_n}{n} = m$ в силу линейности. Тогда, для всякого $\varepsilon > 0$ имеет место неравенство

$$P\Big(\Big|\frac{X_1+\ldots+X_n}{n} - m\Big| \geq \varepsilon\Big) \leq \frac{\mathbb{D}[X_1+\ldots+X_n]}{n^2\varepsilon^2} = \frac{\mathbb{D}X_1}{n\varepsilon} \to 0.$$

□

Предположим теперь, что X_j – независимые бернуллиевские величины с вероятностью успеха p (положительный или отрицательный результат эксперимента). Величины $\frac{X_1+\ldots+X_n}{n}$ — частота успешного исхода эксперимента при проведении n экспериментов. Заметим, что $\mathbb{E}X_n = p, \mathbb{D}X_n = pq, q = 1-p$. Следовательно,

$$P\Big(\Big|\frac{X_1+\ldots+X_n}{n} - p\Big| \geq \varepsilon\Big) \leq \frac{pq}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Таким образом, эмпирическая частота успешного результата эксперимента при n проведенных опытах “стремится” к p, т.е. к теоретическому значению вероятности успешного результата эксперимента.

14. Теорема Муавра–Лапласа.

Пусть проводится серия из независимых испытаний Бернулли X_j с вероятностью «успеха» p. Т.е $P(X_j = 1) = p, P(X_j = 0) = 1-p = q$. Пусть S_n — количество «успешных» испытаний, т.е. $S_n = X_1+\ldots+X_n$. Случайная величина S_n имеет биномиальное распределение, т.е.

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

Исследуем поведение вероятности $P(S_n = k)$ при больших n.

Теорема 35. (локальная теорема Муавра–Лапласа) *Если $n \to \infty$, вероятность $p \in (0,1)$ фиксирована, величина $x_m := \frac{m-np}{\sqrt{npq}}$ ограничена равномерно по m и n ($a \leq x_m \leq b$), то*

$$P(S_n = m) \sim \frac{1}{\sqrt{npq}}\varphi(x_m),$$

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

Функцию φ называют плотностью *стандартного нормального распределения* и как мы увидим далее φ играет ключевую роль в теории вероятностей.

Доказательство. Приведем доказательство только в случае $p = q = 1/2$. Как было сказано выше

$$P(S_n = m) = C_n^m 2^{-n} = e^{-n\ln 2 + \ln n! - \ln m! - \ln(n-m)!}.$$

Так как $m = \frac{n}{2} + x_m\frac{\sqrt{n}}{2} = \frac{n}{2}(1 + \frac{x_m}{\sqrt{n}})$, $n - m = \frac{n}{2} - x_m\frac{\sqrt{n}}{2} = \frac{n}{2}(1 - \frac{x_m}{\sqrt{n}})$, то по формуле Стирлинга:

$$\ln n! = \ln\sqrt{2\pi n} + n\ln n - n + \bar{o}(1);$$

$$\ln m! = \ln\Big(\sqrt{\pi n}\Big(1 + x_m\frac{1}{\sqrt{n}}\Big)^{1/2}\Big) + \Big(\frac{n}{2} + x_m\frac{\sqrt{n}}{2}\Big)\Big(\ln n - \ln 2 + \ln\Big(1 + \frac{x_m}{\sqrt{n}}\Big)\Big) - \frac{n}{2} - x_m\frac{\sqrt{n}}{2} + \bar{o}(1) =$$
$$\ln\sqrt{\pi n} + \bar{o}(1) + \Big(\frac{n}{2} + x_m\frac{\sqrt{n}}{2}\Big)\Big(\ln n - \ln 2 + \frac{x_m}{\sqrt{n}} - \frac{x_m^2}{2n} + \bar{o}(\tfrac{1}{n})\Big) - \frac{n}{2} - x_m\frac{\sqrt{n}}{2} + \bar{o}(1),$$

где мы использовали формулу Тейлора $\ln(1+x) = x - x^2/2 + \bar{o}(x^2)$, ограниченность x_m и тот факт, что $\ln\big(1 + x_m\frac{1}{\sqrt{n}}\big)^{1/2} = \bar{o}(1)$. Аналогично, заменяя x_m на $-x_m$, получаем

$$\ln(n-m)! = \ln\sqrt{\pi n} + \Big(\frac{n}{2} - x_m\frac{\sqrt{n}}{2}\Big)\Big(\ln n - \ln 2 - \frac{x_m}{\sqrt{n}} - \frac{x_m^2}{2n} + \bar{o}(\tfrac{1}{n})\Big) - \frac{n}{2} + x_m\frac{\sqrt{n}}{2} + \bar{o}(1).$$

Приведя подобные, получаем

$$\ln m! + \ln(n-m)! = 2\ln\sqrt{\pi n} + n\ln n - n\ln 2 + x_m^2 - \frac{x_m^2}{2} - n + \bar{o}(1)$$

Наконец,

$$
\begin{aligned}
&-n\ln 2+\ln n!-\ln m!-\ln(n-m)! \\
&\qquad= -n\ln 2+\ln\sqrt{2\pi n}+n\ln n-n-2\ln\sqrt{\pi n}-n\ln n+n\ln 2-\tfrac{x_m^2}{2}+n+\bar{o}(1) \\
&\qquad\qquad= -\ln\sqrt{2\pi}-\ln\sqrt{\tfrac{n}{4}}-\tfrac{x_m^2}{2}+\bar{o}(1).
\end{aligned}
$$

Полученное тождество равносильно эквивалентности из формулировки теоремы в случае $p=q=1/2$. □

Заметим, что $P(a\le \frac{S_n-np}{\sqrt{npq}}\le b)$ "$\sim$" $\sum\limits_{a\le x_m\le b}\frac{1}{\sqrt{npq}}\varphi(x_m)$. Т.к. $x_{m+1}-x_m=\frac{1}{\sqrt{npq}}$ то сумма выше является интегральной для интеграла $\displaystyle\int_a^b \varphi(x)\,dx$. Если в доказательстве выше применить более точные разложения в формуле Стирлинга и для $\ln(1+x)$, можно на самом деле показать, что $|P(S_n=m)-\frac{1}{\sqrt{npq}}\varphi(x_m)|\le\frac{C}{\sqrt{n}}\frac{1}{\sqrt{npq}}\varphi(x_m)$, откуда получаем следующую теорему.

Теорема 36. (интегральная теорема Муавра–Лапласа) *Для любых чисел $a<b$ имеем*

$$\lim_{n\to\infty} P\Big(a\le\frac{S_n-np}{\sqrt{npq}}\le b\Big)=\int_a^b\varphi(x)\,dx.$$

Здесь в левой части написана вероятность того, что число "успехов" лежит в диапазоне от $np+a\sqrt{npq}$ до $np+b\sqrt{npq}$.

Отметим, что разница между вероятностью и интегралом на самом деле оценивается через $\frac{p^2+q^2}{\sqrt{npq}}$ и эта оценка точна. Следовательно, если p близко к нулю или к единице, то вероятность плохо приближается интегралом от φ.

15. Теорема Пуассона.

Как обсуждалось ранее, теорема Муавра–Лапласа обеспечивает хорошее приближение для оцениваемых вероятностей, если величина npq достаточно велика. В связи с этим возникает вопрос, как поступить в задаче, где, скажем, $p=0{,}001$, а $n=1000$. Заметим, что в этом случае $np=1$. В такой ситуации применима следующая теорема Пуассона.

Рассмотрим серию испытаний по схеме Бернулли, причем пусть N-я серия состоит из N испытаний и вероятность успеха в этой серии равна p_N. Потребуем, чтобы произведение $N\cdot p_N=\lambda$ не зависело от N. Нас интересует вероятность $P(S_N=k)$ наступления ровно k успехов в N-ой серии.

Теорема 37. (Пуассон) *Пусть $N\cdot p_N=\lambda$ – не зависит от N. Тогда*

$$P(S_N=k)=C_N^k p_N^k(1-p_N)^{N-k}\to\frac{\lambda^k}{k!}e^{-\lambda},\quad N\to+\infty.$$

Доказательство. Распишем вероятность $P(S_N=k)$ в следующем виде:

$$P(S_N=k)=\frac{\lambda^k}{k!}\Big(1-\frac{1}{N}\Big)\cdots\Big(1-\frac{k-1}{N}\Big)\Big(1-\frac{\lambda}{N}\Big)^{-k}\Big(1-\frac{\lambda}{N}\Big)^{N}.$$

Учитывая, что λ и k не меняются, устремляем $N\to\infty$ и получаем искомое выражение. □

Приведем несколько примеров.

Опечатки при наборе книги

Предположим, что при наборе книги наборщиком в среднем совершается λ опечаток на странице. Какова вероятность того, что данная страница книги будет без опечаток? Какова вероятность обнаружить на данной конкретной странице книги ровно k опечаток?

Предположим, что всего при наборе было совершено N опечаток, а в книге N/λ страниц. Тогда вероятность того, что на данной странице будет присутствовать какая-то одна опечатка равна λ/N. Значит вероятность того, что на данной странице не будет опечаток вычисляется по формуле

$$(1-\lambda/N)^N \to e^{-\lambda}$$

при большом количестве страниц в книге. Вероятность же того, что на странице обнаружено ровно k опечаток, по теореме Пуассона хорошо аппроксимируется значением $\frac{\lambda^k}{k!}e^{-\lambda}$.

Булочка с изюмом

Сколько изюма должны содержать в среднем булочки, для того чтобы вероятность иметь хотя бы одну изюминку в булочке была не меньше $0,99$?

Предположим, что уже изготовлено тесто на некоторое количество булочек. В это тесто добавлено N изюминок так, что отношение числа изюминок к количеству булочек равно λ. Значит количество булочек равно N/λ.

Выделим в тесте кусок, из которого будет изготовлена данная булочка. Вероятность попадания одной изюминки в эту булочку равна λ/N, а вероятность того, что хотя бы одна изюминка попала в булку, равна

$$1-\Big(1-\frac{\lambda}{N}\Big)^N.$$

Поскольку мы рассматриваем серийное производство булочек, то можно предполагать, что $N \to +\infty$, т. е. растет объем теста и количество изюма, но не меняется плотность λ. Как и выше, получаем $\big(1-\frac{\lambda}{N}\big)^N \to e^{-\lambda}$. Для решения задачи надо найти λ такое, что $e^{-\lambda} < 0,01$. Подходит $\lambda = 5$, т. е. плотность изюма должна быть не менее пяти изюминок на булочку.

Пуассоновский процесс

В случайные моменты времени регистрируются некоторые события (например столкновение элементарных частиц или обращения пользователей на сайт). Будем отмечать эти моменты времени точками на луче $[0,+\infty)$. Обозначим через X_t число точек на временном промежутке $(0,t]$ (число событий, произошедших за время t). Пусть $P_k(t) = P\{X_t = k\}$. Будем предполагать, что

(i) $P\{X_{t_2} - X_{t_1} = k\} = P_k(t_2 - t_1)$, т.е. вероятность того, что за данный промежуток времени произошло k событий зависит только от длины этого временного промежутка, но не зависит от его расположения на оси времени;

(ii) для любой конечной системы временных промежутков, которые могут попарно пересекаться лишь концами, попадания точек в каждый из них являются независимыми в совокупности событиями (иными словами, при $0 \le t_1 < t_2 < \ldots < t_n$ случайные величины $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ независимы);

(iii) $P(X_\delta \ge 2) = o(\delta)$, т.е. вероятность того, что за малый промежуток времени δ произошло по крайней мере два события является о-малым от длины δ этого промежутка.

На практике многие явления удовлетворяют этим условиям, например вызовы на телефонной станции или радиоактивный распад. Величина X_t называется пуассоновским процессом. Мы не будем обсуждать существование такого процесса и даже вероятностное пространство, на котором мы работаем в данной модели, а исходя только из свойств $(i), (ii), (iii)$ вычислим вероятности $P_k(t)$.

Рассмотрим сначала $P_0(t)$. Разделим промежуток $[0,t]$ на N промежутков. По свойству (i) вероятность отсутствия событий на каждом промежутке разбиения равна $P_0(t/N)$. По свойству (ii) регистрация события на одном промежутке разбиения не зависит от регистрации событий на других промежутках. Следовательно, мы имеем дело со схемой Бернулли и вероятность отсутствия событий на $[0,t]$ равна $P_0(t) = P_0(t/N)^N$. Положим $P_0(1) = q$. Тогда $P_0\big(\frac{1}{N}\big) = q^{\frac{1}{N}}$ и $P_0\big(\frac{m}{N}\big) = q^{\frac{m}{N}}$. Заметим, что $P_0(t+h) = P_0(t)P_0(h) \le P_0(t)$, т. е. $P_0(t)$ не возрастает. Следовательно, для $\frac{m-1}{N} \le t \le \frac{m}{N}$ выполняются неравенства $q^{\frac{m-1}{N}} \le P_0(t) \le q^{\frac{m}{N}}$. Приближая t последовательностью дробей $\frac{m}{N}$, приходим к равенству $P_0(t) = q^t$. Положим $\lambda = -\ln q > 0$. Тогда $P_0(t) = e^{-\lambda t}$.

Теперь вычислим $P_k(t)$ при $k > 0$. Опять разобьем промежуток $[0, t]$ на N промежутков. Пусть B – событие, состоящее в том, что хотя бы на одном из промежутков зарегистрированы по крайней мере два события. Тогда противоположное событие $\overline{B}$ состоит в том, что в каждом промежутке регистрируется не более одного события. Заметим, что по свойству (iii) вероятность B не превосходит $No(t/N) = t \cdot o(1)$, что стремится к нулю при $N \to \infty$. Пусть событие A_k заключается в том, что есть ровно $N - k$ промежутков разбиения, на которых не зарегистрировано ни одного события. Тогда

$$P_k(t) = P(\{X_t = k\} \cap \overline{B}) + P(\{X_t = k\} \cap B) =$$
$$= P(A_k \cap \overline{B}) + P(\{X_t = k\} \cap B) = P(A_k) - P(A_k \cap B) + P(\{X_t = k\} \cap B).$$

Вероятность события A_k совпадает с вероятностью того, что в схеме Бернулли с вероятностью неудачи $P_0(t/N) = e^{-\lambda t/N}$ произошло ровно $N - k$ неудач. Поэтому

$$P(A_k) = C_N^k \big(e^{-\lambda t/N}\big)^{N-k} \big(1 - e^{-\lambda t/N}\big)^k.$$

Устремляем здесь $N \to \infty$ и в качестве предела получаем $\frac{(\lambda t)^k}{k!} e^{-\lambda t}$.

С учетом сказанного про стремление к нулю $P(B)$ приходим к равенству

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

т. е. получаем распределение Пуассона. Число λ называется интенсивностью или параметром процесса $X(t)$.

16. Общее определение вероятностного пространства.

Перейдем теперь к рассмотрению общего определения вероятности для бесконечного (несчетного) множество элементарных исходов Ω.

Набор множеств, на которых определяется вероятностная мера и которые мы называем событиями, должен удовлетворять некоторым естественным условиям: если A и B события, то к классу событий следует отнести A **и** B, A **или** B, **не** B, т.е. событиями должны быть $A \cap B$, $A \cup B$, $\Omega \setminus B$. Кроме того, нам часто придется работать с пределами случайных величин (например, при рассмотрении различных предельных теорем), поэтому естественно требовать, чтобы для последовательности случайных величин X_n (формальное определение случайной величины в общем случае мы дадим чуть позже) событием было бы и множество $\{\omega\colon \exists \lim\limits_{n\to\infty} X_n(\omega)\}$. Последнее с помощью критерия Коши переписывается в виде $\bigcap\limits_{k\in\mathbb{N}} \bigcup\limits_{N\in\mathbb{N}} \bigcap\limits_{n=N}^{\infty} \bigcap\limits_{m=N}^{\infty} \{\omega\colon |X_n(\omega) - X_m(\omega)| < \frac{1}{k}\}$. Если теперь сделать естественное предположение, что множество $\{\omega\colon |X_n(\omega) - X_m(\omega)| < \frac{1}{k}\}$ является событием, то событиями обязаны быть и счетные пересечения и объединения событий.

Определение 38. Класс $\mathcal{A}_0$ подмножеств пространства Ω называется **алгеброй**, если
(i) $\Omega, \emptyset \in \mathcal{A}_0$;
(ii) $A \in \mathcal{A}_0 \Rightarrow \Omega \setminus A \in \mathcal{A}_0$;
(iii) $A, B \in \mathcal{A}_0 \Rightarrow A \cap B, A \cup B \in \mathcal{A}_0$.

Определение 39. Класс $\mathcal{A}$ подмножеств пространства Ω называется **σ-алгеброй**, если
(i) $\Omega, \emptyset \in \mathcal{A}$;
(ii) $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$;
$(iii)'$ $A_n \in \mathcal{A}\ \forall n \in \mathbb{N} \Rightarrow \bigcap\limits_{n=1}^{\infty} A_n, \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$.

Отметим, что в силу формул $\Omega \setminus \bigcup\limits_{\alpha} A_\alpha = \bigcap\limits_{\alpha}(\Omega \setminus A_\alpha)$ и $\Omega \setminus \bigcap\limits_{\alpha} A_\alpha = \bigcup\limits_{\alpha}(\Omega \setminus A_\alpha)$ в свойствах (iii) и $(iii)'$ достаточно проверять включение любо только для объединений, либо только для пересечений.

Например, множество всех подмножеств 2^Ω, $\{\emptyset, \Omega\}$, $\{\emptyset, B, \Omega \setminus B, \Omega\}$ являются σ алгебрами. Множество всех конечных объединений попарно непересекающихся промежутков $(a, b]$ на $\mathbb{R}$ является алгеброй, но не является σ алгеброй (она не содержит одноточечные множества — пересечения счетного числа полуинтервалов).

Определение 40. Говорят, что σ алгебра **порождена набором множеств** S, если эта σ алгебра является наименьшей по включению среди всех σ-алгебр, которые содержат данный набор множеств S. Такую σ-алгебру обозначают $\sigma(S)$.

Отметим, что $\sigma(S)$ определена для произвольного набора подмножеств S. Действительно, легко видеть, что пересечение произвольного набора σ-алгебр оказывается опять σ-алгеброй, поэтому $\sigma(S)$ просто совпадает с пересечением всех σ-алгебр, содержащих S. Хотя бы одна такая σ-алгебра всегда есть, например σ-алгебра всех подмножеств 2^Ω.

Важнейшим примером является **борелевская** σ-алгебра $\mathcal{B}(\mathbb{R})$ подмножеств прямой $\mathbb{R}$, которая порождена всеми промежутками, т. е. отрезками, интервалами, полуинтервалами. Несложно показать, что в определении не обязательно в качестве порождающего множества брать все промежутки. Например, можно ограничиться только отрезками или только интервалами или только лучами $(-\infty, c]$. Например, проверим, что $\mathcal{B}(\mathbb{R})$ порождена всеми лучами вида $(-\infty, c]$. Действительно, $(-\infty, c] \in \mathcal{B}(\mathbb{R})$, как счетное объединение промежутков вида $(-n, c]$, поэтому $\sigma(\{(-\infty, c]\}) \subset \mathcal{B}(\mathbb{R})$. С другой стороны $(a, b] = (-\infty, b] \setminus (-\infty, a]$, отрезки получаются счетным пересечением промежутков вида $(a - \frac{1}{n}, b]$, интервалы получаются счетным объединением промежутков вида $(a, b - \frac{1}{n}]$, а полуинтервалы вида $[a, b)$ получаются объединением уже полученных отрезков вида $[a, b - \frac{1}{n}]$. Тем самым, все промежутки принадлежат $\sigma(\{(-\infty, c]\})$, а значит имеет место и включение $\mathcal{B}(\mathbb{R}) \subset \sigma(\{(-\infty, c]\})$.

Аналогично определяется $\mathcal{B}([0, 1])$ или $\mathcal{B}(\mathbb{R}^n)$ (в многомерном случае надо промежутки заменить на параллелепипеды, а вместо лучей можно рассматривать множества вида $(-\infty, c_1] \times \ldots \times (-\infty, c_n]$).

17. Общее определение вероятностного пространства: вероятностная мера

Определение 41. Пусть $\mathcal{A}_0$ — алгебра множеств. Функция $P\colon \mathcal{A}_0 \to [0, 1]$ называется **аддитивной**, если для произвольных $A, B \in \mathcal{A}_0$, $A \cap B = \emptyset$ выполнено

$$P(A \cup B) = P(A) + P(B).$$

Функция $P\colon \mathcal{A}_0 \to [0, 1]$ называется **счетно аддитивной**, если для всякого не более чем счетного набора попарно непересекающихся событий $A_n \in \mathcal{A}$, для которых $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}_0$ выполняется

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

Ясно, что в случае, когда $\mathcal{A}_0$ — σ-алгебра, множество $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}_0$ для произвольных $A_n \in \mathcal{A}_0$. Кроме того, для аддитивной функции $P\colon \mathcal{A}_0 \to [0, 1]$ выполнено $P(\emptyset) = 0$, $P(\Omega \setminus A) = P(\Omega) - P(A)$ для произвольного $A \in \mathcal{A}_0$, $P(A) \le P(B)$, если $A \subset B$, $A, B \in \mathcal{A}_0$, $P(A \cup B) \le P(A) + P(B)$ для $A, B \in \mathcal{A}_0$.

Определение 42. Пусть $\mathcal{A}$ – σ алгебра. Функция $P\colon \mathcal{A} \to [0, 1]$ называется **вероятностной мерой** на $\mathcal{A}$, если $P(\Omega) = 1$ и P — счетно аддитивна на $\mathcal{A}$, т.е. для всякого не более чем счетного набора попарно непересекающихся событий $A_n \in \mathcal{A}$ выполняется

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

Тройку $(\Omega, \mathcal{A}, P)$ называют **вероятностным пространством**.

Предложение 43. *Пусть $P\colon \mathcal{A}_0 \to [0,1]$ — аддитивная функция множества на алгебре $\mathcal{A}_0$. Функция P счетно аддитивна на $\mathcal{A}_0$ тогда и только тогда, когда для произвольного набора $A_n \in \mathcal{A}$, $A_{n+1} \subset A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$ выполнено $\lim\limits_{n\to\infty} P(A_n) = 0$.*

Доказательство. Пусть P счетно аддитивна на $\mathcal{A}_0$. Рассмотрим множества $C_n = A_n \setminus A_{n+1}$. Тогда $A_1 = \bigcup\limits_{n=1}^{\infty} C_n$, $A_{N+1} = \bigcup\limits_{k=N+1}^{\infty} C_k$ и $P(A_1) = \sum\limits_{n=1}^{N} P(C_n) + P(A_{N+1})$. Если P счетно аддитивна, то $\sum\limits_{n=1}^{N} P(C_n) \to P(A_1)$ и $P(A_{N+1}) \to 0$.

Наоборот, пусть $C_n \in \mathcal{A}_0$ — набор попарно непересекающихся множеств, причем известно, что $\bigcup\limits_{n=1}^{\infty} C_n = A_1 \in \mathcal{A}_0$. Пусть $A_{N+1} = \bigcup\limits_{k=N+1}^{\infty} C_k$, тогда $A_{N+1} \subset A_N$, причем $\bigcap\limits_{N=1}^{\infty} A_N = \emptyset$. Если $P(A_{N+1}) \to 0$, то $P(A_1) = \sum\limits_{n=1}^{N} P(C_n) + P(A_{N+1})$ и переходя к пределу, получаем $P(A_1) = \sum\limits_{n=1}^{\infty} P(C_n)$. □

Следующие свойства называют непрерывностью меры P.

Следствие 44. *Пусть $(\Omega, \mathcal{A}, P)$ — вероятностное пространство. Тогда*

(i) Если $A_n \in \mathcal{A}$, $A_{n+1} \subset A_n$ и $A = \bigcap\limits_{n=1}^{\infty} A_n$, то $\lim\limits_{n\to\infty} P(A_n) = P(A)$.

(ii) Если $A_n \in \mathcal{A}$, $A_n \subset A_{n+1}$ и $A = \bigcup\limits_{n=1}^{\infty} A_n$, то $\lim\limits_{n\to\infty} P(A_n) = P(A)$.

Доказательство. Ясно, что (ii) следует из (i) переходом к дополнению. Для доказательства свойства (i) достаточно рассмотреть последовательность вложенных множеств $A_n \setminus \bigcap\limits_{k=1}^{\infty} A_k$ с пустым пересечением и применить предыдущую теорему. □

Замети, что, в частности, для вероятностной меры P справедливо следующее свойство **счетной субаддитивности**

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{N\to\infty} P\Big(\bigcup_{n=1}^{N} A_n\Big) \le \sum_{n=1}^{\infty} P(A_n).$$

Важной (и трудной в смысле доказательства) теоремой является следующая теорема о продолжении меры.

Теорема 45 (б/д)**.** *Пусть $\mathcal{A}_0$ есть некоторая алгебра подмножеств пространства Ω и пусть $P_0\colon \mathcal{A}_0 \to [0,1]$* **счетно аддитивная** *функция множества на алгебре $\mathcal{A}_0$. Тогда существует единственная вероятностная мера P на $\sigma(\mathcal{A}_0)$, продолжающая функцию P_0, т.е. $P(A) = P_0(A)$ для произвольного множества $A \in \mathcal{A}_0$.*

Пример 46. Одним из важнейших примеров вероятностной меры является **мера Лебега** λ на $\mathcal{B}([0,1])$. Мера Лебега — обычная длина, т. е. $\lambda([a,b]) = b - a$. Обсудим схему построения такой меры. Рассмотрим алгебру $\mathcal{A}_0$ конечных объединений попарно непересекающихся промежутков вида $(a,b] \subset [0,1]$ и возможно одноточечного множества $\{0\}$.

Для множества $A = \bigsqcup_{j=1}^{m} (a_j, b_j]$ с попарно непересекающимися $(a_j, b_j]$ зададим меру Лебега равенством $\lambda(A) := \sum_{j=1}^{m} (b_j - a_j)$. Нетрудно проверить, что это корректно определенная аддитивная функция множества на $\mathcal{A}_0$. Если теперь проверить, что она оказывается счетно аддитивной на этой алгебре (что верно), то по теореме о продолжении меры существует единственная вероятностная мера на $\mathcal{B}([0,1])$, совпадающая с λ на $\mathcal{A}_0$.

18. Случайная величина в общем случае.

Напомним, что в случае дискретного вероятностного пространства случайными величинами были функции, действующие из этого вероятностного пространства в $\mathbb{R}$. Т.к. случайная величина описывает результаты некоторого эксперимента, нас обычно интересуют вероятности событий типа: значение случайной величины не меньше c, где c некоторое число. Т.к. при определении общего вероятностного пространства не все множества являются событиями (не всем множествам приписана вероятность), то мы не можем рассматривать в качестве случайных величин произвольные функции. Нам необходимо позаботиться, чтобы мы могли говорить про вероятности интересующих нас событий, связанных с рассматриваемой случайной величиной. Т.е. для того, чтобы назвать некоторую функцию $X\colon \Omega \to \mathbb{R}$ случанйо величиной, нам необходимо хотя бы уметь приписывать вероятности событиям вида $\{\omega \in \Omega\colon X(\omega) \le t\}$ для произвольного $t \in \mathbb{R}$ (что соответствует событию: результат эксперимента не меньше заданного значения).

Таким образом, мы пришли к следующему определению.

Определение 47. Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, P)$. Функция $X\colon \Omega \to \mathbb{R}$ называется **случайной величиной**, если для всякого числа $t \in \mathbb{R}$ выполнено

$$X^{-1}((-\infty, t]) = \{\omega \in \Omega\colon\ X(\omega) \le t\} \in \mathcal{A}.$$

Предложение 48. *Если X случайная величина, то $\{\omega\colon X(\omega) \in B\} = X^{-1}(B) \in \mathcal{A}$ для всякого $B \in \mathcal{B}(\mathbb{R})$.*

Доказательство. Напомним следующие соотношения для прообраза функции:

$$X^{-1}\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \bigcap_{n=1}^{\infty} X^{-1}(A_n), \quad X^{-1}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \bigcup_{n=1}^{\infty} X^{-1}(A_n), \quad X^{-1}(\mathbb{R} \setminus B) = \Omega \setminus X^{-1}(B).$$

Рассмотрим систему множеств

$$\mathcal{C} := \{B \subset \mathbb{R}\colon X^{-1}(B) \in \mathcal{A}\}.$$

Эта система образует σ-алгебру. Действительно, $X^{-1}(\emptyset) = \emptyset \in \mathcal{A}$ и $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{A}$. Если $B \in \mathcal{C}$, то $X^{-1}(\mathbb{R} \setminus B) = \Omega \setminus X^{-1}(B) \in \mathcal{A}$. Наконец, если $B_n \in \mathcal{C}$, то

$$X^{-1}\Big(\bigcap_{n=1}^{\infty} B_n\Big) = \bigcap_{n=1}^{\infty} X^{-1}(B_n) \in \mathcal{A}.$$

По условию σ алгебра $\mathcal{C}$ содержит все лучи вида $(-\infty, t]$. Мы знаем, что $\mathcal{B}(\mathbb{R})$ — наименьшая по включению σ алгебра, содержащая все лучи такого вида, поэтому $\mathcal{B}(\mathbb{R}) \subset \mathcal{C}$, что и требовалось. $\square$

Отметим, что если $\mathcal{A} = 2^{\Omega}$, то всякая функция из Ω в $\mathbb{R}$ является случайной величиной.

Замечание 49. Т.к. $\{X^2 \le t\} = \{-\sqrt{t} \le x \le \sqrt{t}\}$ (при $t \ge 0$) и отрезок $[-\sqrt{t}, \sqrt{t}]$ — борелевское множество, получаем, что X^2 — также случайная величина. Можно проверить, что для случайной величины X и для любой «разумной» функции $f\colon \mathbb{R} \to \mathbb{R}$ (например, если f непрерывная), $f(X)$ также будет случайной величиной.

Предложение 50. *Пусть X, Y — случайные величины. Тогда случайными величинами будут $\alpha X + \beta Y$, $X \cdot Y$.*

Доказательство. Ясно, что αX и βY — случайные величины. Проверим, что $X + Y$ — случайная величина:

$$\{X + Y > t\} = \{X > t - Y\} = \bigcup_{r_n \in \mathbb{Q}} (\{X > r_n\} \cap \{r_n > t - Y\}) \in \mathcal{A}.$$

Поэтому и $\{X + Y \le t\} \in \mathcal{A}$, а значит $X + Y$ — случайная величина. Для произведения заметим, что $X \cdot Y = 2^{-1}\big((X+Y)^2 - X^2 - Y^2\big)$. и утверждения следует из уже доказанных. $\square$

Можно показать, что для произвольной «разумной» функции $f\colon \mathbb{R}^2 \to \mathbb{R}$ (например для непрерывной f), $f(X, Y)$ также будет случайной величиной.

Предложение 51. *Пусть X_n – случайные величины и для всякого ω существует предел $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$. Тогда X является случайной величиной.*

Доказательство. Рассмотрим множество $\{\omega\colon X(\omega) \le t\}$. Заметим, что $X(\omega) \le t$ тогда и только тогда, когда для каждого натурального числа k найдется такой номер N, что для всех $n > N$ верно неравенство $X_n(\omega) \le t + \frac{1}{k}$. На языке теории множеств эту фразу можно записать так

$$\{\omega\colon X(\omega) \le t\} = \bigcap_k \bigcup_N \bigcap_{n>N} \{\omega\colon X_n(\omega) \le t + \frac{1}{k}\}.$$

Остается заметить, что $\{\omega\colon X_n(\omega) \le t + \frac{1}{k}\} \in \mathcal{A}$. $\square$

Таким образом, со случайными величинами можно выполнять арифметические операции и переходить к пределу.

Определение 52. Распределением случайной величины X называется вероятностная мера μ_X на $\mathcal{B}(\mathbb{R})$, определяемая равенством

$$\mu_X(B) = P(\{\omega\colon X(\omega) \in B\}) = P(X^{-1}(B)).$$

Обратим внимание, что, как и в дискретном случае, распределение случайной величины это мера на значениях случайной величины, т. е. мера μ_X показывает с какой вероятностью принимаются те или иные значения X.

Определение 53. Функция

$$F_X(t) = \mu_X((-\infty, t]) = P(\{\omega\colon X(\omega) \le t\})$$

называется **функцией распределения** случайной величины X.

Из определения F_X следует, что $P(a < X \le b) = \mu_X((a, b]) = F(b) - F(a)$.

Предложение 54. *Функция F_X удовлетворяет следующим свойствам:*
(i) $F_X\colon \mathbb{R} \to [0, 1]$ не убывает;
(ii) F_X непрерывна справа;
(iii) $\lim\limits_{t\to-\infty} F_X(t) = 0$ и $\lim\limits_{t\to+\infty} F_X(t) = 1$.

Доказательство. Т.к. $\{\omega\colon X(\omega) \le t\} \subset \{\omega\colon X(\omega) \le s\}$ при $t \le s$, то получаем свойство (i).

Обоснуем пункт (ii). Пусть $t_n \to t$, $t_n \ge t$. Заметим, что $\{X \le t\} = \bigcap\limits_{k=1}^{\infty} \{X \le t + \frac{1}{k}\}$. В силу непрерывности вероятностной меры P получаем

$$\lim_{n\to\infty} F_X(t + \tfrac{1}{k}) = \lim_{n\to\infty} P(X \le t + \tfrac{1}{k}) = P(X \le t) = F_X(t).$$

Значит для каждого $\varepsilon > 0$ найдется k, для которого $F_X(t) \le F_X(t + \frac{1}{k}) \le F_X(t) + \varepsilon$. Т.к. $t_n \to t$, $t_n \ge t$, то найдется номер n_0, начиная с которого $t \le t_n < t + \frac{1}{k}$. В силу монотонности $F_x(t) \le F_X(t_n) \le F_X(t + \frac{1}{k}) \le F_X(t) + \varepsilon$ при $n \ge n_0$. Это и означает, что $\lim\limits_{n\to\infty} F_X(t_n) = F(t)$.

Свойство (iii) обосновывается аналогично. □

Теорема 55 (б/д). *Распределение μ_X однозначно определяется функцией распределения F_X. Кроме того, если задана функция F, удовлетворяющая свойствам (i), (ii), (iii), то существует вероятностное пространство $(\Omega, \mathcal{A}, P)$ и случайная величина X с функцией распределения F.*

Эта теорема позволяет говорить о распределении случайной величины без уточнения, на каком вероятностном пространстве задана случайная величина и как именно она задана.

Наметим основные идеи доказательства. Первая часть является прямым следствием теоремы о продолжении меры. Пусть $A = \bigcup_{j=1}^n (a_j, b_j]$, причем $(a_j, b_j] \cap (a_k, b_k] = \emptyset$ при $j \ne k$. Тогда $\mu_X(A) = \sum\limits_j F_X(b_j) - F_X(a_j)$. Кроме того, множества A указанного вида образуют алгебру $\mathcal{A}_0$ подмножеств $\mathbb{R}$. Поэтому, если есть две случайные величины с одной и той же функцией распределения, то по теореме о продолжении меры (часть о единственности продолжения) их распределения также совпадают на всех множествах из $\sigma(\mathcal{A}_0) = \mathcal{B}(\mathbb{R})$.

Доказательство второй части аналогично рассуждению о построении меры Лебега. Будем строить вероятностную меру P на $\Omega = \mathbb{R}$ с $\mathcal{A} = \mathcal{B}(\mathbb{R})$. Рассмотрим алгебру $\mathcal{A}_0$ множеств вида $A = \bigcup\limits_{j=1}^n (a_j, b_j]$, где $(a_j, b_j] \cap (a_k, b_k] = \emptyset$ при $j \ne k$. Для такого множества A положим $P(A) := \sum\limits_{j=1}^n F(b_j) - F(a_j)$. Нетрудно видеть, что корректно определена (т.е. для разных представлений A равенство дает одно и тоже число) аддитивная функция множества на алгебре $\mathcal{A}_0$. Если теперь суметь проверить счетную аддитивность P на $\mathcal{A}_0$, то P продолжается до счетно аддитивной меры на $\sigma(\mathcal{A}_0) = \mathcal{B}(\mathbb{R})$. Если теперь рассмотреть случайную величину $X(\omega) = \omega$, то $F_X(t) = F(t)$ при $t \in \mathbb{R}$ в силу того, что $P((-\infty, t])$, являясь продолжением, стовпадает с $F(t) - F(-\infty) = F(t)$.

19. Абсолютно непрерывные распределения

Определение 56. Говорят, что случайная величина X имеет **абсолютно непрерывное** распределение (или является абсолютно непрерывной), если существует такая неотрицательная (и интегрируемая) функция ϱ_X, что

$$F_X(t) = \int_{-\infty}^t \varrho_X(x)\,dx,$$

Функция ϱ_X называется **плотностью** случайной величины X.

Отметим, что в данном случае

$$\mu_X((a, b]) = F_X(b) - F_X(a) = \int_a^b \varrho_X(x)\,dx,$$

кроме того $P(X = a) = \lim\limits_{n\to\infty} [F_X(a) - F_X(a - 1/n)] = 0$ (непрерывность интеграла с переменным пределом). На самом деле можно доказать, что

$$\mu_X(A) = \int_A \varrho_X(x)\,dx$$

для всякого множества A, для которого имеет смысл интеграл в правой части, т. е. функция $I_A\varrho_X$ интегрируема по Риману, где $I_A(x) = 1$ при $x \in A$ и $I_A(x) = 0$ при $x \notin A$.

Отметим несколько свойств плотности распределения:

1) $\varrho_X \geq 0$,
2) $\int_{-\infty}^{+\infty} \varrho_X(x)\, dx = 1$,
3) $F'_X(x) = \varrho_X(x)$ для любой точки непрерывности функции ϱ_X.

Последнее свойство следует из теоремы о дифференцируемости интеграла с переменным верхнем пределом.

Приведем несколько важных примеров:

(I) **Равномерное распределение**

Случайная величина имеет *равномерное распределение на отрезке* $[a, b]$, если ее распределение задано плотностью

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

Такая случайная величина описывает случайное бросание точки в отрезок $[a, b]$. Вероятность того, что точка попадёт в отрезок $[c, d] \subset [a, b]$ равна $\frac{d-c}{b-a}$.

(II) **Нормальное распределение**

Случайная величина имеет *нормальное распределение* с параметрами a и σ^2, если ее распределение задано плотностью

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

В случае $a = 0$ и $\sigma = 1$ эта плотность появлялась в теореме Муавра–Лапласа.

(III) **Экспоненциальное (показательное) распределение**

Случайная величина имеет *экспоненциальное распределение* (которое еще иногда называется показательным) с параметром $\lambda > 0$, если ее распределение задано плотностью

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

Функция распределения такой случайной величины имеет вид $F(t) = 1 - e^{-\lambda x}$.

Определение 57. Случайная величина X называется **дискретной**, если множество ее значений конечно или счетно. Если $x_1, \dots, x_N, \dots$ — различные значения X, то множества $A_i = \xi^{-1}\{x_i\}$ попарно не пересекаются. Пусть $p_i = P(A_i)$. Тогда распределение μ_ξ имеет вид

$$\mu_\xi = p_1\delta_{x_i} + \ldots + p_N\delta_{x_N} + \ldots$$

и полностью определяется значениями x_i и p_i. В этой формуле $\delta_{x_i}(A) := 1$, если $x_i \in A$ и $\delta_{x_i}(A) = 0$, если $x_i \notin A$ для каждого $A \in \mathcal{B}(\mathbb{R})$.

Свойства дискретных случайных величин ни чем не отличаются от свйоств случайных величин на дискретных пространствах.

20. Совместное распределение случайных величин.

Рассмотрим совместное распределение нескольких случайных величин. Для простоты будем рассматривать случай двух случайных величин, хотя все рассуждения и определения переносятся и на общий случай.

Определение 58. Пусть X и Y — случайные величины. **Совместным распределением** случайных величин X, Y называется вероятностная мера $\mu_{X,Y}$ на $\mathcal{B}(\mathbb{R}^2)$, определяемая следующим образом:

$$\mu_{X,Y}(B) = P(\{\omega\colon (X(\omega), Y(\omega)) \in B\}).$$

Предложение 59. *Определение выше корректно в том смысле, что для* $B \in \mathcal{B}(\mathbb{R}^2)$

$$\{\omega\colon (X(\omega), Y(\omega)) \in B\} \in \mathcal{A}.$$

Доказательство. Рассмотрим отображение $g\colon \Omega \to \mathbb{R}^2$, $g(\omega) = (X(\omega), Y(\omega))$. Аналогично тому, как мы уже делали, проверяется, что система множеств

$$\mathcal{C} := \{B \subset \mathbb{R}^2 \colon g^{-1}(B) \in \mathcal{A}\}$$

является σ-алгеброй. Заметим, что параллелепипеды $[a, b] \times [c, d] \in \mathcal{C}$, т.к.

$$g^{-1}([a, b] \times [c, d]) = \{\omega\colon X(\omega) \in [a, b], Y(\omega) \in [c, d]\} = \{\omega\colon X(\omega) \in [a, b]\} \cap \{\omega\colon Y(\omega) \in [c, d]\}.$$

Тем самым, $\mathcal{C}$ — некотрая σ-алгебра, содержащая все параллелепипеды, а $\mathcal{B}(\mathbb{R}^2)$ — это наименьшая по включению σ-алгебра, содержащая все параллелепипеды. $\square$

21. Функция совместного распределения случайных величин.

Определение 60. Функцию

$$F_{X,Y}(x, y) = P(\{\omega\colon X(\omega) \le x,\, Y(\omega) \le y\}) = \mu_{X,Y}((-\infty, x] \times (-\infty, y]).$$

называют **функцией совместного распределения** случайных величин X и Y или функцией распределения случайного вектора (X, Y).

Заметим, что

$$\mu_{X,Y}((a, b] \times (c, d]) = F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c).$$

Предложение 61. *Функция F совместного распределения пары случайных величин удовлетворяет следующим свойствам:*

(i) $F\colon \mathbb{R}^2 \to [0, 1]$ и $F(b, d) - F(a, d) - F(b, c) + F(a, c) \ge 0$ для всякого прямоугольника $(a, b] \times (c, d]$;

(ii) F непрерывна справа по совокупности переменных;

(iii) $\lim\limits_{(x,y)\to(u,v)} F(x, y) = 0$ если хотя бы одна из переменных u или v равна $-\infty$;

(iv) $\lim\limits_{(x,y)\to(+\infty,+\infty)} F(x, y) = 1$.

Доказательство. Доказательство повторяет рассуждения одномерного случая. Например, докажем (ii). Заметим, что

$$\bigcap_{k\in\mathbb{N}} \{\omega\colon X(\omega) \le x + \tfrac{1}{k}, Y(\omega) \le y + \tfrac{1}{k}\} = \{\omega\colon X(\omega) \le x, Y(\omega) \le y\}.$$

Поэтому $P(X \le x + 1/k, Y \le y + 1/k) \to P(X \le x, Y \le y)$ и для каждого ε найдется такое k, что $P(X \le x + 1/k, Y \le y + 1/k) < P(X \le x, Y \le y) + \varepsilon$. Если теперь $x_n \to x$, $x_n \ge x$, $y_n \to y$, $y_n \ge y$, то для произвольного k найдется номер n_0, начиная с которого выполняется $x \le x_n < x + 1/k$, $y \le y_n < y + 1/k$. Поэтому при $n > n_0$

$$P(X \le x, Y \le y) \le P(X \le x_n, Y \le y_n) \le P(X \le x+1/k, Y \le y+1/k) < P(X \le x, Y \le y)+\varepsilon.$$

Утверждения (iii) и (iv) обосновываются аналогично. $\square$

Теорема 62 (б/д). *Совместное распределение пары случайных величин $\mu_{X,Y}$ однозначно задается функцией совместного распределения $F_{X,Y}$. Кроме того, для всякой функции F, удовлетворяющей свойствам $(i), (ii), (iii), (iv)$, существует вероятностное пространство $(\Omega, \mathcal{A}, P)$ и пара случайных величин X, Y с функцией совместного распределения F.*

Если известно совместное распределение вектора (X, Y), то можно найти распределение каждой из компонент:

$$F_X(x) = P(X \le x, Y \in \mathbb{R}) = \lim_{y\to+\infty} F_{X,Y}(x, y), \quad F_Y(y) = \lim_{x\to+\infty} F_{X,Y}(x, y)$$

или

$$\mu_X(U) = \mu_{X,Y}(U \times \mathbb{R}), \quad \mu_Y(V) = \mu_{X,Y}(\mathbb{R} \times V).$$

Однако, если известны только распределения случайных величин, то найти совместное распределение нельзя.

Пример 63. Пусть в квадрате $[0,1] \times [0,1]$ случайно выбирается точка (x, y). Случайные величины $X(x, y) = x$ и $Y(x, y) = y$ имеют равномерное распределение на $[0,1]$ и их совместное распределение является равномерным на $[0,1] \times [0,1]$, т. е. вероятность попадания в множество B равна площади этого множества. Будем теперь выбирать точку (x, y) случайным образом на диагонале квадрата $[0,1] \times [0,1]$, а случайные величины останутся прежними. Для всякого отрезка $[a, b] \subset [0,1]$ вероятность того, что $(x, y) \in [a, b] \times \mathbb{R}$ равна вероятности попасть в отрезок длины $(b-a)\sqrt{2}$ при бросании точки на отрезок длины $\sqrt{2}$, т. е. равна $b - a$. Таким образом, X и Y опять имеют равномерное распределение на $[0,1]$, но совместное распределение у них совсем другое.

Если известно совместное распределение X и Y, то можно найти распределение $f(X, Y)$, где f — «разумная» (например, непрерывная) функция двух переменных. Действительно, для всякого $B \in \mathcal{B}(\mathbb{R})$

$$\{\omega\colon f(X(\omega), Y(\omega)) \in B\} = \{\omega\colon (X(\omega), Y(\omega)) \in f^{-1}(B)\}.$$

22. Абсолютно непрерывные многомерные распределения

Определение 64. Если существует такая интегрируемая и неотрицательная функция $\varrho_{X,Y}(x, y)$, что

$$F_{X,Y}(x, y) = \iint\limits_{(-\infty,x]\times(-\infty,y]} \varrho_{X,Y}(x, y)\, dx\, dy,$$

то говорят, что совместное распределение случайных величин X, Y **абсолютно непрерывно**. Функцию $\varrho_{X,Y}$ называют **плотностью** совместного распределения случайных величин X, Y.

Ясно, что

$$P(a < X \le b,\ c < Y \le d) = \mu_{X,Y}((a, b] \times (c, d]) = \iint\limits_{(a,b]\times(c,d]} \varrho_{X,Y}(x, y)\, dx\, dy.$$

Можно доказать, что

$$P((X, Y) \in B) = \mu_{X,Y}(B) = \iint\limits_{B} \varrho_{X,Y}(x, y)\, dx\, dy.$$

для всякого множества B, для которого имеет смысл интеграл Римана в правой части. В каждой точке непрерывности плотности $\varrho_{X,Y}$ выполнено равенство

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = \varrho_{X,Y}(x, y).$$

Если известна плотность $\varrho_{X,Y}$ совместного распределения X и Y, то можно найти плотности распределения каждой из случайных величин. Например, для случайной величины X:

$$F_X(t) = P(X \le t, Y \in \mathbb{R}) = \int_{-\infty}^{t} \Big(\int_{-\infty}^{+\infty} \varrho_{X,Y}(x, y)\, dy \Big)\, dx$$

и, следовательно,

$$\varrho_X(x) = \int_{-\infty}^{+\infty} \varrho_{X,Y}(x, y)\, dy.$$

Если распределение каждой из случайных величин задается плотностью, то совместное распределение может не иметь плотность.

Напоминания про кратный интеграл Римана.

Теорема Фубини. Если функция $\varrho(x,y)$ интегрируема по $[a,b]\times[c,d]$ и интегрируема по каждой переменной в отдельности (например, так будет, если ϱ – непрерывная функция), то

$$\iint\limits_{[a,b]\times[c,d]} \varrho(x,y)\,dx\,dy = \int_a^b \Big(\int_c^d \varrho(x,y)\,dy\Big)\,dx = \int_c^d \Big(\int_a^b \varrho(x,y)\,dx\Big)\,dy.$$

Формула замены переменных

Пусть $\varphi\colon U\to V$, $x=x(u,v)$, $y=y(u,v)$ – диффеоморфизм открытых множеств U и V, т. е. φ – биекция, φ и φ^{-1} непрерывно дифференцируемые функции.

Пусть $A\subset U$, тогда для всякой интегрируемой по $\varphi(A)$ функции ϱ верно равенство

$$\iint\limits_{\varphi(A)} \varrho(x,y)\,dx\,dy = \iint\limits_{A} \varrho(x(u,v),y(u,v))\big|J_\varphi(u,v)\big|\,du\,dv,$$

где J_φ определитель матрицы Якоби

$$J_\varphi(u,v) = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{pmatrix}.$$

Интеграл Пуассона

Докажем, что

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

Имеет место следующая цепочка равенств:

$$\Big(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx\Big)^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-x^2/2}\,dx \cdot \int_{-\infty}^{+\infty} e^{-y^2/2}\,dy =$$
$$\frac{1}{2\pi}\iint\limits_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,dx\,dy = \frac{1}{2\pi}\int_0^{+\infty}\int_0^{2\pi} re^{-r^2/2}\,dr\,d\varphi = 1.$$

Теорема 65. *Пусть распределение X,Y задано плотностью $\varrho_{X,Y}$. Рассмотрим две случайные величины $\xi=f(X,Y),\eta=g(X,Y)$ и предположим, что отображение $T:(x,y)\to(f(x,y),g(x,y))$ удовлетворяет условиям теоремы о замене переменных в кратном интеграле Римана (например непрерывно дифференцируемо с невырожденным якобианом). Тогда $\varrho_{\xi,\eta}(u,v)=\varrho_{X,Y}\big(T^{-1}(u,v)\big)\cdot|J\big(T^{-1}(u,v)\big)|^{-1}$, где J — якобиан отображения T.*

Доказательство. Заметим, что

$$P((\xi,\eta)\in A) = P((X,Y)\in T^{-1}(A)) = \iint\limits_{T^{-1}(A)} \rho_{X,Y}(x,y)\,dxdy.$$

Сделаем замену в интеграле $u=f(x,y), v=g(x,y)$, т.е. $(x,y)=T^{-1}(u,v)$. Тогда последний интеграл равен

$$\iint\limits_{A} \rho_{X,Y}\big(T^{-1}(u,v)\big)\cdot|J\big(T^{-1}(u,v)\big)|^{-1}\,dudv,$$

что завершает доказательство. □

Равномерное распределение

Говорят, что вектор (ξ,η) *равномерно распределен* на множестве B, имеющем положительную площадь, если его распределение задано плотностью

$$\varrho(x,y) = \begin{cases} \frac{1}{|B|}, & x\in B,\\ 0, & x\notin B.\end{cases}$$

Аналогичным образом определяется равномерно распределенный вектор с любым конечным числом координат.

23. Независимые случайные величины.

Определение 66. Случайные величины X и Y называются **независимыми**, если

$$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

Предложение 67. *Случайные величины X и Y независимы тогда и только тогда, когда для произвольных $U, V \in \mathcal{B}(\mathbb{R})$ выполнено*

$$P(\{\omega\colon X(\omega) \in U,\, Y(\omega) \in V\}) = P(\{\omega\colon X(\omega) \in U\}) \cdot P(\{\omega\colon Y(\omega) \in V\}),$$

Доказательство. Если $V = (-\infty, y]$, то две меры $U \to \frac{\mu_{X,Y}(U\times V)}{\mu_Y(V)}$ и $U \to \mu_X(U)$ совпадают на всех лучах $(-\infty, x]$, т.е. имеют одинаковые функции распределения, а значит совпадают на всех борелевских множнствах U. Теперь для произвольного борелевского множества U меры $V \to \frac{\mu_{X,Y}(U\times V)}{\mu_X(U)}$ и $V \to \mu_Y(V)$ совпадают на всех лучах $(-\infty, y]$, а значит и на всех борелевских множествах V. □

Определение 68. Функция $f\colon \mathbb{R} \to \mathbb{R}$ называется борелевской, если $f^{-1}((-\infty, t]) \in \mathcal{B}(\mathbb{R})$ для каждого $t \in \mathbb{R}$.

Например, такими функциями будут все монотонные функции или все непрерывные.

Следствие 69. *Пусть X и Y независимы, а f, g — борелевские функции. Тогда $f(X)$ и $g(Y)$ также независимы.*

24. Характеризация независимости в терминах плотностей. Формула Свертки.

Предложение 70. *Пусть распределения X и Y заданы плотностями. Тогда независимость X и Y равносильна тому, что совместное распределение задано плотностью и эта плотность имеет вид*

$$\varrho_{X,Y}(x,y) = \varrho_X(x)\varrho_Y(y).$$

Доказательство. Если, X и Y независимы, то

$$F_{X,Y}(x,y) = F_X(x)\cdot F_Y(y) = \int_{-\infty}^{x} \rho_X(t)\,dt \cdot \int_{-\infty}^{y} \rho_Y(s)\,ds = \iint\limits_{(-\infty,x]\times(-\infty,y]} \rho_X(t)\rho_Y(s)\,dtds.$$

Обратно,

$$F_{X,Y}(x,y) = \iint\limits_{(-\infty,x]\times(-\infty,y]} \rho_X(t)\rho_Y(s)\,dtds = \int_{-\infty}^{x} \rho_X(t)\,dt \cdot \int_{-\infty}^{y} \rho_Y(s)\,ds = F_X(x)\cdot F_Y(y).$$

□

Следствие 71 (Формула свертки)**.** *Предположим, что X и Y независимы и их распределения заданы плотностями ϱ_X и ϱ_Y. Тогда распределение суммы $Z = X + Y$ задано плотностью*

$$\varrho_Z(z) = \int_{-\infty}^{+\infty} \varrho_X(t)\varrho_Y(z-t)\,dt.$$

Доказательство. По определению $F_Z(t) = P(\{\omega\colon X(\omega) + Y(\omega) \le t\})$. С другой стороны, эта вероятность выражается через интеграл

$$\iint\limits_{x+y\le t} \varrho_X(x)\varrho_Y(y)\,dx\,dy.$$

Переходя к новым переменным $u = x+y$, $v = x$, и, применяя теорему Фубини, преобразуем этот интеграл:

$$\int_{-\infty}^{t}\Big(\int_{-\infty}^{+\infty}\varrho_X(v)\varrho_Y(u-v)\,dv\Big)\,du.$$

Следовательно, распределение Z имеет плотность требуемого вида. $\square$

25. Математическое ожидание: ограниченные и неотрицательные случайные величины.

Математическое ожидание в общем случае построим в четыре шага: сначала определим на простых случанйых величинах, затем определим на ограниченных, затем на неотрицательных, и, наконец, определим для общих случайных величин.

Случайные величины с конечным числом значений будем называть **простыми**. Пусть X — простая случайная величина на $(\Omega, \mathcal{A}, P)$, принимающая конечное число значений $\{x_1, \ldots, x_N\}$. Тогда по определению полагаем, что $\mathbb{E}X := \sum\limits_{j=1}^{N} x_j P(X = x_j)$. Ометим, что при построении математического ожидания для случайных величин на дискретных вероятностных пространствах и при доказательстве свойств математического ожидания для таких случайных величин использовалось лишь то, что случайные величины принимают не более чем счетное число значений. Таким образом, для простых случайных величин X, Y на общих вероятностных пространствах справедливы все те же самые свойства математического ожидания:

1) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$

2) если $X \geq 0$ п.н., то $\mathbb{E}X \geq 0$, в частности, если $X \geq Y$ п.н., то $\mathbb{E}X \geq \mathbb{E}Y$

3) если $X \geq 0$ почти наверное и $\mathbb{E}X = 0$, то $X = 0$ почти наверное

4) $|\mathbb{E}X| \leq \mathbb{E}|X|$

5) если случайные величины X и Y независимы и существуют математические ожидания $\mathbb{E}X$ и $\mathbb{E}Y$, то $\mathbb{E}[X \cdot Y] = [\mathbb{E}X] \cdot [\mathbb{E}Y]$

6) $\mathbb{E}\varphi(X) = \sum\limits_{j=1}^{N} \varphi(x_j) P(X = x_j)$.

Предположим теперь, что X — произвольная ограниченаая случайная величина, т.е. существует такое число $R > 0$, что $|X(\omega)| < R\,\forall\omega \in \Omega$.

Лемма 72. *Пусть X — ограниченаая случайная величина. Тогда найдется последовательность простых случайных величин X_n, равномерно сходящаяся к X.*

Доказательство. Пусть $|X(\omega)| < R$ для каждого $\omega \in \Omega$. Рассмотрим случайную величину

$$X_n(\omega) := \sum_{k=1}^{n}(-R + \tfrac{2R}{n}(k-1))I_{\{\omega\colon -R+\frac{2R}{n}(k-1)\leq X(\omega)<-R+\frac{2R}{n}k\}}.$$

Возьмем теперь произвольный элемент $\omega_0 \in \Omega$. Тогда, т.к. $|X(\omega_0)| < R$, то найдется число $k_0 \in \{1, \ldots, n\}$, для которого $-R + \frac{2R}{n}(k_0 - 1) \leq X(\omega_0) < -R + \frac{2R}{n}k_0$. Таким образом $|X(\omega_0) - X_n(\omega_0)| \leq \frac{2R}{n}$. Лемма доказана. $\square$

Определение 73. Пусть X — ограниченная случайная величина, тогда ее **математическим ожиданием** называют предел $\lim\limits_{n\to\infty}\mathbb{E}X_n$ математических ожиданий произвольной последовательности простых случайных величин X_n, равномерно сходящейся к X.

Предложение 74 (корректность предыдущего определния)**.** *Предыдущее определние корректно в том смысле, что для произвольной ограниченной случайной величины X и для* *произвольной последовательности простых случайных величин X_n, равномерно сходящейся к X, существует предел $\lim\limits_{n\to\infty} \mathbb{E}X_n$. Кроме того, для произвольной другой последовательности простых случайных величин Y_n, равномерно сходящейся к X, выполнено $\lim\limits_{n\to\infty} \mathbb{E}Y_n = \lim\limits_{n\to\infty} \mathbb{E}X_n$.*

Доказательство. Заметим что, $|\mathbb{E}X_n - \mathbb{E}X_k| \leq \mathbb{E}|X_n - X_k| \leq \sup\limits_{\omega\in\Omega} |X_n(\omega) - X_k(\omega)|$. Тем самым, последовательность $\{\mathbb{E}X_n\}$ фундаментальна, а значит сходится. Если Y_n другая последовательность простых случайных величин, равномерно сходящаяся к X, то последовательность Z_m, для которой $Z_{2k-1} := X_k$, $Z_{2k} := Y_k$, также образует последовательность простых случайных величин, равномерно сходящуюся к X. Тогда последовательность чисел $\mathbb{E}Z_m$ сходится, а значит $\lim\limits_{n\to\infty} \mathbb{E}Y_n = \lim\limits_{n\to\infty} \mathbb{E}X_n$, как пределы двух подпоследователностей сходящейся последовательности чисел. ☐

Заметим, что для простых случайных величин предыдущее определение дает то же самое значение, что и исходное определение, т.к. к простой случайной величине X равномерно сходится последовательность $X_n = X$.

Предложение 75. *Для ограниченных случайных величин X, Y выполнены свойства:*
1) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$*;*
2) *если $X \geq 0$ п.н., то $\mathbb{E}X \geq 0$, в частности, если $X \geq Y$ п.н., то $\mathbb{E}X \geq \mathbb{E}Y$;*
3) *если $X = 0$ п.н., то $\mathbb{E}X = 0$.*

Доказательство. 1) Если $X_n \rightrightarrows X$, $Y_n \rightrightarrows Y$, X_n, Y_n — простые, то $\alpha X_n + \beta Y_n \rightrightarrows \alpha X + \beta Y$. Отсюда $\mathbb{E}(\alpha X + \beta Y) = \lim\limits_{n\to\infty} \mathbb{E}(\alpha X_n + \beta Y_n) = \lim\limits_{n\to\infty} (\alpha\mathbb{E}X_n + \beta\mathbb{E}Y_n) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$.

2) Пусть сначала $X(\omega) \geq 0$ для каждой $\omega \in \Omega$. Пусть X_n — последовательность простых случайных величн, равномерно сходящаяся к $\sqrt{X}$. Тогда $X_n^2 \rightrightarrows X$, откуда $\mathbb{E}X = \lim\limits_{n\to\infty} \mathbb{E}X_n^2 \geq 0$.

Теперь, для произвольного $X \geq 0$ п.н., выполнено $\mathbb{E}X = \mathbb{E}[XI_{\{X\geq 0\}}] + \mathbb{E}[XI_{\{X<0\}}]$. Покажем, что $\mathbb{E}[XI_{\{X<0\}}] = 0$. По доказанному $\mathbb{E}[-XI_{\{X<0\}}] \geq 0$ и $\mathbb{E}[(M + X)I_{\{X<0\}}] \geq 0$, где $M = \sup\limits_{\omega\in\Omega} |X(\omega)|$. Таким образом, $0 \leq \mathbb{E}[-XI_{\{X<0\}}] \leq MP(X < 0) = 0$.

В случае, когда $X \geq Y$ п.н., получаем, что $X - Y \geq 0$ п.н. и $\mathbb{E}X - \mathbb{E}Y = \mathbb{E}(X - Y) \geq 0$.

3) Если $X = 0$ п.н., то $X \geq 0$ п.н. и $-X \geq 0$ п.н., поэтому, по предыдущему свойству $\mathbb{E}X \geq 0$ и $-\mathbb{E}X = \mathbb{E}[-X] \geq 0$. ☐

Определение 76. Пусть теперь $X \geq 0$ — случайная величина. Скажем, что у нее есть конечное математическое ожидание, если конечен следующий супремум

$$\mathbb{E}X := \sup\{\mathbb{E}U\colon 0 \leq U \leq X;\ U\text{— ограниченная}\}.$$

Отметим, что данное определение согласовано с определнием ожидания для ограниченных случайных величин (т.е. для ограниченных неотрицательных случайных величин получится то же самое значение математического ожидания).

Предложение 77. *Для неотрицательных случайных величин X, Y выполнены свойства:*
1) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$*, если $\alpha, \beta \geq 0$;*
2) *если $X \geq Y \geq 0$, то $\mathbb{E}X \geq \mathbb{E}Y$, в частности, если X имеет конечное математическое ожидание, то и Y также имеет конечное математическое ожидание;*
3) *если $X = 0$ п.н., то $\mathbb{E}X = 0$.*

Доказательство. 1) Достаточно доказать утверждение при $\alpha = \beta = 1$. Если $0 \leq U \leq X$, $0 \leq V \leq Y$, U, V — ограниченные, то $U+V \leq X+Y$, откуда $\mathbb{E}X+\mathbb{E}Y \leq \mathbb{E}(X+Y)$. Наоборот, пусть $0 \leq Z \leq X + Y$, Z — ограниченная случайная величина. Пусть $U := \min(X, Z)$, $V := Z - U$. Тогда $0 \leq U \leq X$, U — ограниченная, $V = (Z - X)I_{\{X<Z\}} \leq X + Y - X = Y$,

V — ограниченная. Таким образом, $\mathbb{E}Z = \mathbb{E}(U+V) = \mathbb{E}U + \mathbb{E}V \leq \mathbb{E}X + \mathbb{E}Y$. А значит аналогичная оценка верна и для $\mathbb{E}(X+Y)$.

2) Следует из определния.

3) Произвольная ограниченная случайная величина U, $0 \leq U \leq X$, также обращается в нуль п.н. Поэтому $\mathbb{E}U = 0$. □

26. Математическое ожидание: общий случай.

Перейдем теперь к построению математического ожидания для произвольной случайной величины X.

Определение 78. Пусть X — случайная величина и пусть $X^+ := \max\{X, 0\} \geq 0$, $X^- := \max\{-X, 0\} \geq 0$ (в частности, $X = X^+ - X^-$). Скажем, что X обладает математическим ожиданием, если X^+ и X^- имеют конечные математические ожидания. В этом случае определим математическое ожидание X равенством: $\mathbb{E}X := \mathbb{E}X^+ - \mathbb{E}X^-$.

Из определния в частности следует, что для случайной величины X, обладающей математическим ожиданием, $|X| = X^+ + X^-$ также будет иметь конечное математическое ожидание. Наоборот, если $|X|$ обладает конечным математическим ожиданием, то $X^+ \leq |X|$, $X^- \leq |X|$, поэтому и X^+ и X^- имеют конечные математические ожидания, а значит и у X определено математическое ожидание.

Заметим, что предыдущее определние корректно в следующем смысле. Предположим, что $U \geq 0, V \geq 0$ — случайные величины с конечными математическими ожиданиями, причем $X = U - V$. Тогда для X определено математическое ожидание и $\mathbb{E}X = \mathbb{E}U - \mathbb{E}V$. Действительно, в этом случае $|X| \leq U + V$, а значит определено математическое ожидание X. Кроме того, $X^+ - X^- = U - V$, т.е. $X^+ + V = U + X^-$, откуда $\mathbb{E}(X^+ + V) = \mathbb{E}(U + X^-)$. В силу того, что все функции U, V, X^+, X^- неотрицательны, получаем, что $\mathbb{E}X^+ + \mathbb{E}V = \mathbb{E}U + \mathbb{E}X^-$, т.е. $\mathbb{E}X^+ - \mathbb{E}X^- = \mathbb{E}U - \mathbb{E}V$.

Предложение 79. *Для случайных величин X, Y, обладающих математическим ожиданием, выполнены свойства:*

1) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$*;*

2) *если $X \geq 0$ почти наверное, то $\mathbb{E}X \geq 0$, в частности, при $X \geq Y$ п.н., $\mathbb{E}X \geq \mathbb{E}Y$;*

3) *если $X = 0$ п.н., то $\mathbb{E}X = 0$;*

4) *(Неравенство Чебышева) если $X \geq 0$ п.н., то $P(X \geq t) \leq t^{-1}\mathbb{E}X$, $\forall t \geq 0$;*

5) *если $X \geq 0$ почти наверное и $\mathbb{E}X = 0$, то $X = 0$ почти наверное;*

6) $|\mathbb{E}X| \leq \mathbb{E}|X|$.

Доказательство. 1) Заметим, что $|\alpha X + \beta Y| \leq |\alpha||X| + |\beta||Y|$, а значит определено математическое ожидание $\alpha X + \beta Y$. Кроме того, $\mathbb{E}[-X] = -\mathbb{E}X$, т.к. для произвольного представления $X = U - V$, $U, V \geq 0$, выполнено $-X = V - U$, откуда $\mathbb{E}[-X] = \mathbb{E}V - \mathbb{E}U = -(\mathbb{E}U - \mathbb{E}V) = -\mathbb{E}X$. Тогда достаточно доказать линейность только в случае $\alpha, \beta \geq 0$. В этом случае $\alpha X + \beta Y = \alpha X^+ + \beta Y^+ - (\alpha X^- + \beta Y^-)$, причем $\alpha X^+ + \beta Y^+ \geq 0$ и $\alpha X^- + \beta Y^- \geq 0$. Поэтому

$$\begin{aligned}\mathbb{E}(\alpha X + \beta Y) = \mathbb{E}(\alpha X^+ + \beta Y^+) - \mathbb{E}(\alpha X^- + \beta Y^-) = \alpha\mathbb{E}X^+ + \beta\mathbb{E}Y^+ - \alpha\mathbb{E}X^- - \beta\mathbb{E}Y^- \\ = \alpha(\mathbb{E}X^+ - \mathbb{E}X^-) + \beta(\mathbb{E}Y^+ - \mathbb{E}\beta Y^-) = \alpha\mathbb{E}X + \beta\mathbb{E}Y.\end{aligned}$$

2) В этом случае $X^- = 0$ п.н., а значит и $\mathbb{E}X^- = 0$. Таким образом, $\mathbb{E}X = \mathbb{E}X^+ \geq 0$.

3) следует из предыдущего рассмотрением X и $-X$.

4) Достаточно заметить, что $t \cdot I_{\{X \geq t\}} \leq X$ п.н. и воспользоваться свойством 2).

5) Заметим, что по неравенствоу Чебышева $P(X \geq k^{-1}) \leq k\mathbb{E}X = 0$, откуда получаем, что $P(X > 0) = P(\bigcup_{k=1}^{\infty}\{X \geq k^{-1}\}) \leq \sum_{k=1}^{\infty} P(X \geq k^{-1}) = 0$.

6) Заметим, что $-|X| \leq X \leq |X|$, откуда по свойствам 2) и 1) получаем, неравенства $-\mathbb{E}|X| \leq \mathbb{E}X \leq \mathbb{E}|X|$. □

Лемма 80. *Пусть случайная величина $X \geq 0$ и пусть $A_n \subset A_{n+1}$, $A_n \in \mathcal{A}$, причем $\bigcup\limits_{n=1}^{\infty} A_n = \Omega$. Тогда X имеет конечное математическое ожидание тогда и только тогда, когда $\sup\limits_n \mathbb{E}[XI_{A_n}] := M < \infty$. При этом $\mathbb{E}X = \lim\limits_{n\to\infty} \mathbb{E}[XI_{A_n}] = M$.*

Доказательство. Пусть U — произвольная ограниченная случайная величина, причем $0 \leq U \leq X$. Тогда $P(\Omega \setminus A_n) \to 0$ (из-за свойства непрерывности вероятностной меры) и $\mathbb{E}[UI_{\Omega\setminus A_n}] \leq [\max U]P(\Omega \setminus A_n) \to 0$. Отсюда $\mathbb{E}U = \lim\limits_{n\to\infty} \mathbb{E}[UI_{A_n}] \leq M$, т.к. имеет место оценка $\mathbb{E}[UI_{A_n}] \leq \mathbb{E}[XI_{A_n}] \leq M$. Значит $\mathbb{E}X \leq M$. С другой стороны $X \geq XI_{A_n}$, откуда $\mathbb{E}X \geq \mathbb{E}[XI_{A_n}]$. □

Предложение 81. *Пусть X — случайная величина, распределение которой имеет плотность ϱ_X. Пусть задана непрерывная функция f. Тогда математическое ожидание $\mathbb{E}f(X)$ существует тогда и только тогда, когда сходится несобственный интеграл*

$$\int_{-\infty}^{+\infty} |f(x)|\varrho_X(x)\, dx.$$

Более того, в случае сходимости

$$\mathbb{E}f(X) = \int_{-\infty}^{+\infty} f(x)\varrho_X(x)\, dx.$$

Доказательство. Пусть $R > 0$. Для непрерывной функции f на отрезке $[-R, R]$ найдется последовательность ступенчатых функций g_n, равномерно сходящаяся к f на $(-R, R]$. Функции g_n имеют вид $g_n = \sum\limits_{j=1}^{N_n} c_j I_{(a_j,b_j]}$, где $\{(a_j, b_j]\}$ — разбиение промежутка $(-R, R]$. Заметим, что

$$\mathbb{E}g_n(X) = \sum_{j=1}^{N_n} c_j \mathbb{E}I_{\{a_j<X\leq b_j\}} = \sum_{j=1}^{N_n} c_j P(X \in (a_j, b_j]) = \sum_{j=1}^{N_n} c_j \int_{a_j}^{b_j} \varrho_X(x)\, dx = \int_{-R}^{R} g_n(x)\varrho_X(x)\, dx.$$

Заметим, что

$$|\mathbb{E}[f(X)I_{\{-R<X\leq R\}}] - \mathbb{E}g_n(X)| \leq \mathbb{E}[|f(X) - g_n(X)|I_{\{-R<X\leq R\}}] \leq \sup_{x\in(-R,R]} |f(x) - g_n(x)| \to 0.$$

Аналогично $\int_{-R}^{R} g_n(x)\varrho_X(x)\, dx \to \int_{-R}^{R} f(x)\varrho_X(x)\, dx$, откуда получаем равенство

$$\mathbb{E}[f(X)I_{\{-R\leq X\leq R\}}] = \int_{-R}^{R} f(x)\varrho_X(x)\, dx.$$

Достаточно доказать исходное утверждение для неотрицательных функций f, для которых оно теперь следует из леммы 80. □

Предложение 82. *Пусть X, Y — независимые случайные величины, имеющие математическое ожидание. Тогда $X \cdot Y$ также обладает математическим ожиданием и $\mathbb{E}[X \cdot Y] = [\mathbb{E}X] \cdot [\mathbb{E}Y]$.*

Доказательство. Заметим, что

$$X \cdot Y = (X^+ - X^-)(Y^+ - Y^-) = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+.$$

Таким образом, достаточно доказать утверждение только для неотрицательных X, Y. Если X, Y ограничены, $|X| < R$, $|Y| < R$, то рассмотрим

$$X_n(\omega) := \sum_{k=1}^{n} (-R + \tfrac{2R}{n}(k-1)) I_{\{\omega\colon -R+\frac{2R}{n}(k-1) \le X(\omega) < -R + \frac{2R}{n}k\}},$$

$$Y_n(\omega) := \sum_{k=1}^{n} (-R + \tfrac{2R}{n}(k-1)) I_{\{\omega\colon -R+\frac{2R}{n}(k-1) \le Y(\omega) < -R + \frac{2R}{n}k\}}.$$

Т.к. X_n имеет вид $f_n(X)$, а Y_n имеет вид $f_n(Y)$ для некоторой функции f_n, то X_n и Y_n также независимы. Кроме того, $X_n \rightrightarrows X$, $Y_n \rightrightarrows Y$. Поэтому

$$[\mathbb{E}X] \cdot [\mathbb{E}Y] = \lim_{n\to\infty} [\mathbb{E}X_n] \cdot [\mathbb{E}Y_n] = \lim_{n\to\infty} \mathbb{E}[X_n \cdot Y_n] = \mathbb{E}[X \cdot Y].$$

Для общих неотрицательных независимых X и Y, рассмотрим независимые ограниченные случайные величины $XI_{\{|X|<R\}}$ и $YI_{\{|Y|<R\}}$. Тогда

$$\mathbb{E}[XI_{\{|X|<R\}} \cdot YI_{\{|Y|<R\}}] = \mathbb{E}[XI_{\{|X|<R\}}] \cdot \mathbb{E}[YI_{\{|Y|<R\}}].$$

Утверждение теперь следует из леммы 80. □

Определения дисперсии, ковариация и коэффициента корреляции совпадают с соответствующими определениями для величин на дискретных пространствах. Все свойства обосновываются аналогично.

В качестве примера рассмотрим вычисление математического ожидания и дисперсии нормально распределенной величины $X \sim \mathcal{N}(a, \sigma^2)$:

$$\mathbb{E}X = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx = a + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2\sigma^2}}\, dy = a,$$

$$\mathbb{D}X = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}}\, dy = \sigma^2,$$

где последнее равенство обосновывается так

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} -y \Big(e^{-\frac{y^2}{2}}\Big)'\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\, dy = 1.$$

27. Сходимости случайных величин.

Определение 83.

Последовательность случайных величин X_n сходится к случайной величине X

1) **почти наверное**, если $P\Big(\lim\limits_{n\to\infty} X_n = X\Big) = 1$ (пишут $X_n \xrightarrow{\text{п.н.}} X$);

2) **по вероятности**, если $\forall \varepsilon > 0\ \lim\limits_{n\to\infty} P(|X_n - X| \ge \varepsilon) = 0$ (пишут $X_n \xrightarrow{P} X$);

3) **по распределению**, если $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$ в каждой точке x, в которой непрерывна функция F_X (пишут $X_n \xrightarrow{d} X$).

Теорема 84. *Если $X_n \xrightarrow{\text{п.н.}} X$, то $X_n \xrightarrow{P} X$.*

Доказательство. Заметим, что в силу сходимости п.н. для произвольного $\varepsilon > 0$ выполнено

$$P\Big(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| \ge \varepsilon\}\Big) = 0.$$

Из-за вложенности, это означает, что $\lim\limits_{N\to\infty} P\big(\bigcup\limits_{n=N}^{\infty}\{|X_n - X| \geq \varepsilon\}\big) = 0$. Но в тоже время $P(|X_N - X| \geq \varepsilon) \leq P\big(\bigcup\limits_{n=N}^{\infty}\{|X_n - X| \geq \varepsilon\}\big)$. Значит $\lim\limits_{N\to\infty} P(|X_N - X| \geq \varepsilon) = 0$ и теорема доказана. □

Замечание 85. Обратное следствие в общем случае не выполнено, что показывает пример Рисса, который подробнее будет обсуждаться на семинарах.

28. Закон больших чисел.

В общем случае также справедливо неравенстов Чебышева.

Предложение 86. *Пусть у неотрицательной случайной величины X определено математическое ожидание. Тогда $P(X \geq t) \leq t^{-1}\mathbb{E}X$ для каждого $t > 0$.*

Пусть у случайной величины X конечный второй момент, т.е. $\mathbb{E}X^2 < \infty$. Тогда $P(|X - \mathbb{E}X| \geq \varepsilon) \leq \varepsilon^{-2}\mathbb{D}X$.

Доказательство. Заметим, что $t \cdot I_{\{X\geq t\}} \leq X$, поэтому

$$tP(X \geq t) = \mathbb{E}[t \cdot I_{\{X\geq t\}}] \leq \mathbb{E}X.$$

Второе неравенство обосновывается рассмотрением случайной величины $|X - \mathbb{E}X|^2$ и применением первого неравенства. □

Следствие 87. (Закон больших чисел в слабой форме) *Пусть $\{X_n\}_n$ — последовательность таких независимых случайных величин, что $\mathbb{E}X_n^2 < \infty$. Обозначим $\mathbb{E}X_n = a_n$ и $\mathbb{D}X_n = \sigma_n^2$. Если*

$$\lim_{n\to\infty} \frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2} = 0,$$

то для всякого $\varepsilon > 0$

$$\lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - \frac{a_1 + \ldots + a_n}{n}\Big| \geq \varepsilon\Big) = 0$$

Доказательство. Для всякого $\varepsilon > 0$ имеет место неравенство

$$P\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - \frac{a_1 + \ldots + a_n}{n}\Big| \geq \varepsilon\Big) \leq \frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2\varepsilon^2},$$

из которого немедленно следует доказываемое утверждение. □

Если $\mathbb{E}X_1 = \mathbb{E}X_2 = \ldots = a$ и $\mathbb{D}X_1 = \mathbb{D}X_2 = \ldots = \sigma^2$, то

$$\frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2} = \frac{\sigma^2}{n} \to 0$$

и выполняется утверждение следствия:

$$P\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - a\Big| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Замечание 88. Если в теореме выше случайные величины имеют одинаковое математическое ожидание $\mathbb{E}X_n = a_n = a$ (например, они одинаково распределены), то теорема утверждает, что $\frac{X_1+\ldots+X_n}{n} \xrightarrow{P} a$.

Теорема 89 (уЗБЧ Колмогорова)**.** *Пусть $\{X_n\}_n$ — последовательность независимых одинаково распределнных случайных величин, имеющих математическое ожидание и пусть $\mathbb{E}X_1 = a$. Тогда*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{п.н.} a.$$

Мы докажем некоторый промежуточный результат, а именно.

Теорема 90 (уЗБЧ). *Пусть $\{X_n\}_n$ — последовательность независимых одинаково распределнных случайных величин, причем $\mathbb{E}[X_n^4] = \mathbb{E}[X_1^4] = C < \infty$. Тогда*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{п.н.}} a = \mathbb{E}X_1.$$

Замечание 91. Заметим, что $(\mathbb{E}X)^2 \leq \mathbb{E}[X^2]$ в силу того, что $\mathbb{D}X \geq 0$. Поэтому в предыдущей теореме математическое ожидание $\mathbb{E}X_1$ определено.

Сначала получим удобное достаточное условие сходимости п.н.

Лемма 92 (Бореля–Кантелли). *Пусть A_n — последовательность событий. Если сходится ряд $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$, то $P\big(\bigcap\limits_{N=1}^{\infty} \bigcup\limits_{n=N}^{\infty} A_n\big) = 0$.*

Событие $A = \bigcap\limits_{N=1}^{\infty} \bigcup\limits_{n=N}^{\infty} A_n$ иногда называют событием таких исходов, для которых выполняется бесконечно много событий из семейства $\{A_n\}_n$.

Доказательство. Заметим, что в силу вложенности, $P\big(\bigcap\limits_{N=1}^{\infty} \bigcup\limits_{n=N}^{\infty} A_n\big) = \lim\limits_{N \to +\infty} P\big(\bigcup\limits_{n=N}^{\infty} A_n\big)$, но, по свойству субаддитивности вероятностной меры, $P\big(\bigcup\limits_{n=N}^{\infty} A_n\big) \leq \sum\limits_{n=N}^{\infty} P(A_n)$, что является остатком сходящегося ряда, а потому стремится к нулю при $N \to +\infty$. □

Предложение 93 (достаточное условие сходимости п.н.). *Пусть $X, X_n, n = 1, 2, \ldots$ — случайные величины, причем при каждом $\delta > 0$ сходится ряд $\sum\limits_{n=1}^{\infty} P(|X_n - X| \geq \delta) < \infty$. Тогда $X_n \xrightarrow{\text{п.н.}} X$.*

Доказательство. Нужно проверить, что $P($посл-ть X_n не сходится к $X) = 0$. Данное событие можно переписать в таком виде

$$\{\text{посл-ть } X_n \text{ не сходится к } X\} = \{\exists k \in \mathbb{N}\ \forall N \in \mathbb{N}\ \exists n \geq N\colon |X_n - X| \geq \tfrac{1}{k}\} = \\ = \bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| \geq \tfrac{1}{k}\}.$$

По лемме Бореля-Кантелли $P(\bigcap\limits_{N=1}^{\infty} \bigcup\limits_{n=N}^{\infty} \{|X_n - X| \geq \frac{1}{k}\}) = 0$. Тогда

$$P(\{\text{посл-ть } X_n \text{ не сходится к } X\}) \leq \sum_{k=1}^{\infty} P\Big(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| \geq \tfrac{1}{k}\}\Big) = 0.$$

Утверждение доказано. □

Перейдем к доказательству уЗБЧ:

Нам достаточно проверить при каждом $\delta > 0$ сходимость ряда $\sum\limits_{n=1}^{\infty} P\big(|\frac{X_1+\ldots+X_n}{n} - a| \geq \delta\big)$.

По неравенству Чебышева:

$$P\big(|\tfrac{X_1+\ldots+X_n}{n} - a| \geq \delta\big) = P\big(|\tfrac{X_1+\ldots+X_n}{n} - a|^4 \geq \delta^4\big) \leq \delta^{-4}\mathbb{E}|\tfrac{X_1+\ldots+X_n}{n} - a|^4.$$

Пусть $Y_j = X_j - a$, тогда Y_j по прежнему независимы и $\mathbb{E}Y_j = 0$. Заметим, что

$$\mathbb{E}|\tfrac{X_1+\ldots+X_n}{n} - a|^4 = n^{-4}\mathbb{E}(Y_1 + \ldots + Y_n)^4 = n^{-4} \sum_{i,j,k,l=1}^{n} \mathbb{E}[Y_i Y_j Y_k Y_l].$$

В сумме есть слагаемые разных видов: 1) все индексы i, j, k, l — совпадают; 2) какая-то парар индекосв совпадает между собой, другая пара индексов также совпадает между собой;

3) есть индекс, который не совпадает ни с каким другим. Заметим, что в случае 3) из-за независимости имеет место равенство $\mathbb{E}[Y_iY_jY_kY_l]=\mathbb{E}[Y_i]\cdot\mathbb{E}[Y_jY_kY_l]=0$ (в предположении, что индекс i не повторяется среди оставшихся). Таким образом

$$\sum_{i,j,k,l=1}^{n}\mathbb{E}[Y_iY_jY_kY_l]=\sum_{i=1}^{n}\mathbb{E}Y_i^4+C_4^2\sum_{i\neq j}\mathbb{E}Y_i^2\cdot\mathbb{E}Y_j^2=n\mathbb{E}Y_1^4+6n(n-1)(\mathbb{E}Y_1^2)^2\le(n+6n(n-1))\mathbb{E}Y_1^4.$$

Остается заметить, что ряд $\sum\limits_{n=1}^{\infty}\frac{n+6n(n-1)}{n^4}$ сходится.

29. Сходимость по распределению.

Перейдем теперь к обсуждению сходимости по распределению. Сначала докажем следующее техническое утверждение.

Лемма 94. *Пусть $\{X_n\}_{n=0}^{\infty}$ — последовательность случайных величин. Пусть $\mathcal{F}:=\{f\}$ и $\mathcal{G}:=\{g\}$ две системы функций на $\mathbb{R}$, причем известно, что для каждой функции $g\in\mathcal{G}$ и для каждого $\varepsilon>0$ существует такая функция $f_\varepsilon\in\mathcal{F}$, что $\mathbb{E}|g(X_n)-f_\varepsilon(X_n)|\le\varepsilon$ для каждого $n\in\mathbb{N}\cup\{0\}$. Тогда, если $\mathbb{E}f(X_n)\to\mathbb{E}f(X_0)$ для каждой $f\in\mathcal{F}$ то верна сходимость $\mathbb{E}g(X_n)\to\mathbb{E}g(X_0)$ для каждой $g\in\mathcal{G}$.*

Доказательство. Для каждой $g\in\mathcal{G}$ и для каждого $\varepsilon>0$ найдется $f_\varepsilon\in\mathcal{F}$, для которой $\mathbb{E}|g(X_n)-f_\varepsilon(X_n)|\le\varepsilon$ для каждого $n\in\mathbb{N}\cup\{0\}$. Кроме того, $\mathbb{E}f_\varepsilon(X_n)\to\mathbb{E}f_\varepsilon(X_0)$, т.е. найдется такой номер $n_0\in\mathbb{N}$, что $|\mathbb{E}f_\varepsilon(X_n)-\mathbb{E}f_\varepsilon(X_0)|<\varepsilon$ при $n>n_0$. Таким образом

$$|\mathbb{E}g(X_n)-\mathbb{E}g(X_0)|\le\mathbb{E}|g(X_n)-f_\varepsilon(X_n)|+|\mathbb{E}f_\varepsilon(X_n)-\mathbb{E}f_\varepsilon(X_0)|+\mathbb{E}|f_\varepsilon(X_0)-g(X_0)|\le3\varepsilon$$

при $n>n_0$. Утверждение доказано. □

Получим теперь эквивалентное описаное сходимости по распределению.

Теорема 95. *Последовательность случайных величин X_n сходится по распределению к случайной величине X тогда и только тогда, когда для каждой непрерывной и ограниченной функции $g\colon\mathbb{R}\to\mathbb{R}$ выполнено $\lim\limits_{n\to\infty}\mathbb{E}g(X_n)=\mathbb{E}g(X)$.*

Доказательство. Пусть t – точка непрерывности $F_X(t)$. Заметим, что $F_X(t)=\mathbb{E}I_{(-\infty,t]}(X)$. Для всякого $\delta>0$ определим непрерывные функции

$$g_\delta(x)=\begin{cases}1, & x<t-\delta,\\ \delta^{-1}(t-x), & t-\delta\le x\le t,\\ 0, & x>t.\end{cases}\qquad h_\delta(x)=\begin{cases}1, & x<t\\ \delta^{-1}(t+\delta-x), & t\le x\le t+\delta,\\ 0, & x>t+\delta.\end{cases}$$

Ясно, что

$$I_{(-\infty,t-\delta]}\le g_\delta(t)\le I_{(-\infty,t]}\le h_\delta(t)\le I_{(-\infty,t+\delta]}.$$

Следовательно,

$$\mathbb{E}g_\delta(X_n)\le F_{X_n}(t)\le\mathbb{E}h_\delta(X_n).$$

Устремляя $n\to\infty$ получаем

$$\mathbb{E}g_\delta(X)\le\varliminf_{n\to\infty}F_{X_n}(t)\le\varlimsup_{n\to\infty}F_{X_n}(t)\le\mathbb{E}h_\delta(X).$$

Теперь заметим, что

$$F_X(t-\delta)\le\mathbb{E}g_\delta(X),\quad\mathbb{E}h_\delta(X)\le F_X(t+\delta).$$

Устремляя $\delta\to0$ приходим к равенству $\lim\limits_{n\to\infty}F_{X_n}(t)=F_X(t)$.

В обратную сторону, для $X_0=X$ мы хотим показать сходимость $\mathbb{E}g(X_n)\to\mathbb{E}g(X_0)$ для системы $\mathcal{G}$ всех непрепрерывных ограниченных функций на $\mathbb{R}$. Нам известно, что $\mathbb{E}I_{(-\infty,t]}(X_n)=F_{X_n}(t)\to F_X(t)=\mathbb{E}I_{(-\infty,t]}(X)$ для каждой t — точки непрерывности функции F_X. В силу линейности предела верна сходимость $\mathbb{E}f(X_n)\to\mathbb{E}f(X)$ для всех функций

f вида $f(x) = \sum_{k=1}^{m} c_j I_{(a_j,b_j]}(x)$, где a_j, b_j — точки непрерывности функции F_X. Обозначим систему таких функций символом $\mathcal{F}$. По доказанной выше лемме нам достаточно показать, что для каждой непрерывной ограниченной функции g и для каждого $\varepsilon > 0$ найдется функция $f_\varepsilon \in \mathcal{F}$, для которой $\mathbb{E}|g(X_n) - f_\varepsilon(X_n)| \le \varepsilon$ для каждого $n \in \mathbb{N} \cup \{0\}$ $(X_0 = X)$.

Пусть g — произвольная непрерывная ограниченная функция на $\mathbb{R}$ и пусть $\varepsilon > 0$ фиксировано. Т.к. $\lim\limits_{x\to-\infty} F_X(x) = 0$ и $\lim\limits_{x\to+\infty} F_X(x) = 1$, то найдется $A > 0$, для которого $F_X(-A) < \varepsilon$ и $1 - F_X(A) < \varepsilon$. Не ограничивая общности, можно считать, что A и $-A$ — точки непрерывности F_X (у монотонной функции не более чем счетное число точек разрыва). Тогда для некоторого n_0 при каждом $n > n_0$ выполнено $F_{X_n}(-A) < 2\varepsilon$ и $1 - F_{X_n}(A) < 2\varepsilon$. Увеличив A, можно считать, что последние неравенства справедливы при всех n. Таким образом,

$$\mathbb{E}\Big|I_{\{-A<X_n\le A\}} g(X_n) - g(X_n)\Big| \le 4M\varepsilon,$$

где $M = \sup|g|$ и та же оценка верна и для предельной случайной величины X. На промежутке $(-A, A]$ функцию g можно равномерно приблизить ступенчатой функцией f_ε (g непрерывна на отрезке $[-A, A]$, а значит и равномерно непрерывна на нем): $|g(x) - f_\varepsilon(x)| \le \varepsilon \ \forall x \in (-A, A]$. Кроме того, не ограничивая общности, можно считать, что точки разрыва функции f_ε будут точкам непрерывности функции F_X, т.е. $f_\varepsilon \in \mathcal{F}$. Пусть $f_\varepsilon(x) = 0$ при $x \notin (-A, A]$. Тогда,

$$\mathbb{E}\Big|I_{\{-AX_n\le A\}} g(X_n) - f_\varepsilon(X_n)\Big| \le \varepsilon.$$

и та же оценка верна и для предельной случайной величины X. Таким образом

$$\mathbb{E}\Big|g(X_n) - f_\varepsilon(X_n)\Big| \le 4M\varepsilon + \varepsilon.$$

Теорема доказана. □

30. Связь сходимости по распределению со сходимостью по вероятности.

Теперь обсудим связь сходимости по распределению и остальных сходимостей.

Предложение 96. *Если g — непрерывная функция и $X_n \xrightarrow{P} X$, то $g(X_n) \xrightarrow{P} g(X)$.*

Доказательство. Пусть $C > 0$. На отрезке $[-C, C]$ функция g равномерно непрерывна. Для всякого $\varepsilon > 0$ существует $\delta > 0$ такое, что из $|x - y| < \delta$ следует $|g(x) - g(y)| < \varepsilon$ для всех $x, y \in [-C, C]$. Следовательно, имеет место неравенство

$$\begin{aligned}
P(|g(X_n) - g(X)| \ge \varepsilon) \le P(|g(X_n) - g(X)| \ge \varepsilon, X_n, X \in [-C, C]) +& \\
P(|g(X_n) - g(X)| \ge \varepsilon, |X_n| > C) + P(|g(X_n) - g(X)| \ge \varepsilon, |X| > C)& \\
\le P(|X_n - X| \ge \delta) + P(|X_n| \ge C) + P(|X| \ge C) \le& \\
P(|X_n - X| \ge \delta) + P(|X_n - X| \ge C/2) + P(|X| \ge C/2) + P(|X| \ge C)&.
\end{aligned}$$

Устремляем сначала $n \to \infty$, получаем

$$\overline{\lim_{n\to\infty}} P(|g(X_n) - g(X)| \ge \varepsilon) \le P(|X| \ge C/2) + P(|X| \ge C).$$

Устремляем теперь $C \to \infty$. Получаем

$$0 \le \underline{\lim_{n\to\infty}} P(|g(X_n) - g(X)| \ge \varepsilon) \le \overline{\lim_{n\to\infty}} P(|g(X_n) - g(X)| \ge \varepsilon) \le 0,$$

а значит $\lim\limits_{n\to\infty} P(|g(X_n) - g(X)| \ge \varepsilon) = 0$ и утверждение доказано. □

Предложение 97 (абсолютная непрерывность ожидания). *Пусть $X \ge 0$ п.в. и существует математическое ожидание $\mathbb{E}X$. Тогда для каждого $\varepsilon > 0$ найдется такое $\delta > 0$, что для произвольного события A с $P(A) \le \delta$, выполнено $\mathbb{E}[XI_A] < \varepsilon$.*

Доказательство. Заметим, что $\mathbb{E}X = \mathbb{E}X^+ = \sup\{\mathbb{E}U\colon 0 \le U \le X^+,\ U\text{— ограниченная}\}$. Пусть ограниченная $U \ge 0$ выбрана так, что $\mathbb{E}U \le \mathbb{E}X \le \mathbb{E}U + \varepsilon/2$. Предположим, что $U < R$, тогда $\mathbb{E}[XI_A] = \mathbb{E}[(X^+ - U)I_A] + \mathbb{E}[UI_A] \le \mathbb{E}[X^+ - U] + RP(A) \le \varepsilon/2 + RP(A)$. Выбирая $\delta < \frac{\varepsilon}{2R}$ получаем требуемое утверждение. □

Теорема 98 (Лебега о мажорируемой сходимости). *Пусть $X_n \xrightarrow{P} X$ (или $X_n \xrightarrow{\text{п.н.}} X$) и пусть существует случайная величина Y с конечным ожиданием, для которой $|X| \le Y$ п.н. и $|X_n| \le Y$ п.н. $\forall n$. Тогда $\mathbb{E}X_n \to \mathbb{E}X$ при $n \to \infty$.*

Доказательство. Мы замечаем, что

$$|\mathbb{E}X_n - \mathbb{E}X| \le \mathbb{E}|X_n - X| = \mathbb{E}\big[|X_n - X| I_{\{|X_n - X| \ge \varepsilon\}}\big] + \mathbb{E}\big[|X_n - X| I_{\{|X_n - X| < \varepsilon\}}\big] \le 2\mathbb{E}\big[Y I_{\{|X_n - X| \ge \varepsilon\}}\big] + \varepsilon.$$

Остается лишь заметить, что по свойству абсолютной непрерывности ожидания найдется такое $\delta > 0$, что для произвольного события A с $P(A) \le \delta$, выполнено $\mathbb{E}[YI_A] < \varepsilon$, а также, что что для некоторого N при $n > N$ верно $P(|X_n - X| \ge \varepsilon) < \delta$ из-за сходимости случайных величин по вероятности. □

Следствие 99. *Пусть $X_n \xrightarrow{P} X$, тогда $X_n \xrightarrow{d} X$.*

Доказательство. Достаточно проверить, что для каждой непрерывной ограниченной функции g выполнено $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$. Т.к. g — ограничена, то существует число M для которого $|g(x)| \le M$ для каждой точки $x \in \mathbb{R}$. Т.к. $X_n \xrightarrow{P} X$, то $g(X_n) \xrightarrow{P} g(X)$. Применяя теорему Лебега с $Y := M$ (постоянная случайная величина), получаем сходимость $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$. □

31. Характеристические функции. Центральная предельная теорема.

Определение 100. Характеристической функцией случайной величины X называется функция

$$\varphi_X(t) = \mathbb{E}\big[e^{it\cdot X}\big] := \mathbb{E}\big[\cos(t \cdot X)\big] + i\mathbb{E}\big[\sin(t\cdot X)\big].$$

Предложение 101 (Свойства характеристических функций.).
1) $\varphi_X(0) = 1$, $|\varphi_X(t)| \le 1$ $\forall t \in \mathbb{R}$;
2) $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$;
3) если $X_1, \ldots, X_n$ — независимые случайные величины и $S_n = X_1 + \ldots + X_n$, то

$$\varphi_{S_n}(t) = \varphi_{X_1}(t) \cdot \ldots \cdot \varphi_{X_n}(t).$$

Доказательство. 1) Т.к. $\big(\mathbb{E}\big[\cos(t\cdot X)\big]\big)^2 \le \mathbb{E}\big[\cos^2(t\cdot X)\big]$ и $\big(\mathbb{E}\big[\sin(t\cdot X)\big]\big)^2 \le \mathbb{E}\big[\sin^2(t\cdot X)\big]$, получается $|\varphi_X(t)| \le 1$.

2) $\varphi_{aX+b}(t) = \mathbb{E}\big[e^{it\cdot(aX+b)}\big] = e^{itb}\mathbb{E}\big[e^{iat\cdot X}\big] = e^{itb}\varphi_X(at)$.

3) $\mathbb{E}e^{it\cdot S_n} = \mathbb{E}[e^{it\cdot X_1} \cdot \ldots \cdot e^{it\cdot X_n}] = \mathbb{E}e^{it\cdot X_1} \cdot \ldots \cdot \mathbb{E}e^{it \cdot X_n}$, где мы воспользовались теоремой о математическом ожидании произведения независимых случайных величин. □

Пример 102. Пусть $Z \sim \mathcal{N}(0,1)$. Тогда

$$\varphi_Z(t) = \mathbb{E}e^{it\cdot Z} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \cos(tx) e^{-x^2/2}\, dx.$$

Заметим, что

$$\varphi_Z'(t) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \sin(tx) e^{-x^2/2}\, dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\cos(tx) e^{-x^2/2}\, dx = -t\varphi_Z(t).$$

Решая дифференциальное уравнение, находим $\varphi_Z(t) = Ce^{-t^2/2}$, но $\varphi_Z(0) = 1$, а значит $C = 1$.

32. Теоремы Вейерштрасса.

Теорема 103 (Вейерштрасса)**.** *Пусть функция f непрерывна на отрезке $[0,1]$. Тогда для каждого $\varepsilon > 0$ существует такой многочлен P_ε, что*

$$\max_{x\in[0,1]} |f(x) - P_\varepsilon(x)| < \varepsilon.$$

Доказательство. Пусть S_n – число успехов в n бросаниях монеты с вероятностью успеха p. Пусть $\varepsilon > 0$. Оценим разность

$$|f(p) - \mathbb{E}f(S_n/n)|.$$

Положим $M = \max\limits_{x\in[0,1]} |f(x)|$. Так как $f(p)$ – константа, то

$$|f(p) - \mathbb{E}f(S_n/n)| \le \mathbb{E}|f(p) - f(S_n/n)|.$$

Найдем $\delta > 0$ такое, что $|f(x)-f(y)| < \varepsilon$ для всех $|x-y| < \delta$. Пусть $A_\delta = \{\omega\colon |p-S_n/n| \ge \delta\}$. Имеем

$$\mathbb{E}|f(p) - f(S_n/n)| = \mathbb{E}\bigl[|f(p) - f(S_n/n)|I_{A_\delta}\bigr] + \mathbb{E}\bigl[|f(p) - f(S_n/n)|I_{\Omega\setminus A_\delta}\bigr].$$

Второе слагаемое меньше ε так как $|p - S_n/n| < \delta$ и, значит, $|f(p) - f(S_n/n)| < \varepsilon$. Оценим первое слагаемое:

$$\mathbb{E}\bigl[|f(p) - f(S_n/n)|I_{A_\delta}\bigr] \le 2MP(A_\delta).$$

Напомним, что по неравенству Чебышева $P(A_\delta) \le n^{-1}\delta^{-2}$. Мы получили оценку

$$|f(p) - \mathbb{E}f(S_n/n)| \le \varepsilon + \frac{1}{n\delta^2}.$$

Выбираем n столь большим, что $n^{-1}\delta^{-2} < \varepsilon$, и получаем $\max\limits_{p\in[0,1]} |f(p) - \mathbb{E}f(S_n/n)| < 2\varepsilon$. Остается заметить, что $\mathbb{E}f(S_n/n)$ является многочленом от p. □

Линейным преобразованием теорема Вейерштрасса переносится на произвольный отрезок.

Следствие 104 (Вейерштрасса)**.** *Если непрерывная на $\mathbb{R}$ функция f является 2π периодической, то для каждого ε существует такой тригонометрический полином*

$$T_N(x) = \frac{a_0}{2} + \sum_{k=1}^{N} \bigl(a_k\cos(kx) + b_k\sin(kx)\bigr),$$

что $\max\limits_{x\in\mathbb{R}} |f(x) - T_N(x)| < \varepsilon$.

Доказательство. Во первых заметим, что $\cos^k x$ и $\sin^k x$ являются тригонометрическими полиномами (по индукции). Покажем, что если функция f четная, то ее можно приблизить многочленом от $\cos x$. Действительно, функция $y \mapsto f(\arccos y)$ является непрерывной функцией на $[-1,1]$. По теореме Вейерштрасса существует такой многочлен P, что $\max\limits_{y\in[-1,1]} |f(\arccos y) - P(y)| < \varepsilon$. Тогда $\max\limits_{x\in[0,\pi]} |f(x) - P(\cos x)| < \varepsilon$. В силу четности и периодичности

$$\max_{x\in\mathbb{R}} |f(x) - P(\cos x)| = \max_{x\in[-\pi,\pi]} |f(x) - P(\cos x)| = \max_{x\in[0,\pi]} |f(x) - P(\cos x)| < \varepsilon.$$

Теперь научимся сводить общий случай к уже рассмотренному. Каждая функция f раскладывается в сумму четной $f_0(x) := \frac{f(x)+f(-x)}{2}$ и нечетной $f_1(x) := \frac{f(x)-f(-x)}{2}$ частей. Предположим, что f_1 нечетная функция. Тогда $f_1(x)\sin x$ является четной. Значит мы умеем приближать $f(x)\sin^2 x = f_0(x)\sin^2 x + (f_1(x)\sin x)\sin x$, где f – произвольная непрерывная 2π периодическая функция. Тогда можно взять вместо f новую функию $g(y) := f(y-\pi/2)$ и приблизить равномерно тригонометрическим многочленом $T(y)$ функцию $g(y)\sin^2 y$ или, что тоже самое, функцию $f(x)\sin^2(x+\pi/2)$ приблизить тригонометрическим многочленом

$T(x+\pi/2)$. Так как $f(x) = f(x)\cos^2 x + f(x)\sin^2 x = f(x)\sin^2(x+\pi/2) + f(x)\sin^2 x$, то мы научились приближать тригонометрическим многочленом функцию f. $\square$

Линейным преобразованием предыдущая теорема Вейерштрасса переносится на любые непрерывные периодические функции с той поправкой, что надо брать функции $\cos(\frac{\pi k}{T}x), \sin(\frac{\pi k}{T}x)$, где $2T$ – период f.

33. Описание сходимости по распределению в терминах сходимость характеристических функций.

Лемма 105. *Пусть X — случайная величина, тогда φ_X — непрерывная функция.*

Доказательство. Следует из теоремы Лебега. $\square$

Теорема 106. *Последовательность X_n сходится по распределению к X тогда и только тогда, когда для каждого t выполнено $\lim\limits_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t)$.*

Доказательство. Если $X_n \xrightarrow{d} X$, то для каждой непрерывной ограниченной функции g выполнено $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$. Функции $x \mapsto \sin(tx)$ и $x \mapsto \cos(tx)$ непрерывны и ограничены.

Пусть для каждого t выполнено $\lim\limits_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t)$. По линейности предела есть сходимость $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ для каждой функции f вида

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{N} \big(a_k \cos(t_k x) + b_k \sin(t_k x)\big).$$

Обозначим семейство таких функций символом $\mathcal{F}$. Как мы уже знаем, чтобы доказать сходимость $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ для каждой непрерывной ограниченной функции g, достаточно показать, что для каждой такой функции g и для каждого $\varepsilon > 0$ найдется такая $f_\varepsilon \in \mathcal{F}$, что $\mathbb{E}|g(X_n) - f_\varepsilon(X_n)|$ для каждого $n \in \mathbb{N} \cup \{0\}$, где $X_0 = X$.

Проведем доказательство в два шага.

1) Сначала для произвольной случайной величины Y оценим $P(|Y| \geq t)$ при $t > 0$ в терминах характеристической функции φ_Y. Пусть $Z \sim \mathcal{N}(0,1)$. Заметим, что

$$\begin{aligned}
P(|Y| \geq t) &= P\big(\exp(-\tfrac{t^{-2}|Y|^2}{2}) \leq e^{-1/2}\big) = P\big(\varphi_Z(t^{-1}Y) \leq e^{-1/2}\big) = \\
&= P\big(1 - \varphi_Z(t^{-1}Y) \geq 1 - e^{-1/2}\big) \leq \tfrac{1}{1-e^{-1/2}}\mathbb{E}\big[1 - \varphi_Z(t^{-1}Y)\big] = \\
&= \tfrac{1}{1-e^{-1/2}}\Big(1 - \mathbb{E}\Big[\int_{-\infty}^{\infty} e^{izt^{-1}Y} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz\Big]\Big) = \tfrac{1}{1-e^{-1/2}}\Big(1 - \int_{-\infty}^{\infty} \mathbb{E}\big[e^{izt^{-1}Y}\big] \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz\Big) = \\
&= \tfrac{1}{1-e^{-1/2}}\Big(1 - \int_{-\infty}^{\infty} \varphi_Y(t^{-1}z) \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz\Big) = \tfrac{1}{1-e^{-1/2}}\Big(1 - \mathbb{E}\big[\varphi_Y(t^{-1}Z)\big]\Big) = \tfrac{1}{1-e^{-1/2}}\mathbb{E}\big[1 - \varphi_Y(t^{-1}Z)\big].
\end{aligned}$$

2) Пусть теперь $\varphi_{X_n}(t) \to \varphi_X(t)$ для каждой точки $t \in \mathbb{R}$. Пусть $A > 0$ выбрано так, чтобы

$$P(|X| \geq A) \leq \tfrac{1}{1-e^{-1/2}}\mathbb{E}\big[1 - \varphi_X(A^{-1}Z)\big] < \varepsilon.$$

Этого можно добиться в силу теоремы Лебега, т.к. $1 - \varphi_X(A^{-1}Z) \to 0$ при $A \to +\infty$ (при каждом $\omega \in \Omega$) и $|1 - \varphi_X(A^{-1}Z)| \leq 2$. Т.к. $\varphi_{X_n}(A^{-1}Z) \to \varphi_X(A^{-1}Z)$ при $n \to +\infty$ (при каждом $\omega \in \Omega$) и $|1 - \varphi_{X_n}(A^{-1}Z)| \leq 2$, то по теореме Лебега

$$P(|X_n| \geq A) \leq \tfrac{1}{1-e^{-1/2}}\mathbb{E}\big[1 - \varphi_{X_n}(A^{-1}Z)\big] \to \tfrac{1}{1-e^{-1/2}}\mathbb{E}\big[1 - \varphi_X(A^{-1}Z)\big] < \varepsilon.$$

Т.е. найдется такой номер n_0, что $P(|X_n| \geq A) < 2\varepsilon$ при каждом $n > n_0$. Увеличивая A, можно считать, что данная оценка верна для всех номеров.

Пусть g – ограниченная непрерывная функция и $M = \sup|g|$ и пусть $\varepsilon > 0$ фиксировано. Как мы уже поняли, существует такой отрезок $[-A, A]$, что

$$P(|X_n| \geq A) \leq \varepsilon, \quad P(|X| \geq A) \leq \varepsilon.$$

Пусть непрерывная функция g_ε совпадает с g на $[-A, A]$, на концах отрезка $[-A-1, A+1]$ равна нулю, а затем продолжена как периодическая функция с периодом $T = 2A + 2$. Конечно можно построить g_ε так, что $\max|g_\varepsilon| \leq M$ (просто соединив линейно значения g на концах отрезка $[-A, A]$ с нулевыми значениями в точках $\pm(A+1)$). Заметим, что

$$\mathbb{E}\big|g_\varepsilon(X_n) - g(X_n)\big| \leq 2M\varepsilon, \quad \mathbb{E}\big|g_\varepsilon(X) - g(X)\big| \leq 2M\varepsilon.$$

По теореме Вейерштрасса непрерывную периодическую функцию g_ε можно равномерно приблизить тригонометрическим многочленом $f_\varepsilon(x) = \frac{a_0}{2} + \sum\limits_{k=1}^{N}\big(a_k\cos(\frac{\pi k}{A+1}x) + b_k\sin(\frac{\pi k}{A+1}x)\big)$, т.е. выполнено $\sup\limits_{x\in\mathbb{R}}|g_\varepsilon(x) - f_\varepsilon(x)| < \varepsilon$. Таким образом

$$\mathbb{E}\big|f_\varepsilon(X_n) - g(X_n)\big| \leq 2M\varepsilon + \varepsilon$$

для каждого $n \in \mathbb{N} \cup \{0\}$, где $X_0 = X$. $\square$

Следствие 107. *Если у двух случайных величин совпадают характеристические функции, то эти величины имеют одинаковые распределения.*

Доказательство. Действительно, пусть $\varphi_X(t) = \varphi_Y(t)$ при каждом t. Рассмотрим последовательность случайных величин $X_n := X$. Тогда $\varphi_{X_n}(t) = \varphi_X(t) \to \varphi_Y(t)$ при каждом t, что дает сходимость $F_X(x) := F_{X_n}(x) \to F_Y(x)$ в каждой точке x непрерывности функции F_Y. Т.е. $F_X(x) = F_Y(x)$ в каждой точке x непрерывности функции F_Y. Т.к. точек разрыва у монотонной функции не более чем счетное количество, то к каждой точке разрыва функции F_Y сходится справа некоторая последовательность точек непрерывности этой функции, в которых F_X и F_Y совпадают. Т.к. функции распределения непрерывны справа, то они совпадают во всех точках $x \in \mathbb{R}$, а значит совпадают и распределения случайных величин X и Y. $\square$

34. Центральная предельная теорема.

Предложение 108. *Пусть случайная величина X обладает конечным k-ым моментом, т.е. $\mathbb{E}|X|^k < \infty$. Тогда φ_X имеет непрерывную k-ю производную, и*

$$\varphi^{(k)}(0) = i^k\mathbb{E}[X^k].$$

Доказательство. Заметим, что для произвольного $h \in \mathbb{R}$ выполняется оценка

$$\Big|\frac{e^{ihy} - 1}{h}\Big| \leq |y|.$$

По теореме Лебега для произвольной последовательности $h_n \to 0$ выполнено

$$\lim_{h_n\to 0}\frac{\varphi_X(t+h_n) - \varphi_X(t)}{h_n} = \lim_{h\to 0}\mathbb{E}\Big[e^{it\cdot X}\Big(\frac{e^{ih_nX} - 1}{h_n}\Big)\Big] = \mathbb{E}[e^{it\cdot X}iX],$$

т.к. $\Big|e^{it\cdot X}\big(\frac{e^{ih_nX}-1}{h_n}\big)\Big| \leq |X|$ и т.к. $\lim\limits_{h_n\to 0} e^{it\cdot X}\big(\frac{e^{ih_nX}-1}{h_n}\big) = e^{it\cdot X}iX$. Производные более высокого порядка рассматриваются аналогично. $\square$

Теорема 109. (Центральная предельная теорема) *Пусть $\{X_n\}$ — последовательность независимых одинаково распределенных случайных величин, причем $\mathbb{E}X_1 = a$ и $\mathbb{D}X_1 = \sigma^2$. Тогда для всех t*

$$\lim_{n\to\infty} P\Big(\frac{X_1 + \ldots + X_n - na}{\sqrt{n\sigma^2}} \leq t\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}\,dx$$

или, равносильно, $\frac{S_n-\mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}} \xrightarrow{d} Z$, *где* $Z \sim \mathcal{N}(0,1)$ *— стандартная нормальная случайная величина, а* $S_n := X_1 + \ldots + X_n$.

Доказательство. Переходя к величинам $X_j - a$ можно далее считать, что $a = 0$. Пусть φ – характеристическая функция X_1. Тогда характеристическая функция случайной величины

$$\frac{X_1 + \ldots + X_n}{\sigma\sqrt{n}}$$

равна

$$\varphi_n(t) = \Bigl(\varphi\Bigl(\frac{t}{\sigma\sqrt{n}}\Bigr)\Bigr)^n.$$

Заметим, что $\varphi(0) = 1$, $\varphi'(0) = i\mathbb{E}X_1 = 0$, $\varphi''(0) = -\mathbb{E}X_1^2 = -\mathbb{D}X_1 = -\sigma^2$. По формуле Тейлора

$$\varphi\Bigl(\frac{t}{\sigma\sqrt{n}}\Bigr) = 1 - \frac{t^2}{2n} + \bar{o}\bigl(\frac{1}{n}\bigr).$$

Следовательно,

$$\varphi_n(t) = \Bigl(1 - \frac{t^2}{2n} + \bar{o}\bigl(\frac{1}{n}\bigr)\Bigr)^n \to e^{-t^2/2}, \quad n \to \infty.$$

Остается заметить, что $e^{-t^2/2}$ – характеристическая функция стандартной нормальной случайной величины $Z \sim \mathcal{N}(0,1)$. □

35. Теоремы о наследовании сходимости.

Для применения центральной предельной теоремы на практике важную роль играют так называемые теоремы о наследовании сходимости (они же теоремы о непрерывности).

Замечание 110. Мы знаем, что для непрерывной функции g из сходимости $X_n \xrightarrow{P} X$ следует сходимость $g(X_n) \xrightarrow{P} g(X)$.

На самом деле можно слегка ослабить условия на g. Пусть $X_n\colon \Omega \to (a,b)$, $X\colon \Omega \to (a,b)$ и g непрерывна на (a,b). Тогда из $X_n \xrightarrow{P} X$ следует $g(X_n) \xrightarrow{P} g(X)$. Доказательство аналогично.

Предложение 111. *Если последовательность случайных величин* X_n *сходится по распределению к* X*, то для всякой непрерывной функции* g *случайные величины* $g(X_n)$ *сходятся по распределению к* $g(X)$.

Доказательство. Немедленно следует из эквивалентного определения сходимости по распределению. Действительно, сходимость $X_n \xrightarrow{d} X$ равносильна тому, что для каждой непрерывной ограниченной функции f выполнено $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$. Но для непрерывной g функция $f(g(x))$ будет непрерывной ограниченной (если такой была f). Поэтому для каждой непрерывной ограниченной функции f выполнено $\mathbb{E}f(g(X_n)) \to \mathbb{E}f(g(X))$, что равносильно сходимости $g(X_n) \xrightarrow{d} g(X)$. □

Лемма 112. *Пусть* X, Y, Z *— случайные величины. Тогда* $\forall t \in \mathbb{R}$, $\forall \varepsilon > 0$ *выполнено*

$$P(X + Z \le t - \varepsilon) - P(|Y - Z| \ge \varepsilon) \le P(X + Y \le t) \le P(X + Z \le t + \varepsilon) + P(|Y - Z| \ge \varepsilon).$$

Доказательство. Заметим, что

$$\begin{aligned} P(X + Y \le t) = P(X + Y \le t,\ |Y - Z| < \varepsilon) + P(X + Y \le t,\ |Y - Z| \ge \varepsilon) \\ \le P(X + Z - \varepsilon \le t) + P(|Y - Z| \ge \varepsilon). \end{aligned}$$

Нижняя оценка получается из уже доказанной с заменой Y на Z, Z на Y и t на $t - \varepsilon$. □

Предложение 113. *Если* $X_n \xrightarrow{P} a = const$ *и* $Y_n \xrightarrow{d} Y$, *то* $X_n \cdot Y_n \xrightarrow{d} a \cdot Y$ *и* $X_n + Y_n \xrightarrow{d} a + Y$.

Доказательство. Сначала докажем утверждение для суммы. Пусть $\varepsilon > 0$. Тогда по предыдущей лемме (с $Z = a$)

$$F_{Y_n}(t-a-\varepsilon) - P(|X_n - a| \geq \varepsilon) \leq P(X_n + Y_n \leq t) \leq F_{Y_n}(t-a+\varepsilon) + P(|X_n - a| \geq \varepsilon)$$

Если $t - a$ точка непрерывности функции F_Y, то, устремляя сначала $n \to \infty$, а затем $\varepsilon \to 0$, получаем равенство

$$\lim_{n\to\infty} P(X_n + Y_n \leq t) = F_Y(t-a) = F_{a+Y}(t).$$

Остается заметить, что $t-a$ точка непрерывности функции F_Y тогда и только тогда, когда t точка непрерывности функции F_{a+Y}.

Докажем утверждение для произведения. Пусть сначала $a = 0$. Так как для произвольных $\varepsilon > 0$ и $C > 0$ верно включение

$$\{|X_n \cdot Y_n| > \varepsilon\} \subset \{|Y_n| > C\} \bigcup \{|X_n| > \varepsilon C^{-1}\},$$

то

$$P(|X_n \cdot Y_n| > \varepsilon) \leq 1 - F_{Y_n}(C) + F_{Y_n}(-C) + P(|X_n| > \varepsilon C^{-1}).$$

Следовательно, устремляя сначала $n \to \infty$, а затем $C \to \infty$, заключаем, что $X_n \cdot Y_n \to 0$ по вероятности. Остается вспомнить, что сходимость по вероятности влечет сходимость по распределению. Общий случай выводится из утверждения для суммы и того, что величины $(X_n - a)Y_n$ стремятся по вероятности к нулю и aY_n сходится по распределению к aY. $\square$

Замечание 114. Заметим, что в случае, когда $a = 0$ и $X_n \xrightarrow{P} a$ и $Y_n \xrightarrow{d} Y$ нами было доказано, что $X_n \cdot Y_n \xrightarrow{P} 0$ (т.е. имеет место сходимость по вероятности).

Замечание 115. На самом деле из семинарских задач известно, что сходимость $X_n \xrightarrow{P} c$ равносильна сходимости $X_n \xrightarrow{d} c$ в случае, когда c — постоянная.

Типичные примеры последовательностей X_n, сходящихся к константе по вероятности, дает закон больших чисел.

Пример 116 (Выборочная дисперсия)**.** Пусть задана последовательность независимых и одинаково распределенных случайных величин X_j, причем $\mathbb{E}X_j = a$ и $\mathbb{D}X_j = \sigma^2$. Тогда последовательность случайных величин

$$s_n^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X}_n)^2,$$

где $\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}$, сходится по вероятности к σ^2. Действительно,

$$s_n^2 = \frac{n}{n-1}\frac{1}{n}\sum_{j=1}^n X_j^2 - \frac{n}{n-1}\overline{X}_n^2.$$

Теперь остается заметить, что по закону больших чисел

$$\frac{1}{n}\sum_{j=1}^n X_j^2 \to \sigma^2 + a^2, \quad \overline{X}_n \to a$$

по вероятности.

Пример 117. Пусть задана последовательность независимых и одинаково распределенных случайных величин X_j, причем $\mathbb{E}X_j = a$ и $\mathbb{D}X_j = \sigma^2 > 0$. Тогда из центральной предельной теоремы следует, что по распределению

$$\frac{\sqrt{n}(\overline{X}_n - a)}{\sigma} \to Z \sim N(0,1).$$

Более того, т. к. $s_n^2 \to \sigma^2 > 0$ по вероятности, то имеет место сходимость по распределению величин

$$\frac{\sqrt{n}(\overline{X}_n - a)}{\sqrt{s_n^2}} = \frac{\sqrt{n}(\overline{X}_n - a)}{\sigma} \cdot \sqrt{\frac{\sigma^2}{s_n^2}} \to Z \sim N(0,1).$$

Предложение 118. *Пусть $a, h_n \in \mathbb{R}$, $h_n \to 0$ и f — непрерывная на $\mathbb{R}$ и дифференцируемая в точке a функция. Если последовательность случайных величин $X_n \xrightarrow{d} X$, то*

$$\frac{f(a + h_n X_n) - f(a)}{h_n} \xrightarrow{d} f'(a)X.$$

Доказательство. Рассмотрим функцию $g(y) := \frac{f(a+y)-f(a)}{y}$ при $y \neq 0$ и $g(0) = f'(a)$. Т.к. f дифференцируема в точке a, то g — непрерывная на $\mathbb{R}$ функция. Тогда, т.к. $h_n X_n \to 0$ по вероятности (см. доказательство предыдущего предложения и замечание после него), то $g(h_n X_n) \to g(0) = f'(a)$ по вероятности. Кроме того $X_n \xrightarrow{d} X$ по условию. Поэтому $g(h_n X_n) \cdot X_n \xrightarrow{d} f'(a) \cdot X$. Остается заметить, что $g(h_n X_n) \cdot X_n = \frac{f(a+h_n X_n)-f(a)}{h_n}$. □

Рассмотрим типичные примеры применения теорем непрерывности.

Пример 119. Пусть задана последовательность независимых и одинаково распределенных случайных величин X_j, причем $\mathbb{E}X_j = a$ и $\mathbb{D}X_j = \sigma^2 > 0$. Если f — дифференцируемая в точке a функция, то

$$\sqrt{n}(f(\overline{X}_n) - f(a)) \xrightarrow{d} Y \sim N(0, q^2), \quad q = \sigma f'(a).$$

Действительно, имеет место равенство

$$\sqrt{n}\frac{f(\overline{X}_n) - f(a)}{\sigma} = \frac{f\big(a + n^{-1/2}\sigma Z_n\big) - f(a)}{\sigma n^{-1/2}},$$

где

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - a)}{\sigma} \xrightarrow{d} Z \sim N(0,1).$$

По доказанному

$$\sqrt{n}(f(\overline{X}_n) - f(a)) = \sigma \cdot \sqrt{n}\frac{f(\overline{X}_n) - f(a)}{\sigma} \xrightarrow{d} \sigma f'(a) \cdot Z \sim N(0, q^2).$$

Пример 120. Пусть $X_1, \ldots, X_n$ положительные независимые одинаково распределенные случайные величины, $\mathbb{E}X_1 = a$, $0 < \mathbb{D}X_1 = \sigma^2 < \infty$. Рассмотрим случайную величину $S_n = X_1 + \ldots + X_n$ и найдем предел в смысле сходимости по распределению у последовательности случайных величин $\sqrt{n}\big(\frac{n}{S_n} - \frac{1}{a}\big)$.

Первый способ:

$$\sqrt{n}\Big(\frac{n}{S_n} - \frac{1}{a}\Big) = -\frac{1}{a}\frac{n}{S_n}\sqrt{n}\Big(\frac{S_n}{n} - a\Big).$$

По ЦПТ $\sqrt{n}\big(\frac{S_n}{n} - a\big) \xrightarrow{d} Y \sim \mathcal{N}(0, \sigma^2)$, по ЗБЧ $\frac{n}{S_n} \xrightarrow{P} \frac{1}{a}$. Таким образом, по доказанному

$$\sqrt{n}\Big(\frac{n}{S_n} - \frac{1}{a}\Big) \xrightarrow{d} -\frac{1}{a^2}Y \sim \mathcal{N}\Big(0, \frac{\sigma^2}{a^4}\Big).$$

Второй способ: пусть $f(x) = 1/x$, тогда

$$\sqrt{n}\Big(\frac{n}{S_n} - \frac{1}{a}\Big) = \sqrt{n}\Big(f\Big(\frac{S_n}{n}\Big) - f(a)\Big) \to Z \sim \mathcal{N}(0, \sigma^2(f'(a))^2).$$

36. Многомерная характеристическая функция и ЦПТ.

Положим

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \ldots + x_my_m, \quad x = (x_1, x_2, \ldots, x_m), y = (y_1, y_2, \ldots, y_m) \in \mathbb{R}^m.$$

Характеристическая функция случайного вектора $X = (X_1, \ldots, X_m)$ определяется равенством

$$\varphi_X(y) = \mathbb{E}e^{i\langle X,y\rangle}.$$

Определение 121. Последовательность случайных векторов $X^n = (X_1^n, \ldots, X_m^n)$ **сходится по распределению** к случайному вектору $X = (X_1, \ldots, X_m)$, если для каждой непрерывной, ограниченной функции $g\colon \mathbb{R}^m \to \mathbb{R}$ выполнено $\mathbb{E}g(X^n) \to \mathbb{E}g(X)$ (обозначение $X^n \xrightarrow{d} X$).

Теорема 122 (б/д)**.** *Последовательность случайных векторов X^n сходится по распределению к случайному вектору X тогда и только тогда, когда $\varphi_{X^n}(y) \to \varphi_X(y)$ для каждого $y \in \mathbb{R}^m$.*

Следствие 123 (б/д)**.** *Если $\varphi_X = \varphi_Y$, то векторы X и Y имеют одинаковые распределения.*

Следствие 124. *Случайные величины $X_1, \ldots, X_m$ независимы тогда и только тогда, когда*

$$\varphi_X(y_1, \ldots, y_m) = \varphi_{X_1}(y_1)\cdots\varphi_{X_m}(y_m) \quad \forall y \in \mathbb{R}^m,$$

где $X = (X_1, \ldots, X_m)$.

Доказательство. Пусть выполенно описанное равенство. Рассмотрим случайный вектор $Y = (Y_1, \ldots, Y_m)$ с функцией распределения $F_Y(x_1, \ldots, x_m) = F_{X_1}(x_1)\cdot\ldots\cdot F_{X_m}(x_m)$, т.е. Y_j имеет такое же распределение, как и X_j, и компоненты вектора Y независимы. Тогда

$$\varphi_Y(y) = \mathbb{E}e^{i\sum_{j=1}^m y_jY_j} = \prod_{j=1}^m \mathbb{E}e^{iy_jY_j} = \varphi_{Y_1}(y_1)\cdots\varphi_{Y_m}(y_m).$$

Т.к. распределение Y_j такое же, как и X_j, то

$$\varphi_{Y_1}(y_1)\cdots\varphi_{Y_m}(y_m) = \varphi_{X_1}(y_1)\cdots\varphi_{X_m}(y_m) = \varphi_X(y_1, \ldots, y_m) = \varphi_X(y).$$

По предыдущему получаем, что распределение векторов X и Y совпадают, а значит компоненты вектора X независимы.

Если компоненты независимы, то равенство следует из того, что ожидание произведения независимых величин есть произведение ожиданий этих величин. □

Определение 125. Пусть $X = (X_1, \ldots, X_m)$ случайный вектор. Матрица R_X с компонентами $r_{kj} := cov(X_k, X_j)$ называется ковариационной матрицей вектора X.

Лемма 126. *Симметричная неотрицательно определенная матрица R является ковариационной матрицей случайного вектора X тогда и только тогда, когда*

$$\langle Rx, y\rangle = cov(\langle x, X\rangle, \langle y, X\rangle) = \mathbb{E}[\langle x, X - a\rangle\langle y, X - a\rangle],$$

где $a = (a_1, \ldots, a_m)$ — вектор средних, т.е. $a_j = \mathbb{E}X_j$.

Доказательство. Для совпадения двух билинейных функций достаточно их совпадения на базисных векторах. □

Следствие 127. *Пусть случайный вектор X имеет матрицу ковариаций R, тогда для произвольной матрицы A и для произвольного вектора b матрица ковариаций вектора $AX+b$ имеет вид ARA^*.*

Доказательство. Пусть R' — матрица ковариаций вектора $AX+b$, тогда для произвольных векторов x, y выполнено тождество

$$\langle R'x, y\rangle = cov\big[\langle x, AX+b\rangle, \langle y, AX+b\rangle\big] = cov\big[\big(\langle A^*x, X\rangle + \langle x, b\rangle\big), \big(\langle A^*y, X\rangle + \langle y, b\rangle\big)\big] = \\ = cov\big[\langle A^*x, X\rangle, \langle A^*y, X\rangle\big] = \langle RA^*x, A^*y\rangle = \langle ARA^*x, y\rangle.$$

В силу произвольности векторов x и y, получаем, что $R' = ARA^*$. □

Предложение 128. *Для произвольного случайного вектора X и для произвольного числа t выполнено*

$$\varphi_{tX}(y) = \varphi_X(ty) = \varphi_{\langle X,y\rangle}(t).$$

Доказательство. Равенства проверяются подстановкой в определение. □

Теорема 129 (многомерная ЦПТ)**.** *Пусть случайные векторы $X^n = (X_1^n, \ldots, X_m^n)$ независимы, одинаково распределены и имеют конечные*

$$a_j = \mathbb{E}X_j^n, \quad r_{kj} = \operatorname{cov}(X_k^1, X_j^1).$$

Тогда последовательность случайных векторов $Y^n = (Y_1^n, \ldots, Y_m^n)$ с компонентами

$$Y_j^n = \frac{X_j^1 + \ldots + X_j^n - na_j}{\sqrt{n}}.$$

сходится по распределению к вектору Z, характеристическая функция которого имеет вид

$$\varphi_Z(y) = e^{-\frac{1}{2}\langle Ry,y\rangle}, \quad R = (r_{kj}).$$

Доказательство. Пусть $y \in \mathbb{R}^m$. Рассмотрим последовательность случайных величин

$$\xi_n := \frac{\langle X^1, y\rangle + \ldots + \langle X^n, y\rangle - n\langle a, y\rangle}{\sqrt{n}} = \langle Y^n, y\rangle.$$

По ЦПТ $\xi_n \xrightarrow{d} Z_y \sim \mathcal{N}(0, \mathbb{D}[\langle X^1, y\rangle])$. В частности,

$$\varphi_{Y^n}(y) = \varphi_{\langle Y^n, y\rangle}(1) = \varphi_{\xi_n}(1) \to \varphi_{Z_y}(1) = e^{-\frac{1}{2}\mathbb{D}[\langle X^1, y\rangle]}$$

Остается заметить, что $\mathbb{D}[\langle X^1, y\rangle] = \operatorname{cov}(\langle X^1, y\rangle, \langle X^1, y\rangle) = \langle Ry, y\rangle$. □

37. Многомерное нормальное распределение.

Определение 130. Случайный вектор X имеет нормальное распределение или является гауссовским, если характеристическая функция

$$\varphi_X(y) = \mathbb{E}[\exp(i\langle X, y\rangle)] = e^{-\frac{1}{2}\langle Ry,y\rangle + i\langle a,y\rangle},$$

где $a = (a_1, a_2, \ldots, a_m) \in \mathbb{R}^m$, R — симметричная неотрицательно определенная $m \times m$ матрица. Далее кратко пишем

$$X \sim N(a, R).$$

Предложение 131. *Если вектор $X \sim N(a, R)$, то вектор $AX + b \sim N(Aa + b, ARA^*)$.*

Доказательство. Доказываем простой подстановкой в определение:

$$\varphi_{AX+b}(y) = \mathbb{E}[\exp(i\langle AX+b, y\rangle)] = e^{i\langle b,y\rangle}\mathbb{E}[\exp(i\langle X, A^*y\rangle)] = e^{-\frac{1}{2}\langle RA^*y, A^*y\rangle + i\langle a, A^*y\rangle + i\langle b,y\rangle}.$$

Остается заметить, что $\langle RA^*y, A^*y\rangle = \langle ARA^*y, y\rangle$ и $\langle a, A^*y\rangle + \langle b, y\rangle = \langle Aa + b, y\rangle$. □

Теорема 132. *Вектор X имеет нормальное распределение тогда и только тогда, когда для каждого вектора y случайная величина $\langle X, y\rangle$ имеет нормальное распределение.*

Доказательство. Если $X \sim N(a, R)$ нормальный вектор, то

$$\varphi_{\langle X,y\rangle}(t) := \mathbb{E}e^{it\langle X,y\rangle} = e^{-\frac{1}{2}t^2\langle Ry,y\rangle + it\langle a,y\rangle},$$

т.е. $\langle X, y\rangle \sim \mathcal{N}(\langle a, y\rangle, \langle Ry, y\rangle)$. В частности, $\langle a, y\rangle = \mathbb{E}[\langle X, y\rangle]$ и $\langle Ry, y\rangle = \mathbb{D}[\langle X, y\rangle]$.

Наоборот,

$$\varphi_X(y) := \mathbb{E}e^{i\langle X,y\rangle} = \varphi_{\langle X,y\rangle}(1) = e^{-\frac{1}{2}\mathbb{D}[\langle X,y\rangle] + i\mathbb{E}[\langle X,y\rangle]} = e^{-\frac{1}{2}\langle Ry,y\rangle + i\langle a,y\rangle},$$

где R — ковариационная матрица вектора X, а a — вектор средних. □

Следствие 133. *Если $X \sim N(a, R)$, то R — ковариационная матрица вектора X, а a — вектор средних.*

Следствие 134. *Если вектор $(X_1, X_2, \ldots, X_n)$ имеет нормальное распределение, причем $cov(X_k, X_j) = 0$ при $j \neq k$, то случайные величины X_j независимы.*

Доказательство. Если указанные ковариации равны нулю, то

$$\varphi_{(X_1,\ldots,X_n)}(y_1, \ldots, y_n) = e^{-\frac{1}{2}(y_1^2\mathbb{D}[X_1]+\ldots+y_n^2\mathbb{D}[X_n]) + i(y_1\mathbb{E}[X_1]+\ldots+y_n\mathbb{E}[X_n])} = \varphi_{X_1}(y_1)\cdot\ldots\cdot\varphi_{X_n}(y_n).$$

□

Следствие 135. *Если $X \sim N(a, R)$, то найдется такая матрица A, что $X = AZ+a$, где $Z = (Z_1, \ldots, Z_k)$ и случайные величины Z_j независимы и имеют стандартное нормальное распределение, $AA^* = R$.*

Доказательство. Переходя к случанйым величинам $X_j' := X_j - a_j$ переходим к задаче нахождения ортонормированного базиса в линейном пространстве $\mathrm{span}\{X_1', \ldots, X_m'\}$ со скалярным произведением $(X, Y) := \mathbb{E}[XY]$. Эта задача решается методом оротганализации Грамма–Шмидта. Применив данный метод, получим случайные величины $Z_1, \ldots, Z_k$, которые линейно выражаются через $X_1', \ldots, X_m'$, т.е., в частности, вектор $Z = (Z_1, \ldots, Z_k)$ — нормальный и $\mathbb{E}[Z_j] = 0$, кроме того, система $\{Z_1, \ldots, Z_k\}$ является ортонормированным базисом в $\mathrm{span}\{X_1', \ldots, X_m'\}$. Тот факт, что это базис означает, что $X' = AZ$, где $Z = (Z_1, \ldots, Z_k)$. Ортонормированность означает, что $\mathrm{cov}(Z_k, Z_j) = E[Z_kZ_j] = 0$ и $\mathbb{D}[Z_j] = \mathbb{E}[Z_j^2] = 1$. Поэтому случайные величины $Z_j \sim \mathcal{N}(0, 1)$ и независимы. Равенство $AA^* = R$ следует из того, как меняется матрица ковариаций при линейных отображениях. □

Теорема 136. *Если $X \sim N(a, R)$ и $\det R \neq 0$, случайный вектор X имеет плотность*

$$\varrho(x) = \frac{1}{(2\pi)^{m/2}\sqrt{\det R}} e^{-\frac{1}{2}\langle R^{-1}(x-a), x-a\rangle}.$$

Доказательство. Т.к. $X = AZ + a$, где матрица A квадратная и невырожденая (иначе, $R = AA^*$ — вырождена), то

$$P(X \in B) = P(AZ + a \in B) = \frac{1}{(2\pi)^{m/2}} \int\limits_{Ax+a\in B} e^{-\frac{1}{2}|x|^2}\,dx = \frac{1}{(2\pi)^{m/2}\det A} \int\limits_B e^{-\frac{1}{2}|A^{-1}(y-a)|^2}\,dy.$$

Остается лишь заметить, что $|A^{-1}(y-a)|^2 = \langle A^{-1}(y-a), A^{-1}(y-a)\rangle = \langle (AA^*)^{-1}(y-a), y-a\rangle$ и $(\det A)^2 = \det R$. □

38. Условные математические ожидания: дискретный случай.

Предположим, что задана дискретная случайная величина

$$X(\omega) = \sum_{j=1}^{m} x_j I_{A_j}(\omega).$$

Рассмотрим следующую задачу: найти математическое ожидание X, если достоверно известно, что произошло событие B, $P(B) > 0$. Поскольку мы знаем, что событие B произошло, то надо пересчитать вероятности A_j с учетом новой информации, а именно, заменить $P(A_j)$ на $P(A_j|B)$. Таким образом, надо вычислить математическое ожидание не относительно исходной вероятностной меры P, а относительно условной вероятности $P(\,\cdot\,|B)$. Имеем

$$\mathbb{E}(X|B) = \sum_{j=1}^{m} x_j P(A_j|B) = \sum_{j=1}^{m} x_j \frac{\mathbb{E}(I_{A_j} I_B)}{P(B)} = \frac{\mathbb{E}(X I_B)}{P(B)}.$$

Это выражение будем называть условным математическим ожиданием X относительно события B. В частности, $\mathbb{E}(I_A|B) = \frac{\mathbb{E}(I_A I_B)}{P(B)} = P(A|B)$.

Пусть теперь имеется разбиение

$$\Omega = B_1 \cup B_2 \cup \ldots \cup B_n, \quad B_k \cap B_j = \emptyset, \quad P(B_k) > 0.$$

Обозначим разбиение $\{B_k\}$ через $\mathcal{B}$. Удобно собрать вместе значения условных математических ожиданий $\mathbb{E}(X|B_k)$ для разных $B_k \in \mathcal{B}$. Рассмотрим случайную величину:

$$\Lambda(\omega) = \sum_{k=1}^{n} I_{B_k}(\omega)\mathbb{E}(X|B_k).$$

Если $\omega \in B_k$, то эта случайная величина выдает среднее значение X при условии, что произошло событие B_k.

Определение 137. Величину Λ называют **условным математическим ожиданием** X относительно разбиения $\mathcal{B}$ и обозначают через $\mathbb{E}(X|\mathcal{B})$.

Случайную величину

$$P(A|\mathcal{B}) := \mathbb{E}(I_A|\mathcal{B})$$

называют условной вероятностью события A относительно разбиения $\mathcal{B}$. Заметим, что

$$\mathbb{E}(X|\mathcal{B}) = \sum_{k=1}^{n} I_{B_k}(\omega) \sum_{j=1}^{m} x_j P(A_j|B_k) = \sum_{j=1}^{m} x_j \sum_{k=1}^{n} I_{B_k}(\omega) P(A_j|B_k) = \sum_{j} x_j P(A_j|\mathcal{B}).$$

Рассмотрим важный пример, когда $\mathcal{B} = \{B, \overline{B} = \Omega \setminus B\}$. Тогда

$$P(A|\mathcal{B}) = I_B P(A|B) + I_{\overline{B}} P(A|\overline{B}).$$

Если $\omega \in B$, то $P(A|\mathcal{B})(\omega) = P(A|B)$.

Определение условного математического ожидания $\mathbb{E}(X|\mathcal{B})$ дословно переносится на произвольные случайные величины X.

Теорема 138. *Имеют место следующие свойства условного математического ожидания:*

(i) (линейность) $\mathbb{E}(\alpha X + \beta Y|\mathcal{B}) = \alpha\mathbb{E}(X|\mathcal{B}) + \beta\mathbb{E}(Y|\mathcal{B})$,

(ii) (монотонность) $X \le Y$ *п.н.* $\Rightarrow \mathbb{E}(X|\mathcal{B}) \le \mathbb{E}(Y|\mathcal{B})$,

(iii) (аналог формулы полной вероятности) $\mathbb{E}\big(\mathbb{E}(X|\mathcal{B})\big) = \mathbb{E}X$,

(iv) (независимость) если случайная величина X *не зависит от разбиения* $\mathcal{B}$, *т. е. случайные величины* X *и* I_{B_k} *независимы для каждого* k, *то* $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}X$.

(v) для всякой случайной величины $Z = \sum_{k=1}^{n} c_k I_{B_k}$ *выполнено* $\mathbb{E}(ZX|\mathcal{B}) = Z\mathbb{E}(X|\mathcal{B})$.

Доказательство. Свойства (i) и (ii) следуют из того, что они верны для $\mathbb{E}(X|B_k)$ для каждого k (т.к. они верны для математического ожидания относительно произвольной вероятностной меры). Свойство (iii) проверяется непосредственной подстановкой в определение:

$$\mathbb{E}\big(\mathbb{E}(X|\mathcal{B})\big) = \mathbb{E}\Big(\sum_{k=1}^{n} I_{B_k} \frac{\mathbb{E}(XI_{B_k})}{P(B_k)}\Big) = \sum_{k=1}^{n} \mathbb{E}(XI_{B_k}) = \mathbb{E}X.$$

Обоснуем пункт (iv). Так как X и I_{B_k} независимы, то

$$\mathbb{E}(X|B_k) = \frac{\mathbb{E}(XI_{B_k})}{P(B_k)} = \frac{\mathbb{E}X\mathbb{E}I_{B_k}}{P(B_k)} = \mathbb{E}X.$$

Следовательно,

$$\mathbb{E}(X|\mathcal{B}) = \sum_{k=1}^{n} I_{B_k}\mathbb{E}(X|B_k) = \sum_{k=1}^{n} I_{B_k}\mathbb{E}X = \mathbb{E}X.$$

Для обоснования (v) достаточно заметить, что

$$\mathbb{E}(XZ|B_k) = \frac{\mathbb{E}(XZI_{B_k})}{P(B_k)} = c_k\frac{\mathbb{E}(XI_{B_k})}{P(B_k)} = c_k\mathbb{E}(X|B_k).$$

Теорема доказана. □

Наиболее типична ситуация, когда разбиение $\mathcal{B}$ появляется посредством некоторой случайной величины

$$Y = \sum_{k=1}^{n} y_k I_{B_k},$$

где y_k — различные числа и $P(B_k) > 0$. В этом случае $B_k = \{\omega\colon Y(\omega) = y_k\}$ и условное математическое ожидание $\mathbb{E}(X|\mathcal{B})$ обозначают символом $\mathbb{E}(X|Y)$ и называют условным математическим ожиданием случайной величины X относительно случайной величины Y. Несложно предъявить функцию f (это можно сделать несколькими способами) такую, что

$$\mathbb{E}(X|Y)(\omega) = f(Y(\omega)).$$

Например можно взять $f(y_k) = \mathbb{E}(X|Y = y_k)$, а в остальных точках задать f как угодно. Можно заметить, что обязательно $f(y_k) = \mathbb{E}(X|Y = y_k)$.

В случае условного ожидания относительно случайной величины, свойства условного ожидание переформулируются следующим образом.

Теорема 139. (i) *(линейность)* $\mathbb{E}(\alpha X + \beta Y|Z) = \alpha\mathbb{E}(X|Z) + \beta\mathbb{E}(Y|Z)$,
(ii) *(монотонность)* $X \le Y$ *п.н.* $\Rightarrow \mathbb{E}(X|Z) \le \mathbb{E}(Y|Z)$,
(iii) *(аналог формулы полной вероятности)* $\mathbb{E}\big(\mathbb{E}(X|Y)\big) = \mathbb{E}X$,
(iv) *(независимость) если случайные величины X и Y независимы, то* $\mathbb{E}(X|Y) = \mathbb{E}X$.
(v) *для всякой случайной величины* $Z = g(Y)$ *выполнено* $\mathbb{E}(ZX|Y) = Z\mathbb{E}(X|Y)$.

Лемма 140. *Для условного математического ожидания выполнено*

$$\mathbb{E}(g(Y)X) = \mathbb{E}\Big(g(Y)\mathbb{E}(X|Y)\Big)$$

для произвольной функции g. Кроме того, если для какой-то случайной величины вида $Z = f(Y)$ выполнено

$$\mathbb{E}(g(Y)X) = \mathbb{E}(g(Y)Z)$$

для произвольной функции g, то $Z = \mathbb{E}(X|Y)$ п.н.

Доказательство. По уже доказанному

$$\mathbb{E}\Big(g(Y)\mathbb{E}(X|Y)\Big) = \mathbb{E}\Big(\mathbb{E}(g(Y)X|Y)\Big) = \mathbb{E}(g(Y)X).$$

Наоборот, если $Z = f(Y)$ и обладает указанным свойством, то

$$\mathbb{E}\Big(g(Y)\mathbb{E}(X|Y)\Big) = \mathbb{E}(g(Y)Z)$$

для произвольной g. Т.к. $\mathbb{E}(X|Y)$ также имеет вид $h(Y)$, то, взяв $g = f - h$, получаем

$$\mathbb{E}|\mathbb{E}(X|Y) - Z|^2 = 0,$$

что дает равенство $Z = \mathbb{E}(X|Y)$ почти наверное. □

Следующее утверждение раскрывает геометрический смысл условного математического ожидания.

Предложение 141. *Условное математическое ожидание $\mathbb{E}(X|Y)$ среди всех случайных величин вида $g(Y)$ является лучшим среднеквадратическим приближением для X, т. е.*

$$\min_{Z:\ Z=g(Y)} \mathbb{E}|X - Z|^2 = \mathbb{E}|X - \mathbb{E}(X|Y)|^2.$$

Доказательство. Пусть $Z = g(Y)$. Так как по предыдущей лемме

$$\mathbb{E}\big[(X - \mathbb{E}(X|Y))(\mathbb{E}(X|Y) - Z) = 0\big],$$

то

$$\mathbb{E}|X - Z|^2 = \mathbb{E}|(X - \mathbb{E}(X|Y)) + (\mathbb{E}(X|Y) - Z)|^2 = \\ = \mathbb{E}|X - \mathbb{E}(X|Y)|^2 + \mathbb{E}|\mathbb{E}(X|Y) - Z|^2 \geq \mathbb{E}|X - \mathbb{E}(X|Y)|^2.$$

□

Таким образом, с геометрической точки зрения условное математическое ожидание является проекцией X на пространство случайных величин вида $g(Y)$ и полностью характеризуется тем свойством, что вектор $X - \mathbb{E}(X|Y)$ ортогонален указанному пространству, что записывается с помощью равенства $\mathbb{E}(Xg(Y)) = \mathbb{E}\big(\mathbb{E}(X|Y)g(Y)\big)$ для произвольной случайной величины $g(Y)$.

39. Условные математические ожидания: общий случай.

Теперь мы можем сформулировать определение условного математического ожидания $\mathbb{E}(X|Y)$ уже без предположения о дискретности распределения Y.

Случайная величина вида $f(Y)$ называется **условным математическим ожиданием** случайной величины X (обладающей математическим ожиданием) относительно случайной величины Y и обозначается $\mathbb{E}(X|Y)$, если

$$\mathbb{E}(Xg(Y)) = \mathbb{E}\Big(\mathbb{E}(X|Y)g(Y)\Big)$$

для всех ограниченных случайных величин вида $g(Y)$. Любые две случайные величины, удовлетворяющие этому определению, почти наверное совпадают.

Функцию $f(y)$ **обозначают** $f(y) = \mathbb{E}(X|Y = y)$.

Предложение 142. *Сформулированные ранее свойства* (i)–(v) *условных математических ожиданий относительно разбиений и дискретных величин остаются верны и в общем случае.*

Доказательство. Линейность ясна из определения и линейности математического ожидания. Для доказательства монотонности достаточно показать (в силу линейности), что из неравенства $X \geq 0$ следует $\mathbb{E}(X|Y) \geq 0$ почти наверное. Для этого в определении положим $g(Y) = 1 - \operatorname{sign}(\mathbb{E}(X|Y)) \geq 0$. Тогда $\mathbb{E}(X|Y) - |\mathbb{E}(X|Y)| \leq 0$, но

$$\mathbb{E}\big[\mathbb{E}(X|Y) - |\mathbb{E}(X|Y)|\big] = \mathbb{E}\big[X\big(1 - \operatorname{sign}(\mathbb{E}(X|Y))\big)\big] \geq 0.$$

Значит $\mathbb{E}(X|Y) - |\mathbb{E}(X|Y)| = 0$.

Равенство $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}X$ является частным случаем определения (надо взять $g \equiv 1$).

Если X и Y независимы, то

$$\mathbb{E}(Xg(Y)) = [\mathbb{E}X] \cdot [\mathbb{E}g(Y)] = \mathbb{E}([\mathbb{E}X]g(Y)).$$

Если $Z = h(Y)$ (с ограниченной h), то подстановкой в определение проверяется, что $Z\mathbb{E}(X|Y)$ является условным математическим ожиданием ZX относительно Y. Случай общей функции h получается предельным переходом. □

40. Условные математические ожидания в конкретных ситуациях.

Предложение 143. *Предположим, что распределение (X, Y) задано совместной плотностью $\varrho_{X,Y}(x, y)$. Тогда*

$$\mathbb{E}(\Phi(X, Y)|Y = y) = \int_{-\infty}^{\infty} \Phi(x, y) \frac{\varrho_{X,Y}(x, y)}{\varrho_Y(y)}\, dx.$$

Доказательство. Имеет место цепочка равенств:

$$\mathbb{E}\big(\Phi(X, Y)g(Y)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi(x, y)g(y)\varrho_{X,Y}(x, y)\, dx\, dy$$
$$= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{\infty} \Phi(x, y)\frac{\varrho_{X,Y}(x, y)}{\varrho_Y(y)}\, dx\right)\varrho_Y(y)\, dy.$$

□

Функцию

$$\varrho_{X|Y}(x|y) = \frac{\varrho_{X,Y}(x, y)}{\varrho_Y(y)}$$

называют **условной плотностью** X относительно Y (условимся, что $\varrho_{X|Y}(x|y) = 0$ в точках y, в которых $\varrho_Y(y) = 0$). Таким образом, верны равенства:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x\varrho_{X|Y}(x|y)\, dx, \quad \varrho_{X,Y}(x, y) = \varrho_{X|Y}(x|y)\varrho_Y(y),$$

последнее из которых является аналогом хорошо знакомого нам правила умножения вероятностей $P(A \bigcap B) = P(A|B)P(B)$.

Условной вероятностью $P(A|Y)$ называют случайную величину $\mathbb{E}(I_A|Y)$, а случайную величину $\mathbb{E}(I_A|Y = y)$ в этом случае обозначают через $P(A|Y = y)$.

Пример 144. Пусть X и Y — такие случайные величины, что существует измеримая функция $\varrho(x|y)$, для которой выполнено

$$P(X \in B|Y = y) = \int_B \varrho(x|y)\, dx.$$

В этом случае

$$\mathbb{E}(h(X)|Y = y) = \int_{\mathbb{R}} h(x)\varrho(x|y)\, dx.$$

Заметим, что для произвольной ограниченной функции h

$$\mathbb{E}h(X) = \mathbb{E}\big(\mathbb{E}(h(X)|Y)\big) = \int_{\mathbb{R}} h(x)\mathbb{E}\varrho(x|Y)\,dx.$$

Тем самым $\varrho_X(x) = \mathbb{E}\varrho(x|Y)$. Для произвольных ограниченных функций f, g выполнено

$$\mathbb{E}\big[f(X)\mathbb{E}(g(Y)|X)\big] = \mathbb{E}\big[f(X)g(Y)\big] = \mathbb{E}\big[g(Y)\mathbb{E}(f(X)|Y)\big].$$

Левая часть тождества равна

$$\int_{\mathbb{R}} f(x)\mathbb{E}(g(Y)|X = x)\varrho_X(x)\,dx,$$

а правая часть равна

$$\mathbb{E}\Big[g(Y)\int_{\mathbb{R}} f(x)\varrho(x|Y)\,dx\Big] = \int_{\mathbb{R}} f(x)\mathbb{E}\big[g(Y)\varrho(x|Y)\big]\,dx.$$

В силу произвольности f получаем следующую формулу Байеса

$$\mathbb{E}(g(Y)|X = x) = \frac{\mathbb{E}\big[g(Y)\varrho(x|Y)\big]}{\mathbb{E}\varrho(x|Y)}.$$

Пусть теперь Y принимает два значения 0 и 1 с вероятностями p и q соответственно. Тогда

$$P(Y = 0|X = x) = \frac{p\varrho(x|0)}{p\varrho(x|0) + q\varrho(x|1)}, \quad P(Y = 1|X = x) = \frac{q\varrho(x|1)}{p\varrho(x|0) + q\varrho(x|1)}.$$

Предложение 145. *Пусть вектор (ξ, η) имеет нормальное распределение и $\mathbb{D}\eta > 0$. Тогда*

$$\mathbb{E}(\xi|\eta) = \mathbb{E}\xi + \frac{cov(\xi, \eta)}{\mathbb{D}\eta}(\eta - \mathbb{E}\eta).$$

Доказательство. Перейдем к случайным величинам $X = \xi - \mathbb{E}\xi, Y = \eta - \mathbb{E}\eta$ и применим ортогонализацию:

$$X = Z + \frac{cov(\xi, \eta)}{\mathbb{D}\eta}Y,$$

где Z нормальная случайная величина с нулевым средним, независимая с Y. Тогда

$$\mathbb{E}(X|Y) = \frac{cov(\xi, \eta)}{\mathbb{D}\eta}Y.$$

Более общим образом,

$$\mathbb{E}(f(X)|Y = y) = \mathbb{E}\Big(f\Big(Z + \frac{cov(\xi, \eta)}{\mathbb{D}\eta}Y\Big)|Y = y\Big) = \int_{\mathbb{R}} f\Big(z + \frac{cov(\xi, \eta)}{\mathbb{D}\eta}y\Big)\frac{\varrho_{Z,Y}(z, y)}{\varrho_Y(y)}\,dz$$
$$= \int_{\mathbb{R}} f\Big(z + \frac{cov(\xi, \eta)}{\mathbb{D}\eta}y\Big)\varrho_Z(z)\,dz = \int_{\mathbb{R}} f(z)\varrho_Z\Big(z - \frac{cov(\xi, \eta)}{\mathbb{D}\eta}y\Big)\,dz.$$

Т.е. условная плотность есть плотность нормального распределения со средним $\frac{cov(\xi,\eta)}{\mathbb{D}\eta}y$ и дисперсией $\frac{\mathbb{D}\xi\mathbb{D}\eta - [cov(\xi,\eta)]^2}{\mathbb{D}\eta}$. $\square$