Конспект лекций по терверу.

Пешехонов Иван. БПМИ195

Дата последней компиляции: 13.12.2020 17:55

Содержание

1 Дискретное вероятностное пространство.

1.1 Вероятностое пространство, события, вероятностная мера, вероятность.

Определение 1. Пусть задано некоторое множество возможных исходов $\Omega = \{\omega_1, \ldots, \omega_n\}$. Это множество называется *множеством элементарных исходов*. Всякое подмножество $A \subseteq \Omega$ называют *событием*.

Функцию $P \colon 2^\Omega \to [0, 1]$, удовлетворяющую следующим свойствам:

- $P(\Omega) = 1$
- $A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B)$ (аддитивность)

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

Для всякого $\omega \in \Omega$ определена вероятностная мера: $P(\{\omega\}) = p_\omega$. Из определения вероятностной меры следует, что $0 \leqslant p_\omega \leqslant 1$. Тогда

$$\sum_{\omega \in \Omega} p_\omega = \sum_{k=1}^{n} P(\{\omega_k\}) = P(\Omega) = 1$$

И тогда вероятность произвольного события A вычисляется по формуле

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega$$

Если все элементарные исходы равновозможны, то $p_{\omega_1} = \ldots = p_{\omega_n} = \dfrac{1}{n}$. В этом случае вероятность события A равна

$$P(A) = \sum_{\omega \in A} p_\omega = \sum_{k=1}^{|A|} \frac{1}{n} = \frac{|A|}{n}$$

Задача 1.1 (Вероятностный алгоритм проверки числа на простоту.)**.** Дано натуральное число $N > 1$. Хотим знать, является ли оно простым быстрее чем за $O(\sqrt{N})$.

Решение: По малой теореме Ферма:
если N — простое число, то для всякого взаимнопростого с N числа b (НОД$(N, b) = 1$) число $b^{N-1} - 1$ делится на N($\iff b^{N-1} \equiv 1 \mod N$).
Предложим следующий алгоритм: выбираем число b случайным образом из промежутка $[2, N-1]$. Если НОД$(b, N) \neq 1$, то N, очевидно, составное. Если НОД$(b, N) = 1$, но $b^{N-1}-1$ не делится на N, то N не простое по малой теореме Ферма. Во всех остальных случаях будем считать, что N простое. Заметим, что пока в нашем алгоритме используются только рандомный выбор (считаем, что можно сделать за константу), проверка делимости (тоже за константу), и подсчёт НОДа (ассимптотика алгоритма Евклида равна $O(\log \min(N, b))$ — быстрее чем полином). Будем говорить, что "число N проходит тест по основанию b", если наш алгоритм определяет число N как простое при выборе случайного b взаимно простого с N.

Предположим теперь, что есть хотя бы одно число a, такое что НОД$(a, N) = 1$, но a^{N-1} не делится на N. Посчитаем вероятность, с которой наш алгоритм выдаст, что N — простое (т.е. посчитаем вероятность ошибки алгоритма):
Пусть $\mathbb{Z}_N^*$ — группа всех взаимнопростых с N чисел из промежутка $[1, N-1]$. Если N проходит тест по основанию $b \in \mathbb{Z}_N^*$, то по основанию ab (где $a \in \mathbb{Z}_N^*$) оно тест не пройдёт, т.к.

$$\begin{cases} (ab)^{N-1} \equiv 1 \mod N \\ (b^{-1})^{N-1} \equiv 1 \mod N \end{cases} \implies a^{N-1} \equiv 1 \mod N$$

Таким образом, для всякого основания b, по которому N проходит тест, существует основание ab, на котором N тест не проходит, а значит оснований, на которых N не проходит тест не меньше чем тех, на которых N тест проходит, а значит вероятность ошибки нашего алгоритма $\dfrac{1}{2}$. Выбирая число b случайным образом k раз можно снизить вероятность ошибки алгоритма до $\dfrac{1}{2^k}$.

1.2 Формула включений и исключений.

Пусть Ω конечное множество элементарных исходов, а $A_1, \cdots, A_n$ — произвольные события.

Предложение. Верно следующее равенство:

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

Пример. Пусть у нас есть множества A_1, A_2, A_3. Распишем для них формулу включений и исключений, чтобы разобраться в её записи. Напомню, что запись $\sum\limits_{i<j}$ эквивалентна записи $\sum\limits_{j=1}^{n}\sum\limits_{i=2}^{j}$.

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

Доказательство. Индукция по n:
База: $n = 2$

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

Предположение индукции: пусть выполнено для $n = k$ множеств.
Шаг: проверим для $n = k + 1$ множеств

$$P(A_1 \cup \ldots \cup A_{k+1}) = \left| B = \bigcup_{i=1}^{k} A_i \right| = P(B \cup A_{k+1}) = P(B) + P(A_{k+1}) - P(B \cap A_{k+1}) = \Diamond$$

$$P(B) = P(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{j=1}^{k} (-1)^{j-1} \sum_{i_1 < \ldots < i_j} P(A_{i_1} \cap \ldots \cap A_{i_j})$$

$$\begin{aligned} P(B \cap A_{k+1}) &= P((A_1 \cup A_2 \cup \ldots \cup A_k) \cap A_{k+1}) = P((A_1 \cap A_{k+1}) \cup \ldots \cup (A_k \cap A_{k+1})) = \\ &= \sum_{i=1}^{k} (-1)^{i-1} \sum_{j_1 < \ldots < j_i} P(A_{j_1} \cap \ldots \cap A_{j_i} \cap A_{k+1}) \end{aligned}$$

$$\begin{aligned} \Diamond &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{i_1 < \ldots < i_k} P(A_{i_1} \cap \ldots \cap A_{i_k}) + P(A_{k+1}) - \sum_{i=1}^{k} (-1)^{i-1} \sum_{j_1 < \ldots < j_i} P(A_{j_1} \cap \ldots \cap A_{j_i} \cap A_{k+1}) = \\ &= \sum_{j=1}^{k+1} (-1)^{j-1} \sum_{i_1 < \ldots < i_j} P(A_{i_1} \cap \ldots \cap A_{i_j}) \end{aligned}$$

■

1.3 Парадокс распределения подарков.

n человек решили подарить друг другу подарки по следующей схеме: каждый человек купил один подарок и положил его в мешок. После этого все люди одновременно сунули руки в мешок и каждый вытащил себе наугад ровно один подарок.

1. Какова вероятность, что каждый человек вытащил подарок, который сам и принёс?

2. Какова вероятность того, что никто не вытащил подарок, который сам принёс?

Решение: Занумеруем все подарки числами от 1 до n. Таким образом мы можем думать о подарках как о перестановках на n элементах, а о событии, при котором каждый человек вытащил конкретно тот подарок, который принёс — как о тождественной перестановке. Пусть событие A — человек вытащил подарок, который сам принёс. Тогда один из n подарков фиксирован, а все остальные переставляются случайным образом. Т.е. $|A| = (n-1)!$, что соответствует перестановкам остальных подарков, а $P(A) = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$. Очевидно, что при достаточных n эта вероятность стремится к 0.

Пусть теперь A_k — событие, соответствующее тому, что k-ый подарок попал к человеку, который его принёс. Тогда $\bigcup\limits_{k=1}^{n} A_k$ — событие, соответсвующее тому, что хотя бы один подарок попал к человеку, который его принёс. Тогда по формуле включений и исключений:

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

Заметим теперь, что вероятность $P(A_{i_1} \cap \ldots \cap A_{i_k})$ на русский язык переводится как "i_1 подарок достался своему человеку, и i_2 подарок достался своему человеку, и …, и i_k подарок достался своему человеку". Т.е. k подарков досталось тем, кто их принёс. Т.е. k подарков фиксированы, а остальные могут быть переставлены как угодно. Таким образом вероястность события $A_{i_1} \cap \ldots \cap A_{i_k}$ равна $\dfrac{(n-k)!}{n!}$.

Теперь заметим, что эта внутренняя сумма занимается только выбором таких k человек, которые вытащат подарки, которые сами и принесли, а мы знаем, что выбрать таких человек можно $\dbinom{n}{k}$ способами.

Таким образом $\sum\limits_{i_1<\ldots<i_k} P(A_{i_1} \cap \ldots \cap A_{i_k}) = \dbinom{n}{k}\dfrac{(n-k)!}{n!}$. А значит

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}(-1)^{k-1} \sum_{i_1<\ldots<i_k} P(A_{i_1} \cap \ldots \cap A_{i_k}) = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{(n-k)!}{n!}$$

Раскроем теперь $\dbinom{n}{k}$ и упростим:

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{(n-k)!}{n!} = \sum_{k=1}^{n}\frac{(-1)^{k-1}}{k!}$$

Таким образом мы посчитали вероятность события, соответсвующего тому, что хотя бы один подарок попал к человеку, который его принёс. Тогда вероятность того, что не один человек не вытащил свой подарок равна:

$$1 - P\left(\bigcup_{k=1}^{n} A_k\right) = 1 - \sum_{k=1}^{n}\frac{(-1)^{k-1}}{k!} = 1 - 1 + \frac{1}{2} - \frac{1}{3!} + \ldots + \frac{(-1)^{n-1}}{n!} \to \frac{1}{e}$$

1.4 Доказательства существования и задача о конференции.

Пусть $A_1, \ldots, A_k$ — произвольные события.

Хотим проверить, что $P\left(\bigcap\limits_{k=1}^{n} A_k\right) \neq 0$.

Пусть B_k — событие, противоположное к A_k. Тогда

$$P\left(\bigcap_{k=1}^{n} A_k\right) = 1 - P\left(\bigcup_{k=1}^{n} B_k\right) \geqslant 1 - \sum_{k=1}^{n} P(B_k)$$

Т.к. для произвольного набора событий $A_1, \ldots, A_k$ справедлива оценка

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

Следует из формулы включений и исключений путём отбрасывания вычитаний попарных пересечений событий и т.д.

Если $\sum\limits_{k=1}^{n} P(B_k) < 1$, то $\sum\limits_{k=1}^{n} P\left(\bigcap\limits_{k=1}^{n} A_k\right) > 0$. Т.к. $P(B_k) = 1 - P(A_k)$, то

$$\sum_{k=1}^{n} P(B_k) < 1 \iff \left(\sum_{k=1}^{n} P(B_k) = \sum_{k=1}^{n} (1 - P(A_k)) = n - \sum_{k=1}^{n} P(A_k) < 1 \implies \sum_{k=1}^{n} P(A_k) > n - 1\right)$$

Так, например если $\forall k$ выполнено $P(A_k) > 1 - \dfrac{1}{n}$, то $\sum\limits_{k=1}^{n} P(A_k) > n - 1$, а значит пересечении A_k имеет не нулевую вероятность, а следовательно не пусто.

Замечание.

$$P(\varnothing) = 1 - P(\Omega) = 1 - 1 = 0$$

С другой стороны

$$P(\varnothing) = \frac{|\varnothing|}{|\Omega|} = \frac{0}{|\Omega|} = 0$$

Задача 1.2 (Задача о конференции)**.** В лаборатории работаю специалисты по 60 направлениям. По каждому направлению ровно 7 человек (каждый человек может быть специалистом по нескольким направлениям). Задача: отправить всех специалистов на две конференции: одна в Канаде, другая в Австралии. На каждой конференции должен быть специалист по каждому направлению.

Решение: Броском монеты будем для каждого специалиста определять его конференцию. Пусть событие A_k соответствует тому, что по k-ому направлению есть специалист на обеих конференциях (т.е. $\bigcap\limits_{k=1}^{60} A_k \neq \varnothing$). Тогда

$$P(A_k) = 1 - \frac{2}{2^7}$$

где 2^7 — количество способов раздать конференцию каждому из 7 специалистов, а 2 — число способов, при которых все специалисты уехали либо в Австралию, либо в Канаду. Тогда

$$P(A_k) = 1 - \frac{1}{2^6} > 1 - \frac{1}{60}$$

Видно, что $P(A_k) > 1 - \dfrac{1}{n}$ при $n = 60$, а значит пересечение A_k не пусто. (Мы не предъявляем способ разослать специалистов по конференциям, мы лишь доказываем, что он существует)

1.5 Бесконечное множество элементарных исходов и счётная аддитивность.

Пусть Ω — бесконечное множество.

Пример. $\Omega = \mathbb{N}$. Определим функцию P, удовлетворяющую свойствам вероятностной меры:
пусть $P(A) = 0$, если A — конечно, и $P(A) = \mathbb{N}$. Тогда

- $P(\Omega) = P(\mathbb{N}) = 1$
- $A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B) = 0$

В чём проблема? А вот в чём:

$$\sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{k \in \mathbb{N}} P(\{k\}) = 0 \neq 1 = P(\mathbb{N}) = P(\Omega)$$

Зададим вероятностную меру следующим множеством: $\{p_{w_j} \mid w_j \in \Omega\}$, причём $\forall\, w_j \in \Omega\colon p_{w_j} \geqslant 0$, $\sum\limits_{j} p_{w_j} = 1$. Тогда вероястность каждого события A считается по формуле

$$P(A) = \sum_{w_j \in A} p_{w_j}$$

Заметим, что для вероятностной меры, заданой таким определением выполнено следующее свойство:

Определение 2. Пусть $A_1, A_2, \dots,$ — произвольный, не более чем счётный набор попарно пересекающихся событий. Тогда

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

Такое свойство мы будем называть *счётной аддитивностью.*

Лемма 1.1 (Ликбез в матан)**.** Пусть $a_{n,m}$ — последовательность положительных чисел. Тогда

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

Где $\sigma\colon \mathbb{N} \to \mathbb{N}^2$ — произвольная перестановка.

Доказательство. Представим последовательность следующей таблицей:

$$\begin{array}{|c|}\hline a_{1,1}\ a_{1,2}\ \dots\ a_{1,n_1} \\ \hline a_{2,1}\ a_{2,2}\ \dots\ a_{2,n_2} \\ \hline \vdots \\ \hline \end{array}$$

Тогда сумму элементов в этой таблице мы можем считать построчно: прибавляя к уже накопленной сумме каждый следующий элемент. Или же мы можем начать считать эту сумму в рандомном порядке, но при этом каждый элемент учитывая только один раз. При этом независимо от того, как именно мы будем складывать, результат не должен измениться. В этом и состоит всё утверждение.

Формально: зафиксируем перестановку $\sigma\colon \mathbb{N} \to \mathbb{N}^2$ — наш способ выбирать элементы из таблицы в рандомном порядке. Тогда для произвольных конечных M, N выполено

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

Тогда устремляя M и N к бесконечности получаем, что

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

С другой стороны для произвольного конечно J выполнено

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

Устремляя J к бесконечности получаем оценку в другую сторону.

Из обеих оценок следует равенстсво. ∎

Предложение. Мы не налажали в определении счётной аддитивности.

Доказательство. Перепишем утверждение о счётной аддитивности следующим образом:

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \iff \sum_{w_i \in \bigcup\limits_{k=1}^{\infty} A_k} p_{w_j} = \sum_{k=1}^{\infty}\sum_{w_j \in A_k} p_{w_j}$$

Справедливость полученного утверждения следует из доказанной леммы. ∎

1.6 Условная вероятность, формула полной вероятности и формула Байеса.

Определение 3. Пусть $P(B) > 0$. Вероятность события A при условии события B, определяемая как

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

называется *условной вероятностью.*

При фиксированном B функция $P(\cdot|B)$ является новой вероятностной мерой.

Определение 4. *Правилом произведения* называется следующее равенство:

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

Теорема 1.2 (Формула полной вероятности.). *Пусть $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$ и $A_i \cap A_j = \varnothing\ \ \forall\, i \neq j$. Пусть так же $\forall\, i\colon A_i \neq \varnothing$. Тогда для всякого события B верно*

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

Доказательство. Имеем

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

■

Теорема 1.3 (Формула Байеса.). *Пусть A, B — два непустых события. Тогда верна формула Байеса:*

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

Доказательство. Имеем

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

■

1.7 Независимые события. Попарная независимость и независимость в совокупности.

Определение 5. События A и B называются *независимыми*, если выполнено

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

иначе события называются *зависимыми.*

Определение 6. События $A_1, A_2, \ldots, A_n$ называются *независимыми в совокупности*, если для некоторого $k \in \{2, \ldots, n\}$ и произвольных $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ выполнено

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

Предложение. Независимость в совокупности не совпадает с попарность независимостью.

Доказательство. Будем два раза подкидывать монету. Пусть событие A — при первом бросании выпал орёл; событие B — при втором бросании выпал орёл; событие C — орёл выпал ровно один раз. Очевидно, что эти события попарно независимы. Однако если при обоих бросках выпали орлы, то событие C становится невозможным, а значит $P(A \cap B \cap C) = 0 \neq P(A) \cdot P(B) \cdot P(C)$. ■

1.8 Задача о билетах к экзамену.

Задача 1.3. Программа коллоквиума состоит из N билетов, а студент выучил только n. На колке студенты по очереди подходят и тянут билет. Зависит ли вероятность вытянуть "хороший" билет от места в очереди?

Решение: Пусть студент стоит на $k+1$ месте в очереди и пусть событие A — студент вытянул "хороший" билет. Положим событие A_j — первые k студентов вытянули j "хороших" билетов; Событие B — вытянуть выученный билет. Тогда очевидно, что

$$P(B) = \sum_j P(B|A_j)P(A_j)$$

Заметим, что вероятность $P(B|A_j)$ считается очень просто: действительно, когда подойдёт наша очередь, Останется всего $n-j$ подходящих нам билетов, а всего билетов останется $N-k$. Таким образом $P(B|A_j) = \dfrac{n-j}{N-k}$.

Найдём теперь $P(A_j)$: заметим, что до нашей очереди билеты были выбраны $\dbinom{N}{k}$ способами. При этом известно, что было вытянуто j выученных билетов, это можно сделать $\dbinom{n}{j}$ способами. Значит, из невыученных билетов (их $N-n$) осталось вытянуть $k-j$ билетов. Это можно сделать $\dbinom{N-n}{k-j}$ способами. Итого имеем:

$$P(A_j) = \frac{\binom{n}{j}\cdot\binom{N-n}{k-j}}{\binom{N}{k}}$$

Подставим в формулу полной вероятности:

$$P(B) = \sum_j P(B|A_j)P(A_j) = \sum_j \frac{n-j}{N-k}\cdot\frac{\binom{n}{j}\cdot\binom{N-n}{k-j}}{\binom{N}{k}} = \sum_j \frac{n-j}{N-k}\cdot\frac{\frac{n!}{(n-j)!j!}\cdot\binom{N-n}{k-j}}{\frac{N!}{(N-k)!k!}} =$$

$$= \sum_j \binom{N-n}{k-j}\cdot\frac{n-j}{N-k}\cdot\frac{n!}{(n-j)!j!}\cdot\frac{(N-k)!k!}{N!} = \sum_j \binom{N-n}{k-j}\frac{n!}{(n-j-1)!j!}\cdot\frac{(N-k-1)!k!}{N!} =$$

$$= \frac{n}{N}\sum_j \binom{N-n}{k-j}\frac{(n-1)!}{(n-j-1)!j!}\cdot\frac{(N-k-1)!k!}{(N-1)!} = \frac{n}{N}\sum_j \underbrace{\binom{N-n}{k-j}\cdot\binom{n-1}{j}\cdot\left[\binom{N-1}{k}\right]^{-1}}_{P(A_j)\text{ для меньшего числа билетов}} = \frac{n}{N}$$

Т.к. A_j попарно не пересекаются, и в объединении дают Ω, то $\sum P(A_j) = 1$ по свойству вероятностной меры.

С другой стороны, если бы билеты раздавались случайным образом, то очевидно, что вероятность получить среди N билетов один из n выученных равна $\dfrac{n}{N}$. Отсюда делаем вывод, что позиция в очереди никак не влияет на вероятность вытянуть выученный билет, и вообще нет разницы просто с рандомной раздачей билетов.

1.9 Задача о сумасшедшей старушке.

Задача 1.4. На посадку в самолёт стоят $N \geqslant 2$ пассажиров, среди которых есть сумасшедшая старушка. Старушка расталкивает остальных пассажиров и садится в самолёт на произвольное место. Затем пассажиры, когда заходят в самолёт, садятся на своё место, если оно свободно, если же место занято, то пасажир садится на рандомное место. Какова вероятность того, что последний пассажир сядет на своё место?

Решение Обозначим вероятность N-ого пассажира сесть на своё место за P_N. Индукция по N:

База: для $N=2$ вероятность, очевидно, $\dfrac{1}{2}$.

Предположение: пусть для $N=k$ верно, что $P_N = \dfrac{1}{2}$.

Шаг: докажем для $N+1$: пусть событие B — последний пассажир сел на своё место; событие A_j — старушка села на место j-ого пассажира. Тогда по формуле полной вероятности:

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

Заметим, что $P(B)$ есть то же самое, что и P_{N+1}. Кроме того, т.к. в самолёт входят $N+1$ пассажир, а старушка садится на любое место, то для любого i верно $P(A_i) = \dfrac{1}{N+1}$. Распишем теперь $P(B)$ следующим образом:

$$P_{N+1} = P(B) = \sum_{i=1}^{N+1} P(B|A_i) \cdot P(A_i) = P(B|A_1)P(A_1) + \sum_{i=2}^{N} P(B|A_i) \cdot P(A_i) + P(B|A_{N+1})P(A_{N+1})$$

Разбираемся: событие A_1 есть событие, при котором старушка садится на своё место (т.к. её саму можно считать первым пассажиром), тогда $P(B|A_1) = 1$, т.к. если старушка села на своё место, то все последующие пассажиры так же сядут на своё место, включая последнего. С другой стороны, если произошло событие A_{N+1}, т.е. старушка села на место последнего пассажира, то, очевидно, $P(B|A_{N+1}) = 0$. Для всех остальных i по предположению индукции верно: $P(B|A_i)P(A_i) = \dfrac{1}{2}$.

$$P_{N+1} = P(B|A_1)P(A_1) + \sum_{i=2}^{N} P(B|A_i) \cdot P(A_i) + P(B|A_{N+1})P(A_{N+1}) = \frac{1}{N+1}\left(1 + \sum_{i=2}^{N}\frac{1}{2} + 0\right) = \frac{1}{N+1}\left(1 + \frac{N-1}{2}\right) =$$

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

1.10 Парадокс Байеса.

Имеется тест для диагностики некоторого редкого заболевания. Известно, что доля больных этим заболеванием равна 0.001. Если человек болен, то тест даёт положительный результат с вероятностью 0.99. Если человек здоров, то тест даёт положительный результат с вероятностью 0.01. Требуется найти вероятность ложноположительного результата. *Решение:* пусть события T_- и T_+ — тест дал отрицательный и положительный результаты соответственно; события Z_- и Z_+ — человек здоров или болен.
Найдём вероятность положительного теста по формуле полной вероятности:

$$P(T_+) = P(T_+|Z_+)P(Z_+) + P(T_+|Z_-)P(Z_-) = 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098$$

Теперь по формуле Байеса найдём вероятность того что человек здоров при условии, что тест дал положительный результат:

$$P(Z_-|T_+) = \frac{P(T_+|Z_-)P(Z_-)}{P(T_+)} = \frac{0.01 \cdot 0.999}{0.01098} \to 1$$

Таким образом видно, что из-за редкости заболевания тест почти гарантированно даст ложноположительный результат.

1.11 Парадокс Монти Холла.

Теперь мы учавствуем в игре, в которой нужно выбрать одну из трёх дверей. За одной из дверей находится автомобиль, а за двумя другими — козы (а что, я бы и от козы не отказался...). После того как мы выбираем одну дверь (пусть будет первая), ведущий равновероятно открывает одну из оставшихся дверей (пусть будет третью), за которой находится коза (важно: если мы изначально выбрали дверь с козой, то ведущий просто откроет другую дверь с козой. Атомобиль он нам не покажет). После этого мы имеем возможность изменить свой выбор на вторую дверь. Следует ли нам это делать?

Решение: Пусть событие A_i — автомобиль находится за i-ой дверью. Очевидно, что $\forall\ i\ P(A_i) = \dfrac{1}{3}$. Пусть событие B — ведущий открыл 3-ю дверь. Тогда если верно событие A_1, то $P(B|A_1) = \dfrac{1}{2}$. Если же событие A_1 не верно, а автомобиль находится, например, за 3 дверью, то $P(B|A_2) = 1$, $P(B|A_3) = 0$. Аналогично для события A_2. Тогда, по формуле Байеса найдём вероятность того, что автомобиль находится за 2 дверью, при условии, что ведущий открыл третью дверь:

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B)} = \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} = \frac{1/3}{1/6 + 1/3 + 0} = \frac{2}{3}$$

С другой стороны изначально шанс угадать, где находится автомобиль, был равен $\frac{1}{3}$, а значит, согласившись изменить выбор двери, мы повысим шанс получить автомобиль.

2 Случайные величины на дискретном вероятностном пространстве.

2.1 Определение случайной величины и её распределение.

Определение 7. *Случайной величиной* на дискретном вероятностном пространстве Ω называют произвольную функцию $X\colon \Omega \to \mathbb{R}$. Если X — случайная величина на дискретном вероятностом пространстве, то множество её значений не более чем счётно (очевидно, т.к. мы для каждого исхода $\omega \in \Omega$ зафиксируем его образ, а таких исходов не более чем счётно).

Пусть X — случайная величина и $x_1, \ldots, x_n, \ldots$ — все её значения.

Определение 8. *Распределением* случайной величины X называется новая вероятностная мера μ_X на пространстве $\{x_1, \ldots, x_n, \ldots\}$, для которой $\mu_X(\{x_j\}) = P(\{w :\ X(w) = x_j\})$. Т.е. вероятность того, что случайная величина примет значение x_j мы считаем как вероятность такого события, элементами которого являются элементарные исходы из Ω, переходящие в x_j.

Положим событие $A_j = \{w\colon X(w) = x_j\}$. Очевидно, что $\forall\ i \neq j \implies A_i \cap A_j = \varnothing$, и $\bigcup\limits_j A_j = \Omega$, а значит μ действительно является вероятностной мерой.

Пусть $p_j = P(w\colon X(w) = x_j)$. Тогда распределение X можно задать таблицей:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
μ_X	p_1	p_2	$\cdots$	p_n	$\cdots$

2.2 Примеры случайных дискретных величин.

Бернуллиевская случайная величина:

Таблица распределение бернулиевской случайной величины имеет вид

X	0	1
P(X)	q	p

Величина моделирует событие с двумя исходами, вероятность одного из которых равна p. Такая случайная величина обычно появляется, как "индикатор" какого-то события A:

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

Тогда $P(A) = p$ и $q = 1 - P(A)$.

Схема Бернулли.

Ω — все возможные наборы длины N из нулей и единиц. Вероятностная мера P задаётся следующим образом: если исход содержит k единиц, то вероятность этого исхода равна $p^k q^{N-k}$, где $p + q = 1$.
Случайная величина $X(w)$ — число единиц в исходе w. Таблица распределения:

X	0	1	$\cdots$	k	$\cdots$	N
P(X)	q^N	Npq^{N-1}	$\cdots$	$\binom{N}{k} p^k q^{N-k}$	$\cdots$	p^N

Геометрическое распределение.

Случайная величина X моделирует событие с двумя исходами, которое повторяется до первого успеха. Таблица распределения:

X	1	2	$\cdots$	k	$\cdots$
P(X)	p	qp	$\cdots$	$q^{k-1}p$	$\cdots$

2.3 Совместное распределение случайных величин.

Пусть X, Y — две случайные величины с множествами значений $M_X = \{x_1, x_2, \ldots\}$ и $M_Y = \{y_1, y_2, \ldots\}$ соответственно.

Определение 9. *Совместным распределением* двух случайных величин назвается вероятностная мера $\mu_{X,Y}(\{x_j, y_k\})$ на вероятностном пространстве $M_X \times M_Y$, для которой

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

Для большего числа случайных величин всё аналогично.

2.4 Независимые случайные величины. Эквивалентное определение независимости случайных величин.

Определение 10. Случайные величины X, Y с множествами значений M_X, M_Y называются *независимыми*, если $\forall\, j, k$ выполнено

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

Для большего числа случайных величин всё аналогично.

Предложение. Пусть X, Y — две случайные величины, и $A, B \subset \mathbb{R}$ — произвольные множества. Тогда

$$X, Y \text{ — независимы} \iff P(\{w\colon X(w) \in A, Y(w) \in B\}) = P(\{w\colon X(w) \in A\}) \cdot P(\{w\colon Y(w) \in B\})$$

Т.е. независимость случайных величин равносильна незвисимости специальных событий.

Доказательство. Очевидно, что достаточно доказать для $A, B \subseteq \operatorname{rng} X, Y$ соответственно. Заметим, что событие $\mathcal{A} = \{w\colon X(w) \in A\}$ можно разбить на объединение следующих событий: $\mathcal{A} = \bigcup\limits_j \{w\colon X(w) = x_j\}$. Причём все эти подсобытия друг с другом не пересекаются, в силу того, что x_j не повторяются. Аналогично можно поступить для события $\mathcal{B} = \{w\colon Y(w) \in B\}$.

Тогда, по свойству счётной аддитивности: $P(\mathcal{A}) = \sum\limits_j \{w\colon X(w) = x_j\}$.

Аналогично $P(\mathcal{B}) = \sum\limits_k \{w\colon Y(w) = y_k\}$.

Аналогично для $P(\mathcal{A} \cap \mathcal{B})$:

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

Теперь имеем:

$$P(\mathcal{A}) \cdot P(\mathcal{B}) = P(\{w\colon X(w) \in A\}) \cdot P(\{w\colon Y(w) \in B\}) =$$

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

Получили независимость событий $\mathcal{A}$ и $\mathcal{B}$. ∎

Замечание. В последнем равенстве во втором равно явно использована абсолютная сходимость рядов

$$\sum_{i\colon x_i \in A} P(\{w\colon X(w) = x_i\}) \text{ и } \sum_{j\colon y_j \in B} P(\{w\colon Y(w) = y_j\})$$

До второго равно написана операция произведения двух рядов, а после второго равно написано что в результате произведения случилось с каждым элементом ряда (по определению произведения абсолютно сходящихся рядов).

2.5 Математическое ожидание случайной величины.

Пусть $\{x_1, \dots, x_n, \dots\}$ — множество всех значений, которые принимает случайная величина X.

Определение 11. *Математическим ожиданием* случайной величины X называется число

$$\mathbb{E}(X) = \sum_j x_j \mu_X(\{x_j\}) = \sum_j x_j P(\{w \colon X(w) = x_j\})$$

В определении мы предполагаем, что ряд сходится абсолютно. Если это не так, то считаем, что случайная величина не имеет конечного математического ожидания.

Обозначение 1. Далее вводим следующее обозначние:

$$P(\{w \colon X(w) = x_j\}) \iff P(X = x_j)$$

Лемма 2.1. Пусть случайная величина X с конечным математическим ожиданием принимает значения y_k на множествах B_k, причём события B_k попарно не пересекаются и в объединении дают всё Ω. Тогда

$$\mathbb{E}(X) = \sum_k y_k P(B_k)$$

Доказательство.

$$\mathbb{E}(X) = \sum_j x_j P(X = x_j) = \sum_j x_j \sum_{k \colon y_k = x_j} P(X = y_k) =$$
$$= \sum_j \sum_{k \colon y_k = x_j} y_k P(X = y_k) = \sum_j \sum_{k \colon y_k = x_j} y_k P(B_k) = \sum_k y_k P(B_k)$$

■

Замечание. Заметим, что по условию у нас величина X имеет конечное математическое ожидание, а значит ряд $\sum\limits_j x_j P(X = x_j)$ сходится абсолютно, а значит мы можем группировать отдельные его слагаемые, так как мы это делали в доказательстве.

2.6 Мат. ожидание функции от случайной величины.

Пусть $\{x_1, \dots, x_n, \dots\}$ — множество всех значений, которые принимает случайная величина X.

Теорема 2.2. *Если $\varphi \colon \mathbb{R} \to \mathbb{R}$ — произвольная функция, то*

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

при условии абсолютной сходимости последнего ряда.

Доказательство. Следует из предыдущей леммы: с одной стороны случайная величина принимает какие-то свои значения ξ_j на множествах $\{w \colon \varphi(X)(w) = \xi_j\}$, а с другой стороны по лемме она принимает значения $y_k = \varphi(x_k)$ на множествах $B_k = \{w \colon X(w) = x_k\}$. ■

2.7 Свойства мат. ожидания.

Определение 12. Если некоторое свойство выполняется с вероятностью единица, то говорят, что оно выполняется *почти наверное.*

Предложение. Мат. ожидание линейно, а именно: пусть X, Y — некоторые случайные величны, $a, b \in \mathbb{R}$. Тогда

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Доказательство. Пусть $\{x_1, x_2, \dots\}, \{y_1, y_2, \dots\}$ — множества значений случайных величин X, Y соответственно. Пусть $B_{k,j} = \{X = x_k, Y = y_j\}$. Заметим, что $B_{k,j}$ попарно не пересекаются, и в объединении дают Ω. Тогда с одной стороны $\mathbb{E}(X) = \sum\limits_j x_j P(X = x_j)$, а с другой стороны $\mathbb{E}(X) = \sum\limits_{k,j} x_j P(B_{k,j})$. Аналогично для $\mathbb{E}(Y)$. Теперь

$$a\mathbb{E}(X) + b\mathbb{E}(Y) = a\sum_j x_j P(X = x_j) + b\sum_k y_k P(Y = y_k) =$$
$$= a\sum_{k,j} x_j P(B_{k,j}) + b\sum_{k,j} y_k P(B_{k,j}) = \sum_{k,j}(ax_j + by_k)P(B_{k,j}) = \mathbb{E}(aX + bY)$$

■

Замечание. Последнее равенство справедливо по лемме: т.к. с одной стороны случайная величина $aX+bY$ принимает какие-то свои значения ξ_i на множествах $\{aX + bY = \xi_i\}$, а сдругой стороны она принимает значения $y_k = ax_j + by_k$ на множествах $B_k = B_{k,j}$.

Предложение. Мат. ожидание монотонно, а именно: если $X \geqslant 0$ почти наверное, то $\mathbb{E}(X) \geqslant 0$.

Доказательство. По определению:

$$\mathbb{E}(X) = \sum_j x_j P(X = x_j)$$

Если $x_j \geqslant 0$, то и вся сумма не меньше нуля. ■

Следствие. Если $X \geqslant Y$ почти наверное, то $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$ почти наверное.

Доказательство. Рассмотрим случайную величину $Z = X - Y$. Видно, что $Z \geqslant 0$ почти наверное. Тогда в силу монотонности

$$0 \leqslant \mathbb{E}(Z) = \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) \iff \mathbb{E}(X) \geqslant \mathbb{E}(Y)$$

■

Предложение. Если $X \geqslant 0$ почти наверное, и $\mathbb{E}(X) = 0$, то $X = 0$ почти наверное.

Доказательство. По определению:

$$\mathbb{E}(X) = \sum_j x_j P(X = x_j) = 0$$

Т.к. $P(X = x_j)$ не может равняться 0 для всех значений X, то $x_j = 0\; \forall\, j$ почти наверное. ■

Предложение. Справедлива оценка

$$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$$

Доказательство. Заметим, что $-|X| \leqslant X \leqslant |X|$. Воспользуемся теперь линейностью и монотонностью мат. ожидания:

$$-|X| \leqslant X \leqslant |X| \iff -\mathbb{E}(|X|) \leqslant \mathbb{E}(X) \leqslant \mathbb{E}(|X|) \iff |\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$$

■

Предложение. Если случайные величины X, Y независимы, и их мат. ожидания определены, то выполнено

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Доказательство.

$$\mathbb{E}(X) \cdot \mathbb{E}(Y) = \left(\sum_i x_i P(X = x_i)\right) \cdot \left(\sum_j y_j P(Y = y_j)\right) =$$
$$= \sum_{i,j} x_i y_j P(X = x_i) P(Y = y_j) = \sum_{i,j} x_i y_j P(\{X = x_1\} \cap \{Y = y_j\}) = \mathbb{E}(XY)$$

■

Замечание. Второй переход справедлив в силу абсолютной сходимости рядов. Четвёртый переход справедлив в силу независимости X, Y.

2.8 Балансировка векторов.

Задача 2.1 (Балансировка векторов.)**.** Пусть $v_1, \ldots, v_n \in \mathbb{R}^n$, и $\forall\ j\ |v_j| = 1$. Всегда ли существует такой набор $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$, что $|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n| \leqslant \sqrt{n}$?

Решение: Если мы будем выбирать ε_i случайным образом и независимо друг от друга, то значение $|\varepsilon_1 v_1 + \ldots + \varepsilon_n v_n|$ будет случайной величиной. Посчитаем мат. ожидание квадрата этой случайной величины:

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

Замечание. По определению длины вектора: $|v| = \sqrt{(v, v)}$, т.е.

$$\left|\sum_j \varepsilon_j v_j\right|^2 = \sqrt{\left(\sum_i \varepsilon_i v_i, \sum_j \varepsilon_j v_j\right)}^2 = \left(\sum_i \varepsilon_i v_i, \sum_j \varepsilon_j v_j\right) = \sum_{i,j}(v_i, v_j)\varepsilon_i\varepsilon_j$$

Заметим так же, что при $i \neq j$ мы имеем $\mathbb{E}(\varepsilon_i\varepsilon_j) = \mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j)$ (в силу независимости выбора ε_i) и $\mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j) = \left(1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2}\right) \cdot 2 = 0$ в силу того, что мы выбираем ε_i равновероятно с вероятностью выбора $\frac{1}{2}$.

С другой стороны, при $i = j$ имеем $\sum_i (v_i, v_i) = |v_i|^2$ и $\mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_i) = \mathbb{E}(\varepsilon_i^2) = 1$.

2.9 Дисперсия, ковариация, коэффициент корреляции.

Определение 13. Пусть X — случайная величина. *Дисперсией* случайной величины X называется число

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

Определение 14. Пусть X, Y — две случайные величины. *Ковариацией* пары случайных величин X, Y называется число

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

Предложение. Ковариация пары случайных величин является билинейной формой.

Доказательство. Приведём определение ковариации к следующему, более удобному виду:

$$\begin{aligned}\operatorname{cov}(X, Y) &= \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \mathbb{E}(XY - \mathbb{E}(Y)X - \mathbb{E}(X)Y - \mathbb{E}(X)\mathbb{E}(Y)) = \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

Проверим линейность по первому аргументу (по второму аналогично):

$$\begin{aligned}\operatorname{cov}(aX_1 + bX_2, Y) &= \mathbb{E}((aX_1 + bX_2)Y) - \mathbb{E}(aX_1 + bX_2)\mathbb{E}(Y) = \\ &= \mathbb{E}(aX_1Y + bX_2Y) - a\mathbb{E}(X_1)\mathbb{E}(Y) - b\mathbb{E}(X_2)\mathbb{E}(Y) = \\ &= a\mathbb{E}(X_1Y) + b\mathbb{E}(X_2Y) - a\mathbb{E}(X_1)\mathbb{E}(Y) - b\mathbb{E}(X_2)\mathbb{E}(Y) = \\ &= [a\mathbb{E}(X_1Y) - a\mathbb{E}(X_1)\mathbb{E}(Y)] + [b\mathbb{E}(X_2Y) - b\mathbb{E}(X_2)\mathbb{E}(Y)] = \\ &= a\operatorname{cov}(X_1, Y) + b\operatorname{cov}(X_2, Y)\end{aligned}$$

■

Следствие. Квадратичная форма $\operatorname{cov}(X, X)$ неотрицательно определённа.

Действительно: $\forall\ X \implies \operatorname{cov}(X, X) = \mathbb{E}(X - \mathbb{E}(X))^2 \geqslant 0$.

Следствие.

$$\mathbb{D}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \operatorname{cov}(X, X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Замечание. Мы доказали, что ковариация является может не самым простым, но всё же линейным объектом, с которыми мы работать умеем. Кроме того мы вывели связь непонятной для нас дисперсии, и довольно понятной ковариации: оказывается, что дисперсия является просто квадратичной формой, ассоциированной с ковариацией. Это позволяет нам считать дисперсию от неочевидных сочетаний случайных величин.

Пример. Пусть например мы всё знаем про величины X, Y по отдельности, и требуется посчитать $\mathbb{D}(X+Y)$. Используем связь дисперсии и ковариации:

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

Получили сумму каких-то выражений, значения которых нам должны быть известны.

Определение 15. Пусть X, Y — случайные величины. *Коэффициентом корреляции* пары случайных величин называется величина

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

2.10 Основные свойства дисперсии и ковариации.

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

Доказательство.

$$\mathbb{D}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = 0 \implies \mathbb{E}(X - \mathbb{E}(X)) = 0 \implies X - \mathbb{E}(X) = 0 \implies X = \mathbb{E}(X)$$

Второй переход верен в силу свойства мат. ожидания: если $X \geqslant 0$ и $\mathbb{E}(X) = 0$, то $X = 0$ почти наверное. ■

Предложение. Для произвольных $a, b \in \mathbb{R}$ верно: $\mathbb{D}(aX+b) = a^2\mathbb{D}(X)$.

Доказательство.

$$\mathbb{D}(aX+b) = \mathbb{E}(aX + b - \mathbb{E}(aX+b))^2 = \mathbb{E}(aX + b - \mathbb{E}(aX) - b)^2 = \mathbb{E}(aX - \mathbb{E}(aX))^2 =$$

$$= \mathbb{D}(aX) = \operatorname{cov}(aX, aX) = a^2\operatorname{cov}(X,X) = a^2\mathbb{D}(X)$$

■

Предложение. Пусть X, Y — независимые случайные величины. Тогда $\operatorname{cov}(X,Y) = 0$

Доказательство.

$$\operatorname{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0$$

■

Следствие. Если X, Y — независимы, то $\mathbb{D}(X+Y) = \mathbb{D}(X) + \mathbb{D}(Y)$.

Действительно, $\mathbb{D}(X+Y) = \mathbb{D}(X) + \underbrace{2\operatorname{cov}(X,Y)}_{0} + \mathbb{D}(Y) = \mathbb{D}(X) + \mathbb{D}(Y)$.

2.11 Неравенство Коши-Буняковского и линал нам в анал.

Предложение. Если определены $\mathbb{E}(X^2)$ и $\mathbb{E}(Y^2)$, и $\exists\, a, b \in \mathbb{R} \colon aX + bY = 0$ почти наверное, то справедливо неравенство Коши-Буняковского

$$|\mathbb{E}(XY)| \leqslant \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$$

Доказательство. Неравенство очевидно верно при $\mathbb{E}(X^2) = 0$ или $\mathbb{E}(Y^2) = 0$. Рассмотрим для случая $\mathbb{E}(X^2) > 0$ и $\mathbb{E}(Y^2) > 0$:

Рассмотрим квадратичную функцию $f(t) = \mathbb{E}(Y - tX)^2 = \mathbb{E}(Y^2) - 2t\mathbb{E}(XY) + t^2\mathbb{E}(X^2)$. Заметим, что $\forall\, t \implies f(t) \geqslant 0$. Отсюда следует неположительность дискриминанта (если бы дискриминант был положителен, то какая-то часть параболы была под осью OX, а значит функция бы не была неотрицательна при любых t):

$$D = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leqslant 0 \iff \mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2) \iff |\mathbb{E}(XY)| \leqslant \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$$

■

Рассмотрим V — векторное пространство случайных величин (тут даже я охуел). Вспомним, что для случайных величин действительно определено умножение на число и сложение, поэтому множество случайных величины с этими двумя операциями действительно будет векторным пространством. Введём скалярное произведение на V по следующему правилу:

$$(X, Y) = \operatorname{cov}(X, Y)$$

Тогда длину векторов (случайных величин) будем измерять по определению:

$$\forall\, X \in V \implies |X| = \sqrt{(X, X)} = \sqrt{\operatorname{cov}(X, X)} = \sqrt{\mathbb{D}(X)}$$

Тогда коэффициент корреляции на случайных величинах из V принимает следующий вид:

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

2.12 Мат. ожидание и дисперсия биномиального распределения.

Пусть случайная величина S_n имеет биномиальное распределение, т.е. $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Найдём мат. ожидание и дисперсию:

Введём $X_1, \ldots, X_n$ — независимые бернулиевские случайные величины, т.е.

$$X_k = \begin{cases} 1 & P(X_k = 1) = p \\ 0 & P(X_k = 0) = 1 - p \end{cases}$$

Очевидно, что $\mathbb{E}(X_k) = p$. Тогда $S_n = \sum\limits_{k=1}^{n} X_k$, и соответственно $\mathbb{E}(S_n) = \sum\limits_{k=1}^{n} \mathbb{E}(X_k) = \sum\limits_{k=1}^{n} p = np$.

Т.к. X_k — независимы, то $\mathbb{D}(S_n) = \sum\limits_{k=1}^{n} \mathbb{D}(X_k) = n\mathbb{D}(X_k) = n\mathbb{E}(X_k - \mathbb{E}(X_k))^2 = n\mathbb{E}(X_k - p)^2 = n\left((1-p)^2 \cdot p + p^2 \cdot (1-p)\right) = np(1-p)$.

2.13 Неравенство Чебышева (Маркова).

Теорема 2.3 (Неравенство Чебышева.). *Если $X \geqslant 0$ почти наверное, то $\forall\, t > 0 \in \mathbb{R}$ выполнено*

$$P(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$$

Понимать это неравенство можно так: при $t > \mathbb{E}(X)$ неравенство даёт некую оценку на вероятность того, что значения случайной величины будут больше её мат. ожидания.

Доказательство. Покажем, что $\mathbb{E}(X) \geqslant tP(X \geqslant t)$: по определению:

$$\mathbb{E}(X) = \sum_j x_j P(X = x_j)$$

Оценим ряд снизу: если $x_j < t$, то заменим x_j на ноль. Иначе если $x_j \geqslant t$, то заменим x_j на t. Имеем:

$$\mathbb{E}(X) = \sum_j x_j P(X = x_j) \geqslant t \sum_{i=1}^{k} p_i, \quad k \in [0, +\infty]$$

Видно, что стоящая справа сумма (или ряд) равносильна $t \cdot P(X \geqslant t)$. ■

Следствие. Если $\mathbb{E}(X^2) < \infty$, то $\forall\, \varepsilon > 0$ верно

$$P(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \frac{\mathbb{D}(X)}{\varepsilon^2}$$

Доказательство. По неравенству Чебышева:

$$P(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(|X - \mathbb{E}(X)|)}{\varepsilon} \leqslant \frac{\mathbb{E}(X - \mathbb{E}(X))^2}{\varepsilon^2} = \frac{\mathbb{D}(X)}{\varepsilon^2}$$

■

3 Закон больших чисел.

3.1 Закон больших чисел в слабой форме.

Лемма 3.1 (Закон больших чисел в слабой форме)**.** Пусть $\{X_n\}_n$ — последовательность независимых одинаково распределённых случайных величин, и $\mathbb{E}(X_n^2) < \infty$. Пусть $\mathbb{E}(X_1) = m$, тогда $\forall\, \varepsilon > 0$ верно

$$\lim_{n\to\infty} P\left(\left|\frac{\sum\limits_{k=1}^{n} X_k}{n} - m\right| \geqslant \varepsilon\right) = 0$$

Доказательство. Т.к. все величины одинаково распределены и $\mathbb{E}(X_1) = m$, то $\forall\, i$ $\mathbb{E}(X_i) = m$. Тогда

$$\mathbb{E}\left(\frac{\sum\limits_{k=1}^{n} X_k}{n}\right) = \sum_{k=1}^{n} \mathbb{E}\left(\frac{X_k}{n}\right) = n \cdot \frac{\mathbb{E}(X_1)}{n} = m$$

Тогда по следствию из неравенства Чебышева:

$$P\left(\left|\frac{\sum\limits_{k=1}^{n} X_k}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{D}\left(\frac{1}{n}\sum\limits_{k=1}^{n} X_k\right)}{\varepsilon^2} = \frac{\mathbb{D}\left(\sum\limits_{k=1}^{n} X_k\right)}{n^2\varepsilon^2} = \frac{n\mathbb{D}(X_1)}{n^2\varepsilon^2} = \frac{\mathbb{D}(X_1)}{n\varepsilon^2}$$

Устремляя n к бесконечности получаем доказательство утверждения. ∎

Замечание. Пусть X_k — независимые бернулиевские случайные величины с вероятностью успеха p. Тогда величина $\frac{\sum\limits_{k=1}^{n} X_k}{n}$ — частота успешного исхода эксперимента при проведении n независимых испытаний. Известно, что $\mathbb{E}(X_k) = p$, тогда $\mathbb{D}(X_k) = \mathbb{E}(X_k^2) - \mathbb{E}(X_k)^2 = p^2 - p = pq$, где $q = (1-p)$. Тогда

$$P\left(\left|\frac{\sum\limits_{k=1}^{n} X_k}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{pq}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

Т.е. при проведении большого числа испытаний, частота успешного результата эсперимента стремится к p.

Можно разобрать смысл этого утверждения на примере: пусть эксперимент у нас состоит в подкидывании монеты, успехом мы считаем выпадение орла. Очевидно, что в одном независимом испытании орёл выпадает с вероятностью $\frac{1}{2}$. При этом в жизни подбрасывая монету $5, 10$ и даже 100 раз мы можем ни разу не получить орла. Но данное утверждение говорит нам о том, что при проведении огромного числа испытаний, вероятность выпадения орла в среднем во всех испытаниях будет примерно $\frac{1}{2}$.

3.2 Теорема Муавра-Лапласа.

Пусть X_k — независимые бернулиевские случайные величины с вероятностью успеха p. Пусть S_n — количество успешных испытаний $\left(S_n = \sum\limits_{k=1}^{n} X_k\right)$. S_n имеет биномиальное распределение:

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}$$

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

Теорема 3.2 (Муавра-Лапласа)**.** *Если $n \to \infty$, вероятность исхода $p \in (0,1)$ фиксирована, величина $x_m = \dfrac{m - np}{\sqrt{npq}}$ ограничена равномерно по m, n ($\exists\, a, b \in \mathbb{R}\colon a \leqslant x_m \leqslant b$), то*

$$P(S_n = m) \sim \frac{1}{\sqrt{npq}}\varphi(x_m), \quad \varphi(x_m) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

где $\varphi(x_m)$ — плотность стандарного нормального распределения (что бы это ни значило).

3.3 Задача о булочках с изюмом.

Задача 3.1 (О булочках с изюмом)**.** Будем заниматься изготовлением булочек с изюмом в промышленных масштабах. Чтобы получить сколько-то готовых булочек, нужно закинуть все ингридиенты (тесто и изюм) в специальную машину, которая сама всё сделает и на выходе у нас будет готовый продукт. Под "готовым продуктом" понимается **готовый продукт**: после того как машина его выдаст, в нём ничего изменить не получится. Мы хотим выяснить, какое наименьшее число изюминок нужно добавить, чтобы с высокой вероятностью (> 0.99) в каждой булочке была бы хотя бы одна изюминка?

Решение: Пусть мы собираемся изготовить N булочек, и рассчитываем тратить не более λ изюма на булочку. Тогда всего нам потребуется засыпать $N\lambda$ изюма. Каждая изюминка попадает в конкретную булочку с вероятностью $\dfrac{1}{N}$ (т.к. каждая добавленная изюминка должна попасть хоть в какую-нибудь булочку). Соответственно, каждая изюминка не попадает в конкретную булочку с вероятностью $1 - \dfrac{1}{N}$. Тогда случайная величина S — количество изюминок, попавших в конкретную булочку, имеет биномиальное распределение, и событие "в конкретную булочку не попала ни одна изюминка" записывается как

$$P(S = 0) = \left(1 - \frac{1}{N}\right)^{N\lambda} = e^{-\lambda}$$

Теперь остаётся только выбрать такое λ, чтобы вероятность такого непрятного события была менее 0.01. Подойдёт $\lambda = 5$: $e^{-5} \approx 0{,}0067 < 0.01$.

3.4 Теорема Пуассона.

Теорема 3.3. *Пусть проводится N серии, в кажой серии ровно N испытаний по схеме Бернулли. Пусть вероятность успеха в N-ой серии равна p_N, и известно, что $p_N \cdot N = \lambda$ (некий показатель среднего результата в каждой серии). Тогда вероятность наступления ровно k успехов в N-ой серии равна*

$$P(S_N = k) \xrightarrow{N\to\infty} \frac{\lambda^k}{k!}e^{-\lambda}$$

Доказательство. Если $p_N \cdot N = \lambda$, то $p_N = \dfrac{\lambda}{N}$. Тогда

$$P(S_N = k) = \binom{N}{k} p_N^k (1 - p_N)^{N-k} = \frac{N!}{k!(N-k)!} \cdot \left(\frac{\lambda}{N}\right)^k \cdot \left(1 - \frac{\lambda}{N}\right)^{N-k} =$$

$$= \frac{\lambda^k}{k!} \cdot \frac{N(N-1)\cdot \ldots \cdot (N-k+1)}{N^k} \cdot \underbrace{\left(1 - \frac{\lambda}{N}\right)^{-k}}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{N}\right)^{N}}_{\to e^{-\lambda}} =$$

$$= \frac{\lambda^k}{k!} \cdot \underbrace{\left(1 - \frac{1}{N}\right)}_{\to 1} \cdot \underbrace{\left(1 - \frac{2}{N}\right)}_{\to 1} \cdot \ldots \cdot \underbrace{\left(1 - \frac{k-1}{N}\right)}_{\to 1} \cdot e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda}$$

■

3.5 Модель Эрдеша-Реньи случайного графа и надёжность сети.

Пусть есть n городов, причём каждый городо соединён дорогой с каждым другим. Известно, что за год конкретная дорога приходит в негодность с вероятностью $1-p$. Соответственно, конкретная дорога за год не изнашивается с вероятностью p. Хотим узнать, с какой вероятностью через год из каждого города можно будет добраться до всех остальных?

Предложение. Пусть V — множество из n вершин графа. Если $p = \dfrac{c\ln(n)}{n}$, и $c > 2$, то граф связен с вероятностью, стремящейся к единице.

Доказательство. Пусть X_n — случайная величина, такая что

$$X_n = \begin{cases} \text{число компонент связности} & \text{граф несвязен} \\ 0 & \text{граф связен} \end{cases}$$

Покажем, что $P(X > 1) \to 0$: рассмотри все k-элементные подмножества множества наших вершин: $K_1^k, K_2^k, \ldots, K_{\binom{n}{k}}^k$, и рассмотрим события следующего вида: $A_{n,k,j}$ — множество K_j^k является компонентой связности. Тогда распишем случайную величину X_n:

$$X_n = \sum_{k=1}^{n-1}\sum_{j=1}^{\binom{n}{k}} I_{A_{n,k,j}}$$

где $I_{A_{n,k,j}} = 1$, если K_j^k — компонента связности, и 0 иначе. Оценим матожидание:

$$\mathbb{E}(X_n) = \sum_{k=1}^{n-1}\sum_{j=1}^{\binom{n}{k}} \underbrace{\mathbb{E}(I_{A_{n,k,j}})}_{P(A_{n,k,j})} \overset{(1)}{\leqslant} \sum_{k=1}^{n-1}\sum_{j=1}^{\binom{n}{k}} P\left(K_j^k \text{ не соединён с } V\setminus K_j^k\right) \overset{(2)}{=} \sum_{k=1}^{n-1}\binom{n}{k}q^{k(n-k)} =$$

$$= \sum_{k=1}^{\frac{n}{2}}\binom{n}{k}q^{k(n-k)} + \sum_{k=\frac{n}{2}+1}^{n-1}\binom{n}{k}q^{k(n-k)} \overset{(3)}{=} 2\sum_{k=1}^{\frac{n}{2}}\binom{n}{k}q^{k(n-k)} \overset{(4)}{\leqslant} 2\sum_{k=1}^{\frac{n}{2}}\binom{n}{k}\left(q^{\frac{n}{2}}\right)^k \leqslant$$

$$\leqslant 2\sum_{k=1}^{n-1}\binom{n}{k}\left(q^{\frac{n}{2}}\right)^k = \leqslant 2\sum_{k=1}^{n-1}\binom{n}{k}(1^{n-k})\left(q^{\frac{n}{2}}\right)^k = 2(1+q^{\frac{n}{2}})^n - 2 - 2q^{\frac{n^2}{2}}$$

Объяснение перехода (1): для события $A_{n,k,j}$ мы требуем, чтобы все вершины в множестве K_j^k были связаны между собой, и ни одна из них не была связана с $V\setminus K_j^k$. Переходя к событиям $\{K_j^k$ не соединён с $V\setminus K_j^k\}$ мы отказываемся от первого требования, и получается, что события $A_{n,k,j}$ являются подмножествами в новых событиях, а значит, их вероятности оцениваются сверху.
Объяснение перехода (2): Вероятность новых событий мы умеем считать: это просто значит, что между каждой из k вершин нашего множества, и каждой из оставшихся $n-k$ вершин его дополнения отсутствует ребро (отсутствует дорога между городами). Для каждого ребра вероятность такого события равна q, и события независимы.
Объяснение перехода (3):

$$\sum_{k=\frac{n}{2}+1}^{n-1}\binom{n}{k}q^{k(n-k)} = \sum_{k=1}^{\frac{n}{2}}\binom{n}{n-k}q^{(n-k)(n-(n-k))}$$

Свойство числа сочетаний: $\dbinom{n}{k} = \dbinom{n}{n-k}$.

Объяснение перехода (4): $k \leqslant \dfrac{n}{2} \implies n-k \geqslant \dfrac{n}{2} \implies q^{n-k} \leqslant q^{\frac{n}{2}}$. Последний переход верен т.к. $|q| \leqslant 1$.

По условию $q = (1-p) = \left(1 - \dfrac{c\ln(n)}{n}\right) \implies q^{\frac{n}{2}} = \left(1-\dfrac{c\ln(n)}{n}\right)^{\frac{n}{2}} = e^{\frac{n}{2}\ln\left(1-\frac{c\ln(n)}{n}\right)} = e^{\frac{n}{2}\left(-\frac{c\ln(n)}{n}+O\left(\frac{\ln^2(n)}{n^2}\right)\right)} =$
$= e^{\frac{c}{2}\ln(n)+o(1)} = e^{o(1)}\cdot n^{-\frac{c}{2}}$.

Тогда теперь

$$(1+q^{\frac{n}{2}})^n = (1+e^{o(1)}\cdot n^{-\frac{c}{2}})^n = e^{n\ln\left(1+e^{o(1)}n^{-\frac{c}{2}}\right)} = e^{n\cdot n^{-\frac{c}{2}}e^{o(1)}+o(1)}$$

Понятно, что $\dfrac{c}{2} > 1$, поэтому $n^{-\frac{c}{2}} \to 0$, и поэтому всё выражение сходится к 1. Тогда $2(1+q^{\frac{n}{2}})^n - 2 - 2q^{\frac{n^2}{2}} \to 0 \iff \mathbb{E}(X_n) \to 0$, и по неравенству Чебышева: $P(X_n > 1) \leqslant \mathbb{E}(X_n) \to 0$. ■

4 Общее определение вероятностного пространства.

4.1 Теория меры в слабой форме.

Attention!!! Вплодь до определения случайной величины в общем случае конспект будет содержать исключительно сухие формулировки, факты и утверждения без доказательств, и мои комментариев. Это связано с тем, что практически всё что есть в этой части курса мной подробно разобрано и расписано в моём конспекте по теории меры Лебега. Там представлено действительно много примеров на эти темы, в частности там можно поппытаться дейтсвительно вникнуть в то, что такое Борелевская алгебра. Конспект пока не закончен и на половину, но сейчас там уже есть всё, что нам преподносилось в этой части курса. Пусть множество X не пусто.

Определение 16. *Алгеброй подмножеств множества* X (алгеброй) называется набор $\mathcal{A} \subset 2^X$, если выполнены следующие свойства:

1. $X, \varnothing \in \mathcal{A}$.
2. $A, B \in \mathcal{A} \implies A \setminus B, A \cap B, A \cup B \in \mathcal{A}$.

Определение 17. *σ-алгеброй* называется набор $\mathcal{A} \subset 2^X$, если выполнены следующие свойства:

1. $\mathcal{A}$ является алгеброй.
2. Если множества $A_1, A_2, \ldots \in \mathcal{A}$, то их счётные объединение и пересечение также там лежат: $\bigcap\limits_{j=0}^{\infty} A_j \in \mathcal{A}, \bigcup\limits_{j=0}^{\infty} A_j \in \mathcal{A}$

Пусть S — какой-то набор подмножеств множества X.

Определение 18. $\sigma(S)$ называется σ-алгеброй, *порождённой* S. $\sigma(S)$ есть пересечение всех σ-алгебр, содержащих S.

Пусть $\mathcal{A}$ — алгебра множеств.

Определение 19. Функция $P\colon \mathcal{A} \to [0,1]$ называется *аддитивной*, если для произвольных $A, B \in \mathcal{A}, A \cap B = \varnothing$ выполнено

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

Функция $P\colon \mathcal{A} \to [0,1]$ называется *счётно-аддитивной*, если для произвольного, не более чем счётного набора $A_n \in \mathcal{A}$, такого что $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$ выполнено

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

Пусть теперь $\mathcal{A}$ — — — σ-алгебра.

Определение 20. Счётно-аддитивная функция $P\colon \mathcal{A} \to [0,1]$ называется *вероятностной мерой* на $\mathcal{A}$, если $P(\Omega) = 1$.

Определение 21. Тройка $(\Omega, \mathcal{A}, P)$ называется *вероятностным пространством.*

Предложение. Пусть $P\colon \mathcal{A} \to [0,1]$ — аддитивная функция на алгебре $\mathcal{A}$. Функция P счётно-аддитивна на $\mathcal{A}$ тогда и только тогда, когда для всякого набора $A_n \in \mathcal{A}$, такого, что каждое следующее множество вложено в предыдущее $(A_{n+1} \subset A_n)$, и $\bigcup\limits_{n=1}^{\infty} A_n = \varnothing$ выполнено $\lim P(A_n) = 0$.

Теорема 4.1. *Пусть $\mathcal{A}$ — некоторая алгебра подмножеств пространства Ω, и пусть $P_0\colon \mathcal{A} \to [0,1]$ — счётно-аддитивная функция множества на алгебре $\mathcal{A}$. Тогда существует единственная вероятностная мера $P\colon \sigma(\mathcal{A}) \to [0,1]$, такая что $P\Big|_{\mathcal{A}} = P_0$.*

4.2 Случайная величина в общем случае.

Пусть у нас есть вероятностное пространство: $(\Omega, \mathcal{A}, P)$. Заметим такую вещь, как только множество Ω стало произвольным, мы начали испытывать некие проблемы с определением события. Если в конечном и счётном случаях мы называли событиями вообще всякое подмножество в Ω, то теперь так не получится, в частности потому что если A, B — события, то нам бы хотелось, чтобы событиями являлись так же $A \cap B, A \cup B, A \setminus B$, а это не всегда возможно, когда мы говорим о континуальных множествах. Чтобы такие ограничения обойти мы придумали какую-то науку и создали *алгебры*. С алгебрами жить становится веселее, т.к. для множеств из алгебры все наши желания исполняются. *σ-алгебры* вообще являются пределом мечтаний, и элементы σ-алгебр мы уже можем воспринимать как полноценные события. Действительно, что бы мы с ними не делали, мы снова получим элемент из σ-алгебры, т.е. снова получим событие.

Теперь нам нужно ввести понятние случайной величины. В конечном и счётном случае случайные величины у нас отображали всевозможные события в числа. Сейчас идея будет примерно та же самая, с поправкой на новое определение события.

Определение 22. Функция $X \colon \Omega \to \mathbb{R}$ называется *случайной величиной*, тогда и только тогда, когда для всякого $t \in \mathbb{R}$ выполняется

$$\{w \in \Omega \colon X(w) \leqslant t\} \in \mathcal{A}$$

Т.е. X — случайная величина, тогда и только тогда, когда множество $\{w \colon X(w) \leqslant t\}$ является событием $\forall\, t$.

Замечание. Заметим следующее:

$$\{w \in \Omega \colon X(w) \leqslant t\} = \{w \in \Omega \colon X(w) \in (-\infty; t]\}$$

Т.е. то множество, которое мы хотим считать событием при любом t состоит из исходов, которые под действием случайной величины X попадают в множество $(-\infty; t]$. Т.е. это множество по определению является полным прообразом множества $\{X(w) \colon X(w) \in (-\infty; t]\} = X((-\infty; t])$, и мы соответственно его будет обозначать как $X^{-1}((-\infty; t])$.

Теперь постараемся расширить наше определение. Перепишем его в следующем виде: X — случайная величина $\iff$ $\{w \colon X(w) \in B\}$. По определению мы умеем найти вероятность такого события когда $B = (-\infty, t]\ \forall\, t$. Но мы хотим как-то расширить класс B чтобы считать вероятности более сложных событий.

Пусть X — случайная величина.

Предложение. Тогда для всякого $B \in \mathcal{B}(\mathbb{R})$ множество $\{w \colon X(w) \in B\} = X^{-1}(B) \in \mathcal{A}$.

Доказательство. Свойства полного прообраза:

- $X^{-1}\left(\bigcap\limits_{k=1}^{\infty} A_n\right) = \bigcap\limits_{k=1}^{\infty} X^{-1}(A_n)$.
- $X^{-1}\left(\bigcup\limits_{k=1}^{\infty} A_n\right) = \bigcup\limits_{k=1}^{\infty} X^{-1}(A_n)$.
- $X^{-1}(\Omega \setminus B) = \mathbb{R} \setminus X^{-1}(B)$.

Рассмотрим систему множеств $C = \{B \subset \mathbb{R} \colon X^{-1}(B) \in \mathcal{A}\}$. $\mathcal{C}$ состоит из таких множеств, прообразы которых лежат в σ-алгебре A. Проверим, что $\mathcal{C}$ является σ-алгеброй, и содержит все лучи вида $(-\infty, t]$. После того как мы это проверим, то т.к. $\mathcal{B}(\mathbb{R}) = \sigma((-\infty, t])$ (доказывалось в пункте 2.3 конспекта по мере Лебега), мы получим включение $\mathcal{B}(\mathbb{R}) \subset \mathcal{C}$, т.е. мы будем знать, что прообраз любого Борелевского множества (по определению $\mathcal{C}$) лежит в $\mathcal{A}$, а это именно то, что мы хотим доказать.

Прежде всего покажем, что $\mathcal{C}$ — σ-алгебра:

$$X^{-1}(\varnothing) = \varnothing \in \mathcal{A} \implies \varnothing \in \mathcal{C}, \quad X^{-1}(\mathbb{R}) = \Omega \in \mathcal{A} \implies \Omega \in \mathcal{C}$$

$$B \in \mathcal{C} \iff X^{-1}(B) \in \mathcal{A} \implies \Omega \setminus X^{-1}(B) \in \mathcal{A} \iff X^{-1}(\mathbb{R} \setminus B) \in \mathcal{A} \iff \mathbb{R} \setminus B \in \mathcal{C}$$

$$B_n \in \mathcal{C} \iff X^{-1}(B_n) \in \mathcal{A} \implies \bigcap_{k=1}^{\infty} X^{-1}(B_n), \bigcup_{k=1}^{\infty} X^{-1}(B_n) \in \mathcal{A} \iff X^{-1}\left(\bigcap_{k=1}^{\infty} B_n\right), X^{-1}\left(\bigcup_{k=1}^{\infty} B_n\right) \in \mathcal{A} \iff$$

$$\iff \bigcap_{k=1}^{\infty} B_n, \bigcup_{k=1}^{\infty} B_n \in \mathcal{C}$$

Таким образом $\mathcal{C}$ — σ-алгебра. Кроме того из условия следует, что $\mathcal{C}$ содержит прообразы лучей, т.к. X — случайная величина по условию. Таким образом имеем $\mathcal{B}(\mathbb{R}) \subset \mathcal{C}$. ∎

Обозначение 2. Введём обозначение, аналогичное обозначению, которое мы вводили для дискретного случая. Если X — случайная величина, то
$$\{X \leqslant t\} \iff \{w \in \Omega \colon X(w) \leqslant t\} = X^{-1}((-\infty, t])$$

Пусть теперь X, Y — случайные величины, $a, b \in \mathbb{R}$.

Предложение. Выражения $aX + bY, X \cdot Y$ являются случайными величинами.

Доказательство. Если $X = const$, то X, очевидно, является случайной величиной. Рассмотрим выражение aX:

$$\{aX \leqslant t\} = \begin{cases} \left\{X \leqslant \dfrac{t}{a}\right\} \in \mathcal{A} & a > 0 \\ \left\{X \geqslant -\dfrac{t}{a}\right\} \in \mathcal{A} & a < 0 \end{cases}$$

Таким образом выражение aX является случайной величиной. Рассмотрим теперь выражение $X + Y$. Покажем, что $X + Y > t$ является событием:

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

Т.к. $\{X > r_n\}$ — дополнение к $\{X \leqslant r_n\} \in \mathcal{A}$, то $\{X > r_n\} \in \mathcal{A}$.
Т.к. $\{r_n > t - Y\}$ — дополнение к $\{-Y < r_n - t\} \in \mathcal{A}$, то $\{r_n > t - Y\} \in \mathcal{A}$.
Из этих двух условий следует, что $\{X + Y > t\} \in \mathcal{A} \implies \{X + Y \leqslant t\} \in \mathcal{A}$ (как дополнение) $\implies X + Y$ — случайная величина.
Кроме того понятно, что X^2 — также является случайной величиной, т.к. $\{X^2 \leqslant t\} = X^{-1}([-\sqrt{t}, \sqrt{t}])$, где $[-\sqrt{t}, \sqrt{t}]$ — Борелевское множество.

Из предыдущих двух утверждений следует, что $X \cdot Y = \dfrac{(X+Y)^2 - X^2 - Y^2}{2}$ является случайной величиной. ■

Предложение. Пусть X_n — случайные величины, и $\forall\ w \implies \lim X_n(w) = X(w)$. Тогда X — случайная величина. Т.е. поточечный предел случайных величин является случайной величиной.

Доказательство. Необходимо показать, что $\forall\ t \implies \lim X_n(w) \leqslant t$. Покажем, что $\lim X_n(w) \leqslant t \iff \forall\ k \in \mathbb{N}\ \exists\ N \in \mathbb{N} \colon \forall\ n > N \implies X_n(w) \leqslant t + \dfrac{1}{k}$.

$\Rightarrow$: Практически очевидно. Для каждой $\dfrac{1}{k}$ начиная с некоторого момента X_n попадают в окрестность предела. Значит, они отличаются от предела менее чем на $\dfrac{1}{k} \implies \lim X_n(w) \leqslant t + \dfrac{1}{k}$.

$\Leftarrow$: Совсем очевидно. Если с некоторого номера $X_n(w) \leqslant t + \dfrac{1}{k}$, то переходим к пределу: $\lim X_n(w) \leqslant t + \dfrac{1}{k}$ — верно для любого $k \implies \lim X_n(w) \leqslant t$.
Перепишем теперь на языке теоретико-множественных операций:

$$\begin{tikzcd}
\lim X_n(w) \leqslant t \arrow[r, Leftrightarrow] \arrow[d, leftrightarrow] & \forall\ k\ \exists\ N \colon \forall\ n > N \implies X_n(w) \leqslant t + \frac{1}{k} \arrow[d, leftrightarrow] \\
\{w \colon X(w) \leqslant t\} \arrow[r, equal] & \bigcup\limits_{k=1}^{\infty} \bigcap\limits_{N=1}^{\infty} \bigcup\limits_{n=N}^{\infty} \{w \colon X_n(w) \leqslant t + \frac{1}{k}\}
\end{tikzcd}$$

Остаётся заметить, что правые множества лежат в $\mathcal{A}$. ■

4.3 Распределение случайных величин.

Пусть X — случайная величина на вероятностном пространстве $(\Omega, \mathcal{A}, P)$.

Определение 23. *Распределением* случайной величины X называется вероятностная мера μ_X на $\mathcal{B}(\mathbb{R})$, определяемая следующим равенством:

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

По предложению мы знаем, что $X^{-1}(B)$ (где B — Борелевское множество) лежит в A, а значит мы можем брать у него вероятность. Счётная аддитивность μ_X следует из счётной аддитивности P и свойств полного прообраза.

Определение 24. *Функцией распределения* случайной величины X называется функция

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

Обозначение 3. Скобками $\langle a, b\rangle$ будем обозначать любой промежуток от a до b, подчёркивая, что нам не важно, является ли он отрезком, полуинтервалом или интервалом.

Предложение. Функция распределения F_X удовлетворяет следующим свойствам:

1. F_X не убывает;
2. F_X непрерывна справа;
3. $\lim\limits_{t\to-\infty} F_X(t) = 0$ и $\lim\limits_{t\to\infty} F_X(t) = 1$;

Доказательство. 1) Пусть $t < s$. Тогда

$$F_X(t) = P((-\infty, t]),\ F_X(s) = P((-\infty, s]) = \underbrace{P((-\infty, t])}_{\geqslant 0} + \underbrace{P((t, s])}_{\geqslant 0} \implies F_X(t) \leqslant F_X(s)$$

2) Пусть $t_n \geqslant t$ и $t_n \to t$. Заметим что

$$\bigcap_{k=1}^{\infty}\{X \leqslant t + \frac{1}{k}\} = \{X \leqslant t\}$$

По непрерывности вероятностной меры: $P\left(X \leqslant t + \dfrac{1}{k}\right) \to P(X \leqslant t)$. В частности

$$\forall\, \varepsilon > 0 \exists\, k \colon P\left(X \leqslant t + \frac{1}{k}\right) \leqslant P(X \leqslant t) + \varepsilon$$

И с другой стороны $P\left(X \leqslant t + \dfrac{1}{k}\right) \geqslant P(X \leqslant t)$. Т.к. $t_n \to t$ и $t_n \geqslant t$, то $\exists\, n_0 \colon \forall\, n > n_0 \implies t \leqslant t_n \leqslant t + \dfrac{1}{k}$. В силу монотонности функции распределения: $F_X(t) \leqslant F_X(t_n) \leqslant F_X\left(t + \dfrac{1}{k}\right) \leqslant F_X(t) + \varepsilon$. Отсюда по определению предела имеем $F_X(t_n) \to F_X(t)$.

3) Доказательство этого пункта аналогично доказательству предыдущего путём устремления t к $-\infty$ или $+\infty$. ∎

Теорема 4.2. *Пусть дана функция $F \colon \mathbb{R} \to [0, 1]$ со свойствами 1)-3). Тогда существует вероятностное пространство $(\Omega, \mathcal{A}, P)$ и случайная величина X на нём, такая что F является функцией распределения X. Кроме того распределение случайной величины μ_X однозначно задаётся функцией распределения F_X.*

Идея доказательства: Пусть $\mathcal{A}$ — алгебра всевозможных попарно непересекающихся промежутков на $\mathbb{R}$. Нетрудно видеть, что $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$. Пусть X, Y — две случайные величины, такие что $F_X = F_Y$. Тогда $\mu_X\mathcal{A} = \mu_Y\mathcal{A}$. И по теореме Лебега $\mu_X = \mu_Y$ на $\sigma(\mathbb{R})$.

Доказательство первой части чуть менее тривиально: рассмотрим ту же алгебру, и положим $P(A) = \sum\limits_{j=1}^{N} F(b_j) - F(a_j)$, $\forall\, A \in \mathcal{A}$. Заметим, что каждое A выглядит как $\bigcup\limits_{j=1}^{N}\langle a_j, b_j\rangle$. Видно, что P — аддитивна. Если бы мы проверили, что P так же счётно-аддитивна (что верно), то по теореме Лебега: $\exists!\ \widetilde{P}$ — счётно-аддитивная мера на $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, такая что $\widetilde{P}\mathcal{A} = P$. И функцией распределения тождественной случайной величина на $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \widetilde{P})$ является F.

4.4 Абсолютно непрерывные и дискретные случайные величины.

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

Определение 25. Случайная величина X имеет *абсолютно непрерывное распределение*, если существует такая неотрицательная и интегрируемая функция ρ_X, что

$$F_X(t) = \int\limits_{-\infty}^{t} \rho_X(x)dx$$

При это функция ρ_X называется *плотностью* случайной величины X.

Понятно, что

$$P(a < X \leqslant b) = \mu_X((a,b]) = F_X(b) - F_X(a) = \int\limits_{a}^{b} \rho_X(x)dx$$

Кроме того

$$P(X = a) = \lim_{n\to\infty} \left(F_X(a) - F\left(a + \frac{1}{n}\right)\right) = 0$$

Свойства плотности:

- ρ_X неотрицательна;
- $\int\limits_{-\infty}^{\infty} \rho_X(t)dt = 1$;
- $F'_X(t) = \rho_X(t)$;

Определение 26. Случайная величина X называется *дискретной*, если множество её значений конечно или счётно. Если $x_1, x_2, \ldots$ — все различные её значения, то множества $A_i = \{w \in \Omega\colon X(w) = x_i\}$ попарно не пересекаются. Пусть $p_i = P(A_i)$, тогда распределение μ_X полностью определено значениями x_i и p_i, и имеет вид

$$F_X = \mu_X = \sum_{x_i\colon x_i \in B} p_i \quad \forall\ B \in \mathcal{B}(\mathbb{R})$$

4.5 Примеры абсолютно непрерывных случайных величин.

Пример. **Равномерное распределение.**

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

$$\rho_X(t) = \begin{cases} \dfrac{1}{b-a} & t \in [a,b] \\ 0 & t \notin [a,b] \end{cases}$$

Пример. **Нормальное распределение.**

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

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

Пример. **Экспоненциальное распределение.**

Случайная величина X имеет *экспоненциальное распределение* с параметром $\lambda > 0$, если её распределение задано плотностью

$$\rho_X(t) = \begin{cases} \lambda e^{-\lambda t} & t \geqslant 0 \\ 0 & t < 0 \end{cases}$$

При этом F_X имеет вид $F_X(t) = 1 - e^{-\lambda t}$.

4.6 Совместное распределение случайных величин.

Пусть $(\Omega, \mathcal{A}, P)$ — вероятностное пространство, и X, Y — случайные величины на нём.

Предложение. Для всякого $B \in \mathcal{B}(\mathbb{R}^2)$ выполнено

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

Доказательство. Рассмотрим отображение $\varphi\colon \Omega \to \mathbb{R}^2$ по правилу $\varphi(w) = (X(w), Y(w))$. Тогда система множеств $\mathcal{C} = \{B \subset \mathbb{R}\colon \varphi^{-1}(B) \in \mathcal{A}\}$ является σ-алгеброй (тривиально). Заметим, что все прямоугольники вида $\langle a, b\rangle \times \langle c, d\rangle \in \mathcal{C}$, т.к.

$$\varphi^{-1}\left(\langle a, b\rangle \times \langle c, d\rangle\right) = \{w\colon X(w) \in \langle a, b\rangle, Y(w) \in \langle c, d\rangle\} = \underbrace{\{w\colon X(w) \in \langle a, b\rangle\}}_{\in \mathcal{C}} \cap \underbrace{\{w\colon Y(w) \in \langle c, d\rangle\}}_{\in \mathcal{C}}$$

Таким образом $\mathcal{C}$ — σ-алгебра, содержащая все прямоугольники $\implies$ $\mathcal{B}(\mathbb{R}^2) \subset \mathcal{C}$. ($\mathcal{B}(\mathbb{R}^2) = \sigma$(все прямоугольники), доказывало в пункте 2.3 конспекта по мере Лебега). ■

Определение 27. *Совместным распределением* случайных величин X, Y называется вероятностная мера $\mu_{X,Y}$ на $\mathcal{B}(\mathbb{R}^2)$, определяемая как

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

Определение 28. Функция

$$F_{X,Y}(t, s) = P(\{w\colon X(w) \leqslant t, Y(w) \leqslant s\}) = \mu_{X,Y}((-\infty, t] \times (-\infty, s])$$

называется *функцией совместного распределения* (функцией распределения случайного вектора) случайных величин X, Y.

Заметим, что

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

Предложение. Для функции совместного распределения выполнены следующие свойства:

1. для всякого $(a, b] \times (c,d]$ выполнено $F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c) \geqslant 0$;
2. F непрерывна справа по совокупности переменных;
3. $\lim\limits_{(x,y)\to(u,v)} F(X, Y) = 0$, если хотя бы одна из переменных u, v равна $-\infty$;
4. $\lim\limits_{(x,y)\to(+\infty,-\infty)} F(X, Y) = 1$;

Доказательство. 1) Доказательство аналогично одномерному случаю: выбираем подпрямоугольник, выражаем вероятностую меру на данном прямоугольнике через меру на выбранном и получаем "неубывание".
2) Заметим, что

$$\bigcup_{k=1}^{\infty}\{w\colon X(w) \leqslant t + \frac{1}{k}, Y(w) \leqslant s + \frac{1}{k}\} = \{w\colon X(w) \leqslant t, Y(w) \leqslant s\}$$

По непрерывности вероятностной меры:

$$P\left(X \leqslant t + \frac{1}{k}, Y \leqslant s + \frac{1}{k}\right) \to P(X \leqslant t, Y \leqslant s)$$

в частности,

$$\forall\, \varepsilon > 0 \exists\, k\colon P\left(X \leqslant t + \frac{1}{k}, Y \leqslant s + \frac{1}{k}\right) \leqslant P(X \leqslant t, Y \leqslant s) + \varepsilon$$

С другой стороны $P\left(X \leqslant t + \frac{1}{k}, Y \leqslant s + \frac{1}{k}\right) \geqslant P(X \leqslant t, Y \leqslant s)$. Если теперь $t_n \to t, t_n \geqslant t$ и $s_n \to s, s_n \geqslant s$, то $\exists\, n_0\colon \forall\, n > n_0 \implies t \leqslant t_n \leqslant t + \frac{1}{k}, s \leqslant s_n \leqslant s + \frac{1}{k}$. В силу монотонности функции распределения:

$$F_{X,Y}(t, s) \leqslant F_{X,Y}(t_n, s_n) \leqslant F_{X,Y}\left(t + \frac{1}{k}, s + \frac{1}{k}\right) \leqslant F_{X,Y}(t, s) + \varepsilon$$

Отсюда, по определению предела имеем $F_{X,Y}(t_n, s_n) \to F_{X,Y}(t, s)$.
3) & 4) аналогично. ■

Теорема 4.3. *Для всякой функции $F\colon \mathbb{R}^2 \to [0,1]$, обладающими свойствами 1)-4) существует вероятностное пространство $(\Omega, \mathcal{A}, P)$, и пара случайных величин X, Y с функцией совместного распределения F. Кроме того распределение пары случайных величин $\mu_{X,Y}$ однозначно задаётся их функцией совместного распределения.*

Заметим также, что если известно совместное распределение величин, то можно найти распределение каждой из компонент:

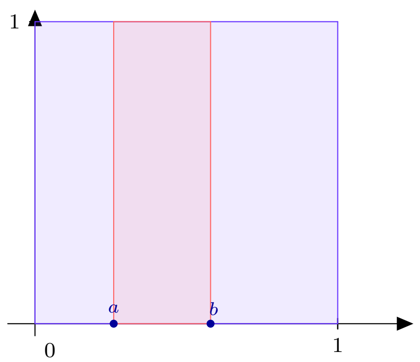
$$F_X(t) = P(X \leqslant t, Y \in \mathbb{R}) = \lim_{s\to\infty} F_{X,Y}(t,s) \qquad\qquad \mu_X(U) = \mu_{X,Y}(U \times \mathbb{R})$$

Однако в другую сторону это, вообще говоря, не верно.

4.7 Пример неоднозначного задания совместного распределения распределениями компонент.

Двумерное равномерное распределение: пара случайных величин (X, Y) имеет равномерное распределение на множестве $D \iff P((X,Y) \in B) = \dfrac{S_B}{S_D}$.

Первый случай: Пусть в квадрате $[0,1]\times[0,1]$ случайно выбирается точка (x, y). Случайные величины $X(x,y) = x, Y(x,y) = y$ имеют равномерное распределение на $[0,1]$, и их совместное распределение является равномерным на квадрате.
Второй случай: Пусть теперь мы случайно выбираем точку (x, y) на диагонали квадрата $[0,1]\times[0,1]$. Тогда распределение каждой из компонент не изменилось, однако их совместное распределение совсем другое.

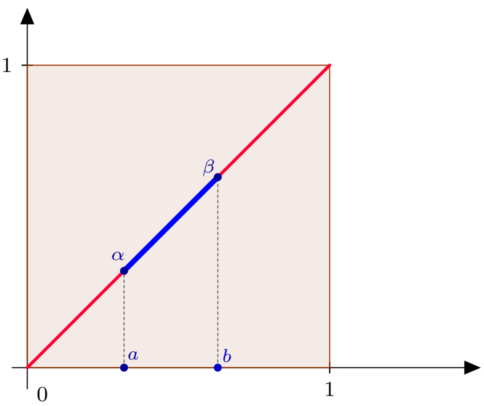


а) Первый случай

$\mu_{X,Y}(B) = S_B$

$P(X \in [a,b]) = b - a$

$P(Y \in [a,b]) = b - a$



б) Второй случай

$\mu_{X,Y} = P((X,Y) \in [\alpha,\beta]) = \dfrac{\beta-\alpha}{\sqrt{2}}$

$P(X \in [a,b]) = \dfrac{(b-a)\sqrt{2}}{\sqrt{2}} = b - a$

$P(Y \in [a,b]) = \dfrac{(b-a)\sqrt{2}}{\sqrt{2}} = b - a$

В этих двух случаях случайные величины X, Y имеют одинаковое распределение, однако их совместные распределения отличаются. Так, например, для первого случая вероятность выбора точки из любого множества, непересекающегося с диагональю не равна нулю, однако это не верно для второго случая. На основании этих случаев мы можем сделать вывод, что если известно распределение каждой из случайных величин, то даже тогда мы не можем узнать их совместное распределение.

4.8 Абсолютно непрерывные случайные распределения.

Пусть $(\Omega, \mathcal{A}, P)$ — вероятностое пространство, в X, Y — случайные величины на нём.

Определение 29. Совместное распределение случайных величин X, Y называется *абсолютно непрерывным*, если существует такая интегрируемая и неотрицательная функция $\rho_{X,Y}(t,s)$, что

$$F_{X,Y}(t,s) = \iint\limits_{(-\infty,t]\times(-\infty,s]} \rho_{X,Y}(t,s)dtds$$

Функция $\rho_{X,Y}$ называется *плотностью* совместного распределения случайных величин X, Y.

Можно показать, что

$$P(a < X \leqslant b, c < Y \leqslant d) = \mu_{X,Y}((a,b]\times(c,d]) = \iint\limits_{(a,b]\times(c,d]} \rho_{X,Y}(t,s)dtds$$

Кроме того в каждой точке непрерывности функции $\rho_{X,Y}$ выполнено равенство

$$\frac{\partial^2}{\partial t\partial s}F_{X,Y}(t,s) = \rho_{X,Y}(t,s)$$

Если известна плотность совместного распределения $\rho_{X,Y}$, то можно найти плотность каждой из компонент:

$$F_X(t) = P(X \leqslant y, Y \in \mathbb{R}) = \int\limits_{-\infty}^{t} dx \int\limits_{-\infty}^{\infty} \rho_{X,Y}(x,y)dy \implies$$

$$\implies F'_X(t) = \rho_X(t) = \int\limits_{-\infty}^{\infty} \rho_{X,Y}(x,y)dy$$

Заметим однако, что совместное распределение может не иметь плотности вообще.

На самом деле можно доказать, что

$$\mu_{X,Y} = P((X,Y)\in B) = \iint\limits_{B} \rho_{X,Y}(t,s)dtds$$

Теорема 4.4. *Пусть распределение X, Y задано плотностью $\rho_{X,Y}$. Рассмотрим случайные величины $\xi = f(X,Y), \eta = g(X,Y)$. Пусть отображение $\varphi\colon (x,y) \to \left(\underbrace{f(x,y)}_{u}, \underbrace{g(x,y)}_{v}\right)$ биективно, φ, φ^{-1} — непрерывно дифференцируемы, и $\det J_\varphi \neq 0$. Тогда $\rho_{\xi,\eta} = \rho_{X,Y}\left(\varphi^{-1}(u,v)\right) \cdot \left|\det J_\varphi^{-1}(\varphi^{-1}(u,v))\right|$.*

Доказательство. Свойство матрицы Якоби: матрица Якоби обратной функции, взятая в точке (u,v) равна обратной матрице Якоби функции, взятая в точке $f^{-1}(u,v)$. Т.е. пусть ψ — функция, тогда

$$J_{\psi^{-1}}(u,v) = \left(J_\psi^{-1}(\psi^{-1}(u,v))\right)$$

Заметим, что

$$P\left((\xi,\eta)\in B\right) = P\left((X,Y)\in\varphi^{-1}(B)\right) = \iint\limits_{\varphi^{-1}(B)} \rho_{X,Y}(x,y)dxdy$$

Сделаем замену: $(u,v) = \varphi(x,y) \implies (x,y) = \varphi^{-1}(u,v)$. Тогда

$$\iint\limits_{\varphi^{-1}(B)} \rho_{X,Y}(x,y)dxdy = \iint\limits_{B} \rho_{X,Y}(\varphi^{-1}(u,v)) \cdot \left|\det\left(J_\varphi^{-1}(\varphi^{-1}(u,v))\right)\right| dudv$$

Таким образом мы получили, что

$$P\left((\xi,\eta)\in B\right)=\iint\limits_B \rho_{X,Y}(\varphi^{-1}(u,v))\cdot\left|\det\left(J_\varphi^{-1}(\varphi^{-1}(u,v))\right)\right|dudv \iff \rho_{\xi,\eta}=\rho_{X,Y}(\varphi^{-1}(u,v))\cdot\left|\det\left(J_\varphi^{-1}(\varphi^{-1}(u,v))\right)\right|$$

■

4.9 Многомерное равномерное распределение.

Пусть $(\Omega,\mathcal{A},P)$ — вероятностное пространство, а $\xi_1,\xi_2,\dots,\xi_n$ — случайные величины на нём.

Определение 30. Вектор $(\xi_1,\xi_2,\dots,\xi_n)$ имеет *равномерное распределение* на множестве $B\subset\Omega$, если

$$P\left((\xi_1,\xi_2,\dots,\xi_n)\in B\right)=\frac{\mathrm{vol}(B)}{\mathrm{vol}(\Omega)}=\int\limits_B\cdots\int\frac{1}{\mathrm{vol}(\Omega)}dx_1dx_2\dots dx_n$$

Иначе говоря, вектор равномерно распределён, если его распределение задано плотностью

$$\rho(x_1,x_2,\dots,x_n)=\begin{cases}\dfrac{1}{\mathrm{vol}(\Omega)} & (x_1,x_2,\dots,x_n)\in B\\ 0 & (x_1,x_2,\dots,x_n)\notin B\end{cases}$$

Под vol везде подразумевается *разумная функция*, отвечающая нашему понимаю измерения объёма в n-мерном случае. В одномерном случае эта функция эквивалентна мере Лебега, в двумерном — площади, в трёхмерном — объёму.

Пусть теперь множество B — произвольное. Тогда вектор $(\xi_1,\xi_2,\dots,\xi_n)$ равномерно на нём распределён, если

$$P\left((\xi_1,\xi_2,\dots,\xi_n)\in B\right)=\frac{\mathrm{vol}(B\cap\Omega)}{\mathrm{vol}(\Omega)}=\int\limits_B\cdots\int\frac{1}{\mathrm{vol}(\Omega)}\cdot I_\Omega(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n$$

где $I_\Omega(x_1,x_2,\dots,x_n)$ возвращает 1, если $(x_1,x_2,\dots,x_n)\in\Omega$ (индикатор). Иначе говоря, вектор равномерно распределён на произвольном множестве B, если его распределение задано плотностью:

$$\rho(x_1,x_2,\dots,x_n)=\begin{cases}\dfrac{1}{\mathrm{vol}(\Omega)} & (x_1,x_2,\dots,x_n)\in B\cap\Omega\\ 0 & (x_1,x_2,\dots,x_n)\notin B\cap\Omega\end{cases}$$

4.10 Независимые случайные величины.

Пусть $(\Omega,\mathcal{A},P)$ — вероятностное пространство, а X,Y — случайные величины на нём.

Определение 31. Случайные величины X,Y называются *независимыми* тогда и только тогда, когда

$$F_{X,Y}(t,s)=F_X(t)F_Y(s)$$

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

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

Доказательство.
$\Leftarrow$: Очевидно по определению (лучи — Борелевские множества).
$\Rightarrow$: Пусть B — луч, если $P(Y\in B)=0$, тогда с левой стороны 0, и справой тоже 0, т.к. $\{X\in A,Y\in B\}\subset\{Y\in B\}$.

Далее считаем, что $P(Y\in B)\neq 0$. Тогда положим $Q(A)=\dfrac{P(X\in A,Y\in B)}{P(Y\in B)}$. Нетрудно понять, что $Q\left(\bigcup\limits_{k=1}^{\infty}A_k\right)=$

$\sum\limits_{k=1}^{\infty} Q(A_k)$. Отсюда следует, что Q — счётно-аддитивная вероятностная мера на $\mathcal{B}(\mathbb{R})$. Известно, что $Q(A) = \mu_X(A)$, когда A — луч. Тогда мы имеем: Q — счётно-аддитивная мера на алгебре лучей. По теореме Лебега: существует единственная мера, которая совпадает с Q на этой алгебре, и продолжает её на σ-алгебру, порождённую этой алгеброй. В конспекте по теории меры Лебега (пункт 2.3) доказывалось, что σ-алгебра, порождённая алгеброй лучей совпадает с $\mathcal{B}(\mathbb{R})$. Значит, μ_X — то самое, единственное продолжение Q на любое Борелевское множество.

Пусть теперь $A \in \mathcal{B}$ — фиксированное Борелевское множество, и $P(X \in A) \neq 0$ (понятно, что если $P(X \in A) = 0$, то всё выполнено). Рассмотрим теперь вероятностную меру

$$Q'(B) = \frac{P(X \in A, Y \in B)}{P(X \in A)}$$

По аналогии с $Q(A), Q'(B)$ счётно-аддитивна. И $Q'(B) = \mu_Y(B)$, если B — луч. Тогда, по аналогии с $Q(A)$ всё снова выполнено для любого Борелевского B. ■

Замечание. Обоснование счётной-аддитивности $Q(A)$:
Заметим, что $\{X \in A, Y \in B\} = \{X \in A\} \cap \{Y \in B\}$. Тогда

$$Q\left(\bigcup_{k=1}^{\infty} A_k\right) = \frac{P\left(\left\{X \in \bigcup\limits_{k=1}^{\infty} A_k, Y \in B\right\}\right)}{P(Y \in B)} = \frac{P\left(\left\{X \in \bigcup\limits_{k=1}^{\infty} A_k\right\} \cap \{Y \in B\}\right)}{P(Y \in B)} = \frac{P\left(\bigcup\limits_{k=1}^{\infty} \{X \in A_k\} \cap \{Y \in B\}\right)}{P(Y \in B)} =$$

$$= \sum_{k=1}^{\infty} Q(A_k)$$

В последнем переходе использована счётная аддитивность P.

Определение 32. Функция $f\colon \mathbb{R} \to \mathbb{R}$ называется *Борелевской*, если $f^{-1}((-\infty, t]) \in \mathcal{B}(\mathbb{R})$ для всякого действительного t.

Следствие. Если f, g — Борелевские функции, а X, Y — независимые случайные величины на вероятностном пространстве, то случайные величины $f(X), g(Y)$ также независимы.
Действительно

$$P(f(X) \in A, g(Y) \in B) = P(X \in f^{-1}(A), Y \in g^{-1}(B)) = P(X \in f^{-1}(A)) \cdot P(Y \in g^{-1}(B)) = P(f(X) \in A) \cdot P(g(Y) \in B)$$

Предложение. Пусть распределения X и Y заданы плотностями. Тогда X и Y независимы тогда и только тогда, когда плотность их совместного распределения имеет вид

$$\rho_{X,Y}(t, s) = \rho_X(t)\rho_Y(s)$$

Доказательство.
$\Rightarrow$:

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

$\Leftarrow$:

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

■

Следствие (Формула свёртки)**.** Пусть X, Y независимы, и их распределения заданы плотностями. Тогда распределение $Z = X + Y$ задано плотностью

$$\rho_Z(z) = \int\limits_{-\infty}^{\infty} \rho_X(t)\rho_Y(z - t)dt$$

Доказательство. По определению функции распределения:

$$F_Z(t) = P(X+Y \leqslant t) = \iint\limits_{x+y\leqslant t} \rho_X(x)\rho_Y(y)dxdy$$

Сделаем замену: $u = x+y, v = x$. Тогда

$$\begin{cases} u = x+y \\ v = x \end{cases} \iff \begin{cases} x = v \\ y = u - v \end{cases}, \quad \det J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

И, подставляя замену в интеграл имеем

$$F_Z(t) = \iint\limits_{x+y\leqslant t} \rho_X(x)\rho_Y(y)dxdy = \iint\limits_{u\leqslant t} \rho_X(v)\rho_Y(u-v)dudv = \int\limits_{-\infty}^{t} du \int\limits_{-\infty}^{\infty} \rho_X(v)\rho_Y(u-v)dv$$

Отсюда $\rho_Z(u) = \int\limits_{-\infty}^{\infty} \rho_X(v)\rho_Y(u-v)dv.$ ■

4.11 Математическое ожидание в общем случае: ограниченные случайные величины.

Определение 33. Пусть X — ограниченная случайная величина (т.е. существует $b > 0\colon \forall w \in \Omega \implies |X(w)| < b$). Тогда её *математическим ожиданием* называется $\lim\limits_{n\to\infty} \mathbb{E}(X_n)$, где X_n — последовательность произвольных случайных величин с конечным числом значений, равномерно сходящаяся к X.

Проверим корректность определения, а именно проверим, что

Лемма 4.5. Для произвольной ограниченной случайной величины X найдётся последовательность случайных величин с конечным числом значений X_n, равномерно сходящаяся к X.

Доказательство. Ограниченность случайной величины X означает что $\exists b > 0\colon \forall w \in \Omega \implies |X(w)| < b$. Это равносильно тому что X принимает значения из интервала $(-b, b)$. Пройдёмся по этому интервалу, и разобьём его на сколько-то дизъюнктных полуинтервалов длины $\dfrac{2b}{n}$. Тогда k-ый полуинтервал будет иметь вид: $\left[-b + (k-1)\dfrac{2b}{n}, -b + k\dfrac{2b}{n}\right)$.

Обозначим k-ый полуинтервал за J_k. Положим случайную величину X_n — для произвольной $w \in \Omega$ выполняется, что $X(w)$ попало в k-ый полуинтервал. Тогда

$$X_n(w) = \sum_{k=1}^{n} \left(-b + (k-1)\frac{2b}{n}\right) \cdot I_{X(w)\in J_k}$$

где $I_{X(w)\in J_k}$ — индикатор соответствующего события. Таким образом X_n принимает не более чем n значений. Выберем теперь произвольную $w_0 \in \Omega$. Известно, что $X(w_0)$ попдает в какой-то полуинтервал, пусть она попала в J_k. Тогда $X_n(w_0) = -b + (k-1)\dfrac{2b}{n}$, т.е. это в точности начало k-ого полуинтервала. Оценим теперь $|X_n(w_0) - X(w_0)|$. Заметим, что $X_n(w_0)$, как уже отмечалось выше, конец k-ого полуинтервала, а $X(w)$ — какая-то точка из этого полуинтервала. Тогда модуль разницы между ними мы можем оценить сверху длиной всего полуинтервала:

$$|X_n(w_0) - X(w_0)| \leqslant |J_k| = \frac{2b}{n}$$

Получили, что в каждой точке разница случайных величин не превосходит $\dfrac{2b}{n}$. Ну тогда выполняется, что

$$\sup |X_n(w) - X(w)| \leqslant \frac{2b}{n} \to 0$$

А это в точности определение равномерной сходимости. ■

Замечание. Если не появилось понимания того что произошло, очень рекомендую посмотреть это доказательство в лекции, там оно получилось на редкость удачное. Таймкод: $26:22$.

Итак, мы теперь знаем, что такие последовательности найдутся. Для полного счастья и уверенности в корректности определения нам теперь нужно понять, что предел их матожиданий существует.

Предложение. Для произвольной ограниченной случайной величины X, и для произвольной последовательности случайных величин, имеющих конечное число значений X_n, равномерно сходящейся к X существует предел $\lim\limits_{n\to\infty} \mathbb{E}(X_n)$. Более того, для другой такой последовательности Y_n, также равномерно сходящейся к X выполнено $\lim \mathbb{E}(X_n) = \lim \mathbb{E}(Y_n$.

Доказательство. Заметим, что

$$|\mathbb{E}(X_n) - \mathbb{E}(X_k)| = \mathbb{E}(|X_n - X_k|) \leqslant \sup_{w\in\Omega} |X_n(w) - X_k(w)|$$

Отсюда следует фундаментальность последовательности $\{\mathbb{E}(X_n)\}$, а значит её предел существует.
Пусть теперь Y_n — другая последовательность случайных величин с конечным числом значений, равномерно сходящаяся к X. Составим последовательность Z_n — по следующему правилу: на чётных местах расположим элементы последовательности X_n, на нечётных Y_n. Тогда Z_n будет обладать теми же свойствами, и, в частности, равномерно сходиться к X. Тогда уже доказали, что существует предел $\{\mathbb{E}(Z_n)\}$, и тогда $\lim \mathbb{E}(Y_n) = \lim \mathbb{E}(X_n)$ как пределы подпоследовательностей сходящейся последовательности. ■

Переходим к свойствам вновь определённого матожидания:

Предложение. Для произвольных ограниченных случайных величин X, Y выполнены свойства матожидания:

1. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
2. Если $X \geqslant 0$ почти наверное, то $\mathbb{E}(X) \geqslant 0$, в частности если $X \geqslant Y$, то $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$.

Доказательство.
1) Если X_n, Y_n — принимают конечное число значений, и $X_n \rightrightarrows X, Y_n \rightrightarrows Y$, то $aX_n + bY_n \rightrightarrows aX + bY$, и $\mathbb{E}(aX + bY) = \lim \mathbb{E}(aX_n + bY_n) = a \lim \mathbb{E}(X_n) + b \lim \mathbb{E}(Y_n) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

2) Пусть $\forall w \in \Omega \implies X(w) \geqslant 0$, а X_n — принимает конечное число значений, и $X_n \rightrightarrows \sqrt{X}$. Тогда $X_n^2 \rightrightarrows X$, и $\mathbb{E}(X) = \lim \mathbb{E}(X_n^2) \geqslant 0$.
Вторая часть утверждения: пусть для произвольного $X \geqslant 0$ почти наверное выполнено $\mathbb{E}(X) = \mathbb{E}(X \cdot I_{X\geqslant 0}) + \mathbb{E}(X \cdot I_{X<0})$. Известно, что при $X < 0, -X$ — неотрицательая случайная величина. Тогда для неё выполнено $\mathbb{E}(-X \cdot I_{X<0}) \geqslant 0$. В силу ограниченности X: $\mathbb{E}((b - -X) \cdot I_{X<0}) = \mathbb{E}((b + X) \cdot I_{X<0}) \geqslant 0$. Теперь следим за руками:

$$\mathbb{E}((b+X) \cdot I_{X<0}) \geqslant 0 \iff \mathbb{E}(b \cdot I_{X<0}) + \mathbb{E}(X \cdot I_{X<0}) \geqslant 0 \iff -\mathbb{E}(X \cdot I_{X<0}) \leqslant b \underbrace{\mathbb{E}(I_{X<0})}_{P(X<0)} \iff$$

$$\iff 0 \leqslant \mathbb{E}(-X \cdot I_{X<0}) \leqslant b \cdot P(X < 0) = 0$$

Соответственно при $X \geqslant Y$ почти наверное имеем $X - Y \geqslant 0$ почти наверное, и тогда $\mathbb{E}(X - Y) \geqslant 0 \iff \mathbb{E}(X) \geqslant \mathbb{E}(Y)$. ■

4.12 Математическое ожидание в общем случае: неотрицательные случайные величины.

Определение 34. Пусть $\forall w \in \Omega \implies X(w) \geqslant 0$ — неотрицательная случайная величина. Её *математическим ожиданием* называется конечное число

$$\mathbb{E}(X) = \sup\{\mathbb{E}(U) \colon 0 \leqslant U \leqslant X, U - \text{ограниченная }\}$$

Предложение. Для неотрицательных случайных величин X, Y выполнены следующие свойства матожидания:

1. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, при $a, b \geqslant 0$.
2. Если $X \geqslant Y \geqslant 0$, то $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$. В частности, если $\mathbb{E}(X)$ — конечно, то и $\mathbb{E}(Y)$ — конечно.
3. Если $X = 0$ почти наверное, то $\mathbb{E}(X) = 0$.

Доказательство.
1) Ясно, что доказать для $a = b = 1$ достаточно. Если $0 \leqslant U \leqslant X$, и $0 \leqslant V \leqslant Y$, то $U + V \leqslant X + Y \iff \mathbb{E}(X) + \mathbb{E}(Y) \leqslant \mathbb{E}(X + Y)$.
В другую сторону: пусть $0 \leqslant Z \leqslant X + Y$, $U = \min(X, Z)$, $V = Z - U$. Тогда $0 \leqslant U \leqslant X, V = (Z - X) \cdot I_{X<Z} \leqslant X + Y - X = Y$. Тогда $\mathbb{E}(Z) = \mathbb{E}(U + V) \leqslant \mathbb{E}(X) + \mathbb{E}(Y) \implies \mathbb{E}(X + Y) \leqslant \mathbb{E}(X) + \mathbb{E}(Y)$.
Все введённые случайные величины считались ограниченными.
2) Очевидно.
3) Пусть $0 \leqslant U \leqslant X$ — ограниченная случайная величина, и $X = 0$ почти наверное. Тогда $U = 0$ почти наверное и

$$\mathbb{E}(U) = \mathbb{E}(U \cdot I_{U \neq 0}) \leqslant \mathbb{E}(\sup U \cdot I_{U \neq 0}) = \sup U \cdot P(U \neq 0) = 0$$

■

4.13 Математическое ожидание в общем случае: реально общий случай.

Обозначение 4. В этой главе мы будем применять следующие два обозначения:

1. $X^+ = \max\{X, 0\} \geqslant 0$.
2. $X^- = \max\{-X, 0\} \geqslant 0$.

При этом заметим, что произвольная случайная величина X обладает свойством $X = X^+ - X^-$.

Пусть X^+, X^- обладают конечным матожиданием.

Определение 35. *Математическим ожиданием* произвольной случайной величины X называется число

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

Проверим корректность определения в следующем смысле: если есть случайные величины $U \geqslant 0, V \geqslant 0$, и X через них выражается как $X = U - V$, то с одной стороны X выражается через них, а с другой стороны $X = X^+ - X^-$, и в этой ситуации нам бы хотелось чтобы выполнялось следующее:

$$\begin{cases} X = U - V \\ X = X^+ - X^- \end{cases} \implies \mathbb{E}(U) - \mathbb{E}(V) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

Ну и действительно имеем $X = U - V = X^+ - X^- \iff U + X^- = X^+ + V \implies \mathbb{E}(U + X^-) = \mathbb{E}(X^+ + V)$. В силу неотрицательности вообще всех случайных величин в этом равенстве имеем $\mathbb{E}(U) + \mathbb{E}(X^-) = \mathbb{E}(X^+) + \mathbb{E}(V)$. Переносим в правильную сторону и получаем то что хотели.

Предложение. Для произвольных случайных величин X, Y выполнены следующие свойства матожидания:

1. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
2. Если $X \geqslant 0$ почти наверное, то $\mathbb{E}(X) \geqslant 0$, и если $X \geqslant Y$, то $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$.
3. Если $X \geqslant 0$ почти наверное, и $\mathbb{E}(X) = 0$, то $X = 0$ почти наверное.
4. $|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$.

Доказательство.
1) Ясно, что достаточно доказать для случая $a, b \geqslant 0$. Известно, что $aX + bY = (aX^+ + bY^+) - (aX^- + bY^-)$, причём каждое из слагаемых — неотрицательная случайная величина. Имеем:

$$\mathbb{E}(aX + bY) = \mathbb{E}(aX^+ + bY^+) - \mathbb{E}(aX^- + bY^-) = a\mathbb{E}(X^+) + b\mathbb{E}(Y^+) - a\mathbb{E}(X^-) - b\mathbb{E}(Y^-) =$$

$$= a\left(\mathbb{E}(X^+) - \mathbb{E}(X^-)\right) + b\left(\mathbb{E}(Y^+) - \mathbb{E}(Y^-)\right) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

2) $X \geqslant 0$ почти наверное $\implies X^- = 0$ почти наверное $\implies \mathbb{E}(X) = \mathbb{E}(X^+) \geqslant 0$.
3)

$$P(X > 0) = P\left(\bigcup_{k=1}^{\infty} X \geqslant \frac{1}{k}\right) = \sum_{k=1}^{\infty} P\left(X \geqslant \frac{1}{k}\right)$$

Рассмотрим случайную величину $\frac{1}{k}\cdot I_{X\geqslant\frac{1}{k}}$. Понятно, что она не превосходит X. Тогда с одной стороны $\mathbb{E}\left(\frac{1}{k}\cdot I_{X\geqslant\frac{1}{k}}\right)\leqslant \mathbb{E}(X)=0$, а с другой стороны $\mathbb{E}\left(\frac{1}{k}\cdot I_{X\geqslant\frac{1}{k}}\right)=\frac{1}{k}P\left(X\geqslant\frac{1}{k}\right)$. Тогда $\frac{1}{k}P\left(X\geqslant\frac{1}{k}\right)=0 \iff P\left(X\geqslant\frac{1}{k}\right)=0$. Из равенства выше получаем требуемое.
4) Аналогично дискретному случаю: понятно, что $-|X|\leqslant X\leqslant |X|$. Теперь по монотонности навешиваем матожидание и получаем требуемое. ■

4.14 Матожидание функции от случайной величины с абсолютно непрерывным распределением.

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

Лемма 4.6. Пусть X — произвольная неотрицательная случайная величина, $A_n\subset A_{n+1}, A_n\in\mathcal{A}$, причём A_n в объединении дают Ω. Тогда $\mathbb{E}(X)<\infty \iff \sup\limits_n \mathbb{E}(X\cdot I_{A_n})=M<\infty$. Более того: $\mathbb{E}(X)=\lim\mathbb{E}(X\cdot I_{A_n})=M$.

Доказательство. Пусть $0\leqslant U\leqslant X$ — произвольная ограниченная случайная величина. Тогда $\mathbb{E}(U\cdot I_{A_n})\leqslant\mathbb{E}(X\cdot I_{A_n})\leqslant M$. Известно, что $\mathbb{E}(U)=\mathbb{E}(U\cdot I_{A_n})+\mathbb{E}(U\cdot I_{\Omega\setminus A_n})$, при этом $\mathbb{E}(U\cdot I_{\Omega\setminus A_n})\leqslant\sup U\cdot P(\Omega\setminus A_n)$. Последняя вероятность стремится к нулю из-за вложенности множеств A_n. Тогда $\mathbb{E}(U)=\lim\limits_{n\to\infty}\mathbb{E}(U\cdot I_{A_n})$. Но т.к. по самой первой оценке при каждом n $U\cdot I_{A_n}\leqslant M$, то предел их ожиданий так же не превосходит M. А также ожидание супремума не превосходит M, но рассматривать супремум U всё равно что рассматривать X. Отсюда $\mathbb{E}(X)\leqslant M$.
В другую сторону: пусть теперь $\mathbb{E}(X)$ — конечно. Тогда очевидно, что $X\geqslant X\cdot I_{A_n}$. Отсюда по монотонности матожидания $\mathbb{E}(X)\geqslant\mathbb{E}(X\cdot I_{A_n})$, а значит это выполнено и для супремума. ■

Предложение. Пусть X — произвольная случайная величина, распределение которой задано плотностью ρ_X (т.е. это абсолютно непрерывная случайная величина). Пусть задана непрерывная функция f. Тогда матожидание $\mathbb{E}(f(X))$ существует тогда и только тогда, когда сходится несобственный интеграл

$$\int\limits_{-\infty}^{+\infty}|f(t)|\rho_X(t)dt$$

Более того, в случае сходимости

$$\mathbb{E}(f(X))=\int\limits_{-\infty}^{+\infty}f(t)\rho_X(t)dt$$

Доказательство. Достаточно доказать для $f\geqslant 0$. Иначе проводим аналогичные рассуждения для $f(X)=f^+(X)-f^-(X)$. Рассмотрим ограниченный отрезок $[-R,R]$. Разбив его на N подотрезков мы можем приблизить функцию f ступенчатыми функциями g_n, которые имеют вид $g_n=\sum\limits_{j=1}^{N_n}c_j\cdot I_{[a_j,b_j]}$. Причём эти ступенчатые функции будут равномерно сходиться к f. Тогда справедлива следующая оценка:

$$\begin{aligned}\left|\mathbb{E}(f(t)\cdot I_{X\in[-R,R]})-\mathbb{E}(g_n(t))\right|=\left|\mathbb{E}\left((f(t)-g_n(t))\cdot I_{X\in[-R,R]}\right)\right|\leqslant\mathbb{E}\left(|f(x)-g_n(x)|\cdot I_{X\in[-R,R]}\right)\leqslant\\ \leqslant\sup_{[-R,R]}|f(t)-g_n(t)|\end{aligned}$$

И такая же оценка справедлива для интегралов:

$$\left|\int\limits_{-R}^{R}f(t)\rho_X(t)dt-\int\limits_{-R}^{R}g_n(t)\rho_X(t)dt\right|=\int\limits_{-R}^{R}|f(t)-g_n(t)|\rho_X(t)dt\leqslant\sup_{[-R,R]}|f-g_n|\cdot\underbrace{\int\limits_{-R}^{R}\rho_X(t)dt}_{1}$$

В силу равномерной сходимости: $g_n \rightrightarrows f \iff \sup|f - g_n| \to 0$. Кроме того можно заметить, что

$$\mathbb{E}(g_n(X)) = \sum_{j=1}^{N_n} c_j \mathbb{E}(I_{X\in[a_j,b_j]}) = \sum_{j=1}^{N_n} c_j P(X \in [a_j, b_j]) = \sum_{j=1}^{N_n} c_j \int\limits_{a_j}^{b_j} \rho_X(t)dt = \int\limits_{-R}^{R} g_n(t)\rho_X(t)dt$$

Тогда из всех предыдущих оценок следует, что

$$\mathbb{E}(f(X) \cdot I_{X\in[-R,R]}) = \int\limits_{-R}^{R} f(t)\rho_X(t)dt$$

Устремляя $R \to \infty$ Получаем утверждение. ■

Замечание. На мой взгляд как-то скомканно, можно посмотреть в лекции по таймкоду $53:00$.

4.15 Матожидание произведения независимых случайных величин.

Предложение. Пусть X, Y — независимые случайные величины, имеющие матожидание. Тогда $X \cdot Y$ также имеет матожидание, и $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

Доказательство. Известно, что если X и Y независимы, то и $f(X), g(Y)$ независимы. Пусть X, Y — неотрицательны и ограничены, т.е. $\exists b\colon 0 \leqslant X, Y < b$. Положим функцию f_n:

$$f_n(X) = \sum_{j=0}^{n} \left(-b + (j-1)\frac{2b}{n}\right) \cdot I_{\left[-b+(j-1)\frac{2b}{n}, -b+j\frac{2b}{n}\right]}$$

Тогда $X_n = f(X), Y_n = f(Y)$ — случайные величины с конечным числом значений. Эти случайные величины независимы, и мы даже знаем, что $X_n \rightrightarrows X, Y_n \rightrightarrows Y$. Тогда по определению матожидания для ограниченных величин: $\mathbb{E}(X) = \lim \mathbb{E}(X_n), \mathbb{E}(Y) = \lim \mathbb{E}(Y_n)$. Т.к. X_n, Y_n принимают конечное число знаений, то про них известно, что

$$\mathbb{E}(X_n \cdot Y_n) = \mathbb{E}(X_n) \cdot \mathbb{E}(Y_n)$$

Т.к. случайные величины равномерно сходились по отдельности, то это же будет выполнено и для их произведения: $X_n \cdot Y_n \rightrightarrows X \cdot Y$. Тогда снова по определению матожидания ограниченных случайных величин:

$$\mathbb{E}(X \cdot Y) = \lim \mathbb{E}(X_n \cdot Y_n) = \lim \mathbb{E}(X_n) \cdot \lim \mathbb{E}(Y_n) = \mathbb{E}(X_n) \cdot \mathbb{E}(Y_n)$$

Пусть теперь X, Y — неотрицательные, неограниченные случайные величины. Тогда рассмотрим соответствующие им ограниченные $X \cdot I_{|X|<b}, Y \cdot I_{|Y|<b}$. Про них мы утверждение уже доказали. Устремляем теперь $R \to \infty$ и получаем требуемое утверждение.
Пусть X, Y — произвольные случайные величины. Тогда

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

Каждое из слагаемых есть неотрицательная случайная величина. ■

4.16 Неравенство Чебышева.

Предложение. Пусть $X \geqslant 0$ почти наверное. Тогда для всякого неотрицательного t выполнено неравенство Чебышева:

$$P(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$$

Доказательство. Понятно, что $t \cdot I_{X\geqslant t} \leqslant X$ почти наверное. Тогда по монотонности матожидания:

$$\mathbb{E}(t \cdot I_{X\geqslant t}) \leqslant \mathbb{E}(X) \iff tP(X \geqslant t) \leqslant \mathbb{E}(X) \iff P(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$$

■

4.17 Дисперсия, ковариация, коэффициент корреляции.

Жмяк (Серьёзно, всё ровно то же самое).