Конспект лекций по терверу.

Пешехонов Иван. БПМИ195

Дата последней компиляции: 30.11.2020 22:37

Содержание

1 Дискретное вероятностное пространство. 2

1.1 Вероятностое пространство, события, вероятностная мера, вероятность. 2

1.2 Формула включений и исключений. 3

1.3 Парадокс распределения подарков. 3

1.4 Доказательства существования и задача о конференции. 4

1.5 Бесконечное множество элементарных исходов и счётная аддитивность. 5

1.6 Условная вероятность, формула полной вероятности и формула Байеса. 7

1.7 Независимые события. Попарная независимость и независимость в совокупности. 7

1.8 Задача о билетах к экзамену. 8

1.9 Задача о сумасшедшей старушке. 8

1.10 Парадокс Байеса. 9

1.11 Парадокс Монти Холла. 9

2 Случайные величины на дискретном вероятностном пространстве. 11

2.1 Определение случайной величины и её распределение. 11

2.2 Примеры случайных дискретных величин. 11

2.3 Совместное распределение случайных величин. 12

2.4 Независимые случайные величины. Эквивалентное определение независимости случайных величин. . . 12

2.5 Математическое ожидание случайной величины. 13

2.6 Мат. ожидание функции от случайной величины. 13

2.7 Свойства мат. ожидания. 13

2.8 Балансировка векторов. 15

2.9 Дисперсия, ковариация, коэффициент корреляции. 15

2.10 Основные свойства дисперсии и ковариации. 16

2.11 Неравенство Коши-Буняковского и линал нам в анал. 16

2.12 Мат. ожидание и дисперсия биномиального распределения. 17

2.13 Неравенство Чебышева (Маркова). 17

3 Закон больших чисел. 19

3.1 Закон больших чисел в слабой форме. 19

3.2 Теорема Муавра-Лапласа. 19

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

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

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

$$P(B \cap A_{k+1}) = P((A_1 \cup A_2 \cup \ldots \cup A_k) \cap A_{k+1}) = P((A_1 \cap A_{k+1}) \cup \ldots \cup (A_k \cap A_{k+1})) =$$

$$= \sum_{i=1}^{k}(-1)^{i-1} \sum_{j_1<\ldots<j_i} P(A_{j_1} \cap \ldots \cap A_{j_i} \cap A_{k+1})$$

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

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

■

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 $P(A_k) > 1 - \frac{1}{n}$, 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 $P(A_k) > 1 - \frac{1}{n}$ 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 $A_1, A_2, \ldots,$ — 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:

$a_{1,1}$	$a_{1,2}$	$\ldots$	a_{1,n_1}
$a_{2,1}$	$a_{2,2}$	$\ldots$	a_{2,n_2}
	$\vdots$		

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 $\{x_1, \ldots, x_n, \ldots\}$ — 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 $\{x_1, \ldots, x_n, \ldots\}$ — 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 $\mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j) = \left(1 \cdot \dfrac{1}{2} + (-1) \cdot \dfrac{1}{2}\right) \cdot 2 = 0$ в силу того, что мы выбираем ε_i равновероятно с вероятностью выбора $\dfrac{1}{2}$.

С другой стороны, при $i = j$ имеем $\sum\limits_i (v_i, v_i) = |v_i|^2$ 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 $\mathbb{D}(S_n) = \sum\limits_{k=1}^{n} \mathbb{D}(X_k) = n\mathbb{D}(X_k) = n\mathbb{E}(X_k - \mathbb{E}(X_k))^2 = n\mathbb{E}(X_k - p)^2 = n\big((1-p)^2 \cdot p + p^2 \cdot (1-p)\big) = np(1-p)$.

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*

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

где $\varphi(x_m) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ *— плотность стандарного нормального распределения (что бы это ни значило).*