

ТВиМС - Коллоквиум 1

Цирк Максимус | [telegram](#)
Артём Мельников | [telegram](#)

Версия от 29.10.2020 07:41

Содержание

1	Вопросы	2
1.1	Дискретное вероятностное пространство. Свойства вероятностной меры на конечных и счётных множествах. Вероятностный алгоритм проверки числа на простоту.	2
1.2	Формула включений-исключений. Парадокс распределения подарков. Задача про конференцию.	2
1.3	Условная вероятность. Формула полной вероятности. Формула Байеса. Независимые события. Отличие попарной независимости от независимости в совокупности. Задача о билетах к экзамену.	3
1.4	Задача о сумасшедшей старушке. Парадокс Байеса. Парадокс Монти Холла.	4
1.5	Случайные величины на дискретном вероятностном пространстве, их распределение. Примеры дискретных распределений. Совместное распределение случайных величин. Независимые случайные величины. Эквивалентное определение независимости случайных величин.	5
1.6	Математическое ожидание случайной величины на дискретном вероятностном пространстве, эквивалентный способ вычисления математического ожидания. Математическое ожидание функции от случайной величины. Свойства математического ожидания: линейность, ожидание неотрицательной случайной величины, неотрицательная случайная величина с нулевым математическим ожиданием, связь модуля ожидания и ожидания модуля случайной величины, математическое ожидание произведения независимых случайных величин. Балансировка векторов.	6
1.7	Дисперсия, ковариация и коэффициент корреляции. Их связь и основные свойства: билинейность ковариации, случайная величина с нулевой дисперсией, дисперсия линейного образа случайной величины, дисперсия суммы независимых случайных величин. Неравенство Коши-Буняковского и геометрическая интерпретация ковариации, дисперсии и коэффициент корреляции. Вычисление ожидания и дисперсии у биномиального распределения.	7
1.8	Неравенство Чебышёва. Закон больших чисел в слабой форме.	8
1.9	Теорема Муавра-Лапласа (формулировка локальной и интегральной теорем, доказательство локальной теоремы в симметричном случае, идея доказательства интегральной теоремы).	9

1 Вопросы

1. Дискретное вероятностное пространство. Свойства вероятностной меры на конечных и счётных множествах. Вероятностный алгоритм проверки числа на простоту.

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

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

(ii) $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$ (правило суммы или аддитивность), называют вероятностной мерой, а значение $P(A)$ - вероятностью события A .

Для бесконечных вероятностных пространств, где $P(A) = \sum_{j: \omega_j \in A} p_{\omega_j}$ (как известно из курса анализа, эта сумма корректно определена, в том смысле, что результат не зависит от порядка суммирования) мы имеем более сильное свойство, чем (ii), а именно (ii)':

$P(\cup_n A_n) = \sum_n P(A_n)$ для произвольного не более чем счётного набора попарно непересекающихся событий A_n .

Тест Ферма проверки числа на простоту.

Пусть дано некоторое натуральное число $N > 1$. Мы хотим проверить является ли это число простым. Можно перебирать все простые делители до \sqrt{N} , но это очень долго. Хотелось бы иметь более быстрый способ проверки.

Если N простое число, то по малой теореме Ферма для всякого натурального числа b такого, что $\text{НОД}(b, N) = 1$, число $b^{N-1} - 1$ делится на N . Следовательно, если для некоторого b , удовлетворяющего условию $\text{НОД}(b, N) = 1$, число $b^{N-1} - 1$ не делится на N , то N не является простым. В этом случае будем говорить, что N не проходит тест Ферма по основанию b . Это наблюдение используют для построения простейшего алгоритма проверки числа на простоту: выберем случайное число b из промежутка $2, \dots, N-1$; если $\text{НОД}(b, N) \neq 1$, то N составное; если $\text{НОД}(b, N) = 1$, но $b^{N-1} - 1$ не делится на N , то N составное. В ином случае N - скорее простое.

Предположим, что существует хотя бы одно число a : $\text{НОД}(a, N) = 1$ и $a^{N-1} - 1$ не делится на N . Посмотрим, с какой вероятностью алгоритм выдаст ответ, что N - скорее простое. Пусть Z_N^* - группа всех чисел из промежутка $1, \dots, N-1$, взаимно простых с N . Если N проходит тест для основания $b \in Z_N^*$, то для основания ab число N уже тест не проходит. В противном случае $(ab)^{N-1} \equiv 1 \pmod{N}$ и $(b^{-1})^{N-1} \equiv 1 \pmod{N}$. Следовательно, $a^{N-1} \equiv (b^{-1})^{N-1} (ab)^{N-1} \equiv 1$, что противоречит предположению. Таким образом, каждому основанию b , для которого N проходит тест, можно сопоставить основание ab , для которого N тест не проходит. Значит, оснований, для которых N не проходит тест, не меньше, чем оснований, для которых N проходит тест на простоту. Поэтому в данной ситуации вероятность получить ответ, что N скорее простое, не более $\frac{1}{2}$.

Если независимым образом повторять описанную процедуру k раз, то вероятность получить неверный ответ не более $\left(\frac{1}{2}\right)^k$. Отметим, что бывают числа, которые проходят тест для всех оснований b . Это числа Кармайкла, например 561. Для них описанный алгоритм по понятным причинам не применим.

2. Формула включений-исключений. Парадокс распределения подарков. Задача про конференцию.

Предложение. Формула включений и исключений. Для произвольных событий $A_1, A_2, A_3, \dots, A_n$ верно равенство

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{i_1 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}).$$

Доказательство. Докажем утверждение по индукции. База: $P(A_1 \cup A_2) = P((A_1 \setminus A_2) \cup (A_2 \setminus A_1) \cup (A_1 \cap A_2)) = P(A_1 \setminus A_2) + P(A_2 \setminus A_1) + P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

Предположим, что утверждение выполняется для n множеств. Проверим, что оно выполнено и для $n+1$.

$$\begin{aligned} P(A_1 \cup \dots \cup A_{n+1}) &= P(A_1 \cup \dots \cup A_n) + P(A_{n+1}) - P((A_1 \cap A_{n+1}) \cup \dots \cup (A_n \cap A_{n+1})) = \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{i_1 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) + P(A_{n+1}) - \sum_{k=1}^n (-1)^{k-1} \sum_{i_1 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap A_{n+1}) = \\ &= \sum_{k=1}^{n+1} (-1)^{k-1} \sum_{i_1 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}). \end{aligned}$$

Парадокс распределения подарков

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

Пространство исходов состоит из всех возможных перестановок чисел $1, 2, \dots, n$, причём все перестановки являются равновероятными. Значит вероятность конкретной перестановки равна $\frac{1}{n!}$. Событие, состоящее в том, что конкретный человек вытащил подарок, который сам принёс, состоит из $(n-1)!$ исходов. Следовательно, вероятность такого события равна $\frac{1}{n}$. При больших n эта вероятность стремится к нулю. Можно было бы думать, что вероятность события: ни один человек не вытащил подарок, который сам принёс, стремится к единице, но это ошибочное мнение.

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

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} C_n^k \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}.$$

Таким образом, вероятность того, что ни один человек не вытащил подарок, который сам принёс, равна $1 - P(A_1 \cup \dots \cup A_n) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots$ и стремится к $\frac{1}{e}$.

Задача про конференцию.

Вероятностные рассуждения можно применять в комбинаторных задачах для доказательства существования объекта с заданными свойствами. Предположим, что заданы события A_1, \dots, A_n . Как проверить, что $P(\cap_k A_k) > 0$? Пусть B_k - противоположное событие к A_k . Тогда $P(\cap_k A_k) = 1 - P(\cup_k B_k) \geq 1 - \sum_k P(B_k)$. Если $\sum_k P(B_k) < 1$, то $P(\cap_k A_k) > 0$. Так как $P(B_k) = 1 - P(A_k)$, то $\sum_k P(B_k) < 1$ тогда и только тогда, когда $\sum_k P(A_k) > n - 1$ и пересечение событий A_k имеет положительную вероятность и, следовательно, непусто.

Перейдём к самой задаче, иллюстрирующей данное рассуждение:

В научном центре работают специалисты по 60 различным разделам компьютерных наук. Известно, что по каждому разделу в центре работает ровно 7 учёных, причём вполне может быть, что один учёный является специалистом сразу по нескольким направлениям. Все учёные должны принять участие в одной (и только одной) из двух конференций, одна из которых проходит в Канаде, а другая в Австралии. Оказывается, что всегда можно так распределить учёных по этим конференциям, что на каждой конференции будут присутствовать специалисты по всем 60 направлениям компьютерных наук.

Будем для каждого учёного выбирать конференцию простым подбрасыванием правильной монеты. Для k -го направления рассмотрим событие A_k , состоящее в том, что среди учёных этого направления окажутся и те, которые поехали в Канаду, и те, которые поехали в Австралию. Вероятность этого события равна $1 - 2^{-6}$ (нас устраивают все исходы кроме двух, когда все отправились на конференцию в одну страну). Остаётся заметить, что количество событий A_k равно 60 и вероятность каждого события больше $1 - \frac{1}{60}$.

3. Условная вероятность. Формула полной вероятности. Формула Байеса. Независимые события. Отличие попарной независимости от независимости в совокупности. Задача о билетах к экзамену.

Определение. Пусть $P(B) > 0$. Условной вероятностью события A при условии B называется число $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Если фиксировать событие B , то функция $P(\cdot|B)$ является новой вероятностной мерой, т.е. удовлетворяет свойствам (i) и (ii) (или (ii)'), если таковой была исходная вероятностная мера P . Равенство из определения часто переписывают в виде $P(A \cap B) = P(A|B)P(B)$ и называют правилом произведения.

Теорема. (Формула полной вероятности)

Пусть $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$ и $A_i \cap A_j = \emptyset$ для всех $i \neq j$. Предположим, что $P(A_i) > 0$. Тогда для каждого события B имеет место равенство $P(B) = \sum_i P(B|A_i)P(A_i)$.

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

Теорема. (Формула Байеса)

Пусть $P(A) > 0$ и $P(B) > 0$. Тогда имеет место равенство $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

Доказательство. Достаточно заметить, что $P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$. ■

Определение. События A и B называются независимыми, если $P(A \cap B) = P(A)P(B)$. В противном случае говорят, что события являются зависимыми.

Заметим, что независимость в совокупности (т.е. в случае, когда для A_1, A_2, \dots, A_n верно $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$ для произвольного $k \in \{2, \dots, n\}$ и произвольных $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$) не совпадает с попарной независимостью. Это иллюстрируется парадоксом независимости:

Бросаем правильную монету два раза. Рассмотрим 3 события A - при первом броске выпал орёл, B - при втором броске также выпал орёл, C - орёл выпал только один раз. Все эти события попарно независимы, но в совокупности они не являются независимыми, поскольку любые два из них однозначно определяют третье, например пересечение A и B исключает C , таким образом $P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$.

Задача о билетах к экзамену

Программа экзамена содержит N билетов, а студент выучил только n . На экзамене студенты по очереди подходят и тянут билет. Зависит ли вероятность вытянуть "хороший" билет от места в очереди?

Пусть студент стоит $k+1$ -м в очереди и пусть событие A заключается в том, что студент вытянул "хороший" билет. Пусть событие A_j заключается в том, что первые k студентов вытянули j "хороших" билетов.

$$\begin{aligned} \text{Тогда } P(A_j) &= \frac{C_n^j C_{N-n}^{k-j}}{C_N^k}. \text{ Таким образом, } P(A) = \sum_{j=0}^{\min\{k,n\}} P(A|A_j)P(A_j) = \sum_{j=0}^{\min\{k,n\}} \frac{n-j}{N-k} \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} = \\ &= \frac{n}{N} \sum_{j=0}^{\min\{k,n\}} \frac{C_{n-1}^j C_{(N-1)-(n-1)}^{k-j}}{C_{N-1}^k} = \frac{n}{N}. \end{aligned}$$

4. Задача о сумасшедшей старушке. Парадокс Байеса. Парадокс Монти Холла.

Задача о сумасшедшей старушке

У нас $N \geq 2$ людей стоят ждут посадки в самолёт. Одним из этих людей оказывается сумасшедшая старушка, которая всех расталкивает и садится первой на случайное место. Дальше заходят по очереди остальные пассажиры. Т.к. все люди очень вежливые (кроме данной старушки), если их место уже было занято, то они садятся на другое случайное место.

Мы хотим посчитать вероятность, с которой N -ый (последний) человек сядет на своё место. Рассмотрим по индукции.

База: $N = 2$

$P_N = \frac{1}{2}$ - шанс, что N человек сел на своё место (старушка села на место последнего пассажира или она села на своё)

Шаг: Для всех $k \leq N$ людей верен факт $P_k = \frac{1}{2}$, рассмотрим для P_{N+1}

Зададим 2 события:

A_i - старушка села на место i человека

B - последний пассажир ($N+1$) сел на своё место (вводим чисто, чтобы удобнее было работать)

Тогда хотим найти: $P_{N+1} = P(B) = \sum_i P(A_i) \cdot P(B|A_i)$ - формула полной вероятности

$P(A_i) = \frac{1}{N+1}$ - старушка садится на произвольное место из $N+1$

$P(B|A_1) = 1$ - если старушка села на своё место (считаю, что она 1 пассажир)

$P(B|A_{N+1}) = 0$ - $N+1$ не сядет на своё место, если там уже сидит старушка

$P(B|A_i) = \frac{1}{2}$ - в остальных случаях

$$\text{Тогда: } P(B) = \sum_i P(A_i) \cdot P(B|A_i) = \frac{1}{N+1} \cdot 1 + \frac{1}{N+1} \cdot 0 + \frac{1}{N+1} \cdot \frac{1}{2} \cdot (N+1-2) = \frac{2+N-1}{2(N+1)} = \frac{1}{2}$$

$$\Rightarrow P_N = \frac{1}{2} \text{ - вероятность, что последний человек сядет на своё место}$$

Парадокс Байеса

Предположим у нас есть очень редкая болезнь (читай ковид)

$P(Z_+) = 0.001$ - шанс, что ты больной

$P(Z_-) = 0.999$ - шанс, что здоров

Теперь также положим, что у нас есть тест на эту болезнь:

$P(T_+|Z_+) = 0.99$ - ты болен, тест показал, что ты болен

$P(T_+|Z_-) = 0.01$ - ты здоров, но тест ошибся и показал, что ты болен

Теперь, скажем, пришёл положительный тест. Посчитаем шанс, с которым ты здоров. (те тест ошибся)

$$P(Z_-|T_+) = \frac{P(T_+|Z_-) \cdot P(Z_-)}{P(T_+)} - \text{по формуле Байеса}$$

Надо посчитать $P(T_+)$:

$$P(T_+) = P(Z_-) \cdot P(T_+|Z_-) + P(Z_+) \cdot P(T_+|Z_+) = 0.999 \cdot 0.01 + 0.001 \cdot 0.99 = 0.01098 - \text{по формуле полной вероятности}$$

$$\text{Тогда: } P(Z_-|T_+) = \frac{P(T_+|Z_-) \cdot P(Z_-)}{P(T_+)} = \frac{0.01 \cdot 0.999}{0.01098} \approx 0.9098 \geq 0.9$$

те если тебе пришёл положительный тест, ты здоров с шансом больше 90% (интуитивно это очень сложно понять, но это объясняется редкостью нашего теоретического заболевания). Поэтому и надо делать больше 1 теста.

Парадокс Монти Холла

Есть игра: 3 двери за 1 из них автомобиль, за 2-мя другими - козы. Ты выбираешь дверь, после чего ведущий открывает 1 из оставшихся, за которой находится коза (он знает где автомобиль) и предлагает тебе изменить выбор. Банальный вопрос: стоит ли это делать? (увеличатся ли шансы найти автомобиль)

Пускай выбрали 1 дверь, а ведущий открыл 3 (чисто для нумерации). Введём события:

A_i - автомобиль за i -ой дверью

B - ведущий открыл 3-ю дверь

Тогда можем определить следующие вероятности:

$$P(B|A_1) = \frac{1}{2} - \text{автомобиль за 1 дверью, ведущий случайно выбирает дверь}$$

$$P(B|A_2) = 1 - \text{если автомобиль в 3, а выбрали 1, то ведущий точно откроет 2.}$$

$$P(B|A_3) = 0 - \text{ведущий не будет открывать автомобиль} - _ -$$

Посмотрим на шансы дверей после открытия 3-ей:

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

P.S. Парадоксом называется т.к. интуитивно многие считают, что вероятности стали $\frac{1}{2}$ и выбор не важен.

5. Случайные величины на дискретном вероятностном пространстве, их распределение. Примеры дискретных распределений. Совместное распределение случайных величин. Независимые случайные величины. Эквивалентное определение независимости случайных величин.

Определение. Случайной величиной на дискретном вероятностном пространстве $\Omega = \omega_1, \omega_2, \dots$ называют произвольную функцию $X : \Omega \rightarrow R$.

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

Пример. Бернуллиевская случайная величина.

0	1
q	p

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

Пример. Схема Бернулли (биномиальное распределение).

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

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

Случайная величина $X(\omega)$ - число единиц в исходе ω . Её таблица представлена выше.

Пример. Геометрическое распределение.

1	2	...	k	...
p	pq	...	pq^{k-1}	...

Данная случайная величина моделирует подбрасывание монетки до первого успеха.

Пример. Распределение Пуассона.

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

Определение. Пусть X, Y - две случайные величины на дискретном вероятностном пространстве с множествами (различных) значений $\{x_1, x_2, \dots, x_k, \dots\}$ и $\{y_1, y_2, \dots, y_k, \dots\}$ соответственно. Их совместным распределением называется вероятностная мера $\mu_{(X,Y)}$ на вероятностном пространстве всех пар (x_j, y_k) , для которой

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

Аналогично определяется совместное распределение трёх и более случайных величин.

Определение. Случайные величины X, Y с множествами значений $\{x_1, \dots, x_k, \dots\}$ и $\{y_1, \dots, y_k, \dots\}$ соответственно называются независимыми, если $\forall k, j : \mu_{(X,Y)}(\{(x_j, y_k)\}) = \mu_X(\{x_j\}) \cdot \mu_Y(\{y_k\})$. Или, другими словами: $\forall k, j : P(\omega : X(\omega) = x_j, Y(\omega) = y_k) = P(\{\omega : X(\omega) = x_j\} \cdot \{\omega : Y(\omega) = y_k\})$.

Аналогично определяется независимость трёх и более случайных величин.

6. Математическое ожидание случайной величины на дискретном вероятностном пространстве, эквивалентный способ вычисления математического ожидания. Математическое ожидание функции от случайной величины. Свойства математического ожидания: линейность, ожидание неотрицательной случайной величины, неотрицательная случайная величина с нулевым математическим ожиданием, связь модуля ожидания и ожидания модуля случайной величины, математическое ожидание произведения независимых случайных величин. Балансировка векторов.

Определение. Пусть случайная величина X принимает значения $\{x_1, x_2, \dots, x_n, \dots\}$ (предполагаем, что значения перечислены без повторов).

Математическим ожиданием случайной величины X называют число: $\mathbb{E}X = \sum_j x_j \cdot P(\omega : X(\omega) = x_j)$, в предположении абсолютной сходимости ряда.

Заметим, что если в определении ряд не сходится абсолютно, то говорят, что случайная величина не имеет конечного мат. ожидания.

Лемма (Эквивалентный способ вычисления математического ожидания). Пусть случайная величина X с конечным математическим ожиданием принимает значения y_k (не обязательно различные) на множествах B_k , причём события B_k попарно не пересекаются, тогда $\mathbb{E}X = \sum_k y_k P(B_k)$.

Доказательство. $\mathbb{E}X = \sum_j x_j \cdot P(X = x_j) = \sum_j \sum_{k: y_k = x_j} y_k \cdot P(B_k) = \sum_k y_k \cdot P(B_k)$ ■

Теорема (Математическое ожидание функции от случайной величины). Пусть X - случайная величина, принимающая значения $\{x_1, x_2, \dots\}$. Рассмотрим случайную величину $Y = \varphi(X)$, тогда: $\mathbb{E}Y = \mathbb{E}\varphi(X) = \sum_k \varphi(x_k) P(X = x_k)$, при условии абсолютной сходимости последнего ряда.

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

Аналогичная формула справедлива и для функции от нескольких случайных величин $\varphi(X_1, X_2, \dots, X_k)$.

Определение. Если некоторое свойство имеет место с вероятностью 1, то говорят, что оно имеет место почти наверное.

Свойства математического ожидания

(i) Линейность: $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y$.

(ii) Ожидание неотрицательной случайной величины: если $X \geq 0$ почти наверное, то $\mathbb{E}X \geq 0$.

(iii) Неотрицательная случайная величина с нулевым математическим ожиданием: если $X \geq 0$ почти наверное $\mathbb{E}X = 0$, то $X = 0$ почти наверное.

Доказательство. Докажем первое свойство:

Пусть $\{x_1, x_2, \dots\}$ и $\{y_1, y_2, \dots\}$ множества значений случайных величин X и Y соответственно. Пусть $A_i = \{X = x_i\}$ и $B_j = \{Y = y_j\}$. Тогда:

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

Так как случайная величина $\alpha X + \beta Y$ принимает значения $\alpha x_i + \beta y_j$ на множествах $A_i \cap B_j$, то получаем
 $(=) \mathbb{E}(\alpha X + \beta Y)$

Второе и третье утверждения следуют из определения. ■

Следствие. Связь модуля ожидания и ожидания модуля случайной величины.

Имеет место оценка: $|\mathbb{E}X| \leq \mathbb{E}|X|$

Доказательство. Достаточно заметить, что $-|X| \leq X \leq |X|$, далее возьмём математическое ожидание с двух сторон неравенства (возможность такого действия можно объяснять линейностью мат. ожидания) и получим:

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

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

Теорема (Математическое ожидание произведения независимых случайных величин). Если случайные величины X и Y независимы, и существуют математические ожидания $\mathbb{E}X$ и $\mathbb{E}Y$, то:

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

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

$$\text{рядов: } \mathbb{E}X \cdot \mathbb{E}Y = \left(\sum_i x_i P(A_i) \right) \cdot \left(\sum_j y_j P(B_j) \right) = \sum_{i,j} x_i y_j P(A_i) \cdot P(B_j) (=)$$

$$\text{Т.к } X \text{ и } Y - \text{независимые случайные величины имеем } (=) \sum_{i,j} x_i y_j P(A_i \cap B_j) = \mathbb{E}[X \cdot Y].$$
■

Балансировка векторов.

Пусть $v_1, v_2, \dots, v_n \in \mathbb{R}^n$, причём $\|v_j\| = 1$. Всегда существует такая расстановка знаков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \in \{-1, 1\}$, что $\|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n\| \leq \sqrt{n}$

Доказательство. Действительно, будем расставлять знаки случайно, тогда:

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

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

- 7. Дисперсия, ковариация и коэффициент корреляции. Их связь и основные свойства: билинейность ковариации, случайная величина с нулевой дисперсией, дисперсия линейного образа случайной величины, дисперсия суммы независимых случайных величин. Неравенство Коши-Буняковского и геометрическая интерпретация ковариации, дисперсии и коэффициент корреляции. Вычисление ожидания и дисперсии у биномиального распределения.**

Определение. Дисперсией случайной величины X называется число $\mathbb{D}X = \mathbb{E}(X - \mathbb{E}X)^2$.

Определение. Ковариацией пары случайных величин X, Y называется число $\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$.

Определение. Коэффициентом корреляции называется величина $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{D}X} \sqrt{\mathbb{D}Y}}$

Теорема (Свойства дисперсии). (i) Если $\mathbb{D}X = 0$, то $X = \mathbb{E}X$ почти наверное

(ii) Для произвольных чисел α, β верно $\mathbb{D}(\alpha X + \beta) = \alpha^2 \mathbb{D}X$

(iii) Если X и Y независимы, то $\text{cov}(X, Y) = 0$ и $\mathbb{D}(X + Y) = \mathbb{D}X + \mathbb{D}Y$

Доказательство. (i) Исходя из свойства (iii) мат. ожидания о неотрицательной случайной величине с нулевым мат. ожиданием, $(X - \mathbb{E}X)^2 = 0 \implies X = \mathbb{E}X$ почти наверное.

(ii) Исходит из линейности математического ожидания

(iii) Так как $\text{cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$, то из независимости X и Y следует $\text{cov}(X, Y) = (\mathbb{E}X)(\mathbb{E}Y) - (\mathbb{E}X)(\mathbb{E}Y) = 0$. Поэтому, $\mathbb{D}(X + Y) = \mathbb{D}X + 2\text{cov}(X, Y) + \mathbb{D}Y = \mathbb{D}X + \mathbb{D}Y$ ■

Следствие. Неравенство Коши-Буняковского.

При условии $\mathbb{E}X^2 < \infty$ и $\mathbb{E}Y^2 < \infty$ имеет место неравенство $|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}$, причём равенство возможно тогда и только тогда, когда $\exists \alpha, \beta$, такие что $\alpha X + \beta Y = 0$ почти наверное.

Доказательство. Если $\mathbb{E}X^2 = 0$, то, исходя из свойства (iii) мат. ожидания, $X^2 = 0 \implies X = 0$ почти наверное, поэтому просто имеем $0 \leq 0$.

Предположим, что $\mathbb{E}X^2 > 0$. Рассмотрим квадратичную функцию $p(t) = \mathbb{E}(Y - tX)^2 = \mathbb{E}Y^2 - 2t\mathbb{E}(XY) + t^2\mathbb{E}X^2$. Из неотрицательности $p(t)$ для любого t , а также того, что $\mathbb{E}X^2 > 0$ следует, что дискриминант этого квадратного многочлена меньше, либо равен 0, то есть:

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

Кроме того, если в неравенстве Коши-Буняковского наблюдается равенство, то $D = 0 \implies \exists t_0 : \mathbb{E}(Y - t_0X)^2 = 0 \implies X + t_0Y = 0$ почти наверное, из чего и следует дополнение к условию теоремы об α и β . ■

Вспомним, что $(X, Y) = \mathbb{E}(XY)$ – неотрицательно определённая билинейная форма. Рассмотрим линейное пространство случайных величин с нулевым мат. ожиданием. На нем введём скалярное произведение $\langle X, Y \rangle = \text{cov}(X, Y)$

Появляется норма $\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{D}X}$. Коэффициент корреляции выступает в роли косинуса угла между векторами X и Y .

Пример. Пусть случайная величина S_n имеет биномиальное распределение, то есть $P(S_n = k) = C_n^k p^k (1-p)^{n-k}$, $k \in \{0, \dots, n\}$. Вычислим ожидание S_n .

Первый способ: по определению $\mathbb{E}S_n = \sum_{k=1}^n k \cdot C_n^k p^k (1-p)^{n-k} = np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np$ (последнее равенство из-за того, что мы по сути суммируем все вероятности в случае, когда у нас в условии задачи $n-1$, а не n , значит сумма равна 1)

Второй способ: Заметим, что ожидание зависит только от распределения случайной величины. Пусть X_1, X_2, \dots, X_n – независимые бернуллиевские случайные величины с вероятностью успеха p . Тогда распределение $X_1 + X_2 + \dots + X_n$ совпадает с биномиальным, значит $\mathbb{E}S_n = \mathbb{E}(X_1 + X_2 + \dots + X_n) = np$.

Вычислим дисперсию: $\mathbb{D}S_n = \mathbb{D}X_1 + \mathbb{D}X_2 + \dots + \mathbb{D}X_n = n((1-p)^2 p + p^2(1-p)) = np(1-p)$. Здесь сразу применили второй способ.

8. Неравенство Чебышёва. Закон больших чисел в слабой форме.

Неравенство Чебышёва. (Фактически, это – неравенство Маркова, но у Косова называется именно так)

Если $X \geq 0$ п.н., то для всякого $t > 0$ верно неравенство $P(\omega : X(\omega) \geq t) \leq \frac{\mathbb{E}X}{t}$.

Доказательство. Пусть $A = \omega : X(\omega) \geq t$. Тогда $X \geq t \cdot I_A$, и значит $\mathbb{E}X \geq tP(A)$. ■

Предложение. (А это уже неравенство Чебышёва) Пусть $\mathbb{E}X^2 < \infty$. Тогда для каждого $\varepsilon > 0$ выполнено $P(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathbb{D}X}{\varepsilon^2}$.

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

Закон больших чисел в слабой форме.

Пусть $\{X_n\}_n$ – последовательность таких независимых одинаково распределённых случайных величин (т.е. таблички, задающие их распределения, совпадают), что $\mathbb{E}X_n^2 < \infty$. Пусть $\mathbb{E}X_1 = m$ (а значит и $\mathbb{E}X_n = m$ при каждом n) тогда для каждого $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| \geq \varepsilon\right) = 0.$$

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

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\mathbb{D}[X_1 + X_2 + \dots + X_n]}{n^2 \varepsilon^2} = \frac{\mathbb{D}X_1}{n \varepsilon^2} \rightarrow 0$$

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

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{pq}{n^2 \varepsilon^2} \rightarrow 0, n \rightarrow \infty.$$

Таким образом, эмпирическая частота успешного результата эксперимента при n проведённых опытах "стремится" к p , т.е. к теоретическому значению вероятности успешного результата эксперимента. ■

9. Теорема Муавра-Лапласа (формулировка локальной и интегральной теорем, доказательство локальной теоремы в симметричном случае, идея доказательства интегральной теоремы).

Теорема 1.1 (Локальная формулировка теоремы Муавра-Лапласа). Если $n \rightarrow \infty$, вероятность $p \in (0, 1)$ фиксирована, величина $x_m := \frac{m - np}{\sqrt{npq}}$ ограничена равномерно по m и n ($a \leq x_m \leq b$), то $P(S_n = m) \sim \frac{1}{\sqrt{npq}} \varphi(x_m)$, где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

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

Доказательство. Приведём доказательство только в случае $p = q = \frac{1}{2}$. Как было сказано выше $P(S_n = m) = C_n^m 2^{-n} = e^{-n \ln 2 + \ln n! - \ln m! - \ln(n-m)!}$. Так как $m = \frac{n}{2} + x_m \frac{\sqrt{n}}{2} = \frac{n}{2} \left(1 + \frac{x_m}{\sqrt{n}}\right)$, $n - m = \frac{n}{2} - x_m \frac{\sqrt{n}}{2} = \frac{n}{2} \left(1 - \frac{x_m}{\sqrt{n}}\right)$, то по формуле Стирлинга $\ln n! = \ln \sqrt{2\pi n} + n \ln n - n + \bar{o}(1)$;

$$\begin{aligned} \ln m! &= \ln \left(\sqrt{\pi n} \left(1 + x_m \frac{1}{\sqrt{n}}\right)^{\frac{1}{2}} \right) + \left(\frac{n}{2} + x_m \frac{\sqrt{n}}{2} \right) \left(\ln n - \ln 2 + \ln \left(1 + \frac{x_m}{\sqrt{n}}\right) \right) - \frac{n}{2} - x_m \frac{\sqrt{n}}{2} + \bar{o}(1) = \\ &= \ln \sqrt{\pi n} + \bar{o}(1) + \left(\frac{n}{2} + x_m \frac{\sqrt{n}}{2} \right) \left(\ln n - \ln 2 + \frac{x_m}{\sqrt{n}} - \frac{x_m^2}{2n} + \bar{o}(1) \right) + \frac{n}{2} - x_m \frac{\sqrt{n}}{2} + \bar{o}(1), \text{ где мы использовали формулу} \\ \text{Тейлора } \ln(1+x) &= x - \frac{x^2}{2} + \bar{o}(x^2), \text{ ограниченность } x_m \text{ и тот факт, что } \ln \left(1 + x_m \frac{1}{\sqrt{n}}\right)^{\frac{1}{2}} = \bar{o}(1). \text{ Аналогично,} \\ \text{заменяя } x_m \text{ на } -x_m, \text{ получаем } \ln(n-m)! &= \\ &= \ln \sqrt{\pi n} + \left(\frac{n}{2} - x_m \frac{\sqrt{n}}{2} \right) \left(\ln n - \ln 2 + \frac{x_m}{\sqrt{n}} - \frac{x_m^2}{2n} + \bar{o}(1) \right) + \frac{n}{2} - x_m \frac{\sqrt{n}}{2} + \bar{o}(1). \end{aligned}$$

Приведём подобные, получаем: $\ln m! + \ln(n-m)! = 2 \ln \sqrt{\pi n} + n \ln n - n \ln 2 + x_m^2 - \frac{x_m^2}{2} - n + \bar{o}(1)$.

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

Полученное нами тождество равносильно эквивалентности из формулировки теоремы в случае $p = q = \frac{1}{2}$. ■

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

Теорема 1.2 (Интегральная формулировка теоремы Муавра-Лапласа). Для любых чисел $a < b$ имеем:

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \int_a^b \varphi(x) dx.$$

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

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