

Комбинаторика и Теория Вероятностей

Конспект курсов “Теория вероятностей – наука о случайности” ([часть I](#), [часть II](#)). и
“Современная комбинаторика” ([ссылка на курс](#))

Содержание

1	Комбинаторика	2
1.1	Основы	2
1.2	Беспорядки	2
1.3	Комбинаторные тождества	3
2	Вероятность	5
2.1	Условная вероятность	5
3	Дискретные СВ	5
3.1	Дискретные распределения	5
3.2	Многомерные дискретные распределения	6
3.3	Условные распределения	6
3.4	Моменты случайной величины. Матожидание	7
3.5	Дисперсия	8
3.6	Задача о беспорядках	9
3.7	Простые распределения	9
3.8	Биномиальное распределение (Binomial distribution)	10
3.9	Геометрическое распределение (Geometric distribution)	10
3.10	Распределение Паскаля (Negative binomial)	11
3.11	Сумма случайного числа НОР случайных величин	12
3.12	Распределение Пуассона (Poisson distribution)	13
4	Непрерывные СВ	13
4.1	Вероятностное пространство	13
4.2	Функции распределения и плотности	14
4.3	Характеристики непрерывных СВ	15
4.4	Экспоненциальное распределение	16
4.5	Процесс Пуассона	16
4.6	Преобразования СВ	17
4.7	Прочие распределения	18
5	Вспомогательные формулы	19



1 Комбинаторика

1.1 Основы

Перестановки (*permutations*): $P_n = n!$

Размещения (*arrangements*) с повторениями: $\overline{A}_n^k = n^k$

Размещения (*arrangements*) без повторений: $A_n^k = (n)_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$

Сочетания (*combinations*) без повторений: $C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{A_n^k}{k!}$

Сочетания (*combinations*) с повторениями: $\overline{C}_n^k = \left(\binom{n+k-1}{k}\right) = C_{n+k-1}^k$

Задача. [[ссылка](#)] Подбрасывают 6 правильных костей. Найти вероятность, что на них выпало ровно 3 разных значения.

Решение. Всего упорядоченных последовательностей 6^6 . Три числа можно выбрать C_6^3 способами. Используя 3 числа, можно составить 3^6 последовательностей, но среди них будут те, где содержится только 2 числа (их количество равно $3 \cdot (2^6 - 2)$, т.е. 3-мя способами выбираем 2 числа из 3-х и из 2^6 вариантов убираем 2 последовательности, содержащие только 1 число), и те, где содержится только 1 число (3 штуки). Искомая вероятность равна $C_6^3(3^6 - 3(2^6 - 1))/6^6 \approx 0.23148$

1.2 Беспорядки

Беспорядок — это перестановка, у которой нет неподвижной точки. Например, письмо не попадает в свой конверт.

Число беспорядков (**субфакториал**) считается по правилу включений-исключений:

$$!n = n! - \frac{n!}{1!} + \frac{n!}{2!} - \frac{n!}{3!} + \dots + (-1)^n \frac{n!}{n!} = \sum_{k=0}^n (-1)^k \frac{n!}{k!}$$

$$!n \approx \frac{n!}{e} \text{ при } n \rightarrow \infty, \text{ поскольку } \sum_{k=0}^{\infty} (-1)^k \frac{1}{k!} = \frac{1}{e}$$

$$!n = \left\lfloor \frac{n! + 1}{e} \right\rfloor$$

$$!1 = 0 \quad !2 = 1 \quad !3 = 2 \quad !4 = 9 \quad !5 = 44 \quad !6 = 265$$

Задача. Если n писем случайным образом положить в n конвертов, то какова *вероятность*, что какое-нибудь из писем попадёт в *свой* конверт?

Решение. $1 - \frac{!n}{n!} \approx 1 - \frac{1}{e} \approx 0.632$

1.3 Комбинаторные тождества

$$C_n^k = C_n^{n-k}$$

$$C_n^0 C_m^k + C_n^1 C_m^{k-1} + \dots + C_n^{k-1} C_m^1 + C_n^k C_m^0 = C_{n+m}^k$$

$$(C_n^0)^2 + (C_n^1)^2 + \dots + (C_n^n)^2 = C_{2n}^n$$

$$(x+y)^n = \sum_{k=0}^n C_n^k x^k y^{n-k}$$

$$C_n^0 + C_n^1 + \dots + C_n^n = \sum_{k=0}^n C_n^k = 2^n$$

$$C_n^0 - C_n^1 + \dots + (-1)^n C_n^n = \sum_{k=0}^n (-1)^k C_n^k = \begin{cases} 1, & \text{при } n = 0 \\ 0, & \text{при } n > 0 \end{cases} \quad (\text{т.к. } (1-1)^n = 0)$$

$$C_n^0 + C_n^2 + C_n^4 + \dots = 2^{n-1}$$

$$C_n^1 + C_n^3 + C_n^5 + \dots = 2^{n-1}$$

$$P(n_1, \dots, n_k) = \frac{(n_1 + \dots + n_k)!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

$$(x_1 + \dots + x_k)^n = \sum_{\substack{0 \leq n_i \leq n \\ n_1 + \dots + n_k = n}} P(n_1, \dots, n_k) x_1^{n_1} \cdot \dots \cdot x_k^{n_k}$$

$$k^n = \sum_{\substack{0 \leq n_i \leq n \\ n_1 + \dots + n_k = n}} P(n_1, \dots, n_k)$$

- $\sqcup A = \{a_1, \dots, a_{n+1}\}$ и V – множество всех m -сочетаний с повторениями из A .
 $\sqcup V_i$ – такое m -сочетание с повторениями из A , которое i раз содержит a_1 .

$$\begin{aligned}
 |V| &= |V_0| + |V_1| + \dots + |V_m| \\
 |V| &= \overline{C}_{n+1}^m = C_{n+m}^m = C_{n+m}^n \\
 |V_0| &= \overline{C}_n^m = C_{n+m-1}^m = C_{m+n-1}^{n-1} \\
 |V_1| &= \overline{C}_n^{m-1} = C_{n+m-2}^{m-1} = C_{m+n-2}^{n-1} \\
 &\dots \\
 |V_m| &= \overline{C}_n^0 = C_{n+0-1}^0 = C_{n-1}^{n-1}
 \end{aligned}$$

Отсюда $C_{m+n}^n = C_{m+n-1}^{n-1} + C_{m+n-2}^{n-1} + \dots + C_{n-1}^{n-1}$

При $n = 2$ получаем $C_{m+2}^2 = \frac{(m+1)(m+1)}{2} = C_{m+1}^1 + C_{m+1}^1 + \dots + C_1^1$
 $\Rightarrow 1 + 2 + \dots + m = \frac{m(m+1)}{2}$.

При $n = 3$ получаем $C_{m+3}^3 = \frac{(m+2)(m+2)(m+1)}{6} = C_{m+2}^2 + C_{m+1}^2 + \dots + C_2^2$
 $\Rightarrow 1^2 + 2^2 + \dots + m^2 = \frac{m(m+1)(2m+1)}{6}$.

$\sqcup A = \{a_1, \dots, a_n\}$ и V – множество всех m -размещений с повторениями из A ,
 Тогда $|V| = N = n^m = (n-0)^m C_n^0$, причём $m < n$.

\sqcup свойство α_i – размещение **не** содержит a_i , и наоборот α'_i – содержит.
 Имеем $N(\alpha_i) = (n-1)^m$, $N(\alpha_i, \alpha_j) = (n-2)^m$, ..., $N(\alpha_1, \dots, \alpha_n) = 0$.
 Поскольку $m < n$, то $N(\alpha'_1, \dots, \alpha'_n) = 0$.

По формуле включений-исключений с учётом $|V_{\alpha_{i_1}, \dots, \alpha_{i_k}}| = C_n^k$:

$$n^m C_n^0 - (n-1)^m C_n^1 + (n-2)^m C_n^2 - \dots + (-1)^{n-1} 1^m C_n^{n-1} + (-1)^n 0^m C_n^n = 0$$

Ещё тождество:

$$\begin{aligned}
 k C_n^k &= n C_{n-1}^{k-1} \\
 1 C_n^1 + 2 C_n^2 + \dots + n C_n^n &= \sum_{k=1}^n k C_n^k = n 2^{n-1}
 \end{aligned}$$

Задача. Сколькими способами можно расселить 10 гостей в 4 различных комнатах так, чтобы ни одна комната не осталась пустой?

Решение. По формуле включений-исключений: $4^{10} - C_4^1 \cdot 3^{10} + C_4^2 \cdot 2^{10} - C_4^3 \cdot 1^{10}$



2 Вероятность

2.1 Условная вероятность

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

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Свойства условной вероятности:

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

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

Формула полной вероятности:

$\sqcup \{H_i\}$ – разбиение $\Omega = \bigsqcup_{i=1}^n H_i$, т.е. $\Omega = \bigcup_{i=1}^n H_i$ и $H_i \cap H_j = \emptyset$ при $i \neq j$. Тогда:

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

Формула Байеса:

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

$$P(H_j|A) = \frac{P(A|H_j)}{\sum_i P(A|H_i)P(H_i)} \quad \{H_i\} \text{ – разбиение } \Omega$$



3 Дискретные СВ

3.1 Дискретные распределения

Случайная величина – это функция (отображение) $X: \Omega \rightarrow \mathbb{R}$

Значение случайной величины – это число $x \in \mathbb{R}$

Дискретное распределение (PMF, probability mass function): $p_X(x) = P(X = x)$

Дискретная СВ: $p(x_1) = p_1, \dots, p(x_n) = p_n$

Условие нормировки: $\sum_i p_i = \sum_i P(X = x_i) = 1$

Функция случайной величины $Y = g(X)$:

$$P(Y = y) = P(g(X) = y) = \sum_{i: g(x_i)=y} P(X = x_i)$$

3.2 Многомерные дискретные распределения

Многомерное дискретное распределение \equiv Multivariate discrete distribution

Совместное распределение X и Y (*joint PMF*): $p_{X,Y}(x, y) = P(X = x, Y = y)$

Нормировка совместного распределения: $\sum_{i,j} p_{ij} = \sum_{i,j} P(X = x_i, Y = y_j) = 1$

Маргинальное распределение: $p_X(x) = \sum_{y \in \Omega_Y} p_{X,Y}(x, y)$ $p_Y(y) = \sum_{x \in \Omega_X} p_{X,Y}(x, y)$

$X \backslash Y$	0	1	2	3	$p_X(\cdot)$
0	0	0	0	1/8	1/8
1	1/8	1/8	1/8	0	3/8
2	2/8	1/8	0	0	3/8
3	1/8	0	0	0	1/8

Marginal distribution of X : $p_X(x)$

X	0	1	2	3
P	1/8	3/8	3/8	1/8

Marginal distribution of Y : $p_Y(y)$

Y	0	1	2	3
P	4/8	2/8	1/8	1/8

X, Y независимы: $\forall x, y : p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$

3.3 Условные распределения

Условное распределение X при условии, что событие A наступило:

$$p_{X|A}(x) = P(X = x|A) = \frac{P(X = x, A)}{P(A)} \quad P(A) \neq 0$$

Условное распределение X при условии, что $Y = y$:

$$A = \{Y = y\} \quad p_Y(y) > 0$$

$$p_{X|Y}(x|y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

$$\forall y \quad \sum_k p_{X|Y}(x_k|y) = 1 \quad (\text{условие нормировки})$$

Формула умножения

$$p_{X,Y}(x, y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x)$$

Формула умножения для трёх СВ:

$$\begin{aligned}
 P(A \cap B \cap C) &= P(A)P(B|A)P(C|A \cap B) \\
 p_{X,Y,Z}(x, y, z) &= P(X = x, Y = y, Z = z) \\
 p_{X,Y,Z}(x, y, z) &= p_X(x)p_{Y|X}(y|x)p_{Z|X,Y}(z|x, y) \\
 p_{Z|X,Y}(z|x, y) &= P(Z = z|X = x, Y = y) = \frac{p_{X,Y,Z}(x, y, z)}{p_{X,Y}(x, y)}
 \end{aligned}$$

X, Y, Z — независимы, если $\forall x, y, z \quad p_{X,Y,Z}(x, y, z) = p_X(x)p_Y(y)p_Z(z)$

3.4 Моменты случайной величины. Матожидание

Матожидание $M(X) = E(X) = \sum_i x_i p_i \quad \exists$ if сходится абсолютно $\sum_i |x_i| p_i < \infty$

Если $Y = g(X)$, то $M(Y) = \sum y_i p_Y(y_i) = \sum g(x_i) p_X(x_i)$

Если $Z = \varphi(X, Y)$, то $M(Z) = \sum_i z_i p_Z(z_i) = \sum_{i,j} \varphi(x_i, y_j) p_{X,Y}(x_i, y_j)$

$$M(X^k) = \sum_i x_i^k p_X(x_i) \quad \text{— } k\text{-й начальный момент}$$

$$M[X - M(X)]^k \quad \text{— } k\text{-й центральный момент}$$

Свойства мат. ожидания:

1. линейность: $M(aX + bY) = a M(X) + b M(Y)$
2. если X и Y независимы, то $M(X \cdot Y) = M(X) \cdot M(Y)$

Условное среднее:

$$M(X|A) = \sum_i x_i p_{X|A}(x_i) \quad \text{или} \quad M(X|Y = y) = \sum_i x_i p_{X|Y}(x_i|y)$$

Формула полной вероятности:

$$M(X) = \sum_i M(X|H_i) P(H_i)$$

Лемма 3.1. Если $X: x_k = k, k = 0, 1, 2, \dots, P(X = k) = p_k$, то $M(X) = \sum_{k=0}^{\infty} P(X > k)$

Док-во.

$$\begin{aligned}
 P(X > k) &= P(X = k+1) + P(X = k+2) + \dots = \sum_{i=k+1}^{\infty} p_i \\
 \sum_{k=0}^{\infty} P(X > k) &= \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} p_i \stackrel{\oplus}{=} \sum_{i=1}^{\infty} \sum_{k=0}^{i-1} p_i = \sum_{i=1}^{\infty} i \cdot p_i = \sum_{i=0}^{\infty} i \cdot p_i = M(X)
 \end{aligned}$$

□

Задача. [ссылка] По каналу связи передаётся либо бесконечная последовательность “0” с вероятностью $\frac{2}{3}$, либо “1” – с вероятностью $\frac{1}{3}$. Каждый символ, независимо от других, воспринимается приёмником с ошибкой (т.е. вместо “1” принимается “0” и наоборот) с вероятностью 0.25. Найти среднее значение номера первой принятой “1”.

Решение. Если передаются “1”, то $P(Rcv_1|Snd_1) = 0.75$ и $M(Pos_1|Snd_1) = \frac{1}{p} = 4/3$.
Если передаются “0”, то $P(Rcv_1|Snd_0) = 0.25$ и $M(Pos_1|Snd_0) = \frac{1}{p} = 4$.
По формуле полной вероятности:
 $M(Pos_1) = M(Pos_1|Snd_1) \cdot P(Snd_1) + M(Pos_1|Snd_0) \cdot P(Snd_0) = \frac{4}{3} \cdot \frac{1}{3} + 4 \cdot \frac{2}{3} = \frac{28}{9}$.

Задача. $\perp X, Y \sim Geom(p)$ – независимые СВ. Найти $M(X|X+Y=c)$.

Решение.

$$P(X+Y=c) = P_{pascal}(m=2, p, k=c) = C_{c-1}^{m-1} p^m q^{c-m} = (c-1)p^2 q^{c-2}$$

$$P(X|X+Y=c) = \frac{P(X=k)P(Y=c-k)}{P(X+Y=c)} = \frac{pq^{k-1}pq^{c-k-1}}{(c-1)p^2q^{c-2}} = \frac{1}{c-1}$$

То есть: $Z = \{X|X+Y=c\}$ – СВ, равномерно распределённая на $k = \overline{1, c-1}$.
Отсюда: $M(X|X+Y=c) = M(Z) = \frac{1+2+\dots+(c-1)}{c-1} = \frac{c}{2}$.

3.5 Дисперсия

$$D(X) = \text{var}(X) = M(X - MX)^2 = M(X^2) - (MX)^2 \geq 0$$

$$D(c) = 0$$

$$D(cX) = c^2 D(X)$$

Среднее квадратическое отклонение (standard deviation): $\sigma(X) = \sqrt{D(X)}$

X и Y независимы $\implies D(X+Y) = D(X) + D(Y)$

В общем случае $D(X+Y) = D(X) + D(Y) + 2\text{cov}(X, Y)$

Ковариация $\text{cov}(X, Y) = M[(X - MX)(Y - MY)] = M(XY) - M(X)M(Y)$

X и Y независимы $\implies \text{cov}(X, Y) = 0$ (но не наоборот)

Коэффициент корреляции (correlation) $r = \rho = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$

Свойства корреляции:

- $-1 \leq \rho \leq 1$
- X, Y – независимы $\implies \rho = 0$
- $Y = aX + c, a > 0 \implies \rho = 1$
- $Y = aX + c, a < 0 \implies \rho = -1$

Размерность: $X(\text{кг}) \ D_X(\text{кг}^2) \ \sigma_X(\text{кг}) \ Y(\text{см}) \ \text{cov}_{X,Y}(\text{кг} \cdot \text{см}) \ \rho_{X,Y}(\text{“1”})$

3.6 Задача о беспорядках

— n писем случайно разбрасываются по n конвертам.

Число комбинаций, когда ни одно письмо не попадёт в свой конверт $!n$.

Вероятность, что ни одно письмо не попадёт в свой конверт $\frac{!n}{n!}$

— X - число писем, попавших в свой конверт. $X = \sum_k I_{A_k}$.

I_{A_k} - индикаторная СВ, A_k - событие “письмо k попало в свой конверт”.

$$\begin{aligned} M(I_{A_k}) &= P(A_k) = \frac{1}{n} & M(I_{A_k}^2) &= \frac{1}{n} \\ M(I_{A_k} \cdot I_{A_m}) &= P(A_k \cap A_m) = \frac{1}{n(n-1)} \end{aligned}$$

Тогда:

$$\begin{aligned} M(X) &= n M(I_{A_k}) = 1 \\ D(X) &= \sum_k D(I_{A_k}) + 2 \sum_{k>m} \text{cov}(I_{A_k}, I_{A_m}) = \\ &= \sum_k (M(I_{A_k}^2) - M^2(I_{A_k})) + 2 \sum_{k>m} (M(I_{A_k} I_{A_m}) - M(I_{A_k}) M(I_{A_m})) = \\ &= n \left(\frac{1}{n} - \frac{1}{n^2} \right) + 2 \frac{n(n-1)}{2} \left(\frac{1}{n(n-1)} - \frac{1}{n^2} \right) = 1 \end{aligned}$$

3.7 Простые распределения

Индикаторная СВ:

$$\begin{aligned} I_A(\omega) &= \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \\ p(I_A = 1) &= p_A \\ M(I_A) &= p_A & M(I_A^2) &= p_A \\ D(I_A) &= p_A - p_A^2 \\ M(I_A \cdot I_B) &= M(I_{A \cap B}) = P(A \cap B) \end{aligned}$$

Распределение Бернулли (Bernoulli distribution):

$$\begin{aligned} P(X = 1) &= p & P(X = 0) &= q & p \in [0, 1] & q = 1 - p \\ M(X) &= p \\ D(X) &= pq \end{aligned}$$

Равномерное дискретное распределение (Discrete uniform distribution):

$$P(X = x_k) = \frac{1}{N} \quad k = \overline{1, N}$$

$$M(X) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$D(X) = \frac{n^2 - 1}{12}$$

Гипергеометрическое распределение (Hypergeometric distribution):

$$P(X = r) = \frac{C_M^r C_N^{n-r}}{C_{M+N}^n} \quad r = 0, \dots, \min(M, n)$$

Урна: M белых + N чёрных шаров. Вынимаем n шаров.

R – число вынутых белых шаров.

3.8 Биномиальное распределение (Binomial distribution)

Схема испытаний Бернулли: A_1, \dots, A_n $P(A_i) = p$ $p \in [0, 1]$ $q = 1 - p$

Биномиальное распределение:

$$X \sim Binomial(n, p) \quad P(X = k) = C_n^k p^k q^{n-k} \quad k = \overline{0, n} \quad p \in [0, 1] \quad q = 1 - p$$

k – число успехов в схеме из n испытаний Бернулли.

$$\text{Условие нормировки: } \sum_{k=0}^n P(k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = 1$$

С ростом числа опытов график уплотняется вокруг точки np .

Симметрично при $p = 0.5$, скошено влево при $p < 0.5$.

Можно представить как сумму НОР СВ Бернулли:

$$X_{binomial} = X_1 + \dots + X_n$$

Следовательно, среднее $M = np$ и дисперсия $D = npq$

Из чего можно получить красивое равенство:

$$\sum_{k=0}^n k C_n^k p^k q^{n-k} = np$$

3.9 Геометрическое распределение (Geometric distribution)

$$X \sim Geom(p) \quad P(X = k) = pq^{k-1} \quad k \in \mathbb{N} \quad p \in [0, 1] \quad q = 1 - p$$

k – номер первого успеха в схеме Бернулли, т. е. до него в $k - 1$ опыте были неудачи (схема Бернулли – ряд *независимых* опытов)

Свойство: $P(X > n) = q^n$, т. к. $X > n \iff n$ опытов окончились неудачей:

$$P(X = \overline{n+1, \infty}) = \sum_{k=n}^{\infty} pq^k = pq^n \sum_{k=0}^{\infty} q^k = pq^n \frac{1}{1-q} = pq^n \frac{1}{p}$$

Успех наступил в 1-м опыте: $P(X = 1) = p$

Успех в 1-м опыте не наступил: $P(X > 1) = q$

Отсутствие памяти (memoryless):

$$P(X = m + n | X > m) = \frac{P(X = m + n)}{P(X > m)} = \frac{pq^{m+n-1}}{q^m} = pq^{n-1} = P(X = n)$$

Геометрическое — единственное среди дискретных распределение с таким свойством.

Следствия из свойства отсутствия памяти:

$$P(X = 1 + n | X > 1) = P(X = 1 + n) \implies$$

$$M(X | X > 1) = M(X + 1) = M(X) + 1$$

$$M(X^2 | X > 1) = M((X + 1)^2)$$

Мат. ожидание $M_{geom}(X) = \frac{1}{p}$

$$\text{число.опытов.до.первого.успеха} \times \text{вероятность.успеха} = 1$$

Док-во.

$$\begin{aligned} M(X) &= M(X | X = 1)P(X = 1) + M(X | X > 1)P(X > 1) \\ &= 1 \cdot p + (1 + M(x))q = 1 + qM(X) \end{aligned}$$

□

Отсюда $\sum_{k=0}^{\infty} kq^k = \frac{1}{p^2}$

$$D(X) = \frac{q}{p^2}$$

3.10 Распределение Паскаля (Negative binomial)

$$Z \sim Pascal(m, p) \quad P(Z = k) = C_{k-1}^{m-1} p^m q^{k-m}$$

$$p \in [0, 1] \quad q = 1 - p \quad k \geq m \quad k, m \in \mathbb{N}$$

k — номер опыта, на котором произошёл m -й успех в схеме Бернулли

Выводится так:

$$\begin{aligned} P_{pascal}(Z = \kappa; m = m) &= P_{binomial}(X = m - 1; n = \kappa - 1) \cdot p \\ &= p \cdot C_{\kappa-1}^{m-1} p^{m-1} q^{(\kappa-1)-(m-1)} = C_{\kappa-1}^{m-1} p^m q^{\kappa-m} \end{aligned}$$

$Z \sim \text{Pascal}(m, p)$ можно представить (в силу отсутствия памяти у Geom) как сумму m независимых одинаково распределённых (НОР) СВ: $Z = X_1 + \dots + X_m$, где $X_i \sim \text{Geom}(p)$. Отсюда получаем мат. ожидание и дисперсию:

$$M(Z) = \frac{m}{p} \quad D(Z) = \frac{mq}{p^2}$$

Замечание: Часто под распределением “Negative binomial” подразумевается число **неудач** до наступления m -го успеха: $M_{\text{неудач}} = M(Z) - m$.

3.11 Сумма случайного числа НОР случайных величин

Лемма 3.2. $\sqcup Z = \sum_{k=1}^N X_k$,

причём $M(X_k) = M_X$, $D(X_k) = D_X$, $M(N) = M_N$, $D(N) = D_N$.

Тогда $M(Z) = M_X M_N$, $D(Z) = D_X M_N + M_X^2 D_N$.

Док-во.

$\sqcup P_N(N = n) = P_n$, $M(X^2) = M_{X^2}$, $M(Z) = M_Z$.

$$\begin{aligned} M(Z) &= \sum_i M(Z|H_i)P(H_i) = \sum_n M(Z|N = n)P(N = n) = \\ &= \sum_n \sum_{k=1}^n M(X_k)P_n = \sum_n P_n n M_X = M_N M_X \end{aligned}$$

$$\begin{aligned} D(Z) &= M(Z^2) - M^2(Z) = \sum_n M(Z^2|N = n)P(N = n) - M_Z^2 = \\ &= \sum_n M\left(\sum_{k=1}^n X_k\right)^2 P_n - M_Z^2 = \sum_n P_n \left(\sum_{i=1}^n X_i\right) \left(\sum_{j=1}^n X_j\right) - M_Z^2 = \\ &= \sum_n P_n \sum_{i=1}^n X_i^2 + \sum_n P_n 2 \sum_{i < j}^n X_i X_j - M_Z^2 = \\ &= \sum_n P_n n M(X^2) + \sum_n P_n 2 \frac{n(n-1)}{2} M(X) M(X) - M_Z^2 = \\ &= M_{X^2} M_N + \sum_n P_n n^2 M_X^2 - \sum_n P_n n M_X^2 - M_N^2 M_X^2 = \\ &= M_N M_{X^2} - M_N M_X^2 + M(N^2) M_X^2 - M_N^2 M_X^2 = M_N D_X + D_N M_X^2 \end{aligned}$$

□

3.12 Распределение Пуассона (Poisson distribution)

Является приближением последовательности биномиальных распределений:

$$S_n \sim \text{Binomial}(n, p_n) \text{ при } n \rightarrow \infty, p_n \rightarrow 0, np_n \rightarrow \lambda \implies P(S_n = k) \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}$$

$$X \sim \text{Poisson}(\lambda) : P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \lambda > 0 \quad k = 0, 1, 2, \dots$$

$$M(X) = D(X) = \frac{1}{\lambda}$$

Симметрично при $\lambda \approx 10$, скошено влево при $\lambda < 10$.



4 Непрерывные СВ

4.1 Вероятностное пространство

Опр-е. 4.1 (σ -алгебра). σ -алгебра над множеством X — это семейство \mathfrak{S} подмножеств множества X , т.ч.:

1. $X \in \mathfrak{S}$ и $\emptyset \in \mathfrak{S}$

2. если $E \in \mathfrak{S}$, то $X \setminus E \in \mathfrak{S}$

3. если \exists семейство $\{A_n\} \in \mathfrak{S}$ (конечное или счётное), то $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{S}$ и $\bigcap_{n=1}^{\infty} A_n \in \mathfrak{S}$

Опр-е. 4.2 (Борелевское множество). \mathbb{B} -множество — такое множество, которое может быть получено из *открытых или замкнутых промежутков* на \mathbb{R} конечным или счетным числом операций $\bigcup A_n$ и $\bigcap A_n$.

Опр-е. 4.3. Борелевская σ -алгебра \equiv минимальная σ -алгебра борелевских множеств на \mathbb{R} .

Задача. Является ли множество всех рациональных точек на прямой борелевским, т.е. верно ли, что $\left\{ \frac{m}{n}, n = 1, 2, \dots; m = 0, 1, 2, \dots \right\} \in \mathbb{B}$?

Решение. Да.

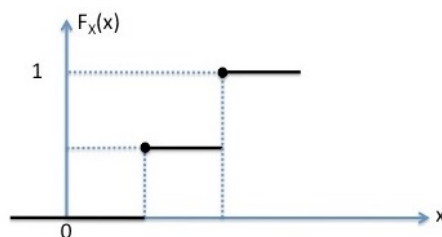
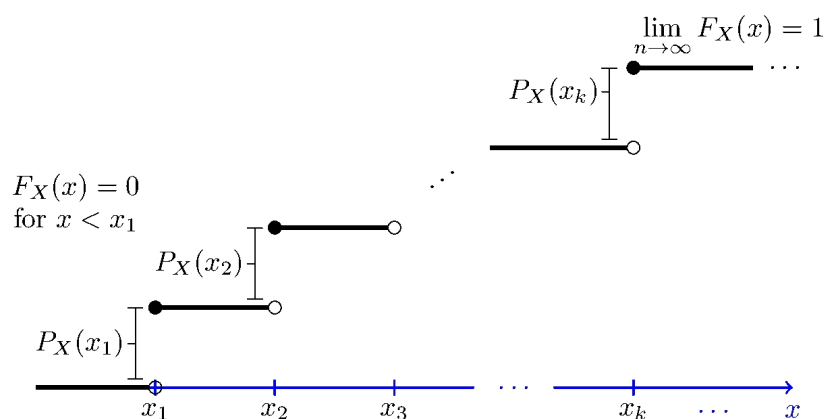
Каждую точку можно рассматривать как отрезок $\left[\frac{m}{n}, \frac{m}{n} \right] = \left(-\infty, \frac{m}{n} \right] \setminus \left(-\infty, \frac{m}{n} \right)$, причём множество всех рациональных точек на прямой счётно.

4.2 Функции распределения и плотности

Функция распределения вероятности (CDF)

1. $F_X(x) = P(X \leq x)$
2. $0 \leq F_X \leq 1$
3. F — монотонно неубывающая
4. $F(-\infty) = 0, F(+\infty) = 1$
5. $P(a < X \leq b) = F(b) - F(a)$
6. F — непрерывна *справа*
7. F — это некоторая вероятность, т.е. безразмерная величина

Замечание: если бы CDF была определена как $P(X < x)$, то была бы непрерывна *слева*.



Функция плотности распределения (PDF):

1. $f(u) \geq 0$
2. $f(x) = F'(x) = \frac{dF(x)}{dx}$
3. $P(X \leq x) = F(x) = \int_{-\infty}^x f(u) du$
4. $\int_{-\infty}^{+\infty} f(u) du = 1$ — условие нормировки
5. $P(X = a) = 0$

6. $P(X \in \langle x, x + \delta \rangle) \approx f(x)\delta$

7. размерность f есть $\frac{1}{\text{размерность}(X)}$ (например, см^{-1} , кг^{-1})

4.3 Характеристики непрерывных СВ

Матожидание $M(x) = \int_{-\infty}^{+\infty} u f(u) du$

$M(x)$ существует \iff интеграл сходится *абсолютно*: $\int_{-\infty}^{+\infty} |u| f(u) du < \infty$.

Матожидание может не существовать. Пример – распределение Коши: $f(x) = \frac{1/\pi}{1+x^2}$.

Для него $\int_{-\infty}^{+\infty} |u| f(u) du = \frac{2}{\pi} \int_0^{\infty} \frac{x dx}{1+x^2} = \frac{1}{\pi} \ln(1+x^2) \Big|_0^{\infty} \rightarrow \infty$

Если $f(x) = f(-x)$ (чётная), то $M(X) = 0$ (если существует).

Матожидание функции $M(\varphi(x)) = \int_{-\infty}^{+\infty} \varphi(u) f(u) du$.

Дисперсия $D(X) = M(X - M(X)) = M(X^2) - M^2(X)$

Медиана (характеристика *положения*): $\text{Me}(X) = \arg \{F_X(x) = 1/2\}$

Медиана существует *всегда*.

Если $f(x) = f(-x)$ (чётная), то для нормировки функции плотности имеем:

$$1 = \int_{-\infty}^{\infty} f(u) du = \int_{-\infty}^0 f(u) du + \int_0^{\infty} f(u) du = F(0) + \int_0^{\infty} f(-u) d(-u) = F(0) + F(0)$$

а значит $F(0) = 1/2$, т.е. $\text{Me}(X) = 0$.

Нижний квартиль: $F(Q_1) = 1/4$

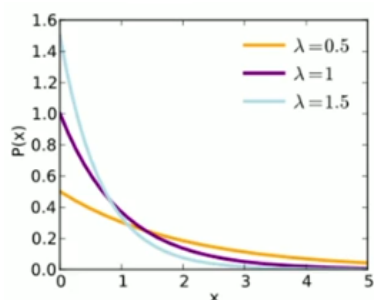
Нижний квартиль: $F(Q_3) = 3/4$

Межквартильный размах: $IQR = Q_3 - Q_1$

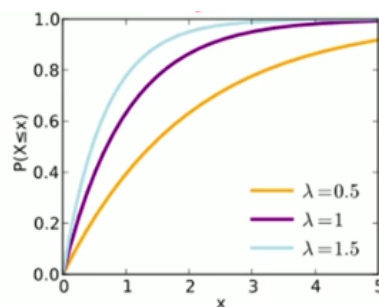
Квантиль уровня α : $F(q_\alpha) = \alpha$

Квантильная функция $Q(p) = F^{-1}(x)$, $p \in [0, 1]$ — функция, обратная к функции распределения

4.4 Экспоненциальное распределение



Probability density function



Cumulative distribution function

$$X \sim \text{Exp}(\lambda)$$

$$\text{PDF: } f(u) = \lambda e^{-\lambda u}, \quad u \geq 0, \lambda > 0$$

$$\text{CDF: } F(X) = P(X \leq x) = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}$$

$$f(u) = \lambda e^{-\lambda u} \quad u \geq 0, \lambda > 0$$

$$F(x) = 1 - e^{-\lambda x} \quad x \geq 0$$

$$M(x) = \frac{1}{\lambda}$$

$$D(x) = \frac{1}{\lambda^2}$$

$$Me(x) = \frac{\ln 2}{\lambda} \quad \text{скошено вправо: } Me < M$$

Свойство отсутствия памяти:

$$P(X > t+s | X > t) = \frac{P(X > t+s, X > t)}{P(X > t)} = \frac{P(X > t+s)}{P(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = P(X > s) \quad (s, t \geq 0)$$

Среди дискретных распределений свойством отсутствия памяти обладает геометрическое. Экспоненциальное распределение – непрерывный аналог геометрического.

4.5 Процесс Пуассона

Случайный процесс – последовательность случайных величин в дискретном времени (обычно НОР, т.е. независимых одинаково распределённых).

Пуассоновский процесс – непрерывный аналог процесса Бернулли, когда интервалы между успехами распределены экспоненциально $P(T > \Delta t) = e^{-\lambda \Delta t} \approx 1 - \lambda \Delta t$:

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

$$M = D = \lambda t$$

Свойства ПП:

1. стационарный (распределение числа событий зависит только от длины интервала)
2. ординарный (вряд ли на очень малом интервале произойдёт больше одного события)
3. поток с отсутствием памяти.

Задача. Будет ли пуассоновский процесс подходящей моделью процесса прибытия пассажиров в пункт выдачи багажа в аэропорту?

Решение. **Нет.**

Не выполняется требование независимости событий, т.к. пассажиры приходят за багажом после посадки самолета и, если мы наблюдаем, к примеру, что в зону получения багажа за последнюю минуту пришли 15 человек, то следует ожидать значительное количество людей в течение следующей минуты. Как получают багаж? Сначала толпа, потом редкие одиночки.

Задача. Будет ли пуассоновский процесс подходящей моделью процесса поступления звонков в регистратуру поликлиники в течение рабочего дня?

Решение. **Нет.**

Не выполняется требование стационарности пуассоновского процесса, т.к. частота звонков намного выше в утренние часы в сравнении с вечерними. К концу дня поток звонков при любом расписании будет снижаться (если это не ночной стационар).

4.6 Преобразования СВ

Лемма 4.1 (Производный поток Бернулли).

$\sqcup X \sim \text{Bernoulli}(p_x), Y \sim \text{Bernoulli}(p_y)$ и $Z = \text{OR}(X, Y)$.
 $\implies Z \sim \text{Bernoulli}(p_z)$ и $p_z = p_x + p_y - p_x p_y$.

Лемма 4.2 (Линейное преобразование).

$\sqcup Y = aX + B, a \neq 0$. Тогда

$$F_Y(t) = \begin{cases} F_X\left(\frac{t-b}{a}\right), & a > 0 \\ 1 - F_X\left(\frac{t-b}{a}\right), & a < 0 \end{cases}$$

$$f_Y(u) = \frac{1}{|a|} f_X\left(\frac{u-b}{a}\right)$$

Лемма 4.3 (Монотонное преобразование).

$\sqcup Y = g(X)$, причём $g \nearrow$. Тогда

$$F_Y(t) = F_X(g^{-1}(t))$$

$$f_Y(u) = \frac{f_X(g^{-1}(u))}{g'(u)}$$

Лемма 4.4 (Монотонно убывающее преобразование).

$\sqcup Y = g(X)$, причём $g \searrow$. Тогда

$$F_Y(t) = 1 - F_X(g^{-1}(t))$$

$$f_Y(u) = -\frac{f_X(g^{-1}(u))}{g'(u)}$$

Лемма 4.5 (Получение заданной СВ из равномерной).

$\sqcup X \sim Uniform(0, 1)$ и $Y = F_*^{-1}(X)$, причём $F_*^{-1}(t) \nearrow: [0, 1] \rightarrow \mathbb{R}$.

Тогда $F_Y(t) = F_*(t)$

4.7 Прочие распределения

Равномерное распределение $X \sim Uniform(a, b)$:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$f(x) = \begin{cases} 0 & x < a \\ 1/(b-a) & a \leq x \leq b \\ 0 & x > b \end{cases}$$

$$Me = M = (a + b)/2$$

$$D = (b - a)^2/12$$

Распределение Лапласа:

$$f = \frac{\lambda}{2} e^{-\lambda|x|}$$

$$M = 0$$

$$D = \frac{2}{\lambda^2}$$

Распределение Коши:

$$f(x) = \frac{1/\pi}{1 + x^2}$$



5 Вспомогательные формулы

$$\begin{aligned}\int_0^\infty e^{-ax} dx &= \frac{1}{a} \\ \int_0^\infty x e^{-ax} dx &= \frac{1}{a^2} \\ \int_0^\infty x^2 e^{-ax} dx &= \frac{2}{a^3} \\ \int_0^\infty x^n e^{-ax} dx &= \frac{n!}{a^{n+1}}\end{aligned}$$

$$\begin{aligned}a^3 - b^3 &= (a - b)(a^2 + ab + b^2) \\ (a - b)^3 &= a^3 - 3a^2b + 3ab^2 - b^3\end{aligned}$$

$$\begin{aligned}\sum_{k=0}^n k C_n^k p^k q^{n-k} &= np \\ \sum_{k=0}^\infty k q^k &= \frac{1}{p^2} \\ \sum_{i=0}^\infty \sum_{j=i+1}^\infty a_{ij} &= \sum_{j=1}^\infty \sum_{i=0}^{j-1} a_{ij} \\ \sum_{n=1}^\infty \frac{n}{2^n} &= 2\end{aligned}$$

