Комбинаторика и Теория Вероятностей

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

Содержание

1 Комбинаторика

1.1 Основы

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$(x_1 + ... + x_k)^n = \sum_{\substack{0 \leqslant n_i \leqslant n \\ n_1 + ... + n_k = n}} P(n_1, ..., n_k) x_1^{n1} \cdots x_k^{n_k}$$

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

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

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

Отсюда $C^n_{m+n} = C^{n-1}_{m+n-1} + C^{n-1}_{m+n-2} + ... + C^{n-1}_{n-1}$

При $n = 2$ получаем $C^2_{m+2} = \frac{(m+1)(m+1)}{2} = C^1_{m+1} + C^1_{m+1} + ... + C^1_1$
$\Longrightarrow\ 1 + 2 + ... + m = \dfrac{m(m+1)}{2}$.
При $n = 3$ получаем $C^3_{m+3} = \frac{(m+3)(m+2)(m+1)}{6} = C^2_{m+2} + C^2_{m+1} + ... + C^2_2$
$\Longrightarrow\ 1^2 + 2^2 + ... + m^2 = \dfrac{m(m+1)(2m+1)}{6}$.

$\lrcorner$ $A = \{a_1, ..., a_n\}$ и V – множество всех m-размещений с повторениями из A,
Тогда $|V| = N = n^m = (n-0)^m C^0_n$, причём $\boldsymbol{m < n}$.
$\lrcorner$ свойство α_i – размещение **не** содержит a_i, и наоборот α'_i – содержит.
Имеем $N(\alpha_i) = (n-1)^m$, $N(\alpha_i, \alpha_j) = (n-2)^m$, ..., $N(\alpha_1, ..., \alpha_n) = 0$.
Поскольку $m < n$, то $N(\alpha'_1, ..., \alpha'_n) = 0$.
По формуле включений-исключений с учётом $|V_{\alpha_{i_1},...,\alpha_{i_k}}| = C^k_n$:

$$n^m C^0_n - (n-1)^m C^1_n + (n-2)^m C^2_n + ... + (-1)^{n-1} 1^m C^{n-1}_n + (-1)^n 0^m C^n_n = 0$$

Ещё тождество:

$$
\begin{aligned}
&k\, C^k_n = n\, C^{k-1}_{n-1} \\
&1\, C^1_n + 2\, C^2_n + ... + n\, C^n_n = \sum_{k=1}^{n} k\, C^k_n = n\, 2^{n-1}
\end{aligned}
$$

Задача. Сколькими способами можно расселить 10 гостей в 4 различных комнатах так, чтобы ни одна комната не осталась пустой?
Решение. По формуле включений-исключений: $4^{10} - C^1_4 \cdot 3^{10} + C^2_4 \cdot 2^{10} - C^3_4 \cdot 1^{10}$

2 Вероятность

2.1 Условная вероятность

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$P(A_1 \cap ... \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot ... \cdot P(A_n|A_1 \cap ... \cap A_{n-1})$$

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:

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

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

Формула Байеса:

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

---

3 Дискретные СВ

3.1 Дискретные распределения

Случайная величина – это функция (отображение) $X : \Omega \to \mathbb{R}$
Значение случайной величины – это число $x \in \mathbb{R}$

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

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

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

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 ≡ Multivariate discrete distribution

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

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

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

X \ 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:

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

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

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

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

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

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:

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

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 $\mathrm{M}(X) = E(X) = \sum\limits_i x_i p_i \quad \exists$ if сходится абсолютно $\sum\limits_i |x_i| p_i < \infty$

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

Если $Z = \varphi(X, Y)$, то $\mathrm{M}(Z) = \sum\limits_i z_i \, p_Z(z_i) = \sum\limits_{i,j} \varphi(x_i, y_j) \, p_{X,Y}(x_i, y_j)$

$\mathrm{M}(X^k) = \sum\limits_i x_i^k p_X(x_i)$ — k-й начальный момент

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

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

Условное среднее:

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

Формула полной вероятности:

$$\mathrm{M}(X) = \sum_i \mathrm{M}(X|H_i) \, P(H_i)$$

Лемма 3.1. Если $X\colon x_k = k,\ k = 0, 1, 2, ...,\ P(X = k) = p_k$, то $\mathrm{M}(X) = \sum\limits_{k=0}^{\infty} P(X > k)$

Док-во.

$$P(X > k) = P(X = k + 1) + P(X = k + 2) + ... = \sum_{i=k+1}^{\infty} p_i$$

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} p_i \overset{\text{☺}}{=} \sum_{i=1}^{\infty} \sum_{k=0}^{i-1} p_i = \sum_{i=1}^{\infty} i \cdot p_i = \sum_{i=0}^{\infty} i \cdot p_i = \mathrm{M}(X)$$

□

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

Задача. $\lrcorner$ $X, Y \sim Geom(p)$ – независимые СВ. Найти $\mathrm{M}(X|X+Y=c)$.
Решение.

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

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

3.5 Дисперсия

$$\mathrm{D}(X) = \mathrm{var}(X) = \mathrm{M}(X - \mathrm{M}\,X)^2 = \mathrm{M}(X^2) - (\mathrm{M}\,X)^2 \geqslant 0$$
$$\mathrm{D}(c) = 0$$
$$\mathrm{D}(cX) = c^2\,\mathrm{D}(X)$$

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

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

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

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

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

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

Свойства корреляции:

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

Размерность: X(кг) D_X(кг^2) σ_X(кг) Y(см) $\mathrm{cov}_{X,Y}$(кг·см) $\rho_{X,Y}$("1")

3.6 Задача о беспорядках

$\lrcorner$ 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 $\dfrac{!n}{n!}$

$\lrcorner$ X - число писем, попавших в свой конверт. $X = \sum\limits_k I_{A_k}$.
I_{A_k} – индикаторная СВ, A_k – событие “письмо k попало в свой конверт”.

$$\mathrm{M}(I_{A_k}) = P(A_k) = \frac{1}{n} \qquad \mathrm{M}(I_{A_k}^2) = \frac{1}{n}$$
$$\mathrm{M}(I_{A_k} \cdot I_{A_m}) = P(A_k \cap A_m) = \frac{1}{n(n-1)}$$

Тогда:

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

3.7 Простые распределения

Индикаторная СВ:

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

Распределение Бернулли (Bernoulli distribution):

$$P(X=1) = p \quad P(X=0) = q \quad p \in [0,1] \quad q = 1-p$$
$$\mathrm{M}(X) = p$$
$$\mathrm{D}(X) = pq$$

Равномерное дискретное распределение (Discrete uniform distribution):

$$P(X = x_k) = \frac{1}{N} \quad k = \overline{1, N}$$
$$\mathrm{M}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\mathrm{D}(X) = \frac{n^2 - 1}{12}$$

Гипергеометрическое распределение (Hypergeometric distribution):

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

Урна: M белых + N чёрных шаров. Вынимаем n шаров.
R – число вынутых белых шаров.

3.8 Биномиальное распределение (Binomial distribution)

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

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.

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

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:

$$X_{binomial} = X_1 + ... + X_n$$

Следовательно, среднее $\mathrm{M} = np$ и дисперсия $\mathrm{D} = npq$

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):

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

Геометрическое — единственное среди дискретных распределение с таким свойством.

Следствия из свойства отсутствия памяти:

$$\begin{aligned} &P(X = 1 + n|X > 1) = P(X = 1 + n) \qquad \implies \\ &\mathrm{M}(X|X > 1) = \mathrm{M}(X + 1) = \mathrm{M}(X) + 1 \\ &\mathrm{M}(X^2|X > 1) = \mathrm{M}\left((X+1)^2\right) \end{aligned}$$

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

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

Док-во.

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

□

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

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

3.10 Распределение Паскаля (Negative binomial)

$$\begin{aligned} &Z \sim Pascal(m, p) \qquad P(Z = k) = C_{k-1}^{m-1} p^m q^{k-m} \\ &p \in [0, 1] \quad q = 1 - p \qquad k \geqslant m \quad k, m \in \mathbb{N} \end{aligned}$$

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

Выводится так:

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

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

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

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

3.11 Сумма случайного числа НОР случайных величин

Лемма 3.2. $\lrcorner$ $Z = \sum\limits_{k=1}^{N} X_k$,
причём $\mathrm{M}(X_k) = \mathrm{M}_X$, $\mathrm{D}(X_k) = D_X$, $\mathrm{M}(N) = \mathrm{M}_N$, $\mathrm{D}(N) = \mathrm{D}_N$.
Тогда $\mathrm{M}(Z) = \mathrm{M}_X \mathrm{M}_N$, $\mathrm{D}(Z) = \mathrm{D}_X \mathrm{M}_N + \mathrm{M}_X^2 \mathrm{D}_N$.
Док-во.
$\lrcorner$ $P_N(N = n) = P_n$, $\mathrm{M}(X^2) = \mathrm{M}_{X^2}$, $\mathrm{M}(Z) = \mathrm{M}_Z$.

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

$$\begin{aligned}
\mathrm{D}(Z) = \mathrm{M}(Z^2) - \mathrm{M}^2(Z) &= \sum_n \mathrm{M}(Z^2|N=n)P(N=n) - \mathrm{M}_Z^2 = \\
&= \sum_n \mathrm{M}\left(\sum_{k=1}^{n} X_k\right)^2 P_n - \mathrm{M}_Z^2 = \sum_n P_n \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{n} X_j\right) - \mathrm{M}_Z^2 = \\
&= \sum_n P_n \sum_{i=1}^{n} X_i^2 + \sum_n P_n 2 \sum_{i<j}^{n} X_i X_j - \mathrm{M}_Z^2 = \\
&= \sum_n P_n n \,\mathrm{M}(X^2) + \sum_n P_n 2\frac{n(n-1)}{2} \mathrm{M}(X)\,\mathrm{M}(X) - \mathrm{M}_Z^2 = \\
&= \mathrm{M}_{X^2} \mathrm{M}_N + \sum_n P_n n^2 \mathrm{M}_X^2 - \sum_n P_n n \,\mathrm{M}_X^2 - \mathrm{M}_N^2 \mathrm{M}_X^2 = \\
&= \mathrm{M}_N \mathrm{M}_{X^2} - \mathrm{M}_N \mathrm{M}_X^2 + \mathrm{M}(N^2)\mathrm{M}_X^2 - \mathrm{M}_N^2 \mathrm{M}_X^2 = \mathrm{M}_N \mathrm{D}_X + \mathrm{D}_N \mathrm{M}_X^2
\end{aligned}$$

$\square$

3.12 Распределение Пуассона (Poisson distribution)

Является приближением последовательности биномиальных распределений:

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

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

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

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 $\varnothing \in \mathfrak{S}$
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 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathfrak{S}$ и $\bigcap\limits_{n=1}^{\infty} A_n \in \mathfrak{S}$

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 $\left\{\frac{m}{n}, n = 1, 2, ...; m = 0, 1, 2, ...\right\} \in \mathbb{B}$?

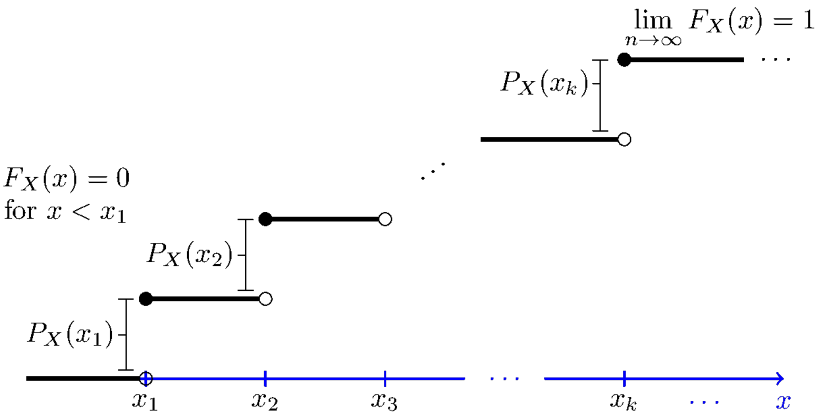
Решение. **Да.**
Каждую точку можно рассматривать как отрезок $\left[\frac{m}{n}, \frac{m}{n}\right] = \left(-\infty, \frac{m}{n}\right] \setminus \left(-\infty, \frac{m}{n}\right)$, причём множество всех рациональных точек на прямой счетно.

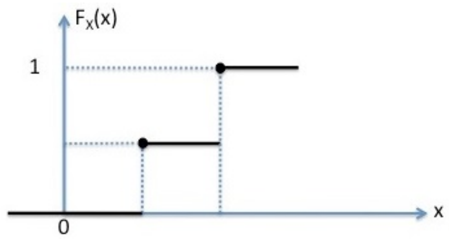
4.2 Функции распределения и плотности

Функция распределения вероятности (CDF)

1. $F_X(x) = P(X \leqslant x)$
2. $0 \leqslant F_X \leqslant 1$
3. F — монотонно неубывающая
4. $F(-\infty) = 0$, $F(+\infty) = 1$
5. $P(a < X \leqslant b) = F(b) - F(a)$
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. $f(u) \geqslant 0$
2. $f(x) = F'(x) = \dfrac{dF(x)}{dx}$
3. $P(X \leqslant x) = F(x) = \displaystyle\int_{-\infty}^{x} f(u)du$
4. $\displaystyle\int_{-\infty}^{+\infty} f(u)du = 1$ — условие нормировки
5. $P(X = a) = 0$

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

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

4.3 Характеристики непрерывных СВ

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

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

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

Для него $\displaystyle\int_{-\infty}^{+\infty} |u|f(u)du = \frac{2}{\pi}\int_0^{\infty} \frac{x\,\mathrm{d}x}{1+x^2} = \frac{1}{\pi}\ln(1+x^2)\Big|_0^{\infty} \to \infty$

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

Матожидание функции $\mathrm{M}\left(\varphi(x)\right) = \displaystyle\int_{-\infty}^{+\infty} \varphi(u)f(u)\mathrm{d}u$.

Дисперсия $\mathrm{D}(X) = \mathrm{M}\left(X - \mathrm{M}(X)\right) = \mathrm{M}(X^2) - \mathrm{M}^2(X)$

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

Медиана существует *всегда*.

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

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

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

Нижний квартиль: $F(Q_1) = 1/4$

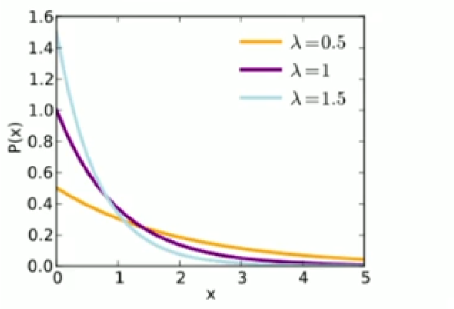
Нижний квартиль: $F(Q_3) = 3/4$

Межквартильный размах: $IQR = Q_3 - Q_1$

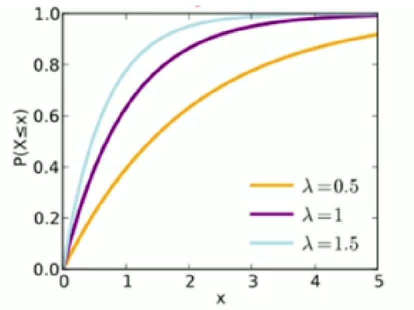
Квантиль уровня α: $F(q_\alpha) = \alpha$

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

$X \sim Exp(\lambda)$

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

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

$$\begin{aligned} f(u) &= \lambda e^{-\lambda u} \qquad u \geqslant 0,\ \lambda > 0 \\ F(x) &= 1 - e^{-\lambda x} \qquad x \geqslant 0 \\ \mathrm{M}(x) &= \frac{1}{\lambda} \\ \mathrm{D}(x) &= \frac{1}{\lambda^2} \\ \mathrm{Me}(x) &= \frac{\ln 2}{\lambda} \qquad \text{скошено вправо: } \mathrm{Me} < \mathrm{M} \end{aligned}$$

Свойство отсутствия памяти:

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

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 $P(T > \Delta t) = e^{-\lambda \Delta t} \approx 1 - \lambda \Delta t$:

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

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 распорядке будет снижаться (если это не ночной стационар).

4.6 Преобразования СВ

Лемма 4.1 (Производный поток Бернулли)**.**

$\lrcorner$ $X \sim Bernoulli(p_x)$, $Y \sim Bernoulli(p_y)$ и $Z = \mathrm{OR}(X, Y)$.

$\implies$ $Z \sim Bernoulli(p_z)$ и $p_x = p_x + p_y - p_x p_y$.

Лемма 4.2 (Линейное преобразование)**.**

$\lrcorner$ $Y = aX + B$, $a \neq 0$. Тогда

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

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

Лемма 4.3 (Монотонное преобразование)**.**

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

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

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

Лемма 4.4 (Монотонно убывающее преобразование).
$\lrcorner$ $Y = g(X)$, причём $g \searrow$. Тогда

$$F_Y(t) = 1 - F_X\left(g^{-1}(t)\right)$$
$$f_Y(u) = -\frac{f_X\left(g^{-1}(u)\right)}{g'(u)}$$

Лемма 4.5 (Получение заданной СВ из равномерной).
$\lrcorner$ $X \sim Uniform(0,1)$ и $Y = F_*^{-1}(X)$, причём $F_*^{-1}(t) \nearrow : [0,1] \to \mathbb{R}$.

Тогда $F_Y(t) = F_*(t)$

4.7 Прочие распределения

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

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & x > b \end{cases}$$
$$f(x) = \begin{cases} 0 & x < a \\ 1/(b-a) & a \leqslant x \leqslant b \\ 0 & x > b \end{cases}$$
$$\mathrm{Me} = \mathrm{M} = (a+b)/2$$
$$\mathrm{D} = (b-a)^2/12$$

Распределение Лапласа:

$$f = \frac{\lambda}{2}e^{-\lambda|x|}$$
$$\mathrm{M} = 0$$
$$\mathrm{D} = \frac{2}{\lambda^2}$$

Распределение Коши:

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

4.8 Многомерные непрерывные распределения

Опр-е. 4.4 (Совместная функция распределения "join distribution function").

$$
\begin{aligned}
&F_{X,Y}(x,y) = P(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v\\
&x_1 \leqslant x_2 \implies F_{X,Y}(x_1,y) \leqslant F_{X,Y}(x_2,y) \quad \forall y\\
&y_1 \leqslant y_2 \implies F_{X,Y}(x,y_1) \leqslant F_{X,Y}(x,y_2) \quad \forall x\\
&F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = 0\\
&F_{X,Y}(+\infty,+\infty) = 1
\end{aligned}
$$

Опр-е. 4.5 (Совместная функция плотности "join density function").

$$
\begin{aligned}
&f_{X,Y}(x,y) = \tfrac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y)\\
&f(x,y) \geqslant 0\\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v = 1
\end{aligned}
$$

Опр-е. 4.6 (Маргинальная плотность).

$$
\begin{aligned}
&f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}v\\
&f_Y(v) = \int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}u
\end{aligned}
$$

Опр-е. 4.7 (Маргинальная функция распределения).

$$
\begin{aligned}
&F_X(x) = P(X \leqslant x) = F_{X,Y}(x,\infty)\\
&F_Y(y) = P(Y \leqslant y) = F_{X,Y}(\infty,y)
\end{aligned}
$$

Опр-е. 4.8 (Совместное мат. ожидание).

$$
\mathrm{M}\left(\varphi(X,Y)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(u,v)\, f(u,v)\,\mathrm{d}u\,\mathrm{d}v
$$

Опр-е. 4.9 (X и Y независимы).

$$
\begin{aligned}
&\forall\, x,\, y\\
&P(X \leqslant x,\, Y \leqslant y) = P(X \leqslant x)\, P(Y \leqslant y)\\
&F_{X,Y}(x,y) = F_X(x)\, F_Y(y)\\
&f_{X,Y}(x,y) = f_X(x)\, f_Y(y)\\
&\mathrm{M}(X\,Y) = \mathrm{M}(X)\ \mathrm{M}(Y)\\
&\mathrm{D}(X+Y) = \mathrm{D}(X) + \mathrm{D}(Y)
\end{aligned}
$$

4.9 Условные непрерывные распределения

Опр-е. 4.10 (Условное распределение при условии события A).

$$P(A) > 0$$
$$F_{X|A}(x|A) = P(X \leqslant x\,|A) = \frac{P(X \leqslant x)}{P(A)}$$
$$f_{X|A}(x|A) = F'_{X|A}(x|A) = \frac{f_X(x)}{P(A)}$$
$$\mathrm{M}(X|A) = \int_{-\infty}^{\infty} x\, f_{X|A}(x|A)\,\mathrm{d}x$$

Опр-е. 4.11 (Условное распределение при условии $Y = y$).

$$F_{X|Y}(x|y) = \lim_{\Delta y \to 0} \frac{P(X \leqslant x,\ y < Y \leqslant y + \Delta y)}{P(y < Y \leqslant y + \Delta y)} = \frac{\frac{\partial}{\partial y} F_{X,Y}(x, y)}{f_Y(y)}$$
$$f_{X|Y}(x|y) = \tfrac{\partial}{\partial x} F_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_{X,Y}(x, y)}{\displaystyle\int_{-\infty}^{+\infty} f_{X,Y}(x, y)\mathrm{d}x}$$

Опр-е. 4.12 (Формула умножения).

$$f_{X,Y}(x, y) = f_Y(y)\, f_{X|Y}(x|y) = f_X(x)\, f_{Y|X}(y|x)$$

Опр-е. 4.13 (Формула Байеса).

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)\, f_Y(y)}{\displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y}$$

Опр-е. 4.14 (Формула Байеса для дискретно-непрерывных распределений).

$$p_{N|X}(n|x) = \frac{p_N(n)\, f_{X|N}(x|n)}{f_X(x)} = \frac{p_N(n)\, f_{X|N}(x|n)}{\sum_n p_N(n)\, f_{X|N}(x|n)}$$
$$f_{X|N}(x|n) = \frac{f_X(x)\, p_{N|X}(n|x)}{p_N(n)} = \frac{f_X(x)\, p_{N|X}(n|x)}{\displaystyle\int_{-\infty}^{\infty} f_X(x)\, p_{N|X}(n|x)\,\mathrm{d}x}$$

4.10 Условное среднее (regression function)

Опр-е. 4.15 (Условное среднее).

$$\mathrm{M}(X|Y = y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\mathrm{d}x$$

Лемма 4.6 (Формула полного среднего)**.** $\mathrm{M}\left(\mathrm{M}(X|Y)\right) = \mathrm{M}(X)$
Док-во.

$$\begin{aligned}
\mathrm{M}(X|Y=y) &= \int_{\mathbb{R}} x\, f_{X|Y}(x|y)\,\mathrm{d}x = \varphi(Y) \\
\mathrm{M}\left(\varphi(Y)\right) &= \int_{\mathbb{R}} \varphi(y) f_Y(y)\mathrm{d}y = \int_{\mathbb{R}}\int_{\mathbb{R}} x\, f_{X|Y}(x|y) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} x\, f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}y\right) x\,\mathrm{d}x = \\
&= \int_{\mathbb{R}} x\, f_X(x)\,\mathrm{d}x = \mathrm{M}(X)
\end{aligned}$$

□

Лемма 4.7 (Свойства условного среднего)**.**

$$\begin{aligned}
&\mathrm{M}(aX + bY\,|Z) = a\,\mathrm{M}(X|Z) + b\,\mathrm{M}(Y|Z) \\
&\mathrm{M}(g(y)\cdot X\,|Y) = g(y)\cdot \mathrm{M}(X|Y)
\end{aligned}$$

Теор. 4.1. $\lrcorner$ $\varphi(Y) = \mathrm{M}(X|Y)$. Тогда для $\forall$ функции $g(Y)$

$$\mathrm{M}\left(X - g(Y)\right)^2 \geqslant \mathrm{M}\left(X - \varphi(Y)\right)^2$$

5 Вспомогательные формулы

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

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

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

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

$$\sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} a_{ij}$$

$$\sum_{n=1}^{\infty} \frac{n}{2^n} = 2$$