Теория Вероятности

Содержание

Конспект курса “Теория вероятностей – наука о случайности” (часть I, часть II).

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: $\overline{C_n^k} = \left(\!\!\dbinom{n}{k}\!\!\right) = C_{n+k-1}^k$

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 $\displaystyle\sum_{k=0}^{\infty} (-1)^k \frac{1}{k} = \frac{1}{e}$

$$!n = \left\lceil \frac{n!+1}{e} \right\rceil$$

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

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

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

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:

$\lrcorner$ $\{H_i\}$ – разбиение Ω, т.е. $\Omega = H_1 \cup ... \cup H_n$ и $H_i \cap H_j = \varnothing$. Тогда:

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

Формула Байеса:

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

3 Дискретные СВ

3.1 Дискретные распределения

Случайная величина – это функция (отображение) $X : \Omega \to \mathbb{R}$
Значение случайной велечины – 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 $\equiv$ Multivariate discrete distribution

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

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

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

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

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

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

Док-во.

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

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

□

3.5 Простые распределения

Распределение Бернулли (Bernoulli distribution):

$$P(X = 1) = p \quad P(X = 0) = q \quad p \in [0, 1] \quad q = 1 - p$$

$$\mathrm{M}(X) = p$$

Равномерное дискретное распределение (Discrete uniform distribution):

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

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

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.

3.6 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 $M = np$

Из чего можно получить красивое равенство:

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

3.7 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.

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

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

Док-во.

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

□

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

3.8 Распределение Пуассона (Poisson distribution)

Является приближением последовательности биномиальных распределений:

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

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

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

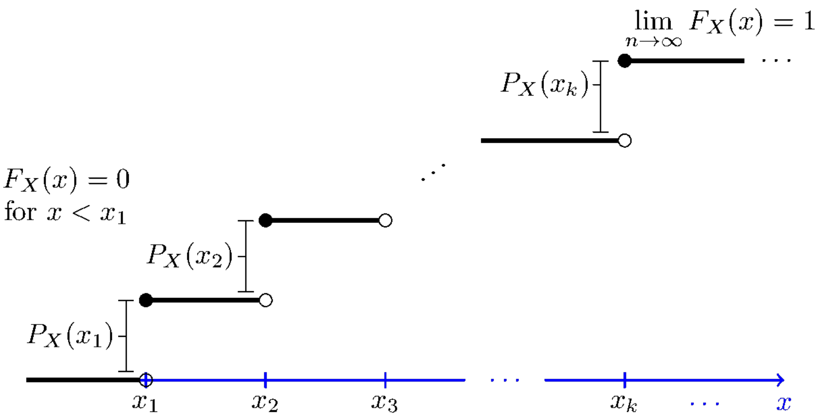
Решение. **Да.**
Каждую точку можно рассматривать как отрезок $\left[\dfrac{m}{n}, \dfrac{m}{n}\right] = \left(-\infty, \dfrac{m}{n}\right] \setminus \left(-\infty, \dfrac{m}{n}\right)$, причём множество всех рациональных точек на прямой счетно.

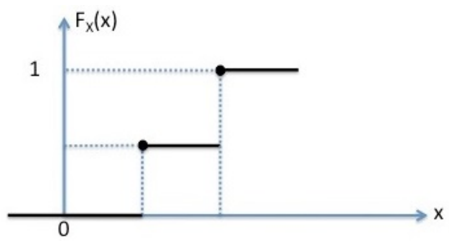
4.2 Функции распределения и плотности

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

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: $F(Q_1) = 1/4$

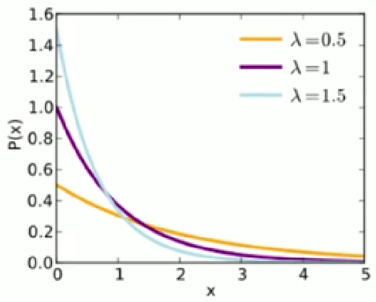
Нижний квартиль: $F(Q_3) = 3/4$

Межквартильный размах: $IQR = Q_3 - Q_1$

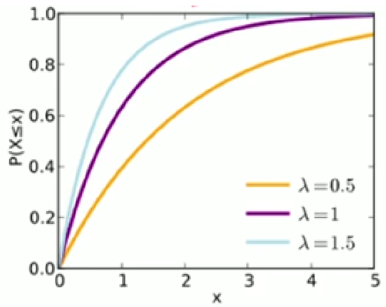
Квантиль уровня α: $F(q_\alpha) = \alpha$

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: $F(X) = P(X \leqslant x) = \int_{0}^{x} \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}$

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

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:

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

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 4.2. 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 4.3. 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 Прочие распределения

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

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

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

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

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

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

$$\int_0^\infty e^{-ax} dx = \frac{1}{a}$$

$$\int_0^\infty x e^{-ax} dx = \frac{1}{a^2}$$

$$\int_0^\infty x^2 e^{-ax} dx = \frac{2}{a^3}$$

$$\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$$

$$a^3 - b^3 = (a-b)(a^2 + ab + b^2)$$
$$(a-b)^3 = a^3 - 3a^2b + 3ab^2 - b^3$$

$$\sum_{k=0}^{n} kC_n^k p^k q^{n-k} = np$$
$$\sum_{k=0}^{\infty} kq^k = \frac{1}{p^2}$$
$$\sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} a_{ij}$$