Финансовый университет
при правительстве Российской Федерации

Шамраева
Виктория Викторовна

кандидат физико-математических наук,
доцент кафедры
математики и анализа данных

Теория вероятностей и математическая статистика

НАПРАВЛЕНИЕ ПОДГОТОВКИ: «Прикладная математика - ПМ»

КВАЛИФИКАЦИЯ (СТЕПЕНЬ): бакалавр

Часть 1 – Теория вероятностей
Раздел 2 – Математическая статистика

Абсолютно-непрерывные случайные величины

Условные распределения и условные плотности.
Условное математическое ожидание и его свойства.
Формула полного математического ожидания.
Условная дисперсия.
Формула полной дисперсии.
Условная ковариация случайных величин X и Y относительно случайной величины Z.
Формула полной ковариации.

Распределение функций от случайных величин и векторов с абсолютно-непрерывным распределением.

Абсолютно-непрерывные случайные векторы

Пример. Пусть двумерная случайная величина имеет функцию плотности $f(x,y) = x + y$, сосредоточенная в квадрате

$$0 \le x \le 1, 0 \le y \le 1.$$

Найти **условные плотности** и проверить **независимость** случайных величин $X,\ Y$.

Решение.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,\ f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx,$$

$$f(x|y) = \frac{f(x,y)}{f_Y(y)};\ f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Абсолютно-непрерывные случайные векторы

Числовые характеристики абсолютно-непрерывного случайного вектора

Абсолютно-непрерывные случайные векторы

Математическое ожидание функции случайного вектора вычисляется по формуле:

$$E[\varphi(x,y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x,y)\, f_{X,Y}(x,y)\,dxdy$$

от компонент

$$E[X] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f_X(x,y)\,dxdy,\quad E[Y] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f_Y(x,y)\,dxdy,$$

если только интегралы в правой части абсолютно сходятся.

Абсолютно-непрерывные случайные векторы

Дисперсия от компонент

$$Var(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E(X))^2\, f_{X,Y}(x,y)dxdy;$$
$$Var(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - E(Y))^2\, f_{X,Y}(x,y)dxdy,$$

если только интегралы в правой части абсолютно сходятся.

Абсолютно-непрерывные случайные векторы

Понятие **условного математического ожидания** можно распространить на абсолютно непрерывные случайные величины, при этом сохраняются все перечисленные свойства, которые нами были введены для дискретного случайного вектора.

Абсолютно-непрерывные случайные векторы

Пусть на одном и том же пространстве элементарных событий Ω заданы две **абсолютно непрерывные случайные** величины X и Y.

Условным законом распределения случайной величины X при условии $Y = y$, так же как и в случае **дискретных случайных величин**, называется любое соотношение, ставящее в соответствие значениям случайной величины X условные вероятности их принятия при условии $Y = y$.

Абсолютно-непрерывные случайные векторы

Условным математическим ожиданием с.в. $\boldsymbol{X}$ **относительно с.в.** $\boldsymbol{Y}$ называется с.в. $E(X|Y)$, которая принимает значение

$$E(X\,|Y = y)$$

при $Y = y$.

Условное математическое ожидание $E(X\,|Y = y)$, как функция от y называется **функцией регрессии** X на Y.

Абсолютно-непрерывные случайные векторы

Условное математическое ожидание непрерывной случайной величины при известном Y вычисляется по формуле:

$$x(y) = E(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx = \int_{-\infty}^{+\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx.$$

Аналогично,

$$y(x) = E(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy = \int_{-\infty}^{+\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)} dy.$$

Абсолютно-непрерывные случайные векторы

Неслучайные условные числовые характеристики X можно получить через подходящую функцию $\varphi(x)$ в виде интеграла

$$E(\varphi(X)\,|Y=y)=\int_{\mathcal{X}}\varphi(X)f_{X|Y}(x|y)\,dx.$$

Случайная X величина с вероятностью 1 принимает значения в промежутке $\mathcal{X}$, $P(X\in\mathcal{X})=1$, а Y – в промежутке $\mathcal{Y}, P(Y\in\mathcal{Y})=1$, $\mathcal{X}\times\ \mathcal{Y}$ – декартово произведение промежутков $\mathcal{X}$ и $\mathcal{Y}$, случайный вектор (X,Y) имеет плотность $f_{X,Y}(x,y)$, равную нулю вне области $\mathcal{X}\times\mathcal{Y}$.

Абсолютно-непрерывные случайные векторы

Случайные условные числовые характеристики X, получаются, если значение y выбирается случайным образом в соответствии с распределением Y. Этот факт можно записать в виде

$$E(\varphi(X)\,|Y) = \int_{\mathcal{X}} \varphi(X) f_{X|Y}(x|Y)\,dx.$$

Случайная X величина с вероятностью 1 принимает значения в промежутке $\mathcal{X}$, $P(X \in \mathcal{X}) = 1$, а Y − в промежутке $\mathcal{Y}, P(Y \in \mathcal{Y}) = 1$, $\mathcal{X} \times \mathcal{Y}$ − декартово произведение промежутков $\mathcal{X}$ и $\mathcal{Y}$, случайный вектор (X, Y) имеет плотность $f_{X,Y}(x, y)$, равную нулю вне области $\mathcal{X} \times \mathcal{Y}$.

Абсолютно-непрерывные случайные векторы

Таким образом, $E(\varphi(X)\,|Y)$ - случайная величина, которая при каждом элементарном исходе $\omega \in \Omega$, принимает значение

$$E(\varphi(X)|Y)(\omega) = \int_{\mathcal{X}} \varphi(X) f_{X|Y}(x|Y(\omega))dx.$$

Случайная X величина с вероятностью 1 принимает значения в промежутке $\mathcal{X}$, $P(X \in \mathcal{X}) = 1$, а Y – в промежутке $\mathcal{Y}, P(Y \in \mathcal{Y}) = 1$, $\mathcal{X} \times \mathcal{Y}$ – декартово произведение промежутков $\mathcal{X}$ и $\mathcal{Y}$, случайный вектор (X, Y) имеет плотность $f_{X,Y}(x, y)$, равную нулю вне области $\mathcal{X} \times \mathcal{Y}$.

Абсолютно-непрерывные случайные векторы

Например, **случайный условный момент порядка k** можно представить как

$$E\left(X^k \middle| Y\right) = \int_{\mathcal{X}} x^k f_{X|Y}(x|Y)\, dx.$$

Если условная плотность $f_{X|Y}$ совпадает с обычной плотностью некоторого распределения $\mathcal{L}$, то используется обозначение

$$X|Y \sim \mathcal{L}.$$

Случайная X величина с вероятностью 1 принимает значения в промежутке $\mathcal{X}$, $P(X \in \mathcal{X}) = 1$, а Y – в промежутке $\mathcal{Y}, P(Y \in \mathcal{Y}) = 1$, $\mathcal{X} \times \mathcal{Y}$ – декартово произведение промежутков $\mathcal{X}$ и $\mathcal{Y}$, случайный вектор (X, Y) имеет плотность $f_{X,Y}(x, y)$, равную нулю вне области $\mathcal{X} \times \mathcal{Y}$.

Абсолютно-непрерывные случайные векторы

Например, если, скажем, Y – положительная случайная величина, а условная плотность X имеет вид

$$f_{X|Y}(x|y) = \begin{cases} 0, & x \notin [0, y]; \\ y^{-1}, & x \in [0, y]; \end{cases}$$

то говорим, что X при фиксированном Y **равномерно распределена на отрезке $[\mathbf{0}, \boldsymbol{Y}]$**, и используем обозначение

$$X|Y \sim \mathrm{Unif}[0, Y].$$

Если условная плотность $f_{X|Y}$ совпадает с обычной плотностью некоторого распределения $\mathcal{L}$, то используется обозначение

$$X|Y \sim \mathcal{L}.$$

Абсолютно-непрерывные случайные векторы

Случайный вектор (X, Y) называется **равномерно распределённым** в области $G \subset \mathbb{R}^2$, если для него существует плотность распределения вида

$$f_{X,Y}(x, y) = \begin{cases} 0, & (x, y) \notin G, \\ |G|^{-1}, & (x, y) \in G, \end{cases}$$

где $|G|$ - площадь G.

Абсолютно-непрерывные случайные векторы

Пример. Пусть двумерный случайный вектор (X,Y) **равномерно распределен** внутри треугольника

$$\Delta = \{(x, y): x > 0, y > 0, x + y < 2\}.$$

Найти **условное распределение** X и **условное математическое ожидание** при условии $Y = y$.

Решение.

Абсолютно-непрерывные случайные векторы

Условное математическое ожидание $E(X|Y)$ обладает следующими **свойствами**:

1. $E(c|Y) = c$, где c- const.

2. $E(aX + b|Y) = aE(X|Y) + b$.

3. $E(X + Y|Z) = E(X|Z) + E(Y|Z)$.

4. Если X и Y– независимые с.в., то $E(X|Y) = E(X)$.

5. $E(\varphi(Y) \cdot X|Y) = \varphi(Y) \cdot E(X|Y)$.

Абсолютно-непрерывные случайные векторы

Теорема. Пусть $Geom(p)$ - **геометрическое распределение**, $n \in \mathbb{N}$; $Exp(\lambda)$- **показательное распределение**, $c \in \mathbb{R}_+$.
Тогда:

1) $X \sim Geom(p) \Rightarrow X - n | X > n \sim Geom(p)$;

2) $X \sim Exp(\lambda) \Rightarrow X - c | X > c \sim Exp(\lambda)$.

Абсолютно-непрерывные случайные векторы

Пример. Случайная величина распределена по **показательному закону**, $X \sim Exp(\lambda)$. Найти условное математическое ожидание $E(X|X > c)$, где константа $c > 0$.

Решение.

Теорема. Пусть $Geom(p)$ - **геометрическое распределение**, $n \in \mathbb{N}$; $Exp(\lambda)$- **показательное распределение**, $c \in \mathbb{R}_+$.
Тогда:
1) $X \sim Geom(p) \Rightarrow X - n|X > n \sim Geom(p)$;
2) $X \sim Exp(\lambda) \Rightarrow X - c|X > c \sim Exp(\lambda)$.

Абсолютно-непрерывные случайные векторы

Теорема (формула полного математического ожидания).

$$E(X) = E(E(X|Y)).$$

Абсолютно-непрерывные случайные векторы

Понятие условной дисперсии, как и понятие условного математического ожидания, можно распространить на **абсолютно непрерывные случайные величины**, при этом перечисленные свойства также сохраняются.

Абсолютно-непрерывные случайные векторы

Условной дисперсией с.в. X относительно с.в. Y называется с. в.

$$Var(X|Y) \equiv E[(X - E(X|Y))^2|Y],$$

которая принимает значение $Var(X\ |Y\ =\ y)$ при $Y\ =\ y$.

Для **непрерывного** случайного вектора,

$$Var(X\ |Y\ =\ y) = \int_{\mathcal{X}} (x - E(X|Y = y))^2\ f_{X|Y}(x|y)dx.$$

Случайная X величина с вероятностью 1 принимает значения в промежутке $\mathcal{X}$, $P(X \in \mathcal{X}) = 1$, а Y – в промежутке $\mathcal{Y}, P(Y \in \mathcal{Y}) = 1$, $\mathcal{X} \times \mathcal{Y}$ – декартово произведение промежутков $\mathcal{X}$ и $\mathcal{Y}$, случайный вектор (X, Y) имеет плотность $f_{X,Y}(x, y)$, равную нулю вне области $\mathcal{X} \times \mathcal{Y}$.

Абсолютно-непрерывные случайные векторы

Свойства условной дисперсии

1. $Var(c|Y) = 0.$

2. $Var(aX + b|Y) = a^2 Var(X|Y).$

3. $Var(X|Y) = E(X^2|Y) - E(X|Y)2.$

4. Если X и Y– независимые с.в., то $Var(X|Y) = Var(X).$

5. $Var(\varphi(Y) \cdot X|Y) = \varphi^2(Y) \cdot Var(X|Y).$

Абсолютно-непрерывные случайные векторы

Теорема (**формула полной дисперсии**).

$$Var(X) = E[Var(X \mid Y)] + Var[E(X \mid Y)].$$

$$Var(Y) = E[Var(Y \mid X)] + Var[E(Y \mid X)].$$

Абсолютно-непрерывные случайные векторы

Пример 1. Ежедневное количество покупателей магазина, совершивших покупку, описывается случайной величиной X, распределенной по **биномиальному закону** с параметрами $n = 500$ и $p = 0.54$. А сумма чека (в рублях) каждого из покупателей описывается случайной величиной Y, распределенной по **нормальному закону** с параметрами $m = 5500$ и $\sigma = 80$.

Оцените **методом Монте-Карло** ежедневную среднюю выручку магазина и ее дисперсию. В поля ответов введите полученные значения для среднего (E), дисперсии (Var) и среднеквадратического отклонения (σ) ежедневной выручки.

Решение.

Абсолютно-непрерывные случайные векторы

Пример 2. Средний ущерб от одного пожара составляет $3{,}2$ млн. руб. Предполагается, что ущерб распределен по **показательному закону**, а число пожаров за год - по **закону Пуассона**. Также известно, что за 10 лет в среднем происходит 19 пожаров.

Найдите:

1) математическое ожидание суммарного ущерба от всех пожаров за один год;
2) стандартное отклонение суммарного ущерба от пожаров за год.

Решение.

Абсолютно-непрерывные случайные векторы

Ковариацией с.в. X **и** Y называется математическое ожидание произведения отклонений X и Y от их математических ожиданий:

$$\boxed{\boldsymbol{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))].}$$

Абсолютно-непрерывные случайные векторы

Свойства ковариации с.в. X **и** Y **:**

1. $\boldsymbol{Cov}(X,Y) = E(XY) - E(X) \cdot E(Y)$.

2. Если X и Y независимы, $\boldsymbol{Cov}(X,Y) = 0$.

3. $\boldsymbol{Cov}(X,Y) = \boldsymbol{Cov}(Y,X)$

4. $\boldsymbol{Cov}(aX + b, cY + d) = ac \cdot \boldsymbol{Cov}(X,Y)$, если a, b, c, d – константы.

5. $\boldsymbol{Cov}(X + Y, Z) = \boldsymbol{Cov}(X,Z) + \boldsymbol{Cov}(Y,Z)$.

6. $\boldsymbol{Cov}(X,X) = \boldsymbol{Var}(X)$

Абсолютно-непрерывные случайные векторы

Пример. Для случайной цены Y известны вероятности:

$$P(Y = 7) = 0{,}6 \text{ и } P(Y = 19) = 0{,}4.$$

При условии, что $Y = y$, распределение выручки X является **равномерным** на отрезке $[0{,}6y]$. Найдите:

1) математическое ожидание $E(XY)$;

2) ковариацию $\boldsymbol{Cov}(X, Y)$.

Решение.

Абсолютно-непрерывные случайные векторы

Условной ковариацией с.в. X и Y относительно с.в. Z называется с.в.

Свойства:

1. $\mathbf{Cov}(X, X|Z) = Var(X|Z)$;

2. $\mathbf{Cov}(a, Y|Z) = 0$, если $a = const$.

3. $\mathbf{Cov}(X, Y|Z) = E(XY|Z) - E(X|Z) \cdot E(Y|Z)$.

4. $\mathbf{Cov}(X, E(Y|X)) = \mathbf{Cov}(X, Y)$.

5. $\mathbf{Cov}(X, Y) = E[\mathbf{Cov}(X, Y|Z)] + \mathbf{Cov}(E(X|Z), E(Y|Z))$ –

формула полной ковариации.

Абсолютно-непрерывные случайные векторы

Пример. Вектор (X, Y) **равномерно распределён** в треугольнике Δ: $\left\{\frac{x}{a} + \frac{y}{b} \leq 1; x \geq 0, y \geq 0\right\}$.

Найти корреляционное отношение $\eta_{X|Y}$ и коэффициент корреляции между его компонентами, $\rho_{X,Y}$

Решение.

Для случайных величин X и Y **корреляционное отношение X на Y** задается формулой $\eta_{X|Y} = \frac{\sigma(E(X|Y))}{\sigma(X)}$, где $\sigma(.)$ – стандартное отклонение.

1. $0 \leq \eta_{X|Y} \leq 1$.
2. $\eta_{X|Y} = 0$, если X и Y – независимые случайные отношения.
3. $\eta_{X|Y} = 1$, если X – функция от Y, $X = \varphi(Y)$.
4. $\eta_{X|Y} \geq |\rho_{X,Y}|$, где $\rho_{X,Y}$ - коэффициент корреляции X и Y.
5. $\eta_{X|Y} = |\rho_{X,Y}|$ в том и только в том случае, если найдутся константы α и $\beta \neq 0$, для которых с вероятностью 1 выполняется соотношение $\mathrm{E}(X|Y) = \alpha + \beta Y$.

Дискретные случайные векторы

n-мерные случайные векторы

Дискретные случайные векторы

Упорядоченный набор

$$X = (X_1, X_2, \dots, X_n)$$

скалярных случайных величин $X_1, X_2, \dots, X_n$, заданных на в.п. $(\Omega, \mathcal{F}, P)$ называется **n-мерным случайным вектором**.

Случайные величины

$$X_1, X_2, \dots, X_n$$

при этом называются **компонентами** вектора X.

Дискретные случайные векторы

Напомним, что минимальная $\sigma-$алгебра, содержащая все открытые подмножества в $\mathbb{R}^n$, называется n–мерной борелевской σ –алгеброй и обозначается $\mathcal{B}(\mathbb{R}^n)$.

С помощью борелевской $\sigma-$ алгебры можно сформулировать **эквивалентное определение**.

***n* -мерным случайным вектором** на в.п. $(\Omega, \mathcal{F}, P)$ называется отображение $X: \Omega \longrightarrow \mathbb{R}^n$, измеримое относительно борелевской σ –алгебры $\mathcal{B}(\mathbb{R}^n)$.

Другими словами, $X: \Omega \longrightarrow \mathbb{R}^n$ - **случайный вектор**, если для любого борелевского множества $B \subset \mathbb{R}^n$ его полный прообраз $X^{-1}(B) \in \mathcal{F}$.

Дискретные случайные векторы

Замечание. Как и в случае скалярных случайных величин, полный прообраз $X^{-1}(B)$ имеет смысл множества всех элементарных исходов $\omega \in \Omega$, благоприятных для события, состоящего в том, что случайная величина X приняла значение их множества $B, \{X \in B\}$.

Поскольку событие и множество благоприятных событий суть одно и то же, вместо обозначения $X^{-1}(B)$ используем запись $\{X \in B\}$.

Дискретные случайные векторы

Распределением n-мерного случайного вектора, заданного на в.п. $(\Omega, \mathcal{F}, P)$, называется вероятностная мера P_X на σ −алгебре $\mathcal{B}(\mathbb{R}^n)$, которая всякому множеству B из $\mathcal{B}(\mathbb{R}^n)$ ставит в соответствие вероятность попадания X в B,

$$P_X(B) = P(\{X \in B\}).$$

Случайные векторы X и Y называются **одинаково распределёнными**, если их распределения совпадают (обозначения $X \sim Y$).

Дискретные случайные векторы

Функция распределения n -мерного случайного вектора Xзадаётся формулой

$$F_X(x) = F_X(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Двумерные нормальные векторы

Абсолютно-непрерывные случайные векторы

Из двумерных систем СВ особый интерес представляет **нормальное распределение на плоскости**.

Абсолютно-непрерывные случайные векторы

Непрерывная двухмерная случайная величина (X,Y) имеет **нормальное распределение**, если ее плотность вероятности равна:

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\,e^{-\frac{1}{2(1-\rho_{XY}^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2}-\frac{2\rho_{XY}(x-m_X)(x-m_Y)}{\sigma_X\sigma_Y}+\frac{(y-m_Y)^2}{\sigma_Y^2}\right]},$$

$m_X, m_Y, \sigma_X, \sigma_Y, \rho_{XY}$ - параметры распределения.

Обозначают

$$\boldsymbol{(X,Y)\sim N(m_X, m_Y, \sigma_X^2, \sigma_Y^2,} \rho_{XY}\boldsymbol{)}$$

Абсолютно-непрерывные случайные векторы

Если составляющие X, Y двумерной нормально распределенной случайной величины **некоррелированы**, то они и независимы, т.е. при $\rho_{XY} = 0$:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\, e^{-\frac{1}{2}\left[\frac{(x-m_X)^2}{\sigma_X^2}+\frac{(y-m_Y)^2}{\sigma_Y^2}\right]} =$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}}\, e^{-\frac{(x-m_X)^2}{2\sigma_X^2}}\, \frac{1}{\sigma_Y\sqrt{2\pi}}\, e^{-\frac{(y-m_Y)^2}{2\sigma_Y^2}} = f_X(x)\, f_Y(y)$$

Итак, для нормальных случайных величин понятия независимости и некоррелированности равносильны.

Абсолютно-непрерывные случайные векторы

Говорят, что случайный вектор (X, Y) имеет **стандартное двумерное нормальное распределение**, если

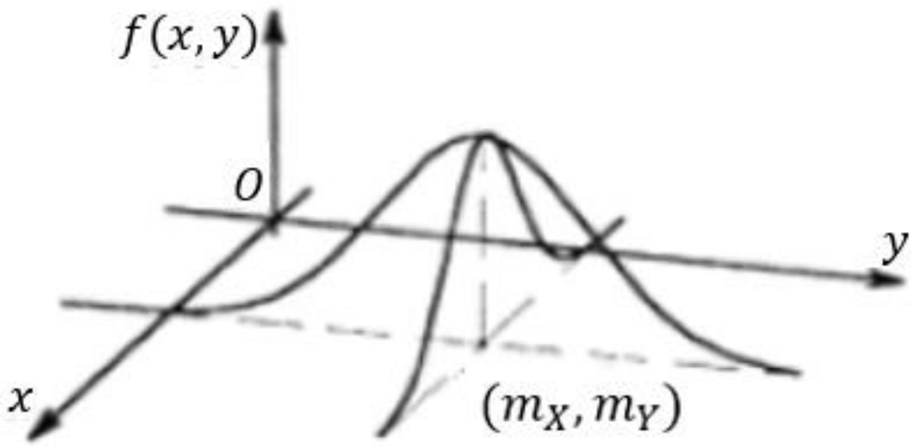
$$(\boldsymbol{X}, \boldsymbol{Y}) \sim \boldsymbol{N}(\mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{0})$$

Плотность распределения такого вектора:

$$f(x, y) = \frac{1}{2\pi}\, e^{-\frac{x^2+y^2}{2}}$$

Абсолютно-непрерывные случайные векторы

Графиком двумерной нормальной плотности является холмообразная поверхность, расположенная над плоскостью xOy , асимптотически приближающаяся к ней на бесконечности, симметричная относительно вертикальной оси, проходящей через центр (m_X, m_Y), и с вершиной в этой точке.



Любое сечение этой поверхности плоскостью, перпендикулярной xOy, является **кривой Гаусса**.

Абсолютно-непрерывные случайные векторы

График плотности вероятности с параметрами, проходящей через центр $(m_X = -1, m_Y = 1), \sigma_X = \sigma_Y = 2, \rho_{XY} = 0$.

Абсолютно-непрерывные случайные векторы

Пример. Случайная точка на плоскости (X, Y) распределена **по нормальному закону** с центром рассеивания $(m_X, m_Y) = (0,1)$ и среднеквадратическими отклонениями $\sigma_X = 1, \sigma_Y = 2$. X и Y **независимы**. Вычислить вероятность попадания случайной точки в прямоугольник с вершинами $(-1; 1)$, $(2; 1)$, $(2; 3), (-1; 3)$.

Решение.

$$\boldsymbol{P}(\, x_1 < X \le x_2, y_1 < Y \le y_2) = F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}\,(x_2, y_1)$$

$$F(x, y) = P(X < x, Y < y) = P(X < x)p(Y < y) = F_X\,(x)F_Y(y)$$

Абсолютно-непрерывные случайные векторы

Свойства двумерного нормального распределения

1. Пятый параметр двумерного нормального распределения $\boldsymbol{\rho_{XY}}$ совпадает с коэффициентом корреляции компонент $\boldsymbol{X}$ и $\boldsymbol{Y}$.

2. Если $(\boldsymbol{X}, \boldsymbol{Y}) \sim \boldsymbol{N}(\boldsymbol{m_X}\ ,\ \boldsymbol{m_Y}\ ,\ \boldsymbol{\sigma_X^2}\ ,\ \boldsymbol{\sigma_Y^2}\ ,\ \boldsymbol{\rho_{XY}})$, то $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{m_X}, \boldsymbol{\sigma_X^2}), \boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{m_Y}, \boldsymbol{\sigma_Y^2})$.

3. Для нормального случайного вектора $(\boldsymbol{X}, \boldsymbol{Y})$ понятия независимости и некоррелированности компонент $\boldsymbol{X}$ и $\boldsymbol{Y}$ эквивалентны.

4. Если случайные величины $\boldsymbol{X}$ и $\boldsymbol{Y}$ таковы, что каждая из них имеет **нормальный закон распределения**, а их совместное распределение является **двумерным нормальным**, то сумма $\boldsymbol{X} + \boldsymbol{Y}$ имеет **нормальный закон распределения**.

Абсолютно-непрерывные случайные векторы

Пример. Пусть $\boldsymbol{X} \sim \boldsymbol{N}(\mathbf{1}, \mathbf{4}), \boldsymbol{Y} \sim \boldsymbol{N}(\mathbf{2}, \mathbf{16})$ независимые случайные величины. Записать плотность распределения случайного вектора $(\boldsymbol{X}, \boldsymbol{Y})$.

Решение.

3. Для нормального случайного вектора $(\boldsymbol{X}, \boldsymbol{Y})$ понятия независимости и некоррелированности компонент $\boldsymbol{X}$ и $\boldsymbol{Y}$ эквивалентны.

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho_{XY}(x-m_X)(x-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]}$$

Абсолютно-непрерывные случайные векторы

Следствие. Пусть (X, Y) — **нормально распределенный случайный вектор**. Тогда при любых $a, b \in \mathbb{R}$ случайная величина

$$aX + bY$$

имеет **нормальный закон распределения**.

Абсолютно-непрерывные случайные векторы

Пример. Известно, что $Z = (X, Y)$, где с.в. X - возраст наугад взятого человека из определённой группы и с.в. Y – его артериальное давление, удовлетворительно описывается двумерным нормальным законом распределения с математическим ожиданием $E(Z) = (58; 137)$ и ковариационной матрицей

$$C = \begin{pmatrix} 94 & 90 \\ 90 & 108 \end{pmatrix}.$$

Найти: $P(Y - X \leq 75)$.

Решение.

---

2. Если $(\boldsymbol{X}, \boldsymbol{Y}) \sim \boldsymbol{N}(\boldsymbol{m_X}, \boldsymbol{m_Y}, \boldsymbol{\sigma_X^2}, \boldsymbol{\sigma_Y^2}, \boldsymbol{\rho_{XY}})$, то $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{m_X}, \boldsymbol{\sigma_X^2})$, $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{m_Y}, \boldsymbol{\sigma_Y^2})$

4. Если случайные величины $\boldsymbol{X}$ и $\boldsymbol{Y}$ таковы, что каждая из них имеет **нормальный закон распределения**, а их совместное распределение является **двумерным нормальным**, то сумма $\boldsymbol{X} + \boldsymbol{Y}$ имеет **нормальный закон распределения**.

Абсолютно-непрерывные случайные векторы

Удобно, плотность двумерного нормального распределения определять **с использованием квадратичных форм**.

Приведем необходимые сведения из этой области.

Абсолютно-непрерывные случайные векторы

Квадратичной формой от двух переменных x_1, x_2 называется выражение

$$Q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + \cdots,$$

в котором a_{11}, a_{12} и a_{22} — действительные числа, не равные нулю одновременно.

Абсолютно-непрерывные случайные векторы

Матрицей квадратичной формы называется матрица

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Квадратичная форма $Q(x_1, x_2)$ называется **положительно определенной**, если при всех значениях переменных x_1, x_2, не равных одновременно нулю, выполняется неравенство

$$Q(x_1, x_2) > 0.$$

Абсолютно-непрерывные случайные векторы

С учётом этого, говорят, что случайный вектор (X_1, X_2) имеет **двумерное нормальное распределение**, если его плотность распределения определяется формулой

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}Q(x_1-m_1, x_2-m_2)},$$

где функция двух переменных

$$Q(x_1-m_1, x_2-m_2) =$$
$$= \frac{1}{1-\rho^2}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)$$

есть положительно определенная квадратичная форма, $\rho^2 = R_{12}^2$.

$Q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$ — **квадратичная форма** от двух переменных x_1, x_2.

Абсолютно-непрерывные случайные векторы

Итак,

$$(\boldsymbol{X_1}, \boldsymbol{X_2}) \sim \boldsymbol{N}(\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{\sigma_1^2}, \boldsymbol{\sigma_2^2}, \boldsymbol{\rho})$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}Q(x_1-m_1, x_2-m_2)}$$

$$Q(x_1 - m_1, x_2 - m_2) = \frac{1}{1-\rho^2}\left(\frac{(x_1 - m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2}\right)$$

Абсолютно-непрерывные случайные векторы

Ещё одно **свойство двумерного нормального распределения**:

5. Пусть C — ковариационная матрица вектора (X_1, X_2) :

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Тогда матрица A квадратичной формы Q, с помощью которой определяется плотность распределения $f(x_1, x_2)$, совпадает с матрицей, обратной к матрице C: $A = C^{-1}$.

Доказательство. Находим определитель матрицы C:

$$\det(C) = \sigma_1^2\sigma_2^2(1 - \rho^2).$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}\, e^{-\frac{1}{2}Q(x_1 - m_1, x_2 - m_2)}$$

$$Q(x_1 - m_1, x_2 - m_2) = \frac{1}{1 - \rho^2}\left(\frac{(x_1 - m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2}\right)$$

Абсолютно-непрерывные случайные векторы

Доказательство (продолжение).

$$C^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

Как несложно убедиться, полученная матрица совпадает с матрицей квадратичной формы Q. □

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}Q(x_1-m_1, x_2-m_2)}$$

$$Q(x_1 - m_1, x_2 - m_2) = \frac{1}{1-\rho^2}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)$$

Пусть C — ковариационная матрица вектора (X_1, X_2) :

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

5. Тогда матрица A квадратичной формы Q, с помощью которой определяется плотность распределения $f(x_1, x_2)$, совпадает с матрицей, обратной к матрице C: $A = C^{-1}$.

Абсолютно-непрерывные случайные векторы

$$Q(x_1 - m_1, x_2 - m_2) = \begin{pmatrix} x_1 - m_1 \\ x_2 - m_2 \end{pmatrix}^T C^{-1} \begin{pmatrix} x_1 - m_1 \\ x_2 - m_2 \end{pmatrix}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \, e^{-\frac{1}{2}Q(x_1 - m_1, x_2 - m_2)}$$

$$Q(x_1 - m_1, x_2 - m_2) = \frac{1}{1-\rho^2}\left(\frac{(x_1 - m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2}\right)$$

Пусть C — ковариационная матрица вектора (X_1, X_2) :

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

5. Тогда матрица A квадратичной формы Q, с помощью которой определяется плотность распределения $f(x_1, x_2)$, совпадает с матрицей, обратной к матрице C: $A = C^{-1}$.

Абсолютно-непрерывные случайные векторы

Пример. Случайный вектор (X_1, X_2) распределен по нормальному закону с плотностью

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{5}}\ e^{-\frac{9}{2}x_1^2+2x_1x_2-\frac{x_2^2}{2}+28x_1-9x_2-\frac{101}{2}}.$$

Требуется найти математическое ожидание $E(X_1)$, дисперсию $D(X_2)$ и коэффициент корреляции ρ_{x_1,x_2}.

Решение.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\ e^{-\frac{1}{2}Q(x_1-m_1,x_2-m_2)}$$

$$Q(x_1-m_1, x_2-m_2) = \frac{1}{1-\rho^2}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)$$

Абсолютно-непрерывные случайные векторы

Теорема. Если случайный вектор (X, Y) имеет нормальное распределение, $(X,Y) \sim N(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$, то

$$(X|Y=y) \sim N\left(m_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_Y); \sigma_X^2(1-\rho_{XY}^2)\right),$$

$$(Y|X=x) \sim N\left(m_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-m_X); \sigma_Y^2(1-\rho_{XY}^2)\right),$$

т. е. условная плотность одной из компонент при фиксированном значении другой является нормальной, причем справедливы формулы

$$f_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho_{XY}^2)}}\, e^{-\frac{1}{2\sigma_X^2(1-\rho_{XY}^2)}\left(x-m_X-\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_Y)\right)^2},$$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho_{XY}^2)}}\, e^{-\frac{1}{2\sigma_Y^2(1-\rho_{XY}^2)}\left(y-m_Y-\rho_{XY}\frac{\sigma_Y}{\sigma_X}(y-m_X)\right)^2},$$

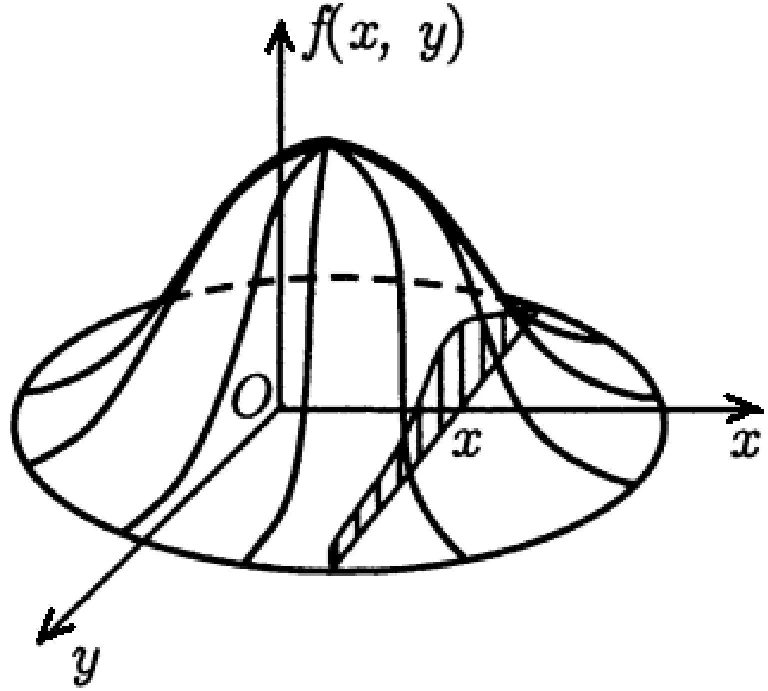
$$E(X|Y=y) = m_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_Y),\ D(X|Y=y) = \sigma_X^2(1-\rho_{XY}^2),$$

$$E(Y|X=x) = m_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-m_X),\ D(Y|X=x) = \sigma_Y^2(1-\rho_{XY}^2).$$

Абсолютно-непрерывные случайные векторы

$(X,Y) \sim N(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho)$,

$z = f(x,y)$ - плотность распределения случайного вектора (X,Y)



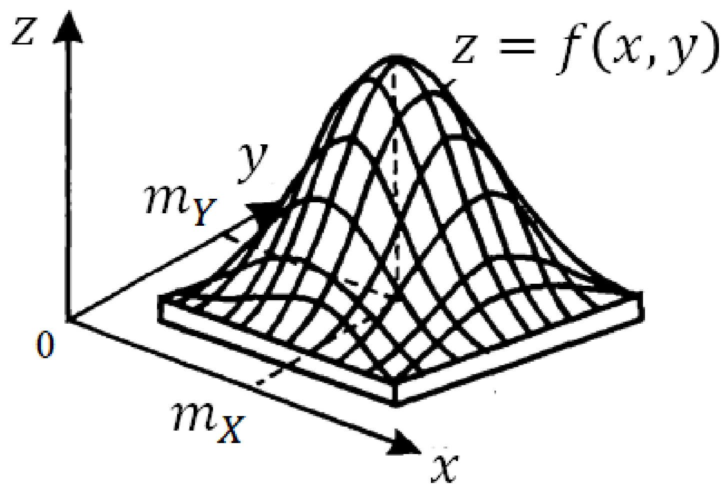
$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]},$$

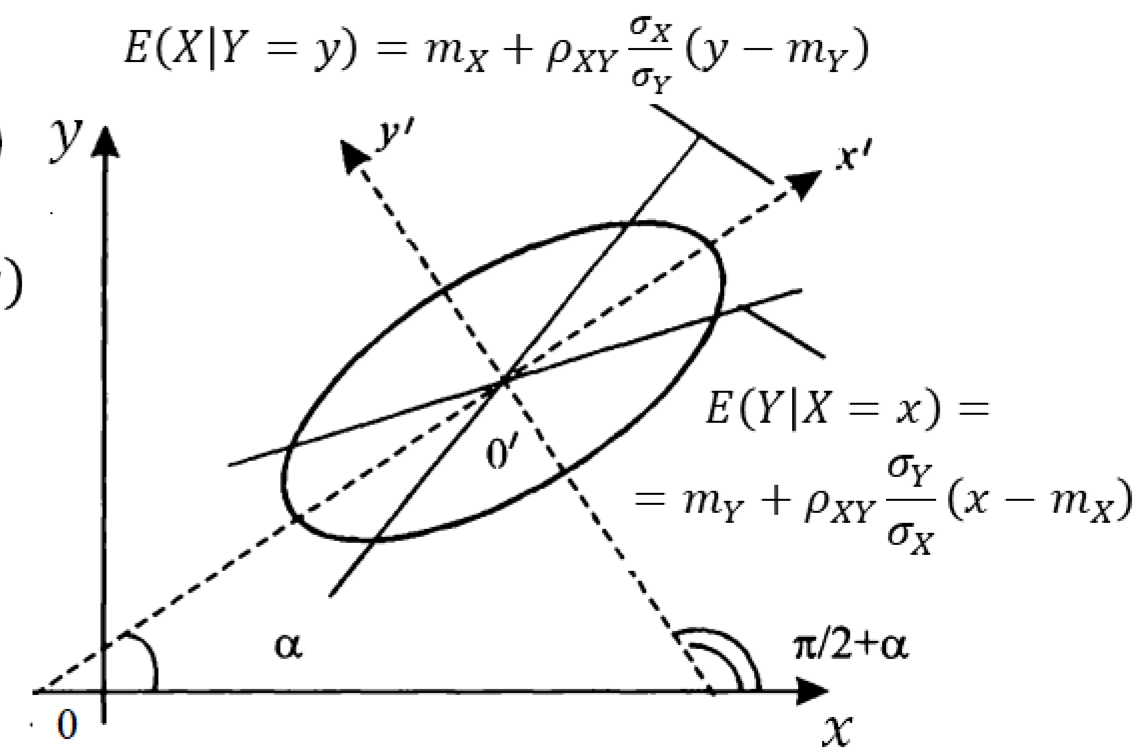
$$f_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho_{XY}^2)}}\, e^{-\frac{1}{2\sigma_X^2(1-\rho_{XY}^2)}\left(x-m_X-\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_Y)\right)^2},$$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho_{XY}^2)}}\, e^{-\frac{1}{2\sigma_Y^2(1-\rho_{XY}^2)}\left(y-m_Y-\rho_{XY}\frac{\sigma_Y}{\sigma_X}(y-m_X)\right)^2}$$

Абсолютно-непрерывные случайные векторы

$(X,Y) \sim N(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho)$





$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]},$$

$$E(X|Y=y) = m_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - m_Y),\ D(X|Y=y) = \sigma_X^2(1-\rho_{XY}^2),$$

$$E(Y|X=x) = m_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - m_X),\ D(Y|X=x) = \sigma_Y^2(1-\rho_{XY}^2).$$

Абсолютно-непрерывные случайные векторы

Пример. Плотность распределения случайного вектора (X, Y) имеет вид

$$f(x,y) = \frac{1}{2\pi}\, e^{-\frac{5}{2}x^2 - 10x - 10 - 3yx - 6y - y^2}.$$

Найдите условное математическое ожидание $E(X|Y = y)$ и $D(X|Y = y)$.

Решение.

$(X,Y) \sim N(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho)$,

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]},$$

$E(X|Y = y) = m_X + \rho\frac{\sigma_X}{\sigma_Y}(y - m_Y)$, $D(X|Y = y) = \sigma_X^2(1-\rho^2)$

Системы n случайных величин

Системы случайных величин

Пусть случайная величина Y является линейной комбинацией с.в. X_1 и X_2:

$$Y = a_1 X_1 + a_2 X_2 .$$

Тогда

$$\mathbf{Var}(Y) = a_1^2 \mathbf{Var}(X_1) + a_2^2 \mathbf{Var}(X_2) + 2a_1 a_2 \operatorname{cov}(X_1, X_2) =$$
$$= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2a_1 a_2 \rho(X_1, X_2) \sigma_1 \sigma_2 .$$

Обобщим эту формулу на n слагаемых.

Системы случайных величин

Пусть случайная величина Y является линейной комбинацией с.в. X_1 , ..., X_n . Тогда справедлива формула

$$\mathbf{Var}(Y) = \sum_{i,j=1}^{n} \operatorname{cov}(X_i , X_j)\, a_i\, a_j = \sum_{i,j=1}^{n} \rho(X_i , X_j)\sigma_i\, \sigma_j\; a_i\, a_j \,. \qquad (*)$$

Упражнение. Доказать формулу (*)

Формуле (*) можно придать более компактный вид, если воспользоваться матричными обозначениями.

Системы случайных величин

Для набора с.в. $X_1, \ldots, X_n$ **ковариационной матрицей** $C = \left(c_{ij}\right)$ и **корреляционной матрицей** $R = \left(\rho_{ij}\right)$ называют квадратные матрицы порядка n, составленные из всех парных ковариаций $c_{ij} = \mathrm{cov}\left(X_i, X_j\right)$ и всех коэффициентов корреляции $\rho_{ij} = \rho\left(X_i, X_j\right), i, j = 1, 2, \ldots, n$.

Системы случайных величин

$$C = \begin{pmatrix} \mathbf{Var}(X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathbf{Var}(X_2) & & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \cdots & \mathbf{Var}(X_n) \end{pmatrix}$$

$$R = \begin{pmatrix} \mathbf{1} & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_2, X_1) & \mathbf{1} & & \rho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \rho(X_n, X_2) & \cdots & \mathbf{1} \end{pmatrix}$$

Системы случайных величин

Свойства.

1. Ковариационная и корреляционные матрицы являются симметричными матрицами.
2. Ковариационная и корреляционные матрицы неотрицательно определены.
3. Определители этих матриц неотрицательны. Более того, $\det R \leq 1$.

Упражнение. Доказать свойства 1)-3).

Для набора с.в. $X_1, \ldots, X_n$ **ковариационной матрицей** $C = \left(c_{ij}\right)$ и **корреляционной матрицей** $R = \left(\rho_{ij}\right)$ называют квадратные матрицы порядка n, составленные из всех парных ковариаций $c_{ij} = \operatorname{cov}\left(X_i, X_j\right)$ и всех коэффициентов корреляции $\rho_{ij} = \rho\left(X_i, X_j\right), i, j = 1,2, \ldots, n$.

Системы случайных величин

Пусть C - ковариационная матрица случайных величин

$$X_1 \,,\ \ldots ,\ X_n$$

и

$$A = (a_1 \,,\ \ldots ,\ a_n \,)^T \text{ -}$$

произвольный ненулевой вектор констант. Тогда для случайной величины

$$Y = \ a_1 X_1 + \ \ldots + \ a_n X_n$$

выполняется соотношение

$$\mathbf{Var}(Y) = A^T C A \geq 0,$$

которое служит **матричной формой** записи формулы (*).

Заметим, что условие $\mathbf{Var}(Y) = 0$ равносильно равенству $CA = 0$, означающему вырожденность матрицы C.

Системы случайных величин

Пример. Найти ковариационную матрицу системы случайных величин (X, Y), если плотность вероятности

$$f(x, y) = \frac{2}{\pi(x^2 + y^2 + 1)^3}.$$

Решение.

Часть 2. Математическая статистика

Оценки параметров
Выборочный метод

Основные задачи математической статистики.

Выборка, эмпирическое распределение и эмпирическая функция распределения, выборочные характеристики, вариационный ряд. Гистограмма.

Сходимость выборочных характеристик к теоретическим.

Моделирование выборок и сравнение выборочных характеристик с теоретическими.

Использование электронных таблиц и библиотек **numpy**, **scipy.stats**, **matplotlib**, **statsmodels** для моделирования выборок, вычисления выборочных и теоретических характеристик и построения эмпирической функции распределения и гистограмм.

Раздел 1. Оценки параметров

Слово «**статистика**» содержится в названии целого ряда дисциплин.

Например,

1) математическая статистика;
2) социальная статистика;
3) экономическая статистика;
4) региональная статистика;
5) статистика отраслей и т.п.

Раздел 1. Оценки параметров

Математической статистикой называется наука, которая, основываясь на методах теории вероятностей, занимается систематизацией, обработкой и использованием статистических данных (то есть результатов наблюдений) для получения научных и практических выводов.

Особая роль математической статистики состоит в том, что она снабжает прочие статистические дисциплины количественными методами.

Раздел 1. Оценки параметров

Различие между теорией вероятностей и математической статистикой заключается в том, что **типичная задача теории вероятностей** – по известным вероятностям простых случайных событий вычислить вероятность более сложного события.

Типичная задача математической статистики – на основании результатов наблюдений оценить вероятность случайного события или характеристики случайной величины.

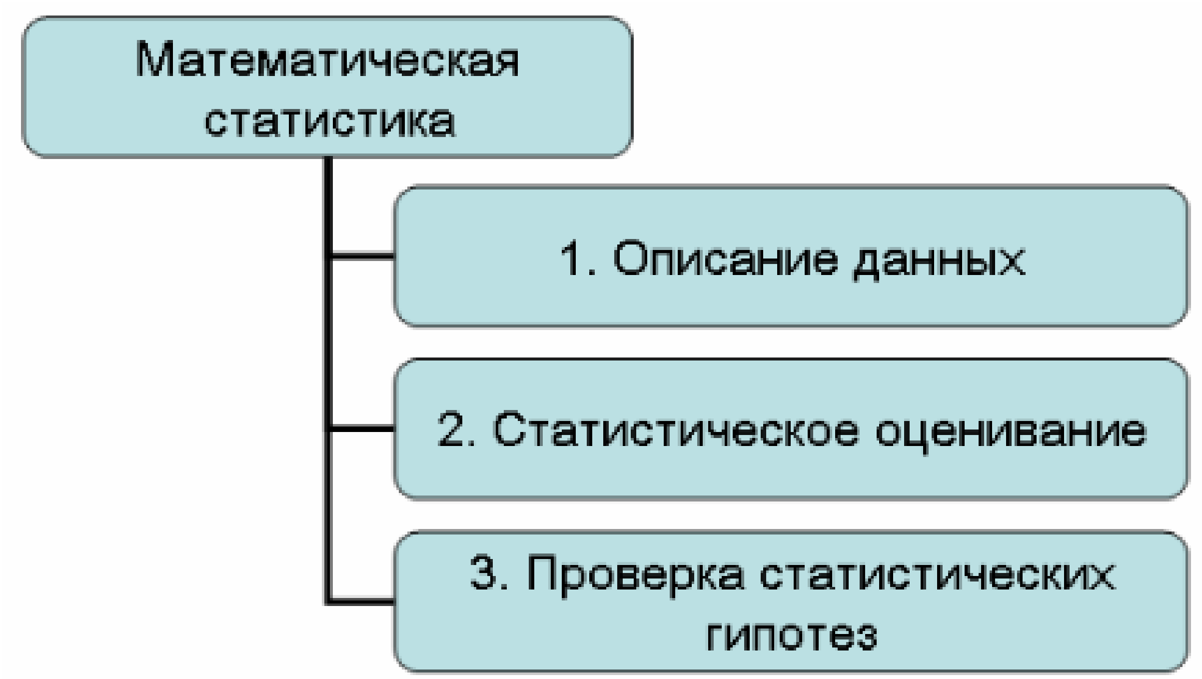
Раздел 1. Оценки параметров

Основные задачи математической статистики состоят в разработке методов:

1. Организации и планирования статистических наблюдений, в том числе способов определения числа необходимых испытаний (планирование эксперимента).
2. Сбора статистических данных.
3. «Свертки информации», то есть методов группировки данных и сведения большого числа данных к небольшому числу параметров.
4. Анализа статистических данных.
5. Принятия решений, рекомендаций и выводов на основе анализа статистических данных.
6. Прогнозирования случайных явлений.

Раздел 1. Оценки параметров

Разделы математической статистики



Раздел 1. Оценки параметров

Совокупностью (статистической совокупностью) называется множество изучаемых объектов, а число ее элементов – **объемом**.

Статистическая совокупность будет обозначаться символом Ω.

Раздел 1. Оценки параметров

Примеры из социальной статистики:

1) совокупность жителей Москвы;
2) совокупность семей в РФ.

Примеры из экономической статистики:

1) совокупность компаний, основная деятельность которых связана с Центральным федеральным округом.

Раздел 1. Оценки параметров

В математической статистике **признак** – то же самое, что функция, но без явной привязки к некоторой области определения — это фиксируемое свойство исследуемого объекта.

Признак – это качество, свойство, типичность всех единиц совокупности.

Признаки, как и с.в., обозначаются буквами X, Y, Z.

Раздел 1. Оценки параметров

Например,

X – цена квартиры; Y – её площадь; Z – тип дома, в котором квартира находится.

Характеристики квартиры X и Y являются примерами **количественных (числовых) признаков**, тогда как Z – пример **качественного признака**.

Очевидно, что **признак** – это некоторое отображение $\Omega \to V$, где V – множество, содержащее все значения признака.

В частности, **числовой признак** – это функция $\Omega \to \mathbb{R}$.

Виды статистических признаков

<table>
<tr><th>Качественные признаки</th><th colspan="2">Количественные признаки</th></tr>
<tr><td rowspan="3">Признаки объекта. Они не поддаются непосредственному измерению (например, цвет одежды, национальность, образование и т.п.).</td><td colspan="2">Результаты подсчета или измерения. В соответствии с этим они делятся на:</td></tr>
<tr><td>дискретные</td><td>непрерывные</td></tr>
<tr><td>могут принимать лишь отдельные значения из некоторого ряда чисел. Например, количество человек в семье, число повторений в опыте.</td><td>могут принимать любые значения в определенном интервале. Например, температура, скорость движения и т. п.</td></tr>
</table>

Раздел 1. Оценки параметров

Заметим, что рассмотрение качественных признаков, в принципе, сводится к изучению количественных признаков путём нумерации всех возможных значений.

Пусть, **например**, имеется некоторая **совокупность** Ω людей и заданный на этой совокупности **признак** «пол»,

$$X: \Omega \to \{\text{«мужской», «женский»}\}.$$

Присвоив значению «мужской», скажем, номер 0, а значению «женский номер 1, можем заменить **качественный** признак X **количественным** признаком

$$X: \Omega \to \{0,1\}.$$

Раздел 1. Оценки параметров

Под **качественной однородностью** совокупности понимается сходство всех единиц совокупности по каким-либо существенным признакам и различие по каким-либо другим признакам.

Каждая единица совокупности обладает индивидуальными особенностями и различиями, отличающими их друг от друга, т.е. существует так называемая **вариация (колеблемость) признака**.

Вариация – это количественные изменения изучаемого признака от одной единицы совокупности к другой.

Раздел 1. Оценки параметров

Пример вариации признака

Предположим, мы изучаем оплату труда на каком либо предприятии.

Статистическая совокупность – численность работников исследуемого предприятия.

Оплата труда – это **признак**.

Единица совокупности – каждый конкретный работник.

Необходимо посмотреть наличие изучаемого признака у всех единиц совокупности: получали они зарплату или нет.

У каждого работника свой размер зарплаты.

Изменение этого размера зарплаты и есть **вариация** признака.

Раздел 1. Оценки параметров

Конечная статистическая совокупность

Раздел 1. Оценки параметров

Рассмотрим признак X, заданный на некотором множестве (статистической совокупности)

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

Например, контролируем размеры производимой на заводе детали или делаем опрос общественного мнения.

Раздел 1. Оценки параметров

Генеральной совокупностью называют совокупность результатов всех мысленно возможных наблюдений над каким либо признаком X (в том числе, и повторяющихся), проводимых в одинаковых условиях.

Выборочной совокупностью (или просто **выборкой**) называют результаты ограниченного числа наблюдений над с.в. X .

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности:

- N – объем генеральной совокупности;
- n – объем выборочной совокупности. Предполагается, что $N>>n$.

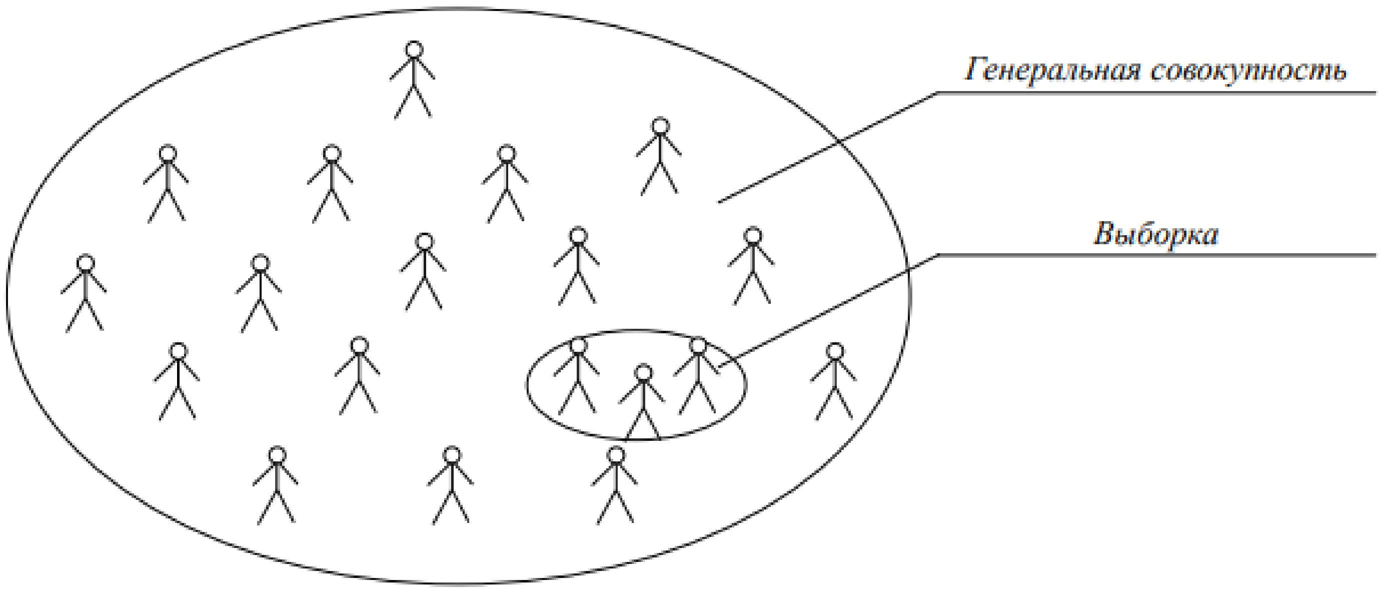
Генеральная совокупность может быть как конечной, так и бесконечной ($N=\infty$).

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

Иначе говоря, **выборка** - это часть элементов генеральной совокупности, отобранная для изучения.

Например, из 1000 деталей отобрано 100 для изучения, тогда **объём генеральной совокупности** $N = 1000$, **объём выборки** $n = 100$.



Раздел 1. Оценки параметров

Статистическое наблюдение – планомерная регистрация признаков у элементов статистической совокупности.

К статистическому наблюдению предъявляются **следующие требования**:

1) полноты и практической ценности статистических данных;

2) достоверности и точности данных;

3) их единообразия и сопоставимости.

Раздел 1. Оценки параметров

К организационным вопросам статистического наблюдения относятся:

- определение единицы наблюдения;
- установление времени и сроков его проведения;
- определение форм и видов наблюдения, а также способов регистрации фактов.

Раздел 1. Оценки параметров

Объект статистического наблюдения – это та совокупность, о которой должны быть собраны сведения (например, студенческая группа).

Единица статистического наблюдения – это элемент объекта, который характеризуется рядом признаков и относительно которого осуществляется наблюдение (например, студент).

Отчетная единица – это субъект, от которого непосредственно получают статистические сведения о единице наблюдения (например, студент, преподаватель).

Раздел 1. Оценки параметров

Проведя n раз эксперимент в одинаковых условиях, получим числа $x_1, \dots, x_n$ — значения наблюдаемой с.в. в первом, втором и т. д. экспериментах (выборка).

Выборки производятся для того, чтобы понять свойства объекта исследования.

В серии уже произведённых экспериментов **выборка** — это набор чисел.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

До того как эксперимент проведён, имеет смысл считать выборку набором случайных величин (независимых и распределённых так же, как X).

Действительно, до проведения опытов мы не можем сказать, какие значения примут элементы выборки: это будут какие-то из значений X.

Поэтому имеет смысл считать, что до опыта X_k — с.в., одинаково распределённая с X, а после опыта x_k — число которое мы наблюдаем в k–ом по счёту эксперименте, т.е. одно из возможных значений с.в. X_k.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

При составлении выборки после того, как объект отобран и над ним произведено наблюдение, он может быть возвращен в генеральную совокупность или не возвращен.

В связи с этим различают **повторные** и **бесповторные** выборки.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

Повторной случайной называется выборка, при которой отобранный объект возвращается в генеральную совокупность перед извлечением следующего объекта.

Бесповторной случайной называется выборка, при которой отобранный объект не возвращается в генеральную совокупность.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

В **повторной случайной выборке** наблюдения $X_1, X_2, \ldots, X_n$ независимы и проводятся в одинаковых (с вероятностной точки зрения) условиях, т. е. распределены по одному и тому же закону:

$$P(X_k \leq x) = P(X \leq x),$$

или

$$F_{Xk}(x)=F_X(x),\ i=1,2,\ldots,n,\ \forall x \in \mathbb{R}.$$

В дальнейшем будем ассоциировать генеральную совокупность с функцией распределения $F(x)$ элемента выборки.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \ldots, \omega_n\}$.

Раздел 1. Оценки параметров

В <u>**бесповторной случайной выборке**</u> наблюдения $X_1, X_2, \dots,$ X_n являются **з а в и с и м ы м и**, что может привести к усложнению статистических вычислений.

Однако <u>при большом объеме генеральной совокупности</u> N и относительно небольшом (по сравнению с объемом генеральной совокупности) объеме бесповторной случайной выборки $n << N$ различиями между бесповторной и повторной случайными выборками <u>можно пренебречь</u>.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

Конкретной выборкой называется конкретный набор чисел $x_1,\ x_2,\ \dots\ ,\ x_n$, полученный в результате наблюдений над признаком X , т.е. набор, состоящий из n **р е а л и з а ц и й** X .

Напомним, что число элементов в выборке называется ее ***объемом***.

Признак X — фиксируемое свойство исследуемого объекта, задан на некотором множестве (статистической совокупности) $\Omega = \{\omega_1, \dots, \omega_n\}$.

Раздел 1. Оценки параметров

Выборочное наблюдение - частичный отбор из статистической совокупности ее отдельных единиц, формируемых в выборку, главным образом случайным способом

(например, социологический опрос прохожих, выборочная проверка качества продукции, микро-переписи населения).

Раздел 1. Оценки параметров

Под **выборочным методом** понимается обследование части совокупности (выборочной совокупности), после чего, на основании полученных результатов, делаются выводы относительно всей совокупности (генеральной совокупности).

Из генеральной совокупности отбирается часть единиц.

По ним проводится исследование, а затем результаты обследования распространяются на всю совокупность с достаточно высокой степенью достоверности, вероятности.

Раздел 1. Оценки параметров

Причины применения:

- Экономия
- Невозможность проведения сплошного исследования

Основная идея выборочного метода состоит в том, что в результате обследования части совокупности можно судить с определенной вероятностью о характеристиках всей изучаемой совокупности (генеральной совокупности).

Раздел 1. Оценки параметров

Задачи выборочного метода

- Определение доверительного интервала, в котором находится характеристика генеральной совокупности
- Определение минимального объема выборки
- Определение доверительной вероятности того, что разность между характеристиками выборочной и генеральной совокупностей не превзойдет наперед заданного числа.

Теория вероятностей и математическая статистика

Конец лекции