

ТЕОРИЯ ВЕРОЯТНОСТЕЙ и МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Часть II. От вероятности к статистике

Еще раз про законы распределения непрерывных случайных величин: равномерное, показательное, нормальное; решение типовых задач (источник засекречен)

Закон больших чисел. Центральная предельная теорема. Еще раз Пуассон, Муавр и Лаплас

Математическая статистика. Три задачи. «Затравка» из Ефимова–Поспелова

Новости экономики

Еще раз про законы распределения непрерывных случайных величин

Равномерное распределение

Непрерывная случайная величина X называется *равномерно распределенной* на отрезке $[a; b]$, если плотность вероятности постоянна на этом отрезке и равна нулю вне этого отрезка

$$f(x) = \begin{cases} 0, & \text{при } x < a, \\ c, & \text{при } a \leq x \leq b, \\ 0, & \text{при } x > b. \end{cases}$$

Из свойств функции плотности известно, что $\int_a^b f(x)dx = 1$. Тогда $\int_a^b cdx = 1$.

Следовательно, $c(b - a) = 1$ и $c = \frac{1}{b - a}$.

Таким образом, плотность $f(x)$ равномерного распределения имеет вид

$$f(x) = \begin{cases} 0, & \text{при } x < a, \\ \frac{1}{b - a}, & \text{при } a \leq x \leq b, \\ 0, & \text{при } x > b. \end{cases}$$

Восстановим функцию распределения $F(x)$ равномерно распределенной случайной величины. Для этого воспользуемся формулой

$$F(x) = \int_{-\infty}^x f(x) dx.$$

Тогда:

$$\text{— при } x < a, \quad F(x) = \int_{-\infty}^x 0 dx = 0;$$

$$\text{— при } a \leq x \leq b, \quad F(x) = \int_{-\infty}^a 0 dx + \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a};$$

$$\text{— при } x > b, \quad F(x) = \int_{-\infty}^a 0 dx + \int_a^b \frac{1}{b-a} dx + \int_b^x 0 dx = \frac{b-a}{b-a} = 1.$$

Таким образом, функция распределения примет вид:

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

Изобразим графики обеих функций (рис.2.5).

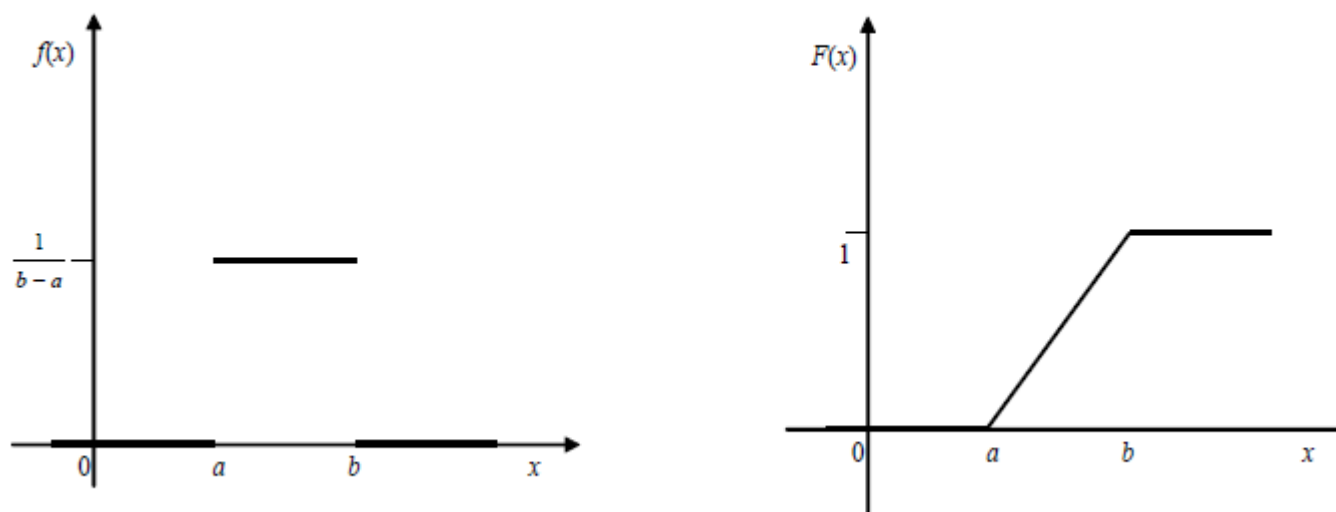


Рис. 2.5.

Теорема. Для вычисления $M(X)$ и $D(X)$ равномерно распределенной случайной величины имеют место равенства

$$M(X) = \frac{a+b}{2}; \quad D(X) = \frac{(b-a)^2}{12}.$$

Доказательство. Согласно определению $M(X) = \int_{-\infty}^{\infty} xf(x)dx$. Тогда

$$M(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Согласно определению $D(X) = M(X^2) - [M(X)]^2$. Тогда

$$\begin{aligned} D(X) &= \int_a^b x^2 \cdot \frac{1}{b-a} dx - \frac{(a+b)^2}{4} = \frac{x^3}{3(b-a)} \Big|_a^b - \frac{(a+b)^2}{4} = \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Очевидно, что $\sigma(X) = \frac{b-a}{2\sqrt{3}}$.

Теорема. Пусть непрерывная случайная величина распределена равномерно на отрезке $[a; b]$, и отрезок $[c; d] \subset [a; b]$, тогда

$$P(c < X < d) = \frac{d - c}{b - a}.$$

Доказательство. Согласно свойствам функции распределения

$$P(c < X < d) = F(d) - F(c) = \frac{d - a}{b - a} - \frac{c - a}{b - a} = \frac{d - c}{b - a}.$$

Таким образом, для того, чтобы полностью описать непрерывную случайную величину, имеющую равномерное распределение, достаточно знать концы отрезка, которому принадлежат все возможные значения этой случайной величины.

Показательное распределение

Показательным (экспоненциальным) называют распределение вероятностей случайной величины X , которое описывается плотностью

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ \lambda e^{-\lambda x} & \text{при } x \geq 0, \end{cases}$$

где λ - постоянная положительная величина.

Показательное распределение определяется одним параметром λ . Эта особенность показательного распределения указывает на его преимущество по сравнению с распределениями, зависящими от большего числа параметров. Обычно параметры неизвестны и приходится находить их оценки (приближенные значения); разумеется, проще оценить один параметр, чем два или три и т.д.

Найдем функцию распределения показательного закона. Воспользуемся формулой

$$F(x) = \int_{-\infty}^x f(x) dx. \text{ Тогда:}$$

$$\text{- при } x < 0, \quad F(x) = \int_{-\infty}^x 0 dx = 0;$$

$$\text{- при } x \geq 0, \quad F(x) = \int_{-\infty}^0 0 dx + \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}.$$

Итак,

$$F(x) = \begin{cases} 0, & \text{при } x < 0; \\ 1 - e^{-\lambda x}, & \text{при } x \geq 0. \end{cases}$$

Графики плотности и функции распределения показательного закона изображены на рис. 2.6.

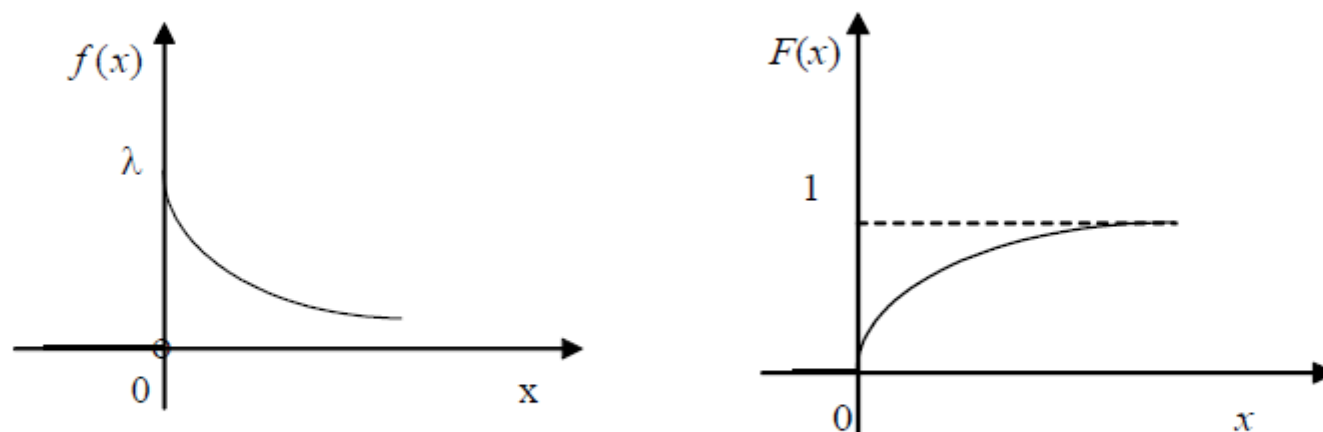


Рис. 2.6.

Теорема. Для $M(X)$ и $D(X)$ показательного распределения справедливы равенства

$$M(X) = \frac{1}{\lambda}; \quad D(X) = \frac{1}{\lambda^2}.$$

Доказательство. Согласно определению $M(X) = \int_{-\infty}^{\infty} xf(x)dx$. Тогда,

$$\begin{aligned} M(X) &= \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} x \cdot e^{-\lambda x} dx = \left| \begin{array}{l} u = x; \quad du = dx \\ dv = e^{-\lambda x}; \quad V = -\frac{1}{\lambda} e^{-\lambda x} \end{array} \right| = \\ &= \lambda \left(-\frac{x}{\lambda} e^{-\lambda x} \Big|_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \right) = \lambda \left(-\frac{1}{\lambda^2} \cdot e^{-\lambda x} \right) \Big|_0^{\infty} = -\frac{1}{\lambda} (0 - 1) = \frac{1}{\lambda}. \end{aligned}$$

Соотношение $D(X) = \frac{1}{\lambda^2}$ предлагается доказать самостоятельно.

Очевидно, что $\sigma(X) = \frac{1}{\lambda}$, т.е. математическое ожидание и среднее квадратическое отклонение показательного распределения равны между собой.

Теорема. Вероятность попадания в интервал $(a; b)$ показательно распределенной случайной величины вычисляется по формуле

$$P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}.$$

Доказательство. Используем формулу

$$P(a < X < b) = F(b) - F(a).$$

Учитывая, что $F(a) = 1 - e^{-\lambda a}$, $F(b) = 1 - e^{-\lambda b}$, получим

$$P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}.$$

Нормальное (гауссовское) распределение

Нормальным называют распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Данное распределение определяется двумя параметрами: a и σ , достаточно знать эти параметры, чтобы задать нормальное распределение.

График плотности нормального распределения называют *нормальной кривой (кривой Гаусса)*.

Исследуем и построим график функции

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

1. Очевидно, функция определена на всей оси абсцисс.
2. При всех значениях x функция принимает положительные значения, т.е. нормальная кривая расположена над осью Ox .
3. Предел функции при неограниченном возрастании x (по абсолютной величине) равен нулю:
 $\lim_{|x| \rightarrow \infty} f(x) = 0$, т.е. ось Ox служит горизонтальной асимптотой графика.

4. Исследуем функцию на экстремум. Найдем первую производную:

$$f'(x) = -\frac{x-a}{\sigma^3 \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Очевидно, что:

– при $x = a$, $f'(x) = 0$;

– при $x < a$, $f'(x) > 0$;

– при $x > a$, $f'(x) < 0$.

Следовательно:

– при $x \in (-\infty; a)$, $f(x)$ - возрастает;

– при $x \in (a; \infty)$, $f(x)$ - убывает;

– при $(x = a)$ функция имеет максимум $f(a) = \frac{1}{\sigma\sqrt{2\pi}}$.

5. Разность $(x - a)$ содержится в аналитическом выражении функции в квадрате, т.е. график функции симметричен относительно прямой $x = a$.
6. Исследуем функцию на точки перегиба. Найдем вторую производную:

$$y'' = -\frac{1}{\sigma^3 \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \left[1 - \frac{(x-a)^2}{\sigma^2} \right].$$

Легко видеть, что при $x = a + \sigma$ и $x = a - \sigma$ вторая производная равна нулю, а при переходе через эти точки она меняет знак (в обеих этих точках значение функции равно $\frac{1}{\sigma\sqrt{2\pi e}}$). Таким

образом, точки графика $\left(a - \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ и $\left(a + \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ являются точками перегиба. На рис.

2.7 изображена нормальная кривая

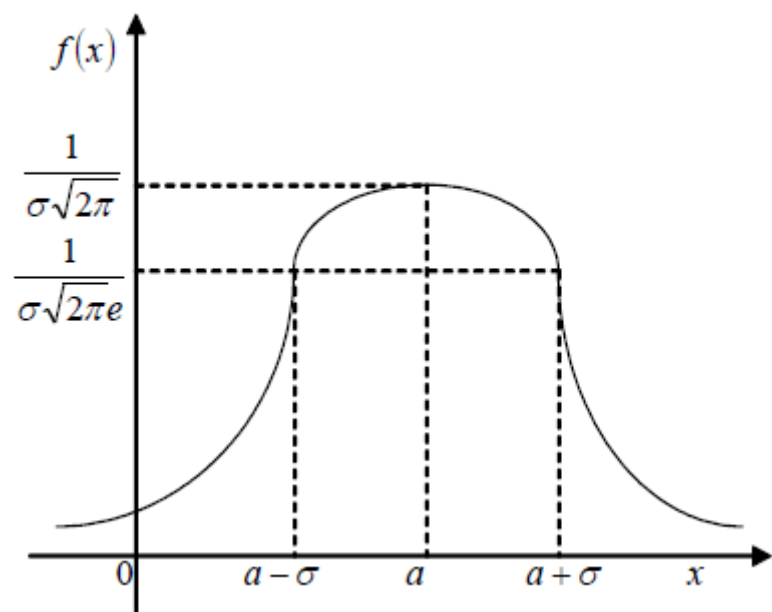


Рис. 2.7.

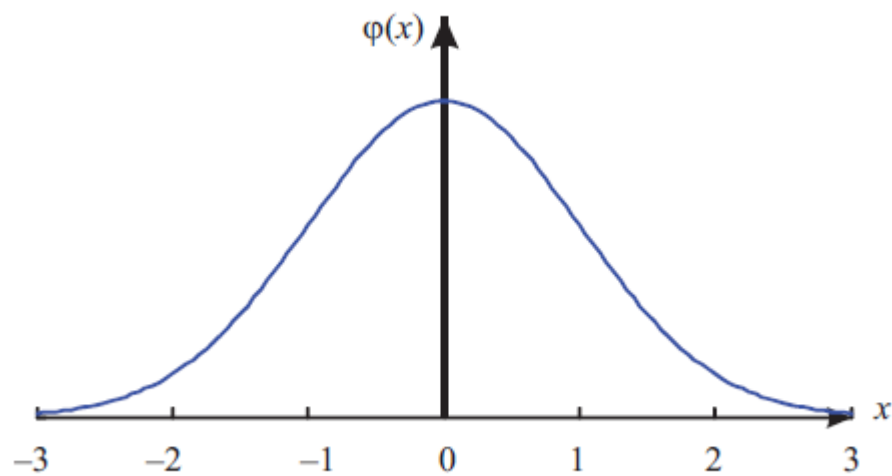


Рис. 1.14 из [1]

Влияние параметров нормального распределения a и σ на форму нормальной кривой

Известно, что графики функций $f(x)$ и $f(x-a)$ имеют одинаковую форму; сдвинув график $f(x)$ в положительном направлении оси x на a единиц масштаба при $a > 0$ или в отрицательном направлении при $a < 0$, получим график $f(x-a)$. Отсюда следует, что изменение величины параметра a не изменяет формы нормальной кривой, а приводит лишь к ее сдвигу вдоль оси Ox : вправо, если a возрастает, и влево, если a убывает.

Рассмотрим форму кривой при изменении параметра σ . Как было указано выше, максимум нормальной кривой равен $\frac{1}{\sigma\sqrt{2\pi}}$. Отсюда следует, что с возрастанием σ максимальная ордината нормальной кривой убывает, а сама кривая становится более пологой, т.е. сжимается к оси Ox ; при убывании σ нормальная кривая становится более «островершинной» и растягивается в положительном направлении оси Oy .

Заметим, что при любых значениях параметров a и σ площадь, ограниченная нормальной кривой и осью x , остается равной единице (третье свойство плотности распределения).

На рис. 2.8 изображены нормальные кривые при различных значениях σ и $a = 0$. Чертеж наглядно иллюстрирует, как изменение параметра σ сказывается на форме нормальной кривой.

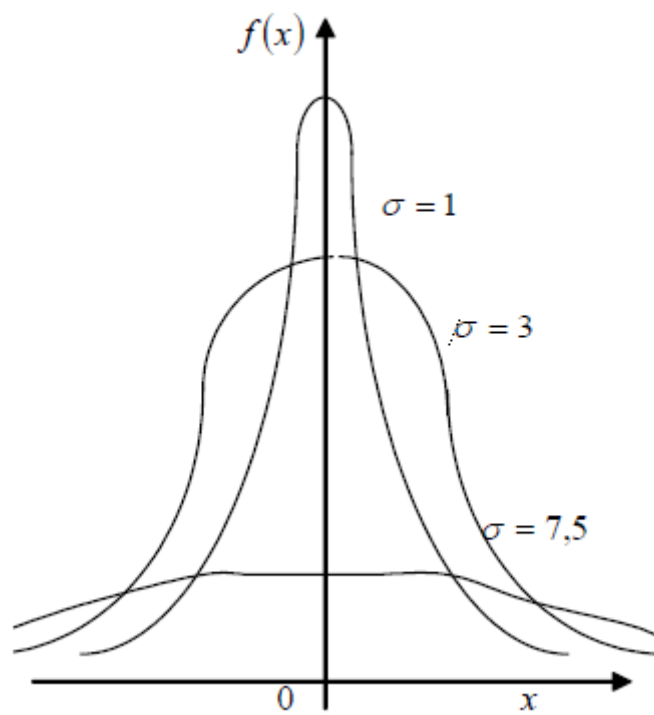


Рис. 2.8.

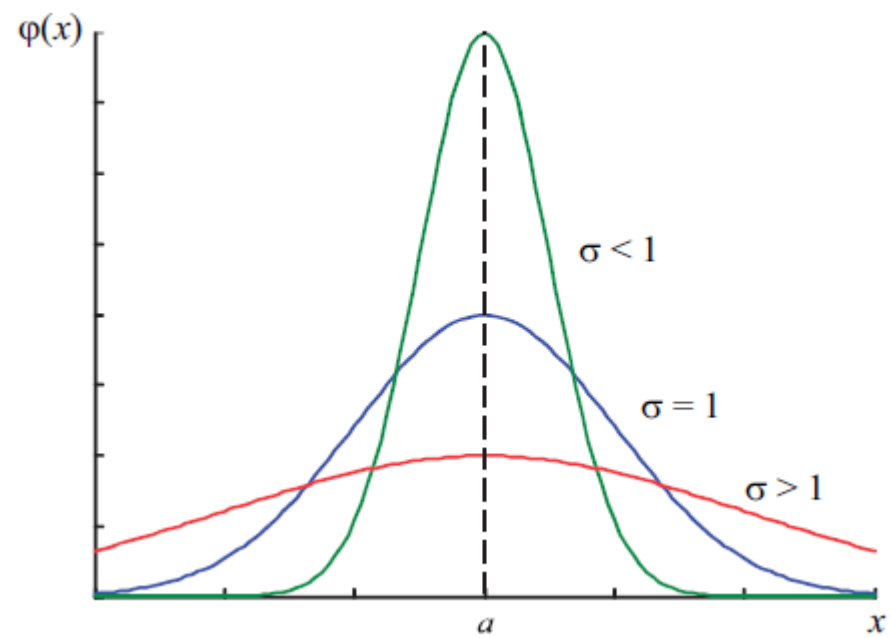


Рис. 1.15 из [1]

Замечание 1. *Общим* называют нормальное распределение с произвольными параметрами a и σ ($\sigma > 0$).

Нормированным (стандартным) называют нормальное распределение с параметрами $a = 0$ и $\sigma = 1$.

Плотность нормированного распределения

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Эта функция табулирована (приложение 1).

Замечание 2. Функция $F(x)$ общего нормального распределения определяется равенством

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

а функция нормированного распределения

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

Легко проверить, что $F(x) = F_0\left(\frac{x-a}{\sigma}\right)$.

Замечание 3. Существует связь между функциями общего и нормированного распределения $F(x)$, $F_0(x)$ и функцией Лапласа $\Phi(x)$:

$$F(x) = 0,5 + \Phi\left(\frac{x-a}{\sigma}\right);$$

$$F_0(x) = 0,5 + \Phi(x).$$

Данные равенства позволяют вычислять значения $F(x)$ и $F_0(x)$ по таблице значений $\Phi(x)$ (приложение 3).

Замечание к замечанию 3. В [9] = [Фаллеб] Функция Лапласа обозначается не Φ , а Φ_0 .

Вероятностный смысл параметров a и σ

Теорема. Для нормально распределенной случайной величины справедливы равенства:

$$M(X) = a; D(X) = \sigma^2$$

Доказательство. 1. По определению математического ожидания непрерывной случайной величины,

$$\begin{aligned} M(X) &= \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left| \begin{array}{l} x = \sigma z + a \\ z = \frac{x-a}{\sigma} \quad dx = \sigma dz \end{array} \right| = \\ &= \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + a) e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma z e^{-\frac{z^2}{2}} dz + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz. \end{aligned}$$

Первое из слагаемых равно нулю (под знаком интеграла нечетная функция; пределы интегрирования симметричны относительно начала координат). Второе из слагаемых равно a

$$\left(\text{интеграл Пуассона } \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi} \right).$$

Итак, $M(X) = a$, т.е. математическое ожидание нормального распределения равно параметру a .

2. По определению дисперсии непрерывной случайной величины, учитывая, что $M(X) = a$, имеем

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

$$= \left| \begin{array}{l} x-a = \sigma z \\ z = \frac{x-a}{\sigma} \quad dx = \sigma dz \end{array} \right| = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \cdot z e^{-\frac{z^2}{2}} dz.$$

Интегрируя по частям, положив $u = z$, $dv = z e^{-\frac{z^2}{2}} dz$, найдем

$$D(X) = \sigma^2.$$

Следовательно,

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\sigma^2} = \sigma.$$

Итак, среднее квадратическое отклонение нормального распределения равно параметру σ .

Вероятность попадания в интервал нормально распределенной сл. в.

Теорема. Вероятность того, что непрерывная случайная величина, распределенная по нормальному закону, примет значение на интервале (α, β) , равна

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right),$$

где $\Phi(x)$ – функция Лапласа.

Доказательство. Из свойств функции распределения известно, что $P(\alpha < X < \beta) = F(\beta) - F(\alpha)$. Так как случайная величина распределена нормально, то ее функция распределения связана с функцией Лапласа равенством $F(x) = 0,5 + \Phi\left(\frac{x - a}{\sigma}\right)$ (замечание 3).

Таким образом, искомая вероятность

$$P(\alpha < X < \beta) = 0,5 + \Phi\left(\frac{\beta - a}{\sigma}\right) - 0,5 - \Phi\left(\frac{\alpha - a}{\sigma}\right) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Вычисление вероятности отклонения нормально распределенной сл. в. X от $a = M(X)$

Теорема. Вероятность того, что отклонение нормально распределенной случайной величины по абсолютной величине меньше положительного числа δ , находится из соотношения

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Доказательство. Заменяем неравенство $|X - a| < \delta$ равносильным ему двойным неравенством

$$-\delta < X - a < \delta, \text{ или } a - \delta < X < a + \delta.$$

Пользуясь предыдущей теоремой

$$\begin{aligned} P(|X - a| < \delta) &= P(a - \delta < X < a + \delta) = \\ &= \Phi\left[\frac{(a + \delta) - a}{\sigma}\right] - \Phi\left[\frac{(a - \delta) - a}{\sigma}\right] = \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right), \end{aligned}$$

приняв во внимание равенство

$$\Phi\left(-\frac{\delta}{\sigma}\right) = -\Phi\left(\frac{\delta}{\sigma}\right)$$

(функция Лапласа – нечетная), окончательно имеем

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Правило трех сигм

Преобразуем формулу

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right),$$

положив $\delta = \sigma t$. В итоге получим

$$P(|X - a| < \sigma t) = 2\Phi(t).$$

Если $t=3$ и, следовательно, $\sigma t = 3\sigma$, то

$$P(|X - a| < 3\sigma) = 2\Phi(3) \approx 2 \cdot 0,499 = 0,998,$$

т.е. вероятность того, что отклонение по абсолютной величине будет меньше утроенного среднего квадратического отклонения, равна 0,998.

Другими словами, вероятность того, что абсолютная величина отклонения превысит утроенное среднее квадратическое отклонение, очень мала, а именно равна 0,002 7. Это означает, что лишь в 0,27 % случаев так может произойти. Такие события исходя из принципа невозможности маловероятных событий можно считать практически невозможными. В этом и состоит сущность правила трех сигм: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основание предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально.

Вопрос: прокомментируйте слова «в противном случае она не распределена нормально».

Решение типовых задач

Задача 1. Поезда метро идут строго по расписанию. Интервал движения – 5 минут. Составить $f(x)$ и $F(x)$ случайной величины X – времени ожидания очередного поезда и построить их графики. Найти $M(X)$, $D(X)$.

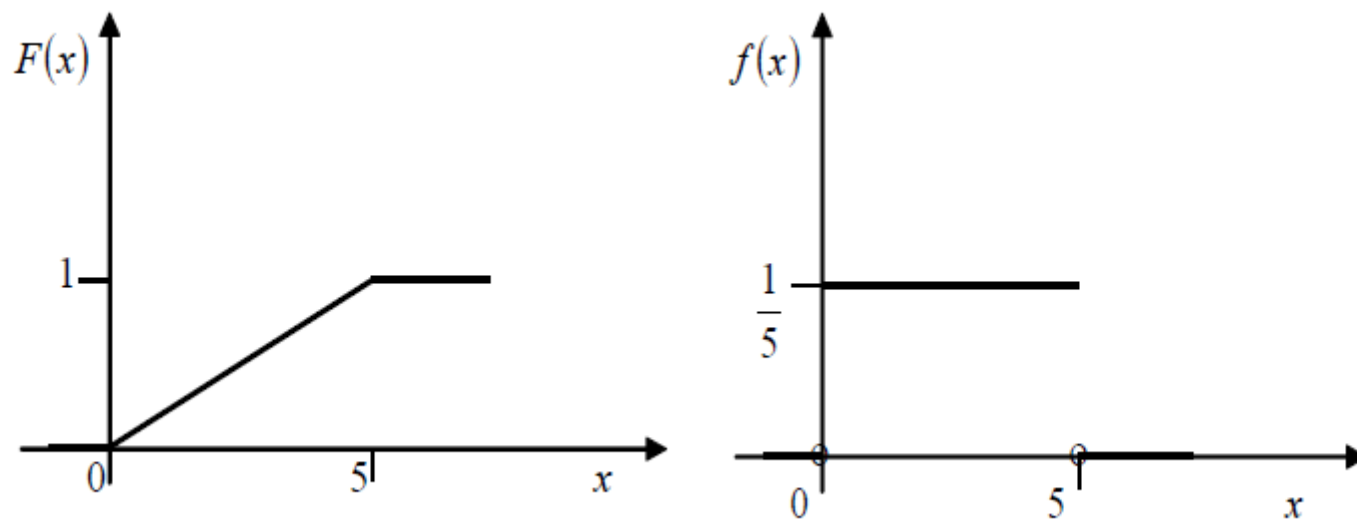
Это тот случай, когда распределение не назначается, а выводится из условий задачи.

Решение. Случайная величина X – время ожидания очередного поезда. Величина X распределена равномерно на отрезке $[0,5]$, поэтому воспользуемся формулами

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

Тогда имеем

$$F(x) = \begin{cases} 0, & x < 0; \\ \frac{x}{5}, & 0 \leq x \leq 5; \\ 1, & x > 5. \end{cases} \quad f(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{5}, & 0 \leq x \leq 5; \\ 0, & x > 5. \end{cases}$$



Математическое ожидание и дисперсия вычисляются по формулам:

$$M(X) = \frac{b+a}{2}; \quad D(X) = \frac{(b-a)^2}{12}.$$

Тогда

$$M(X) = \frac{5+0}{2} = 2,5; \quad D(X) = \frac{(5-0)^2}{12} = \frac{25}{12} \approx 2,08.$$

Задача 2. Непрерывная случайная величина распределена по показательному закону с параметром $\lambda = 2$. Составить функцию распределения, функцию плотности этой случайной величины. Найти числовые характеристики и вероятность того, что случайная величина попадет в интервал $(0,3;1)$.

Решение. Очевидно, искомая плотность распределения

$$f(x) = \begin{cases} 0, & \text{при } x < 0; \\ 2e^{-2x}, & \text{при } x \geq 0. \end{cases}$$

Искомая функция распределения

$$F(x) = \begin{cases} 0, & \text{при } x < 0; \\ 1 - e^{-2x}, & \text{при } x \geq 0. \end{cases}$$

По условию $\lambda = 2$. Следовательно,

$$M(X) = \sigma(x) = \frac{1}{\lambda} = \frac{1}{2} = 0,5; \quad D(X) = \frac{1}{\lambda^2} = \frac{1}{2^2} = 0,25.$$

Для нахождения вероятности $P(0,3 < X < 1)$ воспользуемся формулой $P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}$.

$$\text{Тогда, } P(0,3 < X < 1) = e^{-(2 \cdot 0,3)} - e^{-(2 \cdot 1)} = e^{-0,6} - e^{-2} \approx 0,549 - 0,135 = 0,414.$$

Задача 3. Детали, выпускаемые цехом, по размеру диаметра распределены по нормальному закону. Стандартная длина диаметра детали равна $a=35$, среднее квадратическое отклонение $\sigma = 4$. Требуется:

- а) составить функцию плотности вероятностей;
- б) найти вероятность того, что диаметр наудачу взятой детали будет больше $\alpha = 34$ и меньше $\beta = 40$;
- в) найти вероятность того, что диаметр детали отклонится от стандартной длины не больше чем на $\delta = 2$.

Решение. 1. Так как непрерывная случайная величина X распределена по нормальному закону, есть ее плотность распределения вероятностей выражается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Следовательно,

$$f(x) = \frac{1}{4\sqrt{2\pi}} \cdot e^{-\frac{(x-35)^2}{32}}.$$

2. Для нормально распределенной случайной величины

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Тогда

$$P(34 < X < 40) = \Phi\left(\frac{40 - 35}{4}\right) - \Phi\left(\frac{34 - 35}{4}\right) = \Phi(1,25) + \Phi(0,25) = 0,394\,4 + 0,098\,7 = 0,493\,1.$$

Последнее задание решаем по формуле

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Таким образом,

$$P(|X - 35| < 2) = 2\Phi\left(\frac{2}{4}\right) = 2\Phi(0,5) = 0,3829,$$

где $\Phi(x)$ – интегральная функция Лапласа (приложение 3).

Напоминание. Помним о Φ и Φ_0 .

Закон больших чисел. Центральная предельная теорема

Неравенства Маркова и Чебышева

Теорема 1 (неравенство Маркова). Пусть для случайной величины ξ существует $M|\xi|$, тогда для любого $\varepsilon > 0$ верно:

$$P(|\xi| \geq \varepsilon) \leq \frac{M|\xi|}{\varepsilon}.$$

Следствие 1. Пусть для случайной величины ξ существует $M\xi^2$, тогда для любого $\varepsilon > 0$ верно:

$$P(|\xi| \geq \varepsilon) \leq \frac{M\xi^2}{\varepsilon^2}$$

Доказательство. Заметим, что событие $\{\omega: |\xi| \geq \varepsilon\}$ равносильно событию $\{\omega: \xi^2 \geq \varepsilon^2\}$ и применим неравенство Маркова к случайной величине ξ^2 .

Теорема 2 (неравенство Чебышева). Пусть для случайной величины ξ существует математическое ожидание $M\xi$ и дисперсия $D\xi$, тогда для любого $\varepsilon > 0$ верно:

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}.$$

Доказательство. Применяем следствие неравенства Маркова к случайной величине $\xi - M\xi$, замечая, что $M(\xi - M\xi)^2 = D\xi$.

Основной источник – Фаллеб [9], вспомогательный – [11].

Важное замечание: что открывают нам математическое ожидание и дисперсия

Неравенство Чебышева часто используют в виде:

$$P(|\xi - M\xi| < \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2}$$

Упр. Показать, что неравенства $P(|\xi - M\xi| \geq x) \leq \frac{D\xi}{x^2}$ и $P(|\xi - M\xi| < x) \geq 1 - \frac{D\xi}{x^2}$ эквивалентны.

Важное замечание. Еще раз выпишем неравенство Чебышева:

$$P\{|\xi - M\xi| \geq x\} \leq \frac{D\xi}{x^2}. \quad (32)$$

Неравенство Чебышева (32) показывает, что при малой дисперсии $D\xi$ с вероятностью, близкой к 1, случайная величина ξ концентрируется около математического ожидания $M\xi$:

$$P\{|\xi - M\xi| < x\} \geq 1 - \frac{D\xi}{x^2}. \quad (33)$$

Обобщение неравенства Чебышева на средние арифметические:

Теорема 3. Пусть для случайных величин $\xi_1, \xi_2, \dots, \xi_n$ существуют математические ожидания $M\xi_i$ и дисперсии $D\xi_i$, тогда для любого $\varepsilon > 0$ верно:

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i\right| < \varepsilon\right) \geq 1 - \frac{1}{n^2 \varepsilon^2} \sum_{i=1}^n D\xi_i.$$

Задача 1. В 400 испытаниях Бернулли вероятность успеха в каждом испытании равна 0,8. С помощью неравенства Чебышева оценить вероятность того, что разница между числом успехов в этих испытаниях и средним числом успехов будет меньше 20.

Решение. Число успехов распределено по закону Бернулли, поэтому среднее число успехов равно $M\xi = np = 400 \times 0,8 = 320$, а дисперсия $D\xi = npq = 400 \times 0,8 \times 0,2 = 64$. Тогда в силу неравенства Чебышева имеем:

$$P(|\xi - 320| < 20) \geq 1 - \frac{D\xi}{20^2} = 1 - \frac{64}{400} = 0,84.$$

Вычислим эту же вероятность с помощью приближенной (интегральной) формулы Муавра–Лапласа (см. § 4.3.2):

$$\begin{aligned} P(|\xi - 320| < 20) &= P(|\xi - np| < k) = P\left(-\frac{k}{\sqrt{npq}} < \frac{\xi - np}{\sqrt{npq}} < \frac{k}{\sqrt{npq}}\right) \approx \\ &\approx 2\Phi_0\left(\frac{k}{\sqrt{npq}}\right) = 2\Phi_0\left(\frac{20}{\sqrt{64}}\right) = 2\Phi_0(2,5) = 2 \cdot 0,4938 = 0,9876. \end{aligned}$$

Последнее вычисление показывает, что неравенство Чебышева дает довольно грубые оценки вероятностей.

Вопрос. Эта грубость относится только к данной задаче или ее можно доказать в общем случае?

Упр. Попытайтесь доказать.

Созвучно:

4. Пусть случайная величина ξ имеет нормальное распределение с параметрами $M\xi=a$, $D\xi=\sigma^2$. Найти вероятности $P(|\xi - a| \geq \sigma)$ и $P(|\xi - a| \geq 3\sigma)$, пользуясь таблицами функции Лапласа. Затем оценить вероятности с помощью неравенства Чебышева.
5. Пусть случайная величина ξ имеет распределение Лапласа, т.е. ее плотность равна $p(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \lambda > 0$. Найти вероятности $P(|\xi| < \sigma)$ и $P(|\xi| < 3\sigma)$, где σ – среднее квадратическое отклонение, и сравнить их с оценками, получаемыми с помощью неравенства Чебышева.

Закон больших чисел

Теорема 4 (закон больших чисел). Пусть задана бесконечная последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, для которых существуют математическое ожидание $M\xi_i = a$ и дисперсия $D\xi_i = \sigma^2$. Тогда для любого $\varepsilon > 0$ верно:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right) = 0.$$

Перейдем в книгу Б.А. Севастьянова [11]. Теорема 4 [9] \rightarrow теорема 7 [11]:

Теорема 7 (Теорема Чебышева). Если ξ_1, ξ_2, \dots независимы и существует такая константа $c > 0$, что $D\xi_n \leq c$, $n = 1, 2, \dots$, то при любом $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n} - \frac{M\xi_1 + \dots + M\xi_n}{n}\right| < \varepsilon\right\} = 1. \quad (34)$$

Доказательство. Обозначим $\zeta_n = \xi_1 + \dots + \xi_n$ и применим к ζ_n/n неравенство (33). Имеем при любом $x > 0$:

$$1 \geq P\left\{\left|\frac{\zeta_n}{n} - \frac{M\zeta_n}{n}\right| < x\right\} \geq 1 - \frac{D\zeta_n}{x^2 n^2} \geq 1 - \frac{c}{x^2 n}, \quad (35)$$

так как $D\zeta_n = \sum_{k=1}^n D\xi_k \leq nc$ (см. теорему 4). Из (35) при $n \rightarrow \infty$ имеем (34).

Следствие. Если ξ_1, ξ_2, \dots независимы и одинаково распределены, $M\xi_n = a$, $D\xi_n = \sigma^2 < \infty$, то при любом $x > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| < x \right\} = 1. \quad (36)$$

Предельные утверждения типа (34) и (36) носят название закона больших чисел. Закон больших чисел утверждает, что с вероятностью, приближающейся при $n \rightarrow \infty$ к 1, среднее арифметическое сумм независимых слагаемых при определенных условиях становится близким к константе.

Из (36) получаем закон больших чисел в схеме Бернулли.

Теорема 8. (Теорема Бернулли.) Пусть μ_n — число успехов при n испытаниях в схеме Бернулли с вероятностью $0 < p < 1$ в каждом испытании. Тогда при любом $x > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\mu_n}{n} - p \right| \leq x \right\} = 1. \quad (37)$$

Вспомним I лекцию:

Как выглядит тамошний *Мотивирующий пример* с таблицей в контексте приведенной теоремы 8?

О соотношении (37) из теоремы Бернулли

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\mu_n}{n} - p \right| \leq x \right\} = 1. \quad (37)$$

Соотношение (37) показывает, что при больших n разность между относительной частотой μ_n/n и вероятностью успеха мала с вероятностью, близкой к 1. В условиях, когда справедливо свойство устойчивости частот, можно применять следующий принцип: *при единичном испытании маловероятное событие практически невозможно*. Считая серию в n испытаний в схеме Бернулли за единичное испытание и выбирая x таким, чтобы $\frac{D\mu_n}{x^2 n} = \frac{pq}{x^2 n}$ было мало, мы можем утверждать, что неравенство $|\mu_n/n - p| > x$ практически невозможно. Вопрос о том, какие вероятности считать малыми, зависит от конкретной прикладной задачи.

Вопрос: что такое «свойство устойчивости частот»?

Предельные теоремы в схеме Бернулли (по [11])

Биномиальное распределение

Биномиальное распределение числа успехов μ при n независимых испытаниях в схеме Бернулли с вероятностью успеха p в каждом испытании задается вероятностями

$$P\{\mu = m\} = C_n^m p^m q^{n-m}, \quad m = 0, 1, \dots, n; \quad q = 1 - p. \quad (1)$$

Формула (1) записывается достаточно компактно и просто, однако использование ее для вычисления вероятностей $P\{\mu = m\}$ при больших значениях n и m вызывает значительные трудности.

Таблицы для вероятностей $C_n^m p^m q^{n-m}$ громоздки и очень неудобны для пользования, так как содержат три входа (n , p и m). Еще хуже дело обстоит с вычислением вероятностей

$$P\{m_1 \leq \mu \leq m_2\} = \sum_{m=m_1}^{m_2} C_n^m p^m q^{n-m}, \quad (2)$$

которые зависят уже от четырех параметров: n , p , m_1 и m_2 .

ближенно вычислять вероятности (1) и (2) при больших значениях n , m , m_1 , m_2 . Такие формулы дают нам предельные теоремы.

Теорема Пуассона

Рассмотрим сначала случай больших n и малых p .

Теорема 1. (Теорема Пуассона.) Если $n \rightarrow \infty$ и $p \rightarrow 0$ так, что $np \rightarrow a$, то для любого фиксированного $m = 0, 1, \dots$

$$P\{\mu = m\} = C_n^m p^m q^{n-m} \rightarrow \frac{a^m}{m!} e^{-a}. \quad (3)$$

Локальная предельная теорема Муавра–Лапласа

Биномиальное распределение (1) случайной величины μ имеет $M\mu = np$ и $D\mu = npq$ (см. задачу 3 в гл. 3). Обозначим $\sigma = \sqrt{npq}$ среднее квадратическое отклонение. Доказываемая ниже теорема дает асимптотическую формулу для биномиальной вероятности (1) при p , не близких к 0 или 1.

Теорема 3. (Локальная предельная теорема Муавра — Лапласа.) Если в схеме Бернулли $\sigma = \sqrt{npq} \rightarrow \infty$, то для любого $C > 0$ равномерно по всем $|x| \leq C$ вида $x = \frac{m - np}{\sigma}$, где m — целые неотрицательные числа,

$$P\left\{\frac{\mu - np}{\sqrt{npq}} = x\right\} = \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/2} (1 + o(1)). \quad (13)$$

Интегральная предельная теорема Муавра—Лапласа

Для приближенного вычисления вероятностей $P\{m_1 \leq \mu \leq m_2\}$ можно применять следующую теорему.

Теорема 4. (Интегральная предельная теорема Муавра — Лапласа.) При $\sigma = \sqrt{npq} \rightarrow \infty$ равномерно по $-\infty \leq a < b \leq \infty$

$$P\left\{a \leq \frac{\mu - np}{\sqrt{npq}} \leq b\right\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \rightarrow 0. \quad (16)$$

Применения предельных теорем

Предельные теоремы Пуассона и Муавра — Лапласа применяются для приближенного вычисления вероятностей $P\{\mu = m\}$ и $P\{m_1 \leq \mu \leq m_2\}$ в схеме Бернулли при больших n . Приближение, даваемое теоремой Пуассона, называется *пуассоновским*. Приближение, получаемое с помощью теоремы Муавра — Лапласа, назы-

вается *нормальным*, так как функция $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ есть плотность нормального распределения (см. § 31). Для распределения Пуассона $\frac{a^m}{m!} e^{-a}$ и интеграла

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du, \quad (25)$$

называемого *интегралом Лапласа*, имеются таблицы.

Центральная предельная теорема для одинаково распределенных независимых слагаемых

Ранее мы доказали, что распределение числа успехов μ в схеме Бернулли при $n \rightarrow \infty$ и постоянном $0 < p < 1$ обладает следующим предельным свойством:

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\mu - M\mu}{\sqrt{D\mu}} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du. \quad (1)$$

Функцию нормального распределения будем обозначать

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du. \text{ Функция нормального распределения } \Phi(x) \text{ выражается через интеграл Лапласа}$$

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du, \text{ введенный в гл. 4, следующим}$$

образом: $\Phi(x) = \frac{1}{2} + \Phi_0(x)$. Этот результат является очень частным случаем так называемой центральной предельной теоремы. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых случайных величин. Мы будем говорить, что для этой последовательности выполнена *центральная предельная теорема*, если при любом x справедливо следующее предельное соотношение для сумм $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$:

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\zeta_n - M\zeta_n}{\sqrt{D\zeta_n}} \leq x \right\} = \Phi(x). \quad (2)$$

Так как в схеме Бернулли число успехов можно представить в виде суммы $\mu = \mu_1 + \mu_2 + \dots + \mu_n$ независимых случайных величин с $P\{\mu_i = 1\} = p$,

$P\{\mu_i = 0\} = 1 - p$, то результат (1) есть частный случай центральной предельной теоремы (2).

Для справедливости центральной предельной теоремы (2) на случайные величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ надо налагать те или иные дополнительные условия.

Мы докажем центральную предельную теорему сначала для одинаково распределенных случайных величин.

Теорема 1. Если случайные величины ξ_1, ξ_2, \dots независимы, одинаково распределены и имеют конечные $M\xi_i = a$ и $D\xi_i = \sigma^2 > 0$, то

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\xi_1 + \dots + \xi_n - na}{\sigma \sqrt{n}} \leq x \right\} = \Phi(x).$$

Для завершающего аккорда вернемся к [9].

Пример. Пусть $\xi_1, \xi_2, \dots, \xi_i, \dots$ – последовательность независимых случайных величин, описывающих результаты испытаний Бернулли (1 в случае успеха и 0 в случае неудачи). Тогда сумма $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ есть число успехов m в n испытаниях Бернулли. Из ЦПТ следует, что

$$\lim_{n \rightarrow \infty} P \left(c_1 < \frac{m - np}{\sqrt{npq}} < c_2 \right) = \frac{1}{\sqrt{2\pi}} \int_{c_1}^{c_2} e^{-\frac{x^2}{2}} dx = \Phi_0(c_2) - \Phi_0(c_1),$$

где

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{y^2}{2}} dy$$

– функция Лапласа. Этот результат называется *интегральной теоремой Муавра–Лапласа* и уже встречался раньше (см. § 4.3.2). Таким образом, ЦПТ есть обобщение интегральной теоремы Муавра–Лапласа на случай слагаемых более общего вида (не только 0 и 1).

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Если в **теории вероятностей** мы считаем заданной модель явления и производим расчет возможного реального течения этого явления, то в **математической статистике** мы исходим из известных реализаций каких-либо случайных событий, из так называемых *статистических данных*, которые обычно носят числовой характер.

Математическая статистика разрабатывает различные методы, которые позволяют по этим статистическим данным подобрать подходящую теоретико-вероятностную модель.

Пример. Пусть имеется n независимых наблюдений в схеме Бернулли и пусть в m из них произошло событие A . Модель в схеме Бернулли определяется числом испытаний n и вероятностью $p = P(A)$. Поэтому здесь мы сталкиваемся с одной из задач математической статистики:

Как по m осуществлениям события A в n независимых испытаниях определить вероятность $p = P(A)$?

Перечислим три основные задачи, которые решает математическая статистика, на примере схемы Бернулли.

Проверка гипотез

а) *Проверка статистических гипотез.* Из каких-либо априорных соображений мы можем предполагать, что $p = p_0$, где p_0 — некоторое фиксированное значение. По относительной частоте $\frac{m}{n}$ мы должны решить, справедлива гипотеза $p = p_0$ или нет. Поскольку при больших n относительная частота $\frac{m}{n}$ близка к p , то статистический критерий по проверке гипотезы $p = p_0$ должен основываться на разности $\left| \frac{m}{n} - p_0 \right|$. Если она большая, то, по-видимому, гипотеза неверна, если же она мала, то у нас нет основания отвергать гипотезу $p = p_0$.

Статистическое оценивание

б) *Статистическое оценивание неизвестных параметров.* Иногда нам требуется по наблюдаемому m указать то число \hat{p} , которое можно принять за вероятность p в схеме Бернулли. В нашем примере естественно взять $\hat{p} = \frac{m}{n}$. Оценка должна быть в том или ином смысле близкой к оцениваемому параметру.

Доверительные интервалы

в) *Доверительные интервалы.* Иногда нас интересует не точное значение неизвестного параметра p , а требуется указать тот интервал $\underline{p} \leq p \leq \bar{p}$, в котором с вероятностью, близкой к единице, лежит параметр p . Такой интервал $(\underline{p}(t), \bar{p}(t))$, концы которого случайны и зависят лишь от наблюдаемого значения t , называется *доверительным интервалом*.

«Затравка»

Ефимов-Поспелов, с. 268 по одному из изданий:

2. Проверка гипотез о параметре p биномиального распределения. При статистическом анализе данных, связанных с повторными независимыми испытаниями (схемой Бернулли), обычно рассматривают два вида задач: сравнение вероятности «успеха» p в одном испытании с заданным значением p_0 и сравнение вероятностей «успеха» в двух сериях испытаний.

В первом случае проверяется гипотеза $H_0: p = p_0$. Пусть в n испытаниях по схеме Бернулли «успех» произошел x раз. В качестве статистики критерия выбирают относительную частоту $h = x/n$. При больших значениях n ($n > 50$) и при выполнении условий $nh > 5$,

$n(1 - h) > 5$ распределение случайной величины h с достаточной для практических расчетов точностью аппроксимируется нормальным распределением $N(p, \sqrt{p(1-p)/n})$. Отсюда следует, что если гипотеза H_0 верна, то статистика

$$Z = \frac{h - p_0}{\sqrt{p_0(1 - p_0)/n}} \quad (4)$$

имеет распределение, близкое к нормальному распределению $N(0, 1)$. Критическая область критерия при уровне значимости α определяется неравенствами

$z_{\text{с}} > u_{1-\alpha}$ при альтернативной гипотезе $H_1^{(1)}: p > p_0$,

$z_{\text{с}} < u_{\alpha}$ при альтернативной гипотезе $H_1^{(2)}: p < p_0$,

$|z_{\text{с}}| > u_{1-\alpha/2}$ при альтернативной гипотезе $H_1^{(3)}: p \neq p_0$.

Для проверки гипотезы $H_0 : p = p_0$ также можно использовать доверительные интервалы для параметра p (§ 3, п. 2). При этом гипотеза H_0 принимается на уровне значимости α , если соответствующий односторонний или двусторонний доверительный интервал накрывает значение p_0 ; в противном случае гипотеза H_0 отклоняется.

Пример 6. Предполагается, что большая партия деталей содержит 15 % брака. Для проверки из партии случайным образом отобрано 100 деталей, среди которых оказалось 10 бракованных. Считая, что число бракованных деталей в партии имеет биномиальное распределение, и используя двусторонний критерий при $\alpha = 0,05$, проверить предположение о том, что в партии содержится 15 % бракованных деталей.

◁ Проверяется гипотеза $H_0: p = 0,15$ при альтернативной гипотезе $H_1: p \neq 0,15$. Значение $h = \frac{10}{100} = 0,1$. Так как $n > 50$, $nh = 10$ и $n(1 - h) = 9$, то для проверки гипотезы H_0 можно использовать статистику (4). Выборочное значение этой статистики

$$z_e = \frac{0,1 - 0,15}{\sqrt{\frac{0,15 \cdot 0,85}{100}}} \approx -1,4.$$

По таблице П1 находим $u_{0,975} = 1,96$. Значение $|z_e|$ лежит в области принятия гипотезы H_0 , следовательно, предположение о том, что в партии содержится 15 % брака, согласуется с результатами наблюдений. Этот же результат получим, используя двусторонний доверительный интервал $(0,041; 0,159)$ для p при доверительной вероятности 0,95 (см. пример 4 §3). Так как этот доверительный интервал накрывает значение $p = 0,15$, гипотеза H_0 принимается. ▷

Упр. Разобрать данный пример.

Задачи

19.245. Количество бракованных деталей в партии не должно превышать 5 %. В результате контроля 100 деталей из этой партии обнаружено 6 бракованных. Можно ли считать, что процент брака превосходит допустимый при $\alpha = 0,01$?

Ответ: нет.

19.246. При 600 подбрасываниях игральной кости шестерка появилась 75 раз.

а) Можно ли утверждать, что кость симметрична и однородна? Принять $\alpha = 0,05$.

б) Верна ли гипотеза о том, что вероятность появления шестерки меньше, чем $1/6$, если $\alpha = 0,01$?

Ответ: а) нет; б) да.

Тени прошлого...

26.02.2021 пт 17⁵⁰ 1006 СтоМод 09-065 Казанцев А.В.

Сб-к задач по мат-ке для вузов
Ч.3. ТВиМС. Под ред. Ав.Ефимова.
М.: Наука, 1990.

С. 261, пр-р 6. Предп-ся, что большая партия деталей содержит 15% брака. Для

проверки из партии случайным образом отобрано 100 деталей, среди коих оказалось 10 бракованных. Считаю, что число бракованных деталей в партии имеет биномиальное распределение, и используя двусторонний критерий при $\alpha = 0,05$, проверить предположение о том, что в партии сод-ся 15% бракованных деталей.

Проверяется гип-за $H_0: p = 0,15$ при альтерн. г-зе $H_1: p \neq 0,15$.

"Успех" $X = 10$; число исп-ий $n = 100$; статистика кр-я $h = \frac{x}{n} = 0,1$.

условия: $n = 100 > 50$, $nh = x = 10 > 5$; $n(1-h) = n-x = 90 > 5 \Rightarrow$ для аппр-ии м. исп-ть Норм. распр-е

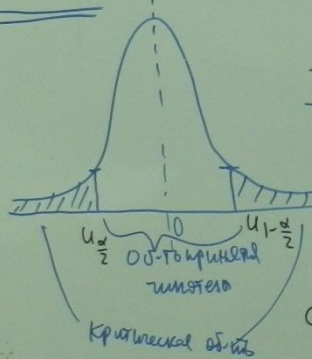
По табл. П1 (Ф-я распр-я Норм. с-та...) находим $\Phi_{0,975} = 1,96$ (с. 407 внизу)

$|z_6| = 1,4 < 1,96 \Rightarrow$ гип-за H_0 принимается, т.е. предположение о том, что в партии сод-я 15% брака, согласуется с факт-ми. Най-й.

Исп-я статистика $Z = \frac{h - p_0}{\sqrt{p_0(1-p_0)/n}}$

выборочное значение данной статистики
 $z_6 = \frac{0,1 - 0,15}{\sqrt{\frac{0,15 \cdot 0,85}{100}}} = -1,4002817 \approx -1,4$

$\Phi(\Phi_{0,975}) \approx P(Z < \Phi_{0,975}) = 0,975 \Rightarrow \Phi_{0,975} = 1,96$



Claim 1. $P(|Z| > u_{1-\frac{\alpha}{2}}) = \alpha$

Claim 2. $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$

$\Phi(u_{\frac{\alpha}{2}}) = \frac{\alpha}{2} = 1 - (1 - \frac{\alpha}{2}) = 1 - \Phi(u_{1-\frac{\alpha}{2}}) = \Phi(-u_{1-\frac{\alpha}{2}}) \Rightarrow u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$

Claim 3. $\Phi(-x) = 1 - \Phi(x) \mid \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$
 $\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-t^2/2} dt = -\frac{1}{\sqrt{2\pi}} \int_x^0 e^{-\tau^2/2} d\tau = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\tau^2/2} d\tau$

Формула Ватсона

$P(Z < -u_{1-\frac{\alpha}{2}})$

$\frac{\alpha}{2} = \Phi(-u_{1-\frac{\alpha}{2}})$

$\Phi(x) + \Phi(-x) = 1$

или $Z > u_{1-\frac{\alpha}{2}}$

$P(Z > u_{1-\frac{\alpha}{2}}) =$

$= 1 - P(Z \leq u_{1-\frac{\alpha}{2}}) =$

$= 1 - P(Z < u_{1-\frac{\alpha}{2}})$, т.к. Р-норм. распр.

$= 1 - \Phi(u_{1-\frac{\alpha}{2}}) =$

$= \Phi(-u_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$

$= \Phi(-u_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$

26.02.2021 пт 17⁵⁰ 1006 СтоМод 09-065 Казанцев А.В.

Сб-к задач по мат-ке для вузов
Ч.3. ТВиМС. Под ред. А.В.Ефимова.
М: Наука, 1990.

15.258. Кол-во ^{бракованных} деталей в партии не должно превышать 5%. В рез-те контроля 100

деталей из этой партии обнаружено 6 бракованных. Можно ли считать, что %-н брака превышает допустимый при $\alpha = 0,01$?

$$H_0: p = 0,05; \text{ альтерн: } H_1: p > 0,05$$

$$\text{"Успехи": } x = 6, \text{ число исп. } n = 100; \text{ ст-ка крп } h = \frac{x}{n} = 0,06$$

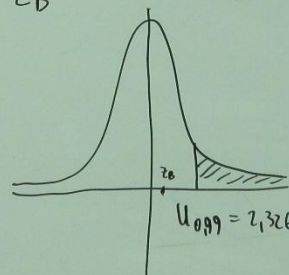
$$n(=100) > 50; nh(=x=6) > 5; n(1-h) = 100 - 6 = 94 > 5 \Rightarrow \text{исп-ся Норм. аппр.}$$

$$Z_B = \frac{h - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,06 - 0,05}{\sqrt{\frac{0,05 \cdot 0,95}{100}}} = 0,145$$

$$\text{Крити. отв-ть: } Z_B > u_{1-\frac{\alpha}{2}} = u_{0,99} = 2,326.$$

Но в нашем случае это не так;
 $Z_B = 0,145 < 2,326$, т.е.
 $Z_B \in \Theta$ -ти применяй гипотезу H_0

Ответ: Нет, при уровне значимости $\alpha = 0,01$
обнаружение 6-ти брак-х деталей в тестовой сотне
не дает оснований считать, что %-н брака превышает
допустимые 5 бракованных на сотню сотно.



15.259. При 600 подбрасываниях игральной
кости 6-рка появилась 75 раз.

а) М. ли утверждать, что Кость симметрична и
однородна? Примем $\alpha = 0,05$.

б) Верна ли гипотеза о том, что вер-сь появления
брок меньше, чем $1/6$,
если $\alpha = 0,01$?

Новости экономики

Задача 8. Случайные приращения цен акций двух компаний за день ξ и η имеют совместное распределение, заданное таблицей:

$\xi \backslash \eta$	-1	+1
-1	0,3	0,2
+1	0,1	0,4

Найти коэффициент корреляции.

Решение. Прежде всего вычисляем $M\xi\eta=0,3-0,2-0,1+0,4=0,4$. Далее находим частные законы распределения ξ и η :

$\xi \backslash \eta$	-1	+1	p_ξ
-1	0,3	0,2	0,5
+1	0,1	0,4	0,5
p_η	0,4	0,6	

Определяем $M\xi=0,5-0,5=0$; $M\eta=0,6-0,4=0,2$; $D\xi=1$; $D\eta=1-0,2^2=0,96$; $\text{cov}(\xi,\eta)=0,4$.

Получаем

$$\rho = \frac{0,4}{\sqrt{1}\sqrt{0,96}} \approx 0,408.$$

Задача 9. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi=1$ и $D\eta=2$, а коэффициент их корреляции $\rho=0,7$. Найти дисперсию приращения цены портфеля из 5 акций первой компании и 3 акций второй компании.

Решение. Используя свойства дисперсии, ковариации и определение коэффициента корреляции, получаем:

$$D(5\xi + 3\eta) = 5^2 D\xi + 3^2 D\eta + 2 \cdot 5 \cdot 3 \rho \sqrt{D\xi} \sqrt{D\eta} = 25 \cdot 1 + 9 \cdot 2 + 30 \cdot 0,7 \cdot 1 \cdot \sqrt{2} \approx 72,7.$$

Замечание 1. Дисперсия приращений цены портфеля акций часто используется на практике как мера риска вложений: чем больше дисперсия, тем больше риск. Поэтому оценка данной величины имеет важное значение для инвесторов.

Отголоски модели Марковица

Задача 10. Случайные приращения цен акций двух компаний³ за день имеют дисперсии $D\xi=4$ и $D\eta=9$, а их коэффициент корреляции $\rho=0,2$. В каких долях следует разделить капитал при вложении в акции, чтобы минимизировать риск?

Решение. Примем весь капитал за единицу. Обозначим долю, вложенную в акции первой компании, за x , тогда на долю акций второй компании останется $1-x$.

Случайное приращение цены пакета акций оказывается пропорционально линейной комбинации $\zeta = x\xi + (1-x)\eta$, и надо минимизировать ее дисперсию. Имеем

$$\begin{aligned} D\zeta &= D(x\xi + (1-x)\eta) = D(x\xi) + D((1-x)\eta) + 2\text{cov}(x\xi, (1-x)\eta) = \\ &= x^2 D\xi + (1-x)^2 D\eta + 2x(1-x)\text{cov}(\xi, \eta) = x^2 D\xi + (1-x)^2 D\eta + 2x(1-x)\rho\sqrt{D\xi}\sqrt{D\eta} = \\ &= 4x^2 + 9(1-x)^2 + 2x(1-x) \times 0,2 \times 2 \times 3 = 4x^2 + 9 - 18x + 9x^2 + 2,4x - 2,4x^2 = \\ &= 10,6x^2 - 15,6x + 9 \end{aligned}$$

Найдем минимум полученной функции, приравняв ее производную к нулю (либо по известной формуле вершины параболы). Получаем $x = 15,6 / (2 \times 10,6) = 39 / 53$, $1-x = 14 / 53$.

Таким образом, следует вложить $39/53$ капитала в акции первой компании и $14/53$ в акции второй компании.

Замечание 2. При решении подобных задач может формально получиться $x > 1$ или $x < 0$. Первое означает, что весь капитал следует вложить в акции первой компании, второе – что весь капитал следует вложить в акции второй компании.

ЛИТЕРАТУРА

1. Карасев В.А., Лёвшина Г.Д. "Теория вероятностей и математическая статистика: теория вероятностей: практикум". М. Изд. Дом МИСиС, 2015. № 2454. (печатное)
2. Карасев В.А., Лёвшина Г.Д. "Теория вероятностей и математическая статистика: математическая статистика: практикум". М. Изд. Дом МИСиС, 2016. № 2770. (электронное)
3. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для вузов. – М.: Издательство Юрайт, 2015.
4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учебное пособие для вузов. – М.: Издательство Юрайт, 2015.
5. Володин И.Н. Лекции по теории вероятностей и математической статистике. Казань, 2006. – 271 с.
6. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х тт. М.: «Мир», 1984.
7. Лекция – 2. ТВиМС Осн.теоремы. Полная вер-ть. Схема Бернулли, предельные теоремы.ppt (Материалы из курса в Moodle).
8. Лекция 2. Предмет теории вероятностей. Аксиомы Колмогорова.pdf (Материалы из курса в Moodle).
9. Лебедев А. В., Фадеева Л. Н. Теория вероятностей и математическая статистика: учебник / А. В. Лебедев, Л. Н. Фадеева. Под ред. А. В. Лебедева. Изд. 4-е, перераб. и доп. – М., 2018. – 480 с.
10. Боровков А.А. Теория вероятностей. – 3-е изд., существенно перераб. и доп. – М.: Эдиториал УРСС, 1999. – 472 с.
11. Севастьянов Б.А. Курс теории вероятностей и математической статистики. – М.: Наука, 1982. – 256 с.
12. Ефимов – Пospelов.

ФОТО НА ЛЕКЦИИ

25.03.2025 вт Верх 12⁴⁰ Л-556 МАТБИМС лекция Казанцев А.В.

Непр. распр-я: Напоминание

3-и больших чисел. Предельные т.мт: понятие о...

К матем. статистике. Немного об экономике.

с. 26

Почему $F(x)$
такое, как там?
Катя знает, е-мт

$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

$$\lambda = \text{const} > 0$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad ? \quad = \lim_{A \rightarrow +\infty} \int_0^A \lambda e^{-\lambda x} dx = \lim_{A \rightarrow +\infty} (e^{-\lambda x}) \Big|_0^A =$$

$$= \lim_{A \rightarrow +\infty} (1 - e^{-\lambda A}) = 1$$

$$\int_{-\infty}^{+\infty} f(x) dx = \underbrace{\int_{-\infty}^0 0 dx}_0 + \int_0^{+\infty} \lambda e^{-\lambda x} dx = \text{КТА}$$

$$M(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

np
тест Шапиро-Уилка
Колмогорова-Смирнова

$$P(\xi = k) = C_n^k p^k (1-p)^{n-k}$$

$$\sum_{k=0}^n P(\xi = k) = \sum_{k=0}^n C_n^k p^k q^{n-k} =$$

$$\sum_{k=0}^n k C_n^k p^k q^{n-k} = M\xi = \sum_{k=0}^n k P(\xi = k) = (p+q)^n = (p+1-p)^n = 1$$

$$k C_n^k = \frac{n! k}{k! (n-k)!} = n \frac{(n-1)!}{(k-1)! (n-1-(k-1))!} = n C_{n-1}^{k-1}$$

$$np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} q^{(n-1)-(k-1)} = np \sum_{m=0}^{n-1} C_{n-1}^m p^m q^{n-1-m}$$

$m = k-1 = 0 \leq m \leq n-1$

$(np) \underbrace{\quad}_{1}$

$$M\xi = np$$

$$(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$$

25.03.2025 вт Верх 12⁴⁰ Л-556 МАТБИМС лекция Казанцев А.В.

$$D(X+Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$$

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)} \sqrt{D(Y)}}$$

$$D(X) = DX$$

$$\text{Cov}(X, Y) = M(XY) - M(X)M(Y)$$

$$\begin{aligned} D(X+Y) &= M(X+Y)^2 - (M(X) + M(Y))^2 = \\ &= \underline{MX^2} + 2M(XY) + \underline{MY^2} - \underline{M(X)^2} - 2M(X)M(Y) - \underline{M(Y)^2} \\ &= DX + DY + 2(\underbrace{M(XY) - M(X)M(Y)}_{\text{Cov}(X, Y)}) \end{aligned}$$

с. 26
Почему $F(x)$
такое, как там?
Катя знает, я - нет

Зад. 10, с. 56.

$$\zeta = x\xi + (1-x)\eta$$

ξ, η - с.в.-мы

$$D\zeta = D(x\xi + (1-x)\eta) = D(x\xi) + D((1-x)\eta) + 2 \operatorname{Cov}(x\xi, (1-x)\eta) = \dots$$

Доразбирать самостоятельно!

Методики
и
лит-ра по

Стат. методы анализа данных
в прикладных решениях...