

Лекция 3. Некоторые распределения, используемые в математической статистике

3.1. Функции Эйлера

Определение 3.1. *Гамма-функцией Эйлера* называется функция

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt. \quad \# \quad (3.1)$$

Несложно проверить, что несобственный интеграл (3.1) сходится при $x \in (0, +\infty)$, поэтому указанное множество и является областью определения гамма-функции.

Приведем без доказательства некоторые свойства гамма-функции.

1°. $\Gamma(x)$ является бесконечное число раз дифференцируемой функцией, при этом

$$\Gamma^{(k)}(x) = \int_0^{+\infty} t^{x-1} e^{-t} \ln^k t dt.$$

2°. $\Gamma(x+1) = x\Gamma(x)$, $x > 0$.

3°. $\Gamma(1) = \int_0^{+\infty} e^{-t} dt = 1$.

4°. (из свойств 2° и 3° следует) $\Gamma(n+1) = n!$, $n \in \mathbb{N}_0$. Таким образом, гамма-функцию можно рассматривать как обобщение понятия факториала на (положительные) вещественные числа.

5°. $\Gamma(1/2) = \sqrt{\pi}$.

6°. $\Gamma(n+1/2) = \frac{(n-1)(n-3) \cdot \dots \cdot 3 \cdot 1}{2^n} \sqrt{\pi}$, $n \in \mathbb{N}$.

7°. Эскиз графика гамма-функции изображен на рис. 3.1. Гамма-функция принимает минимальное значение в точке $x = 3/2$, при этом $\Gamma(3/2) = \sqrt{\pi}/2$. Кроме того,

$$\lim_{x \rightarrow 0+} \Gamma(x) = \lim_{x \rightarrow +\infty} \Gamma(x) = +\infty.$$

Определение 3.2. *Бета-функцией Эйлера* называется функция двух переменных

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt. \quad \# \quad (3.2)$$

Несложно проверить, что интеграл (3.2) сходится при $x > 0$ и $y > 0$, поэтому областью определения бета-функции является множество $(0, +\infty) \times (0, +\infty)$.

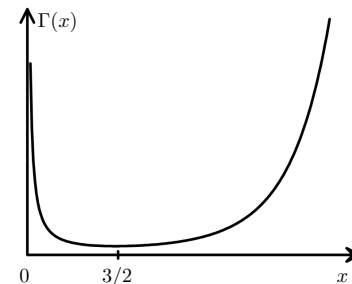


Рис. 3.1

Как и гамма-функция, бета-функция обладает рядом замечательных свойств, из которых помимо очевидной перестановочности аргументов

$$B(x, y) = B(y, x)$$

мы отметим лишь свойство, выражающее ее через гамма-функцию:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

3.2. Гамма-распределение

Определение 3.3. Говорят, что непрерывная случайная величина ξ имеет **гамма-распределение** с параметрами λ и α , если ее функция плотности имеет вид

$$f_{\xi}(x) = \begin{cases} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{иначе.} \end{cases}$$

Обозначение: $\xi \sim \Gamma(\lambda, \alpha)$. #

Замечание 3.1. Экспоненциальное распределение, плотность которого имеет вид

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{иначе,} \end{cases}$$

является частным случаем гамма-распределения для $\alpha = 1$, то есть $\text{Exp}(\lambda) = \Gamma(\lambda, 1)$. #

Теорема 3.1. Пусть случайные величины ξ_1, ξ_2 независимы и $\xi_i \sim \Gamma(\lambda, \alpha_i)$, $i = \overline{1, 2}$. Тогда $\xi_1 + \xi_2 \sim \Gamma(\lambda, \alpha_1 + \alpha_2)$.

◀ Без доказательства. ▶

Следствие 3.1. Если случайные величины ξ_i независимы и $\xi_i \sim \Gamma(\lambda, \alpha_i)$, $i = \overline{1, n}$, то $\xi_1 + \dots + \xi_n \sim \Gamma(\lambda, \alpha_1 + \dots + \alpha_n)$.

◀ Доказательство очевидно. ▶

3.3. Распределение Релея

Определение 3.4. Говорят, что случайная величина η имеет **распределение Релея** с параметром σ , если $\eta = \xi^2$, где $\xi \sim N(0, \sigma^2)$ — нормальная случайная величина с нулевым средним. #

Можно показать, что функция плотности распределения вероятностей такой случайной величины η имеет вид

$$f_{\eta}(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi x}} e^{-x/2\sigma^2}, & x > 0 \\ 0, & \text{иначе.} \end{cases}$$

Это означает, что распределение Релея является частным случаем гамма-распределения для $\lambda = 1/2\sigma^2$ и $\alpha = 1/2$, поэтому корректно использовать обозначение $\eta \sim \Gamma(1/2\sigma^2, 1/2)$.

3.4. Распределение χ^2

Определение 3.5. Говорят, что случайная величина η имеет **распределение хи-квадрат** с n степенями свободы, если

$$\eta = \xi_1^2 + \dots + \xi_n^2,$$

где случайные величины ξ_i независимы и $\xi_i \sim N(0, 1)$, $i = \overline{1, n}$. Обозначение: $\eta \sim \chi^2(n)$. #

Отметим некоторые свойства распределения χ^2 .

1°. $\chi^2(n) = \Gamma(1/2, n/2)$.

◀ Пусть $\eta \sim \chi^2(n)$, тогда $\eta = \xi_1^2 + \dots + \xi_n^2$, где случайные величины ξ_i независимы и каждая из них имеет стандартное нормальное распределение. Это означает, что ξ_i^2 имеет распределение Релея с параметром $\sigma = 1$, то есть $\xi_i^2 \sim \Gamma(1/2, 1/2)$. Так как случайные величины ξ_i независимы, то по свойству гамма-распределения

$$\xi_1^2 + \dots + \xi_n^2 \sim \Gamma\left(\frac{1}{2}, \underbrace{\frac{1}{2} + \dots + \frac{1}{2}}_{n \text{ раз}}\right) = \Gamma\left(\frac{1}{2}, \frac{n}{2}\right).$$

Таким образом, $\eta \sim \Gamma(1/2, n/2)$. ▶

2°. Если η_i — независимые случайные величины и $\eta_i \sim \chi^2(n_i)$, $i = \overline{1, m}$, то

$$\eta_1 + \dots + \eta_m \sim \chi^2(n_1 + \dots + n_m).$$

◀ Случайные величины η_i , $i = \overline{1, n}$, независимы, а в соответствии с предыдущим свойством $\eta_i \sim \Gamma(1/2, n_i/2)$. Использование известного свойства гамма-распределения завершает доказательство. ▶

3°. На рис. 3.2 приведены графики функций плотности для распределений $\chi^2(3)$ (кривая (a)) и $\chi^2(7)$ (кривая (б)). Видно, что при увеличении числа степеней свободы график функции плотности становится более пологим. Кроме того, известно, что при $n \geq 4$ график приобретает вторую точку перегиба, расположенную слева от максимума.

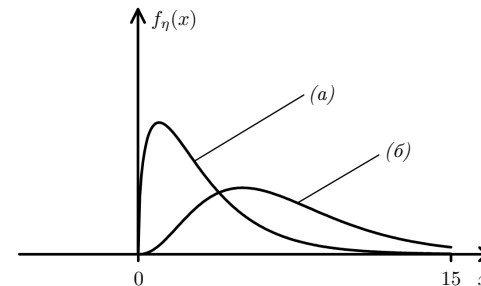


Рис. 3.2

3.5. Распределение Фишера

Определение 3.6. Говорят, что случайная величина ζ имеет **распределение Фишера** со степенями свободы n_1 и n_2 , если

$$\zeta = \frac{n_2 \eta_1}{n_1 \eta_2},$$

где случайные величины η_1 и η_2 независимы и $\eta_i \sim \chi^2(n_i)$, $i = \overline{1, 2}$. Обозначение: $\zeta \sim F(n_1, n_2)$. #

Можно показать, что функция плотности распределения Фишера со степенями свободы n_1 и n_2 имеет вид

$$f_{\zeta}(x) = \begin{cases} C \frac{x^{n_1/2-1}}{\left[1 + \frac{n_1 x}{n_2}\right]^{\frac{n_1+n_2}{2}}}, & x > 0 \\ 0, & \text{иначе,} \end{cases}$$

где

$$C = \frac{(n_1/n_2)^{n_1/2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}$$

(здесь B — бета-функция Эйлера).

Отметим также очевидное свойство распределения Фишера: если $\zeta \sim F(n_1, n_2)$, то $1/\zeta \sim F(n_2, n_1)$.

Замечание 3.2. Распределение Фишера также называют **распределением Фишера-Снедекора** или **F-распределением**. #

3.6. Распределение Стьюдента

Определение 3.7. Говорят, что случайная величина ζ имеет **распределение Стьюдента** с n степенями свободы, если

$$\zeta = \frac{\xi}{\sqrt{\eta}},$$

где случайные величины ξ и η независимы и $\xi \sim N(0, 1)$, $\eta \sim \chi^2(n)$. Обозначение: $\zeta \sim \text{St}(n)$. #

Можно показать, что функция плотности случайной величины ζ , распределенной по закону Стьюдента с n степенями свободы, имеет вид

$$f_{\zeta}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

На рис. 3.3 изображены графики функций плотности распределений $\text{St}(2)$ (кривая (а)) и $\text{St}(20)$ (кривая (б)). Видно, что чем меньше число степеней свободы соответствующего распределения, тем более пологим является график функции плотности. В то же время известно, что при $n \rightarrow \infty$ распределение $\text{St}(n)$ переходит в стандартное нормальное распределение.

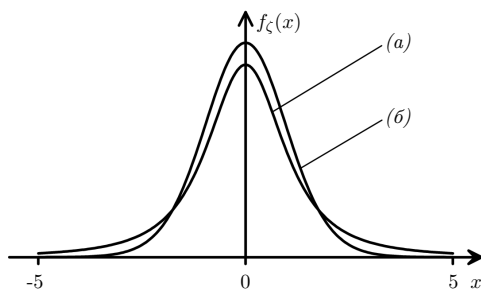


Рис. 3.3

Замечание 3.3. Распределение Стьюдента также называют *t-распределением*. #

Лекция 4. Точечные оценки

4.1. Основные понятия

4.2. Несмещенность точечной оценки

4.3. Состоятельность точечной оценки

4.4. Эффективность точечной оценки

Пусть X — случайная величина, закон распределения которой известен с точностью до параметра θ , а $\hat{\theta}(\vec{X})$ — точечная оценка этого параметра. Будем предполагать, что $\hat{\theta}$, а также все точечные оценки, рассматриваемые в этом разделе, являются несмещенными и обладают конечной дисперсией.

Как известно, дисперсия случайной величины характеризует разброс ее значений относительно математического ожидания. В случае несмещенной оценки $\hat{\theta}$ параметра θ дисперсия $D[\hat{\theta}]$ показывает, насколько кучно выборочные значения статистики $\hat{\theta}$ группируются около теоретического значения параметра: чем меньше $D[\hat{\theta}]$, тем меньше разброс этих значений и тем более удачной будет точечная оценка.

Определение 4.1. Оценка $\hat{\theta}_1$ называется *более эффективной* оценкой параметра θ , чем $\hat{\theta}_2$, если $D\hat{\theta}_1 < D\hat{\theta}_2$. #

Определение 4.2. Оценка $\hat{\theta}$ параметра θ называется *эффективной*, если она обладает наименьшей дисперсией среди всех оценок этого параметра. #

Замечание 4.1. Часто говорят не об эффективной оценке "вообще", а об эффективной оценке в некотором классе. Так, пусть Θ — некоторый класс (несмещенных) оценок параметра θ . Оценка $\hat{\theta} \in \Theta$ называется *эффективной в классе* Θ , если она обладает наименьшей дисперсией среди всех оценок из Θ . #

Пример 4.1. Пусть X — случайная величина, $m = MX$ и $\sigma^2 = DX$. Покажем, что выборочное среднее является эффективной оценкой для m в классе линейных оценок.

Произвольная линейная оценка параметра m имеет вид

$$\hat{m}(\vec{X}) = \sum_{i=1}^n \lambda_i X_i. \quad (4.1)$$

В самом начале настоящего раздела мы договорились рассматривать только несмещенные оценки, поэтому необходимо потребовать выполнение равенства $M[\hat{m}] = m$. Поскольку

$$M\left[\sum_{i=1}^n \lambda_i X_i\right] = \sum_{i=1}^n \lambda_i M[X_i] = \sum_{i=1}^n \lambda_i m = m \sum_{i=1}^n \lambda_i,$$

то требование несмещенности оценки можно записать в виде

$$\sum_{i=1}^n \lambda_i = 1. \quad (4.2)$$

Для нахождения эффективной оценки параметра m в классе линейных оценок попробуем подобрать коэффициенты λ_i , $i = \overline{1, n}$, оценки (4.1) так, чтобы при выполнении равенства (4.2) ее дисперсия была минимальной. Поскольку

$$D[\hat{\mu}] = D\left[\sum_{i=1}^n \lambda_i X_i\right] = \sum_{i=1}^n \lambda_i^2 D[X_i] = \sum_{i=1}^n \lambda_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^n \lambda_i^2,$$

то для достижения поставленной цели необходимо решить задачу условной оптимизации

$$\begin{cases} \sum_{i=1}^n \lambda_i^2 \rightarrow \min, \\ \sum_{i=1}^n \lambda_i - 1 = 0. \end{cases} \quad (4.3)$$

Решение задачи (4.3) проведем стандартным способом, известным из курса "Функции нескольких переменных" и предполагающим использование функции Лагранжа

$$L(\lambda_1, \dots, \lambda_n, \alpha) = \sum_{i=1}^n \lambda_i^2 - \alpha \left(\sum_{i=1}^n \lambda_i - 1 \right),$$

где α — вспомогательная переменная, называемая множителем Лагранжа. Необходимое условие условного экстремума в терминах функции Лагранжа в настоящем примере принимает вид:

$$\begin{aligned} \frac{\partial L}{\partial \lambda_i} &= 2\lambda_i - \alpha = 0, & i = \overline{1, n}, \\ \frac{\partial L}{\partial \alpha} &= -\left(\sum_{i=1}^n \lambda_i - 1 \right) = 0. \end{aligned}$$

Из семейства уравнений, записанных в первой строке, находим

$$\lambda_i = \alpha/2, \quad i = \overline{1, n},$$

и подставляем в уравнение из второй строки:

$$\sum_{i=1}^n \lambda_i = \sum_{i=1}^n \frac{\alpha}{2} = \frac{n\alpha}{2} = 1,$$

откуда

$$\alpha = 2/n$$

и, следовательно,

$$\lambda_i = \frac{1}{n}, \quad i = \overline{1, n}.$$

Проверив достаточное условие условного экстремума, можно показать, что найденные значения λ_i действительно являются точкой условного минимума функции $D[\hat{\mu}]$, поэтому оценка

$$\hat{m}(\vec{X}) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

является эффективной оценкой математического ожидания m в классе линейных несмещенных оценок. При этом дисперсия такой оценки

$$D\left[\hat{m}(\vec{X})\right] = \sigma^2 \left[\sum_{i=1}^n \lambda_i^2 \right] \Big|_{\lambda_i=1/n} = \frac{\sigma^2}{n}. \quad \#$$

Изучим некоторые свойства эффективных оценок, выраженные теоремами 4.1 и 4.2. Обе теоремы приведем без доказательства.

Теорема 4.1. (о единственности эффективной оценки) Пусть $\hat{\theta}_1(\vec{X})$, $\hat{\theta}_2(\vec{X})$ — две эффективные оценки параметра θ . Тогда $\hat{\theta}_1 = \hat{\theta}_2$. ◀ ▶

Перед формулировкой теоремы 4.2 потребуются некоторые дополнительные понятия. Во-первых, при доказательстве этой теоремы используется дифференцирование по параметру под знаком интеграла:

$$\frac{\partial}{\partial \theta} \int_G \varphi(T, \theta) dT = \int_G \frac{\partial \varphi(T, \theta)}{\partial \theta} dT.$$

Параметрические модели, для которых справедлив такой переход, будем называть **регулярными**.

Далее, пусть X — случайная величина, закон распределения которой известен с точностью до параметра θ , а $\vec{X} = (X_1, \dots, X_n)$ — случайная выборка объема n из генеральной совокупности X .

Определение 4.3. **Функцией правдоподобия** случайной выборки \vec{X} называется функция

$$\mathcal{L}(\vec{X}, \theta) = p(X_1, \theta) \cdot \dots \cdot p(X_n, \theta),$$

где

$$p(x, \theta) = \begin{cases} P\{X = x\}, & \text{если } X \text{ — дискретная случайная величина,} \\ f_X(x, \theta), & \text{если } X \text{ — непрерывная случайная величина} \end{cases}$$

(здесь f_X — функция плотности случайной величины X). #

Определение 4.4. **Количеством информации по Фишеру** в n наблюдениях называется число

$$I(\theta) = M \left\{ \left[\frac{\partial \ln \mathcal{L}(\vec{X}, \theta)}{\partial \theta} \right]^2 \right\}. \quad \#$$

Замечание 4.2. Иногда вводят понятие количества информации по Фишеру в i -м наблюдении:

$$I_0(\theta) = M \left\{ \left[\frac{\partial \ln p(X_i, \theta)}{\partial \theta} \right]^2 \right\},$$

где $p(X_i, \theta)$ имеет тот же смысл, что и в определении 4.3. Поскольку все элементы X_i случайной выборки \vec{X} независимы и имеют одинаковое с X распределение, величина I_0 фактически не

зависит от i и может быть записана в виде

$$I_0(\theta) = M \left\{ \left[\frac{\partial \ln p(X, \theta)}{\partial \theta} \right]^2 \right\}.$$

При этом можно показать, что для регулярных моделей $I(\theta) = nI_0(\theta)$, то есть серия из n наблюдений содержит "в n раз больше информации", чем отдельное наблюдение. #

Теорема 4.2. (Рао-Крамера) Пусть рассматриваемая параметрическая модель является регулярной, а $\hat{\theta}(\vec{X})$ — несмещенная точечная оценка параметра θ . Тогда справедливо неравенство

$$D[\hat{\theta}(\vec{X})] \geq \frac{1}{I(\theta)}. \quad \blacktriangleleft \blacktriangleright \quad (4.4)$$

Неравенство (4.4), называемое **неравенством Рао-Крамера**, дает нижнюю оценку значения дисперсии всех несмещенных точечных оценок неизвестного параметра θ в случае, когда соответствующая модель является регулярной.

Пусть $\hat{\theta}(\vec{X})$ — точечная оценка параметра θ , удовлетворяющая условиям теоремы 4.2.

Показателем эффективности по Рао-Крамеру точечной оценки $\hat{\theta}$ называется величина

$$e(\hat{\theta}) = \frac{1}{I(\theta)D[\hat{\theta}]}. \quad \blacktriangleleft \blacktriangleright$$

Согласно неравенству (4.4) справедливо $0 < e(\hat{\theta}) \leq 1$; при этом оценку $\hat{\theta}$, для которой $e(\hat{\theta}) = 1$, называют **эффективной по Рао-Крамеру**.

Замечание 4.3. Очевидно, что оценка, эффективная по Рао-Крамеру, также будет "просто" эффективной (то есть в эффективной смысле определения 4.2). Обратное неверно. Вообще говоря, вопрос о существовании эффективной и, в частности, эффективной по Рао-Крамеру точечной оценки для той или иной параметрической модели является нетривиальным и выходит за рамки настоящего курса. Мы познакомимся с ним лишь на небольшом числе примеров. #

Пример 4.2. Пусть $X \sim N(\theta, \sigma^2)$, где θ — неизвестный параметр, а σ^2 известно. Покажем, что выборочное среднее

$$\hat{\theta}(\vec{X}) = \bar{X}$$

является эффективной по Рао-Крамеру оценкой параметра θ .

В предыдущих примерах было показано, что выборочное среднее является несмещенной оценкой математического ожидания, а

$$D[\bar{X}] = \frac{\sigma^2}{n}.$$

Количество информации по Фишеру в одном испытании (см. замечание 4.2) дается формулой

$$I_0(\theta) = M \left\{ \left[\frac{\partial \ln f(X, \theta)}{\partial \theta} \right]^2 \right\},$$

где

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

— функция плотности распределения случайной величины X . Таким образом,

$$\begin{aligned} I_0(\theta) &= M \left\{ \left[\frac{\partial}{\partial \theta} \left(-\ln \sigma\sqrt{2\pi} - \frac{(X-\theta)^2}{2\sigma^2} \right) \right]^2 \right\} = \\ &= \frac{1}{\sigma^4} M \{ (X-\theta)^2 \} = \left| \frac{MX = \theta \Rightarrow}{\Rightarrow M \{ (X-\theta)^2 \} = DX = \sigma^2} \right| = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}. \end{aligned}$$

С учетом замечания 4.2 количество информации по Фишеру в n испытаниях равно

$$I(\theta) = nI_0(\theta) = n/\sigma^2,$$

поэтому показатель эффективности оценки $\hat{\theta}$

$$e(\hat{\theta}) = \frac{1}{I(\theta)D[\hat{\theta}]} = \frac{1}{\frac{n}{\sigma^2} \frac{\sigma^2}{n}} = 1.$$

Таким образом, в случае нормальной модели с известной дисперсией выборочное среднее является эффективной по Рао-Крамеру (и, следовательно, "просто" эффективной) оценкой математического ожидания. #

4.5. Метод моментов

Пусть X — случайная величина, закон распределения которой известен с точностью до вектора $\vec{\theta} = (\theta_1, \dots, \theta_r)$ неизвестных параметров. Предположим, что для X существуют r первых моментов, которые, очевидно, в общем случае зависят от неизвестных параметров:

$$MX^1 = m_1(\theta_1, \dots, \theta_r),$$

$$\dots \quad \dots \quad \dots$$

$$MX^r = m_r(\theta_1, \dots, \theta_r).$$

Метод моментов основан на том, что выборочные моменты

$$\hat{\mu}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = \overline{1, r},$$

являются состоятельными оценками своих теоретических аналогов. По этой причине в рассматриваемом методе принимают равенства

$$MX^k = \hat{\mu}_k(\vec{X}), \quad k = \overline{1, r},$$

что приводит к системе (в общем случае нелинейной) относительно неизвестных параметров:

$$\begin{cases} m_1(\theta_1, \dots, \theta_r) = \hat{\mu}_1(\vec{X}), \\ \dots \\ m_r(\theta_1, \dots, \theta_r) = \hat{\mu}_r(\vec{X}). \end{cases} \quad (4.5)$$

Решение этой системы

$$\theta_1 = \hat{\theta}_1(\vec{X}), \dots, \theta_r = \hat{\theta}_r(\vec{X})$$

дает искомые точечные оценки.

Замечание 4.4. а) Иногда некоторые уравнения системы (4.5) записывают относительно центральных, а не начальных моментов. В этом случае соответствующее уравнение будет иметь вид:

$$\hat{m}_k(\theta_1, \dots, \theta_r) = \hat{\nu}_k(\vec{X}),$$

где

$$\hat{m}_k(\theta_1, \dots, \theta_r) = M(X - MX)^k, \quad \hat{\nu}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

— соответственно теоретический и выборочный k -е центральные моменты. Выбор между начальным и центральным моментом делают исходя из того, что система (4.5) должна решаться как можно проще.

б) Поскольку выборочные моменты $\hat{\mu}_k$ и $\hat{\nu}_k$ являются состоятельными оценками соответствующих теоретических моментов, то можно показать, что при условии непрерывной зависимости решения системы (4.5) от ее правых частей оценка, полученная методом моментов, является состоятельной и имеет асимптотически нормальное распределение.

в) Поскольку при $k \geq 2$ выборочные моменты $\hat{\mu}_k$, $\hat{\nu}_k$ являются смещенными оценками своих теоретических аналогов, то и полученные с использованием метода моментов точечные оценки неизвестных параметров также могут быть смещенными. #

Пример 4.3. Пусть $X \sim R[a, b]$. С использованием метода моментов построить точечные оценки для параметров a и b .

Закон распределения случайной величины X зависит от двух неизвестных параметров, поэтому система метода моментов будет содержать два уравнения, записанных относительно моментов 1-го и 2-го порядка. Запишем первое уравнение относительно начального момента, а второе относительно центрального. Как известно, для равномерно распределенной на отрезке $[a, b]$ случайной величины X

$$m_1(a, b) = MX = \frac{a+b}{2}, \quad \hat{m}_2(a, b) = DX = \frac{(b-a)^2}{12},$$

поэтому в рассматриваемом случае система (4.5) (с учетом замечания 4.4, а) примет вид

$$\frac{a+b}{2} = \hat{\mu}_1(\vec{X}), \quad \frac{(b-a)^2}{12} = \hat{\nu}_2(\vec{X}), \quad (4.6)$$

где очевидно $\hat{\mu}_1(\vec{X}) = \bar{X}$, $\hat{\nu}_2(\vec{X}) = \hat{\sigma}^2(\vec{X})$. Полученная система уравнений эквивалентна системе

$$a+b = 2\hat{\mu}_1, \quad (b-a)^2 = 12\hat{\nu}_2,$$

решение которой приведено ниже:

$$\begin{aligned} \begin{cases} a = 2\hat{\mu}_1 - b, \\ (2b - 2\hat{\mu}_1)^2 = 12\hat{\nu}_2, \end{cases} & \Rightarrow \begin{cases} a = 2\hat{\mu}_1 - b, \\ (b - \hat{\mu}_1)^2 = 3\hat{\nu}_2, \end{cases} \Rightarrow \\ & \Rightarrow \begin{cases} a = 2\hat{\mu}_1 - b, \\ b - \hat{\mu}_1 = \pm\sqrt{3\hat{\nu}_2}, \end{cases} \Rightarrow \begin{cases} a = 2\hat{\mu}_1 - b, \\ b = \hat{\mu}_1 \pm \sqrt{3\hat{\nu}_2} \end{cases} \Rightarrow \begin{cases} a = \hat{\mu}_1 \mp \sqrt{3\hat{\nu}_2}, \\ b = \hat{\mu}_1 \pm \sqrt{3\hat{\nu}_2}. \end{cases} \end{aligned}$$

Поскольку по содержательному смыслу параметров $a < b$, то, выбирая соответствующее решение из двух найденных, получаем искомые точечные оценки:

$$a = \hat{a}(\vec{X}) = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad b = \hat{b}(\vec{X}) = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}. \quad \#$$

4.6. Метод максимального правдоподобия

Пусть X — случайная величина, общий вид закона распределения которой известен. Выше в определении 4.3 было введено понятие функции правдоподобия случайной выборки в случае, когда закон распределения генеральной совокупности X зависит от одного неизвестного параметра θ . В случае, когда этот закон зависит от вектора $\vec{\theta} = (\theta_1, \dots, \theta_r)$ неизвестных параметров, функция правдоподобия определяется аналогично:

$$\mathcal{L}(\vec{X}, \vec{\theta}) = p(X_1, \vec{\theta}) \cdot \dots \cdot p(X_n, \vec{\theta}),$$

где

$$p(x, \vec{\theta}) = \begin{cases} P\{X = x\}, & \text{если } X \text{ — дискретная случайная величина,} \\ f_X(x, \vec{\theta}), & \text{если } X \text{ — непрерывная случайная величина} \end{cases}$$

(здесь так же f_X — функция плотности случайной величины X).

Замечание 4.5. Пусть, например, $X \sim N(m, \sigma^2)$, где m и σ^2 являются неизвестными параметрами. Преположим, что $m = m_0$, $\sigma^2 = \sigma_0^2$ — теоретические ("истинные") значения этих параметров (разумеется, эти значения неизвестны), а $m = m_1$ и $\sigma^2 = \sigma_1^2$ — некоторые подобранные нами значения этих параметров. При проведении испытаний элементы выборки будут группироваться в районе m_0 (см. рис. 4.4). Из иллюстрации видно, что чем ближе подобранные значения m_1 и σ_1^2 к m_0 и σ_0^2 соответственно, тем больше значения функции плотности f_X в точках X_i и, следовательно, тем больше значения функции правдоподобия $\mathcal{L}(\vec{X}, m_1, \sigma_1^2)$. #

Можно показать, что результат замечания 4.5 справедлив для произвольной параметрической модели: чем более удачно выбраны значения неизвестных параметров, тем больше значения функции правдоподобия. По этой причине в методе максимального правдоподобия в качестве точечных оценок неизвестных параметров принимают такие их значения, которые для каждой реализации \vec{x} случайной выборки \vec{X} максимизируют функцию правдоподобия $\mathcal{L}(\vec{x}, \vec{\theta})$, то есть полагают

$$\hat{\vec{\theta}} = \arg \max_{\vec{\theta}} \mathcal{L}(\vec{X}, \vec{\theta}).$$

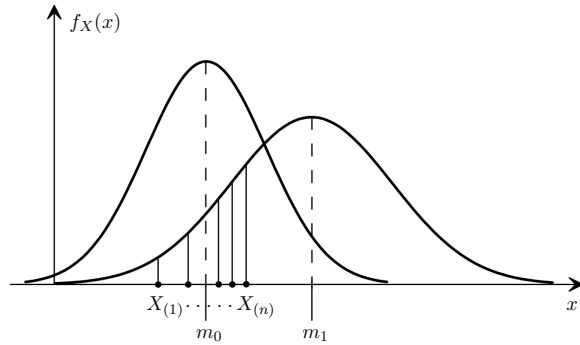


Рис. 4.4

При решении этой задачи оптимизации обычно используют необходимое условие экстремума функции нескольких переменных:

$$\begin{aligned} \frac{\partial \mathcal{L}(\vec{X}, \theta_1, \dots, \theta_r)}{\partial \theta_1} &= 0, \\ &\dots \\ \frac{\partial \mathcal{L}(\vec{X}, \theta_1, \dots, \theta_r)}{\partial \theta_r} &= 0. \end{aligned}$$

Указанные соотношения при этом называют **уравнениями правдоподобия**.

Замечание 4.6. Функция правдоподобия представляет собой произведение n сомножителей, зависящих от вектора $\vec{\theta}$ неизвестных параметров, поэтому вычисление входящих в уравнения правдоподобия производных не очень удобно. По этой причине вместо задачи

$$\mathcal{L}(\vec{X}, \vec{\theta}) \rightarrow \max_{\vec{\theta}}$$

обычно рассматривают эквивалентную задачу

$$\ln \mathcal{L}(\vec{X}, \vec{\theta}) \rightarrow \max_{\vec{\theta}}$$

(задачи и в самом деле эквивалентны, поскольку логарифм является монотонно возрастающей функцией). В этом случае необходимое условие экстремума принимает вид

$$\frac{\partial \ln \mathcal{L}(\vec{X}, \theta_1, \dots, \theta_r)}{\partial \theta_j} = 0, \quad j = \overline{1, r},$$

а указанные соотношения также называют уравнениями правдоподобия. #

Пример 4.4. Пусть $X \sim R[a, b]$. С использованием метода максимального правдоподобия построить точечные оценки параметров a и b .

Равномерно распределенная на отрезке $[a, b]$ случайная величина X является непрерывной, а ее функция плотности имеет вид

$$f_X(x, a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{иначе,} \end{cases}$$

поэтому на первый взгляд в рассматриваемом примере функция правдоподобия записывается как

$$\mathcal{L}(\vec{X}, a, b) = f_X(X_1, a, b) \cdot \dots \cdot f_X(X_n, a, b) = \frac{1}{(b-a)^n}, \quad (4.7)$$

а ее логарифм —

$$\ln \mathcal{L} = -n \ln(b-a). \quad (4.8)$$

Уравнения правдоподобия принимают вид

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial a} &= \frac{n}{b-a} = 0, \\ \frac{\partial \ln \mathcal{L}}{\partial b} &= -\frac{n}{b-a} = 0, \end{aligned}$$

и приводят к противоречивому результату $n = 0$.

Проблема в том, что в настоящем примере границы области, в которой $f_X > 0$, зависят от значений неизвестных параметров a и b , поэтому выражения (4.7) и (4.8) для функции правдоподобия и ее логарифма составлены некорректно. Правильным является выражение (см. рис. 4.5)

$$\mathcal{L}(\vec{X}, a, b) = \begin{cases} \frac{1}{(b-a)^n}, & \text{если } a \leq X_i \leq b, \quad i = \overline{1, n}, \\ 0, & \text{иначе,} \end{cases} \quad (4.9)$$

дифференцирование которого по неизвестным параметрам затруднительно.

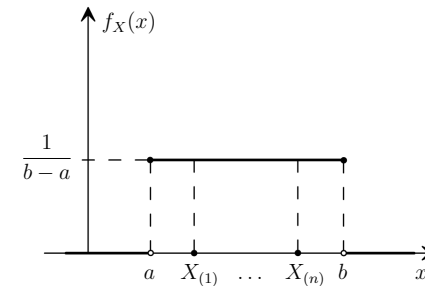


Рис. 4.5

Попробуем подобрать значения a и b , максимизирующие функцию \mathcal{L} , без использования необходимого условия экстремума, выраженного уравнениями правдоподобия. При решении этой задачи параметры a и b являются переменными и их значения можно варьировать. Предположим, что параметр a принял достаточно большое значение и один или несколько элементов выборки оказались меньше a . В этом случае $f_X(X_{(1)}, a, b) = 0$ и, следовательно, $\mathcal{L} = 0$.

Аналогичный результат получается и в случае, когда параметр b принимает достаточно малое значение, для которого $b < X_{(n)}$. Поскольку функция правдоподобия подлежит максимизации, то a и b следует выбирать так, чтобы

$$a \leq X_1 \leq \dots \leq X_{(n)} \leq b, \quad (4.10)$$

в этом случае ее значения, определяемые верхней формулой (4.9), будут положительными. Непосредственный анализ этой формулы показывает, что значения функции \mathcal{L} будут тем больше, чем меньше длина отрезка $[a, b]$. С учетом ограничений (4.10) заключаем, что максимальное значение функции правдоподобия достигается при

$$a = \hat{a}(\vec{X}) = X_{(1)}, \quad b = \hat{b}(\vec{X}) = X_{(n)},$$

что и дает искомые оценки неизвестных параметров. #

Замечание 4.7. Сравнение результатов примеров 4.3 и 4.4шний раз показывает, что для одного и того же неизвестного параметра могут быть предложены различные точечные оценки. #