Math-Net.Ru
Общероссийский математический портал

В. И. Бахтин, Статистические оценки для экспоненциального семейства мер на основе косвенных наблюдений, *Тр. Ин-та матем.*, 2013, том 21, номер 2, 54–62

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
`http://www.mathnet.ru/rus/agreement`

Параметры загрузки:
IP: 18.97.14.81
26 марта 2025 г., 16:47:53

Национальная академия наук Беларуси
Труды Института математики. 2013. Том 21. № 1. С. 54–62

УДК 519.233.22

СТАТИСТИЧЕСКИЕ ОЦЕНКИ ДЛЯ ЭКСПОНЕНЦИАЛЬНОГО СЕМЕЙСТВА МЕР НА ОСНОВЕ КОСВЕННЫХ НАБЛЮДЕНИЙ

В. И. Бахтин

Белорусский государственный университет
e-mail: bakhtin@tut.by
Поступила 05.04.2013

Вводятся понятия косвенной статистической оценки и ее относительной эффективности. В случае простейшего экспоненциального семейства распределений доказывается аналог неравенства Рао–Крамера для косвенных оценок, а также, что косвенная оценка, построенная методом моментов, относительно эффективна.

1. Введение. По определению, всякая статистическая оценка является функцией от выборки независимых случайных наблюдений $x = (x_1, \dots, x_n)$. Однако легко представить себе ситуацию, когда экспериментатору становится известна не вся выборка, а лишь эмпирическое среднее некоторой функции f, определяемое по формуле

$$\bar{f} = \frac{f(x_1) + \ldots + f(x_n)}{n}.$$

Такое эмпирическое среднее мы будем называть *косвенным наблюдением.* Любую статистическую оценку, являющуюся функцией от $\bar{f}$, тоже будем называть косвенной. В случае простейшего экспоненциального семейства распределений в работе доказывается аналог неравенства Рао–Крамера для косвенных оценок, определяется отвечающее ему понятие относительной эффективности и доказывается, что косвенная оценка, построенная методом моментов, относительно эффективна.

В статистической физике экспоненциальные семейства распределений называются распределениями Гиббса, а их нормировочные делители – статистическими суммами. Свойства распределений Гиббса тесным образом связаны с функционалом, равным логарифму статистической суммы. Мы называем этот функционал спектральным потенциалом и начинаем изложение с изучения его свойств.

2. Спектральный потенциал. Будем рассматривать произвольное множество X, на котором задана некоторая σ-алгебра его подмножеств $\mathfrak{A}$. Обозначим через $B(X)$ пространство всех ограниченных измеримых функций $f\colon (X, \mathfrak{A}) \to \mathbb{R}$, через $M(X)$ множество всех положительных мер на $(X, \mathfrak{A})$ и через $M_1(X)$ совокупность всех вероятностных мер на $(X, \mathfrak{A})$.

Каждая мера $\mu \in M(X)$ порождает линейный функционал на $B(X)$ по формуле

$$\mu[f] = \int\limits_X f\, d\mu. \tag{1}$$

Легко видеть, что этот функционал *положителен* (принимает неотрицательные значения на неотрицательных функциях). Если же мера μ вероятностная, то он *нормирован* (принимает единичное значение на единичной функции). Для краткости договоримся использовать обозначение (1) для всех интегрируемых функций $f \in L^1(X, \mu)$.

Фиксируем меру $\mu \in M(X)$. Определим для нее множество измеримых функций

$$D(\mu) = \{\varphi : \ (X, \mathfrak{A}) \to \mathbb{R} |\ \mu[e^\varphi] < +\infty\}.$$

Очевидно, если к функции $\varphi \in D(\mu)$ добавить любую функцию $\psi \in B(X)$, то их сумма тоже будет лежать в $D(\mu)$. В силу неравенства Гёльдера для всяких $\varphi, \psi \in D(\mu)$ выполняется соотношение

$$\mu[e^{(1-t)\varphi+t\psi}] \le \mu[e^\varphi]^{1-t}\mu[e^\psi]^t, \quad t \in [0,1]. \tag{2}$$

Из него следует, что множество $D(\mu)$ выпукло.

Спектральным потенциалом мы будем называть нелинейный функционал

$$\lambda(\varphi, \mu) = \ln \mu[e^\varphi], \quad \mu \in M(X), \quad \varphi : \ (X, \mathfrak{A}) \to \mathbb{R}. \tag{3}$$

Он обладает следующими свойствами.

Во-первых, спектральный потенциал конечен, если $\varphi \in D(\mu)$, и равен $+\infty$ в противном случае. Во-вторых, он монотонен: если $\varphi \ge \psi$, то тогда $\lambda(\varphi, \mu) \ge \lambda(\psi, \mu)$. В-третьих, он аддитивно однороден: $\lambda(\varphi + t, \mu) = \lambda(\varphi, \mu) + t$ для любой константы t. В-четвертых, из неравенства (2) вытекает, что этот функционал выпуклый по φ:

$$\ln \mu[e^{(1-t)\varphi+t\psi}] \le (1-t)\ln \mu[e^\varphi] + t\ln \mu[e^\psi], \quad t \in [0,1].$$

В-пятых, в силу вогнутости логарифмической функции он вогнут по отношению к μ:

$$\ln\{(1-t)\mu[e^\varphi] + t\nu[e^\varphi]\} \ge (1-t)\ln \mu[e^\varphi] + t\ln \nu[e^\varphi], \quad \mu, \nu \in M(X).$$

Наконец, можно заметить, что при всех $\varphi \in D(\mu)$ и $\psi \in B(X)$ выражение $\lambda(\varphi + \psi, \mu)$ аналитически зависит от аргумента ψ.

Для всякой функции $\varphi \in D(\mu)$ определим вероятностную меру $\mu_\varphi \in M_1(X)$ с помощью формулы

$$\mu_\varphi[f] = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]}, \quad f \in B(X).$$

Эта мера эквивалентна μ и имеет по отношению к μ плотность $e^{\varphi - \lambda(\varphi,\mu)}$. С другой стороны, из теоремы Радона–Никодима вытекает, что любая вероятностная мера ν, эквивалентная μ, может быть представлена в виде $\nu = \mu_\varphi$, где $\varphi = \ln\{d\nu/d\mu\}$.

Вычислим первые две производные от спектрального потенциала по отношению к аргументу φ. Предположим, что измеримые функции φ, f удовлетворяют условию $\varphi + tf \in D(\mu)$ при всех достаточно малых вещественных t. В таком случае введем обозначение

$$\lambda'(\varphi, \mu)[f] = \frac{d\lambda(\varphi + tf, \mu)}{dt}\bigg|_{t=0}.$$

Это не что иное, как производная от спектрального потенциала по направлению f в точке φ. Элементарная выкладка показывает, что эта производная существует, конечна и имеет вид

$$\lambda'(\varphi, \mu)[f] = \frac{d\ln \mu[e^{\varphi+tf}]}{dt}\bigg|_{t=0} = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]} = \mu_\varphi[f]. \tag{4}$$

В частности, формула (4) верна для всех функций $\varphi \in D(\mu)$ и $f \in B(X)$.

Пусть теперь измеримые функции φ, f, g удовлетворяют условию $\varphi+tf+sg\in D(\mu)$ при всех достаточно малых t, s. В силу выпуклости множества $D(\mu)$ это условие равносильно тому, что по отдельности $\varphi+tf\in D(\mu)$ и $\varphi+tg\in D(\mu)$ при достаточно малых t. Положим

$$\lambda''(\varphi,\mu)[f,g]=\frac{\partial^2\lambda(\varphi+tf+sg,\mu)}{\partial s\,\partial t}\bigg|_{s,t=0}.$$

Вычислим последнюю производную в явном виде с помощью формулы (4):

$$\lambda''(\varphi,\mu)[f,g]=\frac{\partial}{\partial s}\left(\frac{\mu[e^{\varphi+sg}f]}{\mu[e^{\varphi+sg}]}\right)\bigg|_{s=0}=\frac{\mu[e^{\varphi}fg]}{\mu[e^{\varphi}]}-\frac{\mu[e^{\varphi}f]\mu[e^{\varphi}g]}{(\mu[e^{\varphi}])^2}=\mu_\varphi[fg]-\mu_\varphi[f]\mu_\varphi[g]. \tag{5}$$

В теории вероятностей выражение $\mu_\varphi[f]$ называют *математическим ожиданием* случайной величины f по отношению к вероятностной мере μ_φ, а выражение $\mu_\varphi[fg]-\mu_\varphi[f]\mu_\varphi[g]$ называют *ковариацией* случайных величин f и g. В частности, значение квадратичной формы

$$\lambda''(\varphi,\mu)[f,f]=\frac{d^2\lambda(\varphi+tf,\mu)}{dt^2}\bigg|_{t=0}=\mu_\varphi[f^2]-\mu_\varphi[f]^2 \tag{6}$$

называется дисперсией случайной величины f (по отношению к распределению μ_φ).

Для очистки совести отметим, что мы никак не обосновывали законность дифференцирования под знаком интеграла при выводе формул (4) и (5). Это несложно сделать, применяя теорему Лебега о мажорируемой сходимости к семействам $\mu_{\varphi+tf}[f]$ и $\mu_{\varphi+tf+sg}[fg]$.

3. Большие уклонения эмпирических мер. Пусть задано вероятностное пространство $(X,\mathfrak{A},\mu)$. Рассмотрим последовательность независимых случайных величин $x_i\in X$ с распределением μ. Усиленный закон больших чисел утверждает, что для любой измеримой функции $f:\ X\to\mathbb{R}$ с конечным математическим ожиданием $\mu[f]$

$$\frac{f(x_1)+\ldots+f(x_n)}{n}\longrightarrow\mu[f]\quad\text{почти наверное.}$$

Всякой конечной последовательности $x=(x_1,\ldots,x_n)\in X^n$ сопоставим *эмпирическую меру* $\delta_{x,n}\in M_1(X)$. По определению она сосредоточена на конечном множестве $\{x_1,\ldots,x_n\}$ и приписывает каждой точке x_i меру $1/n$. Интеграл от любой функции f по эмпирической мере имеет вид

$$\delta_{x,n}[f]=\frac{f(x_1)+\ldots+f(x_n)}{n}.$$

В силу закона больших чисел последовательность $\delta_{x,n}[f]$ почти наверное сходится к $\mu[f]$. Поэтому события вида $|\delta_{x,n}[f]-\mu[f]|>\varepsilon$ принято называть *большими уклонениями.* Сейчас мы покажем, как можно легко получать экспоненциальные оценки для вероятностей больших уклонений с помощью спектрального потенциала.

Обозначим через μ^n декартову степень меры μ, определенную на $(X,\mathfrak{A})^n$.

Лемма 1. *Для любой функции* $f\in B(X)$ *и любого числа* $\varepsilon>0$

$$\mu^n\{x\in X^n:\ |\delta_{x,n}[f]-\mu[f]|\ge\varepsilon\}\le 2e^{-n\varepsilon^2/2\sigma^2}, \tag{7}$$

где $\sigma=\sup f(X)-\inf f(X)$ *обозначает осцилляцию функции* f.

Доказательство. Рассмотрим семейство вероятностных мер на X

$$\mu_{tf}=e^{tf-\lambda(tf,\mu)}\mu,$$

зависящее от вещественного параметра t. Очевидно, $d\mu/d\mu_{tf} = e^{\lambda(tf,\mu)-tf}$, и для любой точки $x = (x_1, \ldots, x_n)$ выполняются равенства

$$\frac{d\mu^n(x)}{d\mu^n_{tf}(x)} = \prod_{i=1}^{n} \frac{d\mu(x_i)}{d\mu_{tf}(x_i)} = \prod_{i=1}^{n} e^{\lambda(tf,\mu)-tf(x_i)} = e^{n(\lambda(tf,\mu)-t\delta_{x,n}[f])}.$$

Определим множество

$$Y_n = \{x \in X^n |\ \delta_{x,n}[f] \geq \mu[f] + \varepsilon\}.$$

Тогда для всех $x \in Y_n$ и $t > 0$

$$\frac{d\mu^n(x)}{d\mu^n_{tf}(x)} \leq e^{n(\lambda(tf,\mu)-t\mu[f]-t\varepsilon)},$$

откуда вытекает, что

$$\mu^n(Y_n) = \int\limits_{Y_n} d\mu^n(x) = \int\limits_{Y_n} \frac{d\mu^n(x)}{d\mu^n_{tf}(x)}\, d\mu^n_{tf}(x) \leq e^{n(\lambda(tf,\mu)-t\mu[f]-t\varepsilon)}. \tag{8}$$

Чтобы получить из (8) минимальную оценку сверху для $\mu^n(Y_n)$, нужно подобрать такое $t > 0$, при котором выражение $\lambda(tf,\mu) - t\mu[f] - t\varepsilon$ минимально. Заметим, что в силу (3), (4), (6)

$$\lambda(0,\mu) = 0, \quad \left.\frac{d\lambda(tf,\mu)}{dt}\right|_{t=0} = \mu[f],$$

$$\frac{d^2\lambda(tf,\mu)}{dt^2} = \mu_{tf}[f^2] - \mu_{tf}[f]^2 = \mu_{tf}[(f - \mu_{tf}[f])^2] \leq \sigma^2.$$

Поэтому

$$\lambda(tf,\mu) - t\mu[f] - t\varepsilon \leq \frac{\sigma^2 t^2}{2} - t\varepsilon \leq -\frac{\varepsilon^2}{2\sigma^2} \quad \text{при} \quad t = \frac{\varepsilon}{\sigma^2}. \tag{9}$$

Из (8) и (9) вытекает неравенство

$$\mu^n\{x \in X^n :\ \delta_{x,n}[f] \geq \mu[f] + \varepsilon\} \leq e^{-n\varepsilon^2/2\sigma^2}.$$

Складывая его с точно таким же неравенством для функции $-f$, получаем (7).

Замечание. В случае неограниченной функции f те же самые рассуждения дают оценку

$$\mu^n\{x \in X^n :\ \delta_{x,n}[f] \geq \mu[f] + \varepsilon\} \leq \exp\{n \inf_{t>0}\{\lambda(tf,\mu) - t\mu[f] - t\varepsilon\}\}$$

при условии, что выражение $\mu[e^{tf}]$ конечно при достаточно малых $t > 0$.

Лемма 2. *Для любой функции* $f \in B(X)$ *при всех* $k \in \mathbb{N}$ *справедливы оценки*

$$\mu^n[|\delta_{x,n}[f] - \mu[f]|^k] \leq C(k)\sigma^k n^{-k/2}, \quad \sigma = \sup f - \inf f,$$

где $C(k)$ *— не зависящие от* f *и* n *константы.*

Доказательство. Действительно, в силу (7)

$$\mu^n[|\delta_{x,n}[f] - \mu[f]|^k] \leq \sum_{i=0}^{\infty} \mu^n\left[x \in X^n :\ \frac{\sigma i}{\sqrt{n}} \leq |\delta_{x,n}[f] - \mu[f]| \leq \frac{\sigma(i+1)}{\sqrt{n}}\right] \frac{\sigma^k (i+1)^k}{n^{k/2}} \leq$$

$$\leq \frac{\sigma^k}{n^{k/2}} \left(1 + \sum_{i=1}^{\infty} 2e^{-i^2/2}(i+1)^k\right).$$

4. Неравенство Рао–Крамера для косвенных оценок. Пусть на измеримом пространстве $(X, \mathfrak{A})$ заданы вероятностная мера μ и функция $\varphi \in B(X)$. Определим для них экспоненциальное семейство мер μ_p, зависящее от вещественного параметра p:

$$\mu_p = e^{p\varphi - \lambda(p\varphi, \mu)}\mu.$$

Предположим, что производится случайный эксперимент, результатом которого является выборка $x = (x_1, \ldots, x_n)$, в которой все выборочные значения имеют распределение μ_p и независимы. Этот результат “измеряется” прибором, показывающим значение

$$\delta_{x,n}[f] = \frac{f(x_1) + \ldots + f(x_n)}{n},$$

где f — некоторая ограниченная измеримая функция на X (*калибровочная функция*). Имея только это показание, требуется оценить неизвестное значение параметра p.

Будем рассматривать оценки вида $\hat{p} = F(\delta_{x,n}[f])$, где $F(y)$ — гладкая функция на вещественной оси. Такие оценки естественно называть *косвенными* в отличие от обычных статистических оценок, являющихся функциями от всей выборки x.

Обозначим через $b_n(p)$ смещение статистической оценки $\hat{p}$:

$$b_n(p) = \mathsf{E}\hat{p} - p = \mu_p^n[\hat{p}] - p,$$

а через $\operatorname{Var}\hat{p}$ — ее вариацию относительно истинного значения p:

$$\operatorname{Var}\hat{p} = \mathsf{E}\{(\hat{p} - p)^2\} = \mu_p^n[(\hat{p} - p)^2].$$

Эту вариацию принято рассматривать как меру точности статистической оценки.

Знаменитое неравенство Рао–Крамера утверждает, что если оценка $\hat{p}$ строится как функция от всей выборки $x = (x_1, \ldots, x_n)$, т.е. $\hat{p} = \hat{p}(x)$, то тогда

$$\operatorname{Var}\hat{p} \geq \frac{(1 + b_n'(p))^2}{n\mathcal{I}(p)} + b_n(p)^2, \tag{10}$$

где $\mathcal{I}(p)$ — так называемая информационная функция Фишера, определяемая равенством

$$\mathcal{I}(p) = \operatorname{cov}_p\{\varphi, \varphi\} = \mu_p[(\varphi - \mu_p[\varphi])^2].$$

Для более узкого класса косвенных оценок $\hat{p} = F(\delta_{x,n}[f])$ естественно ожидать, что правая часть в неравенстве Рао–Крамера должна быть больше, поскольку интуитивно ясно, что одна скалярная величина $\delta_{x,n}[f]$ содержит в себе меньше информации, чем целая выборка x, и построенная на ее основе оценка будет иметь меньшую точность.

Оказывается, для косвенных оценок справедлив следующий аналог неравенства Рао–Крамера, который в действительности является равенством:

$$\operatorname{Var}\hat{p} = \frac{(1 + b_n'(p))^2}{\mathcal{I}(p)(n\operatorname{corr}_p\{f, \varphi\}^2 + O(1))} + b_n(p)^2, \tag{11}$$

где $O(1)$ — ограниченная при $n \to \infty$ величина. Точнее, имеет место

Теорема 1. *Если $\varphi, f \in B(X)$ и некоторая функция $F \in C^3(\mathbb{R})$ удовлетворяет условию $F'(\mu_p[f]) \neq 0$, то для косвенной статистической оценки $\hat{p} = F(\delta_{x,n}[f])$ выполняется равенство (11).*

Доказательство. Положим

$$S_n\varphi(x) = \varphi(x_1) + \ldots + \varphi(x_n).$$

Тогда по определению смещения

$$p + b_n(p) = \int\limits_{X^n} F(\delta_{x,n}[f])\, d\mu_p^n(x) = \int\limits_{X^n} F(\delta_{x,n}[f]) e^{pS_n\varphi(x) - n\lambda(p\varphi,\mu)}\, d\mu^n(x).$$

Продифференцируем это равенство по p:

$$1 + b_n'(p) = \int\limits_{X^n} F(\delta_{x,n}[f])(S_n\varphi(x) - n\mu_p[\varphi])\, d\mu_p^n(x) = \mathrm{cov}_p\{\hat{p}, S_n\varphi\},$$

а затем возведем в квадрат:

$$\begin{aligned}(1 + b_n'(p))^2 &= \mathrm{cov}_p\{\hat{p},\hat{p}\}\, \mathrm{cov}_p\{S_n\varphi, S_n\varphi\}\, \mathrm{corr}_p\{\hat{p}, S_n\varphi\}^2 = \\ &= \mathrm{cov}_p\{\hat{p},\hat{p}\}\, n\mathcal{I}(p)\, \mathrm{corr}_p\{\hat{p}, S_n\varphi\}^2. \end{aligned} \tag{12}$$

Далее нам потребуется следующая лемма, которую мы докажем чуть позже.

Лемма 3. *Если $F, G \in C^3(\mathbb{R})$, то для любых функций $f, g \in B(X)$*

$$n\, \mathrm{cov}_p\{F(\delta_{x,n}[f]), G(\delta_{x,n}[g])\} = F'(\mu_p[f]) G'(\mu_p[g])\, \mathrm{cov}_p\{f, g\} + O(1/n), \tag{13}$$

а если еще $F'(\mu_p[f]) \neq 0$ и $G'(\mu_p[g]) \neq 0$, то

$$\mathrm{corr}_p\{F(\delta_{x,n}[f]), G(\delta_{x,n}[g])\}^2 = \mathrm{corr}_p\{f, g\}^2 + O(1/n). \tag{14}$$

Возьмем равенство (14) для функций $g = \varphi$ и $G(y) \equiv y$ (когда функция $G(\delta_{x,n}[g])$ заменяется на $\delta_{x,n}[\varphi]$) и подставим его в (12). У нас получится равенство

$$\mathrm{cov}_p\{\hat{p},\hat{p}\} = \frac{(1 + b_n'(p))^2}{n\mathcal{I}(p)(\mathrm{corr}_p\{f,\varphi\}^2 + O(1/n))}.$$

Складывая его с известным тождеством $\mathrm{Var}\,\hat{p} = \mathrm{cov}_p\{\hat{p},\hat{p}\} + b_n(p)^2$, получаем (11).

Докажем теперь лемму 3. Поскольку функции f и g ограничены, существует отрезок $[A, B]$, содержащий все их значения. В том же отрезке будут находиться значения выражений $\delta_{x,n}[f]$ и $\delta_{x,n}[g]$, а также $\mu_p[f]$ и $\mu_p[g]$. Используем формулу Тейлора с интегральным остаточным членом третьей степени:

$$F(y) = F(\mu) + F'(\mu)(y - \mu) + \frac{F''(\mu)}{2}(y - \mu)^2 + (y - \mu)^3 \int\limits_0^1 \frac{(1-t)^2}{2} F'''((1-t)\mu + ty)\, dt.$$

Из нее следует, что

$$\begin{aligned} F(\delta_{x,n}[f]) &= F(\mu_p[f]) + F'(\mu_p[f])(\delta_{x,n}[f] - \mu_p[f]) + \\ &+ \frac{F''(\mu_p[f])}{2}(\delta_{x,n}[f] - \mu_p[f])^2 + O((\delta_{x,n}[f] - \mu_p[f])^3), \end{aligned} \tag{15}$$

$$\begin{aligned} G(\delta_{x,n}[g]) &= G(\mu_p[g]) + G'(\mu_p[g])(\delta_{x,n}[g] - \mu_p[g]) + \\ &+ \frac{G''(\mu_p[g])}{2}(\delta_{x,n}[g] - \mu_p[g])^2 + O((\delta_{x,n}[g] - \mu_p[g])^3), \end{aligned} \tag{16}$$

где

$$|O((\delta_{x,n}[f] - \mu_p[f])^3)| \leq \frac{1}{2} \sup_{y \in [A,B]} |F'''(y)| |\delta_{x,n}[f] - \mu_p[f]|^3,$$

$$|O((\delta_{x,n}[g] - \mu_p[g])^3)| \leq \frac{1}{2} \sup_{y \in [A,B]} |G'''(y)| |\delta_{x,n}[g] - \mu_p[g]|^3.$$

Вычислим попарные ковариации различных слагаемых в тейлоровских разложениях (15) и (16). Ковариация константы с любой случайной величиной равна нулю. Поэтому начнем с ковариации слагаемых первой степени:

$$\operatorname{cov}_p\{F'(\mu_p[f])(\delta_{x,n}[f]-\mu_p[f]), G'(\mu_p[g])(\delta_{x,n}[g]-\mu_p[g])\} =$$

$$= \frac{1}{n} F'(\mu_p[f]) G'(\mu_p[g]) \operatorname{cov}_p\{f,g\}. \tag{17}$$

Ковариация величин $\delta_{x,n}[f]-\mu_p[f]$ и $(\delta_{x,n}[g]-\mu_p[g])^2$ вычисляется чуть сложнее:

$$\operatorname{cov}_p\{\delta_{x,n}[f]-\mu_p[f], (\delta_{x,n}[g]-\mu_p[g])^2\} =$$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mu_p^n[(f(x_i)-\mu_p[f])(g(x_j)-\mu_p[g])(g(x_k)-\mu_p[g])] =$$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\mu_p^n[(f(x_i)-\mu_p[f])(g(x_i)-\mu_p[g])^2] = \frac{1}{n^2}\mu_p[(f-\mu_p[f])(g-\mu_p[g])^2].$$

Очевидно, она имеет порядок малости $1/n^2$.

Далее оценим ковариацию случайных величин $(\delta_{x,n}[f]-\mu_p[f])^2$ и $(\delta_{x,n}[g]-\mu_p[g])^2$. В силу неравенства Коши–Буняковского и леммы 2 при $k=4$

$$\mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^2(\delta_{x,n}[g]-\mu_p[g])^2] \le$$

$$\le \mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^4]^{1/2}\mu_p^n[(\delta_{x,n}[g]-\mu_p[g])^4]^{1/2} \le \frac{C(4)(B-A)^4}{n^2},$$

где $[A,B]$ — отрезок, содержащий все значения функций f и g. Кроме того,

$$\mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^2] = \frac{1}{n}\operatorname{cov}_p\{f,f\}, \quad \mu_p^n[(\delta_{x,n}[g]-\mu_p[g])^2] = \frac{1}{n}\operatorname{cov}_p\{g,g\}.$$

Отсюда видно, что ковариация

$$\operatorname{cov}_p\{(\delta_{x,n}[f]-\mu_p[f])^2, (\delta_{x,n}[g]-\mu_p[g])^2\} =$$

$$= \mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^2(\delta_{x,n}[g]-\mu_p[g])^2] - \mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^2]\mu_p^n[(\delta_{x,n}[g]-\mu_p[g])^2]$$

является величиной порядка $1/n^2$.

Ковариации всех остальных пар слагаемых тоже имеют порядок малости $1/n^2$ или даже меньше. Мы не будем выписывать все выкладки, доказывающие этот факт. Они хотя и не сложные, но довольно громоздкие. Приведем лишь еще один пример:

$$\operatorname{cov}_p\{\delta_{x,n}[f]-\mu_p[f], O((\delta_{x,n}[g]-\mu_p[g])^3)\} = \mu_p^n[(\delta_{x,n}[f]-\mu_p[f])O((\delta_{x,n}[g]-\mu_p[g])^3)] \le$$

$$\le \frac{1}{2}\sup_{y\in[A,B]}|G'''(y)|\,\mu_p^n[(\delta_{x,n}[f]-\mu_p[f])^4]^{1/4}\mu_p^n[(\delta_{x,n}[g]-\mu_p[g])^4]^{3/4} \le$$

$$\le \frac{1}{2}\sup_{y\in[A,B]}|G'''(y)|\,\frac{C(4)(B-A)^4}{n^2}.$$

В итоге получается, что главный вклад в ковариацию случайных величин $F(\delta_{x,n}[f])$ и $G(\delta_{x,n}[g])$, определяемый формулой (17), дают линейные члены их тейлоровских разложений, а все остальные члены дают погрешность порядка $1/n^2$. Тем самым доказано равенство (13). Из него в свою очередь вытекает (14). Лемма 3 доказана.

5. Относительно эффективные косвенные оценки. В связи с равенством (11) естественно возникает вопрос о существовании косвенных оценок $\hat{p} = F(\delta_{x,n}[f])$, для которых справедлива асимптотика

$$\operatorname{Var}\hat{p} \sim \frac{1}{n\mathcal{I}(p)\operatorname{corr}_p\{f,\varphi\}^2} \quad \text{при} \quad n \to \infty. \tag{18}$$

Мы будем называть такие оценки *относительно эффективными.* Можно сказать, что относительная эффективность обобщает понятие асимптотической эффективности для статистических оценок общего вида $\hat{p} = \hat{p}(x)$, состоящее в том, что

$$\operatorname{Var}\hat{p} \sim \frac{1}{n\mathcal{I}(p)} \quad \text{при} \quad n \to \infty.$$

Известно, что в случае статистических оценок общего вида при некоторых естественных ограничениях на семейство распределений асимптотически эффективными являются оценки максимального правдоподобия. Для косвенных оценок тоже можно доказать некий аналог этого результата. Нужно лишь заменить обычное уравнение правдоподобия на следующее:

$$\mu_{\hat{p}}[f] = \delta_{x,n}[f]. \tag{19}$$

В математической статистике функцию $\mu_p[f]$ называют моментом распределения μ_p, а решение уравнения (19) — оценкой параметра p по методу моментов. Мы докажем ниже, что эта оценка относительно эффективна.

В силу центральной предельной теоремы типичные отклонения случайной величины $\delta_{x,n}[f]$ от ее математического ожидания $\mu_p[f]$ имеют размер порядка $n^{-1/2}$. При таких отклонениях условием локальной разрешимости уравнения (19) относительно переменной $\hat{p}$ является обычное условие теоремы об обратной функции

$$\left.\frac{d\mu_{\hat{p}}[f]}{d\hat{p}}\right|_{\hat{p}=p} \neq 0.$$

Из формул (4), (5) следует, что

$$\left.\frac{d\mu_{\hat{p}}[f]}{d\hat{p}}\right|_{\hat{p}=p} = \operatorname{cov}_p\{f,\varphi\}.$$

Таким образом, если $\operatorname{cov}_p\{f,\varphi\} \neq 0$, то по теореме об обратной функции существуют такие малые окрестности $O_\varepsilon(\mu_p[f])$ и $O_\varkappa(p)$ радиусов ε и $\varkappa$ соответственно, что для всякого значения $\delta_{x,n}[f] \in O_\varepsilon(\mu_p[f])$ уравнение (19) имеет единственное решение $\hat{p} \in O_\varkappa(p)$. Если же значение $\delta_{x,n}[f]$ не попадает в $O_\varepsilon(\mu_p[f])$ (что происходит с экспоненциально малой вероятностью), то для уравнения (19) нельзя гарантировать ни существование, ни единственность решения. В таких исключительных случаях оценку $\hat{p}$ можно определять произвольно, например, положить ее равной нулю. Всюду ниже, говоря о локальном решении уравнения (19), мы будем иметь в виду величину, совпадающую с его единственным решением $\hat{p} \in O_\varkappa(p)$ в случае $\delta_{x,n}[f] \in O_\varepsilon(\mu_p[f])$ и равную нулю в случае $\delta_{x,n}[f] \notin O_\varepsilon(\mu_p[f])$.

Теорема 2. *Если* $\operatorname{cov}_p\{f,\varphi\} \neq 0$, *то в условиях теоремы 1 локальное решение уравнения (19) является относительно эффективной косвенной оценкой.*

Доказательство. В силу теоремы об обратной функции локальное решение уравнения (19) имеет вид

$$\hat{p} = \begin{cases} p + \dfrac{\delta_{x,n}[f] - \mu_p[f]}{\mathrm{cov}_p\{f,\varphi\}} + O((\delta_{x,n}[f] - \mu_p[f])^2), & \text{если} \quad \delta_{x,n}[f] \in O_\varepsilon(\mu_p[f]), \\ 0, & \text{если} \quad \delta_{x,n}[f] \notin O_\varepsilon(\mu_p[f]). \end{cases}$$

Определим функцию $\chi(x)$, принимающую единичное значение при $\delta_{x,n}[f] \in O_\varepsilon(\mu_p[f])$ и обращающуюся в нуль при $\delta_{x,n}[f] \notin O_\varepsilon(\mu_p[f])$. Тогда разность $\hat{p} - p$ можно записать в виде

$$\hat{p} - p = \xi + \eta_1 + \eta_2 + \eta_3, \tag{20}$$

где

$$\xi = \frac{\delta_{x,n}[f] - \mu_p[f]}{\mathrm{cov}_p\{f,\varphi\}}, \quad \eta_1 = O((\delta_{x,n}[f] - \mu_p[f])^2)\chi(x),$$

$$\eta_2 = -\frac{\delta_{x,n}[f] - \mu_p[f]}{\mathrm{cov}_p\{f,\varphi\}}(1 - \chi(x)), \quad \eta_3 = -p(1 - \chi(x)).$$

Очевидно,

$$\mu_p^n[\xi^2] = \frac{\mathrm{cov}_p\{f,f\}}{n\,\mathrm{cov}_p\{f,\varphi\}^2} = \frac{1}{n\mathcal{I}(p)\,\mathrm{corr}_p\{f,\varphi\}^2}. \tag{21}$$

Для остальных слагаемых из лемм 1 и 2 вытекают оценки

$$\mu_p^n[\eta_1^2] = O(n^{-2}), \quad \mu_p^n[\eta_2^2] \le \frac{2\sigma^2 e^{-n\varepsilon^2/2\sigma^2}}{\mathrm{cov}_p\{f,\varphi\}^2}, \quad \mu_p^n[\eta_3^2] \le 2p^2 e^{-n\varepsilon^2/2\sigma^2},$$

где $\sigma = \sup f - \inf f$. Объединяя их с равенствами (20) и (21), получаем требуемую асимптотику (18). Теорема доказана.

На самом деле асимптотика (18) для оценки по методу моментов и ее многомерные аналоги широко известны; их можно найти, например, в работах [1, 2].

В заключение отметим, что при $f = \varphi$ “измерение прибора” $\delta_{x,n}[f]$ превращается в достаточную статистику $\delta_{x,n}[\varphi]$, уравнение (19) становится обычным уравнением правдоподобия, а его решение (то есть оценка максимального правдоподобия) будет асимптотически эффективной оценкой для параметра p. При этом уравнение правдоподобия имеет не более одного решения, поскольку

$$\frac{d\mu_p[\varphi]}{dp} = \mathrm{cov}_p\{\varphi,\varphi\} > 0.$$

Работа выполнена при финансовой поддержке Белорусского республиканского фонда фундаментальных исследований (проект Ф12Р–061).

Литература

1. *Borovkov A.A.* Mathematical Statistics. Gordon & Breach, 1998.
2. *Verbeek M.* A Guide to Modern Econometrics. John Wiley, 2004.

V. I. Bakhtin

Estimators for an exponential family of distributions based on indirect observations

Summary

We introduce the notions of indirect estimator and its relative efficiency. In the case of the simplest exponential family of distributions we prove an analog to the Rao–Cramer inequality for the indirect estimators and show that the indirect estimator obtained by the method of moments is relatively efficient.