ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

О ДЕФЕКТЕ НЕКОТОРЫХ ОЦЕНОК, ОСНОВАННЫХ НА ВЫБОРКАХ СЛУЧАЙНОГО ОБЪЕМА

Бенинг В.Е.
МГУ им. М.В. Ломоносова, Институт проблем информатики РАН, г. Москва

Поступила в редакцию 11.09.2014, после переработки 10.10.2014.

В работе доказаны общие теоремы, позволяющие получать асимптотические дефекты оценок, основанных на выборках случайного объема. Рассмотрены два примера, касающиеся оценок дисперсии, обычно используемых в математической статистике.

Ключевые слова: оценка, асимптотический дефект, выборка случайного объема, распределение Пуассона.

Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 5–14.

1. Введение

Рассмотрим задачу статистического оценивания известной параметрической функции $g(\theta)$, зависящей от неизвестного параметра θ, и обозначим через $m(n)$ необходимое число наблюдений, которое требуется оценке $\delta_{m(n)}(X_1, \ldots, X_{m(n)})$ для достижения такого же качества (например, среднеквадратичного отклонения или дисперсии), что и оценке $\delta_n^*(X_1, \ldots, X_n)$, основанной на n наблюдениях $X_1, \ldots, X_n$. Мы рассматриваем асимптотический подход, означающий, что $n \to \infty$. Под асимптотической относительной эффективностью (АОЭ) оценки $\delta_n(X_1, \ldots, X_n)$ по отношению к оценке $\delta_n^*(X_1, \ldots, X_n)$ понимается предел (в случае его существования и независимости от последовательности $m(n)$) вида (см., например, [3] стр. 305)

$$e \equiv \lim_{n\to\infty} \frac{n}{m(n)}.$$

Например, предположим, что $e = 1/3$, тогда при больших значениях числа наблюдений n величина $m(n)$ приближенно равна $3n$, поэтому оценка $\delta_n(X_1, \ldots, X_n)$ для достижения такого же качества, что и оценка $\delta_n^*(X_1, \ldots, X_n)$, требует примерно в три раза больше наблюдений.

Вместо отношения необходимого числа наблюдений, естественно, можно было бы рассматривать разность вида $m(n) - n$, которая тоже имеет наглядный смысл необходимого дополнительного числа наблюдений, требующихся оценке $\delta_n(X_1, \ldots, X_n)$. Однако, исторически сложилось так, что многие авторы сначала исследовали асимптотические свойства отношения $n/m(n)$ (возможно, в силу относительной простоты его поведения).

Впервые общее асимптотическое исследование поведения разности $m(n) - n$ было предпринято в 1970 году Ходжесом и Леманом (см. работу [1]). Они назвали разность $m(n) - n$ дефектом (deficiency) конкурирующей оценки δ_n относительно оценки δ_n^* и предложили обозначение

$$d_n = m(n) - n. \tag{1.1}$$

Если предел $\lim_{n\to\infty} d_n$ существует, то он называется *асимптотическим дефектом* оценки δ_n относительно оценки δ_n^* и обозначается символом d. Часто d называют просто дефектом δ_n относительно δ_n^*. Заметим, что если АОЭ $e \neq 1$, то $d = \infty$ и этот случай малоинтересен. В работе [1] также было отмечено, что существуют статистические задачи, в которых типичным образом возникает случай $e = 1$ (см., например, книгу [4]), то есть в этом случае понятие АОЭ не дает ответа на вопрос какая оценка лучше и понятие дефекта проясняет эту ситуацию, поскольку в этом случае асимптотический дефект может, в принципе, быть любым.

Предположим, например, что $d = 5$. Тогда для больших значений n величина $m(n)$ равна приближенно $n + 5$. Чтобы получить ту же величину критерия качества оценке δ_n требуется примерно на пять наблюдений больше, чем оценке δ_n^*.

Таким образом, дефект оценки δ_n относительно оценки δ_n^* показывает, сколько добавочных наблюдений требуется, если мы настаиваем на использовании оценки δ_n вместо оценки δ_n^*, и поэтому создает естественный базис для их асимптотического сравнения в случае $e = 1$. Исследование асимптотического поведения дефекта d_n технически более сложно, чем нахождение предела e. Часто оно требует построения асимптотических разложений (а.р.) для соответсвующих функций, характеризующих качество оценок.

Обозначим функции риска оценок δ_n и δ_n^*, соответственно, через

$$R_n(\theta) = \mathsf{E}_\theta \left(\delta_n(X_1,\ldots,X_n) - g(\theta)\right)^2, \quad R_n^*(\theta) = \mathsf{E}_\theta \left(\delta_n^*(X_1,\ldots,X_n) - g(\theta)\right)^2,$$

где $g(\theta)$ – оцениваемая функция, а θ – неизвестный параметр (произвольной природы), тогда по определению величины $d_n(\theta) \equiv d_n = m(n) - n$, для каждого n должно выполняться равенство

$$R_n^*(\theta) = R_{m(n)}(\theta). \tag{1.2}$$

При решении уравнения (1.2) целочисленную величину $m(n)$ следует рассматривать как переменную, принимающую поизвольные действительные значения. Для этого можно определить функцию риска $R_{m(n)}(\theta)$ для нецелых значений $m(n)$ по формуле

$$R_{m(n)}(\theta) = \big(1 - m(n) + [m(n)]\big)\, R_{[m(n)]}(\theta) + \big(m(n) - [m(n)]\big)\, R_{[m(n)]+1}(\theta)$$

(см. работу [1]).

Типичным образом функции риска $R_n^*(\theta)$ и $R_n(\theta)$ неизвестны точно и используются их аппроксимации. Предположим, что для функций риска $R_n^*(\theta)$ и $R_n(\theta)$ справедливы асимптотические разложения вида

$$R_n^* = \frac{a(\theta)}{n^r} + \frac{b(\theta)}{n^{r+s}} + o\big(n^{-r-s}\big), \tag{1.3}$$

$$R_n \;=\; \frac{a(\theta)}{n^r} \;+\; \frac{c(\theta)}{n^{r+s}} \;+\; o\big(n^{-r-s}\big), \tag{1.4}$$

где $a(\theta),\ b(\theta)$ и $c(\theta)$ – некоторые постоянные, не зависящие от n, а $r \;>\; 0,\ s \;>\; 0$ – некоторые константы, определяющие порядок убывания по n этих функций риска. Первый член в этих асимптотических разложениях одинаков и это отражает тот факт, что АОЭ этих оценок равна единице. Из соотношений (1.1) – (1.4) легко получить, что (см. работу [1] или книгу [3], стр. 310)

$$d_n(\theta) \;\equiv\; \frac{c(\theta) \;-\; b(\theta)}{r\; a(\theta)}\; n^{1-s} \;+\; o\big(n^{1-s}\big). \tag{1.5}$$

Таким образом, асимптотический дефект имеет вид

$$d(\theta) \;\equiv\; d \;=\; \begin{cases} \pm\infty, & 0 \;<\; s \;<\; 1, \\ \dfrac{c(\theta) \;-\; b(\theta)}{r\; a(\theta)}, & s \;=\; 1, \\ 0, & s \;>\; 1. \end{cases} \tag{1.6}$$

Аимптотичекий дефект обладает следующим очевидным свойством транзитивности: если дана третья оценка $\bar{\delta}_n$, имеющая функцию риска $\bar{R}_n(\theta)$, которая имеет а.р. типа (1.4), тогда дефект d оценки $\bar{\delta}_n$ относительно оценки δ^*_n удовлетворяет равенству

$$d \;=\; d_1 \;+\; d_2,$$

где d_1 – дефект оценки $\bar{\delta}_n$ относительно оценки δ_n и d_2 – дефект оценки δ_n относительно оценки δ^*_n.

Случай, когда выполняется равенство $s = 1$, представляется наиболее интересным, поскольку в этом случае асимптотический дефект конечен. Ходжес и Леман в работе [1] привели ряд простых примеров, показывающих естественность возникновения этого случая в математической статистике.

В настоящей работе приведены примеры вычисления асимптотического дефекта в задачах оценивания с помощью оценок, основанных на выборках случайного объема. Проведено асимптотическое сравнение качества оценок, основанных на выборках случайного и неслучайного объемов.

2. Оценки, основанные на выборках случайного объема

Рассмотрим случайные величины (с.в.) $N_1, N_2, \ldots$ и $X_1, X_2, \ldots$, заданные на одном и том же вероятностном пространстве $(\Omega,\ \mathcal{A},\ \mathsf{P})$. При этом случайные величины $X_1, X_2, \ldots X_n$ имеют смысл статистических наблюдений, а случайная величина N_n трактуется как случайный объем выборки, зависящий от натурального параметра $n \in \mathbb{N}$. Предположим, что для каждого $n \geqslant 1$ с.в. N_n принимает только натуральные значения (т.е., $N_n \;\in \mathbb{N}$) и не зависит от последовательности с.в. $X_1, X_2, \ldots$. Всюду далее предполагается, что случайные величины $X_1, X_2, \ldots$ независимы, одинаково распределены и имеют распределение, зависящее от неизвестного параметра $\theta \;\in\; \Theta$, при этом множество Θ может иметь произвольную природу.

Для каждого $n \geqslant 1$ обозначим через $T_n = T_n(X_1, \ldots, X_n)$ некоторую статистику, то есть действительную измеримую функцию, зависящую от наблюдений $X_1, \ldots, X_n$. Для каждого $n \geqslant 1$ определим статистику T_{N_n}, зависящую от выборки случайного объема как

$$T_{N_n}(\omega) \;\equiv\; T_{N_n(\omega)}(X_1(\omega), \ldots, X_{N_n(\omega)}(\omega)), \;\; \omega \;\in\; \Omega.$$

Далее под статистикой, указанной выше, будем понимать оценку действительной известной функции $g(\theta)$, зависящей от неизвестного параметра $\theta \;\in\; \Theta$, и будем обозначать ее символами типа $\delta_n(X_1, \ldots, X_n)$. Справедлива следующая теорема.

Теорема 1. *1. Если $\delta_n \;=\; \delta_n(X_1, \ldots, X_n)$ – любая несмещенная оценка функции $g(\theta)$, то есть справедливо тождество*

$$\mathsf{E}_\theta \; \delta_n \;\;\equiv\;\; g(\theta), \;\;\; \theta \;\in\; \Theta$$

и $\delta_{N_n} \;\equiv\; \delta_{N_n}(X_1, \ldots, X_{N_n})$ – оценка, построенная по выборке случайного объема, то она также является несмещенной оценкой функции $g(\theta)$, то есть

$$\mathsf{E}_\theta \; \delta_{N_n} \;\;\equiv\;\; g(\theta), \;\;\; \theta \;\in\; \Theta.$$

2. Предпопложим, что существуют числа $a(\theta)$, $b(\theta)$ и $C(\theta) \;>\; 0$, $\alpha \;>\; 0$, $r \;>\; 0$, $s \;>\; 0$ такие, что для функции риска оценки $\delta_n(X_1, \ldots, X_n)$ справедливо соотношение

$$\Bigl|R_n^*(\theta) \;-\; \frac{a(\theta)}{n^r} \;-\; \frac{b(\theta)}{n^{r+s}}\Bigr| \;\leqslant\; \frac{C(\theta)}{n^{r+s+\alpha}},$$

где

$$R_n^*(\theta) \;=\; \mathsf{E}_\theta\bigl(\delta_n(X_1, \ldots, X_n) \;-\; g(\theta)\bigr)^2,$$

тогда для функции риска оценки, построенной по выборке случайного объема, выполнено неравенство

$$\Bigl|R_n(\theta) \;-\; a(\theta) \; \mathsf{E} \; N_n^{-r} \;-\; b(\theta) \; \mathsf{E} \; N_n^{-r-s}\Bigr| \;\leqslant\; C(\theta) \; \mathsf{E} \; N_n^{-r-s-\alpha},$$

где

$$R_n(\theta) \;=\; \mathsf{E}_\theta\bigl(\delta_{N_n}(X_1, \ldots, X_{N_n}) \;-\; g(\theta)\bigr)^2.$$

Следствие 1. *Пусть существуют числа $a(\theta)$, $b(\theta)$ and $r \;>\; 0$, $s \;>\; 0$ такие, что*

$$R_n^*(\theta) \;=\; \frac{a(\theta)}{n^r} \;+\; \frac{b(\theta)}{n^{r+s}},$$

тогда

$$R_n(\theta) \;=\; a(\theta) \; \mathsf{E} \; N_n^{-r} \;+\; b(\theta) \; \mathsf{E} \; N_n^{-r-s}.$$

Доказательство. Доказательство теоремы непосредственно следует из формулы полной вероятности:

1. $$\mathsf{E}_\theta \; \delta_{N_n} \;=\; \sum_{k=1}^{\infty} \mathsf{E}_\theta \; \delta_k \; \mathsf{P} \; \bigl(N_n \;=\; k\bigr) \;=\;$$

$$= \sum_{k=1}^{\infty} g(\theta)\ \mathsf{P}\left(N_n = k\right) = g(\theta) \sum_{k=1}^{\infty} \mathsf{P}\left(N_n = k\right) = g(\theta), \quad \theta \in \Theta.$$

2. $$\left|R_n(\theta) - a(\theta)\ \mathsf{E}\ N_n^{-r} - b(\theta)\ \mathsf{E}\ N_n^{-r-s}\right| =$$

$$= \Big|\sum_{k=1}^{\infty} \mathsf{E}_\theta \left(\delta_k - g(\theta)\right)^2 \mathsf{P}\left(N_n = k\right) - a(\theta) \sum_{k=1}^{\infty} \frac{1}{k^r}\ \mathsf{P}\left(N_n = k\right) -$$

$$- b(\theta) \sum_{k=1}^{\infty} \frac{1}{k^{r+s}}\ \mathsf{P}\left(N_n = k\right)\Big| =$$

$$= \Big|\sum_{k=1}^{\infty} \left(\mathsf{E}_\theta \left(\delta_k - g(\theta)\right)^2 - \frac{a(\theta)}{k^r} - \frac{b(\theta)}{k^{r+s}}\right) \mathsf{P}\left(N_n = k\right)\Big| \leqslant$$

$$\leqslant \sum_{k=1}^{\infty} \left|\mathsf{E}_\theta \left(\delta_k - g(\theta)\right)^2 - \frac{a(\theta)}{k^r} - \frac{b(\theta)}{k^{r+s}}\right| \mathsf{P}\left(N_n = k\right) \leqslant$$

$$\leqslant \sum_{k=1}^{\infty} \frac{C(\theta)}{k^{r+s+\alpha}}\ \mathsf{P}\left(N_n = k\right) =$$

$$= C(\theta)\ \mathsf{E}\ N_n^{-r-s-\alpha}.$$

□

Приведем пример применения этой теоремы. Предположим, что исходные наблюдения $X_1, \ldots, X_n$ имеют математическое ожидание

$$\mathsf{E}_\theta\ X_1 = g(\theta)$$

и дисперсию

$$\mathsf{D}_\theta\ X_1 = \sigma^2(\theta).$$

Обычная оценка для неизвестного математического ожидания $g(\theta)$, основанная на n наблюдениях, имеет вид

$$\delta_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{2.1}$$

Это несмещенная состоятельная оценка и ее дисперсия равна

$$R_n^*(\theta) = \mathsf{D}_\theta\ \delta_n = \frac{\sigma^2(\theta)}{n}. \tag{2.2}$$

Если рассмотреть эту оценку, но уже использовать выборку случайного объема, то согласно Следствию 1 имеем

$$R_n(\theta) = \mathsf{D}_\theta\ \delta_{N_n}(X_1, \ldots, X_{N_n}) = \sigma^2(\theta)\ \mathsf{E}\ N_n^{-1}. \tag{2.3}$$

Если математическое ожидание $g(\theta)$ известно, то рассмотрим оценку для неизвестной дисперсии $\sigma^2(\theta)$ в виде

$$\bar{\delta}_n \;=\; \frac{1}{n} \sum_{i=1}^{n} (X_i \;-\; g(\theta))^2. \tag{2.4}$$

Это также несмещенная состоятельная оценка, имеющая дисперсию, равную

$$\bar{R}_n^*(\theta) \;=\; \mathsf{D}_\theta \; \bar{\delta}_n \;=\; \frac{\mu_4(\theta) \;-\; \sigma^4(\theta)}{n}, \quad \mu_4(\theta) \;=\; \mathsf{E}_\theta \; (X_1 \;-\; g(\theta))^4. \tag{2.5}$$

Для этой оценки, основанной на выборке случайного объема, имеем

$$\bar{R}_n(\theta) \;=\; \mathsf{D}_\theta \; \bar{\delta}_{N_n}(X_1, \ldots, X_n) \;=\; \big(\mu_4(\theta) \;-\; \sigma^4(\theta)\big) \; \mathsf{E} \; N_n^{-1}. \tag{2.6}$$

Предположим теперь, что математическое ожидание $g(\theta)$ неизвестно и для оценки неизвестной дисперсии вместо оценки (2.4) используется оценка (см. (2.1))

$$\tilde{\delta}_n^{(\gamma)} \;\equiv\; \tilde{\delta}_n \;=\; \frac{1}{n+\gamma} \sum_{i=1}^{n} \big(X_i \;-\; \delta_n\big)^2, \quad \gamma \;\in\; \mathbb{R}. \tag{2.7}$$

Если $\gamma \;\neq\; -1$, то эта оценка не является несмещенной, но может иметь меньшее среднеквадратичное отклонение, чем несмещенная оценка с $\gamma \;=\; -1$.

Нетрудно получить (см. [1], (3.6) и [2]), что справедливо равенство

$$\tilde{R}_n^*(\theta) \;=\; \mathsf{E}_\theta \; \big(\tilde{\delta}_n(X_1, \ldots, X_n) \;-\; \sigma^2(\theta)\big)^2 \;=$$

$$=\; \frac{\sigma^4(\theta)}{n(n \;+\; \gamma)^2} \; \Big((n \;-\; 1) \; \big((\mu_4(\theta)/\sigma^4(\theta) \;-\; 1) \; (n \;-\; 1) \;+\; 2\big) \;+ n \; (\gamma \;+\; 1)^2\Big) \tag{2.8}$$

и значит

$$\tilde{R}_n^*(\theta) \;=\; \sigma^4(\theta) \; \Big(\frac{\mu_4(\theta)/\sigma^4(\theta) \;-1}{n} \;+$$

$$+\; \frac{(\gamma \;+\; 1)^2 \;-\; 2 \; (\mu_4(\theta)/\sigma^4(\theta) \;-1) \;+\; 2 \;-2\gamma(\mu_4(\theta)/\sigma^4(\theta) \;-1)}{n^2}\Big) \;+\; O\big(n^{-3}\big). \tag{2.9}$$

Используя Теорему 1, теперь имеем

$$\tilde{R}_n(\theta) \;=\; \mathsf{E}_\theta \; \big(\tilde{\delta}_{N_n}(X_1, \ldots, X_{N_n}) \;-\; \sigma^2(\theta)\big)^2 \;=$$

$$=\; \sigma^4(\theta) \; \Big((\mu_4(\theta)/\sigma^4(\theta) \;-1) \; \mathsf{E} \; N_n^{-1} \;+$$

$$+ \big((\gamma + 1)^2 \;-\; 2 \; (\mu_4(\theta)/\sigma^4(\theta) \;-1) \;+\; 2 \;-2\gamma(\mu_4(\theta)/\sigma^4(\theta) \;-1)\big) \; \mathsf{E} \; N_n^{-2}\Big) \;+\; O\big(\mathsf{E} \; N_n^{-3}\big). \tag{2.10}$$

3. Дефекты некоторых оценок, основанных на выборках случайного объема

В этом разделе мы применим результаты раздела 2 к вычислению дефектов оценок, рассмотренных выше, основанных на выборках, объем которых имеет распределение Пуассона. Пусть M_n – случайная величина, имеющая распределение

Пуассона с параметром $n - 1$, $n \geqslant 2$, то есть

$$\mathsf{P}\big(M_n = k\big) = e^{1-n} \frac{(n - 1)^k}{k!}, \quad k = 0, 1, \ldots$$

Определим случайный объем выборки по формуле

$$N_n = M_n + 1,$$

тогда в среднем объем случайной выборки равен n (объем неслучайной выборки), то есть

$$\mathsf{E}\, N_n = n$$

и

$$\mathsf{E}\, N_n^{-1} = e^{1-n} \sum_{k=0}^{\infty} \frac{(n - 1)^k}{(k+1)!} = \frac{1 - e^{1-n}}{n - 1}.$$

Поэтому

$$\mathsf{E}\, N_n^{-1} = \frac{1}{n} + \frac{1}{n^2} + o\big(n^{-2}\big). \tag{3.1}$$

Дефект оценки δ_{N_n} относительно оценки δ_n (см. (2.1)), с учетом формул (2.2), (2.3), (3.1) и (1.6) с $r = s = 1$, $a(\theta) = \sigma^2(\theta)$, $b(\theta) = 0$, $c(\theta) = \sigma^4(\theta)$, равен

$$d = 1. \tag{3.2}$$

Аналогично, асимптотический дефект оценки $\bar{\delta}_{N_n}$ относительно оценки $\bar{\delta}_n$ (см. формулу (2.4)), после применения формул (2.5), (2.6), (3.1) и (1.6) с $r = s = 1$, $a(\theta) = c(\theta) = \mu_4(\theta) - \sigma^4(\theta)$, $b(\theta) = 0$, также равняется единице

$$\bar{d} = 1. \tag{3.3}$$

Рассмотрим теперь третий пример из раздела 2 (см. (2.7)). Нетрудно видеть, что справедливы равенства

$$\mathsf{E}\, N_n^{-2} = e^{1-n} \sum_{k=0}^{\infty} \frac{(n - 1)^k}{(k+1)^2 k!} = \frac{e^{1-n}}{n - 1} \sum_{k=1}^{\infty} \frac{(n - 1)^k}{k\, k!} =$$

$$= \frac{e^{1-n}}{n - 1} \int_0^{n-1} \frac{e^x - 1}{x}\, d\, x.$$

Поэтому, применяя правило Лопиталя, получим

$$\int_0^{n-1} \frac{e^x - 1}{x}\, d\, x \sim \frac{e^{n-1}}{n - 1}, \quad n \to \infty$$

и значит справедлива асимптотическая формула

$$\mathsf{E}\, N_n^{-2} \sim \frac{1}{n^2}, \quad n \to \infty. \tag{3.4}$$

Теперь дефект оценки $\tilde{\delta}_{N_n}$ относительно оценки $\tilde{\delta}_n$ (см. формулу (2.7) и см. формулы (2.9), (2.10), (3.4) и (1.6) с $r = s = 1$) равен

$$\tilde{d} = 1 \tag{3.5}$$

и дефект оценки $\tilde\delta_{N_n}^{(\gamma_1)}$ относительно оценки $\tilde\delta_{N_n}^{(\gamma_2)}$ (см. (2.7), (3.1), (3.4) и (1.6) с $r = s = 1$) равен

$$\tilde d_{\gamma_1,\gamma_2} = (\gamma_1 - \gamma_2)\left(\frac{\gamma_1 + \gamma_2 + 2}{\mu_4(\theta)/\sigma^4(\theta) - 1} - 2\right). \tag{3.6}$$

Таким образом, классическая оценка $\tilde\delta_{N_n}^{(0)}$ «лучше» оценки $\tilde\delta_{N_n}^{(-1)}$, когда справедливо неравенство

$$\frac{\mu_4(\theta)}{\sigma^4(\theta)} - 1 > \frac{1}{2},$$

в противном случае она «хуже». В частности, если наблюдение X_1 имеет нормальное распределение, то

$$\frac{\mu_4(\theta)}{\sigma^4(\theta)} - 1 = 2$$

и

$$\tilde d_{\gamma_1,\gamma_2} = \frac{1}{2}\,(\gamma_1 - \gamma_2)\,(\gamma_1 + \gamma_2 - 2). \tag{3.7}$$

Поэтому можно «сэкономить» $3/2$ наблюдений, используя смещенную оценку $\tilde\delta_{N_n}^{(0)}$. Наилучшее (в этом смысле) значение γ в нормальном случае есть $\gamma = 1$, $\tilde d_{0,1} = 2$, которое обеспечивает экономию примерно в $1/2$ наблюдения.

Эти примеры иллюстрируют следующую теорему.

Теорема 2. *Пусть существуют числа $a(\theta)$, $b(\theta)$ и k_1, k_2 такие, что справедливы соотношения*

$$R_n^*(\theta) = \mathsf{E}_\theta\big(\delta_n(X_1,\ldots,X_n) - g(\theta)\big)^2 = $$
$$= \frac{a(\theta)}{n} + \frac{b(\theta)}{n^2} + o\big(n^{-2}\big)$$

и

$$\mathsf{E}\ N_n^{-1} = \frac{1}{n} + \frac{k_1}{n^2} + o\big(n^{-2}\big),$$
$$\mathsf{E}\ N_n^{-2} = \frac{k_2}{n^2} + o\big(n^{-2}\big),$$
$$\mathsf{E}\ N_n^{-3} = o\big(n^{-2}\big),$$

тогда для асимптотического дефекта оценки $\delta_{N_n}(X_1,\ldots,X_{N_n})$ относительно оценки $\delta_n(X_1,\ldots,X_n)$ справедливо равенство

$$d(\theta) = \frac{k_1\ a(\theta) + b(\theta)\ k_2 - b(\theta)}{a(\theta)}.$$

Доказательство близко к доказательству Теоремы 1 (с учетом формулы (1.6)).

Заключение

Таким образом, в работе рассмотрен случай статистического оценивания неизвестного параметра, основанный на оценках, построенных по выборкам случайного объема. С помощью понятия дефекта проведено асимптотическое сравнение

качества таких оценок. Получены явные формулы для асимптотического дефекта, имеющего смысл необходимого добавочного числа наблюдений. Подробно рассмотрен случай, когда случайный объем выборки имеет распределение Пуассона.

Список литературы

[1] Hodges J.L., Lehmann E.L. Deficiency // Annals of the Institute of Statistical Mathematics. 1970. Vol. 41, № 5. Pp. 783 – 801.

[2] Крамер Г. Математические методы статистики. М.: Мир, 1975. 648 с.

[3] Леман Э. Теория точечного оценивания. М.: Наука, 1991. 448 с.

[4] Bening V.E. Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency. VSP, Utrecht, 2000. 277 p.

Библиографическая ссылка

Бенинг В.Е. О дефекте некоторых оценок, основанных на выборках случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 5–14.

Сведения об авторах

1. **Бенинг Владимир Евгеньевич**
профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова, старший научный сотрудник ИПИ РАН.

Россия, 119992, г. Москва, ГСП-1, Воробьевы горы, МГУ им. М.В. Ломоносова.
E-mail: bening@yandex.ru.

ON DEFICIENCIES OF SOME ESTIMATORS BASED ON SAMPLES OF RANDOM SIZE

Bening Vladimir Evgenyevich
Professor of Mathematical Statistics department,
Lomonosov Moscow State University
Russia, 119992, Moscow, GSP-1, Vorobyovi gory, Lomonosov MSU.
E-mail: bening@yandex.ru

Received 10.09.2014, revised 10.10.2014.

In the paper general theorems concerning the asymptotic deficiencies of estimators based on the sample of random size are proved. Two examples of variance estimation are presented.

Keywords: point estimator, asymptotic deficiency, sample with random size, Poisson distribution.

Bibliographic citation

Bening V.E. On deficiencies of some estimators based on samples of random size. *Vestnik TvGU. Seriya: Prikladnaya matematika* [Herald of Tver State University. Series: Applied Mathematics], 2015, no. 1, pp. 5–14. (in Russian)