2. ТОЧЕЧНЫЕ ОЦЕНКИ

Одной из задач математической статистики (см. 1.2) является *оценка* неизвестных параметров выбранной *параметрической модели.*

Очень часто в приложениях рассматривают параметрическую модель. В этом случае предполагают, что закон распределения *генеральной совокупности* принадлежит множеству $\left\{F\left(x;\vec{\theta}\right): \vec{\theta} \in \Theta\right\}$, где вид функции распределения задан, а вектор параметров $\vec{\theta} = (\theta_1, \ldots, \theta_r)$ неизвестен. Требуется найти оценку для $\vec{\theta}$ или некоторой функции от него (например, математического ожидания, дисперсии) по *случайной выборке* $(X_1, \ldots, X_n)$ из генеральной совокупности X.

Например, предположим, что масса X детали имеет нормальный закон распределения, но его параметры $\theta_1 = \mu$ и $\theta_2 = \sigma^2$ неизвестны. Нужно найти приближенное значение параметров по результатам наблюдений $x_1, \ldots, x_n$, полученным в эксперименте (по реализации случайной выборки).

Как уже отмечалось (см. 1.2), в математической статистике существуют два вида *оценок*: *точечные* и *интервальные.* В этой главе будут рассмотрены точечные оценки, а интервальным оценкам посвящена следующая глава.

2.1. Состоятельные, несмещенные и эффективные оценки

Пусть $\vec{X}_n = (X_1, \ldots, X_n)$ — *случайная выборка* из *генеральной совокупности* X, функция распределения $F\left(x;\theta\right)$ которой известна, а θ — неизвестный параметр, т.е. рассматривается *параметрическая модель* $\left\{F\left(x;\theta\right), \theta \in \Theta\right\}$ (для простоты изложения будем считать пока, что θ — скаляр).

Требуется построить *статистику* $\widehat{\theta}(\vec{X}_n)$, которую можно было бы принять в качестве *точечной оценки* параметра θ.

Интуитивно ясно, что в качестве *оценки* параметра θ можно использовать различные статистики. Например, в качестве точечной оценки для $\mu = \mathrm{M}\,X$ можно предложить такие статистики:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \theta^*(\vec{X}_n) = \frac{X_{(1)} + X_{(n)}}{2},$$

$$\widetilde{\theta}(\vec{X}_n) = \begin{cases} \frac{1}{2}\Big(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\Big), & n \text{ — четное;} \\ X_{(\frac{n+1}{2})}, & n \text{ — нечетное.} \end{cases}$$

Какую же из этих статистик предпочесть? В общем случае нужно дать ответ на вопрос: какими свойствами должна обла-дать статистика $\widehat{\theta}(X_1, \ldots, X_n) = \theta(\vec{X}_n)$, чтобы она была в неко-тором смысле наилучшей оценкой параметра θ?

Заметим, что в дальнейшем, как правило, будем говорить об оценке параметра θ параметрической модели, хотя все ска-занное можно перенести и на функцию от θ.

Определение 2.1. Статистику $\widehat{\theta}(\vec{X}_n)$ называют ***состоятельной оценкой*** параметра $\theta \in \Theta$, если с ростом *объема выборки* n она сходится по вероятности к оцениваемому параметру θ, т .е.

$$\widehat{\theta}(\vec{X}_n) \xrightarrow[n\to\infty]{\mathbf{P}} \theta.$$

Иными словами, для состоятельной оценки $\widehat{\theta}(\vec{X}_n)$ отклонение ее от θ на величину ε и более становится маловероятным при большом объеме выборки. Это свойство оценки является очень важным, ибо несостоятельная оценка практически бесполезна. Однако следует отметить, что на практике приходится оценивать неизвестные параметры и при малых объемах выборки.

Естественным является то требование, при выполнении которого оценка не дает систематической погрешности в сторону завышения (или занижения) истинного значения параметра θ.

Определение 2.2. Статистику $\widehat{\theta}(\vec{X}_n)$ называют ***несмещенной оценкой*** параметра θ, если ее математическое ожидание совпадает с θ, т.е. $\mathbf{M}\widehat{\theta}(\vec{X}_n) = \theta$ для любого фиксированного n.

Если оценка является ***смещенной*** (т.е. последнее равенство не имеет места), то величина смещения $b_n(\theta) = \mathbf{M}\widehat{\theta}(\vec{X}_n) - \theta$. Как мы увидим далее, смещение оценки часто можно устранить, введя соответствующую поправку.

Говорят также, что ***оценка*** $\widehat{\theta}(\vec{X}_n)$ является ***асимптотически несмещенной***, если при $n \to \infty$ она сходится по вероятности к своему математическому ожиданию, т.е. для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\widehat{\theta}(\vec{X}_n) - \mathbf{M}\widehat{\theta}(\vec{X}_n)\right| < \varepsilon\right\} = 1.$$

Предположим, что имеются две несмещенные оценки $\widehat{\theta}(\vec{X}_n)$ и $\widetilde{\theta}(\vec{X}_n)$ для параметра θ. Если дисперсии $\mathbf{D}\widehat{\theta}(\vec{X}_n)$ и $\mathbf{D}\widetilde{\theta}(\vec{X}_n)$ удовлетворяют условию

$$\mathbf{D}\widehat{\theta}(\vec{X}_n) \leqslant \mathbf{D}\widetilde{\theta}(\vec{X}_n) \tag{2.1}$$

для любого фиксированного n и $\theta \in \Theta$, то следует предпочесть оценку $\widehat{\theta}(\vec{X}_n)$, поскольку разброс статистики $\widehat{\theta}(\vec{X}_n)$ относительно параметра θ меньше, чем разброс статистики $\widetilde{\theta}(\vec{X}_n)$.

Определение 2.3. Если в некотором классе несмещенных оценок параметра θ, имеющих конечную дисперсию, существует такая оценка $\widehat{\theta}(\vec{X}_n)$, что неравенство (2.1) выполняется для всех оценок $\widetilde{\theta}(\vec{X}_n)$ из этого класса, то говорят, что ***оценка*** $\widehat{\theta}(\vec{X}_n)$ является ***эффективной в данном классе оценок***.

Иными словами, дисперсия эффективной оценки параметра в некотором классе является минимальной среди дисперсий всех оценок из рассматриваемого класса несмещенных оценок.

Замечание 2.1. Эффективную оценку в классе всех несмещенных оценок будем называть ***эффективной оценкой***, не добавляя слов „в классе несмещенных оценок".

Замечание 2.2. В литературе по математической статистике при рассмотрении параметрических моделей вместо термина „эффективная оценка" в классе всех несмещенных оценок используют и другие: „несмещенная оценка с минимальной дисперсией", „оптимальная оценка".

Теорема 2.1. Оценка

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(*выборочное среднее*) математического ожидания

$$\theta = \mathbf{M}X = \mu$$

генеральной совокупности X с конечной дисперсией является несмещенной, состоятельной и эффективной в классе всех ***линейных оценок***, т.е. оценок вида

$$\widetilde{\theta}(\vec{X}_n) = \sum_{i=1}^{n} \alpha_i X_i,$$

где $\sum\limits_{i=1}^{n} \alpha_i = 1$, для произвольной параметрической модели.

◀ Напомним, что элементы X_i, $i = \overline{1, n}$, случайной выборки $\vec{X}_n$ являются независимыми случайными величинами и распределенными так же, как и сама генеральная совокупность X. Следовательно, $\mathbf{M}X_i = \mathbf{M}X = \mu$ и $\mathbf{D}X_i = \mathbf{D}X = \sigma^2$, $i = \overline{1, n}$.

В силу свойств математического ожидания имеем

$$\mathbf{M}\overline{X} = \mathbf{M}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}X_i = \frac{1}{n}n\mu = \mu,$$

что и доказывает несмещенность оценки $\overline{X}$.

Далее, поскольку последовательность $X_1, \ldots, X_n$ состоит из независимых одинаково распределенных случайных величин с конечной дисперсией, то в силу закона больших чисел в форме Чебышева для любого $\varepsilon > 0$

$$\mathbf{P}\big\{|\overline{X} - \mu| < \varepsilon\big\} \to 1, \quad n \to \infty,$$

т.е. оценка $\overline{X}$ сходится по вероятности к оцениваемому параметру, а это и означает ее состоятельность.

Покажем теперь, что

$$\mathbf{D}\,\widetilde{\theta}(\vec{X}_n) = \mathbf{D}\sum_{i=1}^{n}\alpha_i X_i = \sum_{i=1}^{n}\mathbf{D}(\alpha_i X_i) = \sum_{i=1}^{n}\alpha_i^2\,\mathbf{D}\,X_i = \sigma^2\sum_{i=1}^{n}\alpha_i^2$$

достигает своего минимального значения при $\alpha_i = 1/n$, т.е. когда оценка $\widetilde{\theta}(\vec{X}_n) = \overline{X}$, что и означает эффективность оценки $\overline{X}$ в классе линейных оценок.

Для отыскания условного минимума функции

$$g(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n}\alpha_i^2$$

при ограничении

$$\sum_{i=1}^{n}\alpha_i = 1$$

составим функцию Лагранжа

$$L(\alpha_1, \ldots, \alpha_n; \lambda) = \sum_{i=1}^{n}\alpha_i^2 + \lambda\Big(\sum_{i=1}^{n}\alpha_i - 1\Big),$$

где λ — множитель Лагранжа. Необходимые условия существования условного экстремума имеют вид

$$\begin{cases} \dfrac{\partial L}{\partial \alpha_i} = 2\alpha_i + \lambda = 0, \qquad i = \overline{1,n}, \\ \dfrac{\partial L}{\partial \lambda} = \sum\limits_{i=1}^{n} \alpha_i - 1 = 0. \end{cases}$$

Решив эту систему, находим $\lambda = -2/n$ и $\alpha_i = 1/n$, $i = \overline{1,n}$, и убеждаемся в том, что при этих значениях аргументов функция $g(\alpha_1, \ldots, \alpha_n)$ имеет условный минимум. ▶

Замечание 2.3. Можно доказать состоятельность оценки $\overline{X}$ для математического ожидания (если оно существует), не предполагая существования конечной дисперсии $\mathbf{D}X$. #

Свойства *выборочной дисперсии*

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

отражены в следующей теореме.

Теорема 2.2. Если $\vec{X}_n$ — случайная выборка из генеральной совокупности X с конечной дисперсией σ^2, то выборочная дисперсия $\hat{\sigma}^2(\vec{X}_n)$ — смещенная состоятельная оценка σ^2.

◀ Действительно,

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}\Big((X_i - \mu) - (\overline{X} - \mu)\Big)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2(\overline{X} - \mu)\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) + \frac{1}{n}\sum_{i=1}^{n}(\overline{X} - \mu)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2(\overline{X} - \mu)^2 + \frac{1}{n}n(\overline{X} - \mu)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\overline{X} - \mu)^2.$$

Используя свойства математического ожидания, получим

$$\mathbf{M}\,\widehat{\sigma}^2(\vec{X}_n) = \frac{1}{n}\mathbf{M}\sum_{i=1}^{n}(X_i-\mu)^2 - \mathbf{M}(\overline{X}-\mu)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}(X_i-\mu)^2 - \mathbf{M}(\overline{X}-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{D}\,X_i - \mathbf{D}\,\overline{X} =$$

$$= \frac{1}{n}n\sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2,$$

т.е. $\widehat{\sigma}^2(\vec{X}_n)$ — смещенная оценка для дисперсии.

Докажем, что $\widehat{\sigma}^2(\vec{X}_n)$ является состоятельной оценкой. Доказательство проведем для случая, когда генеральная совокупность имеет моменты до четвертого порядка включительно и нулевое математическое ожидание. Последнее допущение не является принципиальным, так как дисперсия не зависит от значения ее математического ожидания (от точки отсчета). Применяя второе неравенство Чебышева, имеем

$$\mathbf{P}\left\{\left|\widehat{\sigma}^2(\vec{X}_n) - \frac{n-1}{n}\sigma^2\right| < \varepsilon\right\} \geqslant 1 - \frac{\mathbf{D}\,\widehat{\sigma}^2(\vec{X}_n)}{\varepsilon^2}.$$

Найдем дисперсию $\widehat{\sigma}^2(\vec{X}_n)$:

$$\widehat{\sigma}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X\overline{X} + \overline{X}^2) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n}X_i + \frac{1}{n}n\overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2.$$

Воспользуемся известным равенством, согласно которому дисперсия скалярной случайной величины равна математическому ожиданию ее квадрата минус квадрат ее математического ожидания:

$$\mathbf{D}\,\widehat{\sigma}^2(\vec{X}_n) = \mathbf{M}\left(\widehat{\sigma}^2(\vec{X}_n)\right)^2 - \left(\mathbf{M}\,\widehat{\sigma}^2(\vec{X}_n)\right)^2.$$

Поскольку

$$(\widehat{\sigma}^2(\vec{X}_n))^2=\Bigl(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\Bigr)^2=$$
$$=\frac{1}{n^2}\Bigl(\sum_{i=1}^{n}X_i^2\Bigr)^2-2\overline{X}^2\frac{1}{n}\sum_{i=1}^{n}X_i^2+\overline{X}^4,$$

заключаем, что

$$\mathrm{M}\bigl(\widehat{\sigma}^2(\vec{X}_n)\bigr)^2=\frac{1}{n^2}\mathrm{M}\Bigl(\sum_{i=1}^{n}X_i^2\Bigr)^2-\frac{2}{n}\mathrm{M}\Bigl(\overline{X}^2\sum_{i=1}^{n}X_i^2\Bigr)+\mathrm{M}\overline{X}^4.$$

Получим выражения для математических ожиданий трех слагаемых, используя свойства математического ожидания и независимость случайных величин X_i, $i=\overline{1,n}$, каждая из которых имеет нулевое математическое ожидание и дисперсию σ^2. Для первого слагаемого имеем

$$\mathrm{M}\Bigl(\sum_{i=1}^{n}X_i^2\Bigr)^2=\mathrm{M}\Bigl(\sum_{i=1}^{n}X_i^4+\sum_{\substack{i,j=1\\i\neq j}}^{n}X_i^2X_j^2\Bigr)=\sum_{i=1}^{n}\mathrm{M}X_i^4+$$
$$+\sum_{\substack{i,j=1\\i\neq j}}^{n}\mathrm{M}X_i^2\,\mathrm{M}X_j^2=\sum_{i=1}^{n}\mathring{m}_4+\sum_{\substack{i,j=1\\i\neq j}}^{n}\sigma^2\sigma^2=n\,\mathring{m}_4+n(n-1)\sigma^4.$$

Чтобы вычислить второе слагаемое, преобразуем его:

$$\mathrm{M}\Bigl(\overline{X}^2\sum_{i=1}^{n}X_i^2\Bigr)=\frac{1}{n^2}\mathrm{M}\Bigl(\Bigl(\sum_{j=1}^{n}X_j\Bigr)^2\sum_{i=1}^{n}X_i^2\Bigr)=$$
$$=\frac{1}{n^2}\mathrm{M}\Bigl(\Bigl(\sum_{j=1}^{n}X_j^2+\sum_{\substack{j,k=1\\j\neq k}}^{n}X_jX_k\Bigr)\sum_{i=1}^{n}X_i^2\Bigr)=$$
$$=\frac{1}{n^2}\mathrm{M}\Bigl(\sum_{j=1}^{n}X_j^2\sum_{i=1}^{n}X_i^2\Bigr)+\frac{1}{n^2}\Bigl(\sum_{i=1}^{n}X_i^2\sum_{\substack{j,k=1\\j\neq k}}^{n}X_jX_k\Bigr).$$

Так как

$$\mathbf{M}\sum_{\substack{i=1\\ i\neq k}}^{n} X_i X_k = \sum_{\substack{i=1\\ i\neq k}}^{n} \mathbf{M}\, X_i \mathbf{M}\, X_k = 0, \quad k = 1, n,$$

то

$$\mathbf{M}\sum_{i=1}^{n} X_i^2 \sum_{\substack{j,k=1\\ j\neq k}}^{n} X_j X_k =$$

$$= \sum_{\substack{i,j,k=1\\ i\neq j,\, i\neq k,\, j\neq k}}^{n} \mathbf{M}(X_i^2 X_j X_k) + 2\sum_{\substack{i,j=1\\ i\neq j}}^{n} \mathbf{M}(X_i^3 X_j) = 0.$$

Следовательно,

$$\mathbf{M}\Big(\overline{X}^2 \sum_{i=1}^{n} X_i^2\Big) = \frac{1}{n^2}\mathbf{M}\Big(\sum_{j=1}^{n} X_j^2 \sum_{i=1}^{n} X_i^2\Big) =$$

$$= \frac{1}{n^2}\mathbf{M}\Big(\sum_{i=1}^{n} X_i^4 + \sum_{\substack{i,j=1\\ i\neq j}}^{n} X_i^2 X_j^2\Big) = \frac{1}{n^2}\big(n\,\mathring{m}_4 + n(n-1)\sigma^4\big).$$

Аналогично можно показать, что

$$\mathbf{M}\overline{X}^4 = \frac{1}{n^4}\Big(\sum_{i=1}^{n} X_i\Big)^4 = \frac{\mathring{m}_4 + 3(n-1)\sigma^4}{n^3}.$$

В итоге получаем

$$\mathbf{M}\big(\widehat{\sigma}^2(\vec{X}_n)\big)^2 = \frac{n\,\mathring{m}_4 + n(n-1)\sigma^4}{n^2} - \frac{2(n\,\mathring{m}_4 + n(n-1)\sigma^4)}{n^3} +$$

$$+ \frac{\mathring{m}_4 - 3\sigma^4}{n^3} = \sigma^4 + \frac{\mathring{m}_4 - 3\sigma^4}{n} - \frac{2\,\mathring{m}_4 - 5\sigma^4}{n^2} + \frac{\mathring{m}_4 + 3(n-1)\sigma^4}{n^3}.$$

Поскольку $\mathbf{M}\widehat{\sigma}^2(\vec{X}_n) = \sigma^2 - \sigma^2/n$, окончательно находим

$$\mathbf{D}\widehat{\sigma}^2(\vec{X}_n) = \frac{\mathring{m}_4 - \sigma^4}{n} - \frac{2(\mathring{m}_4 - 2\sigma^4)}{n^2} + \frac{\mathring{m}_4 - 3\sigma^4}{n^3},$$

откуда с учетом второго неравенства Чебышева и следует состоятельность оценки $\widehat{\sigma}^2(\vec{X}_n)$ для дисперсии σ^2 генеральной совокупности X. ▶

Замечание 2.4. Из теоремы 2.2 следует, что статистика

$$S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

является несмещенной и состоятельной оценкой дисперсии σ^2 генеральной совокупности. Ее называют ***исправленной выборочной дисперсией***.

Действительно,

$$S^2(\vec{X}_n) = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}\sigma^2(\vec{X}_n).$$

Имеем

$$\mathbf{M}S^2(\vec{X}_n) = \frac{n}{n-1}\mathbf{M}\widehat{\sigma}^2(\vec{X}_n) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2$$

и

$$\mathbf{D}S^2(\vec{X}_n) = \frac{n^2}{(n-1)^2}\mathbf{D}\widehat{\sigma}^2(\vec{X}_n) \to 0$$

при $n \to \infty$, откуда и следует несмещенность и состоятельность $S^2(\vec{X}_n)$.

Отметим, что в дальнейшем ее выборочное значение будем обозначать S^2.

Замечание 2.5. Можно доказать, что *выборочные начальные* и *центральные моменты* являются состоятельными оценками соответствующих моментов генеральной совокупности, если только они существуют*. Однако эти оценки, кроме X, являются смещенными.

Пример 2.1. Пусть n — число испытаний по схеме Бернулли с неизвестной вероятностью успеха θ. Рассмотрим случайную выборку $(X_1, \ldots, X_n)$, где X_i, $i = \overline{1, n}$, — случайная величина, которая с вероятностью θ принимает значение 1 („успех" в i-м испытании) и с вероятностью $1 - \theta$ — значение 0 („неудача" в i-м испытании).

В качестве оценки θ возьмем относительную частоту успехов, т.е. $\widehat{\theta}(\vec{X}_n) = k(\vec{X}_n)/n$, где

$$k(\vec{X}_n) = \sum_{i=1}^{n} X_i$$

есть суммарное число успехов в n испытаниях. Эта оценка является несмещенной, так как

$$\mathrm{M}\widehat{\theta}(\vec{X}_n) = \frac{1}{n}\mathrm{M}(X_1 + \ldots + X_n) = \frac{1}{n}(\mathrm{M}X_1 + \ldots + \mathrm{M}X_n) = \frac{1}{n}n\theta = \theta,$$

и состоятельной, что непосредственно вытекает из закона больших чисел в форме Бернулли, согласно которому для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{k(\vec{X}_n)}{n} - p\right| < \varepsilon\right\} = 1. \quad \#$$

В дальнейшем в соответствии с установившейся традицией статистику $k(\vec{X}_n)$, так же как и ее значение, часто будем обозначать просто символом k. В каждом конкретном случае должно быть ясно, о чем идет речь: о случайной величине или ее реализации.

Пример 2.2. Пусть $X_1, \ldots, X_n$ — случайная выборка из генеральной совокупности X, имеющей нормальное распределение с неизвестным средним значением θ и известной дисперсией σ^2.

Оценка $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n) = X_1$ является несмещенной для θ, ибо $\mathrm{M}X_1 = \mathrm{M}X = \theta$, но не является состоятельной, так как, во-первых, X_1 не зависит от объема выборки и, следовательно, ее распределение не меняется с ростом n, а во-вторых,

$$\mathbf{P}\{|X_1-\theta|<\varepsilon\}=\frac{2}{\sigma\sqrt{2\pi}}\int\limits_0^{\varepsilon} e^{-\frac{t^2}{2\sigma^2}}\,dt\neq 1.$$

Пример 2.3. Имеем случайную выборку $\vec{X}_n$ из генеральной совокупности X с равномерным законом распределения

$$p(t;\theta)=\begin{cases}\dfrac{1}{b-a}, & t\in[a,b],\\ 0, & t\notin[a,b],\end{cases}$$

где $b-a=l$ — известная величина, $\theta=(a+b)/2$ — неизвестный параметр.

Возьмем в качестве оценки параметра θ среднее арифметическое *крайних членов вариационного ряда*

$$\theta^*(\vec{X}_n)=\frac{X_{(1)}+X_{(n)}}{2}.$$

Убедимся, что $\theta^*(\vec{X}_n)$ является несмещенной оценкой параметра θ и в классе всех несмещенных оценок $\overline{X}$ не является эффективной оценкой параметра θ для заданной параметрической модели.

Плотности распределения $X_{(1)}$ и $X_{(n)}$ на отрезке $[a,b]$ соответственно равны

$$p_{X_{(1)}}(x)=n\left(1-\frac{x-a}{b-a}\right)^{n-1}\frac{1}{b-a},\qquad p_{X_{(n)}}(x)=n\left(\frac{x-a}{b-a}\right)^{n-1}\frac{1}{b-a}$$

(см. пример 2.20). Вычислив

$$\mathbf{M}X_{(1)}=\int\limits_a^b xn\left(\frac{b-x}{b-a}\right)^{n-1}\frac{1}{b-a}\,dx=b-\frac{n}{n+1}(b-a),$$

$$\mathbf{M}X_{(n)}=\int\limits_a^b xn\left(\frac{x-a}{b-a}\right)^{n-1}\frac{1}{b-a}\,dx=a+\frac{n}{n+1}(b-a),$$

получим

$$\mathbf{M}\theta^*(\vec{X}_n) = \frac{1}{2}\big(\mathbf{M}X_{(1)} + \mathbf{M}X_{(n)}\big) = \frac{a+b}{2},$$

что и доказывает несмещенность оценки $\theta^*(\vec{x}_n)$.

Далее, используя совместную плотность распределения вероятностей случайных величин* $X_{(1)}$ и $X_{(n)}$

$$p_{X_{(1)}X_{(n)}}(x,y) = \frac{n(n-1)(y-x)^{n-2}}{(b-a)^n}, \quad a < x \leqslant y \leqslant b,$$

и равенство

$$\mathbf{D}\,\theta^*(\vec{X}_n) = \frac{1}{4}\mathbf{D}\big(X_{(1)} + X_{(n)}\big) =$$
$$= \frac{1}{4}\big(\mathbf{D}\,X_{(1)} + \mathbf{D}\,X_{(n)}\big) + \frac{1}{2}\,\mathbf{cov}\big(X_{(1)}, X_{(n)}\big),$$

можно получить

$$\mathbf{D}\,\theta^*(\vec{X}_n) = \frac{(b-a)^2}{2(n+1)(n+2)}.$$

Поскольку

$$\mathbf{D}\,\theta^*(\vec{X}_n) < \mathbf{D}(\overline{X}) = \frac{\sigma^2}{n} = \frac{(b-a)^2}{12n}, \quad n \geqslant 3,$$

то, следовательно, в классе всех несмещенных оценок $\overline{X}$ не является эффективной оценкой параметра θ для рассматриваемой параметрической модели.

Теорема 2.3 (о единственности эффективной оценки). Пусть $\widehat{\theta}(\vec{X}_n)$ и $\widetilde{\theta}(\vec{X}_n)$ — две эффективные оценки для параметра θ рассматриваемой параметрической модели. Тогда

$$\widehat{\theta}(\vec{X}_n) = \widetilde{\theta}(\vec{X}_n),$$

где равенство следует понимать в вероятностном смысле:

$$\mathbf{P}\Big\{\vec{X}_n \in \{\vec{x}_n\colon \widehat{\theta}(\vec{x}_n) \neq \widetilde{\theta}(\vec{x}_n)\}\Big\} = 0.$$

◀ Действительно, рассмотрим статистику

$$\theta^*(\vec{X}_n) = \frac{1}{2}\big(\widehat{\theta}(\vec{X}_n) + \widetilde{\theta}(\vec{X}_n)\big).$$

По условию $\mathbf{D}\widehat{\theta}(\vec{X}_n) = \mathbf{D}\widetilde{\theta}(\vec{X}_n)$. Значит,

$$\mathbf{D}\theta^*(\vec{X}_n) = \frac{1}{4}\Big(\mathbf{D}\widehat{\theta}(\vec{X}_n) + \mathbf{D}\widetilde{\theta}(\vec{X}_n)\Big) + \frac{1}{2}\,\mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big) = \\ = \frac{1}{2}\Big(\mathbf{D}\widehat{\theta}(\vec{X}_n) + \mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big)\Big).$$

Поскольку

$$\Big|\mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big)\Big| \leqslant \sqrt{\mathbf{D}\widehat{\theta}(\vec{X}_n)\,\mathbf{D}\widetilde{\theta}(\vec{X}_n)} = \mathbf{D}\widehat{\theta}(\vec{X}_n),$$

то

$$\mathbf{D}\theta^*(\vec{X}_n) = \frac{1}{2}\Big|\mathbf{D}\widehat{\theta}(\vec{X}_n) + \mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big)\Big| \leqslant \\ \leqslant \frac{1}{2}\Big(\mathbf{D}\widehat{\theta}(\vec{X}_n) + \mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big)\Big) \leqslant \mathbf{D}\widehat{\theta}(\vec{X}_n).$$

А так как $\widehat{\theta}(\vec{X}_n)$ — эффективная оценка, то

$$\mathbf{D}\theta^*(\vec{X}_n) = \mathbf{D}\widehat{\theta}(\vec{X}_n) = \mathbf{D}\widetilde{\theta}(\vec{X}_n)$$

и, как следствие,

$$\mathbf{cov}\big(\widehat{\theta}(\vec{X}_n), \widetilde{\theta}(\vec{X}_n)\big) = \mathbf{D}\widehat{\theta}(\vec{X}_n) = \mathbf{D}\widetilde{\theta}(\vec{X}_n) = \sqrt{\mathbf{D}\widehat{\theta}(\vec{X}_n)\,\mathbf{D}\widetilde{\theta}(\vec{X}_n)}.$$

Из последнего равенства следует [XVI], что

$$\widehat{\theta}(\vec{X}_n) = k\widetilde{\theta}(\vec{X}_n) + b.$$

Так как

$$\mathbf{D}\,\widehat{\theta}(\vec{X}_n) = \mathbf{cov}\big(k\widetilde{\theta}(\vec{X}_n) + b, \widetilde{\theta}(\vec{X}_n)\big) = k\mathbf{D}\,\widetilde{\theta}(\vec{X}_n) = k\mathbf{D}\,\widehat{\theta}(\vec{X}_n),$$

то получаем $k = 1$. Из условия несмещенности оценок следует, что $b = 0$:

$$\mathbf{M}\,\widetilde{\theta}(\vec{X}_n) = \mathbf{M}\,\widehat{\theta}(\vec{X}_n) = \mathbf{M}\big(\widetilde{\theta}(\vec{X}_n) + b\big) = \mathbf{M}\,\widetilde{\theta}(\vec{X}_n) + b.$$

Таким образом, $\widehat{\theta}(\vec{X}_n) = \widetilde{\theta}(\vec{X}_n)$. ▶

В дальнейшем изложении при рассмотрении параметрических моделей будем использовать дифференцирование по параметру под знаком *интеграла, зависящего от параметра*. Параметрические модели, для которых выполнены условия, обеспечивающие законность указанных операций, называют ***регулярными моделями***.

Теорема 2.4 (***неравенство Рао — Крамера****). Пусть рассматриваемая параметрическая модель является регулярной и $\widehat{\theta}(\vec{X}_n)$ — несмещенная оценка неизвестного параметра θ. Тогда имеет место неравенство

$$\mathbf{D}\,\widehat{\theta}(\vec{X}_n) \geqslant \frac{1}{nI(\theta)}, \tag{2.2}$$

где

$$I(\theta) = \mathbf{M}\left(\frac{\partial \ln p(X;\theta)}{\partial \theta}\right)^2.$$

Здесь $I(\theta)$ — ***количество информации по Фишеру**** в одном наблюдении, а $p(t;\theta)$ — плотность распределения генеральной совокупности X в случае *непрерывной статистической модели* и вероятность события $\{X = t\}$ в случае *дискретной статистической модели*.

◀ Доказательство проведем для непрерывной модели. Пусть $p(t;\theta) > 0$ при $t \in A \subset \mathbb{R}$ и $p(t;\theta) = 0$ при $t \notin A$. Тогда плотность распределения

$$p_{\vec{X}_n}(T,\theta) \equiv p_{\vec{X}_n}(t_1,\ldots,t_n,\theta) = \prod_{i=1}^{n} p(t_i;\theta)$$

случайной выборки $\vec{X}_n$ отлична от нуля на множестве

$$B = A \times A \times \ldots \times A \subset \mathbb{R}^n,$$

где $T = (t_1, \ldots, t_n)$ — векторный аргумент. Поскольку

$$\int\limits_{\mathbb{R}^n} p_{\vec{X}_n}(T,\theta)\,dT = \int\limits_{B} p_{\vec{X}_n}(T,\theta)\,dT = 1,$$

имеем

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} p_{\vec{X}_n}(T,\theta)\,dT = \frac{\partial}{\partial\theta}\int\limits_{B} p_{\vec{X}_n}(T,\theta)\,dT = \int\limits_{B} \frac{\partial p_{\vec{X}_n}(T,\theta)}{\partial\theta}\,dT = 0,$$

или

$$\int\limits_{B} \frac{\partial \ln p_{\vec{X}_n}(T,\theta)}{\partial\theta}\, p_{\vec{X}_n}(T,\theta)\,dT = 0. \qquad (2.3)$$

Так как $\widehat{\theta}(X_1,\ldots,X_n)$ — несмещенная оценка параметра θ, то

$$\mathrm{M}\,\widehat{\theta}(\vec{X}_n) = \int\limits_{\mathbb{R}^n} \widehat{\theta}(T)\, p_{\vec{X}_n}(T,\theta)\,dT = \int\limits_{B} \widehat{\theta}(T)\, p_{\vec{X}_n}(T,\theta)\,dT = \theta.$$

Таким образом,

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} \widehat{\theta}(T)\, p_{\vec{X}_n}(T,\theta)\,dT = \int\limits_{B} \widehat{\theta}(T)\, p_{\vec{X}_n}(T,\theta)\,dT = 1,$$

или, что то же самое,

$$\int\limits_B \widehat{\theta}(T)\,\frac{\partial \ln p_{\vec{X}_n}(T,\theta)}{\partial \theta}\,p_{\vec{X}_n}(T,\theta)\,dT = 1. \qquad (2.4)$$

Умножив равенство (2.3) на параметр θ и вычтя его из равенства (2.4), приходим к равенству

$$\int\limits_B (\widehat{\theta}(T)-\theta)\,\frac{\partial \ln p_{\vec{X}_n}(T,\theta)}{\partial \theta}\,p_{\vec{X}_n}(T,\theta)\,dT = 1. \qquad (2.5)$$

Согласно неравенству Коши — Буняковского, имеем

$$1 \leqslant \int\limits_B (\widehat{\theta}(T)-\theta)^2 p_{\vec{X}_n}(T,\theta)\,dT \int\limits_B \left(\frac{\partial \ln p_{\vec{X}_n}(T,\theta)}{\partial \theta}\right)^2 p_{\vec{X}_n}(T,\theta)\,dT =$$
$$= \mathbf{D}\,\widehat{\theta}(\vec{X}_n)\,\mathbf{M}\left(\frac{\partial \ln p_{\vec{X}_n}(\vec{X}_n,\theta)}{\partial \theta}\right)^2,$$

откуда и следует неравенство (2.2), так как

$$\mathbf{M}\left(\frac{\partial \ln p_{\vec{X}_n}(\vec{X}_n,\theta)}{\partial \theta}\right)^2 = \int\limits_B \left(\frac{\partial \ln p_{\vec{X}_n}(\vec{X}_n,\theta)}{\partial \theta}\right)^2 p_{\vec{X}_n}(T,\theta)\,dT =$$
$$= \int\limits_B \left(\sum_{i=1}^{n} \frac{\partial \ln p(t_i,\theta)}{\partial \theta}\right)^2 p_{\vec{X}_n}(T,\theta)\,dT =$$
$$= \sum_{i=1}^{n} \int\limits_B \left(\frac{\partial \ln p(t_i,\theta)}{\partial \theta}\right)^2 p_{\vec{X}_n}(T,\theta)\,dT = \sum_{i=1}^{n} \mathbf{M}\left(\frac{\partial \ln p(X_i,\theta)}{\partial \theta}\right)^2 =$$
$$= \sum_{i=1}^{n} \mathbf{M}\left(\frac{\partial \ln p(X,\theta)}{\partial \theta}\right)^2 = n\mathbf{M}\left(\frac{\partial \ln p(X,\theta)}{\partial \theta}\right)^2 = nI(\theta). \qquad \blacktriangleright$$

Неравенство (2.2) определяет нижнюю границу дисперсий несмещенных оценок параметра θ для регулярных моделей.

Величину

$$e(\theta) = \frac{1}{nI(\theta)\, \mathbf{D}\,\widehat{\theta}(\vec{X}_n)}$$

называют ***показателем эффективности по Рао — Крамеру.*** Из (2.2) следует, что для любой несмещенной оценки параметра θ величина $e(\theta)$ удовлетворяет условию $0 < e(\theta) \leqslant 1$.

Определение 2.4. Несмещенную ***оценку*** $\widehat{\theta}(\vec{X}_n)$ параметра $\theta \in \Theta \subset \mathbb{R}$ называют ***эффективной по Рао — Крамеру***, если показатель эффективности $e(\theta) = 1$.

Замечание 2.6. Равенство

$$\mathbf{D}\,\widehat{\theta}(\vec{X}_n) = \frac{1}{nI(\theta)}$$

имеет место тогда и только тогда, когда

$$\frac{\partial \ln p_{\vec{X}_n}(\vec{X}_n, \theta)}{\partial \theta} = a(\theta)\,\big(\widehat{\theta}(\vec{X}_n) - \theta\big),$$

что является необходимым и достаточным условием обращения неравенства Коши — Буняковского в равенство. Следовательно, это равенство является ***критерием эффективности для регулярных моделей.*** При этом из равенства (2.5) следует, что $a(\theta) = 1/\mathbf{D}\,\widehat{\theta}(\vec{X}_n)$.

Замечание 2.7. Эффективная оценка по Рао — Крамеру для рассматриваемой регулярной модели является эффективной (см. определение 2.3). Утверждение следует из теоремы 2.3 о единственности эффективной оценки в классе несмещенных оценок. Обратное утверждение неверно, поскольку не любая параметрическая модель является регулярной (см. пример 2.21).

Пример 2.4. Рассмотрим *нормальную модель* $N(\theta, \sigma^2)$ в предположении, что дисперсия σ^2 известна. Оценка

$$\widehat{\theta}(\vec{X}_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

является несмещенной для неизвестного среднего значения $\theta = \mu$ (см. теорему 2.1). Убедимся в ее эффективности по Рао — Крамеру. Во-первых, в силу независимости элементов случайной выборки $\vec{X}_n = (X_1, \ldots, X_n)$ имеем

$$\mathbf{D}(\vec{X}_n) = \mathbf{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\,X_i = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Во-вторых,

$$I(\theta) = \mathbf{M}\left(\frac{\partial}{\partial\theta}\ln\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\theta)^2}{2\sigma^2}}\right)\right)^2 =$$

$$= \mathbf{M}\left(\frac{\partial}{\partial\theta}\left(\ln\frac{1}{\sqrt{2\pi}} - \ln\sigma - \frac{(X-\theta)^2}{2\sigma^2}\right)\right)^2 = \mathbf{M}\,\frac{(X-\theta)^2}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

Следовательно,

$$e(\theta) = \frac{1}{nI(\theta)\,\mathbf{D}\overline{X}} = \frac{n}{n\frac{1}{\sigma^2}\cdot\sigma^2} = 1,$$

т.е. для нормальной модели $\overline{X}$ — эффективная оценка параметра μ.

Пример 2.5. Рассмотрим модель $N(\mu,\theta)$ в предположении, что среднее значение μ генеральной совокупности известно, а $\theta = \sigma^2$ — неизвестный параметр.

Покажем, что

$$\widetilde{S}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

является несмещенной и эффективной по Рао — Крамеру оценкой параметра σ^2. Действительно,

$$\mathbf{M}\widetilde{S}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}(X_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{D}\,X_i = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2,$$

т.е. $\widetilde{S}^2(\vec{X}_n)$ — несмещенная оценка.

Вычислим дисперсию $S^2(\vec{X}_n)$:

$$\mathbf{D}\,\widetilde{S}^2(\vec{X}_n) = \mathbf{M}\,\widetilde{S}^4(\vec{X}_n) - \left(\mathbf{M}\,\widetilde{S}^2(\vec{X}_n)\right)^2 =$$
$$= \mathbf{M}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right)^2 - \sigma^4 =$$
$$= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbf{M}(X_i-\mu)^4 + \sum_{\substack{i,j=1\\ i\neq j}}^{n}\mathbf{M}\left((X_i-\mu)^2(X_j-\mu)^2\right)\right) - \sigma^4 =$$
$$= \frac{n\,\mathring{m}_4}{n^2} + \frac{n(n-1)}{n^2}\sigma^4 - \sigma^4 = \frac{3\sigma^4}{n} + \frac{n-1}{n}\sigma^4 - \sigma^4 = \frac{2\sigma^4}{n}.$$

Затем определим информацию по Фишеру:

$$I(\theta) = \mathbf{M}\left(\frac{\partial}{\partial\sigma^2}\left(\ln\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(X-\mu)^2}{2\sigma^2}}\right)\right)^2 =$$
$$= \mathbf{M}\left(\frac{\partial}{\partial\sigma^2}\left(\ln\frac{1}{\sqrt{2\pi}} - \frac{1}{2}\ln\sigma^2 - \frac{(X-\mu)^2}{2\sigma^2}\right)\right)^2 =$$
$$= \mathbf{M}\left(-\frac{1}{2\sigma^2} + \frac{(X-\mu)^2}{2\sigma^4}\right)^2 = \frac{1}{4\sigma^4} - \frac{\mathbf{M}(X-\mu)^2}{2\sigma^6} + \frac{\mathbf{M}(X-\mu)^4}{4\sigma^8} =$$
$$= \frac{1}{4\sigma^2} - \frac{\sigma^2}{2\sigma^6} + \frac{3\sigma^4}{4\sigma^8} = \frac{1}{2\sigma^4},$$

поскольку для нормальной модели $\mathring{m}_4 = 3\sigma^4$. В результате получим

$$\mathbf{D}\,\widetilde{S}^2(\vec{X}_n) = \frac{1}{nI(\theta)} \quad \text{и} \quad \mathrm{e}(\theta) = 1,$$

т.е. $\widetilde{S}^2(\vec{X}_n)$ — эффективная оценка параметра θ для нормальной модели. #

Заметим, что

$$S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

является несмещенной оценкой параметра σ^2 (см. замечание 2.4), но для нормальной модели $N(\mu,\theta)$ эта оценка не является эффективной. Это вытекает из теоремы 2.3 о единственности существования эффективной оценки. Можно показать, что

$$\mathbf{D}\widetilde{S}^2(\vec{X}_n) = \frac{2\sigma^4}{n-1}.$$

Следовательно,

$$e(\theta) = \frac{1}{nI(\theta)\,\mathbf{D}S^2(\vec{X}_n)} = \frac{n-1}{n} < 1.$$

Пример 2.6. Рассмотрим экспоненциальную модель

$$p(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{1}{\theta}x}, & x \geqslant 0; \\ 0, & x < 0. \end{cases}$$

Покажем, что $\overline{X}$ является эффективной по Рао — Крамеру оценкой неизвестного параметра θ. Действительно,

$$\mathbf{D}(\overline{X}) = \mathbf{D}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\mathbf{D}X = \frac{\theta^2}{n},$$

$$I(\theta) = \mathbf{M}\bigg(\frac{\partial}{\partial\theta}\Big(\ln\frac{1}{\theta}e^{-\frac{X}{\theta}}\Big)\bigg)^2 = \mathbf{M}\Big(-\frac{1}{\theta}+\frac{X}{\theta^2}\Big)^2 =$$

$$= \mathbf{M}\Big(\frac{X-\theta}{\theta^2}\Big)^2 = \frac{\mathbf{M}(X-\theta)^2}{\theta^4} = \frac{\mathbf{D}X}{\theta^4} = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2},$$

откуда заключаем, что

$$e(\theta) = \frac{1}{n\frac{1}{\theta^2}\cdot\frac{\theta^2}{n}} = 1.$$