

Понятие статистической оценки

Пусть имеется выборка $\mathbb{X}_n = (X_1, \dots, X_n)$ из распределения случайной величины $\xi \in \{F(x; \theta), \theta \in \Theta\}$. В общем случае задача оценивания заключается в том, что используя статистическую информацию, доставляемую выборкой $\mathbb{X}_n$ сделать статистические выводы об истинном значении неизвестного параметра θ .

Определение.

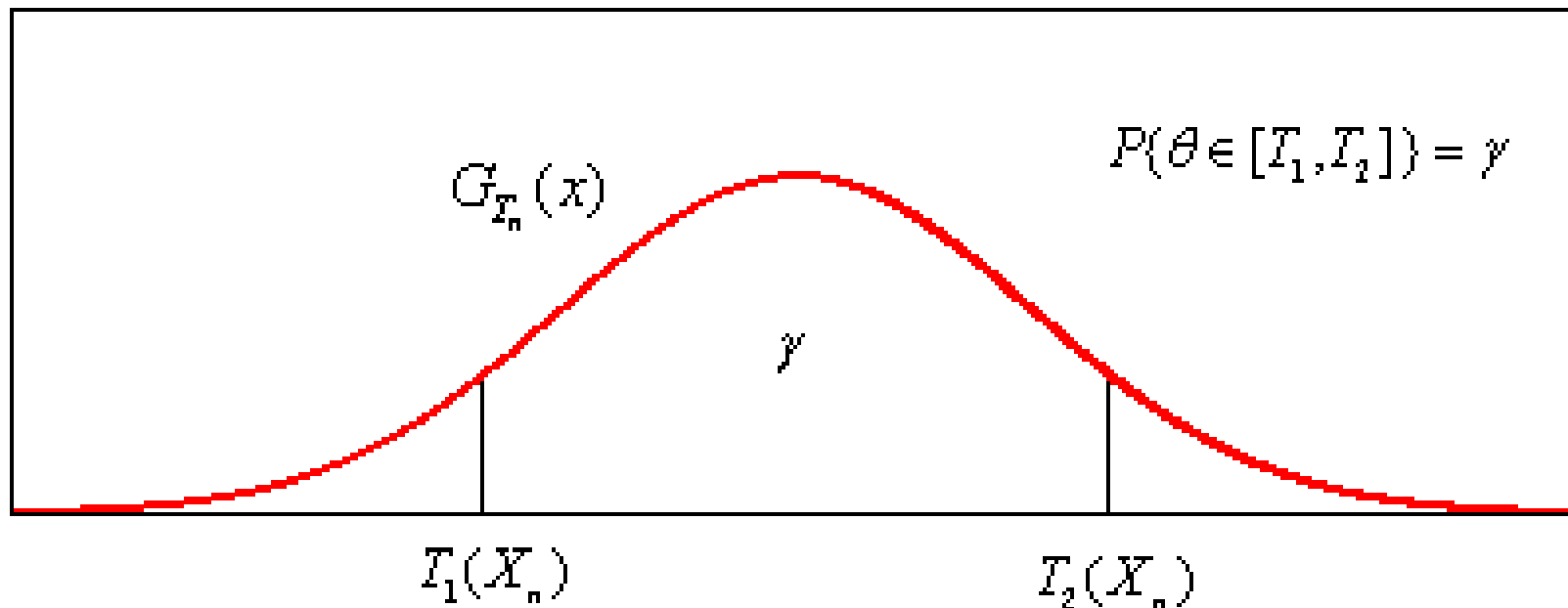
Точечной оценкой неизвестного параметра θ по выборке $\mathbb{X}_n$ называется значение некоторой статистики $T_n = T(\mathbb{X}_n)$, которое приближенно равно значению параметра θ :

$$\hat{\theta} = T_n(\mathbb{X}_n).$$

Так как любая статистика является случайной величиной (имеющей некоторое распределение $G_{T_n}(x)$), то для каждой новой реализации выборки $\mathbb{X}_n$ будет получаться другое значение оценки, в общем случае отличное от истинного значения параметра θ .

Определение.

Интервальной оценкой параметра θ называют интервал $[T_1(X_n), T_2(X_n)]$, содержащий истинное значение параметра θ с вероятностью γ .



Критерии сравнения оценок

Определение.

Статистика $T(\mathbb{X}_n)$ называется *несмещенной* оценкой параметра θ , если выполняется условие: $M[T(\mathbb{X}_n)] = \theta, \forall \theta \in \Theta$.

Для смещенных оценок можно найти величину $b(\theta) = M[T(\mathbb{X}_n)] - \theta$, называемую смещением оценки $T(\mathbb{X}_n)$.

Определение.

Величину $M[T(\mathbb{X}_n) - \theta]^2 = D[T(\mathbb{X}_n)] + b^2(\theta)$ называют *среднеквадратической ошибкой* оценки T .

Для несмещенных оценок среднеквадратическая ошибка совпадет с дисперсией оценки (так как $b(\theta) = 0$).

Несмещенные оценки с равномерно минимальной дисперсией

Определение.

Несмещенной оценкой с равномерно минимальной дисперсией называется такая оценка $T^*(X_n)$, что $D[T^*(X_n)] \leq D[T(X_n)] : \forall \theta \in \Theta, \forall T(X_n) : M[T(X_n)] = \theta$.

Требование равномерной минимальности дисперсии сильное и не всегда имеет место. Может оказаться, что двух несмещенных оценок T_1 и T_2 , для одних значений θ минимальна дисперсия одной оценки, а для других – дисперсия другой оценки.

Теорема 2.1. (О единственности НОРМД)

Если существует $T^*(X_n)$, то она единственная.

Доказательство:

Предположим, что существуют две НОРМД $T_1(X_n)$ и $T_2(X_n)$ функции θ .
 $M[T_1(X_n)] = M[T_2(X_n)] = \theta$ и $D[T_1(X_n)] = D[T_2(X_n)] = v = v(\theta) \inf_{T_n} D[T_n]$.

Рассмотрим оценку $T_3 = \frac{T_1 + T_2}{2}$:

$$M[T_3(X_n)] = \frac{M[T_1] + M[T_2]}{2} = \frac{2\theta}{2} = \theta$$

$$D[T_3(X_n)] = D\left[\frac{T_1 + T_2}{2}\right] = \frac{1}{4}(D[T_1] + D[T_2] + 2\text{cov}(T_1, T_2)) = \frac{v + \text{cov}(T_1, T_2)}{2}$$

Воспользуемся неравенством Коши–Буняковского $\forall \eta_1, \eta_2 : |\text{cov}(\eta_1, \eta_2)| \leq \sqrt{D\eta_1 D\eta_2}$,
 причем равенство достигается тогда и только тогда, когда $\eta_1 = k\eta_2 + a$.

Но так как v – минимальное значение дисперсии оценки, то T_3 – НОРМД и
 $\Rightarrow \text{cov}(T_1, T_2) = \sqrt{vv} = v \Rightarrow T_1$ и T_2 линейно связаны, то есть $T_1 = kT_2 + a$

Из несмещенности

$$\Rightarrow MT_1 = kMT_2 + a \Rightarrow \theta = k\theta + a$$

$$a = \theta(1-k) \Rightarrow T_1 = T\eta_2 + \theta(1-k) \Rightarrow T_1 - \theta = k(T_2 - \theta)$$

теперь найдем k :

$$v = \text{cov}(T_1, T_2) = M(T_1 - \theta)(T_2 - \theta) = kM(T_2 - \theta)^2 = kDT_2 = kv$$

$$\Rightarrow k \equiv 1 \Rightarrow T_1 \equiv T_2$$



Состоятельность оценок. Критерий состоятельности

Определение.

Оценка $T_n(X_n)$ некоторой функции $\tau(\theta)$ называется *состоятельной*, если при $n \rightarrow \infty, T_n \xrightarrow{P} \tau(\theta), \forall \theta \in \Theta$.

То есть $\forall \varepsilon > 0: P\{|T_n(X_n) - \tau(\theta)| > \varepsilon\} \rightarrow 0, n \rightarrow \infty$.

Свойство состоятельности обязательно для любого правила оценивания, однако оно является асимптотическим и не связано со свойствами оценки при фиксированном объеме выборки (в отличие от свойств несмещенности и минимальной дисперсии).

Теорема 2.2 (критерий состоятельности).

Пусть $M_\theta T_n = \tau(\theta) + \varepsilon_n, D_\theta T_n = \delta_n$ и $\varepsilon_n = \varepsilon_n(\theta) \rightarrow 0, \delta_n = \delta_n(\theta) \rightarrow 0$ при $n \rightarrow \infty$.

Тогда T_n – состоятельная оценка функции $\tau = \tau(\theta)$.

Функция правдоподобия. Информационное количество Фишера

Пусть $f(x, \theta)$ - плотность случайной величины ξ в непрерывном случае или вероятность в дискретном случае и $\mathbb{X}_n = (X_1, \dots, X_n)$ - выборка.

Так как все наблюдения в выборке независимы, то совместная плотность распределения вектора $\mathbb{X}_n$:

$$L_{\mathbb{X}_n}(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta), \text{ при этом } \int L(x_1, \dots, x_n, \theta) dx_1 \dots dx_n = 1.$$

Определение.

Функция $L(\mathbb{X}_n, \theta)$, рассматриваемая при фиксированной $\mathbb{X}_n$ как функция θ называется функцией *правдоподобия*.

Если $L(\mathbb{X}_n, \theta) > 0, \forall \mathbb{X}_n \in X, \theta \in \Theta$, то $\ln L(\mathbb{X}_n, \theta) = \sum_{i=1}^n \ln f(X_i, \theta)$.

Определение.

Функция $U(\mathbb{X}_n, \theta) = \frac{\partial \ln L(\mathbb{X}_n, \theta)}{\partial \theta}$ называется *вкладом выборки*.

Найдем, чему равно $MU(\mathbb{X}_n, \theta)$.

$$\begin{aligned}\int_{\mathbb{X}} L(\mathbb{X}_n, \theta) d\bar{x} &= 1 \\ 0 &= \frac{\partial}{\partial \theta} \int_{\mathbb{X}} L(\bar{x}, \theta) d\bar{x} = \int_{\mathbb{X}} \frac{\partial \ln L(\bar{x}, \theta)}{\partial \theta} L(\bar{x}, \theta) d\bar{x} = M \left[\frac{\partial \ln L(\mathbb{X}_n, \theta)}{\partial \theta} \right] = \\ &= M[U(\mathbb{X}_n, \theta)].\end{aligned}$$

Значит $M[U(\mathbb{X}_n, \theta)] = 0$.

Определение.

Информацией Фишера о параметре θ содержащейся в выборке X_n называется дисперсия вклада выборки:

$$I_n(\theta) = D[U(X_n, \theta)] = D\left[\frac{\partial \ln L(X_n, \theta)}{\partial \theta}\right] = D\left[\sum_{i=1}^n \frac{\partial \ln f(x_i, \theta)}{\partial \theta}\right] = ni(\theta)$$

$$i(\theta) = \int_x \left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2 f(x, \theta) dx.$$

где $i(\theta)$ - называют количество информации Фишера, содержащейся в одном наблюдении.

Если продифференцируем выражение $M[U(X_n, \theta)] = 0$ еще раз, то получим:

$$0 = \int \frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} + \int \left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2 f(x, \theta) dx$$

$$\Rightarrow i(\theta) = -M\left[\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2}\right].$$

Пример 1.

Найти $i(\theta)$ для модели $N(\theta, \sigma^2)$.

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

$$\frac{\partial f(x, \theta)}{\partial \theta} = -\frac{x-\theta}{\sigma^2}; \frac{\partial^2 f(x, \theta)}{\partial \theta^2} = -\frac{1}{\sigma^2}$$

$$i(\theta) = -M \left[\frac{\partial^2 f(x, \theta)}{\partial \theta^2} \right] = -\frac{1}{\sigma^2}$$

Пример 2 (нерегулярная модель)

Рассмотрим равномерное распределение $f(x, \theta) = \begin{cases} \frac{1}{\theta}, & x \in [0, \theta] \\ 0, & x \notin [0, \theta] \end{cases}$. Здесь из $\int_0^\theta \frac{1}{\theta} dx = 1$ не

следует $\int_0^\theta \frac{\partial}{\partial \theta} \left(\frac{1}{\theta} \right) dx = 0$, так как при дифференцировании интеграла по верхнему

пределу появляется еще одно слагаемое. В данном случае модель является не регулярной, так как выборочное пространство зависит от неизвестного параметра θ .

Неравенство Рао – Крамера и эффективные оценки

Теорема 2.3. (неравенство Рао – Крамера).

Для любой несмещенной оценки $T(X_n)$ параметрической функции $\tau(\theta)$ справедливо неравенство:

$$D[T(X_n)] \geq \frac{[\tau'(\theta)]^2}{ni(\theta)}$$

Равенство имеет место тогда и только тогда, когда $T(X_n) = \tau(\theta) + a(\theta)U(X_n, \theta)$, где $a(\theta)$ - некоторая функция от θ .

Неравенство Рао – Крамера определяет нижнюю границу дисперсий всех несмещенных оценок заданных параметрической функцией $\tau(\theta)$ для регулярных моделей.

Определение.

Оценка, при которой достигается нижняя граница неравенства Рао – Крамера называется *эффективной*.

Так как эффективная оценка является НОРМД, то по теореме 2.1 она является единственной.

Критерий эффективности оценки.

Пусть условие линейной связи между $T(X_n)$ и $U(X_n, \theta) = \frac{\partial \ln L(X_n, \theta)}{\partial \theta}$, то есть

$$T(X_n) = a(\theta)U(X_n, \theta) + b(\theta)$$

$$\text{Найдем } b(\theta): M(T(X_n)) = a(\theta) \underbrace{M(U(X_n, \theta))}_{=0} + \underbrace{M(b(\theta))}_{=\tau(\theta)} = \tau(\theta) \Rightarrow$$

$$\boxed{T(X_n) - \tau(\theta) = a(\theta)U(X_n, \theta)} \text{ - критерий эффективности,}$$

где $a(\theta)$ - некоторая функция от θ .