Глава 14

Точкови оценки

14.1 Точкови оценки за средно и дисперсия

За целите на настоящата глава ще напомним някои понятия отново.

Нека имаме извадка от независими наблюдения над дадена сл.в.: $X_1, X_2, \ldots X_n \sim X \sim f(x|\theta)$, т.е. от разпределение, чиято вероятностна плътност зависи от неизвестен параметър θ. Основавайки се на нея, какво можем да заключим за параметъра на разпределението θ, или за някаква функция от него $\tau(\theta)$?

Определение 14.1 *Статистика наричаме всяка функция $W(X_1, X_2, \ldots X_n)$ от извадката. Всяка статистика е точкова оценка.*

Оценката на параметъра θ ще отбелязваме с $\hat{\theta}$.

Нека отново да дадем дефинициите на извадъчното средно и извадъчната дисперсия.

Определение 14.2 *Извадъчно средно (оценка на очакването) наричаме:*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Определение 14.3 *Извадъчна дисперсия (оценка на дисперсията) наричаме:*

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

Теорема 14.1 *Нека е дадена извадка $x_1, x_2, \ldots, x_n$ от произволно разпределение. Тогава:*
a) $\min\limits_{a} \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2,.$
b) $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

Доказателство: За да докажем a) прибавяме и изваждаме $\bar{x}$:

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - a)^2$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(x_i - \bar{x})(\bar{x} - a) + \sum_{i=1}^{n}(\bar{x} - a)^2$$
$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 0 + \sum_{i=1}^{n}(\bar{x} - a)^2.$$

Очевидно е, че този израз е минимален за $a = \bar{x}$. За да докажем b), трябва само да положим $a = 0$ в a).

Теорема 14.2 *Нека е дадена извадка $X_1, X_2, \ldots, X_n$ от разпределение с очакване μ и дисперсия σ^2. Тогава:*
a) $\mathbf{E}\bar{X} = \mu$,
b) $Var\bar{X} = \frac{\sigma^2}{n}$,
c) $\mathbf{E}s^2 = \sigma^2$.

Доказателство: За точка a) имаме:

$$\mathbf{E}\bar{X} = \mathbf{E}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\mathbf{E}\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}n\mathbf{E}X_1 = \mu.$$

Аналогично за b):

$$Var\bar{X} = Var\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}Var\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}nVarX_1 = \frac{\sigma^2}{n}.$$

За c) ще използваме резултата от Теорема 14.1:

$$\mathbf{E}s^2 = \mathbf{E}\Big(\frac{1}{n-1}\Big[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\Big]\Big) = \frac{1}{n-1}(n\mathbf{E}X_1^2 - n\mathbf{E}\bar{X}^2)$$
$$= \frac{1}{n-1}\Big(n(\sigma^2 + \mu^2) - n\Big(\frac{\sigma^2}{n} + \mu^2\Big)\Big) = \sigma^2.$$

14.2 Методи за получаване на точкови оценки

14.2.1 Метод на моментите

Нека $X_1, X_2, \ldots X_n \sim f(x|\theta_1, \theta_2, \ldots, \theta_k)$. Определяме всички емпирични m_i (както са дефинирани по-долу) и теоретични μ_i' моменти по следния начин:
$m_1 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^1, \quad \mu_1' = EX^1,$
$m_2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2, \quad \mu_2' =^2,$
$\ldots$
$m_k = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k, \quad \mu_k' =^k.$
Приравняваме ги за съответните степени и образуваме толкова уравнения, колкото са необходими за намиране на неизвестните параметри:
$m_1 = \mu_1'(\theta_1, \theta_2, \ldots, \theta_k),$
$m_2 = \mu_2'(\theta_1, \theta_2, \ldots, \theta_k),$
$\ldots$

$m_k = \mu'_k(\theta_1, \theta_2, \dots, \theta_k)$,
Решението на получената система спрямо $\theta_1, \theta_2, \dots, \theta_k$ е т.нар. оценка по метода на моментите.

Пример 14.1 *Да се намери оценка по метода на моментите за параметрите μ и σ^2 по извадка от нормално разпределение $N(\mu, \sigma^2)$.*
Решение: $m_1 = \bar{X}$, $\mu_1 = EX = \mu \Rightarrow \hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$.

14.2.2 Метод на максималното правдоподобие (ММП)

Определение 14.4 *Функция на правдоподобие на извадката $x_1, x_2, \dots, x_n$ от разпределение с плътност $f(x|\theta_1, \theta_2, \dots, \theta_k)$ наричаме:*

$$L(\theta|\mathbf{x}) = L(\theta_1, \theta_2, \dots, \theta_k | x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \theta_2, \dots, \theta_k).$$

Определение 14.5 *Максимално-правдоподобна оценка (МПО) наричаме стойността на параметъра θ, за която стойността на функцията на правдоподобие е максимална:*

$$\hat{\theta} = \arg\max_{\theta} L(\theta|\mathbf{x}).$$

Решаването на тази задача е с методи от анализа: решаване на уравненията $\frac{\partial}{\partial \theta_i} L(\theta|\mathbf{x}) = 0, \quad i = 1, 2, \dots, k$ за определяне на локални екстремуми, проверка в границите на областта, втори производни и т.н.

Пример 14.2 *Да се намери оценка по ММП за параметъра μ по извадка от нормално разпределение $N(\mu, 1)$.*
Решение:

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu)^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n}(x_i - \mu)^2}$$

Вместо максимума на $\frac{d}{d\mu}L(\mu|\mathbf{x})$, може да търсим този на $\frac{d}{d\mu}\ln L(\mu|\mathbf{x})$, тъй като $\ln$ *е строго растяща функция. Тогава от:*

$$\frac{d}{d\mu}[-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2] = 0$$

следва

$$\sum_{i=1}^{n}(\mu - x_i) = 0,$$

откъдето

$$\sum_{i=1}^{n}\mu = \sum_{i=1}^{n} x_i,$$

или $n\mu = \sum\limits_{i=1}^{n} x_i$, *откъдето* $\mu = \bar{X}$.
Остава да проверим дали $\frac{d^2}{d\mu^2}L(\mu|\mathbf{x}) < 0$ и какви са границите на $L(\mu|\mathbf{x})$ в $\pm\infty$, за да довършим решението на задачата.
Окончателно $\hat{\mu} = \bar{X}$.

14.3 Свойства на точковите оценки

Определение 14.6 *Средноквадратична грешка на дадена оценка W на параметъра θ е функция на θ, дефинирана по следния начин:* $\mathbf{E}_\theta(W-\theta)^2$.

Средноквадратичната грешка може да се разложи на сума от две компоненти - дисперсията на оценката плюс (квадрата на) нейното изместване:

$$\mathbf{E}_\theta(W-\theta)^2 = Var_\theta W + (\mathbf{E}_\theta W - \theta)^2 = Var_\theta W + (Bias_\theta W)^2.$$

Определение 14.7 *Изместване ($Bias_\theta W$) на т.о. W от параметъра θ наричаме разликата $\mathbf{E}_\theta W - \theta$. Ако $Bias_\theta W = 0$, то оценката е неизместена, т.е. $\mathbf{E}_\theta W = \theta$.*

Пример 14.3 *Нека е дадена извадка $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$. Разглеждаме оценките $\bar{X}$ и s^2. За тях от Теорема 14.2 имаме, че $\mathbf{E}\bar{X} = \mu$ и $\mathbf{E}s^2 = \sigma^2$, т.е. те са неизместени. Оттук средноквадратичната грешка и за двете оценки ще бъде равна на дисперсията им: $\mathbf{E}(\bar{X}-\mu)^2 = Var\bar{X} = \frac{\sigma^2}{n}$ и $\mathbf{E}(s^2-\sigma^2)^2 = Var s^2 = \frac{2\sigma^4}{n-1}$ (докажете последното).*

Пример 14.4 *Разглеждаме оценката $\hat{\sigma}^2 = \frac{n-1}{n}s^2$. За нея $\mathbf{E}\hat{\sigma}^2 = \frac{n-1}{n}\sigma^2$ и $Var\hat{\sigma}^2 = \left(\frac{n-1}{n}\right)^2 Var s^2 = \frac{2(n-1)}{n^2}\sigma^4$. Тогава:*

$$\mathbf{E}(\hat{\sigma}^2-\sigma^2)^2 = \frac{2(n-1)}{n^2}\sigma^4 + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \frac{2n-1}{n^2}\sigma^4$$
$$< \frac{2}{n-1}\sigma^4 = E(s^2-\sigma^2)^2.$$

Виждаме, че неизместената оценка s^2 се оказа с по-голяма дисперсия от изместената σ^2, а оттам и по-голяма средноквадратична грешка. Тази "размяна" между дисперсия и изместеност е една от причините за трудностите при определянето на оценка с минимална средноквадратична грешка. Изобщо класът от всички оценки е твърде широк и за да можем да определим по-добри и най-добри оценки е необходимо да го ограничим. Един възможен подход е да разгледаме класа от неизместените оценки.

Определение 14.8 *Дадена оценка W^* се нарича най-добра неизместена оценка (или равномерно неизместена оценка с минимална дисперсия) за $\tau(\theta)$, ако удовлетворява $\mathbf{E}_\theta W^* = \tau(\theta)$, за $\forall\theta$ и за всяка друга оценка W, за която $\mathbf{E}_\theta W = \tau(\theta)$, имаме $Var_\theta W^* \le Var_\theta W, \quad \forall\theta$.*

Възможно е да ограничим отдолу дисперсията на коя да е неизместена оценка за даден параметър. Такава граница ни дава неравенството на Рао - Крамер. Ако намерим оценка, която достига тази граница, то тя ще бъде най-добра неизместена оценка за този параметър.

Теорема 14.3 (неравенство на Рао - Крамер) *Нека е дадена извадка $X_1, X_2, \ldots, X_n \sim f(\mathbf{x}|\theta)$ и нека $W(\mathbf{X}) = W(X_1, X_2, \ldots, X_n)$ е коя да е оценка, удовлетворяваща условията:*

$$\frac{d}{d\theta}\mathbf{E}_\theta W(\mathbf{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}[W(\mathbf{x})f(\mathbf{x}|\theta)]d\mathbf{x}$$

и

$$Var_\theta(\mathbf{X}) < \infty.$$

Тогава

$$Var_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta}\mathbf{E}_\theta W(\mathbf{X}))^2}{\mathbf{E}_\theta\big((\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta))^2\big)}.$$

Доказателство: Доказателството се базира на неравенството на Коши за сл.в., т.е. факта, че ако имаме две сл.в. X и Y, то за тях е изпълнено:

$$[Cov(X,Y)]^2 \leq (VarX)(VarY), \tag{14.3.1}$$

или по друг начин:

$$VarX \geq \frac{[Cov(X,Y)]^2}{VarY}, \tag{14.3.2}$$

като долна граница на дисперсията на X. Ще положим в (14.3.2) $X = W(\mathbf{X})$ и $Y = \frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)$ и ще приложим неравенството на Коши.

Да отбележим, че:

$$\frac{d}{d\theta}\mathbf{E}_\theta W(\mathbf{X}) \qquad = \int_{\mathcal{X}} W(\mathbf{x})\left[\frac{\partial}{\partial\theta}f(\mathbf{x}|\theta)\right]d\mathbf{x}$$

$$= \mathbf{E}_\theta\left[W(\mathbf{X})\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right] \quad \text{след като сме вмъкнали втория множител под диференциала}$$

$$= \mathbf{E}_\theta\left[W(\mathbf{X})\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right], \tag{14.3.3}$$

което е първият елемент от ковариацията на $W(\mathbf{X})$ и $\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)$. За да я пресметнем, ни трябва произведението на двете очаквания. Но, ако в (14.3.3) положим $W(\mathbf{x}) = 1$, то получаваме, че:

$$\mathbf{E}_\theta(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)) = \frac{d}{d\theta}\mathbf{E}_\theta[1] = 0,$$

и така за ковариацията получихме, че е равна на очакването на произведението. Оттук:

$$Cov_\theta\left(W(\mathbf{X}), \frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right) = \mathbf{E}_\theta\left[W(\mathbf{X})\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right] = \frac{d}{d\theta}E_\theta W(\mathbf{X}).$$

Освен това, от $\mathbf{E}_\theta(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)) = 0$ следва, че:

$$Var_\theta\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right) = \mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right).$$

Прилагайки неравенството на Коши (14.3.1), получаваме:

$$Var_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta}\mathbf{E}_\theta W(\mathbf{X}))^2}{\mathbf{E}_\theta\big((\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta))^2\big)}.$$

Следствие: Ако $X_1, X_2, \ldots, X_n \sim f(x|\theta)$ са н.е.р. сл.в., то неравенството се опростява:

$$Var_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta}\mathbf{E}_\theta W(\mathbf{X}))^2}{n\mathbf{E}_\theta\big((\frac{\partial}{\partial\theta}\log f(X|\theta))^2\big)}.$$