Lecture 12

参数估计

ruanyl@buaa.edu.cn
September 2016, Beihang University

12.1　统计推断

统计推断是建立在已有观测事实的基础上对未知量进行概率估计的过程. 例如通过抽样对一个未知分布的期望, 方差, 分为函数, 相关事件概率进行估计.

例题 12.1.1 某电池生产商认为, 所生产的电池使用寿命 X 服从参数为 θ 的指数分布, 其中参数 θ 为电池消耗率 θ(产品数/总使用寿命). 于是生产商检测了一列电池, 并得到其使用寿命 $X_1, \ldots, X_n$, 他们希望通过检测到的数据来

(1) 估计产品使用寿命分布中的参数 θ

(2) 对将来生产的电池使用寿命进行预测

如果参数 θ 已知, 那么电池的平均使用寿命 $E(X) = 1/\theta$, 可见 θ 的含义也正是电池生产商认为的电池消耗率.

12.2 点估计方法

矩估计

假设随机变量 $X_1, ..., X_n$ 为随机样本, 总体 X 的密度函数为 $f(x|\theta), \theta = (\theta_1, ..., \theta_k)^T$. 矩估计方法是用样本矩代替总体矩, 并将参数用各阶样本矩进行表示. 具体地令

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \approx E(X|\theta_1, ..., \theta_k)$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \approx E\left(X^2|\theta_1, ..., \theta_k\right)$$

$$\cdots\cdots$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \approx E\left(X^k|\theta_1, ..., \theta_k\right)$$

然后求解 $(\theta_1, ..., \theta_k)^T$ 并作为估计量.

例题 12.2.1 假设随机变量 $X_1, ..., X_n$ 为随机样本, 总体 $X \sim \mathcal{N}(\mu, \sigma^2)$. 试求出 (μ, σ^2) 的矩估计量.

- 注意 $E(X) = \mu, E(X^2) = \sigma^2 + \mu^2$. 建立方程组

$$m_1 = E(X) = \mu, m_2 = E(X^2) = \sigma^2 + \mu^2.$$

求解得到

$$\mu = m_1, \sigma^2 = m_2 - m_1^2.$$

因此 (μ, σ^2) 的矩估计量可写作 $\tilde{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$,

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

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

极大似然估计

假设随机变量 $X_1, ..., X_n$ 为随机样本, 总体 X 的密度函数为 $f(x|\theta), \theta = (\theta_1, ..., \theta_k)^T$. 给定样本值 $x = (x_1, ..., x_n)^T$, **似然函数**定义为

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$$

可以找到参数 $\tilde{\theta}(x)$ 使得 $L(\theta|x)$ 最大化.

定义 12.2.1 $\tilde{\theta}(x)$ 称为基于随机样本 $X_1, ..., X_n$ 的**最大似然估计 (MLE)**.

例题 12.2.2 假设随机变量 $X_1, ..., X_n$ 为随机样本, 总体 $X \sim \mathcal{N}(\mu, 1)$. 试求出 μ 的最大似然估计.

- 似然函数为

$$L(\mu|x) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x_i-\mu)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

由极值的充分条件

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

从而 $\tilde{\mu}=\overline{x}_n$. 由于 $L(\pm\infty|x)=0$,

$$\frac{d^2L(\mu|x)}{d\mu^2}=\frac{1}{(2\pi)^{n/2}}\left[-n+\left(\sum_{i=1}^{n}(\mu-x_i)\right)^2\right]e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

可见

$$\left.\frac{d^2L(\mu|x)}{d\mu^2}\right|_{\mu=\tilde{\mu}}<0$$

因此 $\tilde{\mu}=\overline{x}_n$ 是最大值.

例题 12.2.3 假设随机变量 $X_1,\ldots,X_n$ 为随机样本, 总体 $X\sim\mathcal{N}(\mu,\sigma^2)$. 试求出 μ,σ^2 的 MLE.

■ 似然函数为

$$L\left(\mu,\sigma^2|x\right)=\prod_{i=1}^{n}\frac{1}{\left(2\pi\right)^{n/2}\sigma^n}e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}=\frac{1}{\left(2\pi\right)^{n/2}\sigma^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

为了求得最大值, 我们考虑

$$\ln L\left(\mu,\sigma^2|x\right)=-\frac{n}{2}\left[\ln\left(2\pi\right)+\ln\left(\sigma^2\right)\right]-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(x_i-\mu\right)^2}{\sigma^2}$$

显然 $L\left(\mu,\sigma^2|x\right)$ 与 $\ln L\left(\mu,\sigma^2|x\right)$ 将同时达到最大值. 分别对 μ,σ^2 求导数得到

$$\frac{\partial\ln L\left(\mu,\sigma^2|x\right)}{\partial\mu}=\sum_{i=1}^{n}\frac{\mu-x_i}{\sigma^2}=0$$

$$\frac{\partial\ln L\left(\mu,\sigma^2|x\right)}{\partial\left(\sigma^2\right)}=-\frac{n}{2\sigma^2}+\frac{1}{2}\sum_{i=1}^{n}\frac{\left(x_i-\mu\right)^2}{\sigma^4}=0$$

解之得到

$$\tilde{\mu}=\overline{x}_n,\tilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2$$

为了验证 $\tilde{\mu},\tilde{\sigma}^2$ 确实为最大值点, 注意到

$$\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2\leqslant\sum_{i=1}^{n}\left(x_i-\mu\right)^2,\forall\mu,$$

从而

$$\frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}\leqslant\frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}_n}{\sigma}\right)^2},\forall\sigma^2$$

因此只需证明上述不等式右侧在 $\tilde{\sigma}^2$ 处达到最大值即可, 而这可以从 $\frac{\partial\ln L\left(\mu,\sigma^2|x\right)}{\partial(\sigma^2)}=0$ 的方程式看出. 因此 $\tilde{\mu},\tilde{\sigma}^2$ 确实为 MLE.

12.3 无偏估计的优良性

定义 12.3.1 **统计量** $T(X_1,\ldots,X_n)$ 称为待估参数 θ 的**无偏估计 (UE)**, 如果 $E(T(X_1,\ldots,X_n))=\theta,\ \forall\theta.$

均方误差

定义 12.3.2 无偏估计的**均方误差 (MSE)** 定义为

$$E\left[(T(X_1,\ldots,X_n)-\theta)^2\right]$$

注意

$$\begin{aligned} E\left[(T-\theta)^2\right] &= E\left[(T-E(T)+E(T)-\theta)^2\right] \qquad (12.1)\\ &= Var(T)+(E(T)-\theta)^2 \end{aligned}$$

因此对无偏估计量,

$$E\left[(T-\theta)^2\right]=Var(T)$$

即对无偏估计量来说, 其方差就是 MSE.

例题 12.3.1 假设随机变量 $X_1, \ldots, X_n$ 为随机样本, 总体 $X \sim \mathcal{N}(\mu, \sigma^2)$. 那么

$$\tilde{\mu}_{UE} = \overline{X}_n, \tilde{\sigma}^2_{UE} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

分别是 μ, σ^2 的无偏估计.

习题 12.3.1 试证明 $\tilde{\sigma}^2_{UE} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$ 的方差

$$Var\left(\tilde{\sigma}^2_{UE}\right) = \frac{2\sigma^4}{n-1}$$

例题 12.3.2 假设随机变量 $X_1, \ldots, X_n$ 为随机样本, 总体 $X \sim \mathcal{N}(\mu, \sigma^2)$. 那么 σ^2 的最大似然估计

$$\tilde{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

具有比 $\tilde{\sigma}^2_{UE}$ 更小的 MSE.

- 要说明

$$E\left[\left(\tilde{\sigma}^2_{MLE} - \sigma^2\right)^2\right] \leqslant E\left[\left(\tilde{\sigma}^2_{UE} - \sigma^2\right)^2\right]$$

$\tilde{\sigma}^2_{UE}$ 的 MSE 正是它的方差

$$E\left[\left(\tilde{\sigma}^2_{UE} - \sigma^2\right)^2\right] = Var\left(\tilde{\sigma}^2_{UE}\right) = \frac{2\sigma^4}{n-1}$$

另外

$$E\left(\tilde{\sigma}^2_{MLE}\right) = E\left(\frac{n-1}{n}\tilde{\sigma}^2_{UE}\right) = \frac{n-1}{n}\sigma^2$$

$$Var\left(\tilde{\sigma}^2_{MLE}\right) = Var\left(\frac{n-1}{n}\tilde{\sigma}^2_{UE}\right) = \frac{2(n-1)\sigma^4}{n^2}$$

因此 $\tilde{\sigma}^2_{MLE}$ 的 MSE 是 (Equation 12.1)

$$E\left[\left(\tilde{\sigma}^2_{MLE} - \sigma^2\right)^2\right] = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2$$

$$= \frac{(2n-1)\,\sigma^4}{n^2}$$

可见

$$E\left[\left(\tilde{\sigma}^2_{MLE} - \sigma^2\right)^2\right] = \frac{(2n-1)\,\sigma^4}{n^2} < \frac{2n\sigma^4}{n^2} = \frac{2\sigma^4}{n}$$

$$< \frac{2\sigma^4}{n-1} = E\left[\left(\tilde{\sigma}^2_{UE} - \sigma^2\right)^2\right]$$

最优无偏估计

定义 12.3.3 假设 $T_1(X_1, ..., X_n)$ 和 $T_2(X_1, ..., X_n)$ 均为待估参数 θ 的无偏估计, 如果

$$Var(T_1|\theta) \leqslant Var(T_2|\theta), \forall\theta$$

那么称估计量 T_1 优于 T_2.

12.4 置信区间

正态分布的枢轴量及其分布

待估参数	其他参数	枢轴量 V	V 的分布
μ	σ^2 已知	$\dfrac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$
μ	σ^2 未知	$\dfrac{\bar{X}_n-\mu}{S/\sqrt{n}}$	$t(n-1)$
σ^2	μ 已知	$\sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2$	$\chi^2(n)$
σ^2	μ 未知	$\sum_{i=1}^{n}\left(\dfrac{X_i-\bar{X}_n}{\sigma}\right)^2$	$\chi^2(n-1)$

总体分布为 $\mathcal{N}(\mu,\sigma^2)$ 时的枢轴量. $\bar{X}_n=\dfrac{X_1+X_2+\cdots+X_n}{n}$, $S=\left[\dfrac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{n-1}\right]^{1/2}$

枢轴法

定义 12.4.1 假设随机变量 $X_1,\ldots,X_n$ 为随机样本, 总体 X 的分布依赖于参数 $\theta=(\theta_1,\ldots,\theta_k)^T$. g 是实数值函数. A,B 为满足下述关系式的统计量

$$P\left(A<g\left(\theta\right)<B\right)\geqslant\gamma,\forall\theta$$

那么称 (A,B) 为 $g\left(\theta\right)$ 的置信度为 γ 的置信区间. 若

$$P\left(g\left(\theta\right)<B\right)\geqslant\gamma,\forall\theta$$

那么称 B 为 $g\left(\theta\right)$ 的置信度为 γ 的置信上限. 若

$$P\left(A<g\left(\theta\right)\right)\geqslant\gamma,\forall\theta$$

那么称 A 为 $g\left(\theta\right)$ 的置信度为 γ 的置信下限.

定义 12.4.2 假设随机变量 $X_1,\ldots,X_n$ 为随机样本, 总体 X 的分布依赖于参数 $\theta=(\theta_1,\ldots,\theta_k)^T$. $V\left(X_1,\ldots,X_n,\theta\right)$ 是随机变量, 它的分布不依赖于 θ, 那么称 $V\left(X_1,\ldots,X_n,\theta\right)$ 为枢轴量.

例题 12.4.1 某车间生产滚珠, 由于机器性能不同, 致使实际滚珠直径 X 有所差异, 经验表明 $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, 其中 μ 由设计决定, σ^2 由机器性能决定. 为了调查机器性能, 我们抽取了 6 粒滚珠, 测量得到直径 (mm) 为: 14.7, 15.21, 14.90, 14.91, 15.32, 15.32. 试求出 σ^2 的置信度 0.95 的置信上限.

- 由于 μ, σ^2 均未知, 可选区枢轴量

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

对 $\alpha > 0$, 令 $\chi^2_\alpha(n-1)$ 为 $\chi^2(n-1)$ 对应于 α 的分位数,

$$P\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 > \chi^2_\alpha(n-1)\right) = 1-\alpha$$

即

$$P\left(\sigma^2 < \frac{1}{\chi^2_\alpha(n-1)}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right) = 1-\alpha$$

根据题意, 取 $\alpha = 0.05$, $n = 6$, 计算得 $\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = 0.3554$, 查表得 $\chi_{\alpha}^2(n-1) = 1.1455$, 因此 σ^2 的置信度 0.95 的置信上限为 $0.3554/1.1455 = 0.3103$.