

## §7.4 无偏估计的优良性

- 假设  $T = T(X_1, \dots, X_n)$  是  $g(\theta)$  的估计量, 称

$$R(\theta, T) := E_{\theta}(T - g(\theta))^2$$

为  $T$  的均方误差/风险函数.

- 定义4.1. 假设  $T = T(X_1, \dots, X_n)$  是  $g(\theta)$  的估计量. 如果

(1)  $T$  是  $g(\theta)$  的无偏估计,

(2) 对于  $g(\theta)$  的任意无偏估计  $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$ , 都有

$$\text{var}_{\theta}(T) \leq \text{var}_{\theta}(\tilde{T}), \quad \forall \theta \in \Theta.$$

则称  $T$  为  $g(\theta)$  的(一致)最小方差无偏估计(Uniformly Minimum Variance Unbiased, UMVU).

- 定义4.2. 假设统计量  $T = T(X_1, \dots, X_n)$  满足:

对任意统计量  $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$  都有,

在  $T = t$  的条件下,  $\tilde{T}$  的分布与参数  $\theta$  无关,

那么, 称  $T$  为充分统计量.

- 即, 在  $T = t$  的条件下,  $(X_1, \dots, X_n)$  的分布与参数  $\theta$  无关.
- 定理4.1. 若  $T = T(X_1, \dots, X_n)$  满足:

$$\prod_{i=1}^n p(x_i, \theta) = q_{\theta}(T(x_1, \dots, x_n))h(x_1, \dots, x_n),$$

则  $T$  是充分统计量.

例4.3. 总体:  $B(1, p)$ , 样本量:  $n$ . 考虑  $T = X_1 + \cdots + X_n$ .

- 对  $t = 0, 1, \dots, n$ , 令

$$S_t := \{(x_1, \dots, x_n) : x_i \in \{0, 1\}, \forall i; \text{ 且 } x_1 + \cdots + x_n = t\}.$$

- 那么,  $\forall (x_1, \dots, x_n) \in S_t$ ,

$$P_p(X_1 = x_1, \dots, X_n = x_n | T = t)$$

$$\stackrel{=}=\frac{P_p(X_1 = \mathbf{x}_1, \dots, X_n = \mathbf{x}_n)}{P_p(T = t)} \stackrel{=}=\frac{p^t(1-p)^{n-t}}{C_n^t p^t (1-p)^{n-t}} \stackrel{=}=\frac{1}{C_n^t}.$$

- $\forall t$ , 在  $T = t$  的条件下,  $(X_1, \dots, X_n)$  服从  $S_t$  上的均匀分布. 该分布与  $p$  无关. 因此,  $T$  是充分统计量.

例4.6. 总体:  $X \sim \text{Exp}(\lambda)$ , 样本量:  $n$ .

- 联合密度为

$$p_{\lambda}(x_1, \dots, x_n) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \cdot \mathbf{1}_{\{x_1, \dots, x_n > 0\}}.$$

- 令  $T = T(X_1, \dots, X_n) = X_1 + \dots + X_n$ . 则  $T$  是充分统计量.
- 则  $\forall t > 0$ , 在  $T = t$  的条件下,  $(X_1, \dots, X_n)$  服从

$$S_t = \{(x_1, \dots, x_n) : x_i > 0, \forall i; \text{ 且 } x_1 + \dots + x_n = t\}$$

上的均匀分布.

例4.4. 总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ .

- 联合密度为

$$\begin{aligned}\prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right\}.\end{aligned}$$

因此,  $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$  是充分统计量.

- 若总体改为  $X \sim N(\mu, 1)$ , 则联合密度改为

$$\begin{aligned}\prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi}^n} \exp\left\{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2}\right\} \\ &= \frac{1}{\sqrt{2\pi}^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} \exp\left\{\mu \sum_{i=1}^n x_i\right\} \exp\left\{-\frac{n}{2} \mu^2\right\}.\end{aligned}$$

因此,  $\bar{X}$  是充分统计量.

例4.4.(续) & 4.7. 总体:  $X \sim N(\mu, 1)$ , 样本量:  $n$ .

- 已有,  $T_1 = \bar{X}$  是充分统计量.
- $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$  也是充分统计量.
- $T_2 = (\bar{X}, X_1 - X_2)$  也是充分统计量.
- $\psi_1(T_1) := T_1$  是  $\mu$  的无偏估计.
- 令  $\phi(T_2) = X_1 - X_2$ , 则  $E_\mu \phi(T_2) = 0, \forall \mu$ .  
因此,  $\psi_2(T_2) := T_1 + X_1 - X_2$  也是  $\mu$  的无偏估计.
- 用  $T_2$  造的无偏估计不好:

$$\begin{aligned} & \text{var}_\mu(T_1 + X_1 - X_2) \\ &= \text{var}_\mu \left( \frac{n+1}{n} X_1 - \frac{n-1}{n} X_2 + \frac{1}{n} X_3 + \cdots + \frac{1}{n} X_n \right) \\ &= \frac{(n+1)^2 + (n-1)^2 + n - 2}{n^2} = 2 + \frac{1}{n} > \frac{1}{n} = \text{var}_\mu(T_1). \end{aligned}$$

- 定义4.3. 设 $T = T(X_1, \dots, X_n)$  为充分统计量. 若对任意 $\phi$ ,

$$E_{\theta}\phi(T) = 0, \forall \theta \in \Theta \text{ 可推出 } P_{\theta}(\phi(T) = 0) = 1, \forall \theta \in \Theta.$$

则称 $T$  为完全充分统计量.

- 定理4.2. 设 $T = T(X_1, \dots, X_n)$  为完全充分统计量. 若

$$E_{\theta}\phi(T) = g(\theta), \forall \theta,$$

则 $\phi(T)$  是 $g(\theta)$  的UMVU 估计.

- 定义4.4. 若密度或分布列 $p(x, \theta)$  能进行如下分解:

$$p(x, \theta) = S(\theta)h(x) \exp \left\{ \sum_{k=1}^m C_k(\theta)T_k(x) \right\},$$

则称 $p(x, \theta), \theta \in \Theta$  为指数族分布.

- 注:  $x$  可为高维向量, 于是 $p(x, \theta)$  为联合密度/联合分布列.
- 引理4.1. 若总体 $X$  是指数族, 则样本 $(X_1, \dots, X_n)$  也是.
- 定理4.3. 总体分布如上;  $\Theta \in \mathbb{R}^m$  且含内点;  $(C_1, \dots, C_m)$  在 $\Theta$  上一对一、连续; 诸 $C_i$  之间( $T_i$  之间)无线性关系. 则

$$\left( \sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

是完全充分统计量.



例4.9 & 4.14. 总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ .

- $\theta = (\mu, \sigma^2)$ .  $m = 2$ :

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2}.$$

- $T_1 = \sum_{i=1}^n X_i$ ,  $T_2 = \sum_{i=1}^n X_i^2$ , 则  $(T_1, T_2)$  是完全充分统计量.
- $\bar{X}$ ,  $S^2$  是  $\mu$ ,  $\sigma^2$  的UMVU 估计:

$$\bar{X} = \frac{1}{n}T_1, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} (T_2 - \frac{1}{n}T_1^2)$$

是  $(T_1, T_2)$  的函数, 且是  $\mu$ ,  $\sigma^2$  的无偏估计.

例4.9 & 4.14(续).

- 改为已知 $\mu$  (例如, 已知 $\mu = 1$ ). 则 $\theta = \sigma^2$ ,  $m = 1$ :

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-1)^2}.$$

- $T_1 = \sum_{i=1}^n (X_i - \mu)^2$  是完全充分统计量.
- $\hat{\sigma}^2$  是 $\sigma^2$  的UMVU 估计, 其中

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

例4.15. 总体:  $X \sim N(\mu, 1)$ , 样本量:  $n$ , 求  $\mu^2$  的UMVU 估计.

- $m = 1, \theta = \mu$ :

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2} e^{-\frac{1}{2}x^2} e^{\mu x}.$$

- $T_1 = \sum_{i=1}^n X_i$  是完全充分统计量, 因此  $\bar{X}$  也是.
- 由  $\text{var}(Y) = EY^2 - (EY)^2$  知,

$$\mu^2 = (E_{\mu} \bar{X})^2 = E_{\mu} \bar{X}^2 - \text{var}_{\mu}(\bar{X}) = E_{\mu} \bar{X}^2 - \frac{1}{n} = E_{\mu} \left( \bar{X}^2 - \frac{1}{n} \right).$$

- 因此,  $\bar{X}^2 - \frac{1}{n}$  是  $\mu^2$  的UMVU 估计.

例4.10 & 4.13. 某工人生产20 件产品, 其中1 件为次品.  
求: 次品率的UMVU 估计.

- 总体:  $Y \sim B(1, p)$ ,  $p = \theta \in [0, 1]$ , 样本量:  $n = 20$ .
- 分布列: (记  $k = y_1 + \cdots + y_n$ )

$$\begin{aligned} P_p(Y_1 = y_1, \cdots, Y_n = y_n) &= \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} \\ &= p^k (1-p)^{n-k} = e^{k \log p + (n-k) \log(1-p)} = e^{n \log(1-p)} e^{(\log p - \log(1-p))k}. \end{aligned}$$

- 或者, 总体:  $X \sim B(20, p)$ ,  $p = \theta \in [0, 1]$ , 样本量: 1. 分布列:

$$P_p(X = k) = C_n^k p^k (1-p)^{n-k} = e^{n \log(1-p)} C_n^k e^{(\log p - \log(1-p))k}.$$

- $T_1 = X = Y_1 + \cdots + Y_{20}$  是完全充分统计量.
- 因此,  $\hat{p} = X/20$  是UMVU 估计.