复习.

- 总体$X \sim F_\theta$, $\theta \in \Theta$.
- 样本$\vec{X} = (X_1, \cdots, X_n)$,
- 样本(值) $\vec{x} = (x_1, \cdots, x_n)$.
- 目标: 用统计量$T(X_1, \cdots, X_n)$ 作为$\theta$ (或$g(\theta)$) 的估计.
- 最大似然估计(ML估计):

  $L(\theta) = \prod_{i=1}^n p_\theta(x_i)$ 在$\Theta$ 中的最大值点$\hat{\theta}$.

总体$X \sim F_\theta$. 目标: 给出$\theta$ 的估计值$\hat{\theta}$.

思想: 样本矩= 真实矩. 支撑: LLN.

- 总体矩: $m_k(\theta) = EX^k$, 参数的函数.

  样本矩: $a_k = \frac{1}{n}\sum_{i=1}^n X_i^k$ 或$\frac{1}{n}\sum_{i=1}^n x_i^k$, 数据的函数(统计量).

  令:

  $$m_k(\hat{\theta}) = a_k, \quad k = 1, \cdots r.$$

  (其中$r$是使得方程组有唯一解的最小整数), 唯一解$\hat{\theta}$称为$\theta$的矩估计.

- **习题一、11(续).** $N$ 条鱼中有80条有标记, 100条中4条有标记. 估计$N$.

  $n = 1$, 总体矩$m_1 = m_1(N) = 100 * \frac{80}{N}$, 样本矩$a_1 = 4$.

  求解$m_1(\hat{N}) = a_1$, 得$\hat{N} = 2000$.

**例1.1(续)** 飞机最大飞行速度$X \sim N(\mu, \sigma^2)$, 数据$x_1, \cdots, x_{15}$. 估计$\mu$, $\sigma^2$.

- 两个参数, 则用前两阶矩:
- 总体矩: $m_1 = \mu$, $m_2 = \sigma^2 + \mu^2$.
  样本矩: $a_1 = \bar{x}$, $a_2 = \frac{1}{n}\sum_{i=1}^n x_i^2$.
- 求解

$$\begin{cases} m_1 = \bar{x}, \\ m_2 = \frac{1}{n}\sum_{i=1}^n x_i^2, \end{cases} \quad \text{即} \begin{cases} \hat{\mu} = \bar{x}, \\ \widehat{\sigma^2} + \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2. \end{cases}$$

  解得$\hat{\mu} = \bar{x}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$.
- 定义2.1. 若待估量$g(\theta) = \phi(m_1, \cdots, m_k)$(不一定是参数本身). 定义$g(\theta)$的矩估计为$\widehat{g(\theta)} := \phi(a_1, \cdots, a_k)$.

## §7.3 估计的无偏性

称$T(X_1,\cdots,X_n)$ 是$g(\theta)$ 的无偏估计, 如果

$$E_\theta T(X_1,\cdots,X_n)=g(\theta),\ \forall\theta\in\Theta.$$

- 样本均值$\hat{\mu}=\bar{X}=\frac{1}{n}(X_1+\cdots+X_n)$.

  $E_\theta\hat{\mu}=\frac{1}{n}(E_\theta X_1+\cdots+E_\theta X_n)=\mu$, 无偏.
- 样本方差$\widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2$.
  - $\frac{1}{n}\sum_{i=1}^n(x_i-\mu)^2=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2+(\mu-\bar{x})^2$.
  - $\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2=\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2-(\mu-\bar{X})^2$.
  - $E_\theta\widehat{\sigma^2}=\mathrm{var}(X_1)-\mathrm{var}(\bar{X})=\sigma^2-\frac{1}{n}\sigma^2=\frac{n-1}{n}\sigma^2$.
  - 估小了: $Y_i=X_i-\bar{X}$, $\forall i$ 不独立, 一个约束条件$\sum_i Y_i=0$.
- $S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ 是$\sigma^2$ 的无偏估计. (定理3.1)

  有些书称$S^2$ 为样本方差.

**例1.4+2.4.** $X \sim U[0,\theta]$, 数据$x_1,\cdots,x_n$. 求$\theta$ 的最大似然估计$\hat{\theta}_1$ 与矩估计$\hat{\theta}_2$.

**最大似然估计:**

- 似然函数$L(\theta)=\prod_{i=1}^n \frac{1}{\theta}1_{0\leqslant x_i\leqslant\theta}$.
- 其最大值点$\hat{\theta}_1$ 需使得$0\leqslant x_i\leqslant\hat{\theta}_1$, $\forall i$ 都成立, 即$\hat{\theta}_1\geqslant \max_i x_i$.
- 进一步, $\hat{\theta}_1$ 使得$\frac{1}{\theta^n}$ 达到最大, 即$\hat{\theta}_1$ 尽量小, 故$\hat{\theta}_1=\max_i x_i$.
- 分析$\hat{\theta}_1=\max_i X_i$:
  - 缺点: $\hat{\theta}_1=\max_i X_i<\theta$, 故$E_\theta\hat{\theta}_1<\theta$, 估小了!
  - 进一步计算:

$$\begin{aligned}
E_\theta\hat{\theta}_1 &= E_\theta \max_i X_i = \int_0^\theta P(\max_{1\leqslant i\leqslant n} X_i > x)dx\\
&= \int_0^\theta (1-P(\max_{1\leqslant i\leqslant n} X_i\leqslant x))dx\\
&= \theta-\int_0^\theta(\tfrac{1}{\theta}x)^n dx=(1-\tfrac{1}{n+1})\theta.
\end{aligned}$$

## 矩估计：

- 总体矩$m_1(\theta) = \frac{1}{2}\theta$, 样本矩$\bar{x}$.
- 求解$m_1(\hat{\theta}_2) = \bar{x}$, 得$\hat{\theta}_2 = 2\bar{x}$.
- 分析$\hat{\theta}_2 = 2\bar{X}$：
  - $E_\theta \hat{\theta}_2 = 2E_\theta \bar{X} = 2E_\theta X_1 = 2 \times \frac{1}{2}\theta = \theta$, 无偏.
  - 缺点: 有可能$2\bar{x} < \max_i x_i$, 明显不合理.
- 不能单纯追求无偏性.

## §7.4 无偏估计的优良性

假设$\hat{\theta}_1$, $\hat{\theta}_2$ 都是$\theta$ 的无偏估计, 即$E_\theta\hat{\theta}_1 = E_\theta\hat{\theta}_2 = \theta$.

若$\mathrm{var}_\theta(\hat{\theta}_1) \leqslant \mathrm{var}_\theta(\hat{\theta}_2)$, 则认为$\hat{\theta}_1$ 更好.

**例.** $X \sim \mathrm{Exp}(\lambda)$. 样本$X_1, \cdots, X_n$. 估计$\theta = \frac{1}{\lambda} = EX$.

- 最大似然估计与矩估计: $\hat{\theta}_1 = \bar{X}$.
- 令$\hat{\theta}_2 = n \min_{1\leqslant i\leqslant n} X_i$.
    - 理由: $\min_{1\leqslant i\leqslant n} X_i \sim \mathrm{Exp}(n\lambda)$, (第七章, 例4.8)

      $\hat{\theta}_2 \sim \mathrm{Exp}(\lambda)$, 因为

      $$P(\hat{\theta}_2 > x) = P(\min_{1\leqslant i\leqslant n} X_i > \frac{1}{n}x)^n = e^{n\times(-\lambda\frac{1}{n}x)}.$$

      $E_\theta\hat{\theta}_2 = \frac{1}{\lambda} = \theta$. 无偏!
- $\mathrm{var}_\theta(\hat{\theta}_1) = \frac{1}{n}\mathrm{var}(X) = \frac{1}{n}\frac{1}{\lambda^2} = \frac{1}{n}\theta^2$.

  $\mathrm{var}_\theta(\hat{\theta}_2) = \frac{1}{\lambda^2} = \theta^2$.

  故, $\hat{\theta}_1$ 比$\hat{\theta}_2$ 好.

## (一致)最小方差无偏估计

**(Uniformly Minimum Variance Unbiased Estimator, UMVUE):**

假设$T = T(X_1, \cdots, X_2)$ 是$\theta$ 的无偏估计, 且对于$\theta$ 的任意无偏估计$\tilde{T} = \tilde{T}(X_1, \cdots, X_n)$ 都有

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

则称$T = T(X_1, \cdots, X_2)$ 是$\theta$ 的(一致)最小方差无偏估计.

指数族分布: 分布密度(或分布列)为

$$p_\theta(x) = S(\theta)h(x)\exp\{\sum_{k=1}^m C_k(\theta)T_k(x)\}.$$

- $\theta$ 与$x$ 分离. **例4.8, 4.9, 4.10.**
- 指数: $p_\lambda(x) = \lambda e^{-\lambda x}1_{\{x>0\}} = \lambda 1_{\{x>0\}}e^{-\lambda x}$.
- 正态: $p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\mu^2}{2\sigma^2}}e^{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x}$
- 二项: $P_{n,p}(X=k) = C_n^k p^k(1-p)^{n-k} =$ $C_n^k e^{k\log p+(n-k)\log(1-p)} = C_n^k e^{(\log p-\log(1-p))k+n\log(1-p)}$
- 泊松: $P_\lambda(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}1_{\{k\geqslant 0\}} = \frac{1}{k!}1_{\{k\geqslant 0\}}e^{(\log\lambda)k-\lambda}$.

指数族分布特性: 假设总体$X\sim p_\theta(x)$, $X_1,\cdots X_n$为$X$的样本, 则$(X_1,\cdots,X_n)$的密度(或分布列)为

$$\prod_{i=1}^n p_\theta(x_i) = S^n(\theta)\prod_{i=1}^n h(x_i)\exp\{\sum_{k=1}^m C_k(\theta)\sum_{i=1}^n T_k(x_i)\}$$

假设总体$X$密度(或分布列)为

$$p_\theta(x) = S(\theta)h(x)\exp\{\sum_{k=1}^{m} C_k(\theta)T_k(x)\}.$$

**定理(4.2+4.3)**. 若

- $\Theta$ 是$\mathbb{R}^m$ 中有内点的集合.
- $(C_1, \cdots C_m) : \Theta \to \mathbb{R}^m$ 一一对应, 连续.
- $C_k(\theta)$, $1 \leqslant k \leqslant m$, 线性无关; $T_k(x)$, $1 \leqslant k \leqslant m$, 线性无关. (例如, 当$T_1(x) = x$, $T_2(x) = x^2$, $T_3(x) = x + x^2$ 时, 应化简为$(C_1(\theta) + C_3(\theta))x + (C_2(\theta) + C_3(\theta))x^2$.)
- $\hat{\theta} = \phi(T_1(\vec{X}), \cdots, T_m(\vec{X}))$ 是$\theta$ (或$g(\theta)$) 的无偏估计.

则$\hat{\theta}$ 是UMVUE.

注: $T_k(\vec{X}) := \sum_{i=1}^{n} T_k(X_i)$, $\quad k = 1, \cdots, m$.

**例4.14.** $X \sim N(\mu, \sigma^2)$.

- $\theta = (\mu, \sigma^2)$ 二维, $m = 2$.
- $p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x}$.

  $C_1(\theta) = \frac{\mu}{\sigma^2}$, $C_2(\theta) = -\frac{1}{2\sigma^2}$ 线性无关.

  $T_1(x) = x$, $T_2(x) = x^2$ 线性无关.
- $T_1(\vec{X}) = \sum_{i=1}^n X_i = n\bar{X}$, 即$\frac{1}{n}T_1(X) = \bar{X}$.

$$T_2(\vec{X}) = \sum_{i=1}^n X_i^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n\bar{X}^2 = (n-1)S^2 + n\bar{X}^2.$$

即$\frac{1}{n}T_2(X) =$ 样本方差 $+$ 样本均值$^2$ $= \frac{n-1}{n}S^2 + \bar{X}^2$.

$\phi(T_1(\vec{X}), T_2(\vec{X})) = \varphi(\bar{X}, S^2)$,

(注: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$)

- 若$\varphi(\bar{X}, S^2)$ 是$\varphi(\mu, \sigma^2)$ 无偏估计, 则是UMVUE.

  特别地, $\bar{X}$, $S^2$ 分别是$\mu$, $\sigma^2$ 的UMVUE.