# 数理统计 · 第二次习题课

2023 年 4 月 15 日

1. 数据压缩原则之充分性原则
   - 充分统计量
   - 辅助统计量
   - 完全统计量
2. 指数型分布族与群族
   - 指数型分布族的定义与性质
   - 指数型分布族中的充分性与完全性
   - 群族
3. 点估计
   - 点估计及其评价
   - 矩估计
   - 极大似然估计
   - 最小差异估计与估计方程

## 充分性原则概览

**充分性原则**

如果 $T(\boldsymbol{X})$ 是参数 $\theta$ 的充分统计量，任何基于样本 $\boldsymbol{X}$ 的对 $\theta$ 进行的推断应该从 $T(\boldsymbol{X})$ 出发. 也就是说，如果 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是两个满足 $T(\boldsymbol{x})=T(\boldsymbol{y})$ 的样本点，那么对 $\theta$ 的推断无论观测到 $\boldsymbol{X}=\boldsymbol{x}$ 或 $\boldsymbol{X}=\boldsymbol{y}$ 都应该是一样的.

- **充分统计量:** 捕获了样本中有关 $\theta$ 的所有信息;
- **辅助统计量:** 没有任何有关 $\theta$ 的信息;
- **完全统计量:** 没有与 $\theta$ 无关的冗余信息;
- **最小充分统计量:** 在保留有关 $\theta$ 的所有信息的同时又进行了最大程度的样本数据的缩减.

**定义 1 (充分统计量, sufficient statistic)**

设 $\boldsymbol{X} = (X_1, \ldots, X_n)$ 是来自未知总体 $P \in \mathcal{P}$ 的一个样本, 其中 $\mathcal{P}$ 是一族总体. 统计量 $T(\boldsymbol{X})$ 对于 $P \in \mathcal{P}$ (或者对 $\theta \in \Theta$, 当 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ 是参数族时) 称为**充分的**, 当且仅当给定 $T$ 时 $\boldsymbol{X}$ 的条件分布已知 (不依赖于 $P$ 或 $\theta$).

- 充分统计量 $T(\boldsymbol{X})$ 包含了 $\boldsymbol{X}$ 中关于 $P$ 的所有信息, 且当 $T$ 不是一对一映射时简化了数据.
- 充分统计量不是唯一的: 设 $T$ 为充分统计量, 可测函数 $g$ 为一一映射 (不依赖于 $P$), 则 $g(T)$ 也是一个充分统计量.
- 次序统计量 $T(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$ 是一个充分统计量.

**定理 1 (因子分解定理)**

假设 $\boldsymbol{X} = (X_1, \ldots, X_n)$ 是来自未知总体 $P \in \mathcal{P}$ 的一个样本，其中 $\mathcal{P}$ 是一族总体. 那么 $T(\boldsymbol{X})$ 是 $P \in \mathcal{P}$ 的充分统计量，当且仅当存在非负 Borel 函数 $h$ (不依赖于 $P$) 和定义在 $T$ 的值域上的 $g_P$ (依赖于 $P$), 使得

$$f_P(\boldsymbol{x}) = g_P(T(\boldsymbol{x})) \cdot h(\boldsymbol{x}).$$

特别地，当 $\mathcal{P}$ 是 $\theta$ 参数族时，也可以写为

$$f(\boldsymbol{x}; \theta) = g(T(\boldsymbol{x}); \theta) \cdot h(\boldsymbol{x}).$$

证明或证伪充分统计量的两大思路:

- 按定义证明 $f(\boldsymbol{x} \mid T(\boldsymbol{x}) = t; \theta)$ 与 $\theta$ 无关;
- 按因子分解定理 $f(\boldsymbol{x}; \theta) = g(T(\boldsymbol{x}); \theta) \cdot h(\boldsymbol{x})$ 进行拆分.

**定义 2 (最小充分统计量, minimal sufficient statistic)**

令 $T$ 是 $P \in \mathcal{P}$ 的充分统计量. $T$ 称为**最小充分统计量**, 当且仅当对任意其它的对 $P \in \mathcal{P}$ 的充分统计量 $S$, 存在一个可测函数 $\psi$, 使得

$$T = \psi(S) \quad \text{a.e. } \mathcal{P}.$$

**定理 2**

假设总体 $P$ 有密度函数 $f_P(x)$, 且存在统计量 $T(\boldsymbol{X})$ 使得对 $\boldsymbol{X}$ 的任意可能值 $\boldsymbol{x}, \boldsymbol{y}$,

$$f_P(\boldsymbol{x}) = f_P(\boldsymbol{y}) C(\boldsymbol{x}, \boldsymbol{y}),\ \forall P \in \mathcal{P} \iff T(\boldsymbol{x}) = T(\boldsymbol{y}),$$

其中 $C(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}$ 与 $P$ 无关, 那么 $T(\boldsymbol{X})$ 是 $P \in \mathcal{P}$ 的最小充分统计量.

**定义 3 (辅助统计量, ancillary statistic)**

一个统计量 $V(\boldsymbol{X})$ 称为**辅助的**, 若它的分布不依赖于总体 $P$; 称为**一阶辅助的**, 若 $E[V(\boldsymbol{X})]$ 不依赖于总体 $P$.

- 一个平凡的辅助统计量是常数统计量 $V(\boldsymbol{X}) \equiv c \in \mathbb{R}$.
- 若 $V(\boldsymbol{X})$ 是一个非平凡辅助统计量, 则它不包含任何有关 $P$ 的信息.
- 如果 $S(\boldsymbol{X})$ 是一个统计量且 $V(S(\boldsymbol{X}))$ 是一个非平凡的辅助统计量, 这意味着 $S(\boldsymbol{X})$ 中有不包含任何有关 $P$ 信息的部分, 这就意味着 $S(\boldsymbol{X})$ 可进一步被简化.
- 对充分统计量 $T(\boldsymbol{X})$ 而言, 如果不存在 $T$ 的非常值函数是辅助的或一阶辅助的, 则 $T(\boldsymbol{X})$ 最成功地压缩了数据.

**定义 4 (完全统计量, complete statistic)**

统计量 $T(X)$ 称为对 $P \in \mathcal{P}$ 是**完全的**, 当且仅当对任一 Borel 函数 $h$,

$$E[h(T)] = 0,\ \forall P \in \mathcal{P} \implies h(T) \equiv 0 \text{ a.e. } \mathcal{P}.$$

若该命题对任一有界 Borel 函数 $h$ 成立, 则称 $T$ 是**有界完全的**.

- 完全统计量一定是有界完全的.
- 若 $T$ 是 (有界) 完全的且对可测函数 $\psi$ 有 $S = \psi(T)$, 则 $S$ 是 (有界) 完全的.
- (Lehman and Scheffé) 若最小充分统计量存在, 那么一个充分且完全的统计量一定是最小充分统计量, 反之不然.

**定理 3 (Basu)**

设 $V$ 和 $T$ 是两个来自总体 $P \in \mathcal{P}$ 的 $\boldsymbol{X}$ 的统计量. 如果 $V$ 是辅助的, $T$ 关于 $P \in \mathcal{P}$ 是有界完全且充分的, 则关于任意 $P \in \mathcal{P}$, $V$ 与 $T$ 独立.

**Proof.**

令 $C$ 是 $V$ 值域上的一个事件.

- 因为 $V$ 是辅助的, 所以 $p_C = \mathrm{P}(V \in C)$ 是一个常数, 与 $P$ 无关.
- 因为 $T$ 是充分的, 所以 $q_C(T) = \mathrm{P}(V \in C \mid T)$ 是 $T$ 的函数 (不依赖于 $P$).

注意到 $\mathrm{E}[q_C(T)] = p_C$, 于是由有界完全性

$$q_C(T) = \mathrm{P}(V \in C \mid T) \equiv p_C \quad \text{a.e. } \mathcal{P}.$$

这就证明了 $T$ 和 $V$ 的独立性. $\square$

**例 1 (2.48)**

设 $X_1, \ldots, X_n$ 是从下列总体中抽取的样本, 其密度函数为

$$f_\theta(x) = \frac{1}{2\theta} \exp\left\{-\frac{|x|}{\theta}\right\}, \quad -\infty < x < +\infty, \ \theta > 0,$$

证明: $T = \sum_{i=1}^{n} |X_i|$ 是 $\theta$ 的充分完全统计量.

- 充分性:

- 完全性:

**例 2 (2.49)**

设 $X_1, \ldots, X_n$ 为取自下列指数分布总体的简单样本,

$$f_\theta(x) = \exp\{-(x-\theta)\}, \quad x > \theta,$$

其中 $-\infty < \theta < +\infty$ 为位置参数. 证明: $T(\boldsymbol{X}) = X_{(1)}$ 为充分完全统计量.

- 充分性:

- 完全性:

**例 3 (2.46)**

设 $X_1, \ldots, X_n$ i.i.d. $\sim N(\theta, \theta^2)$, $\theta > 0$, 问 $\overline{X}$ 是否为充分统计量?

- 充分性:

- 最小充分性:

- 完全性:

**例 4 (2.50)**

设 $\boldsymbol{X}=(X_1,\ldots,X_n)$ 为从均匀分布 $U(-\theta/2,\theta/2)$, $0<\theta<\infty$ 中抽取的简单样本, 证明: $(X_{(1)},X_{(n)})$ 为充分统计量, 但不是完全的.

- 充分性:

- 完全性:

**例 5 (补充作业)**

在均匀分布 $U(0,\theta)$ (其中 $\theta > 1$) 下, 统计量 $T = X_{(n)}$ 不是 $\theta$ 的完全统计量.

**例 6 (2.53)**

设 $X_1, \ldots, X_n$ i.i.d. $\sim N(a, \sigma^2)$, 证明: $\overline{X}$ 和 $X_{(n)} - X_{(1)}$ 独立.

**例 7 (补充题)**

设 $X_1, \ldots, X_n$ 是 i.i.d. 随机变量, 当 $\theta = 0$ 时服从正态分布 $N(\theta, 2)$, 当 $\theta \in \mathbb{R}$ 且 $\theta \neq 0$ 时服从正态分布 $N(\theta, 1)$. 证明: 样本均值 $\overline{X}$ 是 $\theta$ 的完全统计量, 但不是 $\theta$ 的充分统计量.

- **完全性:** 当 $\theta \neq 0$ 时, $\overline{X} \sim N(\theta, 1/n)$, 于是对任一 Borel 函数 $h$, 有

$$\mathrm{E}_\theta[g(\overline{X})] = \int_{-\infty}^{+\infty} g(t)\sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n(t-\theta)^2}{2}\right\} \mathrm{d}t = 0.$$

从而有

$$\int_{-\infty}^{+\infty} g(t) \exp\left\{-\frac{nt^2}{2}\right\} \cdot \exp\{nt\theta\} \, \mathrm{d}t = 0.$$

由 Fourier 变换的逆变换公式可知

$$g(t) \exp\left\{-\frac{nt^2}{2}\right\} = 0 \text{ a.e. } \mathrm{P}_\theta.$$

- 完全性: 所以

$$g(t) = 0 \text{ a.e. } \mathrm{P}_\theta, \quad \theta \in \mathbb{R} \text{ 且 } \theta \neq 0.$$

而 $m\{\theta = 0\} = 0$, 因此 $g(t) = 0$ a.e. $\mathrm{P}_\theta$ 对任意 $\theta \in \mathbb{R}$ 成立.

- 充分性: 作正交变换 $(Y_1, \ldots, Y_n)^\mathsf{T} = \boldsymbol{A}(X_1, \ldots, X_n)^\mathsf{T}$, 其中 $\boldsymbol{A}$ 的第一行元素全为 $1/\sqrt{n}$, 且其为正交矩阵. 于是
  - 当 $\theta = 0$ 时, $Y_1 = \sqrt{n}\overline{X} \sim N(\sqrt{n}\theta, 2)$, $Y_2, \ldots, Y_n \sim N(0, 2)$,
  - 当 $\theta \neq 0$ 时, $Y_1 = \sqrt{n}\overline{X} \sim N(\sqrt{n}\theta, 1)$, $Y_2, \ldots, Y_n \sim N(0, 1)$,

  且 $Y_1, \ldots, Y_n$ 独立. 给定 $Y_1 = y_1$ 时, $Y_1, \ldots, Y_n$ 的条件联合 p.d.f. 为

$$\begin{aligned} f_\theta(\boldsymbol{y} \mid y_1) = (4\pi)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{4}\sum_{i=2}^n y_i^2\right\} I\{\theta = 0\} \\ + (2\pi)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=2}^n y_i^2\right\} I\{\theta \neq 0\}. \end{aligned}$$

  这仍然与 $\theta$ 有关, 从而 $\overline{X}$ 不是充分统计量.

**定义 5 (指数型分布族, exponential family of distributions)**

*设有参数分布族 $\{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ 有密度 (p.d.f. 或 p.m.f.)$f(x;\boldsymbol{\theta})$. 如果 $f(x;\boldsymbol{\theta})$ 可表示成如下形式*

$$f(x;\boldsymbol{\theta}) = \exp\left\{\sum_{i=1}^{k} Q_i(\boldsymbol{\theta})\, T_i(x) - A(\boldsymbol{\theta})\right\} h(x),$$

*则称此分布族为**指数型分布族**. 其中 $A(\boldsymbol{\theta})$, $Q_i(\boldsymbol{\theta})$ $(i=1,\cdots,k)$ 为定义在参数空间 $\Theta$ 上的函数, 与 $x$ 无关, $h(x)>0$, $T_i(x)$ $(i=1,\cdots,k)$ 为与 $\boldsymbol{\theta}$ 无关的函数. 特别地, 当 $k=1$ 时, 此分布族称为**单参数指数型分布族**.*

- 若总体分布族是指数型分布族, 则样本的分布族也是指数型分布族;
- 指数型分布族分布的支撑集与参数 $\boldsymbol{\theta}$ 无关, 即有公共支撑;
- 指数型分布族具有良好的解析性质.

现在我们重参数化 $\eta_i = Q_i(\boldsymbol{\theta})$, $i = 1, \ldots, k$, 而把 $A(\boldsymbol{\theta})$ 表示成 $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^{\mathsf{T}}$ 的函数 $A^*(\boldsymbol{\eta})$, 于是我们得到指数型分布族的**自然形式**

$$f(x; \boldsymbol{\eta}) = \exp\left\{\sum_{i=1}^{k} \eta_i T_i(x) - A^*(\boldsymbol{\eta})\right\} h(x).$$

**自然参数空间**为

$$\Theta^* = \left\{\boldsymbol{\eta} : \int_{\mathcal{X}} h(x) \exp\left\{\sum_{i=1}^{k} \eta_i T_i(x)\right\} \mathrm{d}x < \infty\right\},$$

它是 $\mathbb{R}^k$ 上的凸集. 如果 $\dim(\Theta^*) = k$, 则称它是满秩的指数型分布族.

## 定理 4

设 $X$ 来自指数型分布族，其自然形式为

$$f(x;\boldsymbol{\eta}) = \exp\left\{\sum_{i=1}^{k}\eta_i T_i(x) - A^*(\boldsymbol{\eta})\right\}h(x).$$

设 $\boldsymbol{\eta}$ 为自然参数空间 $\Theta^*$ 的一个内点，则

- 如果 $\boldsymbol{s}$ 满足 $\boldsymbol{\eta}+\boldsymbol{s}$ 为 $\Theta^*$ 的一个内点，则 $(T_1(X),\ldots,T_k(X))^T$ 的联合矩母函数存在且等于

$$M_{\boldsymbol{\eta}}(\boldsymbol{s}) = \exp\{A(\boldsymbol{\eta}+\boldsymbol{s}) - A(\boldsymbol{\eta})\}.$$

- $\exp\{A(\boldsymbol{\eta})\}$ 对分量 $\eta_i$ 任意阶偏导存在，且可在积分号下求得.
- $E_{\boldsymbol{\eta}}[T_i(X)] = \dfrac{\partial}{\partial \eta_i}A(\boldsymbol{\eta})$, $Cov_{\boldsymbol{\eta}}(T_i(X), T_j(X)) = \dfrac{\partial^2}{\partial \eta_i \partial \eta_j}A(\boldsymbol{\eta})$, $1 \le i,j \le k$.

## 定理 5

设 $X$ 来自指数型分布族，其自然参数空间有内点，记内点集为 $\Theta_0$. 设 $g(x)$ 是任一实函数，使得积分

$$G(\boldsymbol{\eta}) = \int_{\mathcal{X}} g(x) \exp\left\{\sum_{i=1}^{k} \eta_i T_i(x)\right\} h(x) dx$$

在 $\Theta_0$ 内存在且有限，则 $G(\boldsymbol{\eta})$ 的任意阶偏导在 $\Theta_0$ 内存在，且可在积分号下求得，即

$$\frac{\partial^m G(\boldsymbol{\eta})}{\partial \eta_1^{m_1} \cdots \partial \eta_k^{m_k}} = \int_{\mathcal{X}} \frac{\partial^m}{\partial \eta_1^{m_1} \cdots \partial \eta_k^{m_k}} \left[ g(x) \exp\left\{\sum_{i=1}^{k} \eta_i T_i(x)\right\} h(x) \right] dx,$$

其中 $m_1 + \cdots + m_k = m$.

**定理 6**

设样本 $\boldsymbol{X}$ 来自满秩的指数型分布族，则 $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X}))$ 是参数 $\theta$ 的最小充分统计量.

**定理 7**

设样本 $\boldsymbol{X}$ 来自满秩的指数型分布族，$T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X}))$ 是参数 $\theta$ 的最小充分统计量，且自然参数空间具有内点 (有 $\mathbb{R}^k$ 上的开子集)，则 $T(\boldsymbol{X})$ 是完全统计量.

**定义 6 (位置-尺度分布族)**

设 $f(x)$ 为密度函数, 于是对任意 $\mu \in \mathbb{R}, \sigma > 0$, 分布族

$$\left\{ \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma > 0 \right\}$$

称为具有标准密度 $f(x)$ 的参数为 $(\mu, \sigma)$ 的***位置-尺度分布族***, $\mu$ 称为***位置参数***, $\sigma$ 称为***尺度参数***.

- 设 $Z \sim f(z)$, $X \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, 则

$$X \stackrel{d}{=} \sigma Z + \mu.$$

特别地, 有 $\mathrm{E}(X) = \sigma \mathrm{E}(Z) + \mu$, $\mathrm{Var}(X) = \sigma^2 \mathrm{Var}(Z)$.

- 对于确定的 $\sigma = 1$ 时为位置分布族, $\mu = 0$ 时为尺度分布族.

## 例 8 (补充作业)

证明: 柯西分布 $C(\mu, \lambda)$ 不属于指数型分布族, 其密度函数为

$$f(x \mid \mu, \lambda) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x-\mu)^2}, \quad x \in \mathbb{R}.$$

## 例 9 (2.47)

设 $X_1,\ldots,X_m$ i.i.d. $\sim N(a,\sigma^2)$, $Y_1,\ldots,Y_n$ i.i.d. $\sim N(b,\sigma^2)$ 且两组样本独立. 记 $\overline{X}=\sum_{i=1}^{m} X_i/m$, $\overline{Y}=\sum_{j=1}^{n} Y_i/n$, 而

$$S^2=\frac{1}{n+m-2}\left[\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right].$$

证明: $(\overline{X},\overline{Y},S^2)$ 为充分完全统计量.

- 充分性:
- 完全性:

**定义 7 (点估计, point estimator)**

设 $\boldsymbol{X}=(X_1,\dots,X_n)$ 为来自总体 $F_{\boldsymbol{\theta}}$, $\boldsymbol{\theta}\in\Theta\subset\mathbb{R}^d$ 的样本, 统计量 $\widehat{g}_n(\boldsymbol{X})$ 称为对感兴趣参数 $g(\boldsymbol{\theta})$ 的**点估计**.

- **无偏性:** $\mathrm{E}_{\boldsymbol{\theta}}[\widehat{g}_n(\boldsymbol{X})]=g(\boldsymbol{\theta})$, $\forall\boldsymbol{\theta}\in\Theta$.
- **渐近无偏性:** $\lim\limits_{n\to\infty}\mathrm{E}_{\boldsymbol{\theta}}[\widehat{g}_n(\boldsymbol{X})]=g(\boldsymbol{\theta})$, $\forall\boldsymbol{\theta}\in\Theta$.
- **有效性:** $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{g}_n(\boldsymbol{X}))\geq\mathrm{Var}_{\boldsymbol{\theta}}(\widetilde{g}_n(\boldsymbol{X}))$, $\forall\boldsymbol{\theta}\in\Theta$ 且严格不等号至少对一个 $\theta$ 成立, 这里 $\widehat{g}_n(\boldsymbol{X})$ 和 $\widetilde{g}_n(\boldsymbol{X})$ 都是 $\boldsymbol{\theta}$ 的无偏估计.
- **均方误差准则:** $\mathrm{Mse}_{\boldsymbol{\theta}}(\widehat{g}_n(\boldsymbol{X}))\geq\mathrm{Mse}_{\boldsymbol{\theta}}(\widetilde{g}_n(\boldsymbol{X}))$, $\forall\boldsymbol{\theta}\in\Theta$ 且严格不等号至少对一个 $\theta$ 成立.
- **相合性:** $\widehat{g}_n(\boldsymbol{X})\xrightarrow{\text{rule}}g(\boldsymbol{\theta})$ as $n\to\infty$.
- **渐近正态性:** $[B_n(\boldsymbol{\theta})]^{-1}[\widehat{g}_n(\boldsymbol{X})-g(\boldsymbol{\theta})]\xrightarrow{\mathcal{L}}N(\boldsymbol{0},\boldsymbol{I})$ as $n\to\infty$.

**定义 8 (矩估计)**

设 $g(\boldsymbol{\theta}) = G(\alpha_1, \ldots, \alpha_k; \mu_2, \ldots, \mu_s)$ 表达了感兴趣参数 $g(\boldsymbol{\theta})$ 与总体原点矩和中心矩之间的关系，那么 $g(\boldsymbol{\theta})$ 的***矩估计***定义为

$$\widehat{g}_n(\boldsymbol{X}) = G(a_{n1}, \ldots, a_{nk}; m_{n2}, \ldots, m_{ns}),$$

其中 $a_{n1}, \ldots, a_{nk}$ 是样本原点矩，$m_{n2}, \ldots, m_{ns}$ 是样本中心矩.

- 矩估计一般不具有无偏性 (除原点矩的线性组合)，但具有渐近无偏性 ($G$ 满足适当条件).
- 矩估计具有相合性 ($G$ 是连续函数).
- 矩估计具有渐近正态性 ($G$ 是连续可微函数).

**例 10 (3.5)**

设总体 $X$ 服从均匀分布 $U(0,2\theta)$, 令 $X_1,\ldots,X_n$ 为从总体 $X$ 中抽取的简单样本,

(1) 证明: $\widehat{\theta}_1=\overline{X}$ 和 $\widehat{\theta}_2=(n+1)X_{(n)}/(2n)$ 为 $\theta$ 的无偏估计.

(2) 证明: $\widehat{\theta}_1$ 为 $\theta$ 的强相合估计, $\theta_2^*=X_{(n)}/2$ 为 $\theta$ 的弱相合估计.

(3) 求 $\widehat{\theta}_1$ 和 $\widehat{\theta}_2$ 的方差, 问哪一个更有效?

**例 11 (3.6)**

设总体 $X$ 服从二项分布 $B(k, p)$, $k$ 是正整数, $0 < p < 1$, 两者都是未知参数, $X_1, \ldots, X_n$ 为从中抽取的简单样本, 求 $k$ 和 $p$ 的矩估计.

**例 12 (3.9)**

设 $X_1, \ldots, X_n$ 是来自正态总体 $N(a, \sigma^2)$ 的简单样本, 求 $\mathrm{P}(X > 1)$ 的矩估计量.

**例 13 (3.10)**

设 $X_1, \ldots, X_n$ i.i.d. $\sim$ 伽马分布 $\Gamma(r, \lambda)$, 其中 $r$ 已知, 求 $\lambda$ 的矩估计量, 并讨论它的无偏性.

## 例 14 (补充题)

一个盒子里有未知的奇数个球, 依次标为 $1, 2, \ldots, (2\theta+1)$, 其中 $\theta$ 是一个未知的非负整数. $X_1, \ldots, X_n$ 为一不放回抽取到的简单随机样本, $X_i$ 是第 $i$ 个抽出的球的标号且 $n < 2\theta+1$.

(1) 试给出 $\theta$ 的一个无偏估计.

(2) 若球标号为 $-\theta, -(\theta-1), \ldots, -2, -1, 0, 1, 2, \ldots, (\theta-1), \theta$, 试给出 $\theta$ 的一个无偏估计.

(1) 注意到 $\mathrm{E}(X_1) = \theta + 1$, 因此 $\theta$ 的一个无偏估计是 $\widehat{\theta} = X_1 - 1$.

(2) 此时 $\mathrm{E}(X_1) = 0$, 上述方法失效. **怎么办?** 考虑

$$\mathrm{E}|X_1| = \sum_{i=1}^{\theta} \frac{2i}{2\theta+1} = \frac{1}{2}\theta + \frac{\theta}{2(2\theta+1)},$$

而且 $\mathrm{P}(|X_1| \neq 0) = \dfrac{2\theta}{2\theta+1}$, 因此 $\mathrm{E}\left(2|X_1| - \dfrac{1}{2} I\{X_1 \neq 0\}\right) = \theta$.

(2) 所以 $\theta$ 的一个无偏估计是

$$\widehat{\theta} = 2|X_1| - \frac{1}{2} I\{X_1 \neq 0\}.$$

**另一种想法:** 基于次序统计量 $X_{(i)}$.

(1) 注意到

$$\mathsf{P}(X_{(i)} = x) = \frac{\binom{x-1}{i-1}\binom{2\theta+1-x}{n-i}}{\binom{2\theta+1}{n}}, \quad x = i, i+1, \ldots, i+2\theta+1-n.$$

所以

$$\mathsf{E}(X_{(i)}) = \sum_{x=i}^{i+2\theta+1-n} \frac{x\binom{x-1}{i-1}\binom{2\theta+1-x}{n-i}}{\binom{2\theta+1}{n}} = i \cdot \frac{2\theta+2}{n+1}.$$

从而 $\theta$ 的一个无偏估计是 $\dfrac{n+1}{2i} X_{(i)} - 1$.

(2) 类似地,

$$\mathrm{P}(X_{(i)}=x)=\frac{\binom{x+\theta}{i-1}\binom{\theta-x}{n-i}}{\binom{2\theta+1}{n}},\quad x=-\theta+i-1,\ldots,\theta-n+i.$$

考虑计算 $X_{(i)}+\theta+1$ 的期望 (为什么?) 为

$$\mathrm{E}(X_{(i)}+\theta+1)=i\cdot\frac{2\theta+2}{n+1}.$$

从而 $\theta$ 的一个无偏估计是 $\dfrac{n+1}{2i-n-1}X_{(i)}-1.$

**定义 9 (极大似然估计, MLE)**

设 $f(\boldsymbol{x};\theta)$ 是样本 $\boldsymbol{X}$ 的概率函数. 当观测到 $\boldsymbol{X}=\boldsymbol{x}$ 时, 记

$$L_n(\theta)=f(\boldsymbol{x};\theta),\quad \theta\in\Theta.$$

为 $\theta$ 的似然函数. 对每个样本观测值 $\boldsymbol{x}$, 令

$$\widehat{\theta}(\boldsymbol{x})=\arg\max_{\theta\in\Theta} L_n(\theta).$$

称 $\widehat{\theta}(\boldsymbol{X})$ 为参数 $\theta$ 的极大似然估计.

- 具有显式解: (1) 单调似然函数; (2) 基于对数似然函数求极值.
- 没有显式解: (1) 梯度方法; (2) EM 算法.

## 例 15 (3.12)

设 $X_1,\ldots,X_n$ 是从总体 $X$ 中抽取的简单样本, $X$ 的分布为下列之一, 试分别用矩估计法和极大似然估计法求 $\theta$ 的估计量.

(1) $f_\theta(x)=\begin{cases}\theta(\theta+1)x^{\theta-1}(1-x), & 0<x<1,\\ 0, & \text{其他},\end{cases}\quad \theta>0.$

(2) $f_\theta(x)=(\theta+1)x^\theta,\ 0<x<1,\ \theta>-1.$

(3) $f_\theta(x)=\begin{cases}\dfrac{\theta^{-r}}{\Gamma(r)}x^{r-1}\exp\left\{-\dfrac{x}{\theta}\right\}, & x>0,\\ 0, & \text{其他},\end{cases}\quad r>0\ \text{已知},\ \theta>0.$

### 例 16 (3.13)

设总体 $X$ 的密度为

$$f(x;a,\sigma) = \frac{1}{2\sigma}\exp\{-|x-a|/\sigma\},$$

其中 $\sigma > 0$ 和 $-\infty < a < \infty$ 为未知参数. 设 $X_1,\ldots,X_n$ 为抽自此总体的简单随机样本, 求 $a$ 和 $\sigma$ 的矩估计和极大似然估计.

## EM 算法

基于观测数据 $\boldsymbol{X}$ 的对数似然函数 $\ell(\theta) = \log f(\boldsymbol{x} \mid \theta)$ 在计算上不容易，但如果问题中存在“隐变量” $\boldsymbol{Y}$，基于“完全观测数据”下的对数似然函数 $\ell_c(\theta) = \log f(\boldsymbol{x}, \boldsymbol{y} \mid \theta)$ 在计算上可行.

- E 步：计算

$$Q(\theta \mid \theta^{(t)}) = \mathrm{E}_{f(\boldsymbol{y} \mid \boldsymbol{x}, \theta^{(t)})}[\log f(\boldsymbol{x}, \boldsymbol{y} \mid \theta)].$$

- M 步：最大化

$$\theta^{(t+1)} = \arg\max_{\theta \in \Theta} Q(\theta \mid \theta^{(t)}).$$

EM 算法保证了算法每次迭代后基于观测数据 $\boldsymbol{X}$ 的对数似然函数 $\ell(\theta)$ 都是递增的，因此可以收敛到局部最优值，算法效率依赖于初值的选取.

## 极大似然估计的性质

**定理 8 (变换不变性)**

如果 $\widehat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的 MLE, 则对于定义在 $\Theta$ 上的函数 $h$, $h(\widehat{\boldsymbol{\theta}})$ 是 $h(\boldsymbol{\theta})$ 的 MLE.

**定理 9 (MLE 是充分统计量的函数)**

设 $T(\boldsymbol{X})$ 是 $\theta$ 的充分统计量. 如果 $\boldsymbol{\theta}$ 的 MLE 存在, 则它必为 $T$ 的函数.

**定理 10 (存在性与唯一性)**

如果基于样本 $\boldsymbol{X}$ 的对数似然函数

$$\ell_n(\boldsymbol{\theta}) = \log f(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i; \boldsymbol{\theta})$$

是严格凹的, 并且 $\lim\limits_{\boldsymbol{\theta} \to \partial\Theta} \ell_n(\boldsymbol{\theta}) = -\infty$, 则 MLE $\widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ 存在且唯一.

### 定理 11 (指数族下 MLE 的唯一性)

*设样本 $\boldsymbol{X}$ 来自满秩的自然参数指数族，具有密度函数*

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \exp\left\{\sum_{i=1}^{k}\theta_i T_i(\boldsymbol{x}) - A(\boldsymbol{\theta})\right\} h(x), \quad \boldsymbol{x} \in \mathcal{X}.$$

*如果*

$$\frac{\partial A(\boldsymbol{\theta})}{\partial \theta_i} = \frac{1}{n}\sum_{j=1}^{n} T_i(X_i), \quad i = 1, \ldots, k$$

*在自然参数空间的内点集上有解，则该解必为 $\boldsymbol{\theta}$ 唯一的 MLE.*

### 例 17 (3.19)

设 $X_1, \ldots, X_m$ 和 $Y_1, \ldots, Y_n$ 分别来自总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 的两组独立样本，求 $\mu_1$, $\mu_2$ 和 $\sigma^2$ 的 MLE.

**定理 12 (相合性)**

设样本 $\boldsymbol{X}$ 来自密度函数为 $f(x;\boldsymbol{\theta})$ 的总体, 参数空间 $\Theta$ 是 $\mathbb{R}^d$ 上的开集. 假设

(1) 参数可识别: 对 $\forall\boldsymbol{\theta}'\neq\boldsymbol{\theta}$, 存在 $x$ 使得 $f(x,\boldsymbol{\theta}')\neq f(x,\boldsymbol{\theta})$,

(2) 公共支撑且可微: $f(\cdot;\boldsymbol{\theta})$ 的支撑不依赖于 $\boldsymbol{\theta}$, 且 $f(x;\boldsymbol{\theta})$ 对 $\boldsymbol{\theta}\in\Theta$ 可微,

(3) 真实参数是参数空间的内点: $\boldsymbol{\theta}_0\in\Theta^0$,

则似然函数存在最大值点 $\widehat{\boldsymbol{\theta}}$, 且 $\widehat{\boldsymbol{\theta}}\xrightarrow{P}\boldsymbol{\theta}_0$.

**例 18 (3.23)**

设 $X_1,\ldots,X_n$ i.i.d. $\sim$ 均匀分布 $U(0,\theta)$, 求 $\theta$ 的 MLE $\widehat{\theta}_n$, 证明: $\widehat{\theta}_n$ 是 $\theta$ 的强相合估计和 $r\,(r>0)$ 阶矩相合估计.

**例 19 (3.24)**

设 $X_1, \ldots, X_n$ i.i.d. $\sim N(\sigma, \sigma^2)$, $\sigma > 0$, 试求 $\sigma^2$ 的 MLE $\widehat{\sigma}^2$, 并证明 $\widehat{\sigma}$ 为 $\sigma$ 的弱相合估计.

**例 20 (3.25)**

设总体 $X$ 为下列 0-1 分布族

$$\mathrm{P}_\theta(X=1) = 1 - \mathrm{P}_\theta(X=0) = \begin{cases} \theta, & \theta \text{ 为有理数}, \\ 1-\theta, & \theta \text{ 为无理数}. \end{cases}$$

设 $X_1, \ldots, X_n$ 是来自总体 $X$ 的简单随机样本, 证明: $\theta$ 的 MLE 不是相合估计.

**定理 13 (渐近正态性)**

设 $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta\}$ 是一个概率函数族, $\Theta$ 为直线上的开区间, $X_1, \ldots, X_n$ 是从 $\mathcal{F}$ 中某个总体 $X \sim f(x;\theta_0)$ 产生的样本. 当 $f(x;\theta)$ 满足适当正则条件时, 在参数 $\theta$ 的未知真值 $\theta_0$ 为 $\Theta$ 的一个内点的情况下, 似然方程有一个解 $\widehat{\theta}_n$ 满足

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I(\theta_0)}\right), \quad as\ n \to \infty,$$

其中

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f(X;\theta)}{\partial \theta}\right)^2\right]$$

是 Fisher 信息量.

## 正则条件:

(1) 对一切 $\theta \in \Theta$, 偏导数

$$\frac{\partial \log f(x;\theta)}{\partial \theta}, \quad \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}, \quad \frac{\partial^3 \log f(x;\theta)}{\partial \theta^3}$$

存在.

(2) 对一切 $\theta^* \in \Theta$, 存在 $\theta^*$ 的一个邻域 $U_{\theta^*}$ 和可积函数 $F_1(x)$, $F_2(x)$, 以及函数 $H(x)$, 使得

$$\left|\frac{\partial f(x;\theta)}{\partial \theta}\right| \le F_1(x), \quad \left|\frac{\partial^2 f(x;\theta)}{\partial \theta^2}\right| \le F_2(x), \quad \left|\frac{\partial^3 \log f(x;\theta)}{\partial \theta^3}\right| \le H(x),$$

对 $\theta \in U_{\theta^*}$ 成立, 其中 $H(x)$ 满足 $\mathrm{E}_\theta[H(X)] \le M$, $\forall \theta \in U_{\theta^*}$.

(3) 对一切 $\theta \in \Theta$, 有

$$0 < I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial \log f(X;\theta)}{\partial \theta}\right)^2\right] < \infty.$$

### 推论 1

设总体分布族是单参数指数族, 其中密度函数具有形式

$$f(x;\theta) = C(\theta)\exp\{Q(\theta)T(x)\}h(x),$$

其中 $C(\theta)$ 和 $Q(\theta)$ 有一至三阶连续的导数, 则定理的结论成立.

**例 21 (3.26)**

设 $X_1,\ldots,X_n$ 是来自伽马分布 $\Gamma(\alpha,\lambda)$ 的简单样本, 试证明:
(1) 当 $\lambda=1$ 时, 记 $\alpha$ 的 MLE 为 $\widehat{\alpha}$, 则

$$\sqrt{n}(\widehat{\alpha}-\alpha)\xrightarrow{\mathscr{L}} N\left(0,\left[\mathrm{E}\left(\frac{\partial\log g(x,\alpha)}{\partial\alpha}\right)^2\right]^{-1}\right),$$

此处 $g(x,\alpha)$ 是参数分别为 $\alpha$, $\lambda=1$ 的伽马分布的密度函数;
(2) 当 $\alpha$ 已知时, $\lambda$ 的 MLE $\widehat{\lambda}$ 近似服从 $N(\lambda,\lambda^2/(n\alpha))$, 即

$$\sqrt{n}(\widehat{\lambda}-\lambda)\xrightarrow{\mathscr{L}} N(0,\lambda^2/\alpha).$$

## 最小差异估计

考虑定义差异函数 $\rho : \mathcal{X} \times \Theta \to \mathbb{R}$, 以及 $D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ 来衡量 $\boldsymbol{\theta}$ 和真值 $\boldsymbol{\theta}_0$ 之间的差距:

$$D(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}_0}[\rho(X, \boldsymbol{\theta})].$$

- 要求 $D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ 在 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ 处具有唯一最小值.
- 通常 $\rho(\boldsymbol{X}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \boldsymbol{\theta}) \to D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ as $n \to \infty$.
- 假设真值 $\boldsymbol{\theta}_0$ 是参数空间的内点, 并且 $D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ 是光滑的, 那么自然希望

$$\nabla_{\boldsymbol{\theta}} D(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 0.$$

- 进而我们自然有 $\boldsymbol{\theta}$ 的估计 $\widehat{\boldsymbol{\theta}}$ 满足

$$\nabla_{\boldsymbol{\theta}} \rho(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}) = 0.$$

- 基于熵, MLE 是最小差异估计.

## 估计方程

如果 $g(X,\boldsymbol{\theta})$ 是总体 $X$ 和待估参数 $\boldsymbol{\theta}$ 的函数，满足

$$S(\boldsymbol{\theta}_0,\boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}_0}[g(X,\boldsymbol{\theta})] = \int_{\mathcal{X}} g(x,\boldsymbol{\theta})\mathrm{d}F_{\boldsymbol{\theta}_0}(x) = 0.$$

那么将 $F_{\boldsymbol{\theta}_0}(x)$ 用样本经验分布函数 $F_n(x)$ 替代，就可以得到方程

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i,\boldsymbol{\theta}) = \int_{\mathcal{X}} g(x,\boldsymbol{\theta})\mathrm{d}F_n(x) = 0.$$

上述方程的解 $\widehat{\theta}_n$ 也就是 $\theta$ 的估计方程的解.

- 最小二乘估计: $\widehat{\boldsymbol{\theta}} = \arg\min\limits_{\boldsymbol{\theta}} \sum_{i=1}^{n}[y_i - g(\boldsymbol{X}_i,\beta)]^2$.
- 最小一乘估计: $\widehat{\boldsymbol{\theta}} = \arg\min\limits_{\boldsymbol{\theta}} \sum_{i=1}^{n}|y_i - g(\boldsymbol{X}_i,\beta)|$.

**例 22 (3.20)**

为了估计湖中有多少条鱼, 从中捞出 1000 条, 标上记号后放回湖中, 然后再捞出 150 条鱼, 发现其中有 10 条鱼有记号. 问湖中有多少条鱼, 才能使 150 条鱼中出现 10 条带记号的鱼的概率最大?

**例 23 (3.21)**

一个罐子中装有黑白两种球, 今有放回地抽取一个大小为 $n$ 的样本, 其中有 $k$ 个白球, 求罐中黑白球之比的极大似然估计.

# Thanks for your time and attention!