

Chap 6 参数估计

- 引入 检测某厂一大批电子元件寿命指标是否合格.
- 总体
 - 所需检测的电子元件寿命;
 - 统计分析问题中研究的对象全体的某个数字特征 X .
 - 元件寿命的分布;
 - 变量 X 的分布.
- 统计总体 一个概率分布.
- 注
 - 有限总体;
 - 无限总体.
- 统计模型
 - 一族概率分布;
 - 参数模型 (正态模型);
 - 非参数模型.
- 样本
 - $(X_1, \dots, X_n)$, 其中 X_i 来自总体 X , n 为样本容量;
 - 获取方式:
 - 试验;
 - 观察 (完全 / 不完全).
- 简单随机抽样
 - 总体个数 N 有限;
 - 无放回;
 - 容量为 n 的样本抽取出来的概率为 $p = \frac{1}{C_N^n}$.
- 随机样本
 - $X_1, \dots, X_n$ 独立同分布, 例: 有放回 / 近似有放回 ($n \ll N$).
- 统计量
 - $T(X_1, \dots, X_n)$ - 完全由样本决定, 是一种数据简化方式.
 - 样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

- 样本方差

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

● 统计推断

- 总体为因, 样本为果;
- 总体决定样本, 样本推断总体;
- 经典方法 (频率) v.s. **Bayes**方法.
- 例 $Y_i = aX_i + \varepsilon_i$.
 - 求 a . $(X_i, Y_i)(1 \leq i \leq n) \rightarrow a \approx \hat{a}$ (模型推断 参数估计)
 - 已知 $a = \hat{a}$, 观测到 Y_i , 求 X_i (变量推断)
- 例 元件寿命.
 - 假设 $X \sim \text{Exp}(\lambda)$, 求 λ (参数估计)
 - 假设合格标准 $E(X) \geq L$, 需建立可操作的检验标准: $X \geq l$, μ 未知, 求 l (假设检验)

1 矩估计

- 定义(样本矩) $X_1, \dots, X_n$ 独立同分布, 定义其 k 阶原点矩:

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{\text{LLN}} E(X^k).$$

其 k 阶中心矩:

$$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \xrightarrow{\text{LLN}} E[(X - \mu)^k].$$

- 例 $X \sim N(\mu, \sigma^2)$. 考虑其均值、方差对应的样本矩.
- 解答

$$\begin{aligned} \mu &= E(X) \approx \bar{X} = a_1; \\ \sigma^2 &= E[(X - \mu)^2] \approx \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = m_2. \end{aligned}$$

- 例 $X \sim \text{Exp}(\lambda)$. 考虑其均值、方差对应的样本矩, 并给出 λ 的矩估计.
- 解答

$$\frac{1}{\lambda} = E(X) \approx \bar{X} = a_1 \Rightarrow \lambda = \frac{1}{a_1};$$

$$\frac{1}{\lambda^2} = E[(X - E(X))^2] \approx \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = m_2 \Rightarrow \lambda = \left(\frac{1}{m_2}\right)^{\frac{1}{2}}.$$

我们通常使用**低阶矩**进行估计.

2 极大似然估计 (MLE)

- **定义(似然函数)** 假设 $X_1, \dots, X_n$ 的联合分布 $f(x_1, \dots, x_n; \theta)$ (**PMF\PDF**), 其中 θ 为参数. 对于观测 $(X_1, \dots, X_n)$ 的似然函数为

$$L(\theta) := f(X_1, \dots, X_n; \theta).$$

- **注**

- 具体观测数据通常记为 $(x_1, \dots, x_n)$, 视为 $(X_1, \dots, X_n)$ 的具体的实现值;
- 离散情形下, $L(\theta)$ 为出现 $(X_1, \dots, X_n)$ 的概率;
- 若 $X_1, \dots, X_n$ 独立同分布, 总体分布为 $f_1(x; \theta)$ (**PMF\PDF**), 那么

$$L(\theta) = f_1(X_1; \theta) f_1(X_2; \theta) \cdots f_1(X_n; \theta).$$

- **例** $X_i \sim N(\mu, \sigma^2)$ 独立同分布, 其中 μ, σ^2 未知. 考虑观测的似然函数.
- **解答**

$$L(\mu, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \mu)^2}{2\sigma^2}} \right).$$

- **定义(MLE)** 定义 θ 的 MLE 为

$$\theta^* = \arg \max_{\theta} L(\theta).$$

其中 $\theta^* = \theta^*(X_1, \dots, X_n)$.

- **解答续** 考虑方程组

$$\begin{cases} \frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} = 0; \\ \frac{\partial \log L(\mu, \sigma^2)}{\partial \sigma^2} = 0. \end{cases}$$

解得

$$\begin{cases} \mu^* = \bar{X}, \\ (\sigma^2)^* = \sum_{i=1}^n (X_i - \bar{X})^2. \end{cases}$$

经验证 $(\mu^*, (\sigma^2)^*)$ 为 (μ, σ^2) 的 MLE.

- 注 $(\sigma^2)^* = (\sigma^*)^2$ (MLE 的不变性).
- 例 随机变量 $X_i \sim U(0, \theta)$ 独立同分布, θ 未知. 求 θ 的 MLE.
- 解答

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & X_i \in (0, \theta), \theta \geq \max\{X_1, \dots, X_n\} \\ 0, & \text{其他.} \end{cases}$$

我们得到

$$\theta^* = \max\{X_1, \dots, X_n\}.$$

- 例 总体 PDF 为 $f(x; \theta) = \frac{1}{\pi(1 + (x - \theta)^2)}$, $x \in \mathbb{R}$ (Cauchy 分布).
- 解答
 - 无矩 $\Rightarrow$ 无矩估计;
 - 似然方程 $\sum_{i=1}^n \frac{X_i - \theta}{1 + (X_i - \theta)^2} = 0$, 不易求解;
 - 合理估计, $\hat{\theta}$ 为样本中位数.

3 优良性准则

- 定义(无偏性) 定义 $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ 为 θ 的估计量 (随机变量). 我们称一个估计 $\hat{\theta}$ 的偏差 (Bias):

$$E_{\theta}(\hat{\theta} - \theta) = E_{\theta}(\hat{\theta}) - \theta.$$

若 $E_{\theta}(\hat{\theta} - \theta) = 0, \forall \theta$, 则称 $\hat{\theta}$ 为 θ 的一个无偏估计量.

- 注
 - 一般地, $E_{\theta}(\hat{g}(X_1, \dots, X_n) - g(\theta)) = 0, \forall \theta \Leftrightarrow \hat{g}$ 为 $g(\theta)$ 的无偏估计;
 - 无偏 $\Leftrightarrow$ 无系统偏差;
 - $\frac{1}{N} \sum_{m=1}^N \hat{\theta}(X_1^{(m)}, \dots, X_n^{(m)}) \xrightarrow{a.s.} E(\hat{\theta}) \stackrel{\text{无偏}}{=} \theta$;
 - 无偏性的重要性视情形而定.
- 例 $E(X) = \mu, \text{Var}(X) = \sigma^2$.
- 解答

- $E(\bar{X}) = \mu \Rightarrow \bar{X}$ 为 μ 的无偏估计;
- $E(m_2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n} \sigma^2$;
- $E(S^2) = E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \sigma^2$.

• 例 $X_i \stackrel{iid}{\sim} U(0, \theta)$.

• 解答

- 矩估计 $\hat{\theta} = 2\bar{X}$;
- MLE $\theta^* = \max\{X_1, \dots, X_n\}$;
- 计算可得 $E(\hat{\theta}) = \theta$, $E(\theta^*) = \frac{n}{n+1}\theta$.

• 注 MLE 可也是有偏的 (本例中即为系统偏小).

• 定义(均方误差准则) 定义一个估计 $\hat{\theta}$ 的均方误差 (MSE):

$$E_{\theta}[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + E_{\theta}^2(\hat{\theta} - \theta).$$

其中 $\text{Var}(\hat{\theta})$ 代表着精确度, $E_{\theta}^2(\hat{\theta} - \theta)$ 代表着准确度.

• 定义 假定 $\hat{\theta}_1, \hat{\theta}_2$ 均为 θ 的无偏估计, 若

$$\forall \theta, \text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$$

且

$$\exists \theta_0, \text{Var}_{\theta_0}(\hat{\theta}_2) < \text{Var}_{\theta_0}(\hat{\theta}_1).$$

则称在均方误差意义下 $\hat{\theta}_1$ 优于 $\hat{\theta}_2$.

• 例 $E(X) = \mu$.

• 解答

$$E(\bar{X}) = E(X_1) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \mu.$$

$$\text{Var}(\bar{X}) = \frac{1}{n}\sigma^2, \text{Var}(X_1) = \sigma^2, \text{Var}\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}\sigma^2.$$

• 注 有时也可称为“有效性准则”.

• 定义 假定 $\hat{\theta}_0$ 为 θ 的无偏估计, 若对任意 θ 与 θ 的无偏估计 $\hat{\theta}$

$$\text{Var}(\hat{\theta}_0) \leq \text{Var}(\hat{\theta})$$

则称 $\hat{\theta}_0$ 为最小方差无偏估计 (MVUE).

• 例 $X \sim N(\mu, \sigma^2)$.

• 解答 由于

$$E(m_2) = \frac{n-1}{n}\sigma^2, E(S^2) = \sigma^2.$$

因此计算可得

$$E[(m_2 - \sigma^2)^2] = \dots < E[(S^2 - \sigma^2)^2] = \dots$$

这是用低偏差换了方差.

- 定义

- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ 的分布称为抽样分布;
- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ 的标准差称为标准误差, 记为 $se(\hat{\theta})$.

- 定义(大样本性质) 估计量 $\hat{\theta}$ 当 $n \rightarrow \infty$ 时的性质.

- 定义(渐进无偏性) 有 $\lim_{n \rightarrow \infty} E(\hat{\theta} - \theta) = 0$;

- 定义(相合性) 若 $\forall \varepsilon > 0$, 有 $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \geq \varepsilon) = 0$, 则称 $\hat{\theta}$ 为 θ 的一个相合估计, 即 $\hat{\theta} \xrightarrow{P} \theta$.

- $WLLN \Rightarrow \bar{X}$ 是 μ 的一个相合估计;
- 相合性是良好点估计的自然要求.

- 例 我们得到

$$\begin{aligned} m_2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu - \bar{X} + \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2 \\ &\xrightarrow{P} E[(X - \mu)^2] \\ &= \sigma^2. \end{aligned}$$

进而有

$$S^2 = \frac{n}{n-1} m_2 \xrightarrow{P} \sigma^2.$$

- 定义(渐进正态性) $\frac{\hat{\theta} - \theta}{se(\hat{\theta})} \xrightarrow{d} Z \sim (0, 1)$ (渐进正态估计).

- $\hat{\theta} \overset{\text{近似}}{\sim} N(0, se^2(\hat{\theta}))$;
- $CLT \Rightarrow \bar{X}$ 为 μ 的一个渐进正态估计, $se(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

4 置信区间 (Confidence Interval)

- 定义(置信区间) $\forall \alpha \in (0, 1), \forall \theta$, 有 $\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n)$, 其中 $i = 1, 2$, 使得

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1 - \alpha.$$

则称 $(\hat{\theta}_1, \hat{\theta}_2)$ 为 θ 的 $(1 - \alpha)$ 置信的区间估计 (双侧).

- 注

- α 通常取为 0.05, 0.01, 0.1;
- 置信 (水平、系数、度) 是针对方法的;
- 可靠性优先原则 (先保证 $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1 - \alpha$), 再尽量提升精度 (通常用 $E(\hat{\theta}_2 - \hat{\theta}_1)$ 刻画).

● 例 $X \sim N(\mu, \sigma^2)$, 已知 σ^2 , 给出 μ 的 $(1 - \alpha)$ 置信区间估计.

● 解答 注意到

$$\begin{aligned}\bar{X} - \mu &\sim N\left(0, \frac{\sigma^2}{n}\right) \\ \iff \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} &\sim N(0, 1) \\ \implies P\left(\left|\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| < Z_{\frac{\alpha}{2}}\right) &= 1 - \alpha.\end{aligned}$$

其中 $Z_{\frac{\alpha}{2}}$ 为 $N(0, 1)$ 的上侧 $\frac{\alpha}{2}$ 分位数. 因此所求 $(1 - \alpha)$ 置信的区间估计为

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right).$$

● 注

- 若 $\alpha = 0.05$, 则 $Z_{\frac{\alpha}{2}} = Z_{0.025} \approx 1.96 \approx 2$;
- 若用 $\bar{X}$ 估计 μ , 则有 $(1 - \alpha)$ 置信, 误差绝对值 $|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$;
- 给定 $\varepsilon > 0$, 则 $n \geq \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{\varepsilon}\right)^2$ 时, 有 $(1 - \alpha)$ 置信, 误差绝对值 $\leq \varepsilon$.

● 例 $X \sim N(\mu, \sigma^2)$, 未知 σ^2 , 给出 μ 的 $(1 - \alpha)$ 置信区间估计.

● 解答 注意到

$$\begin{aligned}\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} &\sim N(0, 1), \\ \frac{(n-1)S^2}{\sigma^2} &\sim \chi^2(n-1).\end{aligned}$$

注意到 $\bar{X}, S^2$ 相互独立, 由陈书 P_{93-94} 可得

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} / \frac{S}{\sigma} \sim t(n-1).$$

即

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

因此

$$P\left(\left|\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\right| < t_{\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha.$$

其中 $t_{\frac{\alpha}{2}}(n-1)$ 为 $t(n-1)$ 分布上侧 $\frac{\alpha}{2}$ 分位数, 所求 $(1-\alpha)$ 置信区间估计为

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)\right).$$

- 注 $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}, \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ 成为枢轴变量.
- 例 $X \sim N(\mu, \sigma^2)$, 未知 μ , 给出 σ^2 的 $(1-\alpha)$ 置信区间估计.
- 解答 注意到

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

因此所求 $(1-\alpha)$ 置信区间估计为

$$\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}\right).$$

- 注 我们有

$$\begin{aligned}\frac{(n-1)S^2}{\sigma^2} &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \\ &= \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \\ &\sim \chi^2(n-1).\end{aligned}$$

- 例 $X \sim N(\mu_1, \sigma^2), Y \sim N(\mu_2, \sigma^2)$ 相互独立, 未知 μ_1, μ_2, σ^2 , 给出 $\mu_1 - \mu_2$ 的 $(1-\alpha)$ 置信区间估计.
- 解答 设 $X_1, \dots, X_n, Y_1, \dots, Y_m$ 为随机样本, S_1^2 为 X 的样本方差, S_2^2 为 Y 的样本方差, 则

$$(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2) \sim N\left(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right).$$

因为

$$\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(n+m-2).$$

记

$$S^2 = \frac{n-1}{n+m-2}S_1^2 + \frac{m-1}{n+m-2}S_2^2.$$

从而得到

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} / \frac{S}{\sigma} \sim t(n + m - 2).$$

也即

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2).$$

这是我们用到的枢轴变量, 所求 $(1 - \alpha)$ 置信区间估计为

$$\left(\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}}(l) S \sqrt{\frac{m+n}{mn}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}}(l) S \sqrt{\frac{m+n}{mn}} \right).$$

其中 $l = n + m - 2$.

- **定义(渐进置信区间)** 通过大样本方法进行区间估计.
- **例(选举问题)** 真实支持率 p 未知, $n = 1200$, 观测比例 $\frac{684}{1200} \approx 0.57$. 给出 p 的一个 95% 置信的区间估计.
- **解答** $X_i \sim B(p)$ ($1 \leq i \leq n$) iid (近似有放回, $n \ll N$). 我们有

$$P_n = P_n(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

因此

$$E(P_n) = p, \text{Var}(P_n) = \frac{p(1-p)}{n}.$$

由 CLT, 得到

$$\frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1) \Rightarrow P\left(\left|\frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}}\right| < Z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

[1] 用 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 来估计 $\sigma^2 = p(1-p)$. 由 CLT, 得到

$$\frac{P_n - p}{\sqrt{\frac{S^2}{n}}} \overset{\text{近似}}{\sim} N(0, 1) \Rightarrow P\left(\left|\frac{P_n - p}{\sqrt{\frac{S^2}{n}}}\right| < Z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

其中 $S^2 \approx 0.2475$, 区间估计为 $(0.542, 0.598)$.

[2] 用 $m_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 来估计 $\sigma^2 = p(1-p)$.

实际上, 我们有

$$m_2 = \frac{1}{n}[(1 - P_n)^2 n P_n + (0 - P_n)^2 (n - n P_n)] = P_n(1 - P_n).$$

由 **CLT**, 得到

$$\frac{P_n - p}{\sqrt{\frac{P_n(1-P_n)}{n}}} \stackrel{\text{近似}}{\sim} N(0, 1) \Rightarrow p \in (0.542, 0.598).$$

[3] 用 $p(1 - p)$ 的最大值来估计 $\sigma^2 = p(1 - p)$. 由 **CLT**, 得到

$$\frac{P_n - p}{\sqrt{\frac{1}{4n}}} \stackrel{\text{近似}}{\sim} N(0, 1) \Rightarrow p \in (0.542, 0.599).$$

- 注 近似分布情况下, 置信度 $\approx 1 - \alpha$, 近似程度取决于 n 及总体分布.

- 定义(利用 MLE 构建置信区间)

总体分布 f 满足一定的光滑性条件, 存在 $\sigma_n > 0$, 使 $\frac{\theta^* - \theta}{\sigma_n} \rightarrow N(0, 1)$.

- 定义(Fisher 信息量) $X_1, \dots, X_n$ iid, 其 PDF 为 $f(x; \theta)$. 定义对数似然函数

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta).$$

定义

$$I_n(\theta) = E\left[\left(\frac{\partial l(\theta)}{\partial \theta}\right)^2\right] = E\left[\left(\sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)^2\right].$$

注意到

$$\begin{aligned} E\left(\frac{\partial \log f(X_i; \theta)}{\partial \theta}\right) &= E\left(\frac{f_\theta(X_i; \theta)}{f(X_i; \theta)}\right) \\ &= \int \frac{f_\theta(X_i; \theta)}{f(X_i; \theta)} f(X_i; \theta) dx \\ &= \int f_\theta(X_i; \theta) dx \\ &= \frac{d}{d\theta} \int f(X_i; \theta) dx = 0. \end{aligned}$$

因此

$$\begin{aligned} I_n(\theta) &= E\left[\left(\sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)^2\right] \\ &= \sum_{i=1}^n E\left[\left(\frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)^2\right] \\ &:= nI(\theta). \end{aligned}$$

由极大似然估计的定义, 我们有

$$0 = l'(\theta^*) \approx l'(\theta) + (\theta^* - \theta)l''(\theta) \Rightarrow \theta^* - \theta \approx \frac{l'(\theta)}{-l''(\theta)}$$

$$\Rightarrow \sqrt{n}(\theta^* - \theta) \approx \frac{\frac{1}{\sqrt{n}}l'(\theta)}{-\frac{1}{n}l''(\theta)}.$$

研究上式分子: 设随机变量 $Y_i = \frac{\partial \log f(X_i; \theta)}{\partial \theta} = \frac{f_\theta(X_i; \theta)}{f(X_i; \theta)}$, 那么

$$E(Y_i) = 0, \text{Var}(Y_i) = E(Y_i^2) - E^2(Y_i) = I(\theta).$$

从而

$$\frac{1}{\sqrt{n}}l'(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \frac{\bar{Y}}{\frac{1}{\sqrt{n}}} \xrightarrow{CLT} N(0, I(\theta)).$$

研究上式分母:

$$-\frac{1}{n}l''(\theta) = -\frac{1}{n} \sum_{i=1}^n \left[\frac{f_{\theta\theta}(X_i; \theta)}{f(X_i; \theta)} - \left(\frac{f_\theta(X_i; \theta)}{f(X_i; \theta)} \right)^2 \right].$$

注意到

$$E\left(\frac{f_{\theta\theta}(X_i; \theta)}{f(X_i; \theta)}\right) = 0.$$

因此

$$-\frac{1}{n}l''(\theta) \xrightarrow{WLLN} E\left[\frac{1}{n} \sum_{i=1}^n \left(\frac{f_\theta(X_i; \theta)}{f(X_i; \theta)}\right)^2\right] = \frac{1}{n} \sum_{i=1}^n E(Y_i^2) = I(\theta).$$

结合上述讨论可得

$$\begin{aligned} \sqrt{n}(\theta^* - \theta) &\rightarrow N(0, \frac{1}{I(\theta)}) \\ &\Leftrightarrow \frac{\theta^* - \theta}{\frac{1}{\sqrt{nI(\theta)}}} \overset{\text{近似}}{\sim} N(0, 1) \\ &\Rightarrow \frac{\theta^* - \theta}{\frac{1}{\sqrt{nI(\theta^*)}}} \overset{\text{近似}}{\sim} N(0, 1). \end{aligned}$$

这说明可以取 $\sigma_n = \frac{1}{\sqrt{nI(\theta)}}$ 或 $\frac{1}{\sqrt{nI(\theta^*)}}$.

- **例(选举问题)** 真实支持率 p 未知, $n = 1200$, 观测比例 $\frac{684}{1200} \approx 0.57$. 给出 p 的一个 95% 置信的区间估计.
- **解答续** 注意到 $X_i \overset{iid}{\sim} B(p)$, p 的 MLE 为 $p^* = P_n$. 有似然函数

$$f(X; p) = p^X(1-p)^{1-X} \Rightarrow \frac{\partial \log f}{\partial p} = \frac{X-p}{p(1-p)}.$$

Fisher 信息量为:

$$I(p) = E\left[\left(\frac{X-p}{p(1-p)}\right)^2\right] = \frac{1}{p(1-p)}.$$

得到 $\sigma_n = \frac{1}{\sqrt{p(1-p)}}$ 或 $\frac{1}{\sqrt{p^*(1-p^*)}}$. 这与前例 [2] 完全一致.

- 例(两总体) $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ 独立, 参数未知. 估计 $\mu_1 - \mu_2$.
- 解答 得到

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1).$$

因此

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \stackrel{\text{近似}}{\sim} N(0, 1).$$

5 Bayes 估计

- 定义(Bayes 估计)
 - 在搜集数据之前, 对参数 θ 有先验知识, 可用一个概率分布来刻画 (先验分布).
 - Θ — 随机变量, θ — Θ 的实现值.
 - X — 试验观测, x — 具体观测结果.
 - $f_\Theta(\theta)$ — 先验分布, $f_{X|\Theta}(x | \theta) = f_X(x, \theta)$ — 搜集的样本分布.
 - 更新先验分布为后验分布:

$$f_{\Theta|X}(\theta | x) = \frac{f(x, \theta)}{f_X(x)} = \frac{f_{X|\Theta}(x | \theta)f_\Theta(\theta)}{f_X(x)}.$$

其中

$$f_X(x) = \int_{\mathbb{R}} f(x, \theta) d\theta.$$

可通过后验期望或后验众数得到 Bayes 估计值.

- 例 投掷硬币, 正面向上的概率为 θ (未知). 抛 n 次观测到 x 次正面向上.
- 解答 对 θ 的先验认知: $\theta \sim U(0, 1)$ (Bayes 法则). 此时有 $f_\Theta(\theta) = 1$, $\theta \in (0, 1)$.

令 $X = n$ 次试验正面向上的次数, 给定 θ 时, $X \sim B(n, \theta)$. 即

$$f_{X|\Theta}(x|\theta) = P(X=x|\Theta=\theta) = C_n^x \theta^x (1-\theta)^{n-x}, x=0,1,\dots,n.$$

故 (X, Θ) 的联合分布为

$$f(x, \theta) = C_n^x \theta^x (1-\theta)^{n-x}, \theta \in (0, 1), x=0,1,\dots,n.$$

得到

$$\begin{aligned} f_X(x) &= \int_0^1 f(x, \theta) d\theta = C_n^x \int_0^1 \theta^x (1-\theta)^{n-x} d\theta \\ &= \frac{n!}{x!(n-x)!} \cdot \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} = \frac{1}{n+1} \end{aligned}$$

进而

$$f_{\Theta|X}(\theta|x) = \frac{f(x, \theta)}{f_X(x)} = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \cdot \theta^x (1-\theta)^{n-x}.$$

注意到这恰好为 $\beta(x+1, n-x+1)$ 的 **PDF**.

后验均值估计值为

$$\hat{\theta} = E(\Theta|X) = \int_0^1 \theta f_{\Theta|X}(\theta|x) d\theta = \frac{x+1}{n+2}.$$

后验众数估计值为 $f_{\Theta|X}(\theta|x)$ 的最大值点

$$\theta^* = \frac{x}{n}.$$

● 注

- 若 $n=20, x=13$, 后验分布服从 $\beta(14, 8)$.
- 此时 $P(\Theta > \frac{1}{2}) \approx 0.91$, 而 $\Theta < \frac{1}{4}$ 基本不可能.

● 注(合理的点估计)

- 后验众数;
- 后验均值.

● 注(先验分布选取)

- 本问题先验分布服从 $U(0, 1) = \beta(1, 1)$, 后验分布服从 $\beta(x+1, n-x+1)$;
- 一般地, 先验分布服从 $\beta(a, b)$, 后验分布服从 $\beta(a+x, b+n-x)$.

- **定义(Bayes 区间估计)** 得到参数 θ 的后验分布 $f_{\Theta}(\theta|x)$, 给定观测值 x 和实数 $\alpha \in (0, 1)$, 若对参数 θ 的任意可能取值, 都成立

$$P(a < \Theta < b|x) \geq 1 - \alpha.$$

则称 (a, b) 为 θ 的 $(1 - \alpha)$ 置信的 **Bayes** 区间估计.

- 注
 - 最大后验区间 (一般单峰);
 - 等尾可信区间.
- 例 $X \sim N(\mu, \sigma^2)$, σ^2 已知, 给出 μ 的 $(1 - \alpha)$ 置信的 Bayes 区间估计.
- 解答 取 μ 的先验分布 $f(\mu) \propto 1$, 有后验分布为 $N(\bar{X}, \frac{\sigma^2}{n})$.

进一步可得

$$\frac{\mu - \bar{X}}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

得到所求 Bayes 估计区间为

$$(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}).$$

6 Review

6.1 样本

- $X_1, \dots, X_n$ 为样本, 通常为 iid.
- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ 为统计量.
- 样本分布: $f(x_1, \dots, x_n; \theta)$ (PDF\PMF).
- 抽样分布: $\hat{\theta}$ 的分布.
- 标准误差: $se = se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$.

6.2 经典估计的优良性

- n 固定: 无偏性 & 有效性 (MSE 的分解)
- $n \rightarrow \infty$: 渐近无偏 & 相合性 (WLLN 一致性) & 渐近正态性 (CLT)

6.3 置信区间

- $\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n) (i = 1, 2)$.
- $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1 - \alpha$.
- 置信 (水平、系数、度) 是针对方法.
- 枢轴变量: $H(\hat{\theta}, \theta)$ 表示分布可用.

6.4 置信区间 v.s. Bayes 区间

- (置信区间) 对具体的样本观测值 $(x_1, \dots, x_n)$, 代入 $\hat{\theta}_1, \hat{\theta}_2$ 得到具体区间.

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1 - \alpha.$$

- (Bayes 区间) 对具体的样本观测值 $(x_1, \dots, x_n)$, 代入 a, b 得到具体区间.

$$P(a < \Theta < b \mid x_1, \dots, x_n) \geq 1 - \alpha.$$

6.5 小样本方法 v.s. 大样本方法

- 精确分布 v.s. 近似分布

6.6 Fisher 信息量

$$\frac{\theta^* - \theta}{\sigma_n} \underset{\sim}{\text{近似}} N(0, 1).$$

$$\sigma_n = \sqrt{\frac{1}{nI(\theta)}} \text{ 或 } \sqrt{\frac{1}{nI(\theta^*)}}.$$