

下界（信息不等式）的完整推导

曹植

SNNU, IEHP

2026 年 8 月 31 日

内容提要

- 1 问题设定
- 2 正则条件
- 3 得分函数及其性质
- 4 协方差的关键公式
- 5 应用 Cauchy-Schwarz 不等式
- 6 推导信息不等式
- 7 无偏估计的特例
- 8 等号成立的条件
- 9 有效估计量的性质
- 10 多参数情形的推广
- 11 应用举例
- 12 总结

问题设定

- 总体 X 的概率密度函数为 $f(x; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$ 是未知参数。
- 考虑来自该总体的一个样本（先考虑样本量为 1 的情况，对于独立同分布样本，只需将 Fisher 信息量乘以样本量 n 即可推广）。
- 设 $S = S(X)$ 是 θ 的一个估计量，其期望为：

$$\psi(\theta) = \mathbb{E}_{\theta}[S] = \int S(x) f(x; \theta) dx$$

- 估计量的偏差定义为：

$$B(\theta) = \psi(\theta) - \theta$$

- 目标：推导 $\text{Var}_{\theta}(S)$ 的下界。

正则条件 1-2

正则条件 (支撑集不变)

支撑集与 θ 无关, 即集合 $\{x: f(x; \theta) > 0\}$ 不依赖于 θ 。

正则条件 (积分与微分可交换)

对 $T(x) = 1$ 和 $T(x) = S(x)$, 以下交换成立:

$$\frac{\partial}{\partial \theta} \int T(x) f(x; \theta) dx = \int T(x) \frac{\partial}{\partial \theta} f(x; \theta) dx$$

- 这一条件保证我们可以将导数符号移入积分内部
- 是后续推导中交换积分与微分的理论基础

正则条件 3-4

正则条件 (Fisher 信息量存在且有限)

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

正则条件 (期望的导数等于导数的期望)

这已隐含在条件 2 中, 对 $T(x) = 1$ 和 $T(x) = S(x)$ 成立。

- 这些条件是 Cramér–Rao 下界成立的基础
- 在实际应用中, 多数常见分布 (指数族) 都满足这些条件

得分函数的定义

定义 (得分函数)

得分函数定义为对数似然函数对参数的导数：

$$\ell'(\theta) = \frac{\partial}{\partial \theta} \log f(X; \theta) = \frac{\frac{\partial}{\partial \theta} f(X; \theta)}{f(X; \theta)} = \frac{f'(X; \theta)}{f(X; \theta)}$$

其中记 $f'(X; \theta) = \frac{\partial}{\partial \theta} f(X; \theta)$ 。

- 得分函数衡量了似然函数对参数的敏感度
- 它是后续推导中的核心工具

得分的期望为零

性质 (得分的期望为零)

由条件 2 (对 $T(x) = 1$ 交换积分与微分):

$$\int f'(x; \theta) dx = \frac{\partial}{\partial \theta} \int f(x; \theta) dx = \frac{\partial}{\partial \theta}(1) = 0$$

因此:

$$\begin{aligned}\mathbb{E}_{\theta}[\ell'(\theta)] &= \int \frac{f'(x; \theta)}{f(x; \theta)} f(x; \theta) dx \\ &= \int f'(x; \theta) dx = 0\end{aligned}$$

- 这一性质在后续推导中至关重要
- 它简化了协方差的计算

得分的方差

性质 (得分的方差)

由于期望为零, 得分的方差即为二阶矩:

$$\begin{aligned}\mathrm{Var}_{\theta}[\ell'(\theta)] &= \mathbb{E}_{\theta}[(\ell'(\theta))^2] - \mathbb{E}_{\theta}^2[\ell'(\theta)] \\ &= \mathbb{E}_{\theta}[(\ell'(\theta))^2] - 0 \\ &= \mathbb{E}_{\theta}[(\ell'(\theta))^2] = I(\theta)\end{aligned}$$

这正是 Fisher 信息量的定义。

- Fisher 信息量 $I(\theta)$ 度量了分布中包含的关于参数的信息量
- 信息量越大, 参数估计的潜在精度越高

协方差的定义

协方差用于衡量两个变量总体误差的期望，其定义式为：
总体协方差：

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

展开可得：

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

对于样本数据，样本协方差的计算公式为：

$$q_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

- 分母使用 $n - 1$ 是为了得到无偏估计

估计量与得分的协方差

考虑估计量 S 与得分函数 $\ell'(\theta)$ 的协方差。

根据协方差定义，由于 $\mathbb{E}_\theta[\ell'(\theta)] = 0$ ，协方差简化为：

$$\begin{aligned}\text{Cov}_\theta(S, \ell'(\theta)) &= \mathbb{E}_\theta[S\ell'(\theta)] - \mathbb{E}_\theta[S] \cdot \mathbb{E}_\theta[\ell'(\theta)] \\ &= \mathbb{E}_\theta[S\ell'(\theta)] - \mathbb{E}_\theta[S] \cdot 0 \\ &= \mathbb{E}_\theta[S\ell'(\theta)]\end{aligned}$$

接下来计算 $\mathbb{E}_\theta[S\ell'(\theta)]$ 。

计算 $\mathbb{E}_\theta[S\ell'(\theta)]$

$$\begin{aligned}\mathbb{E}_\theta[S\ell'(\theta)] &= \int S(x)\ell'(\theta)f(x;\theta) dx \\ &= \int S(x)\frac{f'(x;\theta)}{f(x;\theta)}f(x;\theta) dx \\ &= \int S(x)f'(x;\theta) dx\end{aligned}$$

由条件 2 (对 $T(x) = S(x)$ 交换积分与微分):

$$\int S(x)f'(x;\theta) dx = \frac{\partial}{\partial\theta} \int S(x)f(x;\theta) dx$$

而 $\int S(x)f(x;\theta) dx = \psi(\theta)$, 因此:

$$\int S(x)f'(x;\theta) dx = \frac{\partial}{\partial\theta}\psi(\theta) = \psi'(\theta)$$

关键等式

综合以上推导，我们得到**关键等式**：

$$\text{Cov}_{\theta}(S, \ell'(\theta)) = \psi'(\theta)$$

这个等式建立了估计量的期望导数与得分函数的协方差之间的联系，是整个 Cramér–Rao 下界推导的核心。

- 左边是估计量与得分的协方差
- 右边是估计量期望的导数
- 这一联系使得我们可以将估计量的方差与 Fisher 信息量关联起来

Cauchy-Schwarz 不等式

对任意两个随机变量 U, V , Cauchy-Schwarz 不等式给出:

$$[\text{Cov}(U, V)]^2 \leq \text{Var}(U) \cdot \text{Var}(V)$$

等号成立当且仅当 U 与 V 几乎处处线性相关, 即存在常数 a, b 使得:

$$U = aV + b \quad (\text{几乎处处成立})$$

这个不等式是推导方差下界的数学基础。

代入已知结果

取 $U = S$, $V = \ell'(\theta)$, 代入 Cauchy-Schwarz 不等式:

$$[\text{Cov}_\theta(S, \ell'(\theta))]^2 \leq \text{Var}_\theta(S) \cdot \text{Var}_\theta(\ell'(\theta))$$

代入已知结果:

- $\text{Cov}_\theta(S, \ell'(\theta)) = \psi'(\theta)$
- $\text{Var}_\theta(\ell'(\theta)) = I(\theta)$

得到:

$$[\psi'(\theta)]^2 \leq \text{Var}_\theta(S) \cdot I(\theta)$$

信息不等式的一般形式

从上式整理得：

$$\text{Var}_{\theta}(S) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}$$

又 $\psi(\theta) = \theta + B(\theta)$ ，两边对 θ 求导：

$$\psi'(\theta) = 1 + B'(\theta)$$

代入即得信息不等式（Cramér–Rao 下界）的一般形式：

$$\boxed{\text{Var}_{\theta}(S) \geq \frac{[1 + B'(\theta)]^2}{I(\theta)}}$$

这个不等式给出了估计量方差的下界，是数理统计中最基本的不等式之一。

推导的详细步骤回顾

让我们回顾整个推导过程：

- ① 从估计量 S 和得分函数 $\ell'(\theta)$ 的定义出发
- ② 计算协方差 $\text{Cov}_\theta(S, \ell'(\theta)) = \psi'(\theta)$
- ③ 应用 Cauchy-Schwarz 不等式： $[\text{Cov}(U, V)]^2 \leq \text{Var}(U)\text{Var}(V)$
- ④ 代入： $[\psi'(\theta)]^2 \leq \text{Var}_\theta(S) \cdot I(\theta)$
- ⑤ 整理得： $\text{Var}_\theta(S) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}$
- ⑥ 代入 $\psi'(\theta) = 1 + B'(\theta)$ 得到最终形式

无偏估计的 Cramér–Rao 下界

若估计量无偏，即 $B(\theta) \equiv 0$ ，则 $B'(\theta) \equiv 0$ ，于是：

$$\text{Var}_{\theta}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

这就是最常见的 Cramér–Rao 下界形式。

- 表明无偏估计的方差不能低于 Fisher 信息量的倒数
- Fisher 信息量越大，方差下界越小，估计越精确
- 为评估估计量的优劣提供了理论基准

等号成立的充要条件

Cauchy-Schwarz 不等式等号成立当且仅当 S 与 $\ell'(\theta)$ 几乎处处线性相关, 即存在常数 (可能依赖于 θ) $A(\theta)$ 和 $C(\theta)$ 使得:

$$\ell'(\theta) = A(\theta)S + C(\theta) \quad (\text{几乎处处成立})$$

由 $\mathbb{E}_\theta[\ell'(\theta)] = 0$, 得:

$$0 = A(\theta)\mathbb{E}_\theta[S] + C(\theta) \Rightarrow C(\theta) = -A(\theta)\psi(\theta)$$

代入上式:

$$\ell'(\theta) = A(\theta)[S - \psi(\theta)]$$

有效估计量的特征方程

对于无偏估计, $\psi(\theta) = \theta$, 因此:

$$\ell'(\theta) = A(\theta)(\hat{\theta} - \theta)$$

这就是 CRLB 等号成立的**充要条件**。

满足这一条件的估计量称为**有效估计量**, 它达到了方差下界:

$$\text{Var}(\hat{\theta}) = \frac{1}{I(\theta)}$$

从特征方程解出估计量

当 $\ell'(\theta) = A(\theta)(\hat{\theta} - \theta)$ 时, 可以解得:

$$\hat{\theta} - \theta = \frac{\ell'(\theta)}{A(\theta)}$$

进一步, 我们可以得到估计量的具体形式。例如, 对于指数族分布:

$$f(x; \theta) = h(x) \exp\{\eta(\theta) T(x) - A(\theta)\}$$

得分函数为:

$$\ell'(\theta) = \eta'(\theta)[T(x) - \mathbb{E}_{\theta} T(X)]$$

此时 $T(x)$ 就是有效估计量。

有效估计量的重要性质

当估计量达到 Cramér–Rao 下界时，具有以下重要性质：

- **方差达到下界：** $\text{Var}(\hat{\theta}) = \frac{1}{I(\theta)}$
- **充分性：** $\hat{\theta}$ 是充分统计量的函数
- **指数族分布：** 有效估计量通常出现在指数族分布中，此时得分函数具有线性形式
- **一致性：** 在正则条件下，有效估计量通常也是一致估计量
- **渐近正态性：** 有效估计量通常是渐近正态的

有效估计量与充分统计量

有效估计量与充分统计量之间存在密切联系：

定理 (Fisher-Neyman 因子分解定理)

统计量 $T(X)$ 是充分的当且仅当联合密度可以分解为：

$$f(x; \theta) = g(T(x); \theta)h(x)$$

对于有效估计量， $\hat{\theta}$ 通常是充分统计量的函数，且得分函数可以写成：

$$\ell'(\theta) = \frac{\partial}{\partial \theta} \log g(T(x); \theta)$$

这正是有效估计量的特征形式。

多参数 Cramér–Rao 下界

对于多维参数 $\theta = (\theta_1, \dots, \theta_k)^T$, Cramér–Rao 下界可以推广为矩阵形式。设 $\hat{\theta}$ 是 θ 的一个无偏估计量, 则其协方差矩阵满足:

$$\text{Cov}(\hat{\theta}) \geq \mathbf{I}(\theta)^{-1}$$

其中 $\mathbf{I}(\theta)$ 是 Fisher 信息矩阵。

Fisher 信息矩阵 $\mathbf{I}(\boldsymbol{\theta})$ 的 (i, j) 元素为:

$$I_{ij}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial \log f(X; \boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \log f(X; \boldsymbol{\theta})}{\partial \theta_j} \right]$$

或者等价地:

$$I_{ij}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial^2 \log f(X; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]$$

不等式 $\text{Cov}(\hat{\boldsymbol{\theta}}) - \mathbf{I}(\boldsymbol{\theta})^{-1}$ 是半正定矩阵, 在矩阵意义下理解。

多参数情形的推导要点

多参数情形的推导与单参数类似：

- ① 得分向量 $\nabla \log f(X; \theta)$ 期望为零
- ② 得分向量的协方差矩阵就是 Fisher 信息矩阵
- ③ 考虑估计量向量与得分向量的协方差矩阵
- ④ 应用矩阵形式的 Cauchy-Schwarz 不等式（或多元 Cauchy-Schwarz）

对于无偏估计 $\hat{\theta}$ ，有：

$$\text{Cov}(\hat{\theta}) \geq \mathbf{I}(\theta)^{-1}$$

等号成立当且仅当 $\nabla \log f(X; \theta) = \mathbf{A}(\theta)(\hat{\theta} - \theta)$ 。

例 1: 正态分布均值估计

设 $X \sim N(\mu, \sigma^2)$, σ^2 已知, 估计 $\theta = \mu$ 。

概率密度函数:

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

对数似然函数:

$$\log f(x; \mu) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

得分函数:

$$\ell'(\mu) = \frac{\partial}{\partial \mu} \log f(x; \mu) = \frac{x-\mu}{\sigma^2}$$

例 1: 正态分布均值估计 (续)

Fisher 信息量:

$$I(\mu) = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma^2} \right)^2 \right] = \frac{1}{\sigma^4} \mathbb{E}[(X - \mu)^2] = \frac{1}{\sigma^2}$$

Cramér-Rao 下界:

$$\text{Var}(\hat{\mu}) \geq \frac{1}{I(\mu)} = \sigma^2$$

对于样本量为 n 的情况:

$$\text{Var}(\hat{\mu}) \geq \frac{1}{nI(\mu)} = \frac{\sigma^2}{n}$$

样本均值 $\bar{X}$ 的方差为 $\frac{\sigma^2}{n}$, 恰好等于 CRLB, 因此是有效估计量。

例 1: 验证有效估计量条件

检验样本均值 $\bar{X}$ 是否满足有效估计量的特征方程:
得分函数为:

$$\ell'(\mu) = \frac{x - \mu}{\sigma^2}$$

对于样本量为 n 的情况, 得分函数为:

$$\ell'(\mu) = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} = \frac{n(\bar{x} - \mu)}{\sigma^2}$$

这正是 $\ell'(\mu) = A(\mu)(\bar{X} - \mu)$ 的形式, 其中 $A(\mu) = \frac{n}{\sigma^2}$ 。
因此 $\bar{X}$ 是有效估计量。

例 2: 泊松分布参数估计

设 $X \sim \text{Poisson}(\lambda)$, 估计 $\theta = \lambda$ 。

概率质量函数:

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

对数似然函数:

$$\log f(x; \lambda) = -\lambda + x \log \lambda - \log(x!)$$

得分函数:

$$\ell'(\lambda) = \frac{\partial}{\partial \lambda} \log f(x; \lambda) = -1 + \frac{x}{\lambda}$$

例 2: 泊松分布参数估计 (续)

Fisher 信息量:

$$I(\lambda) = \mathbb{E} \left[\left(-1 + \frac{X}{\lambda} \right)^2 \right] = \text{Var} \left(\frac{X}{\lambda} \right) = \frac{1}{\lambda^2} \text{Var}(X) = \frac{1}{\lambda^2} \cdot \lambda = \frac{1}{\lambda}$$

Cramér-Rao 下界:

$$\text{Var}(\hat{\lambda}) \geq \frac{1}{I(\lambda)} = \lambda$$

对于样本量为 n 的情况:

$$\text{Var}(\hat{\lambda}) \geq \frac{1}{nI(\lambda)} = \frac{\lambda}{n}$$

样本均值 $\bar{X}$ 的方差为 $\frac{\lambda}{n}$, 恰好等于 CRLB, 因此是有效估计量。

例 2: 验证有效估计量条件

检验样本均值 $\bar{X}$ 是否满足有效估计量的特征方程:

对于单个观测, 得分函数为:

$$\ell'(\lambda) = \frac{x}{\lambda} - 1$$

对于样本量为 n 的情况, 得分函数为:

$$\ell'(\lambda) = \sum_{i=1}^n \left(\frac{x_i}{\lambda} - 1 \right) = \frac{n\bar{x}}{\lambda} - n = \frac{n}{\lambda}(\bar{x} - \lambda)$$

这正是 $\ell'(\lambda) = A(\lambda)(\bar{X} - \lambda)$ 的形式, 其中 $A(\lambda) = \frac{n}{\lambda}$ 。
因此 $\bar{X}$ 是有效估计量。

例 3: 伯努利分布参数估计

设 $X \sim \text{Bernoulli}(p)$, 估计 $\theta = p$ 。

概率质量函数:

$$f(x; p) = p^x(1 - p)^{1-x}, \quad x = 0, 1$$

对数似然函数:

$$\log f(x; p) = x \log p + (1 - x) \log(1 - p)$$

得分函数:

$$\ell'(p) = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)}$$

例 3: 伯努利分布参数估计 (续)

Fisher 信息量:

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

Cramér-Rao 下界:

$$\text{Var}(\hat{p}) \geq \frac{1}{I(p)} = p(1-p)$$

对于样本量为 n 的情况:

$$\text{Var}(\hat{p}) \geq \frac{1}{nI(p)} = \frac{p(1-p)}{n}$$

样本比例 $\hat{p} = \bar{X}$ 的方差为 $\frac{p(1-p)}{n}$, 恰好等于 CRLB, 因此是有效估计量。

完整推导步骤总结

- ① 建立得分函数: $\ell'(\theta) = \frac{\partial}{\partial \theta} \log f(X; \theta)$
- ② 得分性质: $\mathbb{E}_{\theta}[\ell'(\theta)] = 0$, $\text{Var}_{\theta}[\ell'(\theta)] = I(\theta)$
- ③ 关键等式: 利用积分与微分可交换, 推导出

$$\text{Cov}_{\theta}(S, \ell'(\theta)) = \psi'(\theta)$$

- ④ Cauchy-Schwarz 不等式:

$$[\text{Cov}_{\theta}(S, \ell'(\theta))]^2 \leq \text{Var}_{\theta}(S) \cdot \text{Var}_{\theta}(\ell'(\theta))$$

- ⑤ 代入得:

$$[\psi'(\theta)]^2 \leq \text{Var}_{\theta}(S) \cdot I(\theta)$$

完整推导步骤总结（续）

⑥ 整理得：

$$\text{Var}_{\theta}(S) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}$$

⑦ 代入 $\psi'(\theta) = 1 + B'(\theta)$ ：

$$\text{Var}_{\theta}(S) \geq \frac{[1 + B'(\theta)]^2}{I(\theta)}$$

⑧ 无偏估计特例：

$$\text{Var}_{\theta}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

⑨ 等号成立条件：

$$\ell'(\theta) = A(\theta)(\hat{\theta} - \theta)$$

- **理论基准**：为评估估计量的优劣提供了理论下界
- **有效性判定**：可以用来判断一个估计量是否达到了最优
- **渐近理论**：极大似然估计在正则条件下是渐近有效的，即渐近达到 CRLB
- **实验设计**：Fisher 信息量越大，CRLB 越小，可以通过实验设计提高信息量
- **局限性**：仅适用于无偏或偏差已知的估计量；小样本情况下可能有更好的估计量（如 MSE 准则）

进一步的学习方向

- **多元参数**: 研究 Fisher 信息矩阵和多元 CRLB
- **渐近理论**: 极大似然估计的渐近正态性和渐近有效性
- **其他下界**: Chapman-Robbins 下界、Bhattacharyya 下界等
- **非正则情形**: 当正则条件不满足时的下界 (如均匀分布)
- **贝叶斯观点**: 后验方差与 Cramér-Rao 下界的关系

- Cramér, H. (1946). *Mathematical Methods of Statistics*.
- Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters.
- Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*.
- 茆诗松、王静龙、濮晓龙 (2012). 高等数理统计.

谢谢!