

AI 中的数学

第十七讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

① 极大似然估计

② 矩估计

③ 估计的无偏性

① 极大似然估计

② 矩估计

③ 估计的无偏性

§7.1 最大似然估计

- 总体 $X \sim F_\theta$. 目标: 给出 θ 的估计值 $\hat{\theta}$.
- 思想: 大概率事件发生. 支撑: 概率的主观置信度含义.
- 似然函数:

$$L(\theta) = p(x_1, \theta) \cdots p(x_n, \theta),$$

其中, $p(\theta, x)$ 为总体的分布列/密度.

- 定义 1.1. θ 的最大似然估计 $\hat{\theta} = \hat{\theta}(x_1, \cdots, x_n)$ 指 $L(\theta)$ 的最大值点. 即, $\hat{\theta} \in \Theta$, 且

$$L(\hat{\theta}) = \max_{\theta \in \Theta} p(x_1, \theta) \cdots p(x_n, \theta).$$

离散统计模型：设 (X_1, \dots, X_n) 为独立重复观察得到的样本，其中 $X_i (i = 1, \dots, n)$ 为离散型随机变量，样本分布列具有下列一般性质：

$$P_{\theta}((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \prod_{i=1}^n P_{\theta}(X_i = x_i) \quad (\theta \in \Theta),$$

此时 θ 为参数，对于固定的样本值 (x_1, \dots, x_n) ，作为 θ 的函数

$$L(\theta) = \prod_{i=1}^n P_{\theta}(X_i = x_i)$$

称为似然函数。

连续统计模型：此时 $X_i (i = 1, \dots, n)$ 为连续型随机变量，样本 (X_1, \dots, X_n) 具有联合密度

$$\prod_{i=1}^n p(x_i, \theta) \quad (\theta \in \Theta).$$

对于固定的样本值 (x_1, \dots, x_n) ， θ 的函数

$$L(\theta) = \prod_{i=1}^n p(x_i, \theta) \quad (1)$$

也称为似然函数。

定义：设 $\theta \in \Theta$ 为统计模型 $(X_1, \dots, X_n) \sim P_\theta$ 的参数，统计模型可为连续型或者离散型，又设 x_1, \dots, x_n 为总体的样本值，若存在 $\hat{\theta}(x_1, \dots, x_n)$ ，使得

$$L(\hat{\theta}(x_1, \dots, x_n)) = \max_{\theta \in \Theta} L(\theta),$$

其中 $L(\cdot)$ 为离散统计模型或者连续统计模型的似然函数，则称 $\hat{\theta}(x_1, \dots, x_n)$ 为 θ 的最大似然估计（简称 ML 估计）。若 $\hat{\theta}$ 为参数 θ 的 ML 估计，则 θ 的函数 $g(\theta)$ 的 ML 估计定义为 $g(\hat{\theta})$ 。

- 在似然函数 $L(\theta) = p(x_1, \theta) \cdots p(x_n, \theta)$ 中, 视数据 \mathbf{x} 为已知, 视参数 θ 为未知.
- 用 $\hat{\theta}(x_1, \cdots, x_n)$ 则强调计算, 用 $\hat{\theta}(X_1, \cdots, X_n)$ 则强调理论.
- $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ 是统计量.
- 必须有 $\hat{\theta} \in \Theta$.
- $g(\hat{\theta})$ 是 $g(\theta)$ 的最大似然估计.
- $g(\theta)$ 是一个一一变换, $\eta = g(\theta)$ 可以视为模型的一个新参数化.
- 最大似然估计 (最大值点) 可能不唯一.
- 机器学习中的损失函数对应于负的对数似然函数

正态模型：考虑随机变量 $X \sim N(\mu, \sigma^2)$ ，有独立重复观察得到的样本 $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ ，希望由这些观察值求出参数 μ, σ^2 的 ML 估计，对于样本值 (x_1, \dots, x_n) ，似然函数为

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\},$$

其中 $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ 为参数。

正态模型：考虑随机变量 $X \sim N(\mu, \sigma^2)$ ，有独立重复观察得到的样本 $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ ，希望由这些观察值求出参数 μ, σ^2 的 ML 估计，对于样本值 (x_1, \dots, x_n) ，似然函数为

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\},$$

其中 $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ 为参数。

为求似然函数 $L(\theta)$ 的最大值点，首先固定二维向量 $\theta = (\mu, \sigma^2)$ 的 σ^2 ，求 $L(\theta) = L(\mu, \sigma^2)$ 相对于变量 μ 的最大值点 $\mu^*(\sigma^2)$ ，随后代入 $L(\theta)$ ，得到 $L(\mu^*(\sigma^2), \sigma^2)$ ，再求 σ^2 的最大值点 σ^{2*} ，可得

$$L(\mu^*(\sigma^{2*}), \sigma^{2*}) \geq L(\mu^*(\sigma^2), \sigma^2) \geq L(\mu, \sigma^2),$$

因此 $(\mu^*(\sigma^{2*}), \sigma^{2*})$ 为 $\theta = (\mu, \sigma^2)$ 的 ML 估计。

首先求 $\mu^*(\sigma^2)$ 。注意到，求 μ 的值 μ^* 使得 $L(\mu, \sigma^2)$ 达到最大，等价于求 μ^* 使得 $\sum_{i=1}^n (x_i - \mu)^2$ 达到最小。由于

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2,$$

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ，可知当 $\mu = \bar{x}$ 时， $\sum_{i=1}^n (x_i - \mu)^2$ 达到最小。

将 $\mu = \bar{x}$ 代入 $L(\mu, \sigma^2)$ 的表达式中, 得

$$L(\bar{x}, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} \right\}.$$

由于 $\sigma^2 \in (0, +\infty)$, L 作为 σ^2 的函数恒取正值。现等价的求解方程 $\frac{\partial \ln L}{\partial \sigma^2} = 0$, 即

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

解得 $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 。由此可知, $\mu = \bar{x}$, $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 使得 $L(\mu, \sigma^2)$ 在它的定义域上达到最大, 再由 ML 估计定义知

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

分别为 μ 和 σ^2 的 ML 估计。

例：设对飞机的最大飞行速度进行测试，测得 15 个数据（单位 m/s ）如下：

422.2, 418.7, 425.6, 420.3, 425.8, 423.1, 431.5,

428.2, 434.0, 438.3, 412.3, 417.2, 413.5, 441.3, 423.7.

试估计飞机最大速度的均值。

解：将飞机最大飞行速度的观察值 x_1, \dots, x_{15} 视为随机变量 X_1, \dots, X_{15} 的观察值，随机变量的共同分布为正态分布 $N(\mu, \sigma^2)$ 。

将数据代入上面的结果，得

$$\hat{\mu} = \bar{x} = 425.05$$

例如实际问题中有下列两种常见的正态模型：一种是 $X \sim N(\mu_0, \sigma^2)$ ，其中 $\sigma^2 > 0$ ， μ_0 为已知的参数值。

例如实际问题中有下列两种常见的正态模型：一种是 $X \sim N(\mu_0, \sigma^2)$ ，其中 $\sigma^2 > 0$ ， μ_0 为已知的参数值。

此时带估参数只有 σ ，求解方程 $\frac{\partial \ln L}{\partial \sigma^2} = 0$ ，即

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu_0)^2 = 0$$

解得 σ^2 的 ML 估计为 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$ 。

另一种是 $X \sim N(\mu, \sigma_0^2)$, 其中 $\mu \in \mathbb{R}$, 而 σ_0^2 为已知的参数值, 此时求解 $L(\theta)$ 的关于 μ 的最大值仍等价于求解 μ^* 使 $\sum_{i=1}^n (x_i - \mu)^2$ 达到最小, 因此 μ 的 ML 估计为 $\hat{\mu} = \bar{x}$ 。

例 1.2 (次品率的估计). 某工人生产 20 件产品, 检查出恰有一件为次品. 估计该工人生产的次品率.

- 总体 $X \sim B(1, p), p \in [0, 1]$. 样本量: $n = 20$.
- 似然函数:

$$L(p) = C_n^s p^s (1-p)^{n-s}, \text{ 其中 } s = x_1 + \cdots + x_n.$$

- \hat{p} 也为 $\ln L(p) = s \ln p + (n-s) \ln(1-p)$ 的最大值点:

$$\frac{d}{dp} \ln L(p) = \frac{s}{p} - \frac{n-s}{1-p} \Rightarrow \hat{p} = \frac{s}{n}.$$

- $n = 20, s = 1$, 因此, $\hat{p} = \frac{1}{20}$.

例 1.4. 总体: $X \sim U(0, \theta)$, 数据: X_1, \dots, X_n , (样本量: n). 求: θ 的最大似然估计.

例 1.4. 总体: $X \sim U(0, \theta)$, 数据: X_1, \dots, X_n , (样本量: n). 求: θ 的最大似然估计.

- 似然函数:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} 1_{\{0 \leq x_i \leq \theta\}} = \frac{1}{\theta^n} 1_{\{0 \leq x_1, \dots, x_n \leq \theta\}}.$$

- 仅当 $\theta \geq \max_{1 \leq i \leq n} x_i$ 时, $L(\theta) > 0$.
- 当 $\theta \geq \max_{1 \leq i \leq n} x_i$ 时, $L(\theta) = \frac{1}{\theta^n}$, 关于 θ 单调下降.
- 从而, $\hat{\theta} = \max_{1 \leq i \leq n} x_i$, 即 $\max_{1 \leq i \leq n} X_i$.

例 1.5. 总体: $X \sim \text{Exp}(\lambda)$, 样本量: n . 求: EX 的最大似然估计.

- 似然函数:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} = \exp \{ n(\ln \lambda - \lambda \bar{x}) \}.$$

- $\hat{\lambda}$ 是 $\ln \lambda - \lambda \bar{x}$ 的最大值点:

$$\frac{d}{d\lambda} (\ln \lambda - \lambda \bar{x}) = \frac{1}{\lambda} - \bar{x} \Rightarrow \hat{\lambda} = 1/\bar{x}.$$

- $EX = 1/\lambda$, 因此, $\widehat{EX} = 1/\hat{\lambda} = \bar{x}$, 或 \bar{X} .

例 1.6. 总体: $X \sim \mathcal{P}(\lambda)$, 样本量: n . 求: λ 的最大似然估计.

例 1.6. 总体: $X \sim \mathcal{P}(\lambda)$, 样本量: n . 求: λ 的最大似然估计.

- 似然函数:

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{1}{\prod_{i=1}^n x_i!} \lambda^{n\bar{x}} e^{-\lambda n} = \frac{1}{\prod_{i=1}^n x_i!} e^{n(\bar{x} \ln \lambda - \lambda)}.$$

- $\hat{\lambda}$ 是 $\bar{x} \ln \lambda - \lambda$ 的最大值点:

$$\frac{d}{d\lambda}(\bar{x} \ln \lambda - \lambda) = \frac{\bar{x}}{\lambda} - 1 \Rightarrow \hat{\lambda} = \bar{x}, \text{ 或 } \bar{X}.$$

例：设总体 X 的分布函数为 $f(x, \theta) = \frac{2x}{\theta^2}, 0 \leq x \leq \theta$ ，设 (x_1, \dots, x_n) 为来自总体 X 的一个样本值，求分布中位数的 ML 估计。

例：设总体 X 的分布函数为 $f(x, \theta) = \frac{2x}{\theta^2}, 0 \leq x \leq \theta$ ，设 (x_1, \dots, x_n) 为来自总体 X 的一个样本值，求分布中位数的 ML 估计。

解：设分布中位数为 a ，则

$$\int_0^a f(x, \theta) dx = \int_a^\theta f(x, \theta) dx = \frac{1}{2}$$

其中

$$\int_0^a f(x, \theta) dx = \frac{a^2}{\theta^2} = \frac{1}{2} \Rightarrow a = \frac{\theta}{\sqrt{2}}$$

似然函数为

$$L(\theta) = \prod_{i=1}^n \frac{2x_i}{\theta^2} = \frac{2 \prod_{i=1}^n x_i}{\theta^{2n}}$$

由于 $\theta \geq x_i, i = 1, \dots, n$ ，上式关于 θ 单调减，故 θ 的 ML 估计为 $\hat{\theta} = \max\{x_1, \dots, x_n\}$ 。中位数 a 的 ML 估计为 $\frac{\max\{x_1, \dots, x_n\}}{\sqrt{2}}$ 。

① 极大似然估计

② 矩估计

③ 估计的无偏性

矩估计

- 思想: 样本矩 \approx 真实矩. 理论: LLN.
- 总体矩: $\alpha_k = \alpha_k(\theta) = E_{\theta} X^k$, 是参数的函数.
- 样本矩: $a_k = \overline{X^k} = \frac{1}{n} \sum_{i=1}^n X_i^k$ 或 $\overline{x^k}$, 是统计量.
- 定义 2.1:
 - (1) 称 a_k 为 α_k 的矩估计.
 - (2) 若待估量 $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$, 其中 ϕ 为连续函数, 则称 $\phi(a_1, \dots, a_k)$ 为 $g(\theta)$ 的矩估计.

定义设 X_1, \dots, X_n 为来自总体 $X \sim F_\theta (\theta \in \Theta)$ 的一个样本, 通常 $\alpha_l \triangleq E_\theta(X^l)$ 称为 l 阶总体矩, 而 $a_l \triangleq \frac{1}{n} \sum_{i=1}^n X_i^l$ 称为 l 阶样本矩。所涉及的矩存在且有限。

(1) l 阶总体矩 $\alpha_l = E_\theta(X^l)$ 的矩估计定义为相应的样本矩, 即

$$\hat{\alpha}_l = a_l = \frac{1}{n} \sum_{i=1}^n X_i^l, \quad l = 1, 2, \dots;$$

(2) 若存在连续函数 ϕ 使 $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$ 成立, 则 $g(\theta)$ 的矩估计定义为

$$\hat{g}(\theta) = \phi(a_1, \dots, a_k).$$

- 尽量使用低阶矩
- 将问题转化为解方程问题
- 估计值 $\hat{g}(\theta)$ 可能不在定义域内

例 2.1 (续例 1.1) 飞机最大飞行速度 $X \sim N(\mu, \sigma^2)$, 样本量: n .
求 μ, σ^2 的矩估计.

- Step 1. 将待估量写为矩的函数:

$$\mu = \alpha_1, \quad \sigma^2 = \alpha_2 - \alpha_1^2.$$

- Step 2. 求涉及到的样本矩: $a_1 = \bar{x}, a_2 = \overline{x^2}$.
- Step 3. 代入函数:

$$\hat{\mu} = \bar{x}, \quad \widehat{\sigma^2} = \overline{x^2} - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

例 2.4. 总体: $X \sim U(0, \theta)$. 样本量: n . 求 θ 的矩估计.

- $\alpha_1 = \frac{1}{2}\theta$, 即 $\theta = 2\alpha_1$. 故 $\hat{\theta}_1 = 2\bar{X}$ 是 θ 的矩估计.
- $\alpha_2 = \frac{1}{3}\theta^2$, 即 $\theta = \sqrt{3\alpha_2}$. 从而 $\hat{\theta}_2 = \sqrt{3\bar{X}^2}$ 也是 θ 的矩估计.
- 比较 $\hat{\theta}_1$ 与最大似然估计 $\hat{\theta} = \max_{1 \leq i \leq n} X_i$.

(1) 当 $2\bar{X} < \max_{1 \leq i \leq n} X_i$ 时, $\hat{\theta}_1$ 不合理. 但, $\hat{\theta}$ 总是合理.

(2) 期望:

$$E_{\theta}\hat{\theta}_1 = 2E\bar{X} = \theta, \quad E_{\theta}\hat{\theta} = \frac{n}{n+1}\theta$$

(3) 方差:

$$\begin{aligned} \text{var}_{\theta}(\hat{\theta}_1) &= \frac{4}{n} \text{var}_{\theta}(X) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{1}{3n} \theta^2 \\ \text{var}_{\theta}\left(\frac{n+1}{n}\hat{\theta}\right) &= \frac{(n+1)^2}{n^2} \cdot \frac{n}{(n+1)^2(n+2)} \theta^2 = \frac{1}{n(n+2)} \theta^2. \end{aligned}$$

如果总体 $\xi \sim B(N, p)$, N, p 为未知参数。

如果总体 $\xi \sim B(N, p)$, N, p 为未知参数。

因为 $E(\xi) = Np$, $D(\xi) = Np(1 - p)$, 由方程组

$$\begin{cases} Np = \bar{\xi} \\ Np(1 - p) = S^2 \end{cases}, \text{ 解得 } N, p \text{ 的矩估计为}$$

$$\begin{cases} \hat{N} = \frac{\bar{\xi}^2}{\bar{\xi} - S^2}, \\ \hat{p} = 1 - \frac{S^2}{\bar{\xi}}. \end{cases}$$

(2) 如果总体 $\xi \sim P(\lambda)$, λ 为未知参数, 因为 $E(\xi) = \lambda$, $D(\xi) = \lambda$, 所以 λ 的矩估计为

$$\hat{\lambda} = \bar{\xi} \text{ 或 } \hat{\lambda} = S^2.$$

一般使用 $E(\xi) = \lambda$ 。

(3) 如果总体服从几何分布, $P(\xi = k) = p(1 - p)^{k-1}$, p 为未知参数, 则因为 $E(\xi) = \frac{1}{p}$, 所以解得 p 的矩估计为

$$\hat{p} = \frac{1}{\bar{\xi}}.$$

(4) 如果总体 $\xi \sim U(\theta_1, \theta_2)$, θ_1, θ_2 均为未知参数, 因为

$$E(\xi) = \frac{\theta_1 + \theta_2}{2}, \quad D(\xi) = \frac{(\theta_1 - \theta_2)^2}{12}.$$

所以由方程

$$\frac{\theta_1 + \theta_2}{2} = \bar{\xi}, \quad \frac{(\theta_1 - \theta_2)^2}{12} = S^2$$

解得 θ_1, θ_2 的矩估计分别为

$$\hat{\theta}_1 = \bar{\xi} - \sqrt{3}S, \quad \hat{\theta}_2 = \bar{\xi} + \sqrt{3}S.$$

例：设总体 ξ 的分布为

$p(\xi = k) = (k-1)\theta^2(1-\theta)^{k-2}, k = 2, 3, \dots, 0 < \theta < 1.$, 求 θ 的矩估计。

例：设总体 ξ 的分布为

$p(\xi = k) = (k-1)\theta^2(1-\theta)^{k-2}, k = 2, 3, \dots, 0 < \theta < 1.$, 求 θ 的矩估计。

解：

$$\begin{aligned} E(\xi) &= \sum_{k=2}^{\infty} kp(\xi = k) = \sum_{k=2}^{\infty} k(k-1)\theta^2(1-\theta)^{k-2}, \\ &= 2 \sum_{k=2}^{\infty} C_k^2 \theta^2 (1-\theta)^{k-2} = 2\theta^2 \sum_{k=0}^{\infty} C_{k+2}^2 (1-\theta)^k \end{aligned}$$

注意到由展开式

$(1-x)^{-3} = \sum_{k=0}^{\infty} (-1)^k \frac{(k+1)(k+2)}{2} x^k = \sum_{k=0}^{\infty} (-1)^k C_{k+2}^2 x^k$, 因此

$$E(\xi) = 2\theta^2 \sum_{k=0}^{\infty} C_{k+2}^2 (1-\theta)^k = 2\theta^2 (1 - (1-\theta))^{-3} = \frac{2}{\theta}$$

因此解得

$$\hat{\theta} = \frac{2}{\bar{\xi}}.$$

① 极大似然估计

② 矩估计

③ 估计的无偏性

- 定义 3.1. 若统计量 $T = T(X_1, \dots, X_n)$ 满足

$$E_{\theta} T = g(\theta), \quad \forall \theta \in \Theta.$$

则称 T 为 $g(\theta)$ 的无偏估计.

- 例 3.1. 样本均值 \bar{X} 是期望 μ 的无偏估计.

$$\begin{aligned} E_{\theta} \bar{X} &= E_{\theta} \frac{1}{n} (X_1 + \dots + X_n) \\ &= \frac{1}{n} (E_{\theta} X_1 + \dots + E_{\theta} X_n) = \mu. \end{aligned}$$

例 3.1(续). 方差 σ^2 的无偏估计.

- 样本方差 $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 不是 σ^2 的无偏估计.
- $x_i - \mu = (x_i - \bar{x}) + (\bar{x} - \mu)$:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + (\bar{x} - \mu)^2.$$

- $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2.$
- $E_{\theta} \widehat{\sigma^2} = \text{var}(X) - \text{var}(\bar{X}) = \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2.$
- 定理 3.1. 若总体方差 σ^2 存在, 则 S^2 是 σ^2 的无偏估计, 其中,

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

例 3.2. 总体: $X \sim \text{Exp}(\lambda)$. 样本量: n . 寻找 λ 的无偏估计.

- 最大似然估计 & 矩估计: $\hat{\lambda} = 1/\bar{X} = \frac{n}{S_n}$, 其中,

$$S_n = X_1 + \cdots + X_n \sim \Gamma(n, \lambda), \quad p_{S_n}(y) = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}, y > 0.$$

- 于是,

$$\begin{aligned} E\hat{\lambda} &= E\frac{n}{S_n} = n \int_0^\infty \frac{1}{y} \cdot \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy \\ &= n \frac{\lambda \Gamma(n-1)}{\Gamma(n)} \int_0^\infty \frac{\lambda^{n-1}}{\Gamma(n-1)} y^{n-2} e^{-\lambda y} dy = \frac{n}{n-1} \lambda \end{aligned}$$

- $n \geq 2$ 时, $\frac{n-1}{n} \hat{\lambda}$ 为 λ 的无偏估计.
- $n = 1$ 时, $E\hat{\lambda} = \int_0^\infty \frac{1}{x} \times \lambda e^{-\lambda x} dx = \infty$.

例：设总体 X 服从参数为 λ 的泊松分布。现在求 $\exp\{-2\lambda\}$ 的估计。

解： $E(X) = \lambda$ 知 X 是 λ 的一个无偏估计。显然 $g_1(X)$ 是一个可能的估计，但不是无偏估计 ($E[g_1(x)] = e^{-\lambda(1-e^{-2})} > e^{-2\lambda}$)。因此考虑另一个估计 $g_2(X)$ ，令

$$g_2(x) = \begin{cases} 1, & x \text{ 是偶数,} \\ -1, & x \text{ 是奇数,} \end{cases}$$

此时此时

$$E(g_2(X)) = \sum_{k=0}^{\infty} (-1)^k P(X = k) = \exp\{-\lambda\} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} = \exp\{-2\lambda\}.$$

确实是一个无偏估计，但是这个估计是荒谬的，当 X 为奇数时，它取负值。因此，在样本量很小的时候不能片面追求无偏性。

例：设总体 $\xi \sim N(\mu, \sigma^2)$ ，求 c 使得 $c \sum_{i=1}^n (\xi_{i+1} - \xi_i)^2$ 为 σ^2 的无偏估计。

例：设总体 $\xi \sim N(\mu, \sigma^2)$ ，求 c 使得 $c \sum_{i=1}^n (\xi_{i+1} - \xi_i)^2$ 为 σ^2 的无偏估计。

解：

$$E\left(\sum_{i=1}^n (\xi_{i+1} - \xi_i)^2\right) = \sum_{i=1}^n E(\xi_{i+1}^2 - 2\xi_{i+1}\xi_i + \xi_i^2) = \sum_{i=1}^n 2\sigma^2 = 2n\sigma^2$$

因此可以取 $c = \frac{1}{2n}$ 。

例：设 $\hat{\theta}$ 为参数 θ 的无偏估计，且有 $\text{var}(\hat{\theta}) > 0$ ，试证明 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计。

例：设 $\hat{\theta}$ 为参数 θ 的无偏估计，且有 $\text{var}(\hat{\theta}) > 0$ ，试证明 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计。

证明：由方差的定义可知，

$$\text{var}(\hat{\theta}) = E(\hat{\theta}^2) - (E(\hat{\theta}))^2 > 0$$

由于 $\hat{\theta}$ 为参数 θ 的无偏估计，即 $E(\hat{\theta}) = \theta$ ，因此

$$E(\hat{\theta}^2) = \text{var}(\hat{\theta}) + (E(\hat{\theta}))^2 = \text{var}(\hat{\theta}) + \theta^2 > \theta^2,$$

所以 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计。

例：设 x_1, \dots, x_n 是来自 $N(\theta, 1)$ 的样本，证明 $g(\theta) = |\theta|$ 没有无偏估计。

证明：假设 $T(x_1, \dots, x_n)$ 为 $g(\theta)$ 的无偏估计，则

$$\left(\frac{1}{\sqrt{2\pi}} \right) \int_{-\infty}^{+\infty} T(x_1, \dots, x_n) \exp \left\{ -\sum_{i=1}^n \frac{(x_i - \theta)^2}{2} \right\} dx_1 dx_2 \cdots dx_n = |\theta|.$$

由上式可知，等式左侧关于 θ 处处可导，而等式右侧在 $\theta = 0$ 不存在导数，因此假设不成立。即 $g(\theta) = |\theta|$ 没有无偏估计。

例：设从均值 μ ，方差为 $\sigma^2 > 0$ 的总体中分别抽取容量为 n_1 和 n_2 的两独立样本， \bar{x}_1 和 \bar{x}_2 分别是这两个样本的均值。试证明，对于任意常数 a, b ($a + b = 1$)， $Y = a\bar{x}_1 + b\bar{x}_2$ 都是 μ 的无偏估计，并确定常数 a, b 使得 $\text{var}(Y)$ 达到最小。

例：设从均值 μ ，方差为 $\sigma^2 > 0$ 的总体中分别抽取容量为 n_1 和 n_2 的两独立样本， \bar{x}_1 和 \bar{x}_2 分别是这两个样本的均值。试证明，对于任意常数 a, b ($a + b = 1$)， $Y = a\bar{x}_1 + b\bar{x}_2$ 都是 μ 的无偏估计，并确定常数 a, b 使得 $\text{var}(Y)$ 达到最小。

证明：由于 \bar{x}_1 和 \bar{x}_2 分别是容量为 n_1 和 n_2 的两独立样本的均值，故

$$E(\bar{x}_1) = \mu, \quad E(\bar{x}_2) = \mu, \quad \text{var}(\bar{x}_1) = \frac{\sigma^2}{n_1}, \quad \text{var}(\bar{x}_2) = \frac{\sigma^2}{n_2}.$$

因而

$$E(Y) = E(a\bar{x}_1 + b\bar{x}_2) = aE(\bar{x}_1) + bE(\bar{x}_2) = a\mu + b\mu = (a + b)\mu = \mu$$

因此 $Y = a\bar{x}_1 + b\bar{x}_2$ 是 μ 的无偏估计。

又由 $a + b = 1$ 知 $Y = a\bar{x}_1 + b\bar{x}_2 = a\bar{x}_1 + (1 - a)\bar{x}_2$, 从而

$$\text{var}(Y) = \frac{a^2\sigma^2}{n_1} + \frac{(1-a)^2\sigma^2}{n_2} = \sigma^2 \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) a^2 - \frac{2}{n_2}a + \frac{1}{n_2} \right],$$

求导知, 当 $a = \frac{1/n_2}{1/n_1 + 1/n_2} = \frac{n_1}{n_1 + n_2}$ 时, $\text{var}(Y)$ 达到最小, 此时 $b = \frac{n_2}{n_1 + n_2}$ 。

结果表明, 来自同一总体的容量为 n_1 和 n_2 的两独立样本的合样本的均值 $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$ 是线性无偏估计类 $U = \{a\bar{x}_1 + (1 - a)\bar{x}_2\}$ 中方差最小的。