概率论与数理统计（10）

清华大学

2020 年春季学期

最大似然估计

- 设总体的概率函数为 $p(x;\theta)$, θ 是一个未知参数或者未知参数组成的参数向量。$x_1,\ldots,x_n$ 是来自该总体的样本，将样本的联合密度函数看作 θ 的函数，则

$$L(\theta)=L(\theta;x_1,\ldots,x_n):=\prod_{i=1}^{n}p(x_i:\theta),$$

称为样本的似然函数。如果某统计量 $\hat{\theta}=\hat{\theta}(x_1,\ldots,x_n)$ 满足

$$L(\hat{\theta})=\max_{\theta}L(\theta)$$

则称 $\hat{\theta}$ 是 θ 的最大似然估计, 简记为 MLE(maximum likelihood estimator).

最大似然估计

- 由于 $\ln x$ 是 x 的单调函数，$\ln L(\theta)$ 到达最大值和 $L(\theta)$ 到达最大值是等价的。

最大似然估计

- 由于 $\ln x$ 是 x 的单调函数，$\ln L(\theta)$ 到达最大值和 $L(\theta)$ 到达最大值是等价的。
- 当 $\ln L(\theta)$ 是可微函数时，一般通过求导的方式寻找 $\hat{\theta}$.

最大似然估计

- 由于 $\ln x$ 是 x 的单调函数，$\ln L(\theta)$ 到达最大值和 $L(\theta)$ 到达最大值是等价的。
- 当 $\ln L(\theta)$ 是可微函数时，一般通过求导的方式寻找 $\hat{\theta}$.
- 总体是二点分布 $b(1,p)$, $x_1,\ldots,x_n$ 为其样本，则

$$L(p;x_1,\ldots,x_n)=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

最大似然估计

- 由于 $\ln x$ 是 x 的单调函数，$\ln L(\theta)$ 到达最大值和 $L(\theta)$ 到达最大值是等价的。
- 当 $\ln L(\theta)$ 是可微函数时，一般通过求导的方式寻找 $\hat{\theta}$.
- 总体是二点分布 $b(1,p)$, $x_1,\ldots,x_n$ 为其样本，则

$$L(p;x_1,\ldots,x_n)=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

$$\ln L(p)=(\sum_{i=1}^n x_i)\ln p+(n-\sum_{i=1}^n x_i)\ln(1-p).$$

$$\partial_p \ln L(p)=\frac{\sum_{i=1}^n x_i}{p}+\frac{n-\sum_{i=1}^n x_i}{1-p}=0.$$

$$p=\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}.\quad \partial_p^2 \ln L(p)=-\frac{n^2}{\bar{x}}-\frac{n^2}{1-\bar{x}}<0$$

最大似然估计

- 总体是指数分布 $Exp(\lambda)$: $p(x,\lambda)=\lambda e^{-\lambda x}$.

最大似然估计

- 总体是指数分布 $Exp(\lambda)$: $p(x,\lambda) = \lambda e^{-\lambda x}$.
- 似然函数为

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

最大似然估计

- 总体是指数分布 $Exp(\lambda)$: $p(x,\lambda)=\lambda e^{-\lambda x}$.
- 似然函数为

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

$$\frac{n}{\lambda}-\sum_{i=1}^{n}x_i=0.$$

最大似然估计

- 总体是指数分布 $Exp(\lambda)$: $p(x,\lambda)=\lambda e^{-\lambda x}$.
- 似然函数为

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

$$\frac{n}{\lambda}-\sum_{i=1}^{n}x_i=0.$$

$$\hat{\lambda}=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}x_i}=\frac{1}{\bar{x}}.$$

$$\partial_\lambda^2 L(\lambda)=-\frac{n}{\lambda^2}<0.$$

最大似然估计

- 总体是正态分布 $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$:

最大似然估计

- 总体是正态分布 $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$:

$$\begin{aligned}\ln L(\mu, \sigma^2) &= \ln\left(\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)\right) \\ &= -\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} - \frac{n\ln\sigma^2}{2} - \frac{n\ln 2\pi}{2}.\end{aligned}$$

- 总体是正态分布 $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$:

$$\ln L(\mu, \sigma^2) = \ln(\prod_{i=1}^{n}(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}))$$
$$= -\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} - \frac{n\ln\sigma^2}{2} - \frac{n\ln 2\pi}{2}.$$

$$\begin{cases} \partial_\mu \ln L(\mu, \sigma^2) = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} = 0, \\ \partial_{\sigma^2} \ln L(\mu, \sigma^2) = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0. \end{cases}$$

- 总体是正态分布 $N(\mu,\sigma^2)$, $\theta=(\mu,\sigma^2)$:

$$\begin{aligned}\ln L(\mu,\sigma^2) &= \ln\left(\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)\right)\\ &= -\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}-\frac{n\ln\sigma^2}{2}-\frac{n\ln 2\pi}{2}.\end{aligned}$$

$$\begin{cases}\partial_\mu \ln L(\mu,\sigma^2)=\frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2}=0,\\ \partial_{\sigma^2}\ln L(\mu,\sigma^2)=\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^4}-\frac{n}{2\sigma^2}=0.\end{cases}$$

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

最大似然估计

- 总体是正态分布 $N(\mu,\sigma^2)$, $\theta=(\mu,\sigma^2)$:

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

$$\begin{cases} \partial_\mu \ln L(\mu,\sigma^2) = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} = 0, \\ \partial_{\sigma^2} \ln L(\mu,\sigma^2) = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0. \end{cases}$$

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

- 最大似然估计不一定是无偏估计。

最大似然估计

- 设总体的概率密度函数 $p(x;\theta) = \theta c^{\theta} x^{-(\theta+1)}$, $x > c$, $c > 0$ 为已知，$\theta > 1$ 为未知参数，$x_1, \ldots, x_n$ 为样本，求 θ 的最大似然估计。

最大似然估计

- 设总体的概率密度函数 $p(x;\theta) = \theta c^{\theta} x^{-(\theta+1)}$, $x > c$, $c > 0$ 为已知，$\theta > 1$ 为未知参数，$x_1, \ldots, x_n$ 为样本，求 θ 的最大似然估计。
- 对数似然函数为

$$L(\theta) = \ln \theta^n c^{n\theta} (\prod_{i}^{n} x_i)^{-(\theta+1)} = n \ln \theta + n\theta \ln c - (\theta+1)(\sum_{i=1}^{n} \ln x_i).$$

最大似然估计

- 设总体的概率密度函数 $p(x;\theta) = \theta c^\theta x^{-(\theta+1)}$, $x > c$, $c > 0$ 为已知，$\theta > 1$ 为未知参数，$x_1, \ldots, x_n$ 为样本，求 θ 的最大似然估计。
- 对数似然函数为

$$L(\theta) = \ln \theta^n c^{n\theta} (\prod_i^n x_i)^{-(\theta+1)} = n \ln \theta + n\theta \ln c - (\theta+1)(\sum_{i=1}^n \ln x_i).$$

$$\partial_\theta L(\theta) = \frac{n}{\theta} + n \ln c - \sum_{i=1}^n \ln x_i = 0$$

最大似然估计

- 设总体的概率密度函数 $p(x;\theta) = \theta c^\theta x^{-(\theta+1)}$, $x > c$, $c > 0$ 为已知，$\theta > 1$ 为未知参数，$x_1, \ldots, x_n$ 为样本，求 θ 的最大似然估计。
- 对数似然函数为

$$L(\theta) = \ln \theta^n c^{n\theta} (\prod_i^n x_i)^{-(\theta+1)} = n \ln \theta + n\theta \ln c - (\theta+1)(\sum_{i=1}^n \ln x_i).$$

$$\partial_\theta L(\theta) = \frac{n}{\theta} + n \ln c - \sum_{i=1}^n \ln x_i = 0$$

$$\hat{\theta} = (\frac{1}{n} \sum_{i=1}^n \ln x_i - \ln c)^{-1}.$$

最大似然估计

- 设总体的概率密度函数 $p(x;\theta)=\theta c^{\theta}x^{-(\theta+1)}$, $x>c$, $c>0$ 为已知，$\theta>1$ 为未知参数，$x_1,\ldots,x_n$ 为样本，求 θ 的最大似然估计。
- 对数似然函数为

$$L(\theta)=\ln\theta^{n}c^{n\theta}(\prod_{i}^{n}x_i)^{-(\theta+1)}=n\ln\theta+n\theta\ln c-(\theta+1)(\sum_{i=1}^{n}\ln x_i).$$

$$\partial_\theta L(\theta)=\frac{n}{\theta}+n\ln c-\sum_{i=1}^{n}\ln x_i=0$$

$$\hat{\theta}=(\frac{1}{n}\sum_{i=1}^{n}\ln x_i-\ln c)^{-1}.$$

$$\partial_\theta^2 L(\theta)=-\frac{n}{\theta^2}<0.$$

最大似然估计

- 总体是均匀分布 $U(0,\theta)$, 似然函数为

最大似然估计

- 总体是均匀分布 $U(0,\theta)$, 似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n}I_{\{0<x_i\leqslant\theta\}}=\frac{1}{\theta^n}I_{\{x_{(n)}\leqslant\theta\}}.$$

最大似然估计

- 总体是均匀分布 $U(0,\theta)$, 似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n}I_{\{0<x_i\leqslant\theta\}}=\frac{1}{\theta^n}I_{\{x_{(n)}\leqslant\theta\}}.$$

- 并不是处处可导的函数。

- 总体是均匀分布 $U(0,\theta)$，似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n}I_{\{0<x_i\leqslant\theta\}}=\frac{1}{\theta^n}I_{\{x_{(n)}\leqslant\theta\}}.$$

- 并不是处处可导的函数。
- $\theta>0$ 时，$\frac{1}{\theta^n}$ 是关于 θ 单调递减，所以 θ 越小，似然函数越大。

最大似然估计

- 总体是均匀分布 $U(0,\theta)$, 似然函数为

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{\{0<x_i\leqslant\theta\}} = \frac{1}{\theta^n} I_{\{x_{(n)}\leqslant\theta\}}.$$

- 并不是处处可导的函数。
- $\theta > 0$ 时，$\frac{1}{\theta^n}$ 是关于 θ 单调递减，所以 θ 越小，似然函数越大。
- 示性函数 $I_{\{x_{(n)}\leqslant\theta\}}$：如果 θ 小于 $x_{(n)}$，则等于 0，所以 θ 不能小于 $x_{(n)}$。

- 总体是均匀分布 $U(0,\theta)$, 似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n} I_{\{0<x_i\leqslant\theta\}}=\frac{1}{\theta^n}I_{\{x_{(n)}\leqslant\theta\}}.$$

- 并不是处处可导的函数。
- $\theta>0$ 时，$\frac{1}{\theta^n}$ 是关于 θ 单调递减，所以 θ 越小，似然函数越大。
- 示性函数 $I_{\{x_{(n)}\leqslant\theta\}}$：如果 θ 小于 $x_{(n)}$，则等于 0，所以 θ 不能小于 $x_{(n)}$。
- 所以最大似然估计为 $\hat{\theta}=x_{(n)}$.

最大似然估计

- 总体为 $U(\theta,(k+1)\theta)$，$\theta>0$，求 θ 的最大似然估计。

最大似然估计

- 总体为 $U(\theta, (k+1)\theta)$，$\theta > 0$, 求 θ 的最大似然估计。
- 似然函数 $L(\theta) = \frac{1}{(k\theta)^n} I_{\theta \leqslant x_{(1)} \leqslant x_{(n)} \theta (k+1)\theta}$.
- 让似然函数大，θ 应该尽可能的小。由 $\theta \leqslant x_{(1)} \leqslant x_{(n)} \theta (k+1)\theta$, 有 $\frac{x_{(n)}}{k+1} \leqslant \theta \leqslant x_{(1)}$.

最大似然估计

- 总体为 $U(\theta,(k+1)\theta)$，$\theta>0$，求 θ 的最大似然估计。
- 似然函数 $L(\theta)=\frac{1}{(k\theta)^n}I_{\theta\leqslant x_{(1)}\leqslant x_{(n)}\theta(k+1)\theta}$.
- 让似然函数大，θ 应该尽可能的小。由 $\theta\leqslant x_{(1)}\leqslant x_{(n)}\theta(k+1)\theta$，有 $\frac{x_{(n)}}{k+1}\leqslant\theta\leqslant x_{(1)}$.
- 所以最大似然估计为 $\frac{x_{(n)}}{k+1}$.

- 总体为 $U(\theta,(k+1)\theta)$，$\theta>0$，求 θ 的最大似然估计。
- 似然函数 $L(\theta)=\frac{1}{(k\theta)^n}I_{\theta\leqslant x_{(1)}\leqslant x_{(n)}\theta(k+1)\theta}$.
- 让似然函数大，θ 应该尽可能的小。由 $\theta\leqslant x_{(1)}\leqslant x_{(n)}\theta(k+1)\theta$，有 $\frac{x_{(n)}}{k+1}\leqslant\theta\leqslant x_{(1)}$.
- 所以最大似然估计为 $\frac{x_{(n)}}{k+1}$.
- 若总体 X 的密度函数如下：

$$f(x;\theta)=\begin{cases}1, & \text{if } \theta=0, 0<x<1;\\ \frac{1}{2\sqrt{x}}, & \text{if } \theta=1, 0<x<1;\\ 0, & else.\end{cases}$$

求 θ 的最大似然估计。

最大似然估计的函数不变性

- 函数不变性：如果 $\hat{\theta}$ 是 θ 的最大似然估计，则对任一函数 $g(\theta)$，其最大似然估计为 $g(\hat{\theta})$.

最大似然估计的函数不变性

- 函数不变性：如果 $\hat{\theta}$ 是 θ 的最大似然估计，则对任一函数 $g(\theta)$，其最大似然估计为 $g(\hat{\theta})$.
  $\theta = \hat{\theta}(x_1, \ldots, x_n)$ 时似然函数 $L(x_1, \ldots, x_n; \theta)$ 取到最大，也就是说 $g(\theta) = g(\hat{\theta})$ 时似然函数
  $\bar{L}(x_1, \ldots, x_n; g(\theta)) = L(x_1, \ldots, x_n; \theta)$ 时取到最大。

最大似然估计的函数不变性

- 函数不变性：如果 $\hat{\theta}$ 是 θ 的最大似然估计，则对任一函数 $g(\theta)$，其最大似然估计为 $g(\hat{\theta})$.
  $\theta = \hat{\theta}(x_1, \ldots, x_n)$ 时似然函数 $L(x_1, \ldots, x_n; \theta)$ 取到最大，也就是说 $g(\theta) = g(\hat{\theta})$ 时似然函数
  $\bar{L}(x_1, \ldots, x_n; g(\theta)) = L(x_1, \ldots, x_n; \theta)$ 时取到最大。
- 总体为 $N(\mu, \sigma^2)$，$\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s_n^2$ 分别为 μ 和 σ^2 的最大似然估计。

- 函数不变性：如果 $\hat{\theta}$ 是 θ 的最大似然估计，则对任一函数 $g(\theta)$，其最大似然估计为 $g(\hat{\theta})$.
  $\theta = \hat{\theta}(x_1, \ldots, x_n)$ 时似然函数 $L(x_1, \ldots, x_n; \theta)$ 取到最大，也就是说 $g(\theta) = g(\hat{\theta})$ 时似然函数
  $\bar{L}(x_1, \ldots, x_n; g(\theta)) = L(x_1, \ldots, x_n; \theta)$ 时取到最大。
- 总体为 $N(\mu, \sigma^2)$，$\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s_n^2$ 分别为 μ 和 σ^2 的最大似然估计。则
  - 标准差 σ 的最大似然估计为 $\hat{\sigma} = s_n$.
  - 概率 $P(X < 3) = \Phi(\frac{3-\mu}{\sigma})$ 的最大似然估计为 $\Phi(\frac{3-\bar{x}}{s_n})$.
  - 总体的 0.9 分位数位 $x_{0.9} = \mu + \sigma u_{0.9}$，它的最大似然估计为 $\bar{x} + u_{0.9} s_n$.

最大似然估计的渐进正态性

- 参数 θ 的相合估计 $\hat{\theta}_n$ 称为渐进正态的，如果存在趋向于 0 的非负常数序列 $\sigma_n(\theta)$，使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布。也称 $\hat{\theta}_n$ 服从渐进正态分布 $N(\theta, \sigma_n^2(\theta))$，记为 $\hat{\theta}_n \sim AN(\theta, \sigma_n^2(\theta))$. $\sigma_n^2(\theta)$ 称为 $\hat{\theta}_n$ 的渐进方差。

最大似然估计的渐进正态性

- 参数 θ 的相合估计 $\hat{\theta}_n$ 称为渐进正态的，如果存在趋向于 0 的非负常数序列 $\sigma_n(\theta)$，使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布。也称 $\hat{\theta}_n$ 服从渐进正态分布 $N(\theta,\sigma_n^2(\theta))$，记为 $\hat{\theta}_n \sim AN(\theta,\sigma_n^2(\theta))$. $\sigma_n^2(\theta)$ 称为 $\hat{\theta}_n$ 的渐进方差。
- 定理：若总体有密度函数 $p(x;\theta)$，$\partial_\theta^i \ln p(x;\theta)$，$i=1,2,3$ 均存在（关于未知参数三阶可导），且满足

$$\sup_\theta\{\int_{-\infty}^{\infty}(|\partial_\theta p|+|\partial_\theta^2 p|)\,dx,\quad E|\partial_\theta^3 \ln p|\}<\infty.$$

- 参数 θ 的相合估计 $\hat{\theta}_n$ 称为渐进正态的，如果存在趋向于 0 的非负常数序列 $\sigma_n(\theta)$，使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布。也称 $\hat{\theta}_n$ 服从渐进正态分布 $N(\theta,\sigma_n^2(\theta))$，记为 $\hat{\theta}_n \sim AN(\theta,\sigma_n^2(\theta))$. $\sigma_n^2(\theta)$ 称为 $\hat{\theta}_n$ 的渐进方差。
- 定理：若总体有密度函数 $p(x;\theta)$, $\partial_\theta^i \ln p(x;\theta)$, $i=1,2,3$ 均存在 (关于未知参数三阶可导), 且满足

$$\sup_\theta\{\int_{-\infty}^{\infty}(|\partial_\theta p|+|\partial_\theta^2 p|)\,dx, \quad E|\partial_\theta^3 \ln p|\}<\infty.$$

同时 $0<I(\theta):=\int_{-\infty}^{\infty}(\partial_\theta \ln p)^2 p(x;\theta)\,dx<\infty.$

最大似然估计的渐进正态性

- 参数 θ 的相合估计 $\hat{\theta}_n$ 称为渐进正态的，如果存在趋向于 0 的非负常数序列 $\sigma_n(\theta)$，使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布。也称 $\hat{\theta}_n$ 服从渐进正态分布 $N(\theta,\sigma_n^2(\theta))$，记为 $\hat{\theta}_n \sim AN(\theta,\sigma_n^2(\theta))$. $\sigma_n^2(\theta)$ 称为 $\hat{\theta}_n$ 的渐进方差。
- 定理：若总体有密度函数 $p(x;\theta)$，$\partial_\theta^i \ln p(x;\theta)$，$i=1,2,3$ 均存在 (关于未知参数三阶可导)，且满足

$$\sup_\theta\{\int_{-\infty}^{\infty}(|\partial_\theta p|+|\partial_\theta^2 p|)\,dx, \quad E|\partial_\theta^3 \ln p|\}<\infty.$$

同时 $0<I(\theta):=\int_{-\infty}^{\infty}(\partial_\theta \ln p)^2 p(x;\theta)\,dx<\infty$.则存在参数 θ 的最大似然估计

$$\hat{\theta}_n=\hat{\theta}(x_1,\ldots,x_n)\sim AN(\theta,\frac{1}{nI(\theta)}).$$

最大似然估计的渐进正态性

- 参数 θ 的相合估计 $\hat{\theta}_n$ 称为渐进正态的，如果存在趋向于 0 的非负常数序列 $\sigma_n(\theta)$，使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布。也称 $\hat{\theta}_n$ 服从渐进正态分布 $N(\theta,\sigma_n^2(\theta))$，记为 $\hat{\theta}_n \sim AN(\theta,\sigma_n^2(\theta))$. $\sigma_n^2(\theta)$ 称为 $\hat{\theta}_n$ 的渐进方差。
- 定理：若总体有密度函数 $p(x;\theta)$，$\partial_\theta^i \ln p(x;\theta)$，$i=1,2,3$ 均存在 (关于未知参数三阶可导)，且满足

$$\sup_\theta\{\int_{-\infty}^{\infty}(|\partial_\theta p|+|\partial_\theta^2 p|)\,dx, \quad E|\partial_\theta^3 \ln p|\}<\infty.$$

同时 $0<I(\theta):=\int_{-\infty}^{\infty}(\partial_\theta \ln p)^2 p(x;\theta)\,dx<\infty$.则存在参数 θ 的最大似然估计

$$\hat{\theta}_n=\hat{\theta}(x_1,\ldots,x_n)\sim AN(\theta,\frac{1}{nI(\theta)}).$$

$I(\theta)=E\Big[(\partial_\theta \ln p(x;\theta))^2\Big]$ 为费希尔信息量或 Fisher 熵。

费希尔信息量

- 总体的概率函数 $p(x;\theta)$, 且 $\partial_\theta^2 p(x,\theta)$ 存在，则

$$I(\theta) = E[(\partial_\theta \ln p(x:\theta))^2] = -E[\partial_\theta^2 \ln p(x;\theta)].$$

费希尔信息量

- 总体的概率函数 $p(x;\theta)$，且 $\partial_\theta^2 p(x,\theta)$ 存在，则

$$I(\theta) = E[(\partial_\theta \ln p(x:\theta))^2] = -E[\partial_\theta^2 \ln p(x;\theta)].$$

- 令 $S_\theta = \partial_\theta \ln p(x;\theta)$，则

$$E(S_\theta) = \int \frac{1}{p(x;\theta)} \partial_\theta p(x;\theta) p(x:\theta)\, dx = \partial_\theta \int p(x;\theta) = 0$$

费希尔信息量

- 总体的概率函数 $p(x;\theta)$，且 $\partial_\theta^2 p(x,\theta)$ 存在，则

$$I(\theta) = E[(\partial_\theta \ln p(x:\theta))^2] = -E[\partial_\theta^2 \ln p(x;\theta)].$$

- 令 $S_\theta = \partial_\theta \ln p(x;\theta)$，则

$$E(S_\theta) = \int \frac{1}{p(x;\theta)} \partial_\theta p(x;\theta) p(x:\theta)\, dx = \partial_\theta \int p(x;\theta) = 0$$

$$\begin{aligned} 0 = \partial_\theta E(S_\theta) &= \int [\partial_\theta S_\theta p(x;\theta) + S_\theta \partial_\theta p(x;\theta)]\, dx \\ &= \int [\partial_\theta^2 \ln p(x;\theta) p(x;\theta) + (\partial_\theta \ln p(x;\theta))^2 p(x;\theta)]\, dx \\ &= E[\partial_\theta^2 \ln p(x;\theta)] + E(\partial_\theta \ln p(x;\theta))^2. \end{aligned}$$

最大似然估计的渐近正态性：大概的证明

1. $l(\theta) = \sum_{i=1}^{n} \log p(x_i; \theta)$. $\hat{\theta}$ 是最大似然估计，则 $l'(\hat{\theta}) = 0$.

最大似然估计的渐近正态性：大概的证明

1. $l(\theta) = \sum_{i=1}^{n} \log p(x_i;\theta)$. $\hat{\theta}$ 是最大似然估计，则 $l'(\hat{\theta}) = 0$.
2. 泰勒展开: $l'(\hat{\theta}) = l'(\theta) + l''(\theta)(\hat{\theta} - \theta) + \frac{l'''(\theta)}{2}(\hat{\theta} - \theta)^2 + \ldots$

最大似然估计的渐近正态性：大概的证明

1. $l(\theta) = \sum_{i=1}^{n} \log p(x_i;\theta)$. $\hat{\theta}$ 是最大似然估计，则 $l'(\hat{\theta}) = 0$.
2. 泰勒展开: $l'(\hat{\theta}) = l'(\theta) + l''(\theta)(\hat{\theta}-\theta) + \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 + \ldots$
3. $-l''(\theta)(\hat{\theta}-\theta) - \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 = l'(\theta) \Rightarrow (\hat{\theta}-\theta) = \frac{l'(\theta)}{-l''(\theta)-\frac{l'''(\theta)}{2}(\hat{\theta}-\theta)}$.

最大似然估计的渐近正态性：大概的证明

1. $l(\theta)=\sum_{i=1}^n \log p(x_i;\theta)$. $\hat\theta$ 是最大似然估计，则 $l'(\hat\theta)=0$.
2. 泰勒展开: $l'(\hat\theta)=l'(\theta)+l''(\theta)(\hat\theta-\theta)+\frac{l'''(\theta)}{2}(\hat\theta-\theta)^2+\ldots$
3. $-l''(\theta)(\hat\theta-\theta)-\frac{l'''(\theta)}{2}(\hat\theta-\theta)^2=l'(\theta)\Rightarrow(\hat\theta-\theta)=\frac{l'(\theta)}{-l''(\theta)-\frac{l'''(\theta)}{2}(\hat\theta-\theta)}$.
4. 两边都乘上 $\sqrt{n}$ 有： $\sqrt{n}(\hat\theta-\theta)=\frac{\frac{1}{\sqrt{n}}l'(\theta)}{-\frac{1}{n}l''(\theta)-\frac{l'''(\theta)}{2n}(\hat\theta-\theta)}$.
5. 由中心极限定理
   $\frac{1}{\sqrt{n}}l'(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\partial_\theta\log(p(x_i;\theta))\to N(0,I(\theta))$

最大似然估计的渐近正态性：大概的证明

1. $l(\theta) = \sum_{i=1}^{n} \log p(x_i;\theta)$. $\hat{\theta}$ 是最大似然估计，则 $l'(\hat{\theta}) = 0$.
2. 泰勒展开: $l'(\hat{\theta}) = l'(\theta) + l''(\theta)(\hat{\theta}-\theta) + \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 + \dots$
3. $-l''(\theta)(\hat{\theta}-\theta) - \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 = l'(\theta) \Rightarrow (\hat{\theta}-\theta) = \frac{l'(\theta)}{-l''(\theta) - \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)}$.
4. 两边都乘上 $\sqrt{n}$ 有：$\sqrt{n}(\hat{\theta}-\theta) = \frac{\frac{1}{\sqrt{n}}l'(\theta)}{-\frac{1}{n}l''(\theta) - \frac{l'''(\theta)}{2n}(\hat{\theta}-\theta)}$.
5. 由中心极限定理
$\frac{1}{\sqrt{n}}l'(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \partial_\theta \log(p(x_i;\theta)) \to N(0, I(\theta))$
6. 由大数定律 $-\frac{l''(\theta)}{n} \to I(\theta)$，$-\frac{l'''(\theta)}{2n}(\hat{\theta}-\theta) \to 0$.

最大似然估计的渐近正态性：大概的证明

1. $l(\theta) = \sum_{i=1}^n \log p(x_i;\theta)$. $\hat{\theta}$ 是最大似然估计，则 $l'(\hat{\theta}) = 0$.
2. 泰勒展开: $l'(\hat{\theta}) = l'(\theta) + l''(\theta)(\hat{\theta}-\theta) + \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 + \dots$
3. $-l''(\theta)(\hat{\theta}-\theta) - \frac{l'''(\theta)}{2}(\hat{\theta}-\theta)^2 = l'(\theta) \Rightarrow (\hat{\theta}-\theta) = \frac{l'(\theta)}{-l''(\theta)-\frac{l'''(\theta)}{2}(\hat{\theta}-\theta)}$.
4. 两边都乘上 $\sqrt{n}$ 有：$\sqrt{n}(\hat{\theta}-\theta) = \frac{\frac{1}{\sqrt{n}}l'(\theta)}{-\frac{1}{n}l''(\theta)-\frac{l'''(\theta)}{2n}(\hat{\theta}-\theta)}$.
5. 由中心极限定理
   $\frac{1}{\sqrt{n}}l'(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \partial_\theta \log(p(x_i;\theta)) \to N(0, I(\theta))$
6. 由大数定律 $-\frac{l''(\theta)}{n} \to I(\theta)$，$-\frac{l'''(\theta)}{2n}(\hat{\theta}-\theta) \to 0$.
7. 所以 $\sqrt{n}(\hat{\theta}-\theta) \to N(0, \frac{1}{I(\theta)})$.

最大似然估计的渐进正态性

- 总体为 $N(\mu, \sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu} = \bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。

最大似然估计的渐进正态性

- 总体为 $N(\mu, \sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu} = \bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu (\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

最大似然估计的渐进正态性

- 总体为 $N(\mu, \sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu} = \bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu (\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

$$I(\mu) = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^2}$$

最大似然估计的渐进正态性

- 总体为 $N(\mu, \sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu} = \bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu (\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

$$I(\mu) = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^2} \Rightarrow \hat{\mu} \sim AN(\mu, \sigma^2/n).$$

最大似然估计的渐进正态性

- 总体为 $N(\mu, \sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu} = \bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu (\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

$$I(\mu) = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^2} \Rightarrow \hat{\mu} \sim AN(\mu, \sigma^2/n).$$

- 在 μ 已知时，σ 的 MLE 为 $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2} \sim AN(\sigma, \frac{16\sigma^2}{27n})$。

最大似然估计的渐进正态性

- 总体为 $N(\mu,\sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu}=\bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu(\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

$$I(\mu) = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^2} \Rightarrow \hat{\mu} \sim AN(\mu,\sigma^2/n).$$

- 在 μ 已知时，σ 的 MLE 为 $\hat{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^n(x_i-\mu)^2} \sim AN(\sigma,\frac{16\sigma^2}{27n})$。

$$\partial_\sigma \ln p = \frac{1}{\sigma}+\frac{(x-\mu)^2}{4\sigma^3}$$

最大似然估计的渐进正态性

- 总体为 $N(\mu,\sigma^2)$,
- 当 σ^2 已知时，μ 的最大似然估计为 $\hat{\mu}=\bar{x}$, 则 $\hat{\mu}$ 服从渐进正态分布。关于 μ 的费希尔信息量为 $I(\mu)$,

$$\partial_\mu \ln p(x) = \partial_\mu(\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}) = \frac{x-\mu}{\sigma^2}.$$

$$I(\mu) = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^2} \Rightarrow \hat{\mu} \sim AN(\mu,\sigma^2/n).$$

- 在 μ 已知时，σ 的 MLE 为 $\hat{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^n(x_i-\mu)^2} \sim AN(\sigma,\frac{16\sigma^2}{27n})$。

$$\partial_\sigma \ln p = \frac{1}{\sigma}+\frac{(x-\mu)^2}{4\sigma^3} \Rightarrow I(\sigma) = E\Big[\frac{1}{16\sigma^2}(4+(\frac{x-\mu}{\sigma})^2)^2\Big] = \frac{27}{16\sigma^2}$$

最大似然估计的渐进正态性

- 总体为参数为 $\lambda > 0$ 的泊松分布：$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \ldots$.

- 总体为参数为 $\lambda > 0$ 的泊松分布：
  $P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, \ldots$. 似然函数为

$$L(\lambda) = \ln\left(\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}e^{-n\lambda}\right) = -n\lambda + \ln\lambda\sum_{i=1}^n x_i - \ln\prod_{i=1}^n x_i!.$$

最大似然估计的渐进正态性

- 总体为参数为 $\lambda > 0$ 的泊松分布：
  $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, \ldots$. 似然函数为

$$L(\lambda) = \ln\left(\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}e^{-n\lambda}\right) = -n\lambda + \ln\lambda\sum_{i=1}^n x_i - \ln\prod_{i=1}^n x_i!.$$

$$\partial_\lambda L(\lambda) = -n + \frac{1}{\lambda}n\bar{x} = 0 \Rightarrow \hat{\lambda} = \bar{x}.$$

最大似然估计的渐进正态性

- 总体为参数为 $\lambda > 0$ 的泊松分布：
  $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \ldots$. 似然函数为

$$L(\lambda) = \ln\left(\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\lambda}\right) = -n\lambda + \ln\lambda \sum_{i=1}^n x_i - \ln \prod_{i=1}^n x_i!.$$

$$\partial_\lambda L(\lambda) = -n + \frac{1}{\lambda} n\bar{x} = 0 \Rightarrow \hat{\lambda} = \bar{x}.$$

  费希尔信息量 $I(\lambda)$ 为

$$\partial_\lambda \ln p(\lambda, k) = \partial_\lambda (k \ln\lambda - \lambda - \ln k) = \frac{k}{\lambda} - 1$$

$$\Rightarrow I(\lambda) = E\frac{(k-\lambda)^2}{\lambda^2} = \frac{1}{\lambda}$$

- $\hat{\lambda} \sim AN(\lambda, \frac{\lambda}{n})$, 这也和中心极限定理相符合。

最大似然估计的渐进正态性

- 总体分布为指数分布 $Exp(\frac{1}{\lambda})$, $p(\lambda, x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} x}$.
- $\frac{1}{\theta}$ 的最大似然估计为 $\frac{1}{\bar{x}}$, 所以 θ 的为 $\bar{x}$.
- 费希尔信息量:

$$\partial_\lambda(\ln p) = \partial_\lambda(-\ln\theta - \frac{x}{\theta}) = -\frac{1}{\theta} + \frac{x}{\theta^2} = \frac{x-\theta}{\theta^2}.$$

$$I(\theta) = E(\frac{x-\theta}{\theta^2})^2 = \frac{Var(X)}{\theta^4} = \frac{1}{\theta^2}.$$

最大似然估计的渐进正态性

- 总体分布为指数分布 $Exp(\frac{1}{\lambda})$, $p(\lambda, x) = \frac{1}{\lambda}e^{-\frac{1}{\lambda}x}$.
- $\frac{1}{\theta}$ 的最大似然估计为 $\frac{1}{\bar{x}}$, 所以 θ 的为 $\bar{x}$.
- 费希尔信息量:

$$\partial_\lambda(\ln p) = \partial_\lambda(-\ln\theta - \frac{x}{\theta}) = -\frac{1}{\theta} + \frac{x}{\theta^2} = \frac{x-\theta}{\theta^2}.$$

$$I(\theta) = E(\frac{x-\theta}{\theta^2})^2 = \frac{Var(X)}{\theta^4} = \frac{1}{\theta^2}.$$

- $\hat{\theta} = \bar{x} \sim AN(\theta, \frac{\theta^2}{n})$. 和中心极限定理相符。

- 设某一个试验有三种可能的情况，其发生的概率分别为

$$p_1=\theta^2,\quad p_2=2\theta(1-\theta),\quad p_3=(1-\theta)^2$$

有 n 次试验，三种情况发生的次数分别为 n_1,n_2,n_3.

- 矩估计：有三种可能 $\theta=\sqrt{p_1}$, $\theta=1-\sqrt{p_3}$, $\theta=p_1+\frac{p_2}{2}$，即

$$\hat\theta_1=\sqrt{n_1/n},\quad \hat\theta_2=1-\sqrt{n_3/n},\quad \hat\theta_3=n_1/n+n_2/2/n.$$

- 最大似然估计：$L(\theta)=\ln(\theta^{2n_1}(2\theta(1-\theta))^{n_2}(1-\theta)^{2(n-n_1-n_2)}$.
$\partial_\theta L(\theta)=\frac{2n_1+n_2}{\theta}-\frac{2n-2n_1-n_2}{1-\theta}=0$.

- 设某一个试验有三种可能的情况，其发生的概率分别为

$$p_1 = \theta^2, \quad p_2 = 2\theta(1-\theta), \quad p_3 = (1-\theta)^2$$

有 n 次试验，三种情况发生的次数分别为 n_1, n_2, n_3.

- 矩估计：有三种可能 $\theta = \sqrt{p_1}, \theta = 1 - \sqrt{p_3}, \theta = p_1 + \frac{p_2}{2}$，即

$$\hat{\theta}_1 = \sqrt{n_1/n}, \quad \hat{\theta}_2 = 1 - \sqrt{n_3/n}, \quad \hat{\theta}_3 = n_1/n + n_2/2/n.$$

- 最大似然估计：$L(\theta) = \ln(\theta^{2n_1}(2\theta(1-\theta))^{n_2}(1-\theta)^{2(n-n_1-n_2)}$.
$\partial_\theta L(\theta) = \frac{2n_1+n_2}{\theta} - \frac{2n-2n_1-n_2}{1-\theta} = 0$. $\hat{\theta} = (2n_1+n_2)/n$.

- 设某一个试验有三种可能的情况，其发生的概率分别为

$$p_1 = \theta^2, \quad p_2 = 2\theta(1-\theta), \quad p_3 = (1-\theta)^2$$

有 n 次试验，三种情况发生的次数分别为 n_1, n_2, n_3.

- 矩估计：有三种可能 $\theta = \sqrt{p_1}$, $\theta = 1 - \sqrt{p_3}$, $\theta = p_1 + \frac{p_2}{2}$, 即

$$\hat{\theta}_1 = \sqrt{n_1/n}, \quad \hat{\theta}_2 = 1 - \sqrt{n_3/n}, \quad \hat{\theta}_3 = n_1/n + n_2/2/n.$$

- 最大似然估计：$L(\theta) = \ln(\theta^{2n_1}(2\theta(1-\theta))^{n_2}(1-\theta)^{2(n-n_1-n_2)}$.
$\partial_\theta L(\theta) = \frac{2n_1+n_2}{\theta} - \frac{2n-2n_1-n_2}{1-\theta} = 0$. $\hat{\theta} = (2n_1+n_2)/n$.

$$\hat{\theta}_1 \to N(\theta, \frac{1-\theta^2}{4n}), \ \hat{\theta}_2 \to N(\theta, \frac{1-(1-\theta)^2}{4n}), \ \hat{\theta}_3 \to N(\theta, \frac{\theta(1-\theta)}{2n})$$

均方误差

- $\hat{\theta}$ 为 θ 的一个统计估计量，则该估计量的均方误差为

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

均方误差

- $\hat{\theta}$ 为 θ 的一个统计估计量，则该估计量的均方误差为

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - E\hat{\theta}) + E\hat{\theta} - \theta]^2 \\ &= E(\hat{\theta} - E\hat{\theta})^2 + (E\hat{\theta} - \theta)^2 + 2E[(\hat{\theta} - E\hat{\theta})(E\hat{\theta} - \theta)] \\ &= Var(\hat{\theta}) + (E\hat{\theta} - \theta)^2. \end{aligned}$$

均方误差

- $\hat{\theta}$ 为 θ 的一个统计估计量，则该估计量的均方误差为

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - E\hat{\theta}) + E\hat{\theta} - \theta]^2 \\ &= E(\hat{\theta} - E\hat{\theta})^2 + (E\hat{\theta} - \theta)^2 + 2E[(\hat{\theta} - E\hat{\theta})(E\hat{\theta} - \theta)] \\ &= Var(\hat{\theta}) + (E\hat{\theta} - \theta)^2. \end{aligned}$$

- 均方误差由点估计统计量的方差和估计偏差的平方组成。一般而言，均方误差越小，估计量越好。

均方误差

- 总体为均匀分布 $U(0,\theta)$, θ 的一个无偏估计为 $\hat{\theta}=\frac{n+1}{n}x_{(n)}$, 它的均方误差为

$$MSE(\hat{\theta})=Var(\hat{\theta})=\frac{\theta^2}{n(n+2)}.$$

均方误差

- 总体为均匀分布 $U(0,\theta)$, θ 的一个无偏估计为 $\hat{\theta}=\frac{n+1}{n}x_{(n)}$, 它的均方误差为

$$MSE(\hat{\theta}) = Var(\hat{\theta}) = \frac{\theta^2}{n(n+2)}.$$

- 考虑 $\hat{\theta}_\alpha = \alpha x_{(n)}$, 其均方误差为

$$\begin{aligned} MSE(\hat{\theta}_\alpha) &= Var(\alpha x_{(n)}) + (\alpha E x_{(n)} - \theta)^2 \\ &= \frac{\alpha^2 n\theta^2}{(n+1)^2(n+2)} + (\frac{n\alpha\theta}{n+1} - \theta)^2. \end{aligned}$$

均方误差

- 总体为均匀分布 $U(0,\theta)$, θ 的一个无偏估计为 $\hat{\theta}=\frac{n+1}{n}x_{(n)}$，它的均方误差为

$$MSE(\hat{\theta})=Var(\hat{\theta})=\frac{\theta^2}{n(n+2)}.$$

- 考虑 $\hat{\theta}_\alpha=\alpha x_{(n)}$，其均方误差为

$$\begin{aligned}MSE(\hat{\theta}_\alpha)&=Var(\alpha x_{(n)})+(\alpha Ex_{(n)}-\theta)^2\\&=\frac{\alpha^2 n\theta^2}{(n+1)^2(n+2)}+(\frac{n\alpha\theta}{n+1}-\theta)^2.\end{aligned}$$

- 当 $\alpha=\frac{n+2}{n+1}$ 时，均方误差最小，为 $\frac{\theta^2}{(n+1)^2}$. 不是无偏估计!

一致最小方差无偏估计

- 定义：设 $\hat{\theta}$ 时未知参数 θ 的无偏估计，如果对另外任意一个 θ 的无偏估计 $\tilde{\theta}$，都有
$$Var(\hat{\theta}) \leqslant Var(\tilde{\theta})$$
则称 $\hat{\theta}$ 的一致最小方差无偏估计，简介为 $UMVUE$.

一致最小方差无偏估计

- 定义：设 $\hat{\theta}$ 时未知参数 θ 的无偏估计，如果对另外任意一个 θ 的无偏估计 $\tilde{\theta}$，都有

$$Var(\hat{\theta}) \leqslant Var(\tilde{\theta})$$

则称 $\hat{\theta}$ 的一致最小方差无偏估计，简介为 *UMVUE*.

- 定理：$X = (x_1, \ldots, x_n)$ 时来自某总体的一个样本，$\hat{\theta}$ 是 θ 的一个无偏估计，$Var(\hat{\theta}) < \infty$, 则 $\hat{\theta}$ 是 θ 的 UMVUE 的充分必要条件为：对任意满足 $E(\varphi(X)) = 0$, $Var(\varphi(X)) < \infty$ 的随机变量 $\varphi(X)$, 都有

$$Cov_\theta(\hat{\theta}, \varphi) = 0.$$

一致最小方差无偏估计

- 定义：设 $\hat{\theta}$ 时未知参数 θ 的无偏估计，如果对另外任意一个 θ 的无偏估计 $\tilde{\theta}$，都有
$$Var(\hat{\theta}) \leqslant Var(\tilde{\theta})$$
则称 $\hat{\theta}$ 的一致最小方差无偏估计，简介为 *UMVUE*.
- 定理：$X = (x_1, \ldots, x_n)$ 时来自某总体的一个样本，$\hat{\theta}$ 是 θ 的一个无偏估计，$Var(\hat{\theta}) < \infty$，则 $\hat{\theta}$ 是 θ 的 UMVUE 的充分必要条件为：对任意满足 $E(\varphi(X)) = 0$, $Var(\varphi(X)) < \infty$ 的随机变量 $\varphi(X)$，都有
$$Cov_\theta(\hat{\theta}, \varphi) = 0.$$
- 充分性：对于 θ 的任一个无偏估计 $\tilde{\theta}$，令 $\varphi = \hat{\theta} - \tilde{\theta}$，则 $E(\varphi) = 0$.

一致最小方差无偏估计

- 定义：设 $\hat{\theta}$ 时未知参数 θ 的无偏估计，如果对另外任意一个 θ 的无偏估计 $\tilde{\theta}$，都有

$$Var(\hat{\theta}) \leqslant Var(\tilde{\theta})$$

则称 $\hat{\theta}$ 的一致最小方差无偏估计，简介为 $UMVUE$.

- 定理：$X = (x_1, \ldots, x_n)$ 时来自某总体的一个样本，$\hat{\theta}$ 是 θ 的一个无偏估计，$Var(\hat{\theta}) < \infty$, 则 $\hat{\theta}$ 是 θ 的 UMVUE 的充分必要条件为：对任意满足 $E(\varphi(X)) = 0$, $Var(\varphi(X) < \infty$ 的随机变量 $\varphi(X)$, 都有

$$Cov_\theta(\hat{\theta}, \varphi) = 0.$$

- 充分性：对于 θ 的任一个无偏估计 $\tilde{\theta}$, 令 $\varphi = \hat{\theta} - \tilde{\theta}$, 则 $E(\varphi) = 0$.
$Var(\tilde{\theta}) = E(\tilde{\theta} - \hat{\theta} + \hat{\theta} - \theta)^2 = E(\varphi^2) + Var(\hat{\theta}) + 2Cov(\varphi, \hat{\theta}) \geqslant Var(\hat{\theta})$.

- 必要性：反证法。假设 $\hat{\theta}$ 是 θ 的 UMVUE, $\varphi(X)$ 满足

$$E(\varphi(X)) = 0, \quad Var(\varphi(X)) < \infty, \quad Cov_\theta(\varphi(X), \hat{\theta}) = a \neq 0.$$

一致最小方差无偏估计

- 必要性：反证法。假设 $\hat{\theta}$ 是 θ 的 UMVUE, $\varphi(X)$ 满足

$$E(\varphi(X)) = 0, \quad Var(\varphi(X)) < \infty, \quad Cov_\theta(\varphi(X), \hat{\theta}) = a \neq 0.$$

取 $b = \frac{-a}{Var(\varphi(X))} \neq 0$, 令 $\tilde{\theta} = \hat{\theta} + b\varphi(X)$, 则

- 必要性：反证法。假设 $\hat{\theta}$ 是 θ 的 UMVUE, $\varphi(X)$ 满足

$$E(\varphi(X)) = 0, \quad Var(\varphi(X)) < \infty, \quad Cov_\theta(\varphi(X), \hat{\theta}) = a \neq 0.$$

取 $b = \frac{-a}{Var(\varphi(X))} \neq 0$, 令 $\tilde{\theta} = \hat{\theta} + b\varphi(X)$, 则

$$\begin{aligned} Var(\tilde{\theta}) &= E(\hat{\theta} + b\varphi - \theta)^2 \\ &= E(\hat{\theta} - \theta)^2 + b^2 E(\varphi^2) + 2bE(\hat{\theta} - \theta)\varphi) \\ &= Var(\hat{\theta}) + b^2 Var(\varphi) + 2ab \\ &= Var(\hat{\theta}) - ab < Var(\hat{\theta}). \end{aligned}$$

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$，$\theta>0,\ x>0.$

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$，$\theta>0,\ x>0.$
- $\hat{\theta}=\bar{x}$ 为 θ 的无偏估计，令 $E\varphi(x_1,\ldots,x_n)=0$，即

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$，$\theta>0,\ x>0.$
- $\hat{\theta}=\bar{x}$ 为 θ 的无偏估计，令 $E\varphi(x_1,\ldots,x_n)=0$，即

$$E(\varphi)=\int_{R^n}\varphi(x_1,\ldots,x_n)\frac{1}{\theta^n}e^{-\frac{x_1+\cdots+x_n}{\theta}}dX=0.$$

即有 $\int_{R^n}\varphi(X)e^{-n\bar{x}/\theta}dX=0.$

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$，$\theta > 0, x > 0$.
- $\hat{\theta} = \bar{x}$ 为 θ 的无偏估计，令 $E\varphi(x_1, \ldots, x_n) = 0$, 即

$$E(\varphi) = \int_{R^n} \varphi(x_1, \ldots, x_n) \frac{1}{\theta^n} e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

即有 $\int_{R^n} \varphi(X) e^{-n\bar{x}/\theta} dX = 0$.对 θ 求导，

$$\int_{R^n} \frac{n\bar{x}}{\theta^2} \varphi(X) e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$，$\theta > 0$, $x > 0$.
- $\hat{\theta} = \bar{x}$ 为 θ 的无偏估计，令 $E\varphi(x_1, \ldots, x_n) = 0$, 即

$$E(\varphi) = \int_{R^n} \varphi(x_1, \ldots, x_n) \frac{1}{\theta^n} e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

即有 $\int_{R^n} \varphi(X) e^{-n\bar{x}/\theta} dX = 0$.对 θ 求导，

$$\int_{R^n} \frac{n\bar{x}}{\theta^2} \varphi(X) e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

$E(\bar{x}\varphi(X)) = 0 \Rightarrow Cov(\bar{x}, \varphi) = E(\bar{x}\varphi(X)) - E(\bar{x})E(\varphi(X)) = 0.$

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$, $p(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, $\theta > 0, x > 0$.
- $\hat{\theta} = \bar{x}$ 为 θ 的无偏估计, 令 $E\varphi(x_1, \ldots, x_n) = 0$, 即

$$E(\varphi) = \int_{R^n} \varphi(x_1, \ldots, x_n) \frac{1}{\theta^n} e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

即有 $\int_{R^n} \varphi(X) e^{-n\bar{x}/\theta} dX = 0.$对 θ 求导,

$$\int_{R^n} \frac{n\bar{x}}{\theta^2} \varphi(X) e^{-\frac{x_1+\cdots+x_n}{\theta}} dX = 0.$$

$E(\bar{x}\varphi(X)) = 0 \Rightarrow Cov(\bar{x}, \varphi) = E(\bar{x}\varphi(X)) - E(\bar{x})E(\varphi(X)) = 0.$

- $\bar{x}$ 为 UMVUE.

一致最小方差无偏估计

- 总体为指数分布 $Exp(\frac{1}{\theta})$，$p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$，$\theta>0,\ x>0$.
- $\hat{\theta}=\bar{x}$ 为 θ 的无偏估计，令 $E\varphi(x_1,\ldots,x_n)=0$, 即

$$E(\varphi)=\int_{R^n}\varphi(x_1,\ldots,x_n)\frac{1}{\theta^n}e^{-\frac{x_1+\cdots+x_n}{\theta}}dX=0.$$

即有 $\int_{R^n}\varphi(X)e^{-n\bar{x}/\theta}dX=0.$对 θ 求导，

$$\int_{R^n}\frac{n\bar{x}}{\theta^2}\varphi(X)e^{-\frac{x_1+\cdots+x_n}{\theta}}dX=0.$$

$E(\bar{x}\varphi(X))=0\Rightarrow Cov(\bar{x},\varphi)=E(\bar{x}\varphi(X))-E(\bar{x})E(\varphi(X))=0.$

- $\bar{x}$ 为 UMVUE. 同时 $\bar{x}$ 也是充分统计量。

一致最小方差无偏估计

- 总体为正态分布 $N(\mu, \sigma^2)$, $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s^2$ 是 μ, σ^2 的无偏估计。

一致最小方差无偏估计

- 总体为正态分布 $N(\mu, \sigma^2)$, $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s^2$ 是 μ, σ^2 的无偏估计。它们也是 UMVUE.

一致最小方差无偏估计

- 总体为正态分布 $N(\mu,\sigma^2)$, $\hat{\mu}=\bar{x}$, $\hat{\sigma}^2=s^2$ 是 μ, σ^2 的无偏估计。它们也是 UMVUE.
- $E(\varphi(x_1,\ldots,x_n))=0$, 即

$$E(\varphi)=\int\varphi(X)(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+n\bar{x}\frac{\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}}dX=0.$$

  - 对 μ 求导，得 $E(\frac{n\bar{x}\varphi}{\sigma^2})=0$, 即 $E(\bar{x}\varphi)=0\rightarrow Cov(\varphi,\bar{x})=0$.

一致最小方差无偏估计

- 总体为正态分布 $N(\mu,\sigma^2)$, $\hat{\mu}=\bar{x}$, $\hat{\sigma}^2=s^2$ 是 μ, σ^2 的无偏估计。它们也是 UMVUE.
- $E(\varphi(x_1,\ldots,x_n))=0$, 即

$$E(\varphi)=\int \varphi(X)(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+n\bar{x}\frac{\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}}dX=0.$$

  - 对 μ 求导，得 $E(\frac{n\bar{x}\varphi}{\sigma^2})=0$, 即 $E(\bar{x}\varphi)=0\to Cov(\varphi,\bar{x})=0$.
  - 对 μ 求两次导，得 $E(\bar{x}^2\varphi)=0$。

一致最小方差无偏估计

- 总体为正态分布 $N(\mu,\sigma^2)$, $\hat{\mu}=\bar{x}$, $\hat{\sigma}^2=s^2$ 是 μ, σ^2 的无偏估计。它们也是 UMVUE.
- $E(\varphi(x_1,\ldots,x_n))=0$, 即

$$E(\varphi)=\int\varphi(X)(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+n\bar{x}\frac{\mu}{\sigma^2}-\frac{n\mu^2}{2\sigma^2}}dX=0.$$

  - 对 μ 求导，得 $E(\frac{n\bar{x}\varphi}{\sigma^2})=0$, 即 $E(\bar{x}\varphi)=0\to Cov(\varphi,\bar{x})=0$.
  - 对 μ 求两次导，得 $E(\bar{x}^2\varphi)=0$。
  - 对 σ^2 求一次导，得 $E(\varphi\sum_{i=1}^n x_i^2)=0$, 从而 $E(s^2\varphi)=0$. 即有 $Cov(s^2,\varphi)=0$.
- $\bar{x}$ 和 $\sum_{i=1}^n x_i^2$ 是未知参数向量 (μ,σ^2) 得充分统计量。

充分性与 UMVUE

- 总体函数为 $p(x;\theta)$, $x_1,\ldots,x_n$ 是其样本，$T=T(x_1,\ldots,x_n)$ 是 θ 的充分统计量，则对于 θ 的任一无偏估计 $\hat{\theta}$，令 $\tilde{\theta}=E(\hat{\theta}|T)$, 则 $\tilde{\theta}$ 是 θ 的无偏估计，而且

$$Var(\tilde{\theta})\leqslant Var(\hat{\theta}).$$

充分性与 UMVUE

- 总体函数为 $p(x;\theta)$, $x_1,\ldots,x_n$ 是其样本， $T=T(x_1,\ldots,x_n)$ 是 θ 的充分统计量，则对于 θ 的任一无偏估计 $\hat{\theta}$，令 $\tilde{\theta}=E(\hat{\theta}|T)$, 则 $\tilde{\theta}$ 是 θ 的无偏估计，而且
$$Var(\tilde{\theta})\leqslant Var(\hat{\theta}).$$
- T 是充分统计量，所以 $E(\hat{\theta}|T)$ 与未知参数 θ 无关，同时
$$E(E(\hat{\theta}|T))=E(\hat{\theta})=\theta.$$

- 总体函数为 $p(x;\theta)$, $x_1,\ldots,x_n$ 是其样本， $T=T(x_1,\ldots,x_n)$ 是 θ 的充分统计量，则对于 θ 的任一无偏估计 $\hat{\theta}$，令 $\tilde{\theta}=E(\hat{\theta}|T)$, 则 $\tilde{\theta}$ 是 θ 的无偏估计，而且

$$Var(\tilde{\theta}) \leqslant Var(\hat{\theta}).$$

- T 是充分统计量，所以 $E(\hat{\theta}|T)$ 与未知参数 θ 无关，同时

$$E(E(\hat{\theta}|T)) = E(\hat{\theta}) = \theta.$$

$$\begin{aligned}
Var(\hat{\theta}) &= E(\hat{\theta}-\tilde{\theta}+\tilde{\theta}-\theta)^2 \\
&= E(\hat{\theta}-\tilde{\theta})^2 + E(\tilde{\theta}-\theta)^2 + 2\underbrace{E(\hat{\theta}-\tilde{\theta})(\tilde{\theta}-\theta)}_{=0}, \\
&= E(\hat{\theta}-\tilde{\theta})^2 + Var(\tilde{\theta}) \geqslant Var(\tilde{\theta}).
\end{aligned}$$

- 总体函数为 $p(x;\theta)$, $x_1,\ldots,x_n$ 是其样本，$T=T(x_1,\ldots,x_n)$ 是 θ 的充分统计量，则对于 θ 的任一无偏估计 $\hat{\theta}$，令 $\tilde{\theta}=E(\hat{\theta}|T)$, 则 $\tilde{\theta}$ 是 θ 的无偏估计，而且
$$Var(\tilde{\theta}) \leqslant Var(\hat{\theta}).$$

- T 是充分统计量，所以 $E(\hat{\theta}|T)$ 与未知参数 θ 无关，同时
$$E(E(\hat{\theta}|T)) = E(\hat{\theta}) = \theta.$$

$$\begin{aligned} Var(\hat{\theta}) &= E(\hat{\theta}-\tilde{\theta}+\tilde{\theta}-\theta)^2 \\ &= E(\hat{\theta}-\tilde{\theta})^2 + E(\tilde{\theta}-\theta)^2 + 2\underbrace{E(\hat{\theta}-\tilde{\theta})(\tilde{\theta}-\theta)}_{=0}, \\ &= E(\hat{\theta}-\tilde{\theta})^2 + Var(\tilde{\theta}) \geqslant Var(\tilde{\theta}). \end{aligned}$$

- 如果无偏估计 $\hat{\theta}$ 不是充分估计量 T 的函数，则可以降低方差。

充分性与 UMVUE

- 总体为 $b(1,p)$，$T=n\bar{x}$ 是 p 的充分估计量，$\theta=p^2$，考虑

$$\theta_1=\begin{cases}1, & x_1=1, x_2=1,\\ 0, & \text{其他}\end{cases}.$$

- 总体为 $b(1,p)$, $T = n\bar{x}$ 是 p 的充分估计量, $\theta = p^2$, 考虑

$$\theta_1 = \begin{cases} 1, & x_1 = 1, x_2 = 1, \\ 0, & \text{其他} \end{cases}.$$

  θ_1 是 θ 的无偏估计量:
  $E(\theta_1) = P(x_1 = 1, x_2 = 1) = p^2 = \theta.$

- 总体为 $b(1,p)$, $T=n\bar{x}$ 是 p 的充分估计量，$\theta=p^2$, 考虑

$$\theta_1=\begin{cases}1, & x_1=1, x_2=1,\\ 0, & \text{其他}\end{cases}.$$

θ_1 是 θ 的无偏估计量:
$E(\theta_1)=P(x_1=1,x_2=1)=p^2=\theta. Var(\theta_1)=\theta(1-\theta).$

- 总体为 $b(1,p)$, $T=n\bar{x}$ 是 p 的充分估计量, $\theta=p^2$, 考虑

$$\theta_1=\begin{cases}1, & x_1=1, x_2=1,\\ 0, & \text{其他}\end{cases}.$$

θ_1 是 θ 的无偏估计量:
$E(\theta_1)=P(x_1=1,x_2=1)=p^2=\theta. Var(\theta_1)=\theta(1-\theta).$

$$\begin{aligned}\hat{\theta}&=E(\theta_1|T=t)=P(\theta_1=1|T=t)\\&=\frac{P(x_1=1,x_2=1,T=t)}{P(T=t)}=\frac{p^2\binom{n-2}{t-2}p^{t-2}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}}\\&=\frac{\binom{n-2}{t-2}}{\binom{n}{t}}=\frac{t(t-1)}{n(n-1)}=\frac{\bar{x}(n\bar{x}-1)}{n-1}.\end{aligned}$$

- 总体为 $b(1,p)$, $T=n\bar{x}$ 是 p 的充分估计量，$\theta=p^2$, 考虑

$$\theta_1=\begin{cases}1, & x_1=1, x_2=1,\\ 0, & \text{其他}\end{cases}.$$

θ_1 是 θ 的无偏估计量：
$E(\theta_1)=P(x_1=1,x_2=1)=p^2=\theta. Var(\theta_1)=\theta(1-\theta).$

$$\begin{aligned}\hat{\theta}&=E(\theta_1|T=t)=P(\theta_1=1|T=t)\\ &=\frac{P(x_1=1,x_2=1,T=t)}{P(T=t)}=\frac{p^2\binom{n-2}{t-2}p^{t-2}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}}\\ &=\frac{\binom{n-2}{t-2}}{\binom{n}{t}}=\frac{t(t-1)}{n(n-1)}=\frac{\bar{x}(n\bar{x}-1)}{n-1}.\end{aligned}$$

$E(\hat{\theta})=\theta, \quad Var(\hat{\theta})<Var(\theta_1).$

- 总体分布为 $p(x;\theta)$，$x_1,\ldots,x_n$ 为该总体的样本，$T=T(x_1,\ldots,x_n)$ 是 $g(\theta)$ 的无偏估计，$g(\theta)$ 关于 θ 可微，即 $g'(\theta)$ 存在；而且

$$g'(\theta)=\int T(x_1,\ldots,x_n)\partial_\theta(\prod_{i=1}^{n}p(x_i;\theta))dX,$$

则有

$$Var(T)\geqslant\frac{[g'(\theta)]^2}{nI(\theta)},$$

其中 $I(\theta)$ 是费希尔信息量，$I(\theta)=E(\partial_\theta\ln p(x:\theta))^2$. 若以上不等式的等号成立，则称 $T=T(x_1,\ldots,x_n)$ 为 $g(\theta)$ 的有效估计，有效估计一定是 UMVUE.

Cramer-Rao 不等式

- $0 = \int \partial_\theta p(x_i;\theta)\,dx_i = \int [\partial_\theta \ln p(x_i;\theta)]p(x_i;\theta)\,dx_i = E[\partial_\theta \ln p(x_i;\theta)].$

Cramer-Rao 不等式

- $0 = \int \partial_\theta p(x_i;\theta)\,dx_i = \int [\partial_\theta \ln p(x_i;\theta)] p(x_i;\theta)\,dx_i = E[\partial_\theta \ln p(x_i;\theta)]$.
- 令 $Z = \partial_\theta \ln \prod_{i=1}^n p(x_i;\theta) = \sum_{i=1}^n \partial_\theta \ln p(x_i;\theta)$,

Cramer-Rao 不等式

- $0 = \int \partial_\theta p(x_i;\theta)\,dx_i = \int [\partial_\theta \ln p(x_i;\theta)] p(x_i;\theta)\,dx_i = E[\partial_\theta \ln p(x_i;\theta)]$.
- 令 $Z = \partial_\theta \ln \prod_{i=1}^n p(x_i;\theta) = \sum_{i=1}^n \partial_\theta \ln p(x_i;\theta)$，则

$$EZ = \sum_{i=1}^n E[\partial_\theta \ln p(x_i;\theta)] = 0,$$

$$EZ^2 = \sum_{i=1}^n Var[\partial_\theta \ln p(x_i;\theta)] = \sum_{i=1}^n E[\partial_\theta \ln p(x_i;\theta)]^2 = nI(\theta).$$

- $0 = \int \partial_\theta p(x_i;\theta)\,dx_i = \int [\partial_\theta \ln p(x_i;\theta)] p(x_i;\theta)\,dx_i = E[\partial_\theta \ln p(x_i;\theta)]$.
- 令 $Z = \partial_\theta \ln \prod_{i=1}^n p(x_i;\theta) = \sum_{i=1}^n \partial_\theta \ln p(x_i;\theta)$，则

$$EZ = \sum_{i=1}^n E[\partial_\theta \ln p(x_i;\theta)] = 0,$$

$$EZ^2 = \sum_{i=1}^n Var[\partial_\theta \ln p(x_i;\theta)] = \sum_{i=1}^n E[\partial_\theta \ln p(x_i;\theta)]^2 = nI(\theta).$$

$$\begin{aligned}[g'(\theta)]^2 &= [E(TZ)]^2 = [E[(T-g(\theta))Z]]^2 \\ &\leqslant E[(T-g(\theta))^2]E(Z^2) = Var(T)Var(Z) = Var(T)nI(\theta).\end{aligned}$$

Cramer-Rao 不等式

- $0 = \int \partial_\theta p(x_i;\theta)\,dx_i = \int [\partial_\theta \ln p(x_i;\theta)] p(x_i;\theta)\,dx_i = E[\partial_\theta \ln p(x_i;\theta)]$.
- 令 $Z = \partial_\theta \ln \prod_{i=1}^n p(x_i;\theta) = \sum_{i=1}^n \partial_\theta \ln p(x_i;\theta)$，则

$$EZ = \sum_{i=1}^{n} E[\partial_\theta \ln p(x_i;\theta)] = 0,$$

$$EZ^2 = \sum_{i=1}^{n} Var[\partial_\theta \ln p(x_i;\theta)] = \sum_{i=1}^{n} E[\partial_\theta \ln p(x_i;\theta)]^2 = nI(\theta).$$

$$\begin{aligned}[g'(\theta)]^2 &= [E(TZ)]^2 = [E[(T-g(\theta))Z]]^2 \\ &\leqslant E[(T-g(\theta))^2]E(Z^2) = Var(T)Var(Z) = Var(T)nI(\theta).\end{aligned}$$

$$Var(T) \geqslant [g'(\theta)]^2/(nI(\theta)).$$

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$.

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$. 为 $b(1,\theta)$,

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$. 为 $b(1,\theta)$, 费希尔信息量为

$$I(\theta)=E[\partial_\theta \ln p(x;\theta)]^2=\frac{1}{\theta^2}\theta+\frac{1}{(1-\theta)^2}(1-\theta)=\frac{1}{\theta(1-\theta)}.$$

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$. 为 $b(1,\theta)$, 费希尔信息量为

$$I(\theta)=E[\partial_\theta \ln p(x;\theta)]^2=\frac{1}{\theta^2}\theta+\frac{1}{(1-\theta)^2}(1-\theta)=\frac{1}{\theta(1-\theta)}.$$

- $\bar{x}$ 是 θ 的无偏估计，且 $Var(\bar{x})=\frac{\theta(1-\theta)}{n}=\frac{1}{nI(\theta)}$. 为有效估计

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$. 为 $b(1,\theta)$, 费希尔信息量为

$$I(\theta)=E[\partial_\theta \ln p(x;\theta)]^2=\frac{1}{\theta^2}\theta+\frac{1}{(1-\theta)^2}(1-\theta)=\frac{1}{\theta(1-\theta)}.$$

- $\bar{x}$ 是 θ 的无偏估计，且 $Var(\bar{x})=\frac{\theta(1-\theta)}{n}=\frac{1}{nI(\theta)}$. 为有效估计
- 总体为指数分布 $Exp(\frac{1}{\theta})$, $p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x>0,\theta>0$. 费希尔信息量为

$$I(\theta)=\int(-\frac{1}{\theta}+\frac{x}{\theta^2})\frac{1}{\theta}e^{-\frac{x}{\theta}}\,dx=\frac{1}{\theta^2}.$$

有效估计

- 总体为 $p(x:\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1$. 为 $b(1,\theta)$, 费希尔信息量为

$$I(\theta)=E[\partial_\theta \ln p(x;\theta)]^2=\frac{1}{\theta^2}\theta+\frac{1}{(1-\theta)^2}(1-\theta)=\frac{1}{\theta(1-\theta)}.$$

- $\bar{x}$ 是 θ 的无偏估计，且 $Var(\bar{x})=\frac{\theta(1-\theta)}{n}=\frac{1}{nI(\theta)}$. 为有效估计
- 总体为指数分布 $Exp(\frac{1}{\theta})$, $p(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x>0,\theta>0$. 费希尔信息量为

$$I(\theta)=\int(-\frac{1}{\theta}+\frac{x}{\theta^2})\frac{1}{\theta}e^{-\frac{x}{\theta}}\,dx=\frac{1}{\theta^2}.$$

- $Var(\bar{x})=\frac{\theta^2}{n}=\frac{1}{nI(\theta)}$. 为有效估计。

有效估计

- 总体为 $N(0,\sigma^2)$, 其费希尔信息量为

$$I(\sigma^2) = E[\frac{x^2}{2\sigma^4} - \frac{1}{2\sigma^2}]^2 = \frac{1}{4\sigma^4} Var(\frac{x^2}{\sigma^2}) = \frac{1}{2\sigma^4}.$$

有效估计

- 总体为 $N(0,\sigma^2)$, 其费希尔信息量为

$$I(\sigma^2) = E[\frac{x^2}{2\sigma^4} - \frac{1}{2\sigma^2}]^2 = \frac{1}{4\sigma^4} Var(\frac{x^2}{\sigma^2}) = \frac{1}{2\sigma^4}.$$

- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ 是 σ^2 的无偏估计，而且 $Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)}$.

有效估计

- 总体为 $N(0,\sigma^2)$, 其费希尔信息量为

$$I(\sigma^2) = E[\frac{x^2}{2\sigma^4} - \frac{1}{2\sigma^2}]^2 = \frac{1}{4\sigma^4} Var(\frac{x^2}{\sigma^2}) = \frac{1}{2\sigma^4}.$$

- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2$ 是 σ^2 的无偏估计，而且 $Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)}$.
- $\hat{\sigma} = \sqrt{\frac{n}{2}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)}\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}$ 为 σ 的无偏估计。其 C-R 下界为

$$\frac{[g'(\sigma^2)]^2}{nI(\sigma^2)} = \frac{\sigma^2}{2n}.$$

- $Var(\hat{\sigma}) = (\sqrt{\frac{n}{2}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)})^2\sigma^2 - \sigma^2 > \frac{\sigma^2}{2n}$,

有效估计

- 总体为 $N(0,\sigma^2)$, 其费希尔信息量为

$$I(\sigma^2) = E[\frac{x^2}{2\sigma^4} - \frac{1}{2\sigma^2}]^2 = \frac{1}{4\sigma^4} Var(\frac{x^2}{\sigma^2}) = \frac{1}{2\sigma^4}.$$

- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2$ 是 σ^2 的无偏估计，而且 $Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)}$.
- $\hat{\sigma} = \sqrt{\frac{n}{2}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)}\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}$ 为 σ 的无偏估计。其 C-R 下界为

$$\frac{[g'(\sigma^2)]^2}{nI(\sigma^2)} = \frac{\sigma^2}{2n}.$$

- $Var(\hat{\sigma}) = (\sqrt{\frac{n}{2}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)})^2\sigma^2 - \sigma^2 > \frac{\sigma^2}{2n}$，$\hat{\sigma}$ 是 UMVUE，但不是有效估计。

有效估计

- 总体为 $N(\mu,1)$, $T=\bar{x}^2-\frac{1}{n}$ 为 μ^2 的 UMVUE, 但不是有效估计。
- $E(T)=E(\bar{x}^2)-\frac{1}{n}=[E(\bar{x})]^2+Var(\bar{x})-\frac{1}{n}=\mu^2+\frac{1}{n}-\frac{1}{n}=\mu^2.$

有效估计

- 总体为 $N(\mu,1)$，$T=\bar{x}^2-\frac{1}{n}$ 为 μ^2 的 UMVUE, 但不是有效估计。
- $E(T)=E(\bar{x}^2)-\frac{1}{n}=[E(\bar{x})]^2+Var(\bar{x})-\frac{1}{n}=\mu^2+\frac{1}{n}-\frac{1}{n}=\mu^2$.
- 对于任意的 $E(\varphi(X))=0$,

$$E(\bar{x}^2\varphi)=\int(\frac{x_1+\cdots x_n}{n})^2\varphi(X)\frac{e^{-\frac{\sum_1^n(x_1-\mu)^2}{2}}}{(\sqrt{2\pi})^n}dX=0$$

  所以 $Cov(T,\varphi)=0$，T 为 UMVUE.
- $Var(T)=Var(\bar{x}^2)=\frac{2}{n^2}+\frac{4}{n}\mu^2$，费希尔信息量 $I(\mu)=1$，于是 C-R 下界为 $\frac{(2\mu)^2}{n}$. 所以 T 不是有效估计。

贝叶斯估计

- 统计推断的基础：总体的信息，样本的信息，

贝叶斯估计

- 统计推断的基础：总体的信息，样本的信息，先验的信息。

贝叶斯估计

- 统计推断的基础：总体的信息，样本的信息，先验的信息。
- 总体的信息：服从何种分布，数学期望、方差为何等等；
- 样本信息：经验分布函数，样本均值，样本方差，统计量等等；
- 先验信息：从以往经验里等到的信息：比如说之前已经做过一轮抽样研究等等。

贝叶斯估计

- 统计推断的基础：总体的信息，样本的信息，先验的信息。
- 总体的信息：服从何种分布，数学期望、方差为何等等；
- 样本信息：经验分布函数，样本均值，样本方差，统计量等等；
- 先验信息：从以往经验里等到的信息：比如说之前已经做过一轮抽样研究等等。
- 我们之前的统计方法都没有用到先验信息。并且假设未知参数是一个常数。

贝叶斯估计

- 统计推断的基础：总体的信息，样本的信息，先验的信息。
- 总体的信息：服从何种分布，数学期望、方差为何等等；
- 样本信息：经验分布函数，样本均值，样本方差，统计量等等；
- 先验信息：从以往经验里等到的信息：比如说之前已经做过一轮抽样研究等等。
- 我们之前的统计方法都没有用到先验信息。并且假设未知参数是一个常数。
- 贝叶斯估计的最大原则是未知参数是一个随机变量，可以用一个概率分布去描述，此分布为先验分布。

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X=(x_1,\ldots,x_n)$ 的联合条件密度函数为 $p(X|\theta_0)=p(x_1,\ldots,x_n|\theta_0)=\prod_{i=1}^n p(x_i|\theta_0)$.

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X = (x_1, \ldots, x_n)$ 的联合条件密度函数为 $p(X|\theta_0) = p(x_1, \ldots, x_n|\theta_0) = \prod_{i=1}^{n} p(x_i|\theta_0)$.
- 样本 X 与未知参数 θ 的联合密度函数为 $h(X, \theta) = \frac{h(X,\theta)}{\pi(\theta)}\pi(\theta) = p(X|\theta)\pi(\theta)$.

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X=(x_1,\ldots,x_n)$ 的联合条件密度函数为 $p(X|\theta_0)=p(x_1,\ldots,x_n|\theta_0)=\prod_{i=1}^n p(x_i|\theta_0)$.
- 样本 X 与未知参数 θ 的联合密度函数为 $h(X,\theta)=\frac{h(X,\theta)}{\pi(\theta)}\pi(\theta)=p(X|\theta)\pi(\theta)$.
- θ 的后验分布为 $\pi(\theta|X)=\frac{h(X,\theta)}{m(X)}=\frac{p(X|\theta)\pi(\theta)}{\int_\Theta p(X|\theta)\pi(\theta)\,d\theta}$.

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X = (x_1, \ldots, x_n)$ 的联合条件密度函数为 $p(X|\theta_0) = p(x_1, \ldots, x_n|\theta_0) = \prod_{i=1}^{n} p(x_i|\theta_0)$.
- 样本 X 与未知参数 θ 的联合密度函数为 $h(X, \theta) = \frac{h(X,\theta)}{\pi(\theta)}\pi(\theta) = p(X|\theta)\pi(\theta)$.
- θ 的后验分布为 $\pi(\theta|X) = \frac{h(X,\theta)}{m(X)} = \frac{p(X|\theta)\pi(\theta)}{\int_{\Theta} p(X|\theta)\pi(\theta)\, d\theta}$.
- 由后验分布 $\pi(\theta|X)$ 来估计 θ:
  - 使用后验密度函数的 $\pi(\theta|X)$ 最大值点作为 θ 的估计。

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X=(x_1,\ldots,x_n)$ 的联合条件密度函数为 $p(X|\theta_0)=p(x_1,\ldots,x_n|\theta_0)=\prod_{i=1}^n p(x_i|\theta_0)$.
- 样本 X 与未知参数 θ 的联合密度函数为 $h(X,\theta)=\frac{h(X,\theta)}{\pi(\theta)}\pi(\theta)=p(X|\theta)\pi(\theta)$.
- θ 的后验分布为 $\pi(\theta|X)=\frac{h(X,\theta)}{m(X)}=\frac{p(X|\theta)\pi(\theta)}{\int_\Theta p(X|\theta)\pi(\theta)\,d\theta}$.
- 由后验分布 $\pi(\theta|X)$ 来估计 θ:
  - 使用后验密度函数的 $\pi(\theta|X)$ 最大值点作为 θ 的估计。
  - 后验分布的数学期望;

贝叶斯公式的密度函数形式

- 总体含有未知参数 θ, 其在给定 θ 的取值时的条件密度函数 $p(x|\theta)$.
- 参数根据先验信息确定的先验分布为 $\pi(\theta)$.
- 样本 $X=(x_1,\ldots,x_n)$ 的联合条件密度函数为 $p(X|\theta_0)=p(x_1,\ldots,x_n|\theta_0)=\prod_{i=1}^n p(x_i|\theta_0)$.
- 样本 X 与未知参数 θ 的联合密度函数为 $h(X,\theta)=\frac{h(X,\theta)}{\pi(\theta)}\pi(\theta)=p(X|\theta)\pi(\theta)$.
- θ 的后验分布为 $\pi(\theta|X)=\frac{h(X,\theta)}{m(X)}=\frac{p(X|\theta)\pi(\theta)}{\int_\Theta p(X|\theta)\pi(\theta)\,d\theta}$.
- 由后验分布 $\pi(\theta|X)$ 来估计 θ:
  - 使用后验密度函数的 $\pi(\theta|X)$ 最大值点作为 θ 的估计。
  - 后验分布的数学期望;
  - 后验分布的中位数。

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X=x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\ldots,n.$$

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

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

- 假设 θ 的先验分布为 $U(0,1)$.

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

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

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X=x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\ldots,n.$$

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。
- X 与 θ 的联合（密度）函数为：
$h(x,\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,\ldots,n,\ 0<\theta<1.$

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X=x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\ldots,n.$$

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。
- X 与 θ 的联合（密度）函数为：
  $h(x,\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,\ldots,n,\ 0<\theta<1.$
- X 的边际分布为
  $m(X)=\binom{n}{x}\int_0^1\theta^x(1-\theta)^{n-x}d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}.$

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X = x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n.$$

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。
- X 与 θ 的联合（密度）函数为：
$h(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0, \ldots, n, \ 0 < \theta < 1.$
- X 的边际分布为
$m(X) = \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}d\theta = \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}.$
- θ 的后验分布为 $\pi(\theta|x) = \frac{h(x|\theta)}{m(x)} = \frac{\Gamma(n+2)\theta^{x+1-1}(1-\theta)^{n-x+1-1}}{\Gamma(x+1)\Gamma(n-x+1)}.$

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X=x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0,1,\ldots,n.$$

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。
- X 与 θ 的联合（密度）函数为：
$h(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0,\ldots,n,\ 0<\theta<1.$
- X 的边际分布为
$m(X) = \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}d\theta = \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}.$
- θ 的后验分布为 $\pi(\theta|x) = \frac{h(x|\theta)}{m(x)} = \frac{\Gamma(n+2)\theta^{x+1-1}(1-\theta)^{n-x+1-1}}{\Gamma(x+1)\Gamma(n-x+1)}.$
- $\theta|x \sim Be(x+1,n-x+1)$，$\hat{\theta} = E(\theta|x) = \frac{x+1}{n+2}.$

贝叶斯估计

- 设某事件发生的概率为 θ, 对试验进行 n 次独立的观测，其中 A 发生的次数为 X. 则 $X|\theta \sim b(n,\theta)$, 即

$$P(X=x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0,1,\ldots,n.$$

- 假设 θ 的先验分布为 $U(0,1)$. 贝叶斯假设：同等无知原则。
- X 与 θ 的联合（密度）函数为：
  $h(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0,\ldots,n,\ 0<\theta<1.$
- X 的边际分布为
  $m(X) = \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}d\theta = \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}.$
- θ 的后验分布为 $\pi(\theta|x) = \frac{h(x|\theta)}{m(x)} = \frac{\Gamma(n+2)\theta^{x+1-1}(1-\theta)^{n-x+1-1}}{\Gamma(x+1)\Gamma(n-x+1)}.$
- $\theta|x \sim Be(x+1,n-x+1)$，$\hat{\theta} = E(\theta|x) = \frac{x+1}{n+2}$. 与 $\bar{x} = \frac{X}{n}$ 不同。

贝叶斯估计

- 总体为 $N(\mu,\sigma_0^2)$, 其中方差 σ_0^2 已知，μ 的先验分布为 $N(\theta,\tau^2)$, 其中的 θ 和 τ 均已知

贝叶斯估计

- 总体为 $N(\mu,\sigma_0^2)$，其中方差 σ_0^2 已知，μ 的先验分布为 $N(\theta,\tau^2)$，其中的 θ 和 τ 均已知
- X 与 μ 的联合分布：$h(X,\mu)=\dfrac{e^{-\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^n x_i^2}{2\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{2\tau^2}\right]}}{(2\pi)^{(n+1)/2}\tau\sigma_0^n}$.

- 总体为 $N(\mu,\sigma_0^2)$，其中方差 σ_0^2 已知，μ 的先验分布为 $N(\theta,\tau^2)$，其中的 θ 和 τ 均已知
- X 与 μ 的联合分布：$h(X,\mu)=\frac{e^{-\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^n x_i^2}{2\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{2\tau^2}\right]}}{(2\pi)^{(n+1)/2}\tau\sigma_0^n}$.
- X 的边际分布为 $m(X)=\int h(X,\mu)d\mu$.
- 后验分布为 $\pi(\mu|X)=\frac{h(X,\mu)}{m(X)}\sim N\left(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{-1}+\tau^{-2}}\right)$.
- 后验均值作为贝叶斯估计

$$\hat{\mu}=\frac{n/\sigma_0^2}{n/\sigma_0^2+1/\tau^2}\bar{x}+\frac{1/\tau^2}{n/\sigma_0^2+1/\tau^2}\theta.$$

贝叶斯估计

- 总体为 $N(\mu,\sigma_0^2)$，其中方差 σ_0^2 已知，μ 的先验分布为 $N(\theta,\tau^2)$，其中的 θ 和 τ 均已知
- X 与 μ 的联合分布：$h(X,\mu)=\frac{e^{-\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^n x_i^2}{2\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{2\tau^2}\right]}}{(2\pi)^{(n+1)/2}\tau\sigma_0^n}$.
- X 的边际分布为 $m(X)=\int h(X,\mu)d\mu$.
- 后验分布为 $\pi(\mu|X)=\frac{h(X,\mu)}{m(X)}\sim N(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{-1}+\tau^{-2}})$.
- 后验均值作为贝叶斯估计

$$\hat{\mu}=\frac{n/\sigma_0^2}{n/\sigma_0^2+1/\tau^2}\bar{x}+\frac{1/\tau^2}{n/\sigma_0^2+1/\tau^2}\theta.$$

- 当 σ_0^2 越小，样本的均值占的比重越大，先验的方差 τ 越小，先验的比重越大。

共轭先验分布

- 定义：设 θ 是总体分布 $p(x;\theta)$ 中的参数，$\pi(\theta)$ 是其先验分布，如果对任意的来自总体 $p(x;\theta)$ 的样本观测值得到的后验分布 $\pi(\theta|X)$ 与 $\pi(\theta)$ 属于同一分布族（类），则称该分布族是 θ 的共轭先验分布族。

共轭先验分布

- 定义：设 θ 是总体分布 $p(x;\theta)$ 中的参数, $\pi(\theta)$ 是其先验分布，如果对任意的来自总体 $p(x;\theta)$ 的样本观测值得到的后验分布 $\pi(\theta|X)$ 与 $\pi(\theta)$ 属于同一分布族（类），则称该分布族是 θ 的共轭先验分布族。
- 贝塔分布是伯努利试验中成功概率的共轭先验分布族：

- 定义：设 θ 是总体分布 $p(x;\theta)$ 中的参数，$\pi(\theta)$ 是其先验分布，如果对任意的来自总体 $p(x;\theta)$ 的样本观测值得到的后验分布 $\pi(\theta|X)$ 与 $\pi(\theta)$ 属于同一分布族（类），则称该分布族是 θ 的共轭先验分布族。
- 贝塔分布是伯努利试验中成功概率的共轭先验分布族：均匀分布是 $Be(1,1)$.

共轭先验分布

- 定义：设 θ 是总体分布 $p(x;\theta)$ 中的参数，$\pi(\theta)$ 是其先验分布，如果对任意的来自总体 $p(x;\theta)$ 的样本观测值得到的后验分布 $\pi(\theta|X)$ 与 $\pi(\theta)$ 属于同一分布族（类），则称该分布族是 θ 的共轭先验分布族。
- 贝塔分布是伯努利试验中成功概率的共轭先验分布族：均匀分布是 $Be(1,1)$.
- 在方差已知时，正态分布时正态总体均值的共轭先验分布族。