数理统计 week 7

学业辅导中心

估计量的评价方法

上节讨论的方法勾画出求参数点估计量的技术轮廓. 不过这里产生了一个困难, 就是我们通常对一个问题可以应用不仅一种方法, 这就使我们时常面临在这些估计量之间进行选择的任务. 当然有可能不同的求估计量的方法导致相同的答案, 这时我们评价稍许容易些, 但是在很多情况下, 不同的方法将导致不同的估计量.

例

对于正态总体, 矩估计和 MLE 给出了两种关于方差的估计量.

这一节我们将介绍一些评价统计量的基本准则并用这些准则来检查几个估计量.

目录

1 均方误差

2 最佳无偏估计

均方误差

定义 (MSE)

参数 θ 的估计量 W 的均方误差 (mean squared error, 简记为 MSE) 是由 $\mathrm{E}_\theta(W-\theta)^2$ 定义的关于 θ 的函数.

偏差-方差分解:

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (\mathrm{E}_\theta W - \theta)^2 = \mathrm{Var}_\theta W + (\mathrm{Bias}_\theta W)^2$$

证明方法: 加一项减一项.
对一个无偏估计量, 我们有

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W$$

因此, 如果一个估计量是无偏的, 它的 MSE 就是它的方差.

正态 MSE

正态总体 $N(\mu, \sigma^2)$ 中两个估计量的 MSE 是

$$\mathrm{E}(\bar{X} - \mu)^2 = \mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n}$$

$$\mathrm{E}\left(S^2 - \sigma^2\right)^2 = \mathrm{Var}^2 = \frac{2\sigma^4}{n-1}$$

正态 MSE

正态总体 $N(\mu, \sigma^2)$ 中两个估计量的 MSE 是

$$\mathrm{E}(\bar{X} - \mu)^2 = \mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n}$$

$$\mathrm{E}\left(S^2 - \sigma^2\right)^2 = \mathrm{Var}^2 = \frac{2\sigma^4}{n-1}$$

$$\mathrm{Var}\,\frac{(n-1)S^2}{\sigma^2} = \mathrm{Var}\,\chi^2_{n-1}$$

$$\frac{(n-1)^2}{\sigma^4}\,\mathrm{Var}\,S^2 = 2(n-1)$$

$$\mathrm{Var}\,S^2 = \frac{2(n-1)\sigma^4}{(n-1)^2}$$

$$= \frac{2\sigma^4}{(n-1)},$$

正态总体中方差的估计

$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{n-1}{n}S^2$. 直接计算得到

$$\hat{\sigma}^2_{\text{MLE}} = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2$$

所以 $\hat{\sigma}^2$ 是 σ^2 的一个有偏的估计量. $\hat{\sigma}^2$ 的方差也可以计算如下

$$\operatorname{Var}\hat{\sigma}^2 = \operatorname{Var}\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 \operatorname{Var} S^2 = \frac{2(n-1)\sigma^4}{n^2}$$

于是, 它的 MSE 是

$$\mathrm{E}\left(\hat{\sigma}^2 - \sigma^2\right)^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4$$

这样我们有

$$\mathrm{E}\left(\hat{\sigma}^2 - \sigma^2\right)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = \mathrm{E}\left(S^2 - \sigma^2\right)^2$$

这说明 $\hat{\sigma}^2$ 具有比 S^2 更小的 MSE. 这样用偏倚抵换方差, MSE 得到改善.

例: Bernoulli

设 $X_1, \cdots, X_n$ 是来自概率质量函数为下式的总体的一组随机样本

$$P_\theta(X = x) = \theta^x(1-\theta)^{1-x}, x = 0 \text{ 或} 1, 0 \leqslant \theta \leqslant \frac{1}{2}$$

(a) 求 θ 的矩估计量和极大似然估计量.
(b) 求以上两种估计量的均方误差.
(c) 哪个估计量被优先选用? 论证你的选择.

解答

矩估计:

$$\mathrm{E}X = \theta = \frac{1}{n}\sum_i X_i = \bar{X} \Rightarrow \tilde{\theta} = \bar{X}$$

极大似然估计:

$$L(\theta \mid X) = \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}$$

对数似然函数:

$$\log L(\theta) = \sum_{i=1}^n X_i \log(\theta) + (n - \sum_{i=1}^n X_i)\log(1-\theta)$$

对数似然函数的导数:

$$\frac{\partial \log L(\theta)}{\partial \theta} = \frac{\sum_{i=1}^n X_i}{\theta} - \frac{n - \sum_{i=1}^n X_i}{1-\theta} \overset{set}{=} 0$$

解答

由于

$$\frac{\partial \log L(\theta)}{\partial \theta} = \frac{\sum_{i=1}^{n} X_i - n\theta}{\theta(1-\theta)}, \quad 0 \leqslant \theta \leqslant \frac{1}{2}.$$

当 $\bar{X} \leq \frac{1}{2}$ 时, 似然函数可以取到极大值点, $\hat{\theta} = \bar{X}$; 当 $\bar{X} > \frac{1}{2}$ 时, 似然函数递增, $\hat{\theta} = \frac{1}{2}$. 于是 $\hat{\theta} = \min\{\bar{X}, 1/2\}$.

解答

矩估计的 MSE:

$$\mathrm{MSE}(\tilde{\theta}) = \mathrm{Var}\,\tilde{\theta} + \mathrm{bias}(\tilde{\theta})^2 = (\theta(1-\theta)/n) + 0^2 = \theta(1-\theta)/n$$

$$= \mathrm{E}(\bar{X} - \theta)^2 = \sum_{y=0}^{n} \left(\frac{y}{n} - \theta\right)^2 \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}.$$

其中 $Y = \sum_{i=1}^{n} X_i \sim \mathit{Binomial}(n, \theta)$.

MLE 的 MSE:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2 = \sum_{y=0}^{n} (\hat{\theta} - \theta)^2 \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}$$

$$= \sum_{y=0}^{[n/2]} \left(\frac{y}{n} - \theta\right)^2 \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n} \left(\frac{1}{2} - \theta\right)^2 \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}.$$

解答

$$\begin{aligned}
\mathrm{MSE}(\tilde{\theta}) - \mathrm{MSE}(\hat{\theta}) &= \sum_{y=[n/2]+1}^{n} \left[\left(\frac{y}{n} - \theta \right)^2 - \left(\frac{1}{2} - \theta \right)^2 \right] \binom{n}{y} \theta^y (1-\theta)^{n-y} \\
&= \sum_{y=[n/2]+1}^{n} \left(\frac{y}{n} + \frac{1}{2} - 2\theta \right) \left(\frac{y}{n} - \frac{1}{2} \right) \binom{n}{y} \theta^y (1-\theta)^{n-y} \\
&> 0
\end{aligned}$$

因此当 $0 < \theta \leq 1/2$ 时, 应选择 MLE.(注意在 $\theta = 0$ 时, 两个估计量的 MSE 都是 0.)

最佳无偏估计量

若 W_1 和 W_2 都是参数 θ 的无偏估计量, 即 $\mathrm{E}_\theta W_1 = \mathrm{E}_\theta W_2 = \theta$, 则它们的均方误差就等于它们的方差, 所以我们应当选择方差比较小的那个估计量. 如果我们能找到一个具有一致最小方差的无偏估计量--最佳无偏估计量--那么我们的任务就完成了.
在开始这个进程之前我们指出, 虽然我们要处理无偏估计量, 但是这里以及下一节的结论实际上更为一般. 假定有 θ 的一个估计量 W^*, 其期望为 $\mathrm{E}_\theta W^* = \tau(\theta) \neq \theta$, 而我们感兴趣于研究 W^* 的价值. 考虑估计类

$$C_\tau = \{W : \mathrm{E}_\theta(W) = \tau(\theta)\}$$

由于对任何 $W_1, W_2 \in C_\tau$, $\mathrm{Bias}_\theta\, W_1 = \mathrm{Bias}_\theta\, W_2$, 于是

$$\mathrm{E}_\theta\,(W_1 - \theta)^2 - \mathrm{E}_\theta\,(W_2 - \theta)^2 = \mathrm{Var}_\theta\, W_1 - \mathrm{Var}_\theta\, W_2$$

这样, 在类 C_τ 中对 MSE 的比较就可以仅基于对方差的比较. 因此, 虽然我们是在用无偏估计量的术语讲话, 而实际上是比较具有相同期望 $\tau(\theta)$ 的估计量.

最佳无偏估计量

定义 (UMVUE)

估计量 W^* 称为 $\tau(\theta)$ 的最佳无偏估计量（best unbiased estimator）如果它满足 $\mathrm{E}_\theta W^* = \tau(\theta)$ 对所有 θ 成立, 并且对任何一个其他的满足 $\mathrm{E}_\theta(W) = \tau(\theta)$ 的估计量 W, 都有 $\mathrm{Var}_\theta\, W^* \leqslant \mathrm{Var}_\theta\, W$ 对所有 θ 成立. W^* 也称为 $\tau(\theta)$ 的一致最小方差无偏估计量 (uniform minimum variance unbiased estimator, 简记 UMVUE).

定理 (Cramér-Rao 不等式)

设 $X_1, \cdots, X_n$ 是具有概率密度函数 $f(\boldsymbol{x} \mid \theta)$ 的样本, 令 $W(\boldsymbol{X}) = W(X_1, \cdots, X_n)$ 是 $\tau(\theta)$ 的一个无偏估计量, 满足积分求导可交换, 即,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \mathrm{E}_\theta W(\boldsymbol{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} [W(\boldsymbol{x}) f(\boldsymbol{x} \mid \theta)] \mathrm{d}\boldsymbol{x}$$

和

$$\mathrm{Var}_\theta\, W(\boldsymbol{X}) < \infty$$

则有

$$\mathrm{Var}_\theta(W(\boldsymbol{x})) \geqslant \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \mathrm{E}_\theta W(\boldsymbol{X})\right)^2}{\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial \theta} \log f(\boldsymbol{X} \mid \theta)\right)^2\right)}$$

CRLB

证明思路:

1. $$\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)^2\right)$$

2. $$\mathrm{Cov}_\theta\left(W(\boldsymbol{X}),\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}_\theta W(\boldsymbol{X})$$

3. $$[\mathrm{Cov}(X,Y)]^2\leqslant(\mathrm{Var}\,X)(\mathrm{Var}\,Y)$$

综合以上三步, CRLB 就证明出来了.

Step 1

根据方差的定义,

$$\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)^2\right)+\left[\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)\right]^2$$

而根据条件的积分求导可交换, 取 $W(\mathbf{X})=1$, 有

$$\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}_\theta[1]=0$$

Step 2

对得分函数做如下变换,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}_\theta W(\boldsymbol{X}) &= \int_{\boldsymbol{X}} W(\boldsymbol{x})\left[\frac{\partial}{\partial\theta}f(\boldsymbol{x}\mid\theta)\right]\mathrm{d}x \\ &= \mathrm{E}_\theta\left[W(\boldsymbol{X})\frac{\frac{\partial}{\partial\theta}f(\boldsymbol{X}\mid\theta)}{f(\boldsymbol{X}\mid\theta)}\right] \qquad (\text{前式乘以} f(\boldsymbol{X}\mid\theta)/f(\boldsymbol{X}\mid\theta)) \\ &= \mathrm{E}_\theta\left[W(\boldsymbol{X})\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right] \qquad (\text{对数的性质})\end{aligned}$$

于是

$$\mathrm{Cov}_\theta\left(W(\boldsymbol{X}),\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\mathrm{E}_\theta\left(W(\boldsymbol{X})\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)=\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}_\theta W(\boldsymbol{X})$$

独立同分布下的 CRLB

定理

在上述定理的假设下, 附加假定 $X_1, \cdots, X_n$ 是 *iid* 的, 具有概率密度函数 $f(x \mid \theta)$, 则

$$\operatorname{Var}_\theta(W(\boldsymbol{X})) \geqslant \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}_\theta W(\boldsymbol{X})\right)^2}{n\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right)}$$

只需证明

$$\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X} \mid \theta)\right)^2\right) = n\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right)$$

证明独立同分布下的 CRLB

根据独立同分布，

$$
\begin{aligned}
\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)^2\right) &= \mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f\left(X_i\mid\theta\right)\right)^2\right) \\
&= \mathrm{E}_\theta\left(\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(X_i\mid\theta\right)\right)^2\right) \qquad \text{(对数的性质)} \\
&= \sum_{i=1}^{n}\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f\left(X_i\mid\theta\right)\right)^2\right) + \qquad \text{(平方展开)} \\
&\quad \sum_{i\neq j}^{n}\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f\left(X_i\mid\theta\right)\frac{\partial}{\partial\theta}\log f\left(X_j\mid\theta\right)\right)
\end{aligned}
$$

证明独立同分布下的 CRLB

对于 $i \neq j$, 根据独立性, 我们有

$$\begin{aligned}
&\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f\left(X_i \mid \theta\right)\frac{\partial}{\partial\theta}\log f\left(X_j \mid \theta\right)\right)\\
=&\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f\left(X_i \mid \theta\right)\right)\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f\left(X_j \mid \theta\right)\right)\\
=&0
\end{aligned}$$

根据同分布, 我们有

$$\sum_{i=1}^{n}\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f\left(X_i \mid \theta\right)\right)^2\right) = n\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right)$$

信息不等式

数量 $\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\mid\theta)\right)^2\right)$ 叫做样本的信息数 (Information number), 或 Fisher 信息量 (Fisher information)。这个术语反映这样一个事实, 信息量为最佳无偏估计量在 θ 处的方差给出了一个界. 当信息量增大, 我们就掌握关于 θ 更多的信息, 从而就有一个较小的对于最佳无偏估计方差的界. 事实上, Cramér-Rao 不等式这个术语也可以换成信息不等式 (Information Inequality).

对于任何可微函数 $\tau(\theta)$, 我们现在对于任何满足 CRLB 且 $\mathrm{E}_\theta W=\tau(\theta)$ 的估计量 W 的方差有一个下界. 这个界仅依赖于 $\tau(\theta)$ 和 $f(x\mid\theta)$ 并且是方差的一致下界. 任何一个估计量 W 满足 $\mathrm{E}_\theta W=\tau(\theta)$ 而且达到了这个下界, 它就是 $\tau(\theta)$ 的一个最佳无偏估计量.

信息论

离散随机变量 X 的熵 (entropy) 定义为

$$H(X) = \mathbf{E}\left(\log \frac{1}{P_X(X)}\right) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{1}{P_X(x)}.$$

设离散随机变量 X 的真实概率分布为 $P_X(x), x \in \mathcal{X}$, 估计概率分布为 $Q_X(x), x \in \mathcal{X}$, X 的交叉熵 (cross entropy) 定义为

$$H(P, Q) = \mathbf{E}\left(\log \frac{1}{Q_X(X)}\right) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{1}{Q_X(x)},$$

相对熵 (relative entropy) 定义为

$$D(P\|Q) = \mathbf{E}\left(\log \frac{P_X(X)}{Q_X(X)}\right) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{P_X(x)}{Q_X(x)}.$$

信息论

一对离散随机变量 (X, Y) 的联合熵 (joint entropy) 定义为

$$H(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log \frac{1}{P(x, y)},$$

条件熵 (conditional entropy) 定义为

$$H(X \mid Y) = \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P(x \mid y) \log \frac{1}{P(x \mid y)}.$$

在机器学习实现中, 一些损失函数往往被称为 cross-entropy loss, 对应的就是数理统计中的对数似然函数.

达到 CRLB

定理

设 $X_1, \cdots, X_n$ 是 *iid* 的, 具有概率密度函数 $f(x \mid \theta)$, 其 $f(x \mid \theta)$ 满足 *Cramér-Rao* 定理的条件. 令 $L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i \mid \theta)$ 表示似然函数. 如果 $W(\boldsymbol{X}) = W(X_1, \cdots, X_n)$ 是 $\tau(\theta)$ 的任意一个无偏估计量, 则 $W(\boldsymbol{X})$ 达到 *Cramér-Rao* 下界当且仅当

$$a(\theta)[W(x) - \tau(\theta)] = \frac{\partial}{\partial \theta} \log L(\theta \mid x)$$

对某一函数 $a(\theta)$ 成立.

证明即根据 Cauchy 不等式的取等条件. 等号成立当且仅当 $W(x) - \tau(\theta)$ 和 $\frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i \mid \theta)$ 成比例.

Bayes 估计

在经典方法中, 参数 θ 被认为是一个未知、但固定的量. 从以 θ 为指标的总体中抽取一组随机样本 $X_1, \cdots, X_n$, 基于样本的观测值来获得关于 θ 的知识. 在 Bayes 方法中, θ 被考虑成一个其变化可被一个概率分布描述的量, 该分布叫做先验分布 (prior distribution). 这是一个主观的分布, 建立在试验者的信念上, 而且见到抽样数据之前就已经用公式制定好了 (因而名为先验分布). 然后从以 θ 为指标的总体中抽取一组样本, 先验分布通过样本信息得到校正. 这个被校正的先验分布叫做后验分布（posterior distribution)。这个校正工作是通过 Bayes 法则完成的, 因而称为 Bayes 统计.

Bayes 估计方法

如果我们把先验分布记为 $\pi(\theta)$ 而把样本分布记为 $f(\boldsymbol{x} \mid \theta)$, 那么后验分布是给定样本 x 的条件下 θ 的条件分布, 就是
$\pi(\theta \mid \boldsymbol{x}) = f(\boldsymbol{x} \mid \theta)\pi(\theta)/m(\boldsymbol{x}), \quad (f(\boldsymbol{x} \mid \theta)\pi(\theta) = f(\boldsymbol{x}, \theta))$
这里 $m(x)$ 是 $\boldsymbol{X}$ 的边缘分布, 由下式得出,

$$m(\boldsymbol{x}) = \int f(\boldsymbol{x} \mid \theta)\pi(\theta)\mathrm{d}\theta$$

注意这个后验分布是一个条件分布, 其条件建立在观测样本上. 现在用这个后验分布来作出关于 θ 的推断, 而 θ 仍被考虑为一个随机的量. 例如, 后验分布的均值就可以被用作 θ 的点估计.

共轭族

定义

设 $\mathcal{F}$ 是概率密度函数或概率质量函数 $f(x \mid \theta)$ 的类 (以 θ 为指标). 称一个先验分布类 Π 为 $\mathcal{F}$ 的一个共轭族 (conjugate family), 如果对所有的 $f \in \mathcal{F}$, 所有的 Π 中的先验分布和所有的 $x \in X$, 其后验分布仍在 Π 中.

我们可以用以下的例子理解: 设 $X \sim N\left(\theta, \sigma^2\right)$, 假定 θ 的先验分布是 $N\left(\mu, \tau^2\right)$. (这里我们假定 σ^2, μ 和 τ^2 都已知.) θ 的后验分布也是正态分布, 均值与方差由以下给出

$$\mathrm{E}(\theta \mid x) = \frac{\tau^2}{n\tau^2 + \sigma^2} \sum_{i=1}^{n} X_i + \frac{\sigma^2}{\sigma^2 + n\tau^2} \mu$$

$$\mathrm{Var}(\theta \mid x) = \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2}$$

Bayes 估计

$\bar{x}, \theta$ 的联合密度是

$$f(\bar{x}, \theta) = f(\bar{x} \mid \theta)\pi(\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(\bar{x}-\theta)^2/\left(2\sigma^2\right)} \frac{1}{\sqrt{2\pi}\tau} e^{-(\theta-\mu)^2/2\tau^2}.$$

对于指数项,

$$\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2 = -\frac{1}{2v^2}(\theta-\delta(\mathbf{x}))^2 - \frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2,$$

其中 $\delta(\mathbf{x}) = \left(\tau^2\bar{x} + \left(\sigma^2/n\right)\mu\right) / \left(\tau^2 + \sigma^2/n\right)$, $v^2 = \left(\sigma^2\tau^2/n\right) / \left(\tau^2 + \sigma^2/n\right)$.
于是

$$f(\mathbf{x}, \theta) = \mathrm{n}\left(\theta, \sigma^2/n\right) \times \mathrm{n}\left(\mu, \tau^2\right) = \mathrm{n}\left(\delta(\mathbf{x}), v^2\right) \times \mathrm{n}\left(\mu, \tau^2 + \sigma^2/n\right).$$

Bayes 估计

$$\mathrm{E}(\theta \mid x) = \frac{\tau^2}{n\tau^2 + \sigma^2} \sum_{i=1}^{n} x_i + \frac{\sigma^2}{\sigma^2 + n\tau^2}\mu$$

$$\mathrm{Var}(\theta \mid x) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}$$

注意正态分布族是自身的共轭族. 再利用后验均值, 我们得到 θ 的 Bayes 估计量是 $\mathrm{E}(\theta \mid x)$.
这个 Bayes 估计量又是先验均值和样本均值的一个线性组合. 注意如果允许先验方差 τ^2 趋于无穷, Bayes 估计量就趋于样本均值. 我们可以对这个估计量作如下的解释, 当先验信息越不明确, 则 Bayes 估计量就倾向于给予样本信息越多的权重. 而另一方面, 当先验信息良好而 $\sigma^2 > \tau^2$ 时, 则更多的权重给予先验均值.