

3.3 贝叶斯估计

一、先验分布和后验分布

二、贝叶斯风险

三、贝叶斯估计

四、共轭先验分布



一、先验分布与后验分布

由 $R(\theta, d) = E_X(L(\theta, d(X))) = E_X(L(\theta, d(X_1, X_2, \dots, X_n)))$

令 $q(x) = q(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i, \theta),$

$$R(\theta, d) = E_X(L(\theta, d(X))) = \int_X L(\theta, d(x)) \mathbf{q(x)} dx$$

知： $R(\theta, d)$ 是决策函数 $d(x)$ 的平均损失，用来衡量决策 $d(x)$ 的好坏，仍然依赖于 θ 和 d ，为一个二元函数，很难进行全面比较。

为此，贝叶斯通过引入先验分布，给出了整体比较的指标。

1、先验信息

在抽取样本之前，人们对所要估计的未知参数所了解的信息，通常称为**先验信息**。



例1 (p89例3. 6) 某学生通过物理试验来确定当地的重力加速度，测得的数据为(m/s²):

9.80, 9.79, 9.78, 6.81, 6.80

试求当地的重力加速度.

解 用样本均值估计其重力加速度应该是合理的，即

$$\bar{X} = 8.596$$

由经验可知，此结果是不符合事实的。因为在估计之前我们知道，重力加速度应该在9.80附近，即

$$X \sim N(9.80, 0.1^2)$$

这个信息就是重力加速度的**先验信息**.

在统计学中，先验信息可以更好的帮助人们解决统计决策问题。

贝叶斯将此思想应用于统计决策中，形成了完整的贝叶斯统计方法。

2、先验分布

把未知参数 θ 看成一个随机变量，对 θ 的先验信息用一个分布 $\pi(\theta)$ 来表示，此分布 $\pi(\theta)$ 称为 θ 的先验分布。

例如 例1中重力加速度的先验分布为

$$X \sim N(9.80, 0.1^2)$$

3、后验分布

抽样之前，人们对未知参数的了解，即为先验分布。抽样之后，由于样本中包含未知参数的信息，而这些关于未知参数新的信息可以帮助人们修正抽样之前的先验分布。

设总体 $X \sim p(x, \theta), \theta \in \Theta, \theta$ 的先验分布为 $\pi(\theta)$, $(X_1, X_2, \dots, X_n)^T$ 为总体 X 的样本，其联合密度为

$$q(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i, \theta),$$

而样本值是在知道 θ 的先验分布的前提下得到的，因而上述分布可以改写为

$$q(x | \theta) = q(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n p(x_i | \theta),$$

θ 和样本 x 的联合分布可以表示为

$$f(x, \theta) = q(x | \theta)\pi(\theta) = m(x)h(\theta | x)$$

由此

$$h(\theta | x) = \frac{q(x | \theta)\pi(\theta)}{m(x)}, \quad (m(x) = \int_{\Theta} q(x | \theta)\pi(\theta) d\theta)$$

称 $h(\theta | x)$ 为 θ 的后验分布.

即加入新的信息以后, 对原有分布进行修正.

所以后验分布综合运用了先验分布与样本信息.

注：给定 $\theta \sim \pi(\theta)$ 时，

$$q(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i, \theta),$$

$$\Rightarrow q(x | \theta) = q(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n p(x_i | \theta),$$

$$R(\theta, d) = E_X(L(\theta, d(X))) = \int_X L(\theta, d(x)) q(x) dx$$



$$R(\theta, d) = E_{X|\theta}(L(\theta, d(X))) = \int_X L(\theta, d(x)) q(x | \theta) dx$$

现在 θ 是随机变量， $R(\theta, d)$ 也是随机变量。

例2 (p90例3.7) 为了提高某产品的质量，公司经理考虑增加投资来改进生产设备，预计需投资90万元，但从投资效果来看，顾问们提出两种不同的意见：

θ_1 : 改进生产设备后，高质量产品可占90%，

θ_2 : 改进生产设备后，高质量产品可占70%，

经理根据以往的经验，两个顾问建议可信度分别为

$$\pi(\theta_1) = 0.4 \quad \pi(\theta_2) = 0.6$$

这两个概率是经理的主观判断（也就是先验率），为了得到更准确的信息，经理决定进行小规模试验，实验结果如下：

A: 试制5个产品，全是正品，

根据这个试验结果，经理对顾问的可信度变为多少？

解: 根据题意，参数 θ 的先验分布 $\pi(\theta)$ 为

θ	θ_1	θ_2
P	0.4	0.6

X: 试制5个产品，正品次数，

$$X \sim B(5, p) \quad p = 0.9, 0.7$$

$$P(A) = C_5^5 p^5 (1-p)^{5-5} = p^5$$

如果相信顾问1的话, 则 $p(A | \theta_1) = (0.9)^5 = 0.590$,

如果相信顾问2的话, 则 $p(A | \theta_2) = (0.7)^5 = 0.168$

由全概率公式可以得到:

$$P(A) = P(A | \theta_1)P(\theta = \theta_1) + P(A | \theta_2)P(\theta = \theta_2) = 0.337$$

其后验概率为:

$$h(\theta_1 | A) = \frac{P(A | \theta_1)P(\theta = \theta_1)}{P(A)} = 0.700$$

$$h(\theta_2 | A) = \frac{P(A | \theta_2)P(\theta = \theta_2)}{P(A)} = 0.300$$

显然经理对二位顾问的看法已经做了修改，为了得到更准确的信息，经理又做了一次试验，结果为

B: 试制10个产品，9个是正品，

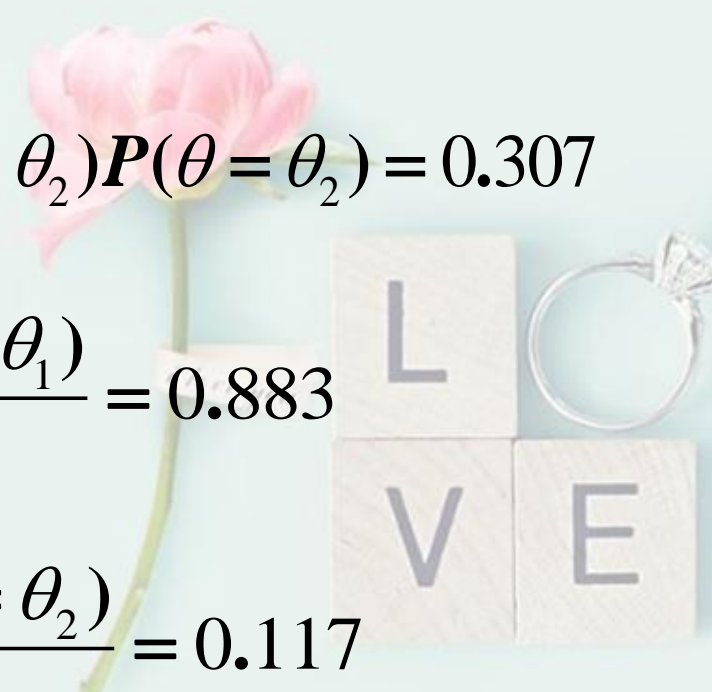
此时先验分布为



θ	θ_1	θ_2
P	0.7	0.3

$$P(B | \theta_1) = 10(0.9)^9 (0.1) = 0.387,$$

$$P(B | \theta_2) = 10(0.7)^9 (0.3) = 0.121$$


$$P(B) = P(B | \theta_1)P(\theta = \theta_1) + P(B | \theta_2)P(\theta = \theta_2) = 0.307$$

$$h(\theta_1 | B) = \frac{P(B | \theta_1)P(\theta = \theta_1)}{P(B)} = 0.883$$

$$h(\theta_2 | B) = \frac{P(B | \theta_2)P(\theta = \theta_2)}{P(B)} = 0.117$$

由此可见后验分布更能准确描述事情真相。

二、贝叶斯风险

已知风险函数

$$R(\theta, d) = E_{X|\theta}(L(\theta, d(X))) = \int_{\mathcal{X}} L(\theta, d(x)) q(x | \theta) dx$$

是随机变量 θ 的函数，也是随机变量。

1、贝叶斯风险的定义

在给定 θ 的先验分布 $\pi(\theta)$ 时，定义

$$R_B(d) = E(R(\theta, d)) = \int_{\Theta} R(\theta, d) \pi(\theta) d\theta$$

为决策函数 d 在给定先验分布 $\pi(\theta)$ 下的贝叶斯风险，简称为 d 的贝叶斯风险。

2、贝叶斯风险的计算

当 X 与 θ 都是连续型随机变量时，贝叶斯风险为

$$\begin{aligned} R_B(d) &= E(R(\theta, d)) = \int_{\Theta} R(\theta, d) \pi(\theta) d\theta \\ &= \int_{\Theta} \int_X L(\theta, d(x)) q(x | \theta) \pi(\theta) dx d\theta \\ &= \int_{\Theta} \int_X L(\theta, d(x)) h(\theta | x) m(x) dx d\theta \\ &= \int_X m(x) \left\{ \int_{\Theta} L(\theta, d(x)) h(\theta | x) d\theta \right\} dx \end{aligned}$$

当 X 与 θ 都是离散型随机变量时，贝叶斯风险为

$$R_B(d) = E(R(\theta, d))$$

$$= \sum_x m(x) \left\{ \sum_{\theta} L(\theta, d(x)) h(\theta | x) \right\}$$

注 由上述计算可以看出，贝叶斯风险为计算2次期望值得到，即

$$R_B(d) = E_{\theta}(E_{X|\theta}(L(\theta, d(X))))$$

$$= E(R(\theta, d)) = E_X(E_{\theta|X}(L(\theta, d(X))))$$

此时风险大小只与决策函数 d 有关，而不再依赖参数 θ 。因此以此来衡量决策函数优良性更合理。


三、贝叶斯估计

1、贝叶斯点估计

定义3.8 若总体 X 的分布函数 $F(x, \theta)$ 中参数 θ 为随机变量， $\pi(\theta)$ 为 θ 的先验分布，若决策函数类 D 中存在一个决策函数使得对决策函数类中的任一决策函数均有

$$R_B(d^*) = \inf_{d \in D} R_B(d), \quad \forall d \in D$$

则称 $d^*(X)$ 为参数 θ 的贝叶斯估计量
也称 $d^*(X)$ 为参数 θ 的贝叶斯决策函数。

- 
- 注** 1、贝叶斯估计是使贝叶斯风险达到最小的决策函数.
- 2、不同的先验分布, 对应不同的贝叶斯估计
- 3、不同的损失函数, 对应不同的贝叶斯估计

由于
$$R_B(d) = E(R(\theta, d)) = E_X(E_{\theta|X}(L(\theta, d(X))))$$

是求两次期望所得。

$$= \begin{cases} \int_{\Theta} L(\theta, d(x)) h(\theta | x) d\theta, & \text{连续型} \\ \sum_{\theta} L(\theta, d(x)) h(\theta | x), & \text{离散型} \end{cases}$$

定义3.9 设 $d=d(x)$ 为决策函数类 D 中任一决策函数，损失函数为 $L(\theta, d(x))$ ，则 $L(\theta, d(x))$ 对后验分布 $h(\theta|x)$ 的数学期望称为**后验风险**，记为

$$R(d | x) = E_{\Theta}[L(\theta, d(x))]$$

$$= \begin{cases} \int_{\Theta} L(\theta, d(x)) h(\theta | x) d\theta, & \theta \text{ 为连续型随机变量,} \\ \sum_i L(\theta_i, d(x)) h(\theta_i | x), & \theta \text{ 为离散型随机变量.} \end{cases}$$

注 如果存在一个决策函数 d^{**} ，使得

$$R(d^{**} | x) = \inf_d R(d | x), \quad \forall d \in D$$

则称此决策为后验风险准则下的最优决策函数，
或称为贝叶斯（后验型）决策函数。



定理3.5 对给定的统计决策问题(包含先验分布给定的情形) 和决策函数类 D ,当贝叶斯风险满足如下条件:

$$\inf_d R_B(d) < \infty, \quad \forall d \in D$$

则贝叶斯决策函数 $d^*(x)$ 与贝叶斯后验型决策函数 $d^{**}(x)$ 是等价的.

证明从略。

(等价性的证明参见《高等数理统计》P361茆诗松等编或者利用法都引理)

定理表明：如果决策函数使得贝叶斯风险最小，此决策函数也使得后验风险最小，反之，也成立。

$$\underline{R_B(d^*)} = E_X(\underline{E_{\theta|X}(L(\theta, d^*(X)))})$$

$$\min R_B(d^*) \Leftrightarrow \min E_{\theta|X}(L(\theta, d^*(X)))$$

2、贝叶斯点估计的计算

平方损失下的贝叶斯估计

定理3.2 设 θ 的先验分布为 $\pi(\theta)$ 和损失函数为

$$L(\theta, d) = (\theta - d)^2$$

则 θ 的贝叶斯估计为

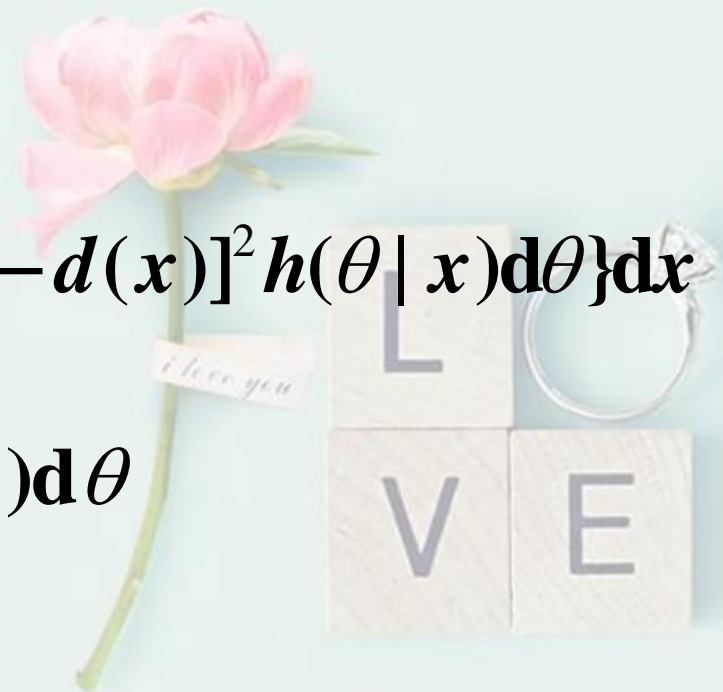
$$d^*(x) = E(\theta | X = x) = \int_{\Theta} \theta h(\theta | x) d\theta$$

其中 $h(\theta | x)$ 为参数 θ 的后验分布. 后验分布期望



证 首先对贝叶斯风险做变换

$$\begin{aligned}\min R_B(d) &= \min \int_X m(x) \left\{ \int_{\Theta} [\theta - d(x)]^2 h(\theta | x) d\theta \right\} dx \\ &\Leftrightarrow \min_a \int_{\Theta} [\theta - d(x)]^2 h(\theta | x) d\theta\end{aligned}$$



又因为

$$E(\theta | x) = \int_{\Theta} \theta h(\theta | x) d\theta$$

$$\int_{\Theta} [\theta - d(x)]^2 h(\theta | x) d\theta$$

$$= \int_{\Theta} [\theta - E(\theta | x) + E(\theta | x) - d(x)]^2 h(\theta | x) d\theta$$

$$= \int_{\Theta} [\theta - E(\theta | x)]^2 h(\theta | x) d\theta + \int_{\Theta} [E(\theta | x) - d(x)]^2 h(\theta | x) d\theta + 2 \int_{\Theta} [\theta - E(\theta | x)][E(\theta | x) - d(x)] h(\theta | x) d\theta \quad (1)$$

$$= [E(\theta | x) - d(x)] \int_{\Theta} [\theta - E(\theta | x)] h(\theta | x) d\theta$$

$$= [E(\theta | x) - d(x)][E(\theta | x) - E(\theta | x)] = 0$$

显然，当 $d^*(x) = E(\theta | x)$ a.s 时，(1) 为 0.

即 $R_B(d)$ 达到最小.

例5(p98例3.11) 设总体 X 服从两点分布 $B(1,p)$, 其中参数 p 未知, 而 p 在 $[0,1]$ 上服从均匀分布, 样本 (X_1, X_2, \dots, X_n) 来自总体 X , 损失函数为平方损失, 试求参数 p 的贝叶斯估计与贝叶斯风险?

解 平方损失下的贝叶斯估计为:

$$d^*(x) = E(p | X = x) = \int_{\Theta} p h(p | x) dp$$

而

$$h(p | x) = \frac{q(x | p)\pi(p)}{m(x)} = \frac{q(x | p)\pi(p)}{\int_0^1 q(x | p)\pi(p) dp}$$

$$= \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \cdot 1}{\int_0^1 p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \cdot 1 dp} = \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}{\beta(\sum_{i=1}^n x_i + 1, n + 1 - \sum_{i=1}^n x_i)}$$

其中 $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, 则

$$h(p | x) = \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \Gamma(n+2)}{\Gamma(\sum_{i=1}^n x_i + 1) \Gamma(n+1 - \sum_{i=1}^n x_i)}$$

← 是β分布

$$= \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} (n+1)!}{(\sum_{i=1}^n x_i)! (n - \sum_{i=1}^n x_i)!}$$

$$d^*(x) = \hat{p} = \int_{\Theta} ph(p | x) dp$$

$$= \frac{(n+1)!}{(\sum_{i=1}^n x_i)!(n - \sum_{i=1}^n x_i)!} \int_0^1 p^{\sum_{i=1}^n x_i + 1} (1-p)^{n - \sum_{i=1}^n x_i} dp$$

$$= \frac{(n+1)!}{(\sum_{i=1}^n x_i)!(n - \sum_{i=1}^n x_i)!} \frac{(\sum_{i=1}^n x_i + 1)!(n - \sum_{i=1}^n x_i)!}{(n+2)!}$$

$$= \frac{\sum_{i=1}^n x_i + 1}{n+2}$$

注：如果知道 β 分布的均值，可直接写出贝叶斯估计。

计算贝叶斯风险，先计算风险函数

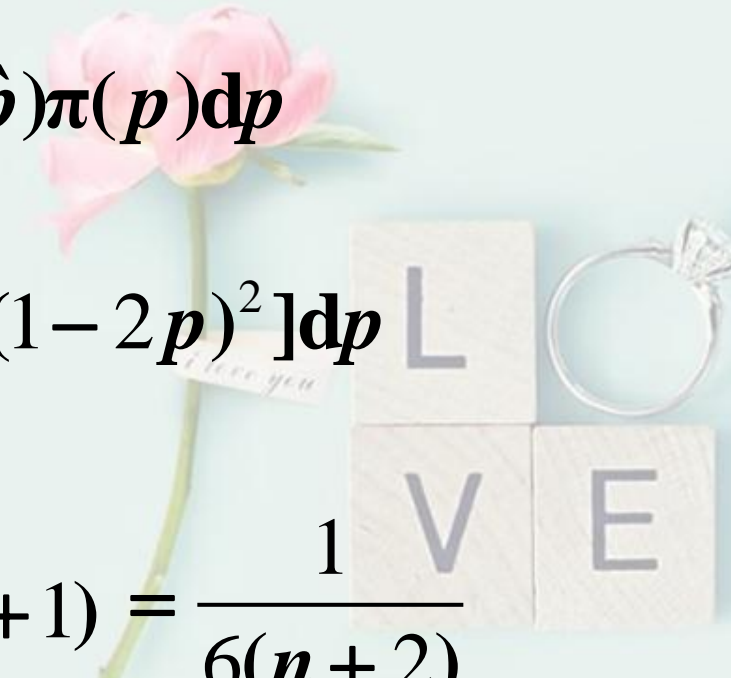
$$\begin{aligned} R(\hat{p}, p) &= E[L(\hat{p}, p)] = E\left(\frac{\sum_{i=1}^n X_i + 1}{n + 2} - p\right)^2 \\ &= \frac{1}{(n + 2)^2} E\left(\sum_{i=1}^n X_i + 1 - (n + 2)p\right)^2 \\ &= \frac{1}{(n + 2)^2} E\left(\sum_{i=1}^n X_i - np + 1 - 2p\right)^2 \\ &= \frac{1}{(n + 2)^2} E\left[\left(\sum_{i=1}^n X_i - np\right)^2 + (1 - 2p)^2\right. \\ &\quad \left.+ 2(1 - 2p)\left(\sum_{i=1}^n X_i - np\right)\right] \end{aligned}$$

又因为 $(\sum_{i=1}^n X_i) \sim B(n, p)$

则 $E(\sum_{i=1}^n X_i) = np, \quad D(\sum_{i=1}^n X_i) = np(1-p)$

所以上式

$$\begin{aligned} &= \frac{1}{(n+2)^2} (D(\sum_{i=1}^n X_i) + (1-2p)^2 + 0) \\ &= \frac{1}{(n+2)^2} (np(1-p) + (1-2p)^2) \end{aligned}$$



$$\begin{aligned}
 R_B(\hat{p}) &= E(R(p, \hat{p})) = \int_{\Theta} R(p, \hat{p}) \pi(p) dp \\
 &= \frac{1}{(n+2)^2} \int_0^1 [np(1-p) + (1-2p)^2] dp \\
 &= \frac{1}{(n+2)^2} \left(\frac{4-n}{3} + \frac{n-4}{2} + 1 \right) = \frac{1}{6(n+2)}
 \end{aligned}$$

而 p 的最大似然估计为 \bar{X} ,其贝叶斯风险为 $\frac{1}{6n} > \frac{1}{6(n+2)}$

注：这里按照如下公式计算贝叶斯风险

$$R_B(d) = E(R(\theta, d)) = \int_{\Theta} R(\theta, d) \pi(\theta) d\theta$$

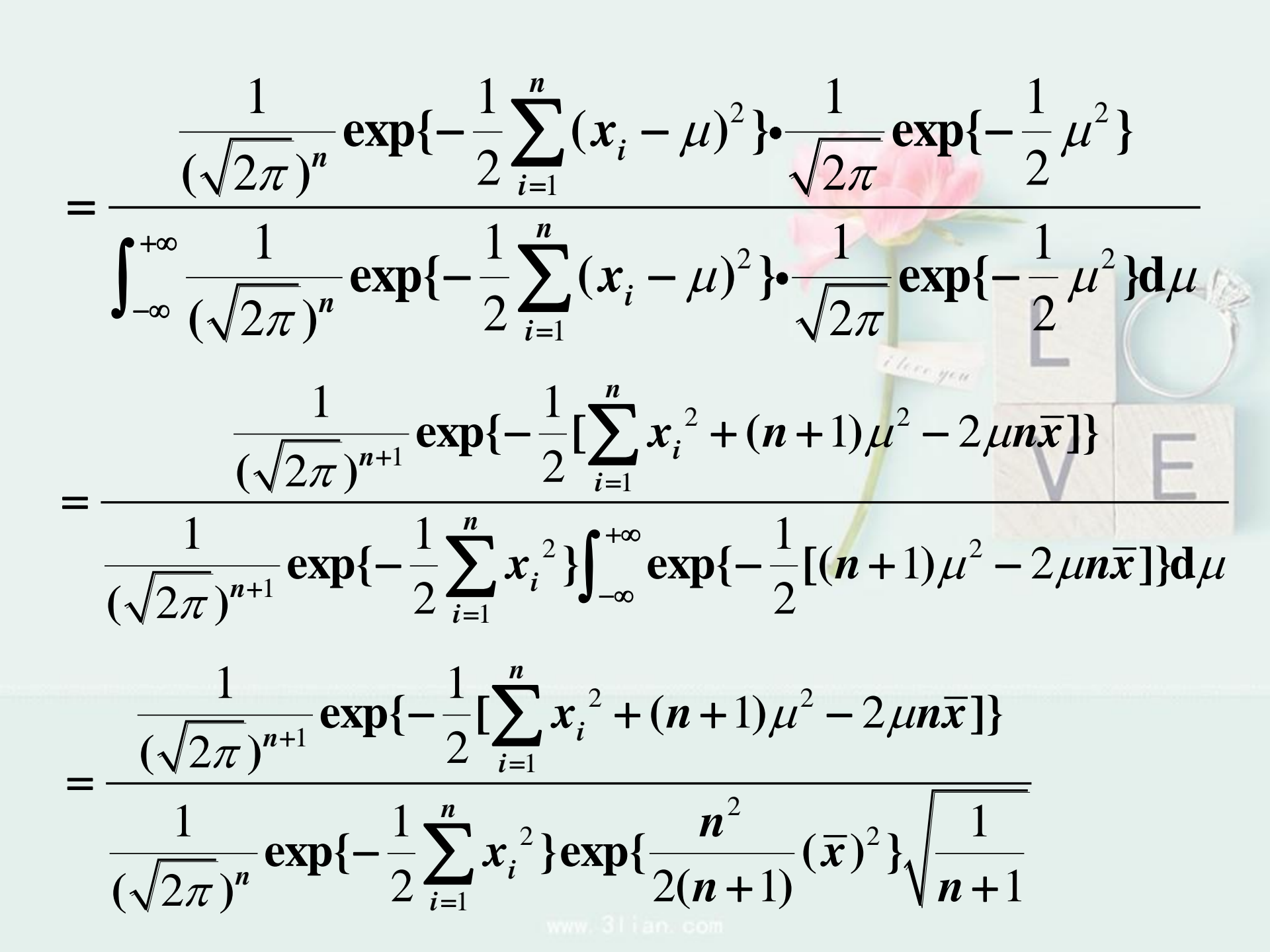
例6(p100例3.12) 设总体 X 服从正态分布 $N(\mu, I)$, 其中参数 μ 未知, 而 μ 服从标准正态分布 $N(0, 1)$, 样本 (X_1, X_2, \dots, X_n) 来自总体 X , 损失函数为平方损失, 试求参数 μ 的贝叶斯估计?

解 平方损失下的贝叶斯估计为:

$$d^*(x) = E(\mu | X = x) = \int_{\Theta} \mu h(\mu | x) d\mu$$

而

$$h(\mu | x) = \frac{q(x | \mu)\pi(\mu)}{m(x)} = \frac{q(x | \mu)\pi(\mu)}{\int_{-\infty}^{+\infty} q(x | \mu)\pi(\mu) d\mu}$$



$$\begin{aligned}
 &= \frac{\frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \mu^2\right\}}{\int_{-\infty}^{+\infty} \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \mu^2\right\} d\mu} \\
 &= \frac{\frac{1}{(\sqrt{2\pi})^{n+1}} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^n x_i^2 + (n+1)\mu^2 - 2\mu n\bar{x}\right]\right\}}{\frac{1}{(\sqrt{2\pi})^{n+1}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2} [(n+1)\mu^2 - 2\mu n\bar{x}]\right\} d\mu} \\
 &= \frac{\frac{1}{(\sqrt{2\pi})^{n+1}} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^n x_i^2 + (n+1)\mu^2 - 2\mu n\bar{x}\right]\right\}}{\frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x_i^2\right\} \exp\left\{\frac{n^2}{2(n+1)} (\bar{x})^2\right\} \sqrt{\frac{1}{n+1}}}
 \end{aligned}$$

$$= \frac{\frac{1}{(\sqrt{2\pi})^{n+1}} \exp\{-\frac{1}{2}[\sum_{i=1}^n x_i^2 + (n+1)\mu^2 - 2\mu n\bar{x}]\}}{\frac{1}{(\sqrt{2\pi})^n} \exp\{-\frac{1}{2}\sum_{i=1}^n x_i^2\} \exp\{\frac{n^2}{2(n+1)}(\bar{x})^2\} \sqrt{\frac{1}{n+1}}}$$

化简得

$$h(\mu | x) = \left(\frac{n+1}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{n+1}{2}\left(\mu - \frac{n\bar{x}}{n+1}\right)^2\right\}$$

$$d^*(x) = \hat{\mu} = E(\mu | x)$$

$$= \frac{n\bar{x}}{n+1} \neq \bar{x}$$

其贝叶斯风险为 $\frac{1}{n+1} < R_B(\bar{x}) = \frac{D(X)}{n} = \frac{1}{n}$

定理3.3 设 θ 的先验分布为 $\pi(\theta)$ 和损失函数为加权平方损失

$$L(\theta, d) = \lambda(\theta)(\theta - d)^2$$

则 θ 的贝叶斯估计为

$$d^*(x) = \frac{E(\lambda(\theta) \cdot \theta | x)}{E(\lambda(\theta) | x)}$$

证明略，此证明与定理3.2的证明类似。

定理3.4 设参数 θ 为随机向量，先验分布为 $\pi(\theta)$ 和损失函数为二次损失函数

$$L(\theta, d) = (d - \theta)^T Q (d - \theta)$$

其中 Q 为正定矩阵，则 θ 的贝叶斯估计为后验分布 $h(\theta|x)$ 的均值向量，即

$$d^*(x) = E(\theta | x) = \begin{bmatrix} E(\theta_1 | x) \\ \vdots \\ E(\theta_p | x) \end{bmatrix}$$

其中参数向量为 $\theta = (\theta_1, \theta_2, \dots, \theta_p)$.

注 定理表明，正定二次损失下， θ 的贝叶斯估计不受正定矩阵 Q 的选取干扰，表现出其稳健性.

定理3.6 设 θ 的先验分布为 $\pi(\theta)$ 和损失函数为**绝对值损失函数**

$$L(\theta, d) = |d - \theta|,$$

则 θ 的贝叶斯估计为

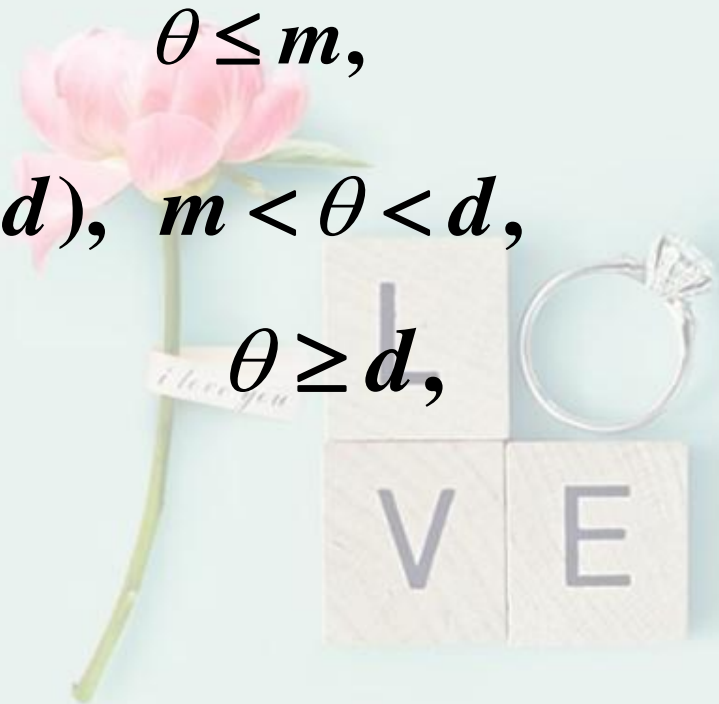
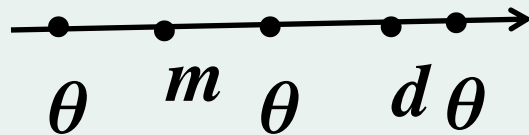
$$d^*(x) = \{\text{后验分布 } h(\theta | x) \text{ 的中位数}\}$$

证 设 m 为 $h(\theta|x)$ 的中位数，又设 $d=d(x)$ 为 θ 的另一估计，为确定起见，先设 $d > m$ ，由绝对损失函数的定义可得

$$d > m$$

$$L(\theta, m) - L(\theta, d) = \begin{cases} m - d, & \theta \leq m, \\ 2\theta - (m + d), & m < \theta < d, \\ d - m, & \theta \geq d, \end{cases}$$

$$= |m - \theta| - |d - \theta|$$



又由于

当 $m < \theta < d$ 时, $2\theta - (m + d) \leq 2d - (m + d) = d - m$

则 $d > m$ 时

$$L(\theta, m) - L(\theta, d) \leq \begin{cases} m - d, & \theta \leq m, \\ d - m, & \theta > m, \end{cases}$$

由于 m 是中位数, 因而 $-P\{\theta \leq m \mid x\} \leq -\frac{1}{2}$

$$P\{\theta \leq m \mid x\} \geq \frac{1}{2}, \quad P\{\theta > m \mid x\} \leq \frac{1}{2},$$

则有 $R(m \mid x) - R(d \mid x) = E(L(\theta, m) - L(\theta, d) \mid x)$

$$\leq (m - d)P\{\theta \leq m \mid x\} + (d - m)P\{\theta > m \mid x\}$$

$$\leq (d - m)[-P\{\theta \leq m \mid x\}] + (d - m)P\{\theta > m \mid x\}$$

$$\leq (m - d) \cdot \frac{1}{2} + (d - m) \cdot \frac{1}{2} = 0$$

于是，当 $d > m$ 时

$$R(m | x) \leq R(d | x)$$

同理可证，当 $d < m$ 时 $R(m | x) \leq R(d | x)$

因而 $d^*(x) = m = \{\text{后验分布 } h(\theta | x) \text{ 的中位数}\}$



定理3.7 设 θ 的先验分布为 $\pi(\theta)$ 和损失函数为**线性损失函数**

$$L(\theta, d) = \begin{cases} k_0(\theta - d), & d \leq \theta, \\ k_1(d - \theta), & d > \theta, \end{cases}$$

则 θ 的贝叶斯估计为

$$d^*(x) = \{\text{后验分布 } h(\theta | x) \text{ 的 } \frac{k_1}{k_1 + k_0} \text{ 上侧分位数}\}$$

证 首先计算任一决策函数 $d(x)$ 的**后验风险**

$$\begin{aligned} R(d | x) &= E[L(\theta, d(x))] = \int_{-\infty}^{+\infty} L(\theta, d(x)) h(\theta | x) d\theta \\ &= \int_{-\infty}^d k_1(d - \theta) h(\theta | x) d\theta + \int_d^{+\infty} k_0(\theta - d) h(\theta | x) d\theta \end{aligned}$$

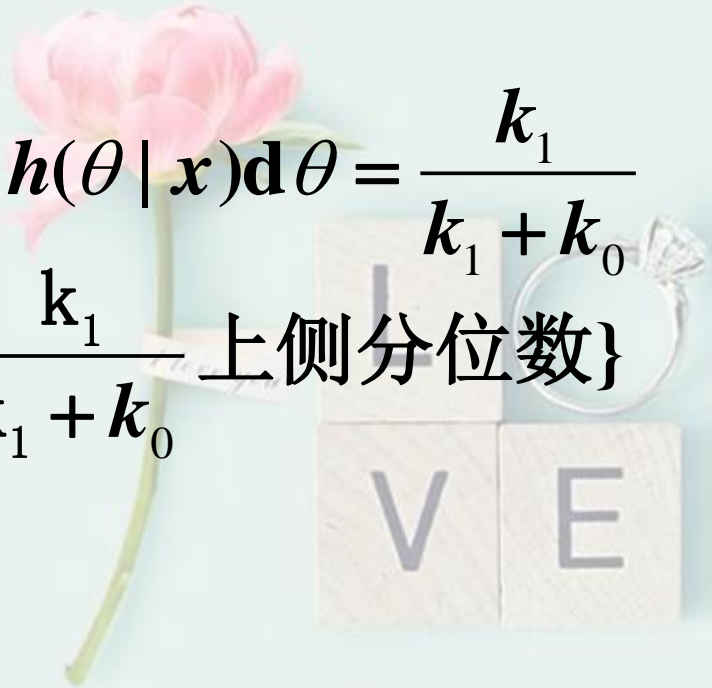
$$\begin{aligned}
 &= \int_{-\infty}^d k_1 (d - \theta) h(\theta | x) d\theta + \int_d^{+\infty} k_0 (\theta - d) h(\theta | x) d\theta \\
 &= (k_1 + k_0) \int_{-\infty}^d (d - \theta) h(\theta | x) d\theta + k_0 (E(\theta | x) - d)
 \end{aligned}$$

为了得到 $R(d|x)$ 的极小值，上式对 d 求导：

$$\begin{aligned}
 \left[\int_{-\infty}^d (d - \theta) h(\theta | x) d\theta \right]'_d &= \int_{-\infty}^d [(d - \theta) h(\theta | x)]'_d d\theta \\
 &= \int_{-\infty}^d h(\theta | x) d\theta
 \end{aligned}$$

$$\frac{dR(d | x)}{d(d)} = (k_1 + k_0) \int_{-\infty}^d h(\theta | x) d\theta - k_0 = 0$$

$$\left[\int_{-\infty}^x f(x, \theta) d\theta \right]'_x = f(x, x) + \int_{-\infty}^x [f(x, \theta)]'_x d\theta$$



即 $\int_{-\infty}^d h(\theta | x) d\theta = \frac{k_0}{k_1 + k_0} \Rightarrow \int_d^{+\infty} h(\theta | x) d\theta = \frac{k_1}{k_1 + k_0}$

则 $d^*(x) = \{\text{后验分布 } h(\theta | x) \text{ 的 } \frac{k_1}{k_1 + k_0} \text{ 上侧分位数}\}$

例7(p101 例3.13) 设总体 X 服从均匀分布 $U(0, \theta)$, 其中参数 θ 未知, 而 θ 服从pareto分布, 其分布函数与密度函数分别为

$$F(\theta) = 1 - \left(\frac{\theta_0}{\theta}\right)^\alpha, \theta \geq \theta_0, \quad \pi(\theta) = \frac{\alpha \theta_0^\alpha}{\theta^{\alpha+1}}, \theta \geq \theta_0,$$

其中 $0 < \alpha < 1$, 和 $\theta_0 > 0$ 为已知, 该分布记为 $Pa(\alpha, \theta_0)$,

θ 的数学期望为 $E(\theta) = \frac{\alpha \theta_0}{\alpha - 1}$, (X_1, X_2, \dots, X_n) 来自总体 X , 损失函数为绝对值损失和平方损失时, 试求参数 θ 的贝叶斯估计?

解 $h(\theta | x) = \frac{q(x | \theta)\pi(\theta)}{m(x)} = \frac{q(x | \theta)\pi(\theta)}{\int_{-\infty}^{+\infty} q(x | \theta)\pi(\theta)d\theta}$

$$= \frac{\left(\prod_{i=1}^n \frac{1}{\theta}\right) \cdot \frac{\alpha \theta_0^\alpha}{\theta^{\alpha+1}}}{\int_{\theta_1}^{+\infty} \left(\prod_{i=1}^n \frac{1}{\theta}\right) \cdot \frac{\alpha \theta_0^\alpha}{\theta^{\alpha+1}} d\theta} = \frac{\frac{\alpha \theta_0^\alpha}{\theta^{n+\alpha+1}}}{\int_{\theta_1}^{+\infty} \frac{\alpha \theta_0^\alpha}{\theta^{n+\alpha+1}} d\theta} \quad (\theta_1 = \max\{x_1, \dots, x_n, \theta_0\})$$

$$= \frac{\frac{\alpha \theta_0^\alpha}{\theta^{n+\alpha+1}}}{\frac{\alpha \theta_0^\alpha}{(\alpha + n)\theta_1^{n+\alpha}}} = \frac{(\alpha + n)\theta_1^{n+\alpha}}{\theta^{n+\alpha+1}}, \quad (\theta > \theta_1)$$

显然 $h(\theta | x)$ 仍为pareto分布 $Pa(\alpha + n, \theta_1)$.

根据定理3.6可知，绝对值损失对应的贝叶斯估计为后验分布的中位数,即

$$P\{\theta \leq \hat{\theta}_B | x\} = \frac{1}{2}, \quad P\{\theta > \hat{\theta}_B | x\} = \frac{1}{2},$$

$$\text{则 } F(\hat{\theta}_B) = 1 - \left(\frac{\theta_1}{\hat{\theta}_B}\right)^{\alpha+n} = \frac{1}{2} \quad d^*(x) = \hat{\theta}_B = 2^{\frac{1}{\alpha+n}} \theta_1$$

根据定理3.2可知，平方损失对应的贝叶斯估计为后验分布的均值,即

$$d^*(x) = \hat{\theta}_\beta = \frac{\alpha + n}{\alpha + n - 1} \theta_1 = \frac{\alpha + n}{\alpha + n - 1} \max\{x_1, \dots, x_n, \theta_0\}$$

θ 的连续函数的贝叶斯估计

定理3.8 设参数 θ 的先验分布为 $\pi(\theta)$, $g(\theta)$ 为 θ 的连续函数, 则在平方损失函数下, $g(\theta)$ 的贝叶斯估计为
$$d(x) = E[g(\theta) | x].$$

证明 在平方损失下, 任一决策函数 $d = d(x)$ 的后验风险为
$$E[(d - g(\theta))^2 | x] = d^2 - 2dE[g(\theta) | x] + E[g^2(\theta) | x].$$

当 $d(x) = E[g(\theta) | x]$ 时, 上述后验风险达到最小. 即在平方损失下, $g(\theta)$ 的贝叶斯估计为 $d(x) = E[g(\theta) | x]$.

例8(p102 例3.14) 设总体 X 服从伽玛分布 $\Gamma(r, \theta)$, 其中参数 r 已知, θ 的先验分布为 $\Gamma(\alpha, \beta)$, (X_1, X_2, \dots, X_n) 来自总体 X , 损失函数取平方损失和损失函数

$$L(\theta, d) = \theta^2 \left(d - \frac{1}{\theta}\right)^2$$

试求参数 θ^{-1} 的贝叶斯估计?

解 由于 $E(X) = \frac{r}{\theta}$, 因此, 人们更感兴趣 θ^{-1} 估计, θ 的后验分布为

$$h(\theta | x) = \frac{q(x | \theta)\pi(\theta)}{m(x)} = \frac{q(x | \theta)\pi(\theta)}{\int_0^{+\infty} q(x | \theta)\pi(\theta)d\theta}$$

$$\begin{aligned}
 &= \frac{\prod_{i=1}^n \frac{\theta^r}{\Gamma(r)} x_i^{r-1} e^{-\theta x_i} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}}{\int_0^{+\infty} \prod_{i=1}^n \frac{\theta^r}{\Gamma(r)} x_i^{r-1} e^{-\theta x_i} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta} \\
 &= \frac{\theta^{nr+\alpha-1} e^{-\theta(\sum_{i=1}^n x_i + \beta)}}{\int_0^{+\infty} \theta^{nr+\alpha-1} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta} \\
 &= \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+nr}}{\Gamma(\alpha + nr)} \theta^{nr+\alpha-1} e^{-\theta(\sum_{i=1}^n x_i + \beta)}
 \end{aligned}$$



则 θ^{-1} 在平方损失下的贝叶斯估计为

$$d^*(x) = \hat{\theta}^{-1} = E(\theta^{-1} | x)$$

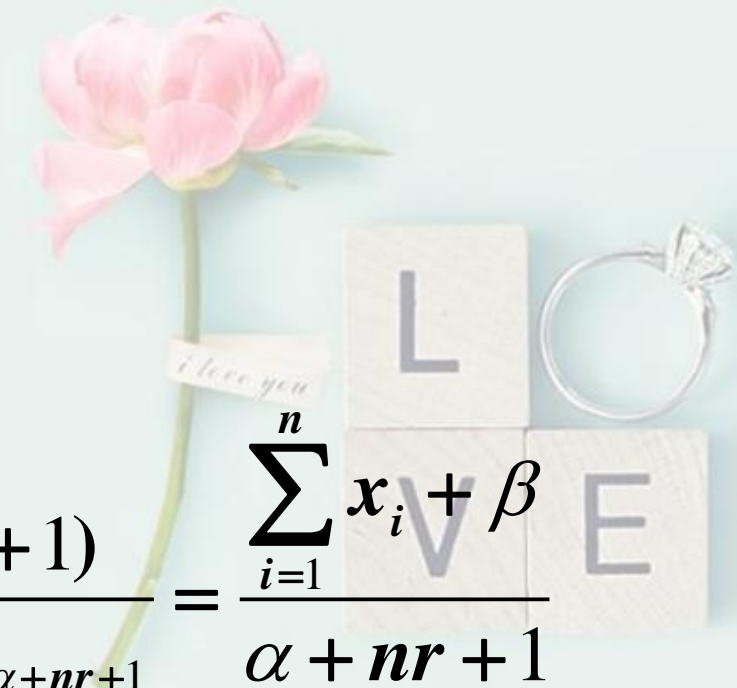
$$\begin{aligned} &= \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha + nr}}{\Gamma(\alpha + nr)} \int_0^{+\infty} \frac{1}{\theta} \cdot \theta^{nr + \alpha - 1} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta \\ &= \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha + nr}}{\Gamma(\alpha + nr)} \cdot \frac{\Gamma(\alpha + nr - 1)}{(\sum_{i=1}^n x_i + \beta)^{\alpha + nr - 1}} = \frac{\sum_{i=1}^n x_i + \beta}{\alpha + nr - 1} \end{aligned}$$

由定理3.3可知, θ^{-1} 在 $L(\theta, d) = \theta^2(d - \frac{1}{\theta})^2$ 下的贝叶斯估计为

$$d^*(x) = \hat{\theta}^{-1} = \frac{E(\theta^2 \cdot \theta^{-1} | x)}{E(\theta^2 | x)}$$

$$= \frac{\frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+nr}}{\Gamma(\alpha+nr)} \int_0^{+\infty} \theta \cdot \theta^{nr+\alpha-1} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta}{\frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+nr}}{\Gamma(\alpha+nr)} \int_0^{+\infty} \theta^2 \cdot \theta^{nr+\alpha-1} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta}$$

$$\begin{aligned}
&= \frac{\int_0^{+\infty} \theta^{nr+\alpha} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta}{\int_0^{+\infty} \theta^{nr+\alpha+1} e^{-\theta(\sum_{i=1}^n x_i + \beta)} d\theta} \\
&= \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+nr+2}}{\Gamma(\alpha + nr + 2)} \cdot \frac{\Gamma(\alpha + nr + 1)}{(\sum_{i=1}^n x_i + \beta)^{\alpha+nr+1}} = \frac{\sum_{i=1}^n x_i + \beta}{\alpha + nr + 1} \\
&= \frac{\alpha + nr - 1}{\alpha + nr + 1} \cdot \frac{1}{\alpha + nr - 1} (\sum_{i=1}^n x_i + \beta)
\end{aligned}$$



四、共轭先验分布

为了使得后验分布计算简单，引入共轭先验分布。

1、共轭分布族

定义3.7 设总体 X 的分布密度为 $p(x|\theta)$, F^* 为 θ 的一个分布族, $\pi(\theta)$ 为 θ 的任意一个先验分布, $\pi(\theta) \in F^*$, 若对样本的任意观测值 x , θ 的后验分布 $h(\theta|x) \in F^*$, 则称 F^* 是关于分布密度 $p(x|\theta)$ 的共轭先验分布族, 简称共轭分布族。



注 1.共轭分布族总是针对分布中的某个参数而言的。

2.共轭分布族可使得数学运算变得简单。因为参数的先验分布和后验分布属于同一个分布族，这对于后面求解贝叶斯估计问题时极其方便。

2、后验分布核

后验分布

$$h(\theta | x) = \frac{q(x | \theta)\pi(\theta)}{m(x)}, \quad (m(x) \text{ 为样本的边缘分布})$$

可以看出， $m(x)$ 不依赖于参数 θ ，因而参数 θ 的后验分布可以写为如下等价形式：

$$h(\theta | x) \propto q(x | \theta)\pi(\theta) = f(x, \theta)$$

则称 $q(x | \theta)\pi(\theta)$ 为后验分布 $h(\theta | x)$ 的核。其中符号 \propto 表示左右两边相差一个不依赖 θ 的常数因子。

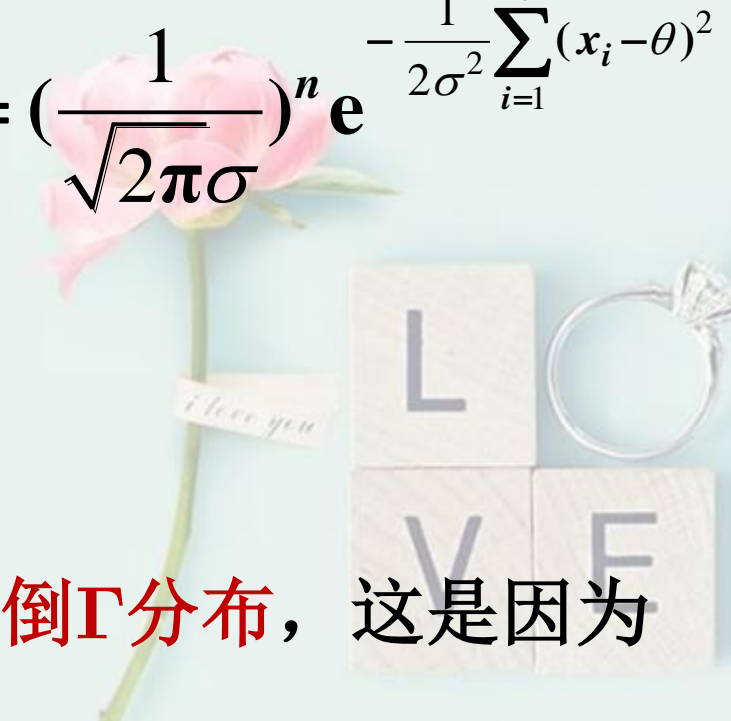
3、共轭先验分布族的构造方法

2种构造方法.

第一种方法 首先计算似然函数 $q(x|\theta)$,根据似然函数所含 θ 的因式情况,选取与似然函数具有相同核的分布作为先验分布.

例3 (p92例3.8) 设 $(X_1, X_2, \dots, X_n)^T$ 是来自正态总体 $N(\theta, \sigma^2)$ 的一个样本,其中 θ 已知,现寻求 σ^2 的共轭先验分布,由于该样本的似然函数为

$$q(x | \theta) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \theta)^2}{2\sigma^2}} \right] = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}$$

$$\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}$$


哪一个分布具有上述核？结论是**倒Γ分布**，这是因为Γ分布的密度函数为

$$f(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

设 $Y = X^{-1}$, 则 Y 的密度函数为

$$f(y; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{y}\right)^{\alpha+1} e^{-\frac{\beta}{y}}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

此分布密度为倒 Γ 分布的密度函数。

设 σ 的先验分布为倒 Γ 分布, 即

$$\pi(\sigma^2) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{\beta}{\sigma^2}}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

则 σ 的后验分布为

$$h(\sigma^2 | x) \propto q(x | \sigma^2) \pi(\sigma^2)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\alpha + \frac{n}{2} + 1} \mathbf{e}^{-\frac{1}{\sigma^2} [\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2]}$$

显然此分布仍为倒 Γ 分布，即先验分布与后验分布都为倒 Γ 分布，因而倒 Γ 分布是 σ 的共轭先验分布族。

例3 (p93例3. 9) 设 $(X_1, X_2, \dots, X_n)^T$ 是来自总体 $B(N, \theta)$ 的一个样本, 现寻求 θ 的共轭先验分布, 由于该样本的似然函数为

$$q(x | \theta) = \prod_{i=1}^n \left[C_N^{x_i} \theta^{x_i} (1 - \theta)^{N - x_i} \right]$$
$$\propto \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{nN - \sum_{i=1}^n x_i}, \quad x_i = 0, 1, \dots, N$$

哪一个分布具有上述核? 结论是 β 分布, 这是因为 β 分布的密度函数为

$$f(x; \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{其他} \end{cases}$$

其期望 $EX = \frac{\alpha}{\alpha + \beta}$, $DX = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

设 θ 的先验分布为 β 分布, 即

$$\pi(\theta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, & 0 < \theta < 1, \\ 0, & \text{其他} \end{cases}$$

则 θ 的后验分布为

$$h(\theta | x) \propto q(x | \theta) \pi(\theta)$$

$$\propto \theta^{\alpha + \sum_{i=1}^n x_i - 1} (1 - \theta)^{\beta + nN - \sum_{i=1}^n x_i - 1}, \quad 0 < \theta < 1$$

显然此分布仍然是 β 分布的核，因而 β 分布是 θ 的共轭先验分布族。经计算可知

$$h(\theta | x) = \beta(\alpha + \sum_{i=1}^n x_i, \beta + nN - \sum_{i=1}^n x_i)$$



第二种方法 设总体 X 的分布密度为 $p(x|\theta)$,统计量

$T(X) = T(X_1, X_2, \dots, X_n)$ 是参数 θ 的充分统计量, 则有
定理3.1 设 $f(\theta)$ 为任一固定的函数, 满足条件

$$(1) f(\theta) \geq 0, \theta \in \Theta,$$

$$(2) 0 < \int_{\Theta} g_n(t|\theta) f(\theta) d\theta < \infty$$

则

$$D_f = \left\{ \frac{g_n(t|\theta) f(\theta)}{\int_{\Theta} g_n(t|\theta) f(\theta) d\theta} : n = 1, 2, \dots \right\}$$

是共轭先验分布族, 其中

$$q(x|\theta) = \prod_{i=1}^n p(x_i|\theta) = g_n(t|\theta) h(x_1, x_2, \dots, x_n)$$

例4 (p94例3. 10) 设 $(X_1, X_2, \dots, X_n)^T$ 是来自总体 $B(1, \theta)$ 的一个样本, 试寻求 θ 的共轭先验分布?

解 其似然函数为

$$\begin{aligned} q(x | \theta) &= \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= g_n(t | \theta) \cdot 1, \end{aligned}$$

其中 $g_n(t | \theta) = \theta^t (1-\theta)^{n-t}$, 选取 $f(\theta) = 1$, 则

$$D_f = \left\{ \frac{\theta^t (1-\theta)^{n-t}}{\int_0^1 \theta^t (1-\theta)^{n-t} d\theta} : n = 1, 2, \dots, t = 0, 1, 2, \dots \right\}$$

显然此共轭分布族为 β 分布的子族, 因而, 两点分布的共轭先验分布族为 β 分布.

常见共轭先验分布

总体分布	参数	共轭先验分布
二项分布	成功概率 p	β 分布 $\beta(\alpha, \beta)$
泊松分布	均值 λ	Γ 分布 $\Gamma(\alpha, \beta)$
指数分布	均值的倒数 λ	Γ 分布 $\Gamma(\alpha, \beta)$
正态分布 (方差已知)	均值 μ	正态分布 $N(\mu, \sigma^2)$
正态分布 (均值已知)	方差 σ^2	倒 Γ 分布

五、贝叶斯估计的误差

在计算 θ 的贝叶斯估计时，用到了 θ 的后验分布，因此考察估计值与真实值之间的误差时，也应考虑 θ 的后验分布，误差定义如下：

定义3.10 参数 θ 的后验分布为 $h(\theta|x)$ ，其贝叶斯估计为 $\hat{\theta}$ ，则 $(\hat{\theta} - \theta)^2$ 的后验期望为

$$MSE(\hat{\theta} | x) = E_{\theta|x} (\hat{\theta} - \theta)^2$$

称其为 $\hat{\theta}$ 的后验均方差，而其平方根 $[MSE(\hat{\theta} | x)]^{\frac{1}{2}}$ 称为后验标准误差，其中符号 $E_{\theta|x}$ 表示对条件分布 $h(\theta | x)$ 求期望。

后验均方差与后验方差的关系

1、当 $\hat{\theta} = E(\theta | x)$ 时，则后验均方差为

$$MSE(\hat{\theta} | x) = E_{\theta|x} (E(\theta | x) - \theta)^2 = \text{Var}(\theta | x)$$

2、当 $\hat{\theta} = E(\theta | x)$ 时，则后验均方差达到最小，
因而后验均值 $E(\theta | x)$ 是较好的贝叶斯估计。这是因为

$$\begin{aligned} MSE(\hat{\theta} | x) &= E_{\theta|x} [(\hat{\theta} - E(\theta | x) + (E(\theta | x) - \theta))^2] \\ &= E_{\theta|x} [\hat{\theta} - E(\theta | x)]^2 + \text{Var}(\theta | x) \\ &= [\hat{\theta} - E(\theta | x)]^2 + \text{Var}(\theta | x) \geq \text{Var}(\theta | x) \end{aligned}$$

后验均方差与后验方差的优点

- 1、二者只依于样本，不依赖参数 θ .
- 2、二者的计算不依赖于统计量的分布，即抽样分布
- 3、贝叶斯估计不考虑无偏性，因为贝叶斯估计只考虑出现的样本，不考虑没出现的样本.

4、贝叶斯区间估计

定义 当 θ 为连续型随机变量时, 给定 $1-\alpha$, 当

$$P\{a \leq \theta \leq b \mid x\} = 1-\alpha,$$

则称区间 $[a, b]$ 为参数 θ 的贝叶斯区间估计.

定义 当 θ 为离散型随机变量时, 给定 $1-\alpha$, 当

$$P\{a \leq \theta \leq b \mid x\} > 1-\alpha,$$

则称区间 $[a, b]$ 为参数 θ 的贝叶斯区间估计.

定义3.11 设参数 θ 的后验分布为 $h(\theta|x)$ ，对给定的样本 $X = (X_1, X_2, \dots, X_n)^T$ 和概率 $1-\alpha$ ($0 < \alpha < 1$)，若存在两个统计量， $\hat{\theta}_L = \hat{\theta}_L(X)$ 和 $\hat{\theta}_U = \hat{\theta}_U(X)$ ，使得

$$P\{\hat{\theta}_L < \theta < \hat{\theta}_U \mid x\} \geq 1 - \alpha$$

则称区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 为参数 θ 的置信度为 $1-\alpha$ 的贝叶斯置信区间，或简称为 θ 的 $1-\alpha$ 的置信区间，而满足 $P\{\theta \geq \hat{\theta}_L \mid x\} \geq 1-\alpha$ 的 $\hat{\theta}_L$ 为 θ 的 $1-\alpha$ 单侧置信下限，满足 $P\{\theta \leq \hat{\theta}_U \mid x\} \geq 1-\alpha$ 的 $\hat{\theta}_U$ 为 θ 的 $1-\alpha$ 单侧置信上限.

注 贝叶斯置信区间依赖于先验分布，不需要抽样分布，计算相对简单.

正态分布均值的贝叶斯置信区间

例9 (p104例3.15) 设 $X = (X_1, X_2, \dots, X_n)^T$ 是来自正态总体 $N(\theta, \sigma^2)$ 的一个样本，其中 σ^2 已知，取 θ 的先验分布为正态分布 $N(\mu, \tau^2)$ ，参数 μ, τ^2 已知，试求参数 θ 的置信度为 $1-\alpha$ 的置信区间.

解 θ 的先验分布为

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{1}{2\tau^2}(\theta - \mu)^2\right\}$$

则样本 X 与 θ 的联合密度函数为

$$f(x, \theta) = k_1 \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(n\theta^2 - 2n\theta\bar{x} + \sum_{i=1}^n x_i^2 \right) + \frac{1}{\tau^2} (\theta^2 - 2\mu\theta + \mu^2) \right] \right\}$$

$$k_1 = (2\pi)^{-(n+1)/2} \tau^{-1} \sigma^{-n}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \sigma_0^2 = \frac{\sigma^2}{n},$$

$$A = \sigma_0^{-2} + \tau^{-2}, \quad B = \bar{x}\sigma_0^{-2} + \mu\tau^{-2}, \quad C = \sigma^{-2} \sum_{i=1}^n x_i^2 + \mu^2\tau^{-2}$$

则有

$$f(x, \theta) = k_1 \exp \left\{ -\frac{1}{2} [A\theta^2 - 2\theta B + C] \right\} = k_2 \exp \left\{ -\frac{1}{2A^{-1}} (\theta - B/A)^2 \right\}$$

$$\text{其中 } k_2 = k_1 \exp \left\{ -\frac{1}{2} (C - B^2/A) \right\}.$$

因此样本 x 的边缘分布为

$$m(x) = \int_{-\infty}^{+\infty} f(x, \theta) d\theta = k_2 \left(\frac{2\pi}{A} \right)^{\frac{1}{2}}$$

因而 θ 的后验分布为

$$h(\theta | x) = \frac{f(x, \theta)}{m(x)} = \left(\frac{A}{2\pi} \right)^{\frac{1}{2}} \exp \left\{ -\frac{(\theta - B/A)^2}{2/A} \right\}$$

这正好是一个正态分布的密度函数。其中

$$\mu_1 = \frac{B}{A} = \frac{\bar{x}\sigma_0^{-2} + \mu\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}}, \sigma_1^2 = A^{-1} = \sigma_0^{-2} + \tau^{-2} = \frac{\sigma_0^2\tau^2}{\sigma_0^2 + \tau^2}$$

据此可知 $\frac{\theta - \mu_1}{\sigma_1}$ 服从标准正态分布，于是可得

$$P\left\{\left|\frac{\theta - \mu_1}{\sigma_1}\right| \leq u_{\alpha/2}\right\} = 1 - \alpha,$$

即

$$P\{\mu_1 - \sigma_1 u_{\alpha/2} \leq \theta \leq \mu_1 + \sigma_1 u_{\alpha/2}\} = 1 - \alpha$$

故可得 θ 的 $1 - \alpha$ 贝叶斯置信区间为

$$[\mu_1 - \sigma_1 u_{\alpha/2}, \mu_1 + \sigma_1 u_{\alpha/2}]$$

$$\mu_1 = \frac{B}{A} = \frac{\bar{x}\sigma_0^{-2} + \mu\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}} \quad \sigma_1^2 = A^{-1} = \sigma_0^{-2} + \tau^{-2} = \frac{\sigma_0^2\tau^2}{\sigma_0^2 + \tau^2}$$

例10 (p105例3. 16) 对某儿童进行智力测验，设测验结果服从 $N(\theta, 100)$, 其中 θ 为心理学中儿童的智商， θ 的先验分布为 $N(100, 225)$, 试求 θ 的置信度为0.95的贝叶斯置信区间.

解 将相关数据代入上述置信区间公式可得： θ 的置信度为0.95的置信区间为

$$[94.07, 126.69]$$

而用经典方法可得 θ 的置信度为0.95的置信区间为

$$[95.4, 134.6]$$

Thank You!

