第3章
先验分布寻求方法

§ 3.1　先验分布类型已知时超参数估计

我们知道先验分布中所含的未知参数称为超参数，而先验信息最充分的情形就是先验分布所属的分布族已知，例如，已知二项分布中成功概率θ的先验分布属于贝塔分布$\mathrm{Beta}(\alpha,\beta)$族（一般地，如果已知共轭先验，先验分布所属的分布族当然已知），这时，只要把超参数估计出来就可以确定先验分布了。估计超参数其实也就是经典统计中的参数估计，因此，可以利用经典统计中的各种方法。我们通过一些例子来说明此时先验分布的确定。

例 3.1 （利用先验矩确定先验分布）某学生报的编辑打算做一个对当前学生会主席的支持率的调查，她需要确定学生会主席的支持率θ的先验分布。根据以往的经验, 她相信可取均值是 0.5，标准差是 0.15 的贝塔分布$\mathrm{Beta}(\alpha,\beta)$作为先验，但她是文科生，没有学过贝叶斯统计，请你帮助她确定符合她的先验信念的先验分布。

解：由贝塔分布的性质和所给条件，可以得到超参数的联立方程组

$$\frac{\alpha}{\alpha+\beta}=0.5,\quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=0.15^2$$

解之得 $\alpha=\beta=5.05$，即先验分布为贝塔分布 $\mathrm{Beta}(5.05, 5.05)$。

注：

1. 如根据历史数据整理、加工可获得支持率（成功率或比率等）θ 的若干相当于样本的估计值 $\theta_1,\theta_2,\ldots,\theta_k$，则可算得先验均值 $\bar{\theta}$ 和先验方差 S_θ^2，其中

$$\bar{\theta}=\frac{1}{k}\sum_{i=1}^{k}\theta_i\,,\quad S_\theta^2=\frac{1}{k-1}\sum_{i=1}^{k}(\theta_i-\bar{\theta})^2$$

然后令其分别等于贝塔分布 $\mathrm{Beta}(\alpha,\beta)$ 的期望与方差，即

$$\frac{\alpha}{\alpha+\beta}=\bar{\theta},\quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=S_\theta^2$$

解这个联立方程组，即可得超参数 α 与 β 的先验矩估计值

$$\hat{\alpha}=\bar{\theta}\left(\frac{(1-\bar{\theta})\bar{\theta}}{\mathrm{S}_\theta^2}-1\right),\quad \hat{\beta}=(1-\bar{\theta})\left(\frac{(1-\bar{\theta})\bar{\theta}}{\mathrm{S}_\theta^2}-1\right)$$

例 3.2（利用先验分位数确定先验分布）设某总体分布的参数θ的先验信息根据历史资料知：先验中位数为 0；先验分布的 0.25 分位数和 0.75 分位数分别为-1 和 1。（1）如果先验分布为正态分布$N(\mu,\tau^2)$，试求出具体先验分布。（2）如果先验分布为柯西分布$Cauchy(\alpha,\beta)$，具体的先验分布又是什么？

解：（1）因为$\theta \sim N(\mu,\tau^2)$，所以估计出超参数$\mu$和$\tau^2$即可。由于正态分布是对称的，故均值和中位数相等，从而μ=0。另外由 0.75 分位数为 1 这个已知条件，可列出方程

$$P(\theta<1)=0.75 \text{或} P(\theta/\sigma<1/\sigma)=0.75$$

其中$\theta/\sigma^2 \sim N(0,1)$，查标准正态分布表或用 R 命令 qnorm(0.75)可求得

$$1/\sigma=0.6745 \text{即} \sigma=1.48$$

这样就得到先验分布为$N(0,1.48^2)$。

（2）若设θ的先验分布为柯西分布$Cauchy(\alpha,\beta)$，则其密度函数为

$$\pi(\theta|\alpha,\beta)=\frac{\beta}{\pi(\beta^2+(\theta-\alpha)^2)},\ -\infty<\theta<\infty$$

由于柯西密度函数是关于α的对称函数，其中位数是α。由已知条件得α=0。另外，由 0.25 分位数为-1，可得方程

$$\int_{-\infty}^{-1}\frac{\beta}{\pi(\beta^2+\theta^2)}d\theta=\frac{1}{\pi}\arctan(-\frac{1}{\beta})+0.5=0.25$$

由此可解得$\beta=1$，即θ的先验分布为标准柯西分布$Cauchy(0,1)$。

注：

1. 本例中，问题（1）似乎没用上先验信息“0.25 分位数为-1”，而问题（2）似乎没用上先验信息“0.75 分位数为 1”，这是由于正态分布和柯西分布都是关于中位数对称的而且这里先验中位数为 0，所以没用上的先验信息在问题（1）和（2）中都自然被各自的先验分布满足了。

2. 这样一来，相同的先验信息有二个先验分布可供选择，那么到底选择哪一个呢？一般而言，假如二个先验分布差异不大，那么对于同样的总体及其样本，由贝叶斯公式算出的后验分布差别也不大，因此可任选一个作先验。假如二个先验分布差异大，我们则应慎重选择，因为不同的选择对后验分布的影响可能很大。在本例中，虽然柯西分布 $Cauchy(0,1)$ 因为厚尾连数学期望都不存在，但是它与正态分布 $N(0,1.48^2)$ 在形状上还是很相似的，都是关于中位数对称的钟型曲线（图 3.1），因此在用同样的总体及其样本进行更新后所得后验密度差别不大（图 3.2）。所以，除非还有另外的先验信息，否则，二个先验可以任选一个。作为比较，现在取均匀分布 $U(1,3)$ 为先验，总体及其样本不变，这时更新所得后验密度就与本例完全不同（图 3.3），这是因为这时的先验与本例的先验完全不同所造成的。最后，图 3.2 和图 3.3 所用 R 程序如下（读者应动手做一下）：

```
library(BayesianStat)
x<-rnorm(500,0,6)
Normgd(x, sigma.x=6, density = "normal" , params = c(0,1.48), n.mu = 100)
prior.dens<- dcauchy(seq(-4,4, by=0.1),0,1)
Normgd(x,sigma.x=6,density="user",mu=seq(-4,4,by=0.1),mu.prior=prior.dens)
Normgd(x, sigma.x=6, density = "uniform" , params = c(1,3), n.mu = 100)
```

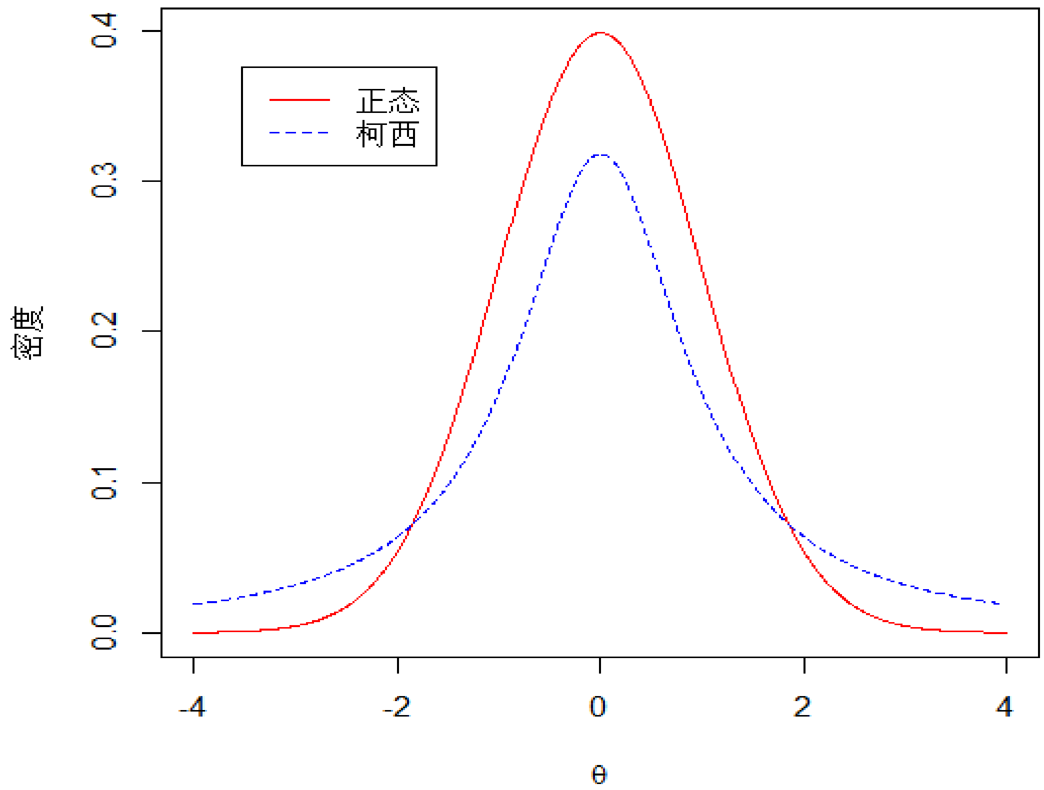

图 3.1 正态密度与柯西密度比较图

图 3.1 正态密度与柯西密度比较图

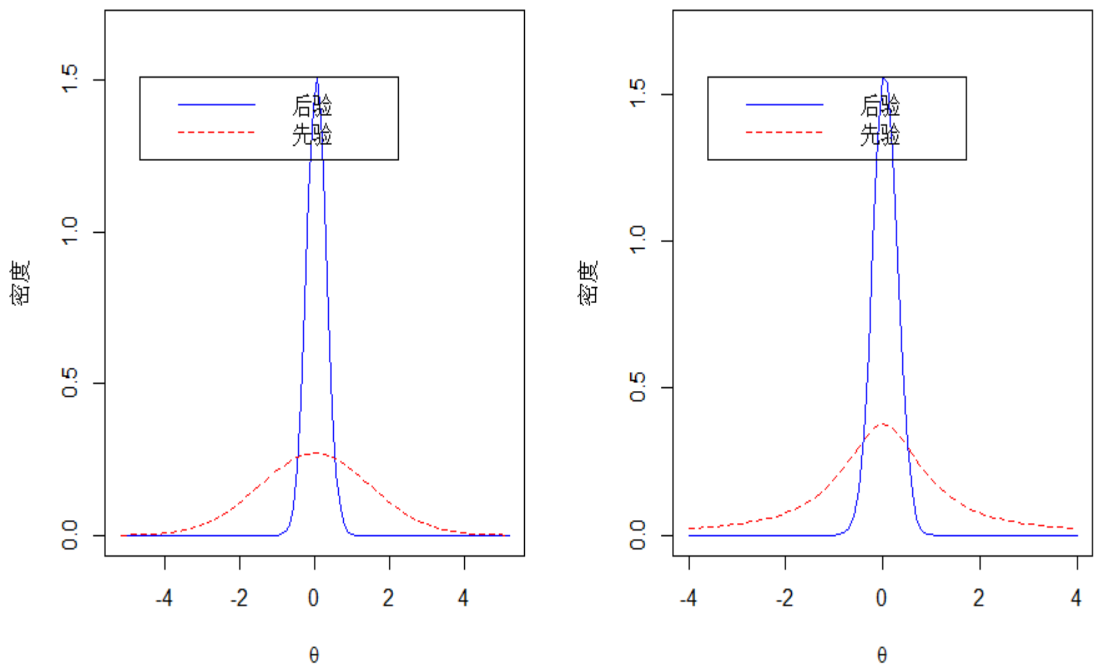

图 3.2 正态先验（左图）与柯西先验（右图）更新的后验比较图

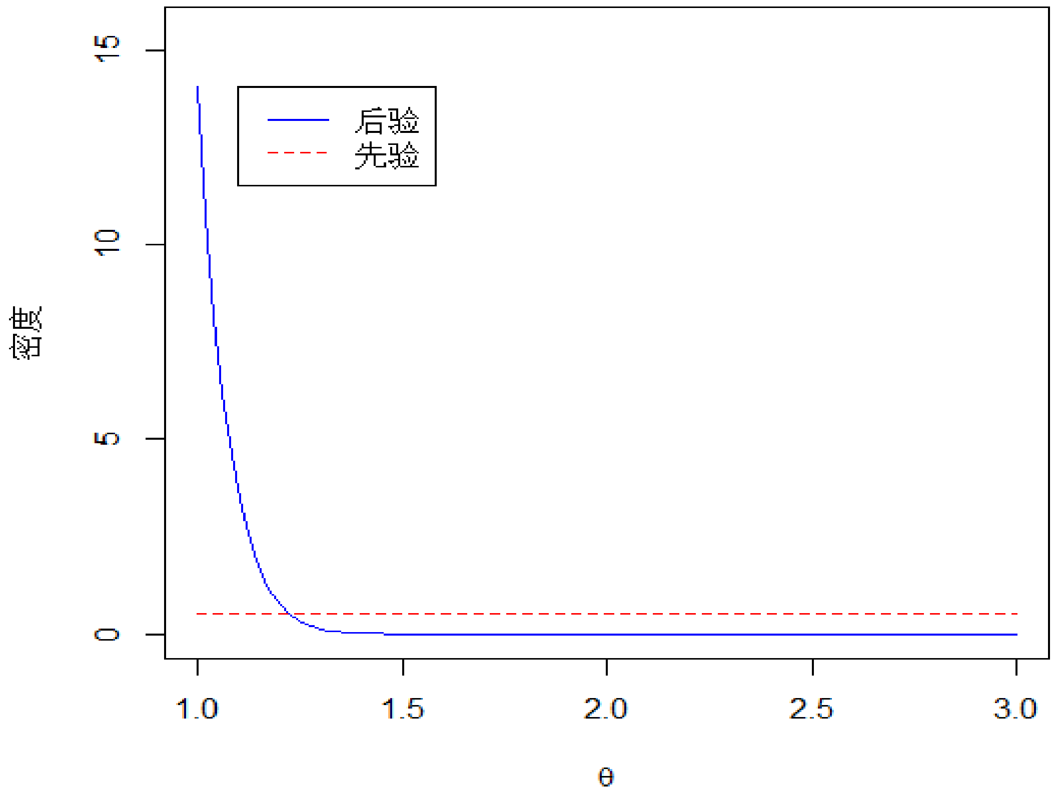

图 3.3 均匀分布先验更新的后验分布图

§ 3.2　由边际分布确定先验分布

另一种先验信息不是直接关于总体分布的未知参数 θ 的，而是关于贝叶斯公式中边际（缘）分布的。我们来讨论此时如何确定先验分布。

如果总体 X 的密度函数为 $p(x|\theta)$，它含有未知参数 θ，我们知道在贝叶斯公式（这里样本量为 1）

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

中，$m(x)$ 是 X 的边际分布，并且按照 θ 是连续的还是离散的而写成

$$m(x)=\begin{cases}\int_{\Theta} p(x\mid\theta)\pi(\theta)d\theta \\ \sum_{\theta_i\in\Theta} p(x\mid\theta_i)\pi(\theta_i)\end{cases}$$

当先验分布含有未知超参数（向量）λ 时，先验 $\pi(\theta)$ 实际上是 $\pi(\theta\mid\lambda)$，从而边际分布 $m(x)$ 也依赖于 λ，

故应记为 $m(x|\lambda)$。这样，如果我们有与边际分布有关的先验信息，就有可能把先验 $\pi(\theta\mid\lambda)$ 确定下来。

3.2.1 混合分布与混合样本

设 $q_k > 0\ (k=1,2,\cdots K)$ 且 $\sum_{k=1}^{K} q_k = 1$，又设 $F(x|\theta_k)\ (k=1,2,\cdots,K)$ 是 K 个分布而 $p(x|\theta_k)\ (k=1,2,\cdots,K)$ 是对应的密度，令

$$F(x)=\sum_{k=1}^{K} q_k F(x|\theta_k),\ p(x)=\sum_{k=1}^{K} q_k p(x|\theta_k)$$

则不难看出 $F(x)$ 也是一个分布，并被称为 $F(x|\theta_k)\ (k=1,2,\cdots,K)$ 的混合分布，而 $p(x)$ 则是对应的混合密度。如果我们定义一个取值为 $\{\theta_k; k=1,2,\cdots,K\}$ 的随机变量 θ 的分布 $\pi(\theta)$ 为

$$\pi(\theta_k)=q_k,\ k=1,2,\cdots,K$$

则

$$F(x)=\sum_{k=1}^{K} \pi(\theta_k) F(x|\theta_k),\ p(x)=\sum_{k=1}^{K} \pi(\theta_k) p(x|\theta_k)$$

所以从混合分布 $F(x)$ 中抽取一个样本 x_1，相当于如下二步抽样：

第 1 步，从 $\pi(\theta)$ 抽取一个 θ；

第 2 步，若 $\theta=\theta_k$，则从 $F(x|\theta_k)$ 中再抽一个样本 x_1。

如此反复进行下去，我们就可以从混合分布 $F(x)$ 抽取一个容量为 n 的样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$，这样的样本被称为混合样本，显然其中约有 $[n\pi(\theta_k)]$（表示方括号里那个数的整数部分）个单样本来自分布 $F(x|\theta_k)$

从贝叶斯公式中的边际分布 $m(x)$ 的公式可以看出它实际上是混合分布的推广。当 θ 为离散随机变量时，$m(x)$ 是由有限个或可数无限个密度函数混合而成，当 θ 为连续随机变量时，$m(x)$ 是由不可数无限个密度函数混合而成。

例 3.3 列举混合样本的两个例子

解： （1）一批产品来自三位工人之手，三位工人产品占比分别为 30%，40%，30%。现在设 θ 是随机变量，可能取值为 1，2，3（分别代表三位工人）而且概率分布为

$$\pi(1)=P(\theta=1)=0.3,\quad \pi(2)=P(\theta=2)=0.4,\quad \pi(3)=P(\theta=3)=0.3$$

又设三位工人生产的产品长度分别服从分布 $p(x|\theta=i)$。现在随机抽取 n 件产品并测得长度分别为 $x_1,x_2,\cdots,x_n$，则样本 $x_1,x_2,\cdots,x_n$ 就可以看作是一个混合样本，来自混合分布

$$p(x)=\sum_{i=1}^{3}p(x|i)\pi(i)$$

（2）设一个大学生的贝叶斯统计课程的考试成绩服从分布 $p(x|\theta)$，其中 θ 是其考试能力参数。依据贝叶斯统计的思想，我们可以假设 θ 服从某个分布 $\pi(\theta)$ 。现在假设有 n 位同学参加某次贝叶斯统计课程的考试，他们的考试能力当然是不同的，假设分别是 $\theta_1,\theta_2 \cdots \theta_n$，那么，$\theta_1,\theta_2 \cdots \theta_n$ 就可看成来自分布 $\pi(\theta)$ 的样本， $p(x|\theta_k)$ 就是第 k 位同学的成绩分布密度。最后，假设这些同学考完的成绩是 $x_1,x_2,\cdots,x_n$ ， 那么， x_i 就可看成是从 $p(x|\theta_i)$ 抽取的样本。这样一来，整个样本 $x_1,x_2,\cdots,x_n$ 就可看作是混合样本。

3.2.2 寻求先验密度的 II 型最大似然法

在边际分布 $m(x)$ 的公式中，除了样本密度 $p(x|\theta)$ 外，就是先验 $\pi(\theta)$，如果 $p(x|\theta)$ 已知，则 $m(x)$ 就依赖于先验 $\pi(\theta)$，因而可记 $m(x)=m(x|\pi)$。现在设所有可能的先验密度的集合为 $\Psi=\{\pi,\pi\text{为先验}\}$（简称为先验族），混合样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 已知，则类似于经典统计中最大似然估计的思想方法，可以把 $m(\mathbf{x}|\pi)$ 看作先验 π 的似然函数，如果对二个不同的先验 $\pi_1\in\Psi$，$\pi_2\in\Psi$ 有

$$m(\mathbf{x}|\pi_1)>m(\mathbf{x}|\pi_2)$$

则可认为当先验取 π_1 时，样本 $\mathbf{x}$ 出现的可能性比先验取 π_2 时大，于是，我们就是要去求那个使 $m(\mathbf{x}|\pi)$ 最大的 π。这种思想方法称为II型最大似然法（或 ML—II 法），所求出的先验称为II型最大似然先验（或 ML—II 先验）。也就是说，II 型最大似然先验 $\hat{\pi}$ 是以下方程的解

$$m(\mathbf{x}|\hat{\pi})=\sup_{\pi\in\Psi}\{m(\mathbf{x}|\pi)\}$$

如果先验族 Ψ 的先验密度函数的形式已知，未知的仅是其中的超参数，即先验密度函数族可以表示如下

$$\Psi = \{\pi(\theta|\lambda), \lambda \in \Lambda\}$$

其中，Λ 是超参数集，这时寻求 ML—II 先验就是寻求这样的超参数 $\hat{\lambda}$ 使得

$$m(\mathbf{x}|\hat{\lambda}) = \sup_{\lambda\in\Lambda}\{m(\mathbf{x}|\lambda)\}$$

从而 ML—II 先验是 $\pi(\theta|\hat{\lambda})$。

注：一般而言，混合样本 $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ 是简单随机样本，因而有

$$m(\mathbf{x}|\hat{\lambda}) = \sup_{\lambda\in\Lambda}\{m(\mathbf{x}|\lambda)\} = \sup_{\lambda\in\Lambda}\left\{\prod_{i=1}^{n} m(x_i|\lambda)\right\}$$

但是，因为

$$m(\mathbf{x} | \boldsymbol{\lambda}) = \int_\Theta p(\mathbf{x} | \theta)\pi(\theta | \boldsymbol{\lambda})d\theta,\ m(x_i | \boldsymbol{\lambda}) = \int_\Theta p(x_i | \theta)\pi(\theta | \boldsymbol{\lambda})d\theta$$

所以，上式最右部分不一定更简单。

例 3.4 设 $X \sim N(\theta, \sigma^2)$，其中 σ^2 已知，又设均值参数 $\theta \sim N(\mu_\pi, \sigma_\pi{}^2)$，其中 $\boldsymbol{\lambda} = (\mu_\pi, \sigma_\pi{}^2)$ 为未知超参数向量，而 $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ 是来自边际分布 $m(x \mid \boldsymbol{\lambda})$ 的混合样本。试求超参数 $\boldsymbol{\lambda} = (\mu_\pi, \sigma_\pi{}^2)$。

解：不难计算知 $m(x \mid \boldsymbol{\lambda}) = m(x \left| \mu_\pi, \sigma_\pi^2 \right.)$ 是正态分布 $N(\mu_\pi, \sigma_\pi{}^2 + \sigma^2)$。

所以当样本 $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ 已知时，超参数 $\boldsymbol{\lambda} = (\mu_\pi, \sigma_\pi{}^2)$ 的似然函数为

$$m(\mathbf{x} \left| \boldsymbol{\lambda} \right.) = m(\mathbf{x} \left| \mu_\pi, \sigma_\pi^2 \right.) = [2\pi(\sigma_\pi^2 + \sigma^2)]^{-\frac{n}{2}} \exp\left\{ -\frac{\sum (x_i - \mu_\pi)^2}{2(\sigma_\pi^2 + \sigma^2)} \right\}$$

$$= [2\pi(\sigma_\pi^2 + \sigma^2)]^{-\frac{n}{2}} \exp\left\{ \frac{-ns^2}{2(\sigma_\pi^2 + \sigma^2)} \right\} \exp\left\{ -\frac{n(\overline{x} - \mu_\pi)^2}{2(\sigma_\pi^2 + \sigma^2)} \right\}$$

其中 $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}),\ s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$

于是，对数似然函数为

$$l(x\left|\mu_\pi,\sigma_\pi^2\right.)=\ln[m(x\left|\mu_\pi,\sigma_\pi^2\right.)]$$

$$=-\frac{n}{2}\ln[2\pi(\sigma_\pi^2+\sigma^2)]-\frac{ns^2}{2(\sigma_\pi^2+\sigma^2)}-\frac{n(\bar{x}-\mu_\pi)^2}{2(\sigma_\pi^2+\sigma^2)}$$

将对数似然函数求偏导数并令其为零，得似然方程

$$\begin{cases}\dfrac{\partial l}{\partial\mu_\pi}=\dfrac{n(\bar{x}-\mu_\pi)}{(\sigma_\pi^2+\sigma^2)}=0\\ \dfrac{\partial l}{\partial\sigma_\pi^2}=-\dfrac{n\sigma_\pi}{(\sigma_\pi^2+\sigma^2)}+\dfrac{ns^2\sigma_\pi}{(\sigma_\pi^2+\sigma^2)^2}+\dfrac{n(\bar{x}-\mu_\pi)^2\sigma_\pi}{(\sigma_\pi^2+\sigma^2)^2}=0\end{cases}$$

解似然方程，可得超参数

$$\begin{cases}\mu_\pi=\bar{x}\\ \sigma_\pi^2=s^2-\sigma^2\end{cases}$$

从而所求的 ML—II 先验为正态分布 $N(\bar{x},s^2-\sigma^2)$。

3.2.3 寻求先验密度的边际矩法

如果来自边际分布 $m(x|\lambda)$ 的混合样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 给定，那么当先验分布 $\pi(\theta|\lambda)$ 的形式已知时，我们还可利用先验矩与边际分布矩之间的关系寻求超参数 λ 的估计 $\hat{\lambda}$，从而获得先验分布 $\pi(\theta|\hat{\lambda})$。这个方法的要点是：首先将边际分布的矩表示成超参数的函数，得到一个方程或方程组，然后将边际分布的矩用相应的混合样本矩代替，这样就得到以超参数为未知量的方程或方程组，解之就得到超参数的估计量 $\hat{\lambda}$，从而获得先验分布 $\pi(\theta|\hat{\lambda})$。其具体的三步骤如下。

（1）计算总体分布 $p(x|\theta)$ 的期望 $\mu(\theta)$ 和方差 $\sigma^2(\theta)$，即

$$\mu(\theta)=E^{x|\theta}(X)=\int_{\chi}xp(x|\theta)dx,\ \sigma^2(\theta)=E^{x|\theta}[X-\mu(\theta)]^2$$

这里符号 $E^{x|\theta}$ 表示对条件分布 $p(x|\theta)$ 求期望（以下类似符号解释类推）。

（2）计算边际密度 $m(x|\lambda)$ 的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m{}^2(\lambda)$。首先

$$\begin{aligned}\mu_m(\lambda)&=E^{x|\lambda}(X)=\int_{\chi}xm(x\mid\lambda)dx=\int_{\chi}x\int_{\Theta}p(x|\theta)\pi(\theta|\lambda)d\theta dx\\&=\int_{\Theta}\int_{\chi}xp(x|\theta)dx\pi(\theta|\lambda)d\theta=\int_{\Theta}\mu(\theta)\pi(\theta|\lambda)d\theta\\&=E^{\theta|\lambda}[\mu(\theta)]\end{aligned}$$

其次，因为

$$\begin{aligned}E^{x|\theta}[x-\mu_m(\lambda)]^2&=E^{x|\theta}[x-\mu(\theta)+\mu(\theta)-\mu_m(\lambda)]^2\\&=E^{x|\theta}[x-\mu(\theta)]^2+E^{x|\theta}[\mu(\theta)-\mu_m(\lambda)]^2\\&=\sigma^2(\theta)+[\mu(\theta)-\mu_m(\lambda)]^2\end{aligned}$$

所以

$$\begin{aligned}
\sigma_m^2(\lambda) &= E^{x|\lambda}[X-\mu_m(\lambda)]^2 = \int_{\chi}[x-\mu_m(\lambda)]^2 m(x|\lambda)dx \\
&= \int_{\chi}[x-\mu_m(\lambda)]^2\int_{\Theta} p(x|\theta)\pi(\theta|\lambda)d\theta dx \\
&= \int_{\Theta}\left\{\int_{\chi}[x-\mu_m(\lambda)]^2 p(x|\theta)dx\right\}\pi(\theta|\lambda)d\theta \\
&= \int_{\Theta} E^{x|\theta}[X-\mu_m(\lambda)]^2\pi(\theta|\lambda)d\theta \\
&= \int_{\Theta}\sigma^2(\theta)\pi(\theta|\lambda)d\theta + \int_{\Theta}[\mu(\theta)-\mu_m(\lambda)]^2\pi(\theta|\lambda)d\theta \\
&= E^{\theta|\lambda}[\sigma^2(\theta)] + E^{\theta|\lambda}[\mu(\theta)-\mu_m(\lambda)]^2
\end{aligned}$$

这样，我们就把边际分布 $m(x|\lambda)$ 的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m{}^2(\lambda)$ 表示成了超参数 λ 的函数。

（3）当先验分布只含有二个超参数，即 $\lambda=(\lambda_1,\lambda_2)$ 时，可用混合样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 得样本均值和样本方差分别为

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

再用样本矩代替边际分布矩，得到如下方程组

$$\begin{cases} \hat{\mu}_m = E^{\theta|\lambda}[\mu(\theta)] \\ \hat{\sigma}_m = E^{\theta|\lambda}[\sigma^2(\theta)] + E^{\theta|\lambda}[\mu(\theta)-\mu_m(\lambda)]^2 \end{cases}$$

解这个方程组，就可得超参数 $\lambda=(\lambda_1,\lambda_2)$ 的估计 $\hat{\lambda}=(\hat{\lambda}_1,\hat{\lambda}_2)$，从而获得先验分布 $\pi(\theta|\hat{\lambda})$ 。

例 3.5 设总体 X 服从正态分布 $N(\theta,1)$，参数 θ 的先验分布取为共轭先验 $N(\mu_\pi,\sigma_\pi^2)$，其中超参数 $\lambda=(\mu_\pi,\sigma_\pi^2)$ 未知。现在设由边际分布 $m(x|\lambda)$ 的样本算得其样本均值和方差分别为 $\hat{\mu}_m=10$， $\hat{\sigma}_m^2=3$。试确定参数 θ 的先验分布。

解：由题目所给条件，总体均值 $\mu(\theta)=\theta$，总体方差 $\sigma^2(\theta)=1$ 为常数。由步骤（2）算出边际分布 $m(x|\lambda)$ 的均值与方差

$$\mu_m(\lambda)=E^{\theta|\lambda}[\mu(\theta)]=\mu_\pi$$
$$\sigma_m^2=E^{\theta|\lambda}[\sigma^2(\theta)]+E^{-\theta|\lambda}[\theta-\mu_m(\lambda)]^2=1+E^{\theta|\lambda}[\theta-\mu_\pi]^2=1+\sigma_\pi^2$$

将样本均值 $\hat{\mu}_m=10$ 和方差 $\hat{\sigma}_m^2=3$ 代入上两式的左边，我们得方程组

$$\mu_\pi=10,\ 1+\sigma_\pi^2=3$$

解之得 $\hat{\mu}_\pi=10$， $\hat{\sigma}_\pi^2=2$，从而参数 θ 的先验分布为 $N(10,2)$。

§ 3.3 用主观概率作为先验概率

在第 1 章中我们提到贝叶斯统计不反对经典统计中概率概念的频率定义（被称为客观概率），但同时认为概率也是个体对事件发生可能性的一种信念，这种信念越强则事件发生的概率越大，反之亦然，这种概率我们称之为主观概率（有时也称为可信度）。历史上第一次正式提出主观概率概念的是英国天才哲学家（同时是数学家和经济学家）拉姆齐 Ramsey (1926)，他在《真理与概率(Truth and Probability)》等文章中详细讨论了主观概率这个概念。贝叶斯统计不同于经典统计的一个重要方面就是它认可主观概率并把主观概率作为先验概率来使用。本节就是讨论为什么需要主观概率以及如何确定主观概率。一旦主观概率确定下来了，则先验概率也就随之确定了。

3.3.1 为什么需要主观概率

众所周知，在经典统计中主要利用频率的稳定性来定义概率，这就有一个非常重要的前提，那就是研究对象必须能在相同的环境下大量重复。由于不能重复的研究对象无法用频率的方法去确定有关事件的概率，而且由于时间的单向性或人力物力的限制，大量自然和社会现象都是无法重复的，这就大大地限制了经典统计学的应用和研究范围。另一方面，虽然大量自然和社会现象无法重复，我们却也经常听到电视台上气象专家的这样的说法：“未来某日（比如 2014 年 12 月 25 日）下雨的概率是 0.78。”这里的概率不能做出频率解释，因为 2014 年 12 月 25 日只有一次，无法重复。但是，气象专家根据专业知识和经验给出的“未来某日下雨的概率是 0.78”这一说法反映了气象专家对未来某日下雨可能性的一种信念。由于气象专家具有专业知识和经验，人们也会认可这一说法并理解其意义，那就是未来某日下雨的可能性比较大。这种基于专业知识和经验总结出的事件发生可能性的个人信念就是主观概率，而且它的数量测度与客观概率一样，都是用区间[0, 1]上的实数来表示主观概率的大小。

贝叶斯统计学既认同客观概率也认同主观概率。这样一来，大量无法重复的随机现象也就可以谈及概率了，于是贝叶斯统计学的应用和研究范围相较于经典统计学大大地得到了扩充。

例 3.6 主观概率的例子

（1）某企业家认为一项新产品在未来市场上畅销的概率是 0.87；

（2）某投资者认为购买某种股票能获得高收益的概率是 0.76；

（3）某脑外科医生认为某病人手术成功的概率是 0.89；

（4）某大学生认为教贝叶斯的老师年龄在 40 到 50 岁之间的概率是 0.9。

（5）某研究员相信明年的失业率小于 8%的概率是 0.88

这些概率都是主观概率，是相关个体利用专业知识和（工作或生活）经验总结出的对事件发生可能性的信念，而个体的专业知识和（工作或生活）经验其实就是一种先验信息，从而主观概率也是先验信息加工提炼得来的。

需要注意的是虽然主观概率是个体确定下来的，但它不是随心所欲的信口开河，它要求当事人对所研究的随机事件有较透彻的了解并根据过去的经验来确定，特别是它必须与所采用的概率公理体系相一致，不产生矛盾。就本书而言，主观概率就是要满足第 1 章第 2 节中的三条关于概率的公理（称为柯尔莫哥洛夫概率公理体系，是前苏联著名数学家柯尔莫哥洛夫于 1933 年建立的）。

3.3.2 确定主观概率的方法

有各种确定主观概率的方法。在实践中也可以根据实际情况和掌握的先验信息创造出新方法。以下是一些常用的方法。

1. 利用对立事件的概率比来确定主观概率。

例 3.7 现在我国的出版社全部都改制为企业了，因此利润是出版商必须考虑的大事。这样，要出一本新书，出版商首先就想知道这本新书畅销的概率是多少，以决定是否与作者签订出版合同。在了解了这本新书的内容后，出版商根据自己多年出书的经验认为该书畅销的可能性较大，并且畅销（记为 A）比不畅销（$\overline{A}$）的可能性要高出一倍，即有 $P(A)=2P(\overline{A})$，再根据概率的性质 $P(A)+P(\overline{A})=1$ 就可以推得 $P(A)=2/3$，即该书畅销的主观概率是 2/3。

2. 咨询专业人员和专家。

当决策者对某事件了解甚少或拿不定主意时，就可去咨询专业人员和专家的意见。这里有两个关键点，一是向专家提出的问题要设计科学，既要使专家容易明白，又要使专家的回答不会模棱两可，这样才能够从专家那里得到真实的信息。二是对专家本人较为了解，以便作出修正，形成决策者自己的主观概率。如果可以咨询多位专业人员和专家，再将他们的意见综合与平衡起来，最后得到自己的主观概率，当然是很好的做法。

例 3.8 某公司在决定是否生产某种新产品时，要进行预测分析，估计该产品在未来市场上畅销的概率。为此公司经理召集设计、财会、销售和质量管理等方面人员开座谈会，仔细分析影响新产品销路的各种因素，大家认为此新产品设计新颖而且质量好，只要定价合理，畅销可能性很大，而影响销路的主要因素是市场竞争。据了解，外面还有一家工厂（简称外厂）也有生产此种新产品的想法，而且该厂技术和设备都比本厂好。经理在听取大家的分析后，向与会人员提出两个问题：

（1）如果外厂不生产此新产品，本公司的新产品畅销的概率有多大？

（2）如果外厂要生产此新产品，本公司的新产品畅销的概率又有多大？

与会人员根据自己的经验和了解的信息对两个问题各写了一个概率，经理在计算了每个问题的概率的平均值后，略加修改，提出自己的观点："对于上述两个问题，本公司新产品畅销的概率各为 0.9 和 0.4。"就在此时，公司情报部门报告，外厂正忙于另一项新产品的开发，很可能无暇顾及此新产品的生产和销售。经理据此认为，外厂将生产此新产品的概率为 0.3，不生产此新产品的概率为 0.7。但是，如何得到最后的本公司此新产品畅销的概率呢？经理百思不得

其解，于是命令统计专业毕业的秘书利用上面 4 个主观概率把本公司此新产品畅销的概率算出来。秘书暗暗叫苦，悔不该当年没有好好学习，但静下心来一想，这不就是要用全概率公式吗？于是，立马计算出了本公司此新产品畅销的主观概率为

$$0.9\times0.7+0.4\times0.3=0.75$$

当秘书把答案告诉经理并向经理解释了如何得到这个答案后，经理很开心并当众表扬了秘书。

为了得到合理的主观概率，本题既综合了专业人员和专家的意见，也包含了决策者的看法，还利用了概率论知识，是一种值得推荐的好做法。

还有两种较细致的专家咨询法，分别叫做定分度法和变分度法。它们能够得到各种主观概率，还可以整理出累积概率分布等等，感兴趣的读者可参见有关著作（茆诗松和汤银才，2012）

3. 借鉴可类比的历史资料。

有的历史资料虽然不是直接关于所研究对象的，但相关的对象与研究对象可以类比，那么我们可以借鉴那些历史资料。

例 3.9 某公司经营儿童玩具多年，产销了几十种玩具，留下了不少有用的资料。现今该公司又设计了一款新式玩具，在投入生产之前要进行市场研究以决定是否生产。但是，公司并没有关于该玩具的直接资料，然而，经理查阅了公司过去 37 种类似新式玩具的销售记录，得知销售状态为畅销（A）、一般（B）、滞销（C）分别有 29，6，2 种，于是估算得类似新式玩具的三种销售状态的概率分别为

$$P(A)=29/37=0.784\text{，}\quad P(B)=6/37=0.162\text{，}\quad P(C)=2/37=0.054$$

考虑到这次设计的玩具在开发儿童智力上有显著效果而且定价适中，经理认为此种新玩具会更畅销一些，故对上述概率作了修改，提出了自己的主观概率

$$P(A)=0.85\text{，}\quad P(B)=0.14\text{，}\quad P(C)=0.01$$

注：这里必须有 $P(A)+P(B)+P(C)=1$。不然，就要修改上述主观概率以满足要求。

§3.4 无信息先验分布

在本章前三节中，用来确定先验分布的先验信息要么信息量越来越少要么信息本身越来越模糊，从而先验分布的确定越来越困难。但是，在实际应用的研究中，往往还会出现无先验信息可用的情形，例如，所研究的是一个全新的问题或者获得先验信息的代价昂贵，此时如何去寻求先验分布呢？其实，从贝叶斯统计诞生之日起就伴随着“无先验信息可用，如何确定先验分布？”的问题。两百多年来，经过一代又一代统计学家的努力，已经提出了多种无信息先验分布的确定方法。本节就是要对有关的内容做一个简要的介绍，因专业知识的限制，我们略去了许多证明和细节。

3.4.1 **非正常先验与贝叶斯假设**

在贝叶斯统计中，为了获得后验分布，常常要借助于非正常先验分布，那么，什么是非正常先验分布呢？我们先来看一个例子。

例 3.10 设 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 是来自正态总体 $N(\theta,1)$ 的样本，如果参数 θ 的先验取为 $\pi(\theta)=C$（大于 0 的常数），$-\infty<\theta<\infty$，试求后验分布。

解： 样本密度为

$$p(\mathbf{x}\mid\theta)=\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2\}$$

根据贝叶斯公式，后验密度

$$\pi(\theta|\mathbf{x})=\frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} p(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

$$=\frac{C\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2\}}{\int_{-\infty}^{\infty} C\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2\}d\theta}$$

$$=\frac{\exp\{-\frac{1}{2}\sum_{i=1}^{n}[(x_i-\overline{x})+(\overline{x}-\theta)]^2\}}{\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}\sum_{i=1}^{n}[(x_i-\overline{x})+(\overline{x}-\theta)]^2\}d\theta}$$

$$=\frac{\exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(\overline{x}-\theta)^2\}}{\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(\overline{x}-\theta)^2\}d\theta}$$

$$=\frac{\exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\}\exp\{-\frac{n}{2}(\overline{x}-\theta)^2\}}{\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\}\exp\{-\frac{n}{2}(\overline{x}-\theta)^2\}d\theta}$$

$$=\frac{\exp\{-\frac{n}{2}(\theta-\overline{x})^2\}}{\int_{-\infty}^{\infty}\exp\{-\frac{n}{2}(\theta-\overline{x})^2\}d\theta}=\sqrt{\frac{n}{2\pi}}\exp\{-\frac{n}{2}(\theta-\overline{x})^2\}$$

其中 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 是样本均值。这就是说后验分布是正态分布 $N(\overline{x},1/n)$ 。

注：先验 $\pi(\theta)=C$ 中的常数 C 显然可以取任意正值而且都对后验的确定没有影响，所以今后我们常取 $\pi(\theta)=1$ 。另外，从下小节知道，这个先验其实就是参数 θ 的无信息先验。

在例 3.10 中，θ 是正态分布的均值参数，因此可取任何一个实数，即参数空间 $\Theta=(-\infty,+\infty)$。而给定的先验 $\pi(\theta)=C$，$-\infty<\theta<\infty$ 显然不是一个正常分布密度，不过，由它和样本密度按贝叶斯公式决定的后验却是正常的分布即正态分布 $N(\bar{x},1/n)$。这样的先验在贝叶斯统计中经常用到，被称为非正常先验。不仅如此，我们还看到后验均值就等于样本均值，所以，如果我们用后验均值作为参数 θ 的估计，那么它和经典统计中的估计是一模一样的，这是很奇妙的事！但它却不是孤立的事情，后面我们会一再看到这种现象。这说明经典统计学中好些估计量可以看作使用贝叶斯统计学中适当的无信息先验的结果。下面我们给出非正常先验的正式定义。

定义 3.1 设样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 来自总体 $p(x|\theta)$，$\theta\in\Theta$。若参数 θ 的函数 $\pi(\theta)$ 满足下列条件：

（1）$\pi(\theta)\geq 0$ 且 $\int_\Theta \pi(\theta)d\theta=+\infty$；（2）由贝叶斯公式确定的后验密度 $\pi(\theta|\mathbf{x})$ 是正常的密度函数，则称 $\pi(\theta)$ 为 θ 的非正常（或广义）先验（分布）密度。

从这个定义可知，非正常（或广义）先验却起正常的作用，因为它与样本密度一起通过贝叶斯公式确定了一个正常的后验密度。例 3.10 中给出的 $\pi(\theta)=1$ 就是一个非正常先验密度，并常称它为直线上的均匀密度。另外，非正常先验 $\pi(\theta)$ 乘以任意一个给定的正数仍然是非正常先验。有时在参数空间上可积的非负函数，只要它的积分值不等于 1，我们也称之为非正常（或广义）先验密度。

定义 3.2 设总体分布为 $p(x\mid\theta)$，则参数 θ 的无信息先验分布是指在仅知 θ 的变动范围（即定义域）Θ 和 θ 在总体分布中的地位这两条信息下寻求到的先验分布。

注：这里不是有意不用其它先验信息而是无法获得它们。

贝叶斯假设：无关于 θ 的任何其它信息就意味着对 θ 的任何可能取值都是同样无知的，因此很自然地把 θ 的定义域上的每个点同等看待，或者说每个点有一样的机会被抽取到，而要做到这点就得取 Θ 上的均匀分布作为 θ 的无信息先验分布，这就是贝叶斯假设。

按照定义域Θ的具体情形分类，我们有

1. 定义域Θ只含有有限个可能取值，即$\Theta=\{\theta_i, i=1,\cdots,n\}$。这时无信息先验就是分布列

$\pi(\theta_i)=1/n, i=1,\cdots,n$

2. 定义域Θ是有界的，这时无信息均匀先验是

$$\pi(\theta)=\begin{cases} C, \theta\in\Theta \\ 0, \theta\notin\Theta \end{cases}$$

其中C是一个常数。例如，$\Theta=(a,b)$，则$\pi(\theta)=1/(b-a), \theta\in\Theta$。（“$\theta\notin\Theta$时$\pi(\theta)=0$”没有写出来是习惯约定，以后都按这个约定做。）

3. 定义域Θ是无界的，这时无信息均匀先验是

$$\pi(\theta)=1, \theta\in\Theta$$

这是一个非正常先验密度，法国数学家拉普拉斯（1812）也研究和应用过这个无信息均匀先验，所以它也叫做拉普拉斯（无信息）先验。

当把非正常先验密度也作为无信息先验后，似乎在任何情况下寻找先验分布的问题都解决了，因为上述三种参数定义域把各种可能的情形都考虑进去了。但是，进一步的研究发现事情没有这么简单。下面我们来看一个例子。

例 3.11 考虑正态分布 $N(0,\sigma^2)$ 的未知标准差参数 σ，它的参数空间显然是 $\Theta=(0,\infty)$。现在没有 σ 的任何信息，那么按照贝叶斯假设，σ 的无信息先验应为 $\pi(\sigma)=1,\sigma\in\Theta$。现在定义 $\eta=\sigma^2$，则 η 是正态方差。在区间 $(0,\infty)$ 上，η 与 σ 是一一对应的，不会损失什么信息。因此，如果 σ 是无信息参数，那么 η 也是无信息参数，且它们的参数空间都是区间 $(0,\infty)$。按照贝叶斯假设，它们的无信息先验分布应都为常数，也就是说 η 的无信息先验应为 $\pi_\eta(\eta)=1$， $\eta\in\Theta$。另一方面，由于 $\eta=\sigma^2$，即 $\sigma=f(\eta)=\sqrt{\eta}$，则按概率论中分布密度的运算法则，$\eta$ 的密度函数

$$\pi_\eta(\eta)=\pi(\sqrt{\eta})\left|\frac{df}{d\eta}\right|=\frac{1}{2\sqrt{\eta}}\pi(\sqrt{\eta})$$

但已知 σ 的无信息先验为 $\pi(\sigma)=1,\sigma\in\Theta$，这样，$\pi(\sqrt{\eta})=1$，从而 η 的无信息先验 $\pi_\eta(\eta)=(1/2)\eta^{-1/2}$，这与 $\pi_\eta(\eta)=1$， $\eta\in\Theta$ 产生矛盾。

3.4.2 **位置参数的无信息先验**

设总体分布为 $p(x|\theta)$，则在考虑其参数 θ 的无信息先验分布时，我们首先要知道该参数 θ 的定义域和在总体分布的地位（作用），也就是说用到了总体分布的信息。本小节讨论参数 θ 为位置参数时其无信息先验分布的确定。

定义 3.3 设总体 X 的密度函数 $p_X(x|\theta)$ 具有形式 $p(x-\theta)$，参数空间 Θ 为实数集 R。那么，密度函数的集合 $\{p(x-\theta),\theta\in\Theta\}$ 称为位置参数族，参数 θ 称为位置参数，而 $p(x)$ 称为这个位置参数族的标准分布密度函数。

例 3.12 位置参数族的两个例子。

（1）正态分布族 $\{N(\theta,\sigma^2),-\infty<\theta<\infty\}$ 当方差 σ^2 已知时是一个位置参数族，均值 θ 为位置参数，分布 $N(0,\sigma^2)$ 的密度为这个位置参数族的标准（分布）密度函数。

（2）分布族 $\{p(x-\theta)=e^{-(x-\theta)},\theta\in\Theta,x\geq\theta\}$ 是另一个位置参数族的例子，称为位置指数分布族，其中 $p(x)=e^{-x},x\geq 0$ 是它的标准密度函数。

对于同一个位置参数族中的密度函数来说，它们的形状一模一样，只是位置不同而已，图 3.4 就是位置参数族 $N(\theta,1)$ 中的三个密度函数图，（从左到右）位置参数 θ 分别为-1，0，1。从该图容易看出，位置参数决定了密度曲线图的位置，而图形本身只是平移，没有任何变化。

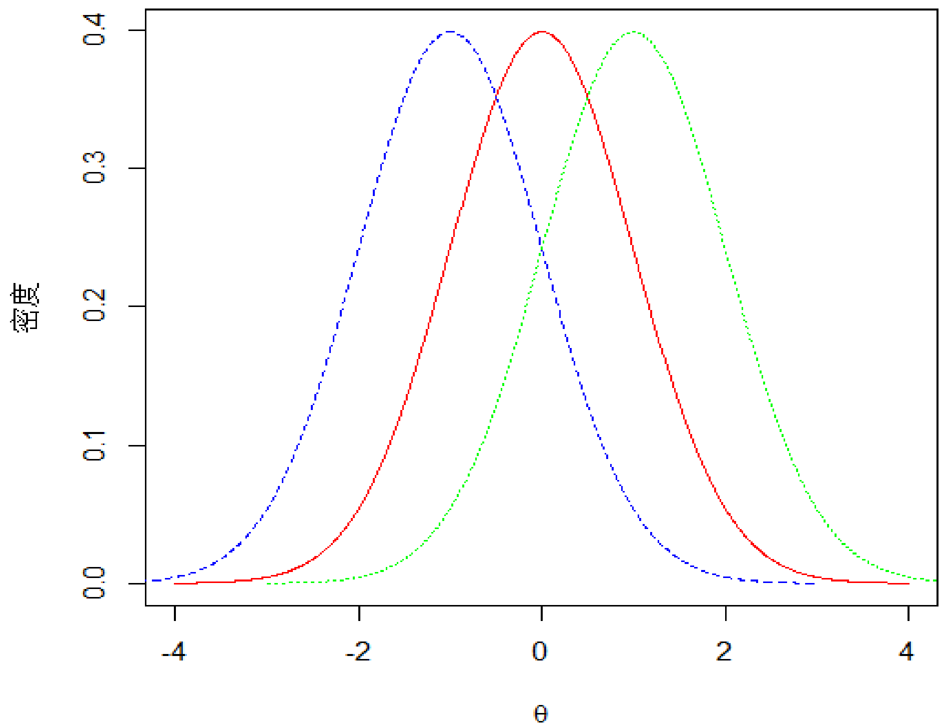

图 3.4　同一个位置参数族中的三条密度曲线

现在讨论位置参数θ的无信息先验分布问题。对X做平移变换，得到$Y=X+c$，让参数θ也做同样的平移变换得到$\eta=\theta+c$，那么Y有密度$p(y-\eta)$。事实上，设Y的密度为$p_Y(y|\eta)$，则

$$p_Y(y|\eta)=\frac{dP(Y<y)}{dy}=\frac{dP(X<y-c)}{dy}=\frac{dF_X(y-c)}{dy}$$
$$=p_X(y-c|\theta)=p(y-c-\theta)=p(y-\eta)$$

其中$F_X(x)$是X的分布函数。显然，Y的密度$p(y-\eta)$仍是该位置参数族的成员，η也是一个位置参数，所以研究对象(X,θ)与(Y,η)的统计结构（或者说，统计规律）完全相同，因此θ与η应有相同的无信息先验分布，即

$$\pi_\theta(t)=\pi_\eta(t)$$

另一方面，由变换$\eta=\theta+c$（即$\theta=\eta-c$）及概率论中分布密度的运算法则，可以算得η的先验分布密度又等于

$$\pi_\eta(\eta)=\pi_\theta(\eta-c)\left|\frac{d\theta}{d\eta}\right|=\pi_\theta(\eta-c)$$

于是

$$\pi_\theta(\eta)=\pi_\eta(\eta)=\pi_\theta(\eta-c)$$

特别地，取$\eta=c$，则有$\pi_\theta(c)=\pi_\theta(0)$。由 c 的任意性，得知先验$\pi(\theta)$为常数，故得$\theta$的无信息先验分布密度为$\pi(\theta)=1$（你可用任意常数）。这就说明当$\theta$是位置参数时，可用贝叶斯假设确定它的无信息先验分布。

例 3.13 样本$\mathbf{x}=(x_1,x_2,\cdots,x_n)$来自指数分布$p(x\,|\,\theta)=e^{-(x-\theta)}$，$x\ge\theta$，参数$\theta\in(-\infty,\infty)$但无先验信息。试求后验分布并判断它是否为正常分布。

解：我们已知指数分布$p(x\,|\,\theta)=e^{-(x-\theta)}$，$\theta\in(-\infty,\infty)$全体构成一个位置参数族（位置指数分布族），所以，可取$\pi(\theta)=1$为无信息先验密度。此外，样本密度

$$p(\mathbf{x}\,|\,\theta)=\exp\{-(\sum_{i=1}^n x_i-n\theta)\}=\exp\{-n(\bar{x}-\theta)\}$$

注意每个样本必须满足$x_i\ge\theta$，令$x_0=\min(x_i,1\le i\le n)$，由贝叶斯公式，后验密度

$$\pi(\theta|\mathbf{x})=\frac{p(\mathbf{x}\,|\,\theta)\pi(\theta)}{\int_R p(\mathbf{x}\,|\,\theta)\pi(\theta)d\theta}=\frac{\exp\{-n(\bar{x}-\theta)\}}{\int_{-\infty}^{x_0}\exp\{-n(\bar{x}-\theta)\}d\theta}=\frac{\exp(n\theta)}{\int_{-\infty}^{x_0}\exp(n\theta)d\theta}=ne^{n(\theta-x_0)}$$

注意到$\theta\le x_0$，容易看出后验密度是正常的指数分布。

3.4.3 尺度参数的无信息先验

定义 3.4 设总体 X 的密度函数 $p_X(x|\sigma)$ 具有形式 $\sigma^{-1}p(x/\sigma)$，参数空间为 $(0,\infty)$。那么，密度函数的集合 $\left\{\sigma^{-1}p(x/\sigma),\sigma\in(0,\infty)\right\}$ 称为一个尺度参数族，参数 σ 称为尺度参数，而 $p(x)$ 称为这个尺度参数族的标准分布密度函数。

例 3.14 尺度参数族的两个例子。

（1）正态分布族 $\{N(\theta,\sigma^2),-\infty<\theta<\infty\}$ 当均值 θ 已知时是一个尺度参数族，标准差 σ 为尺度参数，正态分布 $N(\theta,1)$ 的密度为这个位置参数族的标准（分布）密度函数。

（2）伽玛分布族 $Gamma(\alpha,\lambda)$，$\lambda>0$ 当形状参数 α 已知时是一个尺度参数族，参数 $\lambda>0$ 为尺度参数。事实上，伽玛分布族的密度函数

$$p(x|\alpha,\lambda)=\frac{\lambda^{-\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-x/\lambda}=\lambda^{-1}\left[\frac{1}{\Gamma(\alpha)}(\frac{x}{\lambda})^{\alpha-1}e^{-x/\lambda}\right],x>0$$

具有形式 $\lambda^{-1}p(x/\lambda)$。这个尺度参数族的标准（分布）密度函数则为

$$p(x|\alpha,\lambda=1)=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x},x>0$$

对于同一个尺度参数族中的密度函数来说，它们的图形的形状与这个尺度参数族的标准（分布）密度函数的图形基本相同，只不过对它进行了伸张或收缩。图 3.5 就是尺度参数族 $N(0,\sigma^2)$ 中的三个密度函数图，（从上到下）尺度参数 σ 分别为 1，1.2，2，其中 $\sigma=1$ 时对应的密度函数图就是标准密度函数的图形（图 3.5 中实线者）。

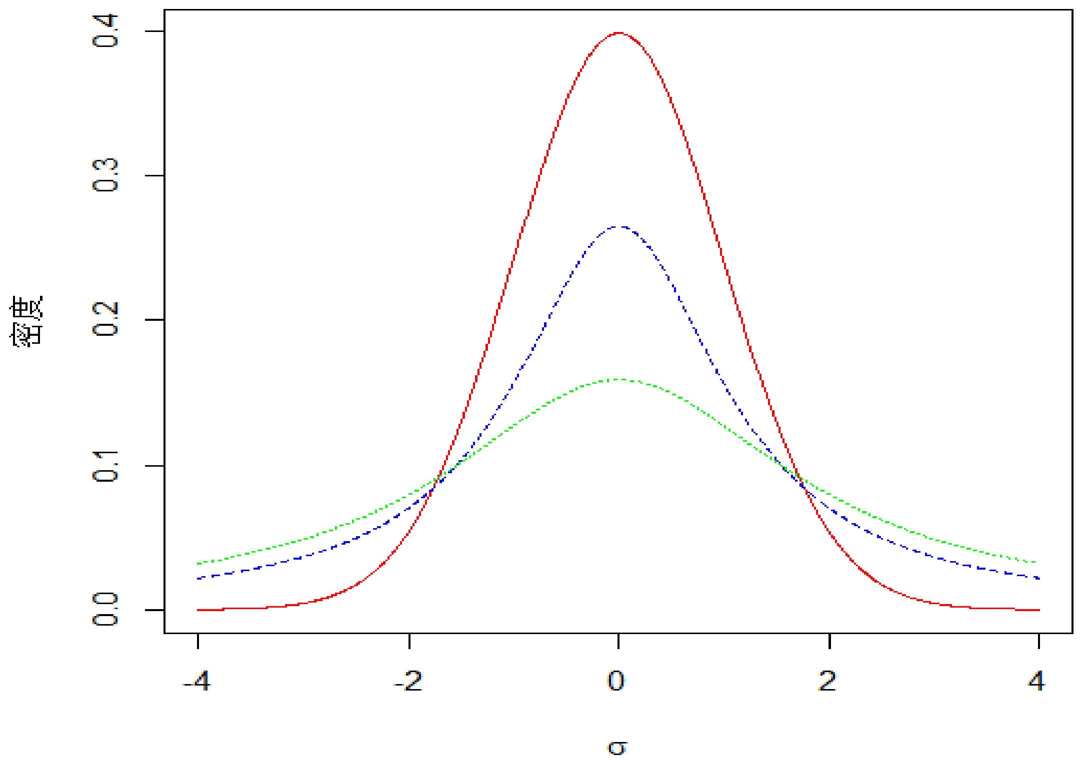

图 3. 5　同一个尺度参数族中的三条密度曲线

现在讨论尺度参数σ的无信息先验问题。对 X 做伸缩变换，得到$Y=cX$，让参数σ也做同样的变换而得到$\eta=c\sigma$，其中变换系数$c>0$。可以证明Y有密度$\frac{1}{\eta}p(\frac{y}{\eta})$（作为练习），显然这个密度函数仍属于给定的尺度参数族，η也是一个尺度参数，σ的参数空间与η的参数空间都为$(0,\infty)$，可见研究对象(X,σ)与(Y,η)的统计结构完全相同，所以σ的无信息先验$\pi_\sigma(\sigma)$与η的无信息先验$\pi_\eta(\eta)$应该相同$\pi_\sigma(t)=\pi_\eta(t)$。另一方面，由变换$\eta=c\sigma$（即$\sigma=\eta/c$）及概率论中分布密度的运算法则，可以得到$\eta$的无信息先验

$$\pi_\eta(\eta)=\pi_\sigma(\eta/c)\left|\frac{d\sigma}{d\eta}\right|=\frac{1}{c}\pi_\sigma(\frac{\eta}{c})$$

从而可得

$$\pi_\sigma(\eta)=\pi_\eta(\eta)=\frac{1}{c}\pi_\sigma(\frac{\eta}{c})$$

取$\eta=c$，则有

$$\pi_\sigma(c)=\frac{1}{c}\pi_\sigma(1)$$

由 $c>0$ 的任意性，得 σ 的无信息先验 $\pi_\sigma(\sigma)=\frac{1}{\sigma}\pi_\sigma(1)$ 。再由贝叶斯公式知可令常数 $\pi_\sigma(1)=1$（任何非零常数乘以先验都不影响后验分布的确定），最后得尺度参数 σ 的无信息先验密度

$$\pi(\sigma)=\frac{1}{\sigma},\sigma>0$$

例 3.15　样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 来自指数分布

$$p(x|\sigma)=\sigma^{-1}\exp\{-x/\sigma\},\sigma>0,x>0$$

但无参数 σ 的先验信息。试求后验分布及后验均值。

解：所给指数分布全体显然构成了一个尺度参数族，而且 σ 是尺度参数，又无参数 σ 的先验信息，因此 σ 的先验取无信息先验 $\pi(\sigma)=\sigma^{-1},\sigma>0$ 。在样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 给定下， σ 的后验密度函数

$$\begin{aligned}\pi(\sigma|\mathbf{x})&\propto p(\mathbf{x}|\sigma)\pi(\sigma)\\&=\sigma^{-(n+1)}\prod_{i=1}^{n}\exp\{-\frac{x_i}{\sigma}\}=\sigma^{-(n+1)}\exp\{-\frac{1}{\sigma}\sum_{i=1}^{n}x_i\},\sigma>0\end{aligned}$$

这是逆伽玛分布 $IGamma(n,\sum_{i=1}^{n}x_i)$，它的后验均值 $E(\sigma|\mathbf{x})=\frac{1}{n-1}\sum_{i=1}^{n}x_i$ 。

3.4.4 杰弗里斯先验

在前两小节中，我们讨论了位置参数族和尺度参数族的无信息先验的确定问题，并且获得了相应的无信息先验。但是，有许多分布族既不是位置参数族也不是尺度参数族，例如，最常见的正态分布族 $\{N(\theta,\sigma^2)\}$ 当两个参数都未知时就是如此。 所以，对它们的无信息先验分布的确定仍然是一个问题。统计学家杰弗里斯（Jeffreys，1961）对此问题做了深入研究，提出了一般情形下确定无信息先验的方法。

设分布族 $\{f(x|\boldsymbol{\theta}),\boldsymbol{\theta}\in\Theta\}$ 满足 Cramer-Rao 正则条件，其中 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_p)$ 是 p 维参数向量。该条件共五条，感兴趣的读者可参考相关高等数理统计的著作，大部分常见的分布族满足该条件。在参数 $\boldsymbol{\theta}$ 无先验信息可利用时，杰弗里斯证明了可用以下步骤来确定 $\boldsymbol{\theta}$ 的无信息先验，这样的无信息先验被后人称为杰弗里斯（Jeffreys）先验。

（1）写出总体密度（概率函数）$f(x|\boldsymbol{\theta})$ 的自然对数并记为

$$l(\boldsymbol{\theta}, x) = l(\boldsymbol{\theta}) = \ln[f(x|\boldsymbol{\theta})]$$

（2）求总体的费雪（Fisher）信息阵（量）

$$I(\boldsymbol{\theta}) = \left[I_{ij}(\boldsymbol{\theta})\right]_{p\times p},\ I_{ij}(\boldsymbol{\theta}) = E^{x|\boldsymbol{\theta}}\left(\frac{\partial l}{\partial \theta_i} \times \frac{\partial l}{\partial \theta_j}\right)\ (i, j = 1, \cdots, p)$$

这里 $E^{x|\boldsymbol{\theta}}$ 表示对总体密度 $f(x|\boldsymbol{\theta})$ 求期望，例如，在单参数（$p=1$）情形下

$$I(\theta) = E^{x|\theta}\left(\frac{dl}{d\theta}\right)^2 = \int \left(dl/d\theta\right)^2 f(x|\boldsymbol{\theta})dx$$

（3）参数向量 $\boldsymbol{\theta}$ 的无信息先验密度为

$$\pi(\boldsymbol{\theta}) = \sqrt{\det[I(\boldsymbol{\theta})]}$$

其中 $\det[I(\boldsymbol{\theta})]$ 表示 $p\times p$ 阶矩阵 $I(\boldsymbol{\theta})$ 的行列式。特别地，在单参数情形下

$$\pi(\theta) = [I(\theta)]^{1/2}$$

注:

1. 如果总体 $f(x|\boldsymbol{\theta})$ 关于参数向量 $\boldsymbol{\theta}$ 的各个二阶导数存在，则有简化公式

$$I_{ij}(\boldsymbol{\theta}) = E^{x|\boldsymbol{\theta}}\left(\frac{\partial l}{\partial \theta_i}\times\frac{\partial l}{\partial \theta_j}\right) = E^{x|\boldsymbol{\theta}}\left(-\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}\right)\ (i, j = 1, \cdots, p)$$

特别地，在单参数情形下

$$I(\theta) = E^{x|\theta}\left(\frac{dl}{d\theta}\right)^2 = E^{x|\theta}\left(-\frac{d^2 l}{d\theta^2}\right)$$

2. 特别要注意这里寻求的是先验密度，不牵涉到任何来自总体的样本，所以，有的著作把 $l(\boldsymbol{\theta}, x) = \ln[f(x|\boldsymbol{\theta})]$ 看成对数似然函数是不妥的。

3. 虽然杰弗里斯先验（以及位置参数和尺度参数先验）是无信息先验，但从某个角度讲，却是客观而不是主观先验分布，因为它们主要是利用概率统计的内在逻辑和运算规则确定下来的。

例 3.16 总体 X 服从泊松分布 $Poisson(\lambda)$，其分布列是

$$p(x|\lambda)=\frac{\lambda^x}{x!}e^{-\lambda},\ x=0,1,2,\cdots$$

但参数 λ 无先验信息。试求总体的 Fisher 信息量及参数 λ 的无信息先验。

解： 总体概率函数的对数为

$$l(\lambda,x)=\ln[p(x|\lambda)]=x\ln\lambda-\lambda-\ln(x!)$$

从而

$$\frac{dl}{d\lambda}=\frac{x}{\lambda}-1,\quad \frac{d^2l}{d\lambda^2}=-\frac{x}{\lambda^2}$$

再由 $E(X)=\lambda$，所以参数 λ 的 Fisher 信息量

$$I(\lambda)=-E\left[\frac{d^2l}{d\lambda^2}\right]=E(\frac{X}{\lambda^2})=\frac{1}{\lambda}$$

参数 λ 的无信息先验 $\pi(\lambda)=\lambda^{-1/2}$。

例 3.17 总体 X 服从正态分布 $N(\mu,\sigma^2)$，但无参数向量 (μ,σ) 的任何先验信息，试求参数向量（μ,σ）的杰弗里斯先验。

解： 容易写出总体密度的对数

$$l(\mu,\sigma)=-\frac{1}{2}\ln(2\pi)-\ln\sigma-\frac{1}{2\sigma^2}(x-\mu)^2$$

它的各个二阶偏导数

$$\frac{\partial^2 l}{\partial\mu^2}=-\sigma^{-2}，\quad \frac{\partial^2 l}{\partial\mu\partial\sigma}=\frac{\partial^2 l}{\partial\sigma\partial\mu}=-2(x-\mu)\sigma^{-3}，\quad \frac{\partial^2 l}{\partial\sigma^2}=\sigma^{-2}-3(x-\mu)^2\sigma^{-4}$$

由于 $E(X)=\mu$， $E(X-\mu)^2=\sigma^2$，故总体的 Fisher 信息阵

$$I(\mu,\sigma)=\begin{pmatrix} E(-\frac{\partial^2 l}{\partial\mu^2}) & E(-\frac{\partial^2 l}{\partial\mu\partial\sigma}) \\ E(-\frac{\partial^2 l}{\partial\mu\partial\sigma}) & E(-\frac{\partial^2 l}{\partial\sigma^2}) \end{pmatrix}=\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

从而 $\det[I(\mu,\sigma)]=2\sigma^{-4}$，所以 (μ,σ) 的杰弗里斯先验为 $\pi(\mu,\sigma)=\sigma^{-2}$。

注：

1. 对于多维参数向量，常用 Fisher 信息阵的行列式 $\det[I(\boldsymbol{\theta})]$ 来表示关于总体的信息量，在本例中 $\det[I(\mu,\sigma)] = 2\sigma^{-4}$，而 σ^2 是总体分布的方差，这就说明总体分布的方差越小（即分布越集中），关于总体的信息量就越大。

2. 当 σ 已知，$I(\mu) = E(-\frac{\partial^2 l}{\partial \mu^2}) = \frac{1}{\sigma^2}$ 为常数，故参数 μ 的先验 $\pi_1(\mu) = 1$。这与位置参数族 $\{N(\mu,\sigma^2), -\infty < \mu < \infty\}$ 下参数 μ 的无信息先验一致。

3. 当 μ 已知，$I(\sigma) = E(-\frac{\partial^2 l}{\partial \sigma^2}) = \frac{2}{\sigma^2}$，故参数 σ 的先验 $\pi_2(\sigma) = \sigma^{-1}$。这与尺度参数族 $\{N(\mu,\sigma^2), 0 < \sigma < \infty\}$ 下参数 σ 的无信息先验一致。另外，类似可求得此时方差 σ^2 的杰弗里斯先验为 $\pi(\sigma^2) = 1/\sigma^2 = (\sigma^2)^{-1}$

4. 当μ和σ独立时，$\pi(\mu,\sigma)=\pi_1(\mu)\pi_2(\sigma)=1/\sigma$。因此，$\mu$和$\sigma$不独立时参数向量$(\mu,\sigma)$的联合先验分布$\sigma^{-2}$与$\mu$和$\sigma$独立时（$\mu,\sigma$）的联合先验分布$\sigma^{-1}$不同。不过，回忆经典概率统计的经典定理：对于来自正态分布$N(\mu,\sigma^2)$的样本$\{X_i,1\le i\le n\}$，样本均值$\bar{X}=\sum X_i/n$和样本标准差(方差)$S=\left[\sum(X_i-\bar{X})^2/(n-1)\right]^{1/2}$相互独立，而且它们分别是$\mu$和$\sigma$的很好的估计量。因此，把$\mu$和$\sigma$看成独立也是有一定道理的，这样，就可取$\pi(\mu,\sigma)=\sigma^{-1}$为参数向量$(\mu,\sigma)$的杰弗里斯联合无信息先验。这样做还有一个原因，那就是当参数是多维时，参数间要不相关用杰弗里斯先验才能有较好的推断结果。正因为如此，杰弗里斯最终推荐的是先验$\pi(\mu,\sigma)=\sigma^{-1}$。同理，$(\mu,\sigma^2)$的杰弗里斯先验是$\pi(\mu,\sigma^2)=(\sigma^2)^{-1}=1/\sigma^2$。

例 3.18 设 X 服从二项分布

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

其中参数θ为成功概率。求θ的杰弗里斯先验。

解： 二项分布概率函数的对数为

$$l=x\ln\theta+(n-x)\ln(1-\theta)+\ln\binom{n}{x}$$

其对参数θ二阶导数

$$\frac{d^{2}l}{d\theta^{2}}=-\frac{x}{\theta^{2}}-\frac{n-x}{(1-\theta)^{2}}$$

因此，Fisher 信息量

$$I(\theta)=E^{x|\theta}(-\frac{d^{2}l}{d\theta^{2}})=\frac{n}{\theta}+\frac{n}{1-\theta}=n\theta^{-1/2}(1-\theta)^{-1/2}$$

所以，成功概率θ的杰弗里斯无信息先验为

$$\pi(\theta)\propto\theta^{-1/2}(1-\theta)^{-1/2}, 0<\theta<1$$

注：这个分布核正是贝塔分布 $Beta(0.5,0.5)$ 的核，所以这个无信息先验是 $Beta(0.5,0.5)$。我们以前还用过均匀分布 $U(0,1)$ 作为成功概率 θ 的无信息先验。那么，这两个无信息先验以及它们通过贝叶斯公式确定的后验分布到底有多大的区别呢？首先，注意均匀分布 $U(0,1)$ 就是贝塔分布 $Beta(1,1)$，不难用 R 命令画出这两个无信息先验的密度曲线（图 3.6），从该图可以看出两个无信息先验似乎差别较大。其次，注意贝塔分布是成功概率 θ 的共轭先验，取样本为 $n=10, x=2$，则不难算出上面两个先验对应的后验分别是 $Beta(2.5,8.5)$ 和 $Beta(3,9)$，画出这两个后验密度的曲线（图 3.7），我们看到它们却几乎是一模一样的（注意图中的尺度是很小的）。

一般而言，无信息先验不是唯一的，但在大多数情形下它们通过贝叶斯公式确定的后验分布的差异很小，从而它们的贝叶斯统计推断的结果的差异也很小，所以，任何合理的无信息先验分布都可以采用。

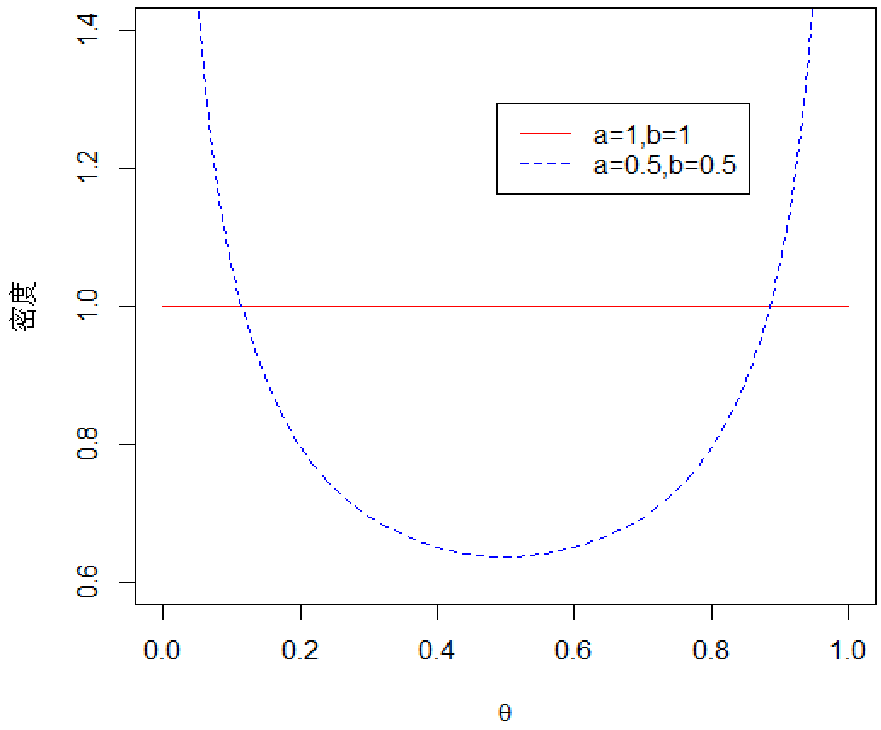

图 3.6　两个先验密度的比较

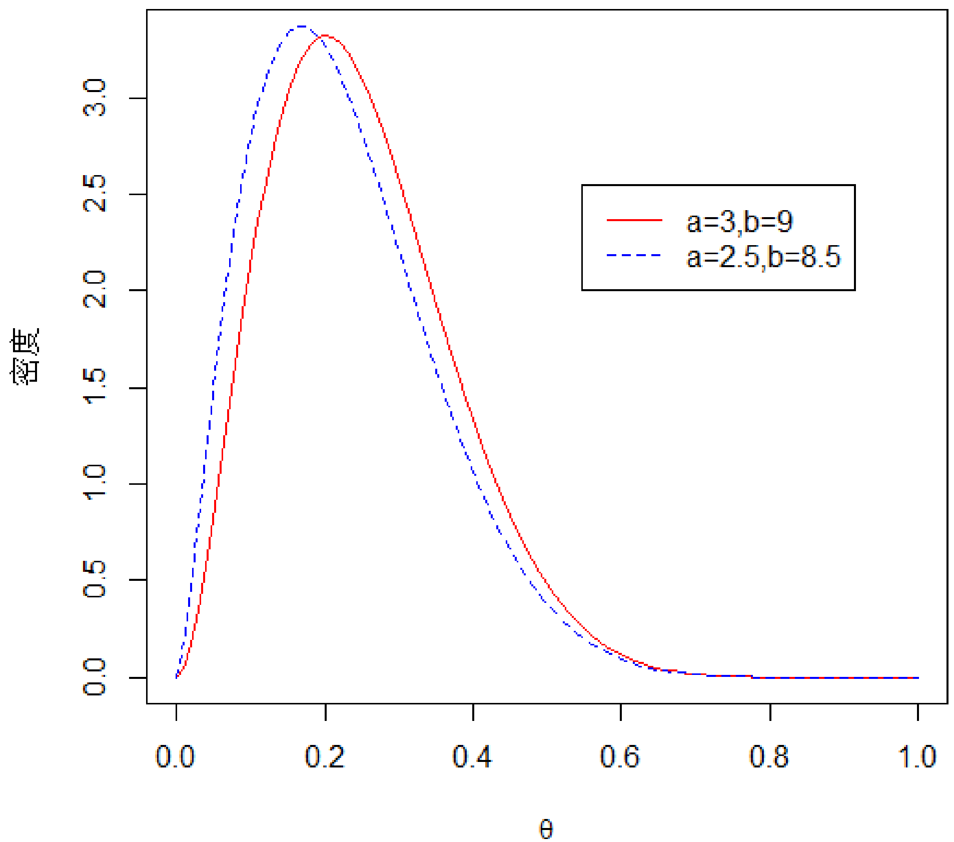

图 3.7　同一总体和样本但不同先验下两个后验密度比较

作业：

PP49-50，

1，2，3，4，5，8，9，13，14，15