第4章 贝叶斯统计推断

在第1章和第2章，我们引入了各种类型的贝叶斯公式并初步讨论了后验分布（密度）的计算问题，并且了解到**贝叶斯统计的所有统计推断都是基于后验分布(密度)来进行的**。那么，当后验分布(密度)计算出来后，如何进行贝叶斯统计推断呢？这就是本章要讨论的主要问题。我们将介绍如何利用后验分布（密度）进行参数的点估计、区间估计、假设检验、模型选择以及统计预测等问题。

§4.1 贝叶斯估计

4.1.1 点估计

设样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 有联合密度（概率函数） $p(\mathbf{x}|\theta)$，其中 θ 是未知的待估参数。为了估计该参数，贝叶斯统计的做法是，依据 θ 的先验信息选择一个适当的先验分布 $\pi(\theta)$，再经由贝叶斯公式算出后验分布 $\pi(\theta|\mathbf{x})$，最后，选择后验分布 $\pi(\theta|\mathbf{x})$ 的某个特征量作为参数 θ 的估计。下面给出正式定义。

定义 4.1 后验密度（概率函数）$\pi(\theta|\mathbf{x})$ 的众数 $\hat{\theta}_{MD}$ 称为参数 θ 的后验众数估计（也称为广义最大似然估计和最大后验估计），后验分布的中位数 $\hat{\theta}_{ME}$ 称为 θ 的后验中位数估计，后验分布的期望（均值）$\hat{\theta}_E$ 称为 θ 的后验期望估计。这三个估计也都可称为 θ 的贝叶斯（点）估计并记为 $\hat{\theta}_B$。

在一般情形下，这三种贝叶斯估计是不同的，但当后验密度函数关于均值左右对称时，这三种贝叶斯估计重合为一个数。另外，一般而言，当先验分布为共轭先验时，贝叶斯估计比较容易求得。

例 4.1 设样本 x（成功次数）来自二项分布

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

其中参数 θ 为成功概率。现取贝塔分布 $Beta(\alpha,\beta)$ 为 θ 的先验分布，试求参数 θ 的后验众数估计和后验期望估计。

解：我们已知贝塔分布 $Beta(\alpha,\beta)$ 是参数 θ 的共轭先验分布，所以，θ 的后验分布为贝塔分布 $Beta(\alpha+x,\beta+n-x)$。因此，θ 的后验众数估计和后验期望估计分别为

$$\hat{\theta}_{MD}=\frac{\alpha+x-1}{\alpha+\beta+n-2}, \quad \hat{\theta}_{E}=\frac{\alpha+x}{\alpha+\beta+n}$$

注：由第 3 章例 3.18 知 θ 的杰弗里斯先验为 $\pi(\theta)\propto\theta^{-1/2}(1-\theta)^{-1/2}$（即贝塔分布 $Beta(0.5,0.)$），而由贝叶斯假设得 θ 的先验分布为均匀分布 $U(0,1)$（即贝塔分布 $Beta(1,1)$），二者都是特殊的贝塔分布，因此，对应这两个无信息先验的贝叶斯估计也一并解决了。

现在我们特别对先验分布取为均匀分布 $Beta(1,1)$ 的情形做深入一点的讨论。显然，此时参数 θ 的两个贝叶斯估计分别为

$$\hat{\theta}_{MD}=\frac{x}{n},\ \hat{\theta}_{E}=\frac{x+1}{n+2}$$

这里令人感到惊奇的是参数 θ 的后验众数估计居然就是经典统计中 θ 的最大似然估计，也就是说，成功概率 θ 的最大似然估计就是取特定的先验分布 $Beta(1,1)$ 下的后验众数估计。这种现象不是孤立的，以后我们会经常遇到。这种现象表明经典统计在自觉或不自觉地使用特定的贝叶斯推断。其次，考察表 4.1 中的数据，不难看出 θ 的后验期望估计 $\hat{\theta}_E$ 要比后验众数估计 $\hat{\theta}_{MD}$（即最大似然估计）更合理一些，而且从下一小节知道后验期望估计在所有的参数 θ 的估计中的后验均方差最小，所以人们经常选用后验期望估计作为 θ 的贝叶斯估计。这样，在这个统计模型中贝叶斯估计就优于经典统计的最大似然估计，而且这里并没有用到先验信息，因为 $Beta(1,1)$ 是无信息先验。换句话说，这里参数 θ 的贝叶斯估计用到的信息与经典统计中 θ 的最大似然估计用到的信息是一样的，但是，结果是前者优于后者，这再一次令人感到惊奇！

表 4.1　成功概率 θ 的二种贝叶斯估计的比较

试验编号	试验次数	成功次数	$\hat{\theta}_{MD}$	$\hat{\theta}_{E}$
1	5	0	0	0.143
2	10	0	0	0.083
3	5	5	1	0.857
4	10	10	1	0.917

4.1.2 贝叶斯估计优良性准则

在经典统计中，比较估计量优良性的一种准则是看均方差的大小，均方差越小，估计量越好。对于贝叶斯统计，我们有类似的准则来评定一个贝叶斯估计的优良性。具体定义如下。

定义 4.2 设参数θ的后验分布为$\pi(\theta|\mathbf{x})$，其中$\mathbf{x}=(x_1,\cdots,x_n)$是已知样本，又设$\hat{\theta}$是$\theta$的一个贝叶斯估计，则$(\theta-\hat{\theta})^2$的后验期望

$$PMSE(\hat{\theta})=E^{\theta|\mathbf{x}}(\theta-\hat{\theta})^2=E[(\theta-\hat{\theta})^2|\mathbf{x}]$$

称为$\hat{\theta}$的后验均方差，其平方根$[PMSE(\hat{\theta})]^{1/2}$称为$\hat{\theta}$的后验标准误。如果$\hat{\theta}_1$和$\hat{\theta}_2$是$\theta$的两个贝叶斯估计且$PMSE(\hat{\theta}_1)<PMSE(\hat{\theta}_2)$，则称在后验均方差准则下$\hat{\theta}_1$优于$\hat{\theta}_2$。

注：当$\hat{\theta}$为θ的后验期望$\hat{\theta}_E=E(\theta|\mathbf{x})$时，有

$$PMSE(\hat{\theta})=E^{\theta|\mathbf{x}}(\theta-\hat{\theta}_E)^2=Var(\theta|\mathbf{x})$$

并称之为θ的后验方差，其平方根$[Var(\theta|\mathbf{x})]^{1/2}$称为后验标准差。对于$\theta$的任一个贝叶斯估计$\hat{\theta}$，其后验均方差与$\theta$的后验方差有如下关系

$$PMSE(\hat{\theta})=E^{\theta|\mathbf{x}}(\theta-\hat{\theta})^2=E^{\theta|\mathbf{x}}[(\theta-\hat{\theta}_E)+(\hat{\theta}_E-\hat{\theta})]^2=Var(\theta|\mathbf{x})+(\hat{\theta}_E-\hat{\theta})^2$$

这表明，当$\hat{\theta}$取后验均值$\hat{\theta}_E=E(\theta|\mathbf{x})$时，后验均方差达到最小。换句话说，后验均值估计$\hat{\theta}_E=E(\theta|\mathbf{x})$是后验均方差准则下的最优估计，所以实际中常取后验均值作为θ的贝叶斯估计。

例 4.2 在例 4.1 中已经知道对于二项分布总体，如果选用贝塔分布 $Beta(\alpha,\beta)$ 为先验分布，那么，成功概率 θ 的后验分布为另一个贝塔分布 $Beta(\alpha+x,\beta+n-x)$。（1）试求 θ 的后验方差；（2）当先验分布为 $Beta(1,1)$ 时，试求 θ 的后验期望估计 $\hat{\theta}_E$ 和后验众数估计 $\hat{\theta}_{MD}$ 的后验均方差并加以比较。

解：（1）根据贝塔分布的性质，不难求得 θ 的后验方差为

$$\operatorname{var}(\theta|x)=\frac{(x+\alpha)(n-x+\beta)}{(n+\alpha+\beta)^2(n+1+\alpha+\beta)}$$

（2）由例 4.1 知，这时 θ 的后验期望估计 $\hat{\theta}_E$ 和后验众数估计 $\hat{\theta}_{MD}$ 分别为

$$\hat{\theta}_E=\frac{x+1}{n+2},\quad \hat{\theta}_{MD}=\frac{x}{n}$$

根据（1），这时 θ 的后验方差为

$$\operatorname{var}(\theta|x)=\frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}$$

显然，$\hat{\theta}_E$ 的后验均方差就是后验方差 $Var(\theta|x)$，而 $\hat{\theta}_{MD}$ 的后验均方差为

$$PMSE(\hat{\theta}_{MD})=\operatorname{var}(\theta|x)+(\hat{\theta}_E-\hat{\theta}_{MD})^2=\frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}+\left(\frac{x+1}{n+2}-\frac{x}{n}\right)^2$$

所以，$PMSE(\hat{\theta}_E)<PMSE(\hat{\theta}_{MD})$，即 θ 的后验期望估计 $\hat{\theta}_E$ 优于 θ 的后验众数估计 $\hat{\theta}_{MD}$。

4.1.3 区间估计

在贝叶斯统计中，区间估计问题处理简明、含义清晰、解释易懂。下面给出正式定义。

定义 4.3 设给定的样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 来自总体 $p(x|\theta)$ 而且参数 θ 的后验分布为 $\pi(\theta|\mathbf{x})$。对于给定的概率 $1-\alpha$（一般而言，α 是小于或等于 0.1 的正数），（1）如果可找到二个统计量 $\hat{\theta}_L=\hat{\theta}_L(\mathbf{x})$ 和 $\hat{\theta}_U=\hat{\theta}_U(\mathbf{x})$，使得

$$P(\hat{\theta}_L\le\theta\le\hat{\theta}_U|\mathbf{x})\ge 1-\alpha$$

则称区间 $[\hat{\theta}_L,\hat{\theta}_U]$ 为**参数 θ 的可信水平（度）为 $1-\alpha$ 的贝叶斯可信区间**（或**区间估计**）也可简称为 θ **的 $1-\alpha$ 可信区间**（**区间估计**）；（2）如果可找到统计量 $\hat{\theta}_L=\hat{\theta}_L(\mathbf{x})$，使得

$$P(\theta\ge\hat{\theta}_L|\mathbf{x})\ge 1-\alpha$$

则称 $\hat{\theta}_L$ 为 θ **的 $1-\alpha$ 可信下限**；（3）如果可找到统计量 $\hat{\theta}_U=\hat{\theta}_U(\mathbf{x})$，使得

$$P(\theta\le\hat{\theta}_U|\mathbf{x})\ge 1-\alpha$$

则称 $\hat{\theta}_U$ 为 θ **的 $1-\alpha$ 可信上限**。

注：

1. 这里术语“可信区间（Credible interval）”等与经典统计中的术语“置信区间（Confidence interval）”等不同，不要混淆。虽然在贝叶斯统计中偶尔也有人用术语“置信区间”，但不是主流。

2. 当θ为连续型随机变量时，定义中的三个不等式可以改为等式。只是当θ为离散型随机变量时，因为对给定的概率$1-\alpha$，可信区间（下限，上限）不一定存在，所以这时略微放大左端后验概率，以便得到可信区间（下限，上限）。

3. 如果求出了θ的可信水平为 0.95 的可信区间（区间估计）$[a,b]$，那你可以写出

$$P(a \le \theta \le b \mid \mathbf{x}) = 0.95$$

并大大方方地说：“θ属于区间$[a,b]$的概率为 0.95。”但是，对经典统计的置信区间就不能这么说，因为经典统计认为θ是未知常量，它要么在区间$[a,b]$内，要么在此区间外，所以不能说：“θ在区间$[a,b]$内的概率为 0.95”，而只能说：“在 100 次重复使用这个置信区间时，大约有 95 次能覆盖住θ。” 这对于非统计专业的人来说，是非常别扭和不易理解的。

例 4.3 在例 4.1 中已经知道对于二项分布总体，如果选用贝塔分布 $Beta(\alpha,\beta)$ 为先验分布，那么，成功概率 θ 的后验分布为另一个贝塔分布 $Beta(\alpha+x,\beta+n-x)$。现在通过 10 次独立试验得到成功次数 $x=9$，而且知道先验分布为 $Beta(0.5,0.5)$，求参数 θ 的后验均值估计和 95%区间估计。

解：（1）当先验分布为 $Beta(0.5,0.5)$ 时，后验分布为 $Beta(9.5,1.5)$，所以参数 θ 的后验均值估计为 $\hat{\theta}_E=9.5/(9.5+1.5)=0.8636$。（2）求 θ 的 95%区间估计就是要找到两个统计量 $\hat{\theta}_L<\hat{\theta}_U$ 使 $\theta\in[\hat{\theta}_L,\hat{\theta}_U]$ 的后验概率等于 0.95，即

$$P(\hat{\theta}_L\le\theta\le\hat{\theta}_U\,|x)=0.95$$

我们分别找 $\hat{\theta}_L$ 和 $\hat{\theta}_U$ 使

$$P(\theta<\hat{\theta}_L\,|x)=0.025,\quad P(\theta\le\hat{\theta}_U\,|x)=0.975$$

即 $\hat{\theta}_L$ 和 $\hat{\theta}_U$ 分别是 0.025 分位数和 0.975 分位数。这样就有

$$P(\hat{\theta}_L\le\theta\le\hat{\theta}_U\,|x)=P(\theta\le\hat{\theta}_U\,|x)-P(\theta<\hat{\theta}_L\,|x)=0.95$$

利用如下 R 命令就可求得 θ 的 95%区间估计为[0.6187, 0.9890]。

```
qbeta(c(0.025,0.975), 9.5,1.5)
[1] 0.6186852 0.9889883
```

§4.2 泊松分布参数的估计

4.2.1 后验分布

设样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 来自泊松分布 $Poisson(\lambda)$，其概率函数是

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

例 2.3 证明了伽玛分布 $Gamma(\alpha,\beta)$ 是均值（方差）λ 的共轭先验分布，且此时的后验分布是 $Gamma(\alpha+n\overline{x},\beta+n)$。例 3.16 证明了 $\pi(\lambda)=\lambda^{-1/2}$ 是 λ 的杰弗里斯无信息先验，此时 λ 的后验分布是

$$\pi(\lambda\mid\mathbf{x})\propto p(\mathbf{x}\mid\lambda)\pi(\lambda)\propto\lambda^{n\overline{x}-1/2}e^{-n\lambda}$$

这正是伽玛分布 $Gamma(n\overline{x}+1/2,n)$， 其中 $\overline{x}$ 是样本均值。

4.2.2 参数估计

1. 当泊松分布的均值（方差）λ 取伽玛分布 $Gamma(\alpha,\beta)$ 为共轭先验时，在例 2.3 中已经求得 λ 的后验期望为

$$E(\lambda|\mathbf{x})=\frac{n\bar{x}+\alpha}{\beta+n}=\frac{n}{\beta+n}\times\bar{x}+\frac{\beta}{\beta+n}\times\frac{\alpha}{\beta}$$

其中 $\bar{x}$ 是样本均值，因此，此时的泊松分布的均值（方差）λ 的后验期望估计为

$$\hat{\lambda}_B=E(\lambda|\mathbf{x})=\frac{n}{\beta+n}\times\bar{x}+\frac{\beta}{\beta+n}\times\frac{\alpha}{\beta}$$

它是样本均值与先验均值之间的一种加权平均。

2. 当先验取为杰弗里斯无信息先验 $\pi(\lambda)=\lambda^{-1/2}$ 时，后验分布是伽玛分布 $Gamma(n\bar{x}+1/2,n)$。因此，此时的泊松分布的均值（方差）λ 的后验期望估计为

$$\hat{\lambda}_B=E(\lambda|\mathbf{x})=\frac{n\bar{x}+\dfrac{1}{2}}{n}=\bar{x}+\frac{1}{2n}$$

案例 4.1 （某国受教育程度不同妇女生育率比较研究）在 20 世纪 90 年代的一项综合社会调查中收集了 155 名妇女受教育程度和她们生育孩子个数的数据。这些妇女在 1970 年代是 20 来岁的年龄，这段时间是某国历史上一段低生育率的时期。在本案例中，我们将比较大学本科及以上学历（以下简称为有大学文凭）的妇女的生育率与那些没有大学文凭的妇女的生育率。设

$$X_{11}, X_{21}, \cdots, X_{m1}$$

表示 m 个没有大学文凭的妇女各自生育的小孩数，而

$$X_{12}, X_{22}, \cdots, X_{n2}$$

表示 n 个有大学文凭的妇女各自生育的小孩数。进一步，设

$$X_{11}, X_{21}, \cdots, X_{m1} \mid \theta_1 \sim i.i.d.\ Poisson(\theta_1)$$

$$X_{12}, X_{22}, \cdots, X_{n2} \mid \theta_2 \sim i.i.d.\ Poisson(\theta_2)$$

根据调查收集到的数据，我们可以得到如下统计量

对于没有大学文凭的妇女： $m = 111, X_1 = \sum_{i=1}^{111} X_{i1} = 217, \bar{X}_1 = 1.95$

对于有大学文凭的妇女： $n = 44, X_2 = \sum_{i=1}^{44} X_{i2} = 66, \bar{X}_2 = 1.50$

根据先验信息，现在取伽玛分布 $Gamma(2,1)$ 作为 θ_1 和 θ_2 的先验分布，那么我们可以分别得到它们的后验分布如下

$$\theta_1 \mid (m=111, X_1=217) \sim Gamma(2+217, 1+111) = Gamma(219,112)$$

$$\theta_2 \mid (n=44, X_2=66) \sim Gamma(2+66, 1+44) = Gamma(68,45)$$

对于 θ_1 容易求得它的后验期望估计 $\hat{\theta}_1 = 219/112 = 1.955$，利用如下 R 命令可求得 95%可信区间为[1.705, 2.223]。

```
> qgamma(c(0.025,0.975),219,112)
[1] 1.704943 2.222679
```

类似可得 θ_2 的后验期望估计 $\hat{\theta}_2 = 68/45 = 1.511$ 和 95%可信区间为[1.173, 1.891]。所以从两个后验期望估计来看，没有大学文凭的妇女平均生育率 θ_1 大约为 2 而有大学文凭的妇女平均生育率 θ_2 大约为 1.5。另外，我们还可以画出 θ_1 和 θ_2 的后验密度曲线（图 4.1），从该图我们可以清晰地看出 θ_1 明显大于 θ_2。

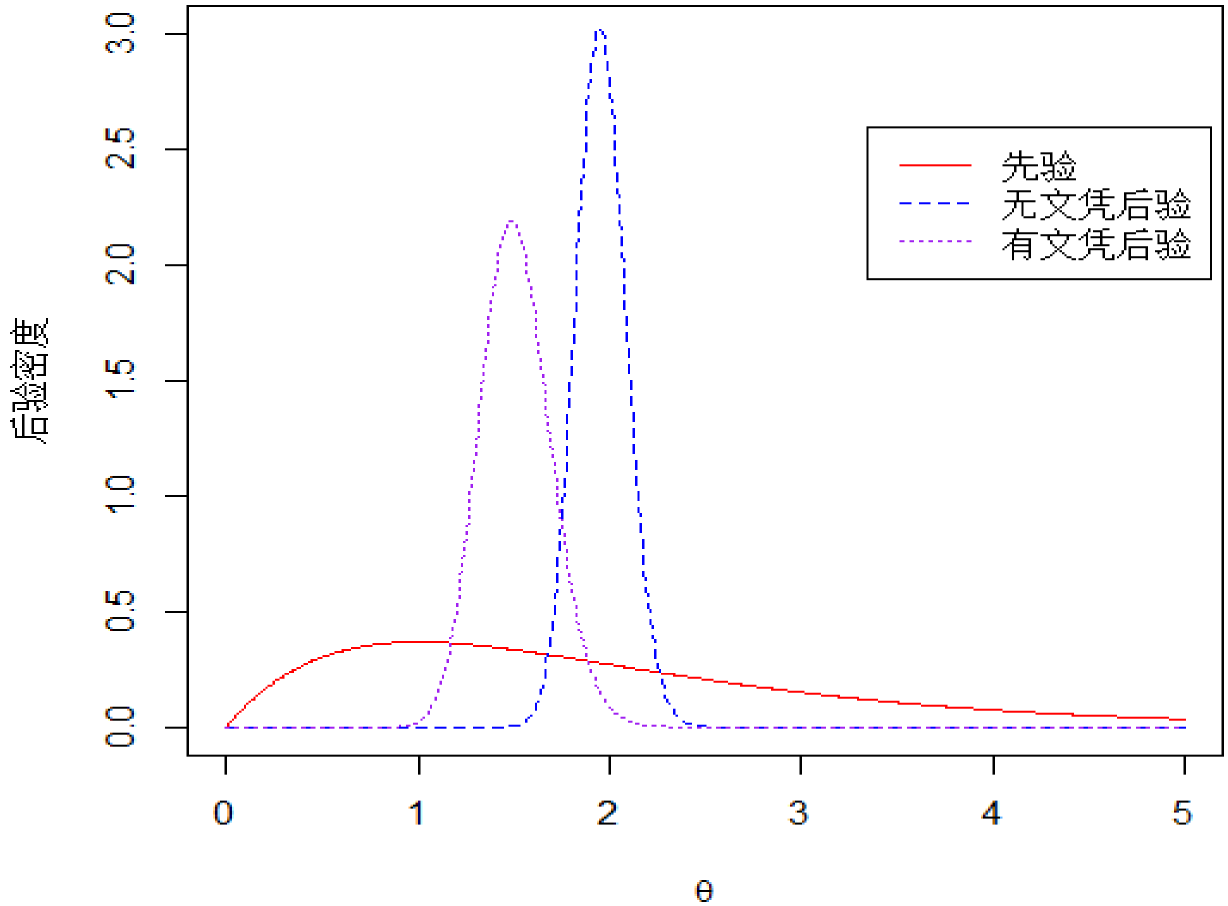


图 4.1 某国有无大学文凭妇女生育率后验密度比较图

§ 4.3 指数分布参数的估计

4.3.1 参数估计

在第 2 章中引入了逆（或倒）伽玛分布 $IGamma(\alpha,\beta)$，其密度函数为

$$p(y|\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}y^{-(\alpha+1)}\exp(\frac{-\beta}{y}),y>0$$

在本节我们要用到它。现在设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自指数分布 $\mathrm{Exp}(\lambda)$的样本，指数分布的密度为

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

（1）不难验证逆伽玛分布 $IGamma(\alpha,\beta)$ 是参数 λ 的共轭先验分布，而且后验分布为逆伽玛分布 $IGamma\left(\alpha+n,\beta+n\bar{x}\right)$。事实上，此时后验分布

$$\pi(\lambda\,|\,\mathbf{x})\propto p(\mathbf{x}\,|\,\lambda)\pi(\lambda)\propto\lambda^{-n}e^{-n\bar{x}\lambda^{-1}}\lambda^{-(\alpha+1)}e^{-\beta\lambda^{-1}}=\lambda^{-(\alpha+n+1)}e^{-(\beta+n\bar{x})\lambda^{-1}}$$

其中 $\bar{x}$ 是样本均值。因此，此时参数 λ 的后验期望估计为

$$\hat{\lambda}_B=E(\lambda\,|\,\mathbf{x})=\frac{\beta+n\bar{x}}{\alpha+n-1}=\frac{\alpha-1}{\alpha+n-1}\times\frac{\beta}{\alpha-1}+\frac{n}{\alpha+n-1}\times\bar{x}$$

（2）当把指数分布的密度

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

全体看成一个尺度参数族时，参数 λ 的无信息先验 $\pi(\lambda)=\lambda^{-1}$ 而且例 3.15 验证了此时后验分布为 $IGamma(n,n\bar{x})$。于是参数 λ 的的后验期望估计为

$$\hat{\lambda}_B=E(\lambda|\mathbf{x})=\frac{n\bar{x}}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}x_i$$

（3）习题 3.14 求出了参数 λ 的杰弗里斯无信息先验为 $\pi(\lambda)=\lambda^{-1}$， 因此，此时有把指数分布的密度全体看成一个尺度参数族的情形一样的结果。

4.3.2 案例：国产彩电的寿命有多长？

案例 4.2（本案例是华东师范大学茆诗松教授等人研究的项目，这里对可信下限的计算进行了化简。）经过早期筛选的彩色电视机（简称彩电）的寿命T服从指数分布，它的密度函数和分布函数分别为

$$p(t|\theta)=\theta^{-1}e^{-t/\theta}, t>0,\quad F(t|\theta)=1-e^{-t/\theta}$$

其中$\theta>0$且$E(T)=\theta$，即θ是彩电的平均寿命。

现在从一批彩电中随机抽取n台进行寿命试验，试验到第$r(\le n)$台失效时为止，记它们的失效时间为$t_1\le t_2\le\cdots\le t_r$，另外$n-r$台彩电直到试验停止时还未失效，这样的试验称为截尾寿命试验，所得样本$\mathbf{t}=(t_1,\cdots,t_r)$称为截尾样本，此截尾样本的联合密度函数为

$$p(\mathbf{t}|\theta)\propto[\prod_{i=1}^{r}p(t_i|\theta)][1-F(t_r|\theta)]^{n-r}=\theta^{-r}\exp\{-s_r/\theta\}$$

其中$s_r=t_1+...+t_r+(n-r)t_r$称为总试验时间。

为寻求彩电平均寿命θ的贝叶斯估计，首先需要确定θ的先验分布，根据国内外的经验，选用逆伽玛分布$IGamma(\alpha,\beta)$作为θ的先验分布$\pi(\theta)$是可行的，于是，参数θ的后验密度

$$\pi(\theta|\mathbf{t})\propto p(\mathbf{t}|\theta)\pi(\theta)\propto\theta^{-(\alpha+r+1)}e^{-(\beta+s_r)/\theta}$$

显然，这是逆伽玛分布的核，故θ的后验分布为$IGamma(\alpha+r,\beta+s_r)$。若取后验均值作为$\theta$的贝叶斯估计，则有

$$\hat{\theta}=E(\theta|\mathbf{t})=\frac{\beta+s_r}{\alpha+r-1}$$

接下来的任务就是要确定超参数α与β的值。我国各彩电生产厂过去做了大量的彩电寿命试验，我们从 15 个彩电生产厂的实验室和一些独立实验室就收集到 13142 台彩电的寿命试验数据，共计 5369812 台时，此外还有 9240 台彩电进行三年现场跟踪试验，总共进行了 5547810 台时试验，在这些试验中总共失效台数不超过 250 台。对如此大量先验信息加工整理后，确认我国彩电平均寿命不低于 30000 小时，它的 10%的分位数$\theta_{0.1}$大约为 11250 小时，经过一些专家认定，这两个数据是符合我国前几年彩电寿命的实际情况，并且是留有余地的。由此以及逆伽玛分布的性质，可列出如下二个方程

$$\begin{cases} \dfrac{\beta}{\alpha-1} = 30000 \\ \int_0^{11250} \pi(\theta)d\theta = 0.1 \end{cases}$$

其中先验$\pi(\theta)$为逆伽玛分布$IGamma(\alpha,\beta)$的密度函数，它的数学期望为$E(\theta)=\beta/(\alpha-1)$。用计算机软件解此方程组，可得

$$\alpha=1.956，\quad \beta=2868$$

这样我们就得到θ的先验分布密度$IGamma(1.956,2868)$以及后验分布密度$IGamma(1.956+r,2868+s_r)$。

现随机抽取 100 台彩电，在规定条件下连续进行 400 小时的寿命试验，结果是没有一台失效。这时总试验时间为

$$s_r = 100 \times 400 = 40000 \text{ 小时}，\quad r = 0$$

从而 θ 的后验分布密度为 $IGamma(1.956, 42868)$，据此，彩电的平均寿命 θ 的贝叶斯估计为

$$\hat{\theta} = \frac{2868 + 40000}{1.956 - 1} = 44841 \text{（小时）}$$

利用后验分布 $IGamma(1.956, 42868)$ 还可获得 θ 的可信下限，下面就来求可信水平为 $1-\gamma = 0.9$ 的可信下限。首先易知 $\theta^{-1} \sim Gamma(1.956, 42868)$。如设 θ_L 为 θ 的 0.9 可信下限，则

$$P(\theta \geq \theta_L \mid \mathbf{t}) = 0.9，\text{ 即 } P(\theta^{-1} \leq \theta_L^{-1} \mid \mathbf{t}) = 0.9$$

也就是说 θ_L^{-1} 是伽玛分布 $Gamma(1.956, 42868)$ 的 0.9 分位数，利用如下 R 命令可得此分位数 $\theta_L^{-1} = 8.9207 \times 10^{-5}$，从而 θ 的可信下限

$$\hat{\theta}_L = 1/(8.9207 \times 10^{-5}) = 11209.88 \text{（小时）}$$

```
 qgamma(0.9,1.956,42868)
[1] 8.920722e-05
```

§ 4.4 正态分布参数的估计

本节讨论在各种不同的情形下，正态分布 $N(\theta,\sigma^2)$ 的均值和方差的贝叶斯估计问题。

4.4.1 方差已知时均值的估计

现在设 $x=(x_1,\cdots,x_n)$ 是来自正态总体 $N(\theta,\sigma^2)$ 的样本，在方差 σ^2 已知时，对均值参数 θ 的先验和后验分布在前面不同的章节中做了讨论，这里我们来总结一下并求出相应的贝叶斯估计。

1. 例 2.1 证明了正态总体 $N(\theta,\sigma^2)$ 的均值 θ 有共轭先验 $N(\mu,\tau^2)$ 而且后验是正态分布 $N(\mu_1,\tau_1^2)$，其中

$$\mu_1=\frac{\bar{x}\sigma_0^{-2}+\mu\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}},\quad \frac{1}{\tau_1^2}=\frac{1}{\sigma_0^2}+\frac{1}{\tau^2},\quad \sigma_0^2=\frac{\sigma^2}{n}$$

从而，这时 θ 的贝叶斯估计为

$$\hat{\theta}_B=E(\theta\,|\,\mathbf{x})=\mu_1=\frac{\bar{x}\sigma_0^{-2}+\mu\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}}=\frac{\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}}\mu+\frac{\sigma_0^{-2}}{\sigma_0^{-2}+\tau^{-2}}\bar{x}$$

2. 若 θ 无先验信息，那么无论把正态总体看作位置参数族还是用杰弗里斯方法求出的无信息先验都是 $\pi(\theta)=1$。这时，样本均值 $\bar{x}$ 是 θ 的充分统计量，从而后验分布

$$\pi(\theta|\mathbf{x})=\pi(\theta|\bar{x})\propto p(\bar{x}|\theta)\pi(\theta)\propto\exp\{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\}$$

这是正态分布 $N(\bar{x},\sigma^2/n)$。故这时 θ 的贝叶斯估计与经典统计中的一样都为样本均值 $\bar{x}=(1/n)\sum_{i=1}^{n}x_i$。

4.4.2 均值已知时方差的估计

1. 共轭先验分布。在例 2.2 中，我们找到了方差σ^2的共轭先验为逆伽玛分布$IGamma(\alpha,\beta)$，而且后验分布为$IGamma(\alpha+\frac{n}{2},\beta+\frac{ns^2}{2})$，其中$ns^2=\sum_{i=1}^{n}(\theta-x_i)^2$。故这时方差$\sigma^2$的后验期望估计

$$\hat{\sigma}^2=E(\sigma^2\,|\,\mathbf{x})=\frac{\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}{\alpha+\frac{n}{2}-1}=\frac{2\beta+\sum_{i=1}^{n}(x_i-\theta)^2}{2\alpha+n-2}$$

2. 无先验信息。对于标准差σ，此时无论作为尺度参数还是用杰弗里斯方法得到的无信息先验分布都是$\pi(\sigma)=\sigma^{-1}$，从而易知方差σ^2的无信息先验是$\pi(\sigma^2)=\sigma^{-2}$。于是，方差$\sigma^2$的后验分布

$$\pi(\sigma^2\,|\mathbf{x})\propto p(\mathbf{x}|\sigma^2)\pi(\sigma^2)$$

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

这正是逆伽玛分布$IGamma(n/2,ns^2/2)$。故这时方差σ^2的后验期望估计

$$\hat{\sigma}^2=E(\sigma^2\,|\,\mathbf{x})=\frac{\frac{ns^2}{2}}{\frac{n}{2}-1}=\frac{n}{n-2}s^2$$

4.4.3 均值和方差的同时估计

对于实际问题而言，总体均值和方差往往是同时未知的，因此都需要进行估计，这就是多参数的估计问题。下面分两种情形考虑。

1. 共轭先验分布。在例 2.5 中，我们已经求出了正态总体 $N(\mu,\sigma^2)$ 两参数即均值与方差 (μ,σ^2) 的（联合）共轭先验分布，它为正态--逆伽玛分布 $N-IGamma(\mu_0,\kappa_0,\upsilon_0,\sigma_0^2)$，其密度是

$$\begin{aligned}\pi(\mu,\sigma^2)&=\pi(\mu\,|\,\sigma^2)\pi(\sigma^2)\\&\propto(\sigma^2)^{-[(\upsilon_0+1)/2+1]}\exp\{-\frac{1}{2\sigma^2}[\upsilon_0\sigma_0^2+\kappa_0(\mu-\mu_0)^2]\}\end{aligned}$$

而对应的后验分布 $\pi(\mu,\sigma^2|\mathbf{x})$ 是 $N-IGamma(\mu_n,\kappa_n,\upsilon_n,\sigma_n^2)$，其中参数

$$\mu_n=\frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{x},\ \kappa_n=\kappa_0+n,\ \upsilon_n=\upsilon_0+n$$

$$\upsilon_n\sigma_n^2=\upsilon_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\mu_0-\bar{x})^2，\ (n-1)s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

对于后验分布

$$\pi(\mu,\sigma^2|\mathbf{x})\propto(\sigma^2)^{-[(\upsilon_n+1)/2+1]}\exp\{-\frac{1}{2\sigma^2}[\upsilon_n\sigma_n^2+\kappa_n(\mu-\mu_n)^2]\}$$

受先验分布构成的启发，它也可看成

$$\pi(\mu,\sigma^2|\mathbf{x})=\pi(\mu|\sigma^2,\mathbf{x})\pi(\sigma^2|\mathbf{x})$$

其中 $\pi(\mu|\sigma^2,\mathbf{x})=N(\mu_n,\sigma^2/\kappa_n)$ ，$\pi(\sigma^2|\mathbf{x})=IGamma(\upsilon_n/2,\upsilon_n\sigma_n^2/2)$ 。于是，我们可以分别得到均值 μ 与方差 σ^2 的一种近似估计

$$\hat{\mu}=\mu_n=\frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{x}，\ \hat{\sigma}^2=\frac{\upsilon_n\sigma_n^2/2}{(\upsilon_n/2)-1}=\frac{\upsilon_n\sigma_n^2}{\upsilon_n-2}$$

2. 无先验信息。由例 3.17 知道当总体服从正态分布 $N(\mu,\sigma^2)$，但无参数 (μ,σ^2) 的先验信息时，(μ,σ^2) 的杰弗里斯先验为 $\pi(\mu,\sigma^2)=1/\sigma^2$。于是，参数向量 (μ,σ^2) 的后验分布

$$\pi(\mu,\sigma^2|\mathbf{x}) \propto p(\mathbf{x}|\mu,\sigma^2)\pi(\mu,\sigma^2)$$

$$\propto (\sigma^2)^{-(\frac{n}{2}+1)}\exp\left[-\frac{\sum_{i=1}^{n}(\mu-x_i)^2}{2\sigma^2}\right]=(\sigma^2)^{-(\frac{n}{2}+1)}\exp\left(-\frac{s^2+n(\mu-\overline{x})^2}{2\sigma^2}\right)$$

其中 $s^2=\sum_{i=1}^{n}(x_i-\overline{x})^2$。为了看出上式是什么分布的核，考虑自由度为 n 的 $\chi^2(n)$ 分布，其密度为

$$p(x)=\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$$

现在设随机变量 $X\sim\chi^2(n)$，再令 $Y_n=X^{-1}$，我们来求 Y_n 的分布 $p_{Y_n}(y)$。首先计算变换 $y=x^{-1}$ 的雅可比行列式的绝对值

$$|dx/dy|=|-y^{-2}|=y^{-2}$$

于是根据概率密度函数的运算法则，Y_n 的分布密度函数

$$p_{Y_n}(y)=p(x)|_{x=y^{-1}}\left|\frac{dx}{dy}\right|=p(y^{-1})y^{-2}=\frac{1}{2^{n/2-1}\Gamma(n/2)}y^{-(n/2+1)}\exp(-\frac{1}{2y})$$

我们称 Y_n 是自由度为 n 的逆 χ^2 变量并记 $Y_n\sim\chi^{-2}(n)$（逆 χ^2 分布）。现在令 $Z=s^2Y_n$，即 $Z\sim s^2\chi^{-2}(n)$，那么其密度

$$p_Z(z)=p_{Y_n}(y)|_{y=s^{-2}z}\left|\frac{dy}{dz}\right|=\frac{s^{n+6}z^{-(n/2+1)}}{2^{n/2-1}\Gamma(n/2)}\exp(-\frac{s^2}{2z})\propto z^{-(n/2+1)}\exp(-\frac{s^2}{2z})$$

另一方面，由

$$\pi(\mu,\sigma^2|\mathbf{x})\propto(\sigma^2)^{-(\frac{n}{2}+1)}\exp\left(-\frac{s^2+n(\mu-\bar{x})^2}{2\sigma^2}\right)$$

$$=\sigma^{-1}\exp\left(-\frac{n(\mu-\bar{x})^2}{2\sigma^2}\right)\times(\sigma^2)^{-(\frac{n-1}{2}+1)}\exp\left(-\frac{s^2}{2\sigma^2}\right)$$

令

$$\pi(\mu|\sigma^2,\mathbf{x})\propto\sigma^{-1}\exp\left(-\frac{n(\mu-\bar{x})^2}{2\sigma^2}\right),\quad \pi(\sigma^2|\mathbf{x})\propto(\sigma^2)^{-(\frac{n-1}{2}+1)}\exp\left(-\frac{s^2}{2\sigma^2}\right)$$

那么，显然有

$$\pi(\mu,\sigma^2|\mathbf{x})=\pi(\mu|\sigma^2,\mathbf{x})\pi(\sigma^2|\mathbf{x})$$

而且

$$\pi(\mu|\sigma^2,\mathbf{x})=N(\bar{x},\sigma^2/n),\quad \pi(\sigma^2|\mathbf{x})=s^2\chi^{-2}(n-1)$$

由此可见参数向量(μ,σ^2)的后验分布是正态分布与逆χ^2分布的积，于是，它们的一种估计是$\hat{\mu}=\bar{x}$，$\hat{\sigma}^2=E(s^2Y_{n-1})=s^2E(Y_{n-1})$，当然，这是一种近似的估计。下面来算$E(Y_{n-1})$。由于这里$Y_{n-1}\sim\chi^{-2}(n-1)$，所以

$$E(Y_{n-1})=\int_{-\infty}^{\infty}x^{-1}p(x)dx=\int_0^{\infty}\frac{1}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}x^{(n-3)/2-1}e^{-x/2}dx$$

$$=\frac{2^{(n-3)/2}\Gamma(\frac{n-3}{2})}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}\int_0^{\infty}\frac{1}{2^{(n-3)/2}\Gamma(\frac{n-3}{2})}x^{(n-3)/2-1}e^{-x/2}dx$$

$$=\frac{\Gamma(\frac{n-3}{2})}{2\Gamma(\frac{n-1}{2})}=\frac{1}{n-3}$$

于是

$$\hat{\sigma}^2=E(s^2Y_{n-1})=s^2E(Y_{n-1})=\frac{s^2}{n-3}=\frac{1}{n-3}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

换句话说，这种估计比经典统计中的常用估计略大，但没有本质不同。

由于直接求参数向量(μ,σ^2)的后验期望估计并不容易，所以在实践中常用随机模拟的方法来求之，下面用一个体育领域的案例来说明具体的做法。

4.4.4 案例：无先验信息如何估计马拉松成绩分布的参数

案例 4.3 在 R 软件包 BayesianStat 中的数据集 marathontime 收集了二十位运动员在一场马拉松比赛中的成绩（时间单位：分）。假设这些成绩服从正态分布 $N(\mu,\sigma^2)$，但对于两参数 (μ,σ^2) 无任何先验信息。为了估计均值和方差 (μ,σ^2)，我们可取无信息先验为杰弗里斯先验 $\pi(\mu,\sigma^2)=1/\sigma^2$，这样，就可利用前面推导出的联合后验分布。我们首先来画出后验密度的等高线，所用 R 命令如下，其中，命令 data 是将数据集 marathontime 引入 R 平台；命令 attach 使得数据集 marathontime 中的抬头 mtime 可以作为代表这个数据列的对象来使用；函数 Normki2post 用来计算联合后验分布的对数密度（为使数量变小些）；命令 icontour 则画出后验密度的等高线（图 4.2），c 数组中前两个是横轴取值范围，后两个是纵轴取值范围，mtime 是已知样本。

```
library(BayesianStat)
data(marathontime)
attach(marathontime)
icontour(Normki2post, c(220, 330, 500, 9000),mtime,xlab="均值",ylab="方差")
```

接下来，我们来模拟参数(μ,σ^2)联合后验分布的样本。(μ,σ^2)的模拟值可以分两步得到，首先用分布$\pi(\sigma^2|\mathbf{x})=s^2\chi^{-2}(n-1)$将$\sigma^2$模拟出来，然后用分布$\pi(\mu|\sigma^2,\mathbf{x})=N(\bar{x},\sigma^2/n)$模拟$\mu$，这样就得到参数$(\mu,\sigma^2)$的模拟样本。现在我们来模拟1000对这样的样本，模拟所用R命令如下，其中第三个命令模拟σ^2；第四个命令模拟μ；最后一个命令是将模拟得到的样本画到图4.2上。

```
s2<-sum((mtime - mean(mtime))^2)
n <-length(mtime)
sigma2<- s2/rchisq(1000, n-1)
mu<-rnorm(1000, mean=mean(mtime),sd =sqrt(sigma2)/sqrt(n))
points(mu, sigma2)
```

现在我们可以利用模拟样本来估计均值和方差这两个参数了。利用 R 命令 mean(mu)和 mean(sigma2)，我们就分别得到均值的后验期望估计 $\hat{\mu}=278.19$ 和方差的后验期望估计 $\hat{\sigma}^2=2712.18$。再利用 R 命令 quantile，我们就可求得两个参数的 95%可信区间分别为(254.3, 301.5)和(1454.4, 5157.9)，R 命令 quantile 的具体用法如下：

```
 quantile(mu, c(0.025, 0.975))
     2.5%      97.5%
254.2789 301.4753
 quantile(sigma2, c(0.025, 0.975))
     2.5%      97.5%
1454.386 5157.870
```

这个案例所使用的随机模拟方法是贝叶斯统计中极为重要的统计推断计算方法（其实在经典统计中也是如此），由于计算机科学技术的巨大进步以及有效算法的出现，模拟方法是近几十年贝叶斯统计学的热门课题，解决了大量的不同领域的复杂统计推断问题，本书将用专章对有关内容做概要介绍。

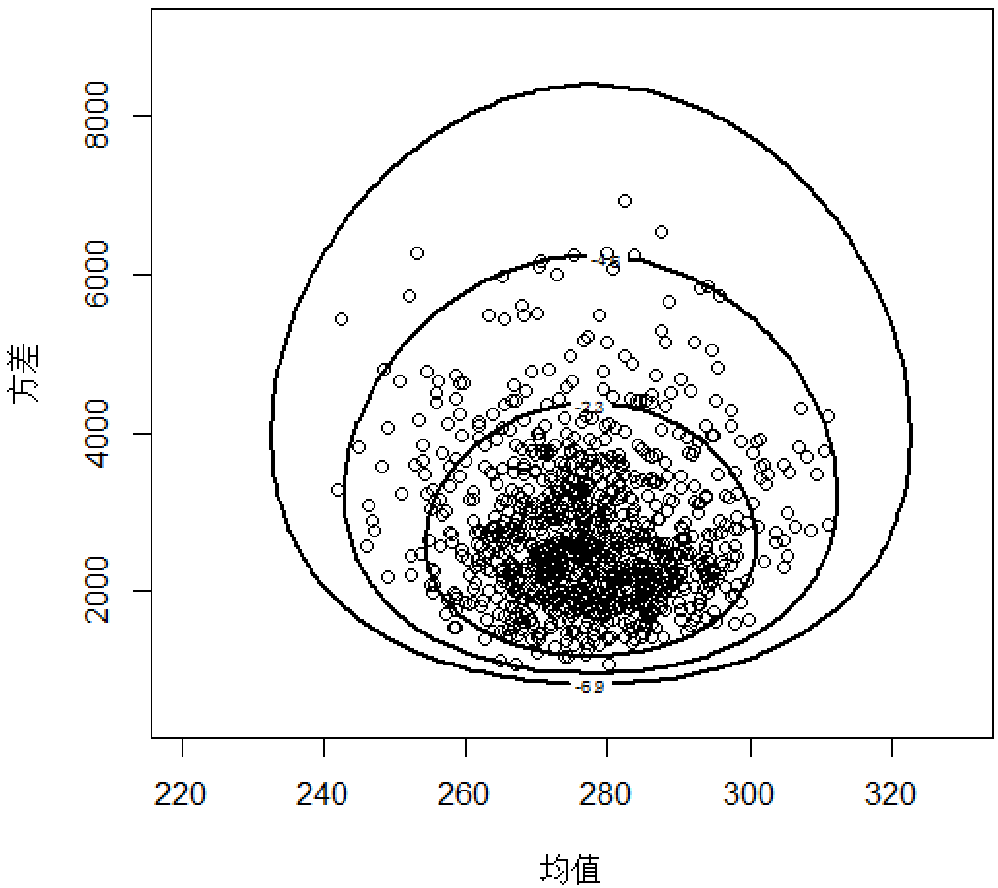
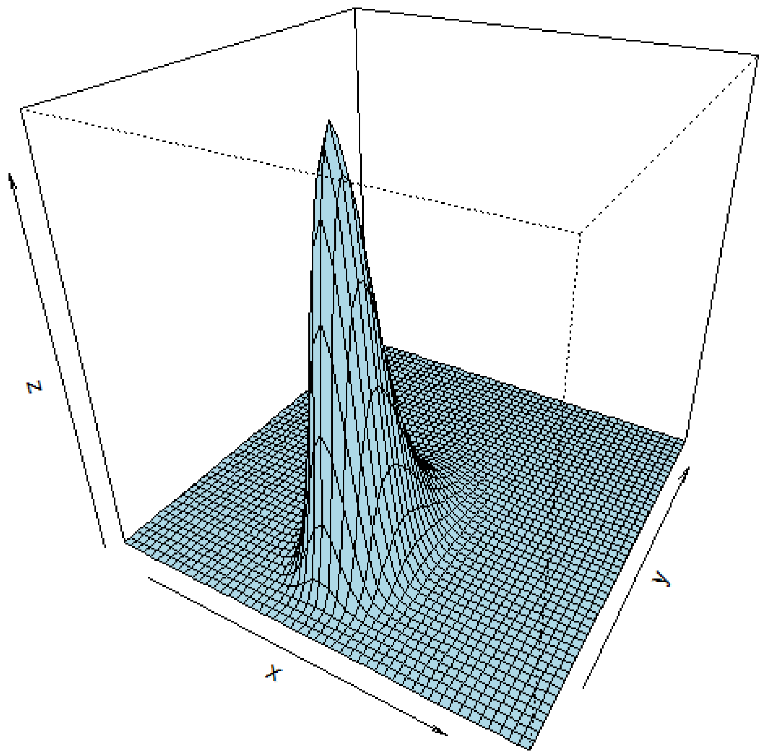


图 4.2 后验密度等高线与模拟样本图

后验密度曲面：



§4.5 贝叶斯假设检验

假设检验是一类重要的统计推断问题，但是，在经典统计中处理假设检验是相当不容易的，它需要选择检验统计量，确定抽样分布等等。然而，在贝叶斯统计中，处理假设检验则容易得多而且直截了当。不仅如此，贝叶斯统计还可以进行多假设的检验。本节将对贝叶斯假设检验做较详细的讨论。

4.5.1 贝叶斯假设检验与贝叶斯因子

我们知道要进行假设检验，首先要建立起假设。设我们观察到来自总体 $p(x|\theta)$ 的样本 $\mathbf{x}=(x_1,\cdots,x_n)$，其中参数 θ 属于参数空间 Θ。现在建立原假设 H_0 与备择假设 H_1 如下：

$$H_0:\theta\in\Theta_0,\quad H_1:\theta\in\Theta_1$$

其中 Θ_0 与 Θ_1 是 Θ 的两个非空子集且满足 $\Theta_0\cup\Theta_1=\Theta$, $\Theta_0\cap\Theta_1=\Phi$（空集）这时称 Θ_0 与 Θ_1 是 Θ 的一个划分。在贝叶斯统计中，当获得后验分布 $\pi(\theta|\mathbf{x})$ 后，我们计算两个假设 H_0 与 H_1 的后验概率

$$\alpha_0=P(H_0|\mathbf{x})=P(\theta\in\Theta_0|\mathbf{x}),\ \alpha_1=P(H_1|\mathbf{x})=P(\theta\in\Theta_1|\mathbf{x})$$

并称 α_0/α_1 为后验概率比（也称为后验机会比）。对于上面的两个假设检验问题，杰弗里斯提出了如下的准则。

杰弗里斯假设检验准则：（1）当$\alpha_0/\alpha_1>1$时，接受H_0；（2）当$\alpha_0/\alpha_1<1$时，接受H_1；（3）当$\alpha_0/\alpha_1\approx1$时，不宜立即下任何结论，需要进一步收集抽样信息和（或）先验信息，然后再次计算后验分布和后验机会比并重新判断接受还是拒绝原假设。

注：

1. 这种假设检验方法简单明了，无需象经典统计中那样去做寻求检验统计量、确定抽样分布等等困难的工作。

2. 显然，后验机会比越大，接受原假设的可信度越强。但是，杰弗里斯假设检验准则是基于先验分布和样本来做出选择的，并不能保证这种选择是完全正确无误的，它只是表明一个假设比另一个假设更可信，这点与经典统计中的假设检验是类似的。

3. 易知$\alpha_0/\alpha_1>1$等价于$\alpha_0>1/2$，所以当$\alpha_0>1/2$时，也接受H_0。

例 4.4 设二项分布 $Bin(n,\theta)$ 的参数 θ 的先验取为均匀分布 $U(0,1)$， x 是取自该分布的一个样本，现考虑如下两个假设

$$H_0:\theta\in\Theta_0=\{\theta;0<\theta\le 1/2\},\ H_1:\theta\in\Theta_1=\{\theta;1/2<\theta<1\}$$

根据例 1.5，参数 θ 的后验分布为贝塔分布 $Beta(x+1,n-x+1)$，于是假设 H_0 与 H_1 的后验概率分别为

$$\alpha_0=P(\theta\in\Theta_0|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^{1/2}\theta^x(1-\theta)^{n-x}d\theta$$

$$\alpha_1=P(\theta\in\Theta_1|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\int_{1/2}^{1}\theta^x(1-\theta)^{n-x}d\theta$$

并且 $\alpha_1=1-\alpha_0$。现在设 $n=5$，则由如下 R 命令容易算出样本 x 各种取值下的后验概率及后验机会比，由计算结果可见当 $x=0,1,2$ 时，接受 H_0； 当 $x=3,4,5$ 时，拒绝 H_0 而接受 H_1。

```
x<-c(0,1,2,3,4,5)
pbeta(0.5,x+1,5-x+1)
[1] 0.984375    0.890625    0.656250    0.343750    0.109375    0.015625
1-pbeta(0.5,x+1,5-x+1)
[1] 0.015625    0.109375    0.343750    0.656250    0.890625    0.984375
pbeta(0.5,x+1,5-x+1)/(1-pbeta(0.5,x+1,5-x+1))
[1] 63.00000000 8.14285714 1.90909091    0.52380952    0.12280702    0.01587302
```

前面我们定义了后验机会比并用它来决定假设的接受或拒绝。如果我们有正常的先验分布 $\pi(\theta)$，显然也可以定义先验机会比（先验概率比）如下

$$\frac{\pi_0}{\pi_1}=\frac{P_\pi(H_0)}{P_\pi(H_1)}=\frac{P_\pi(\theta\in\Theta_0)}{P_\pi(\theta\in\Theta_1)}$$

其中下标 π 表示对先验分布求概率。我们还可以把先验机会比与后验机会比进行比较，从而得到另一个重要概念---贝叶斯因子。

定义 4.4 设两个假设 H_0 与 H_1 的先验概率分别为 π_0 与 π_1，后验概率分别为 α_0 与 α_1，则先验机会比与后验机会比之比

$$B^\pi(\mathbf{x})=\frac{\text{后验机会比}}{\text{先验机会比}}=\frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$

称为**贝叶斯因子**。

从这个定义可见，贝叶斯因子既依赖于数据 $\mathbf{x}=(x_1,\cdots,x_n)$，又依赖于先验分布 $\pi(\theta)$，两种机会比（概率比）相除的结果，反映了当利用数据更新先验分布为后验分布后，两个假设的概率之比的变化，所以，贝叶斯因子 $B^\pi(\mathbf{x})$ 是数据（样本）支持原假设 H_0 的程度的一种度量。事实上，不难推出原假设 H_0 的后验概率可表示为

$$\alpha_0=\frac{\pi_0B^\pi(\mathbf{x})}{\pi_0B^\pi(\mathbf{x})+1-\pi_0}$$

而函数 $f(t)=t/(c+t),c>0$ 是单调增函数，所以，当先验给定后，贝叶斯因子越大，原假设 H_0 的后验概率就越大（而且反之亦然）。另外，容易证明 $B^\pi(\mathbf{x})>\pi_1/\pi_0$ 等价于 $\alpha_0>1/2$，因此，当 $B^\pi(\mathbf{x})>\pi_1/\pi_0$ 时，接受原假设 H_0。

4.5.2 简单假设对简单假设

所谓简单假设对简单假设是指原假设H_0和备择假设H_1为如下的情形

$$H_0:\theta=\theta_0,\quad H_1:\theta=\theta_1$$

在这种情形下，不难推出这两假设的后验概率分别为

$$\alpha_0=P(\theta=\theta_0\mid\mathbf{x})=\frac{\pi_0 p(\mathbf{x}\mid\theta_0)}{\pi_0 p(\mathbf{x}\mid\theta_0)+\pi_1 p(\mathbf{x}\mid\theta_1)}$$

$$\alpha_1=P(\theta=\theta_1\mid\mathbf{x})=\frac{\pi_1 p(\mathbf{x}\mid\theta_1)}{\pi_0 p(\mathbf{x}\mid\theta_0)+\pi_1 p(\mathbf{x}\mid\theta_1)}$$

这时后验机会比为

$$\frac{\alpha_0}{\alpha_1}=\frac{\pi_0 p(\mathbf{x}|\theta_0)}{\pi_1 p(\mathbf{x}|\theta_1)}$$

因此，贝叶斯因子

$$B^{\pi}(\mathbf{x})=\alpha_0\pi_1/(\alpha_1\pi_0)=p(\mathbf{x}|\theta_0)/p(\mathbf{x}|\theta_1)$$

即在简单假设对简单假设这种情形下，贝叶斯因子恰好等于样本的似然比而与先验分布无关。

如果要拒绝原假设H_0，必须且只须$\alpha_0/\alpha_1<1$，亦即数据$\mathbf{x}=(x_1,\cdots,x_n)$必须且只须满足

$$\frac{p(\mathbf{x}|\theta_1)}{p(\mathbf{x}|\theta_0)}>\frac{\pi_0}{\pi_1}$$

这正是经典统计中著名的奈曼---皮尔逊引理。从贝叶斯统计的观点看，这个临界值就是两个先验概率比（先验机会比）。反之，如果要接受原假设H_0，则必须且只须$\alpha_0/\alpha_1>1$，亦即$B^{\pi}(\mathbf{x})=p(\mathbf{x}|\theta_0)\big/p(\mathbf{x}|\theta_1)>\pi_1/\pi_0$。显然此时贝叶斯因子越大，不等式越可能成立，从而接受原假设H_0的证据越强。这就更具体的说明贝叶斯因子的大小表示了样本支持原假设H_0的程度。

例 4.5 设样本$\mathbf{x}=(x_1,\cdots,x_n)$来自总体$N(\theta,1)$，其中$\theta$只有二种可能：非 0 即 1。如果已知$n=10$，$\bar{x}=2$，请检验如下两个假设

$$H_0:\theta=0，\quad H_1:\theta=1$$

解： 由题目条件，不难验证均值$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$是参数$\theta$的充分统计量且其分布为$N(\theta,1/n)$，于是在$\theta=0$和$\theta=1$下的似然函数分别为

$$p(\bar{x}|0)=\sqrt{\frac{n}{2\pi}}\exp\{-\frac{n}{2}\bar{x}^2\}，\quad p(\bar{x}|1)=\sqrt{\frac{n}{2\pi}}\exp\{-\frac{n}{2}(\bar{x}-1)^2\}$$

从而贝叶斯因子为

$$B^{\pi}(\mathbf{x})=p(\bar{x}|0)/p(\bar{x}|1)=\exp\{-\frac{n}{2}(2\bar{x}-1)\}$$

利用已知条件$n=10$，$\bar{x}=2$和 R 命令 exp(-10*(2*2-1)/2)，可算得贝叶斯因子$B^{\pi}(\mathbf{x})=3.06\times10^{-7}$，这是个很小的数，因此样本（数据）支持原假设$H_0$的程度微乎其微。进一步，因为要接受$H_0$就必须

$$\frac{\alpha_0}{\alpha_1}=B^{\pi}(x)\frac{\pi_0}{\pi_1}=3.06\times10^{-7}\frac{\pi_0}{\pi_1}>1$$

注意到$\pi_0+\pi_1=1$即知必须$\pi_0>0.9999997$，但这个不等式可以说是不可能实现的，因此，可以明确地拒绝H_0而接受H_1。

4.5.3 复杂假对复杂假设

所谓复杂假设对复杂假设是指原假设H_0和备择假设H_1为如下的情形

$$H_0:\theta\in\Theta_0\text{，}\quad H_1:\theta\in\Theta_1$$

其中Θ_0与Θ_1都是Θ的至少包含两个元素的子集且组成Θ的一个划分。进一步，设参数θ的先验分布为$\pi(\theta)$，而且原假设H_0和备择假设H_1的先验概率分别为$\pi_0=P_\pi(\theta\in\Theta_0)$与$\pi_1=P_\pi(\theta\in\Theta_1)$。为了研究后验机会比和贝叶斯因子，我们把先验分布$\pi(\theta)$进行分解，为此令

$$g_0(\theta)=\pi_0^{-1}\pi(\theta)I_{\Theta_0}(\theta)\text{，}\quad g_1(\theta)=\pi_1^{-1}\pi(\theta)I_{\Theta_1}(\theta)$$

其中，$I_{\Theta_i}(\theta),i=0,1$是示性函数，则有

$$\int_{\Theta_0}g_0(\theta)d\theta=\int_{\Theta_0}\pi_0^{-1}\pi(\theta)I_{\Theta_0}(\theta)d\theta=\pi_0^{-1}\int_{\Theta_0}\pi(\theta)d\theta=1$$

同理$\int_{\Theta_1}g_1(\theta)d\theta=1$，即$g_0$与$g_1$分别是$\Theta_0$与$\Theta_1$上的概率密度函数，而且先验分布$\pi(\theta)$可分解为

$$\pi(\theta)=\pi_0g_0(\theta)+\pi_1g_1(\theta)=\begin{cases}\pi_0g_0(\theta),\theta\in\Theta_0\\ \pi_1g_1(\theta),\theta\in\Theta_1\end{cases}$$

于是后验机会比（概率比）为

$$\frac{\alpha_0}{\alpha_1}=\frac{P(\theta\in\Theta_0|\mathbf{x})}{P(\theta\in\Theta_1|\mathbf{x})}=\frac{\int_{\Theta_0}\pi(\theta\,|\,\mathbf{x})d\theta}{\int_{\Theta_1}\pi(\theta\,|\,\mathbf{x})d\theta}=\frac{\int_{\Theta_0}p(\mathbf{x}|\theta)\pi_0 g_0(\theta)d\theta}{\int_{\Theta_1}p(\mathbf{x}|\theta)\pi_1 g_1(\theta)d\theta}$$

贝叶斯因子为

$$B^{\pi}(\mathbf{x})=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{\int_{\Theta_0}p(\mathbf{x}|\theta)g_0(\theta)d\theta}{\int_{\Theta_1}p(\mathbf{x}|\theta)g_1(\theta)d\theta}=\frac{m_0(\mathbf{x})}{m_1(\mathbf{x})}$$

因此，虽然这时贝叶斯因子已经不是似然比，但可看作 Θ_0 与 Θ_1 上的加权似然之比，这里通过积分（加权平均的极限）部分地消除了先验分布的影响，强调了样本观察值的作用（上面等式的最后部分表面上看只是样本的函数）。

例 4.6 设从正态总体 $N(\theta,1)$ 中随机抽取一个容量为 10 的样本 $\mathbf{x}$，其样本均值 $\bar{x}=1.5$。若取 θ 的共轭先验分布为 $N(0.5,2)$，试检验如下两假设

$$H_0:\theta\le 1,\quad H_1:\theta>1$$

解：根据所学知识，知 θ 的后验分布为 $N(\mu_1,\sigma_1^2)$，其中 μ_1 与 σ_1^2 的计算如下

$$\mu_1=\frac{1.5\times 10+0.5\times 0.5}{10+0.5}=1.4524,\quad \sigma_1^2=\frac{1}{10+0.5}=0.09524=0.3086^2$$

即 θ 的后验分布实为 $N(1.4524,0.3086^2)$。于是可用如下 R 命令算得 H_0 与 H_1 的后验概率

$$\alpha_0=P(\theta\le 1|\mathbf{x})=0.0713,\quad \alpha_0=P(\theta>1|\mathbf{x})=1-0.0713=0.9287$$

因此后验机会比为 $\alpha_0/\alpha_1=0.0713/0.9287=0.0768\ll 1$（双小于号表示大大小于的意思），故应拒绝 H_0，而接受 H_1，即认为正态均值应大于 1。

```
pnorm(1,1.4524,0.3086)
[1] 0.0713275
1-pnorm(1,1.4524,0.3086)
[1] 0.9286725
pnorm(1,1.4524,0.3086)/(1-pnorm(1,1.4524,0.3086))
[1] 0.07680587
```

注：由先验分布 $N(0.5,2)$，可用 R 命令算得 H_0 与 H_1 的先验概率

$$\pi_0 = 0.6382,\ \pi_1 = 1-0.6382 = 0.3618$$

```
pnorm(1,0.5,sqrt(2))
[1] 0.6381632
```

因此，先验机会比 $\pi_0/\pi_1 = 1.7637$，可见先验信息是支持原假设 H_0 的。换句话说，由于样本带来了新信息，使得后验机会比改变了由先验机会比得出的结论。最后，我们考察一下本题中贝叶斯因子的大小，计算贝叶斯因子即两个机会比之比得

$$B^{\pi}(x) = 0.0768/1.7637 = 0.0435 \ll \pi_1/\pi_0 = 0.5670$$

由此可见，贝叶斯因子的结果也明确拒绝 H_0。

例 4.7 设样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 来自正态总体 $N(\theta,\sigma^2)$，其中 σ^2 已知。如果均值参数 θ 的先验是无信息先验 $\pi(\theta)=1$，试讨论如下两个假设的检验问题

$$H_0:\theta\le\theta_0,\quad H_1:\theta>\theta_0$$

解：由 4.4.1 小节知此时后验分布 $\pi(\theta|\mathbf{x})$ 为正态分布 $N(\bar{x},\sigma^2/n)$，因此原假设的后验概率

$$\alpha_0=P(\theta\le\theta_0|\mathbf{x})=P\left[\sqrt{n}(\theta-\bar{x})/\sigma\le\sqrt{n}(\theta_0-\bar{x})/\sigma\,|\,\mathbf{x}\right]=\Phi\left(\sqrt{n}(\theta_0-\bar{x})/\sigma\right)$$

其中，$\Phi(\bullet)$ 是标准正态分布函数。从而后验机会比

$$\frac{\alpha_0}{\alpha_1}=\frac{\Phi\left(\sqrt{n}(\theta_0-\bar{x})/\sigma\right)}{1-\Phi\left(\sqrt{n}(\theta_0-\bar{x})/\sigma\right)}$$

所以，只要样本给定了，虽然这里的无信息先验是非正常（广义）分布，同样可以进行假设检验。例如，给定 $\sigma^2=1$，$\theta_0=1$，$n=10$，$\bar{x}=1.5$（与例 4.6 的一样），则原假设的后验概率

$$\alpha_0=\Phi\left(\sqrt{n}(\theta_0-\bar{x})/\sigma\right)=\Phi\left(\sqrt{10}(1-1.5)\right)=0.0569\ll 0.5$$

故应拒绝 H_0，而接受 H_1。可见，虽然先验分布不同，但这里的结论与例 4.6 的一样。

4.5.4 简单假设对复杂假设

所谓简单假设对复杂假设是指原假设H_0和备择假设H_1为如下的情形

$$H_0:\theta=\theta_0，\quad H_1:\theta\neq\theta_0$$

这是经典统计中常见的一类假设检验问题。显然，当参数空间是离散情形时，对于这两个假设的先验概率不存在任何问题，即原假设H_0的先验概率会大于零。但是，当参数空间是连续情形时，如果直接选择一个连续型分布，则原假设的先验概率为零，从而后验概率也为零，因此，这样的假设检验问题就不是一个有意义的问题了。对于这种情形的先验概率要如何确定呢？其实，从实际的角度看，既然要检验参数θ是否等于θ_0，那就说明θ_0在决策者心中是个重要而且会以一定概率出现的参数值，因而可以给予θ_0一个正概率π_0（可以认为π_0是主观概率），当然具体的概率有多大必须由决策者依据先验信息和专业知识来判断。而对于$\Theta_1=\Theta-\{\theta_0\}$则可以给一个正常密度$g_1(\theta)$，这样，$\theta$的先验分布就可取为

$$\pi(\theta)=\begin{cases}\pi_0, & \theta=\theta_0\\ \pi_1 g_1(\theta), & \theta\in\Theta_1\end{cases}$$

其中$\pi_1=1-\pi_0$，而且易知$\pi(\theta)$是参数空间Θ上的正常密度函数。

现在设样本$\mathbf{x}=(x_1,\cdots,x_n)$的联合分布密度为$p(\mathbf{x}|\theta)$，则参数的后验分布

$$\pi(\theta\mid\mathbf{x})\propto p(\mathbf{x}|\theta)\pi(\theta)=\begin{cases}\pi_0 p(\mathbf{x}|\theta), & \theta=\theta_0\\ \pi_1 g_1(\theta)p(\mathbf{x}|\theta), & \theta\in\Theta_1\end{cases}$$

再令

$$m_1(\mathbf{x})=\int_{\Theta_1}p(\mathbf{x}|\theta)g_1(\theta)d\theta$$

则由上述先验分布容易得到样本的边际分布

$$m(\mathbf{x})=\int_{\Theta}p(\mathbf{x}|\theta)\pi(\theta)d\theta=\pi_0 p(\mathbf{x}|\theta_0)+\pi_1 m_1(\mathbf{x})$$

从而原假设H_0与备择假设H_1的后验概率分别为

$$\alpha_0=P(\theta=\theta_0|\mathbf{x})==\pi_0 p(\mathbf{x}|\theta_0)/m(\mathbf{x})\text{，}\quad \alpha_1=P(\theta\neq\theta_0|\mathbf{x})=\pi_1 m_1(\mathbf{x})/m(\mathbf{x})$$

于是后验机会比和贝叶斯因子分别为

$$\frac{\alpha_0}{\alpha_1}=\frac{\pi_0}{\pi_1}\frac{P(\mathbf{x}|\theta_0)}{m_1(\mathbf{x})}\text{，}\quad B^{\pi}(\mathbf{x})=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{p(\mathbf{x}|\theta_0)}{m_1(\mathbf{x})}$$

特别值得注意的是，从这个贝叶斯因子的表达式，我们看到先验概率已经约去了，因此，贝叶斯因子不受先验概率的影响。这就是说，在计算这个贝叶斯因子的时候，我们并不需要真正知道先验概率等于多少，故在这种情形下，我们常常先计算和考察贝叶斯因子$B^{\pi}(\mathbf{x})$，然后再考虑原假设H_0的后验概率。

4.5.5 案例：哪个疗效更好？

案例 4.4（Berger，1995）本案例是有关药物疗效的统计分析问题，在医药领域有大量的新药或新的医疗法需要进行统计分析以确定是否有疗效。

一个临床试验有两种治疗方法如下：

治疗 1：服药 A，治疗 2：同时服药 A 与药 B

如今进行了 n 次对照试验，设 x_i 为第 i 次对照试验中治疗 2 与治疗 1 的疗效之差，又设诸 x_i 相互独立同分布于正态分布 $N(\theta,1)$，因此，前 n 次的样本均值 $\bar{x}_n \sim N(\theta,1/n)$ 而且是充分统计量。如今要对如下两个假设进行检验：

$$H_0:\theta=0，\quad H_1:\theta\neq 0$$

显然，原假设表示无疗效而备择假设表示有疗效。由于对两种治疗的疗效情况知之甚少，故对原假设 H_0 和备择假设 H_1 取相等先验概率，即 $\pi_0=\pi_1=1/2$。对于 $\Theta_1=\Theta-\{0\}$ 上的先验密度 $g_1(\theta)$，根据先验信息和专业知识，一般看法是参数 θ（疗效之差）接近于 0 比远离 0 更为可能，因此 $g_1(\theta)$ 取为正态分布 $N(0,2)$。这样，我们确定了

$$p(\bar{x}_n|\theta)=\sqrt{\frac{n}{2\pi}}\exp\{-\frac{n}{2}(\bar{x}_n-\theta)^2\},\ g_1(\theta)=\frac{1}{2\sqrt{\pi}}\exp\{-\frac{\theta^2}{4}\}$$

容易算得$\bar{x}_n$对$g_1(\theta)$的边际密度函数

$$m_1(\bar{x}_n)=\int_{-\infty}^{\infty}p(\bar{x}_n|\theta)g_1(\theta)d\theta=\frac{1}{2\pi}\sqrt{\frac{n}{2}}\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}[n(\bar{x}_n-\theta)^2+\frac{\theta^2}{2}]\}d\theta$$ 将被积函数指数部分进行配方得

$$n(\bar{x}_n-\theta)^2+\frac{\theta^2}{2}=(n+0.5)\left(\theta-\frac{n\bar{x}_n}{n+0.5}\right)^2+\frac{n\bar{x}_n^2}{1+2n}$$

将此式带入被积函数并利用密度函数的性质得

$$m_1(\bar{x}_n)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2+n^{-1}}}\exp\{-\frac{\bar{x}_n^2}{2(2+n^{-1})}\}$$

即$\bar{x}_n$对$g_1(\theta)$的边际分布为正态分布$N(0,2+n^{-1})$。 于是，根据前面刚刚得到的结果，贝叶斯因子

$$B^{\pi}(\bar{x}_n)=\frac{p(\bar{x}_n|0)}{m_1(\bar{x}_n)}=\frac{\sqrt{\frac{n}{2\pi}}\exp\{-nx_n^2/2\}}{\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{1+2n}}\exp\{-\frac{x_n^2}{2(2+n^{-1})}\}}$$

$$=\sqrt{1+2n}\exp\{-\frac{(n\bar{x}_n)^2}{1+2n}\}$$

若现在有样本$n=1,\bar{x}_n=1.63$，则马上可以算出

$$B^{\pi}(1.63)=\sqrt{1+2}\exp\{-\frac{(1.63)^2}{1+2}\}=0.7144<1=\frac{\pi_1}{\pi_0}$$

即我们要拒绝原假设。换言之，应该认为两种治疗的疗效是有差别的。

另外，我们也可以求出原假设的后验概率

$$\alpha_0 = \frac{\pi_0 B^{\pi}(\bar{x}_n)}{\pi_0 B^{\pi}(\bar{x}_n) + 1 - \pi_0} = \frac{B^{\pi}(\bar{x}_n)}{B^{\pi}(\bar{x}_n) + 1} = 0.4167 < 0.5$$

至此，这一统计分析似乎完美地完成了。可是，静下心一想，其实有一个更值得研究的问题：哪种治疗的疗效更好一些呢？

我们在刚开始的时候实际上就应该研究如下三个假设的检验问题：

$$H_0:\theta=0,\quad H_2:\theta<0,\quad H_3:\theta>0$$

其中 H_0 还是表示疗效无差别，H_2 表示治疗 2 的疗效不如治疗 1，H_3 表示治疗 2 的疗效优于治疗 1。医生们显然对这个三假设检验问题更感兴趣，因为他们最终要了解的正是哪种治疗的疗效更好。为了做这个检验，先把 $\theta \neq 0$ 时的后验分布求出来，此时后验分布有

$$\pi(\theta|\bar{x}_n) \propto p(\bar{x}_n|\theta)\pi(\theta) = \pi_1 g_1(\theta) p(\bar{x}_n|\theta)$$

$$\propto \exp\left\{-\frac{1}{2}\left[n(\bar{x}_n-\theta)^2+\frac{\theta^2}{2}\right]\right\} = \exp\left\{-\frac{1}{2}\left[(n+0.5)\left(\theta-\frac{n\bar{x}_n}{n+0.5}\right)^2+\frac{n\bar{x}_n^2}{1+2n}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[(n+0.5)\left(\theta-\frac{n\bar{x}_n}{n+0.5}\right)^2\right]\right\}$$

容易看出在给定$\overline{x}_n$下，θ（不含 0）的后验分布为

$$N(n\overline{x}_n/(n+0.5),(n+0.5)^{-1})\text{。}$$

但是，当$\theta=0$时，我们已算得H_0的后验概率为

$$\alpha_0=B^{\pi}(\overline{x}_n)\big/(B^{\pi}(\overline{x}_n)+1)$$

因此，θ（不含 0）的真实后验分布应为

$$(1-\alpha_0)N(n\overline{x}_n/(n+0.5),(n+0.5)^{-1})\text{。}$$

对于前面给定的样本$n=1,\overline{x}_n=1.63$以及$1-\alpha_0=1\big/(1+B^{\pi}(\overline{x}_n))$可以直接用如下 R 命令求得后验概率

$$\alpha_2=P(\theta<0\,|\,\overline{x}_n)=0.0534\text{，}\quad \alpha_3=P(\theta>0\,|\,\overline{x}_n)=0.5299$$

由此，我们可以判断治疗 2 的疗效优于治疗 1。显然，要想得到更加可信的结论，那就必须再多做几次临床试验。

```
pnorm(0,1.63/1.5,1/sqrt(1.5))/(1+sqrt(3)*exp(-1.63^2/3))
[1] 0.05343751
 (1-pnorm(0,1.63/1.5,1/sqrt(1.5)))/(1+sqrt(3)*exp(-1.63^2/3))
[1] 0.5298606
```

　　这类多假设检验问题很有实用价值，从本案例可见，贝叶斯统计可以容易检验多假设问题，而经典统计对此是难以处理的，这是贝叶斯统计的另一优点。

§ 4.6　模型的比较与选择

在统计建模中，模型的比较与选择一直是个重要的问题，因为对一个数据集可以建立起众多的模型，那么，就必须考虑哪一个模型是最适合的。在本节，我们将简要介绍以贝叶斯因子为工具的模型比较与选择问题。

4.6.1 模型比较与选择

一般而言，模型选择就是在给定的样本下，从侯选模型集合中按一定的准则选择最佳的模型, 这里“最佳”的意思是比较而言更好又切实可行，并不是“最优”的意思，因为模型是只有更好没有最好的。设样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 的联合分布密度为 $p(\mathbf{x}\,|\,\theta)$ 而且参数 θ 的先验为 $\pi(\theta)$，那么，我们称它们构成一个贝叶斯模型并记为

$$\mathbf{x}\sim p(\mathbf{x}\,|\,\theta),\ \ \theta\sim\pi(\theta),\ \ \theta\in\Theta$$

我们知道这时后验分布事实上也被确定

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

其中，分母是样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 的边际分布并记为

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

为了下面的应用，这里先讨论一下边际分布的计算问题。在贝叶斯统计中的大多数情形下，这个边际分布的被积函数没有解析形式的原函数，因此必须进行数值积分。这里介绍一种被称为**拉普拉斯逼近法**的近似算法如下。将对数被积函数 $\ln[p(\mathbf{x}|\theta)\pi(\theta)]$ 在后验众数（或参数的最大似然估计）$\hat{\theta}$ 处进行二阶泰勒展开，得

$$\ln[p(\mathbf{x}|\theta)\pi(\theta)] \approx \ln[p(\mathbf{x}|\hat{\theta})\pi(\hat{\theta})] + \frac{1}{2}(\theta-\hat{\theta})' H(\hat{\theta})(\theta-\hat{\theta})$$

其中, $H(\hat{\theta})$ 是海森矩阵(由对数被积函数的二阶导数组成)在 $\hat{\theta}$ 处的值, 于是

$$p(\mathbf{x}|\theta)\pi(\theta) \approx p(\mathbf{x}|\hat{\theta})\pi(\hat{\theta})\exp\{-\frac{1}{2}(\theta-\hat{\theta})'(-H(\hat{\theta}))(\theta-\hat{\theta})\}$$

注意到上式指数部分正是多元正态分布的核, 我们有

$$\begin{aligned} m(\mathbf{x}) &\approx p(\mathbf{x}|\hat{\theta})\pi(\hat{\theta})\int_{\Theta}\exp\{-\frac{1}{2}(\theta-\hat{\theta})'(-H(\hat{\theta}))(\theta-\hat{\theta})\}d\theta \\ &= p(\mathbf{x}|\hat{\theta})\pi(\hat{\theta})(2\pi)^{d/2}|-H(\hat{\theta})^{-1}|^{1/2} \end{aligned}$$

其中 d 是参数向量的维数。这样，边际分布就容易估计出来了。当然，我们也可以求对数边际密度

$$\log m(\mathbf{x}) \approx \frac{d}{2}\log(2\pi) + \log[p(\mathbf{x}|\hat{\theta})\pi(\hat{\theta})] + \frac{1}{2}\log|-H(\hat{\theta})^{-1}|$$

现在设有 K 个贝叶斯模型

$$M_k : \mathbf{x} \sim p_k(\mathbf{x}|\theta_k),\ \theta_k \sim \pi_k(\theta_k),\ \ \theta_k \in \Theta_k,\ k=1,2,\cdots,K$$

我们想比较它们并选择出一个最佳的模型来。如果记模型 M_k 的先验概率为 $\pi_{0k}=P(M_k)$，则有 $0<\pi_{0k}<1,\ \sum_{k=1}^{K}\pi_{0k}=1$。为了求出后验概率 $\pi_{1k}=P(M_k\,|\,\mathbf{x})$，利用离散型的贝叶斯公式得

$$\pi_{1k}=P(M_k\,|\,\mathbf{x})=\frac{\pi_{0k}\,p(\mathbf{x}\,|\,M_k)}{\sum_{j=1}^{K}\pi_{0j}\,p(\mathbf{x}\,|\,M_j)}$$

利用边际分布密度公式得

$$p(\mathbf{x}\,|\,M_k)=\int_{\Theta_k} p(\mathbf{x},\theta_k\,|\,M_k)\mathrm{d}\theta_k=\int_{\Theta_k} p(\mathbf{x}\,|\,\theta_k,M_k)\,p(\theta_k\,|\,M_k)\mathrm{d}\theta_k$$

在贝叶斯模型 M_k 已知的条件下，有

$$p(\mathbf{x}\,|\,\theta_k,M_k)=p_k(\mathbf{x}\,|\,\theta_k),\ p(\theta_k\,|\,M_k)=\pi_k(\theta_k),\ \ k=1,2,\cdots,K$$

故模型 M_k 的后验概率为

$$\pi_{1k}=P(M_k\,|\,\mathbf{x})=\frac{\pi_{0k}\int_{\Theta_k} p_k(\mathbf{x}|\theta_k)\pi_k(\theta_k)\mathrm{d}\theta_k}{\sum_{j=1}^{K}\pi_{0j}\int_{\Theta_j} p_j(\mathbf{x}|\theta_j)\pi_j(\theta_j)\mathrm{d}\theta_j}$$

在上面这个式子中，分母对于每一个模型的后验概率而言都是一样的，因此，比较模型的后验概率的大小就只要比较上式的分子就可以了。

另一方面，模型 M_i 与 M_j 的贝叶斯因子为

$$B_{ij}^{\pi}(\mathbf{x}) = \frac{\pi_{1i}\pi_{0j}}{\pi_{1j}\pi_{0i}} = \frac{\pi_{0j}\pi_{0i}\int_{\Theta_i} p_i(\mathbf{x}|\theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i}{\pi_{0i}\pi_{0j}\int_{\Theta_j} p_j(\mathbf{x}|\theta_j)\pi_j(\theta_j)\mathrm{d}\theta_j} = \frac{\int_{\Theta_i} p_i(\mathbf{x}|\theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i}{\int_{\Theta_j} p_j(\mathbf{x}|\theta_j)\pi_j(\theta_j)\mathrm{d}\theta_j} = \frac{m_i(\mathbf{x})}{m_j(\mathbf{x})}$$

利用贝叶斯因子的方便之处是可以不知道模型的先验概率，因为通过约分把它们都约去了。

另外，通过贝叶斯因子，模型 M_k 的后验概率又可以表示为

$$\pi_{1k} = \frac{\pi_{0k}\int_{\Theta_k} p_k(\mathbf{x}|\theta_k)\pi_k(\theta_k)\mathrm{d}\theta_k}{\sum_{j=1}^{K}\pi_{0j}\int_{\Theta_j} p_j(\mathbf{x}|\theta_j)\pi_j(\theta_j)\mathrm{d}\theta_j}$$

$$= \left(\sum_{j=1}^{K}\frac{\pi_{0j}\int_{\Theta_j} p_j(\mathbf{x}|\theta_j)\pi_j(\theta_j)\mathrm{d}\theta_j}{\pi_{0k}\int_{\Theta_k} p_k(\mathbf{x}|\theta_k)\pi_k(\theta_k)\mathrm{d}\theta_k}\right)^{-1} = \left(\sum_{j=1}^{K}\frac{\pi_{0j}}{\pi_{0k}}B_{jk}^{\pi}(\mathbf{x})\right)^{-1}$$

用后验概率来比较和选择模型是很自然和恰当的做法，哪个模型的后验概率最大，哪个模型就是最佳模型，但是计算后验概率必须已知先验概率，而先验概率并不容易确定。因此，在实际应用中常常去估算贝叶斯因子，然后通过贝叶斯因子的结果来比较和选择模型。

贝叶斯因子 $B_{ij}^{\pi}(\mathbf{x})$ 只要大于 1，它就倾向于支持模型 M_i，而且它越大，支持模型 M_i 的证据就越强，但是对于贝叶斯因子的估算结果，并不存在完全客观的分类使其成为接受模型 M_i 的标准（贝叶斯因子 $B_{ij}^{\pi}(\mathbf{x})$ 小于 1 时，拒绝模型 M_i）。杰弗里斯（1961）给出了一个贝叶斯因子的结果分类（表 4.2，其中 $B_{ij}^{\pi}(\mathbf{x})$ 简记为 B_{ij}，下同）。另外，卡斯和拉弗特里(Kass and Raftery, 1995) 也给出了一个贝叶斯因子的结果分类（表 4.3）。容易看出表 4.2 和表 4.3 没有本质的区别，所以我们可以参照这两个表来对模型进行比较和选择。

表 4.2　杰弗里斯对贝叶斯因子的值的分类

贝叶斯因子 B_{ij} 值的范围	关于模型 M_i 的证据的强度
$B_{ij} < 1$	拒绝模型 M_i
$1 \le B_{ij} < 3$	支持模型 M_i 的证据弱
$3 \le B_{ij} < 10$	支持模型 M_i 的证据中
$10 \le B_{ij} < 30$	支持模型 M_i 的证据强
$30 \le B_{ij} < 100$	支持模型 M_i 的证据很强
$B_{ij} \ge 100$	肯定支持模型 M_i

表 4.3　卡斯和拉弗特里对贝叶斯因子的值的分类

贝叶斯因子 B_{ij} 值的范围	关于模型 M_i 的证据的强度
$1 \le B_{ij} < 3$	支持模型 M_i 的证据弱
$3 \le B_{ij} < 20$	支持模型 M_i 的证据中
$20 \le B_{ij} < 150$	支持模型 M_i 的证据强
$B_{ij} \ge 150$	支持模型 M_i 的证据很强

4.6.2 案例：足球队进球数量的分布是什么？

案例 4.5 本案例是为一支足球队建立进球数量的模型。假设在美国职业足球大联盟中有支球队是我们感兴趣的，而且我们观看了它的 n 场比赛，得到它的进球个数为 $x_1,x_2,\cdots,x_n$。由于足球进球是相当难得的事，所以可以设进球个数服从均值参数为 λ 的泊松分布 $Poisson(\lambda)$。为了对这支球队有更深入的了解，我们当然想去估计均值参数 λ，但是我们没有关于 λ 的先验信息，所以只好挑选四个主观认定的先验分布：（1）$\lambda \sim Gamma(4.57,1.43)$，这是 λ 的共轭先验，超参数之所以取为(4.57,1.43)，是因为我们认为球队进球的平均数约为 3 个，而且上下四分之一分位数分别是 4.04 和 2.10；（2）$\log\lambda \sim N(1,0.5^2)$，这时的上下四分之一分位数分别是 1.34 和 0.66；（3）$\log\lambda \sim N(2,0.5^2)$，这时的上下四分之一分位数分别是 10.35 和 5.27。之所以取这个先验，是因为我们相信大联盟球队的进球率高；（4）$\log\lambda \sim N(1,2^2)$，这时的上下四分之一分位数分别是 28.50 和 1.92。之所以取这个先验，是表明我们对足球赛的得分方式不甚了解。这样，就形成了四个备选的模型。同时说明一下，如果一个取正值的随机变量的对数服从正态分布，则其本身服从对数正态分布，所以（2）、（3）、（4）实际上是说参数 λ 分别服从不同的对数正态分布。

本案例的样本是这支球队 35 场比赛进球个数的数据，以文件名 football 存放在 R 包 BayesianStat 中。显然，样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 的分布或似然函数为

$$p(\mathbf{x}\mid\lambda)=\lambda^t e^{-35\lambda}\Big/\prod_{i=1}^{35}x_i!\text{，其中 } t=\sum_{i=1}^{35}x_i=57$$

现在用 R 程序计算四个备选贝叶斯模型的对数边际密度 $m_i(\mathbf{x}),i=1,2,3,4$，所用命令如下，其中，命令 list 是将数据 data=goals 和参数 par=c(4.57,1.43)组合成列表 datapar；命令 iLaplace 就是用拉普拉斯逼近法计算模型的对数边际密度，它的三个参变量分别是 $\log[p(\mathbf{x}\mid\theta)\pi(\theta)]$，众数的初始值 0.5 和对应的列表 datapar，它输出三个值分别为后验众数 mode、后验方差 var 以及对数边际密度 int（logmarg)。其余的命令是不难理解的。

```
library(BayesianStat)
data(football)
attach(football)
datapar=list(data=goals,par=c(4.57,1.43))
fit1=iLaplace(Logpoissgamma,.5,datapar)
datapar=list(data=goals,par=c(1,.5))
fit2=iLaplace(Logpoissnormal,.5,datapar)
datapar=list(data=goals,par=c(2,.5))
fit3=iLaplace(Logpoissnormal,.5,datapar)
datapar=list(data=goals,par=c(1,2))
fit4=iLaplace(Logpoissnormal,.5,datapar)
postmode=c(fit1$mode,fit2$mode,fit3$mode,fit4$mode)
postsd=sqrt(c(fit1$var,fit2$var,fit3$var,fit4$var))
logmarg=c(fit1$int,fit2$int,fit3$int,fit4$int)
```

```
cbind(postmode,postsd,logmarg)
      postmode    postsd   logmarg
[1,] 0.5248047 0.1274414 -1.502977
[2,] 0.5207825 0.1260712 -1.255171
[3,] 0.5825195 0.1224723 -5.076316
[4,] 0.4899414 0.1320165 -2.137216
```

计算出对数边际密度后，我们就能计算贝叶斯因子了。利用如下公式

$$B_{ij}=\frac{m_i(\mathbf{x})}{m_j(\mathbf{x})}=\exp[\ln m_i(\mathbf{x})-\ln m_j(\mathbf{x})]$$

可算得这三个贝叶斯因子 $B_{21}=1.28$， $B_{23}=45.7$， $B_{24}=2.42$。从第一、三这两个贝叶斯因子的值可以看出虽然支持模型二的证据不是很强，但还是偏向它的，因为贝叶斯因子都大于 1，而第二个贝叶斯因子的值则强烈支持模型二，总而言之，这四个备选模型中，模型二是最佳的，于是我们可以选择模型二作为我们需要的模型（还可参考习题 4.11）。模型确定后，进一步分析就不太难了。

§ 4.7 统计预测

4.7.1 预测原理

设总体（随机变量）$X \sim p(x|\theta)$，其中参数θ未知但具有先验分布（密度）$\pi(\theta)$。这时如何对总体X的未来值进行预测呢？由于参数θ未知，不能直接利用总体X的分布。然而，虽然θ未知，我们仍可以利用先验分布$\pi(\theta)$得到X的边际分布

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

于是，可以用$m(x)$的期望值或中位数或众数作为X的预测值（点估计）。正因为如此，这个边际分布还有一个更富于含义的名称“先验预测分布”。这里“先验”是指对X预测时没有用任何样本信息（数据）。

如果我们还有观察数据$\mathbf{x} = (x_1, \cdots, x_n)$，那么可以由贝叶斯公式

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

得到后验分布$\pi(\theta|\mathbf{x})$，这时如何对具有密度函数$f(z|\theta)$（其中未知参数θ有一样的先验分布$\pi(\theta)$）的随机变量Z进行预测呢？我们先定义Z的后验预测分布如下

$$g(z|\mathbf{x})=\int_{\Theta} f(z|\theta)\pi(\theta|\mathbf{x})d\theta=\frac{\int_{\Theta} f(z|\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_{\Theta} p(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

显然，当把这个后验预测分布计算出来后，就可以用它的期望值或中位数或众数作为Z的预测值（点估计）了，不仅如此，我们还可以确定可信度为$1-\alpha$的预测区间$[a,b]$，它满足

$$P(a\le Z\le b|\mathbf{x})=\int_a^b g(z|\mathbf{x})dz=1-\alpha$$

注：如果随机变量Z就是X本身，则后验预测分布为

$$g(x|\mathbf{x})=\int_{\Theta} p(x|\theta)\pi(\theta|\mathbf{x})d\theta=\frac{\int_{\Theta} p(x|\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_{\Theta} p(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

4.7.2 统计预测的例

例 4.8 设总体 X 服从参数为 θ 的指数分布，即其分布密度是

$$p(x|\theta)=\theta\exp(-\theta x), x>0, \theta>0$$

而参数 θ 的先验为无信息先验 $\pi(\theta)\propto\theta^{-1}$。现在我们观察到一组样本值 $\mathbf{x}=(x_1,\cdots,x_n)$，（1）求新观测量 x 的后验预测分布；（2）求新观测量 x 的后验预测分布的期望（均值）。

解：（1）显然似然函数（样本联合分布）为

$$p(\mathbf{x}|\theta)=\theta^n\exp(-\theta n\overline{x})$$

其中 $\overline{x}$ 是样本均值。于是，后验分布

$$\pi(\theta\,|\,\mathbf{x})\propto p(\mathbf{x}|\theta)\pi(\theta)=\theta^{n-1}\exp(-\theta n\overline{x})$$

上式右边实际上是伽玛分布 $Gamma(n, n\overline{x})$ 的核，也就是说

$$\pi(\theta\,|\,\mathbf{x})=\frac{(n\overline{x})^n}{\Gamma(n)}\theta^{n-1}\exp(-\theta n\overline{x})$$

故新观测量 x 的后验预测分布为

$$g(x|\mathbf{x}) = \int_{\Theta} p(x|\theta)\pi(\theta|\mathbf{x})d\theta$$

$$= \frac{(n\overline{x})^n}{\Gamma(n)}\int_{\Theta} \theta \exp(-\theta x)\theta^{n-1}\exp(-\theta n\overline{x})d\theta$$

$$= \frac{(n\overline{x})^n \Gamma(n+1)}{\Gamma(n)(x+n\overline{x})^{n+1}} \frac{(x+n\overline{x})^{n+1}}{\Gamma(n+1)}\int_{\Theta} \theta^{n+1-1}\exp[-\theta(x+n\overline{x})]d\theta$$

$$= \frac{n(n\overline{x})^n}{(x+n\overline{x})^{n+1}} \quad (\because \Gamma(n+1) = n\Gamma(n))$$

（2）新观测量 x 的后验预测分布的期望

$$E(X\mid\mathbf{x}) = \int_0^{\infty} xg(x|\mathbf{x})dx = \int_0^{\infty} \frac{n(n\overline{x})^n x}{(x+n\overline{x})^{n+1}}dx$$

$$= \int_0^{\infty} \frac{(n\overline{x})^n}{(x+n\overline{x})^n}dx = \frac{n\overline{x}}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n} x_i$$

其中第三个等号是利用分部积分得到。这样，我们也就预测出了 x 的观测值。

例 4.9 设某试验的一次成功概率为θ，其先验分布取为共轭先验分布$Beta(\alpha,\beta)$。（1）现在在n次独立的贝努里试验中成功了x次，求未来的k次相互独立的贝努里试验中成功次数Z的后验预测概率分布。这里的贝努里试验中的成功可以是零件的合格、射击的命中等等。（2）如果在过去 10 次试验中成功了 3 次，而θ的先验为$Beta(1,1)$，对未来 5 次试验中成功的次数Z作出预测。

解：（1）由以前所学，我们知道样本x的似然函数（概率函数）为

$$p(x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

而参数θ后验密度为

$$\pi(\theta|x)=\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

未来的k次相互独立的贝努里试验中成功次数Z的似然函数为

$$p(z|\theta)=\binom{k}{z}\theta^z(1-\theta)^{k-z}$$

于是Z的后验预测概率分布为

$$
\begin{aligned}
g(z|x) &= \int_\Theta p(z|\theta)\pi(\theta|x)d\theta = \int_0^1 \binom{k}{z}\theta^z(1-\theta)^{k-z}\pi(\theta|x)d\theta \\
&= \binom{k}{z}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\int_0^1 \theta^{z+x+\alpha-1}(1-\theta)^{k-z+n-x+\beta-1}d\theta \\
&= \binom{k}{z}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\frac{\Gamma(z+x+\alpha)\Gamma(k-z+n-x+\beta)}{\Gamma(n+k+\alpha+\beta)}
\end{aligned}
$$

（2）现在的条件是过去 10 次试验中成功了 3 次，而θ的先验为$Beta(1,1)$（即均匀分布$U(0,1)$），要对未来 5 次试验中成功的次数 Z 作出预测。所以，此时 Z 的的后验预测概率分布为

$$
g(z|x=3) = \binom{5}{z}\frac{\Gamma(12)\Gamma(4+z)\Gamma(13-z)}{\Gamma(4)\Gamma(8)\Gamma(17)}
$$

这里 z 可取 0，1，…，5 各个数。利用如下 R 命令容易算出后验预测概率分布列，其中 gamma(x)就是伽玛函数，choose(k, z)是组合函数。

```
g=c()
gg=gamma(4)*gamma(8)*gamma(17)
for(i in 0:5){g[i+1]=choose(5, i)*gamma(12)*gamma(4+i)*gamma(13-i)/gg}
g
[1]0.18131868    0.30219780    0.27472527    0.16483516    0.06410256    0.01282051
```

从后验预测概率分布可见，z=1 是众数，成功次数 Z 在 0 到 3 之间的概率 $P(0 \le Z \le 3 | x) \approx 0.92$ ，这表明[0, 3]是 Z 的 92%预测区间，故在未来 5 次试验中成功的次数基本上不会超过 3 次，而且最可能是 1 次。

本章介绍了贝叶斯统计推断的基础知识，包括点估计、区间估计、假设检验、模型选择、统计预测。从中我们看到贝叶斯统计的许多优点和应用的广泛性。其实，正是因为最近几十年来，贝叶斯统计成功地解决了大量来自不同领域的复杂统计问题，才使得它成为统计专业的必修课，同时也成为其他专业人员喜欢使用的统计工具。

Homework：

PP78-79, 1~10