第2章 共轭先验分布与充分统计量

在前一章我们引入了贝叶斯公式，知道它把总体信息、先验信息以及样本信息综合利用起来，目的是在有了新得到的样本的条件下，对先验分布进行更新调整，以期得到更加符合现实的后验分布。这里的一个关键问题是如何把后验分布算出来，在实际应用中这不是容易的事，并因此成为贝叶斯统计的核心问题。本章将引入共轭先验和充分统计量等概念，对该问题做一个较为详细的讨论。

§ 2.1　共轭先验分布

2.1.1　后验分布的核

在第 1 章里我们了解到在给定先验分布 $\pi(\theta)$ 和样本分布 $p(\mathbf{x}|\theta)$ 后，可用贝叶斯公式计算 θ 的后验分布

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

容易看出

$$\int_{\Theta} \pi(\theta \mid \mathbf{x})d\theta = \frac{\int_{\Theta} p(\mathbf{x}|\theta)\pi(\theta)d\theta}{m(\mathbf{x})} = \frac{m(\mathbf{x})}{m(\mathbf{x})} = 1$$

这就是说，边际分布 $m(\mathbf{x})$ 不依赖于 θ，在计算 θ 的后验分布公式中其实是一个常数，只起到保证后验分布是正常概率分布的作用。因此，可以把 $m(\mathbf{x})$ 省略不写，而把贝叶斯公式改写为如下形式

$$\pi(\theta \mid \mathbf{x}) \propto p(\mathbf{x}|\theta)\pi(\theta)$$

其中符号“$\propto$”表示左边正比于右边或者说两边仅差一个不依赖于θ的常数因子。虽然$p(\mathbf{x}|\theta)\pi(\theta)$不是正常的密度函数，但它是后验分布$\pi(\theta|\mathbf{x})$的最主要部分，被称为后验分布的核，通过它可以容易判断出后验分布是什么分布。如果$p(\mathbf{x}|\theta)\pi(\theta)$中还有常数因子，则可以进一步略去不写。这样，在计算后验分布时就可以不要计算$m(\mathbf{x})$和别的常数因子，从而大大化简了计算。例如，对于第 1 章的例 1.5 中后验分布，我们有

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

显而易见这正是参数为 x+1 和 n-x+1 的贝塔分布的核，这样也就得出后验分布为贝塔分布$Beta(x+1,n-x+1)$的结论。

2.1.2 共轭先验分布

在第 1 章的例 1.5 中后一部分，二项分布的参数θ的先验分布是均匀分布$U(0,1)$，也就是贝塔分布 $Beta(1,1)$，而我们得到其后验分布是另一个贝塔分布 $Beta(x+1,n-x+1)$。换言之，先验分布与后验分布同属于贝塔分布族，只是参数不同而已。这种先验分布被称为参数θ的共轭先验分布，由雷法和施莱弗(H. Raiffa and R. Schlaifer, 1961)首先提出，它能够起到化简后验分布计算以及合理解释先验和样本特征量与后验特征量之间的关系。现在我们给出一个正式的定义。

设样本$\mathbf{x}=(x_1,\cdots,x_n)$来自于总体$p(x|\theta)$，其中$\theta$ 是参数（向量），$\pi(\theta)$是θ的先验密度函数，$p(\mathbf{x}|\theta)$是样本密度函数（即似然函数）。如果后验密度函数$\pi(\theta|\mathbf{x})$与先验密度函数$\pi(\theta)$属于同一个分布族（即有相同的函数形式），则称$\pi(\theta)$共轭于似然函数$p(\mathbf{x}|\theta)$，同时也称$\pi(\theta)$是参数θ的共轭先验分布。

例 2.1 证明如果正态总体 $N(\theta,\sigma^2)$（方差已知）的均值 θ 用正态分布为先验则其是 θ 的共轭先验。

证明：设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自正态分布 $N(\theta,\sigma^2)$ 的一个样本。此样本的联合密度为

$$p(\mathbf{x}|\theta)=(\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\},$$

现取另一个正态分布 $N(\mu,\tau^2)$ 作为 θ 的先验分布，即

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

其中参数 μ 与 τ^2 为已知。这样，样本 $\mathbf{x}$ 与参数 θ 的联合密度函数

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

其中 $k_1=(2\pi)^{-(n+1)/2}\tau^{-1}\sigma^{-n}$， $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 。再引入记号

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

则样本 $\mathbf{x}$ 与参数 θ 的联合密度函数

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

其中 $k_2=k_1\exp\{-\frac{1}{2}(c-B^2/A)\}$。由此易得样本 $\mathbf{x}$ 的边际分布

$$m(\mathbf{x})=\int_{-\infty}^{\infty}h(\mathbf{x},\theta)d\theta=k_2(\frac{2\pi}{A})^{\frac{1}{2}}$$

于是 θ 的后验分布

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

这显然也是一个正态分布 $N(\mu_1,\tau_1^2)$，其均值 μ_1 与方差 τ_1^2 分别为

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

因此证明了正态分布 $N(\mu,\tau^2)$ 是均值参数 θ（方差已知）的共轭先验分布。

注：

1. 在证明过程中，如果我们利用正比符号“ $\propto$ ”进行推导，则可以大大化简计算过程如下

$$\pi(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)\pi(\theta) \propto \exp\{-\frac{1}{2}[\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\sigma^2}+\frac{(\theta-\mu)^2}{\tau^2}]\}$$

$$\propto \exp\{-\frac{1}{2}[A\theta^2-2B\theta]\}$$

$$\propto \exp\{-\frac{A}{2}(\theta-B/A)^2\}$$

2. 后验分布的均值 μ_1 可以改写为

$$\mu_1=\frac{\sigma_0^{-2}}{\sigma_0^{-2}+\tau^{-2}}\bar{x}+\frac{\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}}\mu=\gamma\bar{x}+(1-\gamma)\mu$$

这表明后验均值是样本均值与先验均值之间的一种均衡。

3. 随机变量的方差的倒数可以看成某种取值的精度。我们看到后验方差 τ_1^2 的倒数为

$$\frac{1}{\tau_1^2}=\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}=\frac{n}{\sigma^2}+\frac{1}{\tau^2}$$

这就是说，后验分布的精度是样本均值分布的精度与先验分布的精度之和，另外，增加样本量 n 或减少先验分布的方差都有利于提高后验分布的精度（ σ^2 已知）。

例 2.2 寻找正态总体 $N(\theta,\sigma^2)$（均值 θ 已知）的方差 σ^2 的共轭先验。

设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自正态分布 $N(\theta,\sigma^2)$ 的样本。此样本的联合密度为

$$
\begin{aligned}
p(\mathbf{x}|\sigma^2) &= (\sqrt{2\pi}\sigma)^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\} \\
&\propto (\frac{1}{\sigma^2})^{n/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\}
\end{aligned}
$$

在这里上述密度当然可以看成似然函数，其中 σ^2 的因式将决定 σ^2 的共轭先验分布的形式。换句话说，如果有分布的核也具有这个似然函数的形式，那么这个分布与似然函数的乘积也就有同样的形式，因而这个分布是共轭先验。

现设 X 服从伽玛分布 $Gamma(\alpha,\lambda)$，其中 $\alpha>0$ 是形状参数，$\lambda>0$ 为尺度参数，其密度函数为

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

再令$Y = X^{-1}$，由概率论知识，不难求得Y的密度函数

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

这个分布称为逆（或倒）伽玛分布，并记为$IGamma(\alpha,\lambda)$，它的形式与上述的似然函数形式相似。

假如取这个逆伽玛分布为σ^2的先验分布，则其密度函数为

$$\pi(\sigma^2) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(\frac{1}{\sigma^2})^{\alpha+1}\exp(-\frac{\lambda}{\sigma^2})$$

于是σ^2的后验分布为

$$\pi(\sigma^2|\mathbf{x}) \propto p(\mathbf{x}|\sigma^2)\pi(\sigma^2) \propto (\frac{1}{\sigma^2})^{\alpha+\frac{n}{2}+1}\exp\{-\frac{1}{\sigma^2}[\lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2]\}$$

显然，这仍是逆伽玛分布并具有形式$IGamma(\alpha+\frac{n}{2},\lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2)$。因此逆（或倒）伽玛分布

$IGamma(\alpha,\lambda)$是正态方差σ^2（均值θ已知）的共轭先验分布。

例 2.3 证明伽玛分布 $Gamma(\alpha,\beta)$ 是泊松分布 $Poisson(\lambda)$ 的均值 λ 的共轭先验分布。

证明：泊松分布 $Poisson(\lambda)$ 的分布列是

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

设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自泊松分布 $Poisson(\lambda)$ 样本，此样本的联合概率为

$$p(\mathbf{x}\mid\lambda)=\frac{\lambda^{\sum_{i=1}^n x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda},\quad -\infty<x_1,\cdots,x_n<+\infty$$

现取伽玛分布 $Gamma(\alpha,\beta)$ 作为泊松分布均值 λ 的先验分布，即

$$\pi(\lambda)=\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda},\quad -\infty<\lambda<+\infty$$

于是均值 λ 的后验分布为

$$\pi(\lambda \mid \mathbf{x}) \propto p(\mathbf{x} \mid \lambda)\pi(\lambda) \propto \lambda^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-(\beta+n)\lambda}$$

这实质上就是伽玛分布 $Gamma(\sum_{i=1}^{n} x_i + \alpha, \beta + n)$。所以，伽玛分布是泊松分布的均值 λ 的共轭先验分布。

根据伽玛分布的性质，λ 的后验均值为

$$E(\lambda \mid \mathbf{x}) = \frac{\sum_{i=1}^{n} x_i + \alpha}{\beta + n} = \frac{n}{\beta + n} \times \bar{x} + \frac{\beta}{\beta + n} \times \frac{\alpha}{\beta}$$

由此可见，与前面例 2.1 相类似，后验均值仍是样本均值与先验均值之间的一种均衡。

例 2.4 设一种布料上的断线点个数服从泊松分布 $Poisson(\lambda)$，现在通过检查得到样本 $\mathbf{x}=(3,4,3,2,1)$，取 $Gamma(\alpha,\beta)$ 分布为均值参数 λ 的共轭先验分布。用 R 命令分别做出 $(\alpha,\beta)=(1,0)$，$(\alpha,\beta)=(1,2)$ 时的后验密度曲线图。

解：R 命令如下

```
library(BayesianStat); x<-c(3,4,3,2,1); Poisgamma(x,1,0); Poisgamma(x,1,2)
```

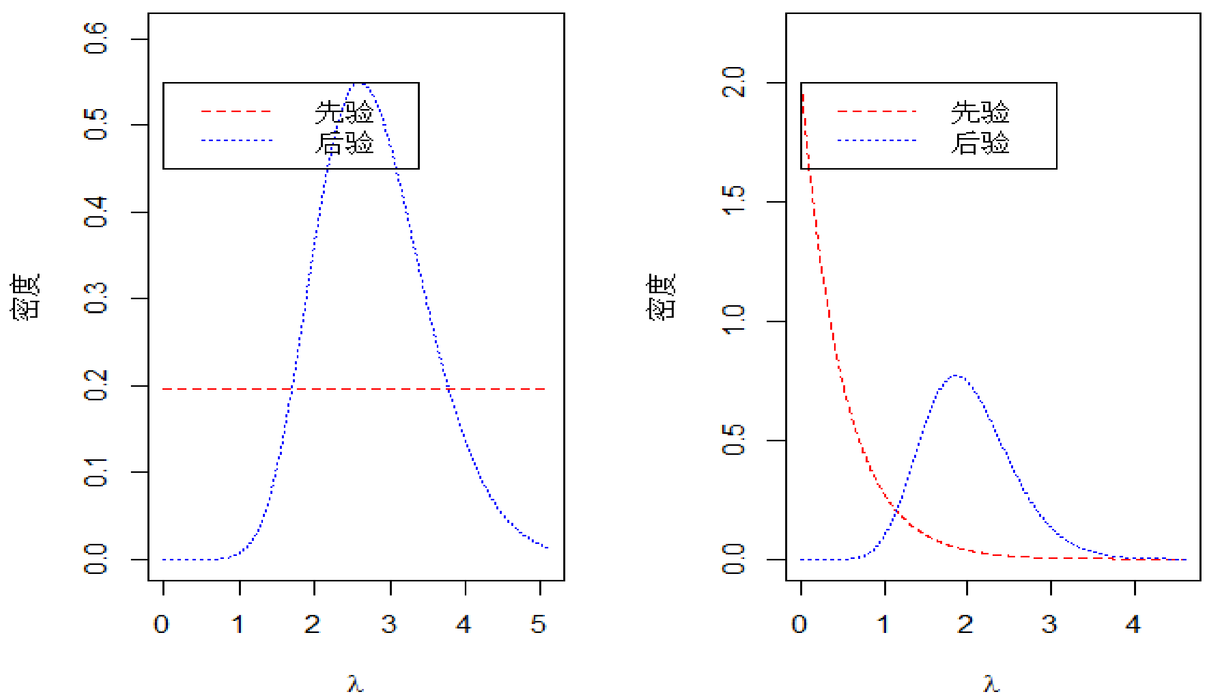

图 2. 1 泊松分布均值的（先）后验密度曲线

§ 2.2　多参数分布

2.2.1　联合先（后）验密度函数

在前面章节中寻求正态总体 $N(\mu,\sigma^2)$ 的参数的后验分布时，我们是假设一个参数已知而去求另一个参数的后验分布。可是对实际问题来说，正态总体 $N(\mu,\sigma^2)$ 的二个参数 (μ,σ^2) 都是未知的，需要同时进行估计等统计推断，从而需要寻找它们的联合先（后）验密度函数。

多个参数的联合后验密度函数的寻求与单参数情形是相似的，即先根据先验信息提炼出参数的联合先验分布，然后按贝叶斯公式算出联合后验分布。我们以二个参数的情形为例，设总体密度 $p(x|\theta)$ 只含二个参数 $\theta=(\theta_1,\theta_2)$ 并给出了联合先验密度 $\pi(\theta_1,\theta_2)$。若有来自该总体的样本 $\mathbf{x}=(x_1,\ldots,x_n)$，而样本分布密度为 $p(\mathbf{x}|\theta)=p(\mathbf{x}|\theta_1,\theta_2)$，则参数向量 (θ_1,θ_2) 的联合后验密度

$$\pi(\theta_1,\theta_2|\mathbf{x})\propto p(\mathbf{x}|\theta_1,\theta_2)\pi(\theta_1,\theta_2)$$

剩下的问题就是怎样把它具体算出来了。

2.2.2 多参数共轭先验的例

例 2.5 寻求正态总体 $\mathrm{N}(\mu,\sigma^2)$ 的两参数均值与方差 (μ,σ^2) 的联合共轭先验分布。

解：设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自正态分布 $\mathrm{N}(\mu,\sigma^2)$ 的样本，则该样本的联合密度函数（即似然函数）为

$$
\begin{aligned}
p(\mathbf{x}|\mu,\sigma^2) &\propto \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\} \\
&= \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-\mu)]^2\} \\
&= \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{x}-\mu)^2]\}
\end{aligned}
$$

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

在前面章节中我们了解到，当σ^2已知时，正态分布是μ的共轭先验，而当μ已知时，逆伽玛分布是σ^2的共轭先验，再考虑到μ与σ^2在样本联合密度函数$\mathrm{p}(\mathbf{x}\,|\,\mu,\sigma^2)$中的位置以及$\mathrm{p}(\mathbf{x}\,|\,\mu,\sigma^2)$本身的形式，我们尝试先把$\sigma^2$当作已知，取$\pi(\mu\,|\,\sigma^2)$为正态分布，而$\pi(\sigma^2)$为逆伽玛分布，然后利用乘法公式

$$\pi(\mu,\sigma^2)=\pi(\mu\,|\,\sigma^2)\pi(\sigma^2)$$

来求得先验分布，其中

$$\mu\big|\sigma^2 \sim N(\mu_0,\sigma^2/\kappa_0),\ \sigma^2 \sim IGamma(\upsilon_0/2,\upsilon_0\sigma_0^2/2)$$

其中$\mu_0,\kappa_0,\upsilon_0,\sigma_0^2$是超参数并且$\mu_0$是先验均值而$\kappa_0$相当于先验样本量，而逆伽玛分布中参数形式是为了下面处理方便。

由此就可以写出(μ,σ^2)的联合先验密度函数

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

这种形式的分布被称为正态--逆伽玛分布并记为$N-IGamma(\mu_0,\kappa_0,\upsilon_0,\sigma_0^2)$。

将(μ,σ^2)的联合先验密度乘以样本联合密度，就得后验密度的核

$$\pi(\mu,\sigma^2\,|\mathbf{x})\propto p(\mathbf{x}|\mu,\sigma^2)\pi(\mu,\sigma^2)$$
$$\propto(\sigma^2)^{-[(\upsilon_0+n+1)/2+1]}\exp\{-\frac{1}{2\sigma^2}[\upsilon_0\sigma_0^2+\kappa_0(\mu-\mu_0)^2+(n-1)s^2+n(\bar{x}-\mu)^2]\}$$

引入记号

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

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

并注意到

$$\kappa_0(\mu-\mu_0)^2+n(\overline{x}-\mu)^2=(\kappa_0+n)\mu^2-2\mu(\kappa_0\mu_0+n\overline{x})+\kappa_0\mu_0^2+n\overline{x}$$

$$=(\kappa_0+n)(\mu-\frac{\kappa_0\mu_0+n\overline{x}}{\kappa_0+n})^2-\frac{(\kappa_0\mu_0+n\overline{x})^2}{\kappa_0+n}+\kappa_0\mu_0^2+n\overline{x}^2$$

$$=(\kappa_0+n)(\mu-\frac{\kappa_0\mu_0+n\overline{x}}{\kappa_0+n})^2+\frac{n\kappa_0(\mu_0-\overline{x})^2}{\kappa_0+n}$$

$$=\kappa_n(\mu-\mu_n)^2+\frac{n\kappa_0(\mu_0-\overline{x})^2}{\kappa_0+n}$$

则(μ,σ^2)的联合后验密度的核可化为

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

这显然是正态--逆伽玛分布$N-IGamma(\mu_n,\kappa_n,\upsilon_n,\sigma_n^2)$的核，因此，正态--逆伽玛分布

是正态均值与方差(μ,σ^2)的（联合）共轭先验分布。

案例 2.1 为了研究摇蚊的翼长，研究人员收集了一个种类的 9 只摇蚊翼长的数据(1.64 ，1.70 ，1.72 ，1.74 , 1.82 , 1.82 , 1.82 , 1.90 , 2.08)（按上升方式排列，单位：毫米 mm）。我们可以假设摇蚊的翼长服从正态分布 $N(\mu,\sigma^2)$，但其中两个参数都未知。根据以往对别的摇蚊种类的研究，摇蚊翼长均值约为 1.9，方差约为 0.01，即可取超参数 $\mu_0 = 1.9$， $\sigma_0^2 = 0.01$。此外，为了不让先验均值和方差影响过大取 $\kappa_0 = \upsilon_0 = 1$。根据例 2.5 的公式，在 R 平台上，通过以下命令就可以求出各后验参数值。

```
mu0<-1.9; k0<-1; s20<-.010; nu0<-1
y<- c( 1.64 , 1.70 , 1.72 , 1.74 , 1.82 , 1.82 , 1.82 , 1.90 , 2.08 )
n<-length(y); ybar<-mean(y); s2<-var(y)
kn<-k0+n; nun<-nu0+n
mun<-(k0*mu0 + n*ybar)/kn
s2n<-(nu0*s20+(n-1)*s2+k0*n*(ybar-mu0)^2/(kn))/(nun)
```

后验参数值如下

$$\mu_n = 1.814, \quad \sigma_n^2 = 0.015, \quad \kappa_n = \upsilon_n = 10$$

于是，得到具体的后验正态--逆伽玛分布 $N - IGamma(1.814, 10, 10, 0.015)$。这样，以后就可以利用这个后验分布进行统计推断了。

例 2.6 设 d 元随机向量 $\mathbf{X} = (X_1, \ldots, X_d)'$ 服从 d 维正态分布 $\mathrm{N_d}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，这里 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$ 是 d 维均值向量，$\boldsymbol{\Sigma}$ 为其 d 阶方差－协方差矩阵，其联合密度函数为

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$$

若从该分布中随机抽取一样本 $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$，则此样本的联合密度函数为

$$p(\mathbf{y}_1, \ldots, \mathbf{y}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})\}$$

现在假设$\mathbf{\Sigma}$已知，要来寻求均值向量$\boldsymbol{\mu}$的共轭先验分布。在 2.1 节，我们知道当 d=1 时正态分布是$\boldsymbol{\mu}$的共轭先验。由于上述联合密度函数的对数是$\boldsymbol{\mu}$的二次型，故 d 元正态分布应是$\boldsymbol{\mu}$的共轭先验分布，即可设$\boldsymbol{\mu}$的先验分布为$N_d(\boldsymbol{\mu}_0,\mathbf{\Lambda}_0)$，其中均值向量$\boldsymbol{\mu}_0$与方差-协方差阵$\mathbf{\Lambda}_0$都假设已给定，于是$\boldsymbol{\mu}$的后验密度有如下形式

$$\pi(\boldsymbol{\mu}|\mathbf{y})\propto\exp\{-\frac{1}{2}[(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\mathbf{\Lambda}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)+\sum_{i=1}^{n}(\mathbf{y}_i-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})]\}$$

进一步

$$\begin{aligned}\pi(\boldsymbol{\mu}|\mathbf{y})&\propto\exp\{-\frac{1}{2}[\boldsymbol{\mu}'\mathbf{\Lambda}_0^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}'\mathbf{\Lambda}_0^{-1}\boldsymbol{\mu}_0+n\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}-2n\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\bar{\mathbf{y}}]\}\\&\propto\exp\{-\frac{1}{2}[\boldsymbol{\mu}'(\mathbf{\Lambda}_0^{-1}+n\mathbf{\Sigma}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}'(\mathbf{\Lambda}_0^{-1}\boldsymbol{\mu}_0+n\Sigma^{-1}\bar{\mathbf{y}})]\}\\&\propto\exp\{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_d)'\mathbf{\Lambda}_d^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_d)\}\end{aligned}$$

这表明$\boldsymbol{\mu}$的后验密度是 d 元正态分布，其均值向量$\boldsymbol{\mu}_d$与方差一协方差矩阵$\boldsymbol{\Lambda}_d$分别为

$$\boldsymbol{\mu}_d = (\boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0 + n\boldsymbol{\Sigma}^{-1}\overline{\mathbf{y}})$$
$$\boldsymbol{\Lambda}_d^{-1} = \boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1}$$

这个结果类似于一元正态分布的结果，其后验均值向量是先验均值向量$\boldsymbol{\mu}_0$与样本均值向量$\overline{\mathbf{y}}$的加权平均，其权数由先验精度矩阵$\boldsymbol{\Lambda}_0^{-1}$与样本精度矩阵$n\boldsymbol{\Sigma}^{-1}$决定，而后验精度矩阵是先验精度矩阵与样本精度矩阵之和。

§2.3　充分统计量与应用

在经典统计中，我们已经了解了充分统计量的概念，它在统计推断中起着简化问题和计算的作用，是经典统计的一个重要概念。这个概念可以平移到贝叶斯统计中，从而成为经典统计和贝叶斯统计共享的一个概念。

2.3.1　充分统计量概念

现在回忆一下经典统计中充分统计量的定义：设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自分布 $F(x|\theta)$ 的一个样本，$T=T(\mathbf{x})$ 是一个统计量。如果在给定 $T(\mathbf{x})=t$ 的条件下，$\mathbf{x}$ 的条件分布与 θ 无关，即 $p(\mathbf{x}|\theta,T(\mathbf{x})=t)=p(\mathbf{x}|T(\mathbf{x})=t)$，则称该统计量为 θ 的充分统计量。

参数的充分统计量$T=T(\mathbf{x})$包含了样本$\mathbf{x}$所包含的有关该参数的全部信息,因此在利用样本进行该参数的统计推断时，就可以利用该参数的充分统计量代替样本来进行。要直接利用上述定义验证一个统计量的充分性一般是很困难的，因为需要计算条件分布。幸运的是有一个判别充分统计量的好方法，它就是著名的因子分解定理。

定理 2.1（费雪--奈曼因子分解定理）　一个统计量$T(\mathbf{x})$是参数θ的充分统计量的充要条件是存在一个 t 与θ的函数$g(t,\theta)$和另一个不含θ的函数$h(\mathbf{x})$，使得对任一样本$\mathbf{x}$和任意θ，样本的联合密度$p(\mathbf{x}|\theta)$可表示为它们的乘积，即

$$p(\mathbf{x}|\theta)=g(T(\mathbf{x}),\theta)h(\mathbf{x})$$

上述公式就是说，样本的联合密度可以用一个单纯的样本函数与一个统计量和参数的函数的乘积来表示。在许多情况下，这种表示不难找到。

定理 2.2 设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自密度函数 $p(x|\theta)$ 的样本，$T=T(\mathbf{x})$ 是统计量且密度函数为 $p(t|\theta)$，又设 $\Pi=\{\pi(\theta)\}$ 是参数 θ 的某个先验分布族。则 $T(\mathbf{x})$ 为 θ 的充分统计量的充要条件是对于 Π 中的任一先验分布 $\pi_s(\theta)$，有

$$\pi_s(\theta|T(\mathbf{x}))=\pi_s(\theta|\mathbf{x})$$

即用统计量 $T(\mathbf{x})$ 的分布算得的后验分布与样本分布 $p(\mathbf{x}|\theta)$ 算得的后验分布是相同的。

注:

1. 定理 2.2 中的条件是充分必要条件，因此可作为充分统计量的另一个定义，我们称之为充分统计量的贝叶斯定义， 它当然与经典定义等价。

2. 如果已知一个统计量 $T(\mathbf{x})$ 是充分的，那么按定理 2.2，后验分布可用该统计量的分布代替样本分布算得，从而简化了后验分布的计算。因此充分统计量可以看成样本的不减信息量的压缩。

3. 如果参数是多维情形，则统计量也应是相应维数的。以二维为例，设 $\theta=(\theta_1,\theta_2)$ 是二维的，则统计量 $T(\mathbf{x})=(T_1(\mathbf{x}),T_2(\mathbf{x}))$。

2.3.2 充分统计量的例

例 2.7 设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自指数分布 $\text{Exp}(\theta)$ 的样本。证明统计量

$T(\mathbf{x})=\sum_{i=1}^{n}x_i$ 是参数 θ 的充分统计量。

证明：指数分布 $\text{Exp}(\theta)$ 的分布密度是

$$f(x\mid\theta)=\theta e^{-\theta x}=\theta\exp(-\theta x),\ x\geq 0$$

于是样本的联合密度

$$f(\mathbf{x}\mid\theta)=\theta^n\exp(-\theta\sum_{i=1}^{n}x_i)$$

我们取

$$g(t,\theta)=\theta^n\exp(-\theta t),\ h(\mathbf{x})=1$$

显而易见

$$f(\mathbf{x}\mid\theta)=g(T(\mathbf{x}),\theta)h(\mathbf{x})$$

因此，根据因子分解定理，$T(\mathbf{x})=\sum_{i=1}^{n}x_i$ 是参数 θ 的充分统计量。

例 2.8 设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，则其联合密度函数为

$$p(\mathbf{x}|\mu,\sigma^2)=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$
$$=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}$$

其中

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ Q=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

我们取

$$g((s,t),\mu,\sigma^2)=\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[t+n(s-\mu)^2]\},\ h(\mathbf{x})=(2\pi)^{-\frac{n}{2}}$$

则显然

$$p(\mathbf{x}|\mu,\sigma^2)=g((\bar{x},Q),\mu,\sigma^2)h(\mathbf{x})$$

于是根据因子分解定理，二维统计量 $T=(\bar{x},Q)$ 是 (μ,σ^2) 的充分统计量。

现在设 $\pi(\mu,\sigma^2)$ 是 (μ,σ^2) 的任意一个先验，则根据贝叶斯公式，(μ,σ^2) 的后验密度为

$$\pi(\mu,\sigma^2|\mathbf{x})=\frac{\sigma^{-n}\pi(\mu,\sigma^2)\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}}{\int_0^{\infty}\int_{-\infty}^{\infty}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}\pi(\mu,\sigma^2)d\mu d\sigma^2}$$

另一方面，根据经典统计知识，$\bar{x}\sim N(\mu,\sigma^2/n),\ Q/\sigma^2\sim\chi^2(n-1)$，即 $\bar{x}$ 与 Q 的分布密度分别为

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

$$p(Q|\mu,\sigma^2)=\frac{1}{\Gamma(\frac{n-1}{2})(2\sigma^2)^{\frac{n-1}{2}}}Q^{\frac{n-3}{2}}\exp\{-Q/2\sigma^2\}$$

同样由经典统计知 $\bar{x}$ 与 Q 独立，所以 $\bar{x}$ 与 Q 的联合密度是各密度之积

$$p(\bar{x},Q\,|\,\mu,\sigma^2)=\frac{\sqrt{n}/\sqrt{2\pi}\sigma}{\Gamma(\frac{n-1}{2})(2\sigma^2)^{\frac{n-1}{2}}}Q^{\frac{n-3}{2}}\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}$$

利用与前面相同的先验分布 $\pi(\mu,\sigma^2)$，我们得在给定 $\bar{x}$ 与 Q 下的后验分布

$$\pi(\mu,\sigma^2\,|\,\bar{x},Q)=\frac{\sigma^{-n}\pi(\mu,\sigma^2)\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}}{\int_0^{\infty}\int_{-\infty}^{\infty}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\}\pi(\mu,\sigma^2)d\mu d\sigma^2}$$

容易看出两个后验分布相等，即

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

由此可见，用充分统计量 $(\bar{x},Q)$ 的分布算得的后验分布与样本分布算得的后验分布是相同的。

例 2.9 设样本 $\mathbf{x}=(x_1,\cdots,x_n)$ 来自泊松分布 $Poisson(\lambda)$， 其分布 $p(x\mid\lambda)=\dfrac{\lambda^x}{x!}e^{-\lambda}, x=0,1,2,\cdots$

用贝叶斯公式验证统计量 $T(\mathbf{x})=\sum_{i=1}^{n}x_i$ 是参数 λ 的充分统计量。

解： 由概率论知识我们知道泊松分布具有可加性， 即统计量 $T(\mathbf{x})=\sum_{i=1}^{n}x_i$ 服从泊松分布

$Poisson(n\lambda)$， 其分布列 $$P(T(\mathbf{x})=t\mid\lambda)=p_T(t\mid\lambda)=\frac{(n\lambda)^t}{t!}e^{-n\lambda}, t=0,1,2,\cdots$$

现设先验分布族为 $\{\pi(\lambda);\lambda\in\Lambda\}$ （ Λ 是参数空间）， 则对任意先验 $\pi(\lambda)$， 由贝叶斯公式，

后验分布 $\pi(\lambda\mid\mathbf{x})\propto p(\mathbf{x}\mid\lambda)\pi(\lambda)\propto\lambda^{\sum_{i=1}^{n}x_i}e^{-n\lambda}\pi(\lambda)$

另一方面， 当统计量 $T(\mathbf{x})=\sum_{i=1}^{n}x_i=t$ 已知时， 由其分布计算的后验分布

$$\pi(\lambda\mid T(\mathbf{x}))\propto p_T(t\mid\lambda)\pi(\lambda)\propto(n\lambda)^t e^{-n\lambda}\pi(\lambda)\propto\lambda^{\sum_{i=1}^{n}x_i}e^{-n\lambda}\pi(\lambda)$$

比较以上两式， 我们看到两个后验的核是相同的， 从而有 $\pi(\lambda\mid T(\mathbf{x}))=\pi(\lambda\mid\mathbf{x})$， 这样， 根据定

理 2.2， 统计量 $T(\mathbf{x})=\sum_{i=1}^{n}x_i$ 是充分统计量。

Homework：

P27，1，2，3，4，5，6，7， 8，10