第一章 先验分布与后验分布

第一章 先验分布与后验分布

一、统计推断中可用的三种信息

二、贝叶斯公式

三、共轭先验分布

四、超参数及其确定

五、多参数模型

六、充分统计量

§1.1 统计推断中可用的三种信息

1.总体信息：总体分布或所属分布族提供给我们的信息

2.样本信息：从总体抽取的样本提供给我们的信息

3.先验信息：在抽样之前有关统计推断的一些信息。

§ 1.2 贝叶斯公式

贝叶斯统计学的基础是著名的贝叶斯公式，它是英国学者贝叶斯（**T.R.Bayes** 1702~1761）在他死后二年发表的一篇论文《论有关机遇问题的求解》中提出的。经过二百年的研究与应用，贝叶斯的统计思想得到很大的发展，目前已形成一个统计学派—贝叶斯学派。

为了纪念他，英国历史最悠久的统计杂志《**Biometrika**》在1958年又全文刊登贝叶斯的这篇论文。

一、贝叶斯公式的三种形式

初等概率论中的贝叶斯公式是用事件的概率形式给出的。可在贝叶斯统计学中应用更多的是贝叶斯公式的密度函数形式。

1.贝叶斯公式的事件形式：

假定 $A_1,\cdots,A_k$ 是互不相容的事件，它们之和$\bigcup_{i=1}^{k}A_i$包含事件B，即 $B\subset\bigcup_{i=1}^{k}A_i$，则有：

$$P(A_i\mid B)=\frac{P(A_i)P(B\mid A_i)}{\sum_{i=1}^{k}P(A_i)P(B\mid A_i)}$$

2.贝叶斯公式的密度函数形式：

在给出贝叶斯公式的密度函数形式之前，先介绍以下贝叶斯学派的一些具体思想或者叫着基本假设：

假设Ⅰ 随机变量X有一个密度函数$p(x;\theta)$，其中θ是一个参数，不同的θ对应不同的密度函数，故从贝叶斯观点看，$p(x;\theta)$是在给定θ后的一个条件密度函数，因此记为 $p(x|\theta)$ 更恰当一些。这个条件密度能提供我们的有关的θ信息就是总体信息。

假设Ⅱ 当给定θ后，从总体$p(x|\theta)$中随机抽取一个样本$X_1,X_2,\cdots,X_n$，该样本中含有θ的有关信息。这种信息就是样本信息。

假设Ⅲ 从贝叶斯观点来看，未知参数$\boldsymbol{\theta}$是一个随机变量。而描述这个随机变量的分布可从先验信息中归纳出来，这个分布称为先验分布，其密度函数用$\pi(\theta)$表示。

（1） **先验分布**

定义1 将总体中的未知参数$\theta\in\Theta$看成一取值于Θ的随机变量，它有一概率分布，记为$\pi(\theta)$，称为参数$\boldsymbol{\theta}$的**先验分布**。

（2） **后验分布**

在贝叶斯统计学中，把以上的三种信息归纳起来的最好形式是在总体分布基础上获得的样本$X_1,X_2,\cdots,X_n$和参数$\boldsymbol{\theta}$的联合密度函数：

$$h(x_1,\cdots,x_n,\theta)=p(x_1,\cdots,x_n|\theta)\pi(\theta)$$

在这个联合密度函数中。当样本 $X_1,\cdots,X_n$ 给定之后，未知的仅是参数 $\boldsymbol{\theta}$ 了，我们关心的是样本给定后，$\boldsymbol{\theta}$ 的条件密度函数，依据密度的计算公式，容易获得这个条件密度函数：

$$\pi(\theta|x_1,\cdots,x_n)=\frac{h(x_1,\cdots,x_n,\theta)}{m(x_1,\cdots,x_n)}=\frac{p(x_1,\cdots,x_n|\theta)\pi(\theta)}{\int_\Theta p(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta}$$

这就是贝叶斯公式的密度函数形式，其中 $\pi(\theta|x_1,\cdots,x_n)$ 称为 $\boldsymbol{\theta}$ 的**<u>后验密度函数</u>**<u>（或</u>**<u>后验分布</u>**<u>）</u>。而：

$$m(x_1,\cdots,x_n)=\int_\Theta p(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta$$

是样本的边际分布，或称样本 $X_1,\cdots,X_n$ 的无条件分布，它的积分区域就是参数 $\boldsymbol{\theta}$ 的取值范围，随具体情况而定。

3.贝叶斯公式的离散形式：

当θ是离散随机变量时，先验分布可用先验分布列$\pi(\theta_i)$，这时后验分布也是离散形式：

$$\pi(\theta_i \mid x) = \frac{p(x \mid \theta_i)\pi(\theta_i)}{\sum_j p(x \mid \theta_j)\pi(\theta_j)},\ i = 1,2,\cdots$$

假如总体X也是离散的，则只须将$p(x \mid \theta)$换成$P(X = x \mid \theta)$即可。

二 后验分布是三种信息的综合

前面的分析总结如下：人们根据先验信息对参数 θ 已有一个认识，这个认识就是先验分布 $\pi(\theta)$。通过试验，获得样本。从而对 θ 的先验分布进行调整，调整的方法就是使用上面的贝叶斯公式，调整的结果就是后验分布 $\pi(\theta|x_1,\cdots,x_n)$。后验分布是三种信息的综合。获得后验分布使人们对 θ 的认识又前进一步，可看出，获得样本的的效果是把我们对 θ 的认识由 $\pi(\theta)$ 调整到 $\pi(\theta|x_1,\cdots,x_n)$。所以对 θ 的统计推断就应建立在后验分布 $\pi(\theta|x_1,\cdots,x_n)$ 的基础上。

例1.4 设事件A的概率为θ，即$P(A)=\theta$。为了估计θ而作n次独立观察，其中事件A出现次数为X，则有X服从二项分布$b(n,\theta)$

即$P(X=x|\theta)=C_n^x\theta^x(1-\theta)^{n-x}, x=0,1,\cdots,n.$ 如何求出后验分布？

解题步骤：1. 作贝叶斯假设。如果此时我们对事件A的发生没有任何了解，对θ的大小也没有任何信息。在这种情况下，贝叶斯建议用区间$(0,1)$上的均匀分布作为θ的先验分布。因为它在$(0,1)$上每一点都是机会均等的。因此：

$$\pi(\theta)=\begin{cases}1, & 0<\theta<1\\ 0, & \text{其他}\end{cases}$$

2.计算样本X与参数θ的联合分布：

$$h(x,\theta)=C_n^x\theta^x(1-\theta)^{n-x}, x=0,1,\cdots,n,\quad 0<\theta<1$$

此式在定义域上与二项分布有区别。

3. 计算X的边际密度为:

$$\boldsymbol{m}(\boldsymbol{x})=\int_0^1 \boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})\boldsymbol{d\theta}=\boldsymbol{C}_n^x\,\frac{\Gamma(\boldsymbol{x}+1)\Gamma(\boldsymbol{n}-\boldsymbol{x}+1)}{\Gamma(\boldsymbol{n}+2)},\boldsymbol{x}=0,1,\cdots,\boldsymbol{n}$$

4. 利用贝叶斯公式可得 θ 的后验分布：

$$\pi(\theta|x)=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x},0<\theta<1$$

即：$X\sim Be(x+1,n-x+1)$

5. 具体算例。拉普拉斯计算过这个概率，研究男婴的诞生比例是否大于0.5? 如果抽了251527个男婴，女婴241945个。他选用 $U(0,1)$ 作为 θ 的先验分布，于是可得 θ 的后验分布 $Be(x+1, n-x+1)$，其中

$$n = 251527 + 241945 = 493472, \qquad x = 251527$$

由此拉普拉斯计算了"$\theta \le 0.5$"的后验概率：

$$P(\theta \le 0.5 \mid x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \int_0^{0.5} \theta^x (1-\theta)^{n-x} d\theta = 1.15 \times 10^{-42}$$

故他断言男婴诞生的概率大于0.5。

注： 1.伽玛分布与贝塔分布简介：

$$\Gamma(s)=\int_0^{\infty}x^{s-1}e^{-x}dx, s>0,$$

$$\Gamma(s+1)=s\Gamma(s), \qquad \Gamma(n+1)=n!$$

$$B(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}dx, \quad p>0,q>0$$

$$B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \quad p>0,q>0$$

定义： 定义在[0，1]上，且用密度函数：

$$p(\theta;p,q)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\theta^{p-1}(1-\theta)^{q-1}, 0\le\theta\le 1, p<0, q>0$$

表示的概率分布称为 β_{I} 型分布，记为 $\beta_{\mathrm{I}}(p,q)$ 或者 $\beta_e(p,q)$ 。

2.特例：

① 当 $p = q = 1$ 时， $\beta_{\mathrm{I}}(1,1)$ 型分布即为区间[0， 1]上的均匀分布；

② 当 $p = q = \frac{1}{2}$ 时，$\beta_{\mathrm{I}}(\frac{1}{2}, \frac{1}{2})$ 型分布称为反正弦分布，密度函数为：

$$p(x) = \frac{1}{\sqrt{\pi x(1-x)}}，\ 0 < x < 1$$

③ 设 $\boldsymbol{x_i} \in \boldsymbol{U}(0,1)$， 则 $\boldsymbol{x_{(k)}}$的密度函数为：

$$p(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}，\ 0 < x < 1$$

即： $x_{(k)} \sim \beta_I(k, n-k+1)$

3. 数字特征:

（1）伽玛分布，其密度函数为

$$p(x;\alpha,\lambda)=\begin{cases}\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, x\geq 0\\ 0, x<0\end{cases} \qquad (\alpha>0,\lambda>0)$$

$$EX=\frac{\alpha}{\lambda},\quad \mathrm{Var}X=\frac{\alpha}{\lambda^2}$$

（2）贝塔分布

$$EX=\frac{p}{p+q},\quad \mathrm{Var}X=\frac{pq}{(p+q)^2(p+q+1)}$$

4. 为什么将贝塔分布作为 θ的先验分布族是恰当的?

(1) 参数 θ是废品率，它仅在(0,1)上取值。因此，必需用区间 (0,1) 上的一个分布去拟合先验信息。β分布正是这样一个分布。

(2) β分布含有两个参数 p 与 q，不同的 p 与 q 就对应不同的先验分布，因此这种分布的适应面较大。

(3)样本X的分布为二项分布 时，假如θ的先验分布为β分布，则用贝叶斯估计算得的后验分布仍然是β分布，只是其中的参数不同。这样的先验分布(β分布)称为参数θ的共轭先验分布。选择共轭先验分布在处理数学问题上带来不少方便。

例1.5 投资决策问题

为了提高某产品的质量，公司经理考虑增加投资来改进生产设备，预计需投资100万元，但从投资效果看，下属部门有两种意见:

- θ_1： 改进生产设备后，高质量产品可占90%
- θ_2： 改进生产设备后，高质量产品可占70%

问：公司经理怎样决策?

分析：

经理当然希望 θ_1 发生，公司效益可得很大提高，投资改进设备也是合算的。但根据下属二个部门过去建议被采纳的情况，经理认为，θ_1 的可信程度只有 40%，θ_2 的可信程度是 60%。即

$$\pi(\theta_1)=0.4, \qquad \pi(\theta_2)=0.6$$

这二个都是经理的主观概率。经理不想仅用过去的经验来决策此事，想慎重一些，通过小规模试验后观其结果再定。为此做了一项试验，试验结果（记为 A）如下：

A：试制五个产品，全是高质量的产品。

分析：

经理对这次试验结果很高兴，希望用此试验结果来修改他原先对 θ_1 和 θ_2 的看法，即要求后验概率 $\pi(\theta_1|A)$ 与 $\pi(\theta_2|A)$。这可用贝叶斯公式的离散形式(1.1.2)来完成。如今已有先验概率 $\pi(\theta_1)$ 与 $\pi(\theta_2)$。还需要二个条件概率 $P(A|\theta_1)$ 与 $P(A|\theta_2)$，这可用二项分布算得，

$$P(A|\theta_1)=0.9^5=0.590, \qquad P(A|\theta_2)=0.7^5=0.168$$

由全概率公式可算得 $P(A)=P(A|\theta_1)\pi(\theta_1)+P(A|\theta_2)\pi(\theta_2)=0.337$。最后由(1.2)式可算得，

$$\pi(\theta_1|A)=P(A|\theta_1)\pi(\theta_1)/P(A)=0.236/0.337=0.700$$

$$\pi(\theta_2|A)=P(A|\theta_2)\pi(\theta_2)/P(A)=0.101/0.337=0.300$$

这表明，经理根据试验 A 的信息调整自己的看法，把对 θ_1 与 θ_2 的可信程度由 0.4 和 0.6 调整到 0.7 和 0.3。后者是综合了经理的主观概率和试验结果而获得的，要比主观概率更有吸引力，更贴近当今的实际，这就是贝叶斯公式的应用。

分析：

经过实验 A 后，经理对增加投资改进质量的兴趣增大。但因投资额大，还想再做一次小规模试验，观其结果再作决策。为此又做了一批试验，试验结果（记为 B）如下：

B：试制 10 个产品，有 9 个是高质量产品。

经理对此试验结果更为高兴。希望用此试验结果对 θ_1 与 θ_2 再作一次调整。为此把上次后验概率看作这次的先验概率，即

$$\pi(\theta_1)=0.7, \pi(\theta_2)=0.3$$

用二项分布还可算得

$$P(B|\theta_1)=10(0.9)^9(0.1)=0.387$$
$$P(B|\theta_2)=10(0.7)^9(0.3)=0.121$$

由此可算得 $P(B)=0.307$ 和后验概率 $\pi(\theta_1|B)=0.883$，$\pi(\theta_2|B)=0.117$。

经理看到，经过二次试验，θ_1（高质量产品可占 90%）的概率已上升到 0.883，到可以下决心的时候了，他能以 88.3% 的把握保证此项投资能取得较大经济效益。

§1.3 共轭先验分布

一、共轭先验分布

定义2 设 θ 是总体分布中的参数（或参数向量），$\pi(\theta)$是 θ 的先验密度函数，假如由抽样信息算得的后验密度函数与 $\pi(\theta)$ 有相同的形式，则称 $\pi(\theta)$是 θ 的（自然）共轭先验分布。

注意：共轭先验分布是对某一分布中的参数而言的。如正态均值、正态方差、泊松均值等。离开指定参数及其所在的分布去谈论共轭先验分布是没有意义的。

例1.6 证明：正态均值（方差已知）的共轭先验分布是正态分布。

证明思路：

(1) 写出样本的似然函数：

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

$$-\infty < x_1, \cdots, x_n < +\infty$$

(2) 确定先验分布：

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

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

其中 μ 与 τ^2 为已知

(3)计算后验分布:

写出样本 $\boldsymbol{x}$ 与参数 θ 的联合密度函数

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

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

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

则有

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

$$=k_2\exp\left\{-\frac{(\theta-B/A)^2}{2/A}\right\}$$

其中 $k_2=k_1\exp\{-\frac{1}{2}(C-B^2/A)\}$。由此容易算得样本 $\boldsymbol{x}$ 的边缘分布

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

上面两式相除，即得 θ 的后验分布

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

这是正态分布 $N(\mu_1,\tau_1^2)$，其均值 μ_1 与方差 τ_1^2 分别为

$$\mu_1=\frac{B}{A}=\frac{\bar{x}\sigma_0^{-2}+\mu\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}},\qquad \frac{1}{\tau_1^2}=\frac{1}{\sigma_0^2}\overset{+}{\cancel{=}}\frac{1}{\tau^2}\tag{1.3.4}$$

这就说明了正态均值(方差已知)的共轭先验分布是正态分布。譬如，设 $X\sim N(\theta,2^2)$，$\theta\sim N(10,3^2)$。若从正态总体 X 抽得容量为 5 的样本，算得 $\bar{x}=12.1$，于是可从(1.3.4)算得 $\mu_1=11.93$ 和 $\tau_1^2=\left(\frac{6}{7}\right)^2$。这时正态均值 θ 的后验分布为正态分布 $N(11.93,\left(\frac{6}{7}\right)^2)$。

补充例题：

设X表示人的胸围，根据经验，胸围是近似服从正态分布的。现测量了 $n=10000$ 个人的胸围，得样本均值为 39.8(cm)，样本方差为 4，假设 θ 的先验分布为 $N(38,9)$，求θ的后验分布。**(答案：** $N(39.8,1/2500)$)

说明：样本较大时，似然函数起决定作用，先验信息几乎不起做用。

二、怎样简化后验分布的计算

——省略常数因子

- 在给定样本分布 $p(x|\theta)$ 和先验分布 $\pi(\theta)$ 后可用贝叶斯公式计算 θ 的后验分布：

$$\pi(\theta|x)=\frac{p(x|\theta)\pi(\theta)}{m(x)}$$

- 由于 $m(x)$ 不依赖于 θ ，在计算 θ 的后验分布中仅起到一个正则化因子的作用。假如把 $m(x)$ 省略，把贝叶斯公式改写成如下等价形式：

$$\pi(\theta|x)\propto p(x|\theta)\pi(\theta)$$

- 其中符号“$\propto$”表示两边仅差一个常数因子，一个不依赖于 θ 的常数因子。上式右端称为后验分布 $\pi(\theta|x)$ 的核。

利用后验分布的核重新证明例1.6

在μ与τ^2已知的情况下，θ后验分布为

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

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

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

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

例1.7 证明：二项分布的成功概率θ的共轭先验分布是贝塔分布。

证明：设总体 $X \sim b(n,\theta)$,则 $\boldsymbol{b}(\boldsymbol{n},\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{x}(1-\boldsymbol{\theta})^{n-x}$ 。再设 θ 的先验分布为贝塔分布,即 $\mathrm{B}e(\boldsymbol{\alpha},\boldsymbol{\beta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$ ，其中参数已知。由此可写出 θ 的后验分布：

$$\boldsymbol{\pi}(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto \boldsymbol{\theta}^{\alpha+x-1}(1-\boldsymbol{\theta})^{\beta+n-x-1}, \quad 0<\boldsymbol{\theta}<1$$

这是贝塔分布的核，其密度函数为：

$$\boldsymbol{\pi}(\boldsymbol{\theta} \mid \boldsymbol{x})=\frac{\Gamma(\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{n})}{\Gamma(\boldsymbol{\alpha}+\boldsymbol{x})\Gamma(\boldsymbol{\beta}+\boldsymbol{n}-\boldsymbol{x})}\boldsymbol{\theta}^{\alpha+x-1}(1-\boldsymbol{\theta})^{\beta+n-x-1}, \quad 0<\boldsymbol{\theta}<1$$

三、共轭先验分布的优缺点

共轭先验分布在很多场合被采用，因为它有两个优点：

（1）计算方便；

（2）后验分布的一些参数可得到很好的解释。

不足：怎样找到合适的先验分布？

例1.8 例1.6中后验均值与后验方差的合理解释

由例1.6知 $$\mu_1 = \frac{B}{A} = \frac{\bar{x}\sigma_0^{-2} + \mu\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}}$$

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

其中 $\gamma = \dfrac{\sigma_0^{-2}}{\sigma_0^{-2} + \tau^{-2}}$ 是用方差倒数组成的权，于是后验均值 μ_1 是样本均值与先验均值 μ 的加权平均。

- 而 $\dfrac{1}{\tau_1^2} = \dfrac{1}{\sigma_0^2} + \dfrac{1}{\tau^2} = \dfrac{n}{\sigma^2} + \dfrac{1}{\tau^2}$ 可解释为：后验分布的精度是样本均
- 值分布的精度与先验分布精度之和，增加样本量n或减少先
- 验分布方差都有利于提高后验分布的精度。

例1.9 对例1.7中后验分布的均值和方差的解释。

分析： 后验分布 $Be(\alpha+x,\beta+n-x)$ 的均值和方差可写为：

$$E(\theta|x)=\frac{\alpha+x}{\alpha+\beta+n}$$

$$=\frac{n}{\alpha+\beta+n}\frac{x}{n}+\frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta}$$

$$=\gamma\cdot\frac{x}{n}+(1-\gamma)\cdot\frac{\alpha}{\alpha+\beta}$$

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

$$=\frac{E(\theta|x)[1-E(\theta|x)]}{\alpha+\beta+n+1}$$

其中 $\gamma=n/(\alpha+\beta+n)$，$x/n$ 是样本均值，$\alpha/(\alpha+\beta)$ 是先验均值，从上述加权平均可见，后验均值是介于样本均值与先验均值之间，它偏向哪一侧由 γ 的大小决定。另外，当 n 与 x 都较大，且 x/n 接近某个常数 θ_0 时，我们有

$$E(\theta|x)\approx\frac{x}{n}$$

$$\mathrm{Var}(\theta|x)\approx\frac{1}{n}\,\frac{x}{n}\left(1-\frac{x}{n}\right)$$

这表明：当样本量增大时，后验均值主要决定于样本均值，而后验方差愈来愈小。这时后验密度曲线的变化可从图 1.3.1 中看到，随着 n 与 z 在成比例地增加时，后验分布愈来愈向比率 x/n 集中，这时先验信息对后验分布的影响将愈来愈小。

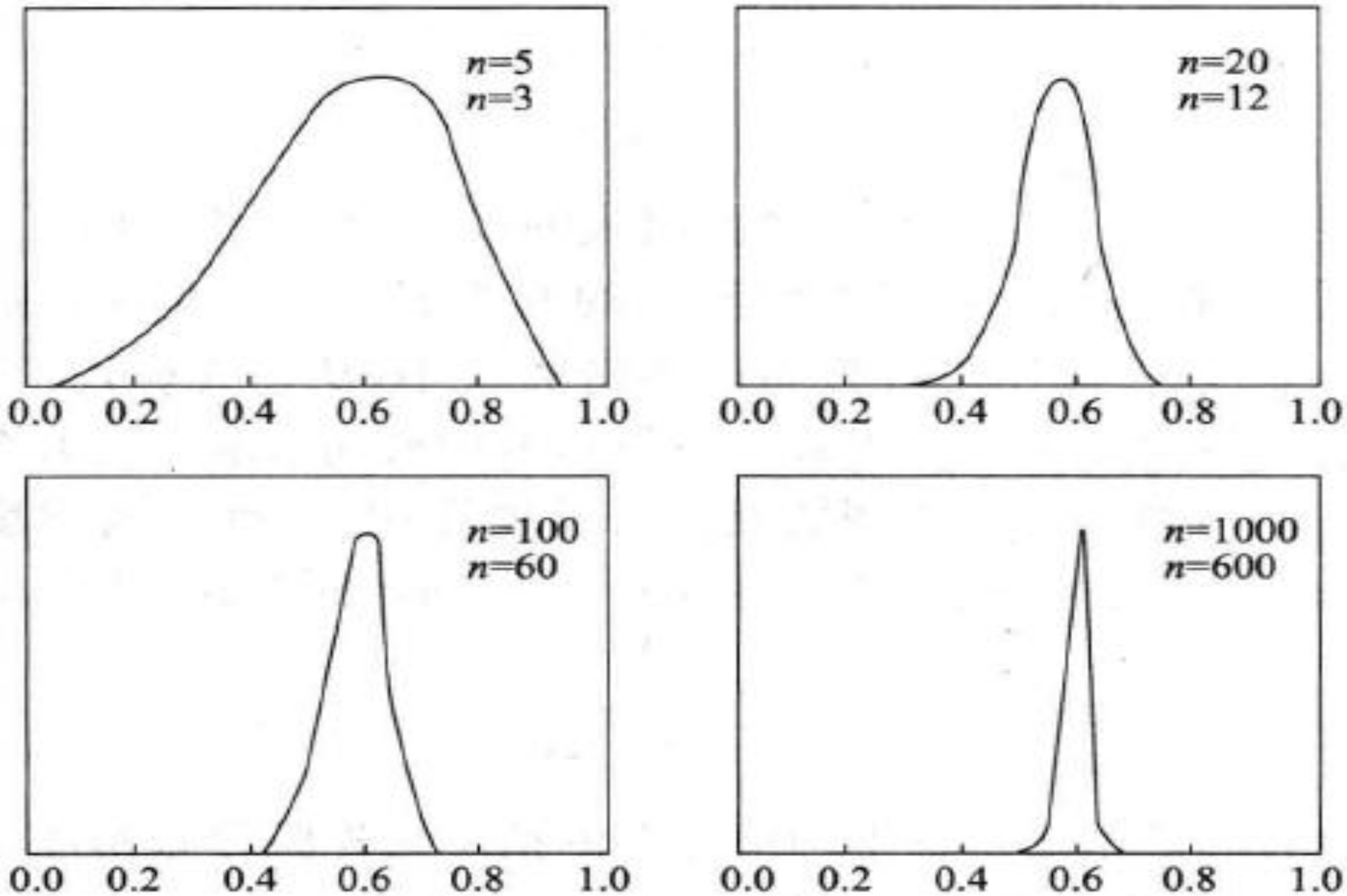

图 1.3.1　当 n 与 x 成比例增加时，后验密度——贝塔分布 $Be(\alpha+x, \beta+n-x)$—变化情况(纵坐标刻度不同)

四、 常用的一些共轭先验分布

共轭先验分布选取的一般原则：是由似然函数 L(θ)=p(x|θ)中所含的因式所决定的，即选与似然函数具有相同核的分布作为先验分布。

例1.10 设 $x_1,\cdots,x_n$ 是来自正态分布$N(\theta,\sigma^2)$ 的一个样本观测值，其中θ已知，求 σ方差的共轭先验分布。

四、 常用的一些共轭先验分布

共轭先验分布选取的一般原则：是由似然函数 $L(\theta)=p(x|\theta)$ 中所含的因式所决定的，即选与似然函数具有相同核的分布作为先验分布。

例 1.10 设 $x_1,\cdots,x_n$ 是来自正态分布 $N(\theta,\sigma^2)$ 的一个样本观测值，其中 θ 已知，求方差 σ^2 的共轭先验分布。

解题的基本思路：

写出样本的似然函数：

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

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

上述似然函数中 σ^2 的因式将决定 σ^2 的共轭先验分布的形式，什么分布具有上述的核呢？

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

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

通过概率运算可以求得 $Y=X^{-1}$ 的密度函数

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

这个分布称为**倒伽玛分布**，记为 $IGa(\alpha,\lambda)$，其均值为 $E(y)=\lambda/(\alpha-1)$。假如取此倒伽玛分布为 σ^2 的先验分布，其中参数 α 与 λ 已知，则其密度函数为

$$\pi(\sigma^2)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}e^{-\lambda/\sigma^2},\sigma^2>0$$

于是 σ^2 的后验分布为

$$\pi(\sigma^2 \mid \boldsymbol{x}) \propto p(\boldsymbol{x} \mid \sigma^2)\pi(\sigma^2)$$

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

容易看出,这仍是倒伽玛分布 $IGa\left(\alpha+\frac{n}{2},\lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right)$,这表明,倒伽玛分布 $IGa(\alpha,\lambda)$是正态方差 σ^2 的共轭先验分布。

有趣的是:这个 σ^2 后验分布的均值可改等为如下加权平均

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

$$= \gamma \cdot \frac{\lambda}{\alpha - 1} + (1-\gamma) \cdot \frac{1}{n\cancel{2}} \sum_{i=1}^{n} (x_i - \theta)^2$$

其中权 $\gamma = \dfrac{\alpha - 1}{\alpha + \frac{n}{2} - 1}$，$\dfrac{\lambda}{\alpha - 1}$ 是 σ^2 的共轭先验分布 $IGa(\alpha, \lambda)$ 的先验均值，$\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \theta)^2$

是在 θ 已知条件下的样本方差(样本对 θ 的偏差平方的平均)，由此可知，在取 σ^2 的共轭先验分布场合，其后验均值是 σ^2 的先验均值与样本方差的加权平均。当样本量 n 足够大时，γ 接受于0，从而后验均值 $E(\sigma^2/x)$ 主要由样本方差决定。而当 n 不大时，后验均值 $E(\sigma^2/x)$ 是介于 σ^2 的先验均值与样本方差之间的某一个数。

常用的一些共轭先验分布

总体分布	参数	共轭先验分布	后验分布的期望
正态分布 $N(\theta,\sigma^2)$	均值	正态分布 $N(\mu,\tau^2)$	$\dfrac{\tau^{-2}\bar{x}+\mu\sigma^{-2}}{\tau^{-2}+\sigma^{-2}}$
正态分布 $N(\theta,\sigma^2)$	方差	倒Γ分布IGa(a,b)	
二项分布 $b(n,p)$	成功概率	β 分布 $\beta(a,b)$	$\dfrac{a+x}{a+b+x+n}$
Poisson分布 $P(\theta)$	均值	Γ分布$Ga(a,b)$	$\dfrac{a+n\bar{x}}{b+\cancel{1}n}$
指数分布	均值的倒数	Γ分布$Ga(a,b)$	

§ 1.4 超参数及其确定

- 一、定义：先验分布中所含的未知参数称为超参数。
- 二、估计方法：共轭先验分布是一种有信息的先验分布，故其中所含的超参数应充分利用各种先验信息来确定它，下面用一个例子来介绍目前国内外文献中对超参数的估计方法：
- 问题：二项分布中成功概率θ的共轭先验分布是贝塔分布$Be(\alpha,\beta)$，怎样确定两个超参数α和β？

1.利用先验矩：

假如根据先验信息能获得成功概率 θ 的若干个估计值，记为 $\theta_1,\theta_2,\cdots,\theta_k$，一般它们是从历史数据整理加工获得的，由此可算得先验均值 $\bar{\theta}$ 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 $Be(\alpha,\beta)$ 的期望与方差，即

$$\begin{cases}\dfrac{\alpha}{\alpha+\beta}=\bar{\theta}\\[2ex]\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=s_\theta^2\end{cases}$$

解之，可得超参数 α 与 β 的估计值

$$\hat{\alpha}=\bar{\theta}\left(\frac{(1-\bar{\theta})\bar{\theta}}{s_\theta^2}-1\right)$$

$$\hat{\beta}=(1-\bar{\theta})\left(\frac{(1-\bar{\theta})\bar{\theta}}{s_\theta^2}-1\right)$$

2.利用先验分位数：

假如根据先验信息可以确定贝塔分布的二个分位数，则可用这二个分位数来确定 α 与 β，譬如用二个上、下四分位数 θ_U 与 θ_L（见图 1.4.1 来确定 α 与 β，θ_U 与 θ_L 分别满足如下二个方程

$$\int_0^{\theta_L} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta = 0.25$$

$$\int_{\theta_U}^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta = 0.25$$

从这二个方程解出 α 与 β 即可确定超参数

2.利用先验分位数：

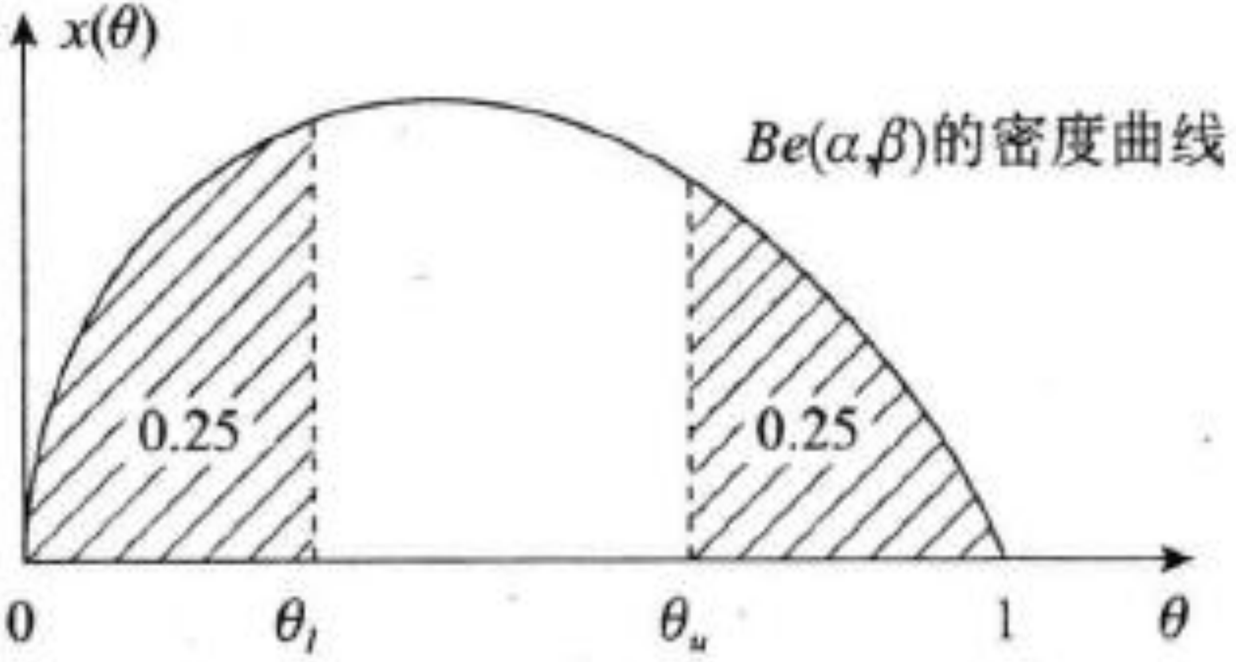

图 1.4.1 贝塔分布的上，下四分位数 θ_U 与 θ_L

2.利用先验分位数：

表 1.4.1 贝塔分布的上、下四分位数 θ_U 与 θ_L

$B(a,b)$	θ_L	θ_U	$B(a,b)$	θ_L	θ_U
(1,1)	0.250	0.750	(4,6)	0.135	0.255
(1,2)	0.133	0.500	(4,21)	0.102	0.264
(1,3)	0.092	0.379			
(1,4)	0.069	0.293	(5,5)	0.350	0.650
(1,9)	0.032	0.143	(5,10)	0.249	0.413
(1,14)	0.020	0.095	(5,15)	0.180	0.312
(1,19)	0.015	0.072	(5,20)	0.142	0.250
(1,24)	0.012	0.056			
			(6,4)	0.498	0.710
(2,1)	0.500	0.867	(6,9)	0.311	0.485
(2,2)	0.326	0.674	(6,11)	0.224	0.368
(2,3)	0.243	0.544	(6,19)	0.180	0.296
(2,8)	0.107	0.274			
(2,13)	0.068	0.180	(7,3)	0.609	0.805
(2,18)	0.051	0.136	(7,8)	0.377	0.554
(2,23)	0.040	0.109	(7,13)	0.274	0.422
			(7,18)	0.216	0.339
(3,1)	0.607	0.909			
(3,2)	0.456	0.757	(8,2)	0.726	0.886
(3,7)	0.196	0.392	(8,7)	0.446	0.623
(3,12)	0.125	0.263	(8,12)	0.322	0.475
(3,17)	0.091	0.196	(8,17)	0.258	0.383
(3,22)	0.072	0.157			
			(9,1)	0.857	0.968
(4,1)	0.707	0.931	(9,6)	0.515	0.689
(4,6)	0.290	0.512	(9,11)	0.372	0.526
(4,11)	0.184	0.339	(9,16)	0.292	0.425

3.利用先验矩和先验分位数

假如根据先验信息可获得先验均值 $\bar{\theta}$ 和 p 分位数 θ_p，则可列出下列方程

$$\begin{cases} \dfrac{\alpha}{\alpha+\beta}=\bar{\theta} \\ \displaystyle\int_0^{\theta_p} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta=p \end{cases}$$

解之，可得超参数 α 与 β 的估计值。

4.其它方法

假如根据先验信息只能获得先验均值 $\bar{\theta}$，这时可令

$$\frac{\alpha}{\alpha+\beta}=\bar{\theta}$$

一个方程不能唯一确定二个参数，这时还要利用其它先验信息才能把 α 与 β 确定下来。譬如，可藉助使用者对先验均值 $\bar{\theta}$ 的可信程度的大小来确定 α 与 β，例如，$\bar{\theta}=0.4$，那么满足方程 $\alpha/(\alpha+\beta)=0.4$ 的 α 与 β 有无穷多组解，表 1.4.2 列出若干组，从表 1.4.2 可见，它们的方差 $Var(\theta)$ 随着 $\alpha+\beta$ 的增大而减小，方差减少意味着概率在向均值 $E(\theta)$ 集中，从而提高人们对 $E(\theta)=0.4$ 的确信程度，这样一来，选择 $\alpha+\beta$ 的问题就转化为决策人对 $E(\theta)=0.4$ 的确信程度大小的问题，若对 $E(\theta)=0.4$ 很确信，那 $\alpha+\beta$ 可选得大一些，若对 $E(\theta)=0.4$ 尚存疑惑，那 $\alpha+\beta$ 就选的小一些，譬如决策人对 $E(\theta)=0.4$ 很确信，从而选 $\alpha+\beta=35$，从表 1.4.2 可见，此时 $\hat{\alpha}=14$，$\hat{\beta}=21$，这样 θ 的先验分布就为贝塔分布 Be(14,21)。

§1.5 多参数模型

- 由以上几节内容可知，求某一个参数的后验分布的基本思想可概括为：先根据先验信息给出参数的先验分布，然后按贝叶斯公式算得后验分布，即：

$$\left.\begin{array}{r}\text{总体的密度函数} p(x\mid\theta) \\ \text{先验密度} \pi(\theta)\end{array}\right\} \Rightarrow \text{参数}\theta\text{的后验密度} \pi(\theta\mid x) \propto p(x\mid\theta)\pi(\theta)$$

- 但在很多实际问题中却包含有多个未知参数的情形，如正态分布、多项分布以及多元正态分布等，此时可采用与单参数相似的方法来求参数的后验分布，而把其它的参数看成是讨厌参数。

例1.12 试求正态均值与正态方差的（联合）共轭先验分布及后验分布。（P24）

1.取先验分布为 $\pi(\theta,\sigma^2)\propto\frac{1}{\sigma^2}$ 的情形

2.关于指数分布族的若干结论

3.取先验分布为共轭先验分布的情形

1.取先验分布为 $\pi(\theta,\sigma^2)\propto\frac{1}{\sigma^2}$ 的情形

设总体的模型为正态分布，即 $X(\theta,\sigma^2)\sim N(\theta,\sigma^2)$

今有 $\mathbf{x}=(x_1,\cdots,x_n)$ 为样本容量为 n 的 iid 样本，贝叶斯统计认为未知参数(θ,σ^2)为一个 2 维随机变量，若取(θ,σ^2)的先验密度为 $\pi(\theta,\sigma^2)\propto\frac{1}{\sigma^2}$，则

(1) i) $(\theta,\sigma^2)|\mathbf{x}$ 的后验密度 $\pi(\theta,\sigma^2\mid\mathbf{x})$ 为：

$$\pi(\theta,\sigma^2\mid\mathbf{x})\propto(\sigma^2)^{-\frac{\gamma+1}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left(S+n(\bar{x}-\theta)^2\right)\right\}$$

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

ii) θ 的边际后验 $\theta|\mathrm{x}$，有：$t=\frac{\theta-\bar{x}}{s/\sqrt{n}}\sim t(\gamma)$

其中 $s^2=\frac{1}{n-1}S$，$t(\gamma)$是自由度为 γ 的 t 分布。

(2)方差$\boldsymbol{\sigma}^2$ 的后验边际分布：

$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\sigma}^2 \mid \mathbf{x}) &= \int \boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{x}) d\boldsymbol{\theta} \\
&= \int_{-\infty}^{+\infty} \left(\boldsymbol{\sigma}^2\right)^{-\frac{\gamma+1}{2}-1} \exp\left\{-\frac{1}{2\boldsymbol{\sigma}^2}\left(\boldsymbol{S} + \boldsymbol{n}(\boldsymbol{\theta} - \bar{\boldsymbol{x}})^2\right)\right\} d\boldsymbol{\theta}
\end{aligned}$$

$$\propto \left(\boldsymbol{\sigma}^2\right)^{-\frac{\gamma}{2}-1} \exp\left\{-\frac{\boldsymbol{S}}{2\boldsymbol{\sigma}^2}\right\} \int_{-\infty}^{+\infty} \left(\frac{2\boldsymbol{\pi}\boldsymbol{\sigma}^2}{\boldsymbol{n}}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\boldsymbol{\sigma}^2} \cdot \boldsymbol{n}(\boldsymbol{\theta} - \bar{\boldsymbol{x}})^2\right\} d\boldsymbol{\theta}$$

$$= \left(\boldsymbol{\sigma}^2\right)^{-\frac{\gamma}{2}-1} \exp\left\{-\frac{\boldsymbol{S}}{2\boldsymbol{\sigma}^2}\right\}$$

(3)均值θ 在给定σ^2 之后的后验条件密度

由 $$\pi(\theta,\sigma^2 \mid \mathbf{x}) = \pi(\sigma^2 \mid \mathbf{x}) \cdot \pi(\theta \mid \sigma^2, \mathbf{x})$$

及 $$\pi(\theta,\sigma^2 \mid \mathbf{x}) \propto (\sigma^2)^{-\frac{\gamma+1}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\left(S + n(\bar{x}-\theta)^2\right)\right\}$$

$$\pi(\sigma^2 \mid \mathbf{x}) = \left(\sigma^2\right)^{-\frac{\gamma}{2}-1} \exp\left\{-\frac{S}{2\sigma^2}\right\}$$

所以 $$\pi(\theta \mid \sigma^2, \mathbf{x}) \propto \left(\sigma^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\right\}$$

即 $$\theta \mid \sigma^2, \mathbf{x} \sim N(\bar{x}, \sigma^2 / n)$$

back

3.取先验分布为共轭先验分布的情形

(1)求 (θ,σ^2) 的共轭先验密度

(2)求 θ 的后验边际密度

(3)求给定 $\boldsymbol{\sigma}^2$ 后 θ 的条件后验密度函数 $\pi(\theta|\sigma^2,\mathbf{x})$

例题

例 有一实验站关于生长小麦的经验为每块样地的均值和标准差分别为100及10的正态分布，现在他们研究施加激素的影响。在12块地施加激素后所得产量如下(单位：千克)：

141,102,73,171,137,91,81,157,146,69,121,134

关于方差的信息是均值、标准差分别约为300及160；

关于均值的信息是均值约为110，约为15即相当于观测了15个观测值。

求: (1) $(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$ 的共轭先验；

(2) $(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$ 的后验密度函数；

(3) $\boldsymbol{\mu}$ 的边际后验；

(4) $\boldsymbol{\mu}$ 对 $\boldsymbol{\sigma}^2$ 已知情况下的条件后验密度函数。

back

§1.6 充分统计量

一、经典统计中充分统计量的回顾

充分性是数理统计中最重要的概念之一，也是数理统计这一学科特有的基本概念之一。它是Fisher在1925年提出的。

充分性的直观定义：不损失信息的统计量。

引例：研究某个运动员的打靶命中率θ，我们对该运动员进行10次测试，发现除第三、六次没有命中外，其余8次都命中，这样的结果包含了哪些信息?

（1）打靶10次命中8次；

（2）2次不命中分别出现在第3次和第6次打靶上。

概率分析：

定义：设$\mathbf{x}=(x_1,\cdots,x_n)$是来自分布函数$F(x|\theta)$的一个样本，$T=T(x)$是统计量，假如在给定$T(x)=t$的条件下，x的条件分布与θ无关的话，则称该统计量为θ的充分统计量。

充分统计量的一个重要特性：当得到充分统计量T的某个取值t之后，而失去原样本的观察值也没有关系。因为我们可以根据上述的条件分布来构造某个随机试验，从中获得来自总体的一个新样本，这个新样本虽不能完全恢复老样本的原状，但它与老样本所含的有关参数θ的信息是一样的。

例题1 设总体为二点分布$b(1,\theta)$，$x_1,\cdots,x_n$ 为样本，令$T=\sum_{i=1}^{n}x_i$ 求在给定T的取值后，X的条件分布。

例题 2 设$x_1,\cdots,x_n$是来自$N(\mu,1)$的样本，验证：$T=\bar{x}$是μ的充分统计量。

因子分解定理：一个统计量T(x)对参数θ是充分的充要条件是：存在一个t与θ的函数g(t,θ)和一个样本x的函数h(x)，使得对任一样本x和任意θ，样本的联合密度p(x|θ)可表示为它们的乘积，即：

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

这个定理表明：假如存在充分统计量T(x)，则样本分布p(x| θ)一定可以分解为两个因子的乘积：一个是与θ无关，仅与样本x有关；另一个是可以与θ有关，但与样本x的关系仅仅通过充分统计量T(x)表现出来。

二、贝叶斯统计中充分统计量的有关结论及应用

贝叶斯统计中充分统计量与经典统计中充分统计量的概念是一致的。

定理1.1 设 $\mathbf{x}=(x_1,\cdots,x_n)$是来自密度函数p(x|θ)的一个样本，T=T(x)是统计量，它的密度函数为p(t|θ),又设H={π(θ)}是θ的某个先验分布族，则T(x)为θ的充分统计量的充要条件是对任一先验分布π(θ)∈H，有：

$$\pi(\theta \mid T(x)) = \pi(\theta \mid x)$$

即用样本分布p(x|θ)算得的后验分布与用统计量T(x)算得的后验分布是相同的。

例题1.14 验证定理1.1及其它的含义

$$p(\boldsymbol{x}|\mu,\sigma^2)=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$=(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\right\}$$

其中

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

设 $\pi(\mu,\sigma^2)$ 是任一个先验分布，则 μ,σ^2 的后验密度为

$$\pi(\mu,\sigma^2|\boldsymbol{x})=\frac{\sigma^{-n}\pi(\mu,\sigma^2)\exp\left\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\right\}}{\int_{-\infty}^{\infty}\int_{0}^{\infty}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}[Q+n(\bar{x}-\mu)^2]\right\}d\mu d\sigma^2}$$

另一方面，二维统计量 $T=(\bar{x},Q)$ 恰好是 (μ,σ^2) 的充分统计量，大家知道，$\bar{x}\sim N(\mu,\sigma^2/n)$，$Q/\sigma^2\sim\chi^2(n-1)$，由此可分别写出 $\bar{x}$ 与 Q 的分布

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

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

由于 $\bar{x}$ 与 Q 独立，所以 $\bar{x}$ 与 Q 的联合密度容易写出

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

利用相同的先验分布 $\pi(\mu,\sigma^2)$，可得在给定 $\bar{x}$ 和 Q 下的后验分布

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

比较这二个后验密度,可得

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

由此可见,用充分统计量$(\bar{x},Q)$的分布算得的后验分布与用样本

分布算得的后验分布是相同的。

- 关于定理1.1的两点说明:
- 1.定理1.1所给出的条件是充分必要的，所以定理1.1的充要条件可以作为充分统计量的贝叶斯定义。
- 2.假如已知统计量T(x)是充分的，那么按定理1.1，其后验分布可用该统计量的分布算得，由于充分统计量可以可简化数据、降低维数，故定理1.1亦可用来简化后验分布的计算。

例1.15　用充分统计量计算正态分布$N(\theta,1)$中参数θ的后验分布。（自学）

补充内容：指数分布族及其先验分布

(1)单参数指数分布族的定义

设总体X或X|θ的分布密度$p(x|\theta)$为：

$$p(x|\theta)=g(x)h(\theta)\exp\{t(x)\varphi(\theta)\}$$

其中g,h,t,φ为一般的函数记号

则称$p(x|\theta)$属于指数分布族。

如，正态分布$N(\mu,\sigma^2)$当σ^2已知时，由

$$p(x|\mu)=\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$=\left[\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right]\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}\exp\left\{\frac{x\mu}{\sigma^2}\right\}$$

因此它属于指数分布族。

显然，$N(\mu,\sigma^2)$当 μ 已知，σ^2未知时，也是单参数σ^2的指数分布族。

又如，泊松分布也属于指数分布族，因为：

$$p(x\mid\lambda)=\frac{1}{x!}\exp\{-\lambda\}\exp\{(\ln\lambda)x\}$$

另外，二项分布也属于指数分布族，因为：

$$p(x\mid\pi)=C_n^x(1-\pi)^n\exp\left\{x\ln\frac{\pi}{1-\pi}\right\}$$

类似的还有gamma分布，beta分布。

Be(α,β)中α,β一个已知，一个未知时都属于指数分布族。因此，指数分布族是一类非常广泛的分布族。

(2)指数分布族中的参数θ的共轭先验密度

若X|θ的分布属于指数分布族:

$$p(x\mid\theta)=g(x)h(\theta)\exp\left\{t(x)\varphi(\theta)\right\}$$

故似然函数$l(\theta\mid\mathbf{x})$(其中$\mathbf{x}=(x_1,\cdots,x_n)$为i.i.d.的样本)为

$$l(\theta\mid\mathbf{x})\propto[h(\theta)]^n\exp\left\{\sum t(x_i)\varphi(\theta)\right\}$$

则共轭先验密度$\pi(\theta)$为: $\pi(\theta)\propto[h(\theta)]^\gamma\exp\left\{\tau\varphi(\theta)\right\}$

即指数分布族的共轭族Π为:

$$\Pi=[h(\theta)]^\gamma\exp\left\{\tau\varphi(\theta)\right\}$$

(3)两参数指数分布族及其共轭先验族

设$X \sim p(x \mid \theta,\phi);\theta,\phi$皆未知，若：

$$p(x \mid \theta,\phi) = g(x)h(\theta,\phi)\exp\left\{t(x)\varphi(\theta,\phi) + u(x)\chi(\theta,\phi)\right\}$$

则称X的分布属于两参数指数族。

若 $\mathbf{x} = (x_1,\cdots,x_n)$为i.i.d.的样本，X的分布为两参数的指数分布族，则似然函数 $l(\theta,\phi \mid \mathbf{x})$ 可写为：

$$l(\theta,\phi \mid \mathbf{x}) \propto [h(\theta,\phi)]^n \exp\left\{\sum t(x_i)\varphi(\theta,\phi) + \sum u(x_i)\chi(\theta,\phi)\right\}$$

共轭先验族 Π 为以下形式：

$$\pi(\theta,\phi) \propto [h(\theta,\phi)]^{\gamma} \exp\left\{\alpha\varphi(\theta,\phi) + \beta\chi(\theta,\phi)\right\}$$

(4)两参数 θ,ϕ 的后验及先验的一种求法

设 $\pi(\theta,\phi\,|\,\mathbf{x})$为已知i.i.d.的样本x的关于 θ,ϕ 的后验，

X的模型为 $p(x\,|\,\theta,\phi)$ ， θ,ϕ皆未知，则：

$$\pi(\theta,\phi\,|\,\mathbf{x})=\pi(\phi\,|\,\mathbf{x})\cdot\pi(\theta\,|\,\phi,\mathbf{x})$$

相应的，也有类似的先验公式，设$\pi(\theta,\phi)$ 为 θ,ϕ的先验密度，则：

$$\pi(\theta,\phi)=\pi(\phi)\cdot\pi(\theta\,|\,\phi)$$

back