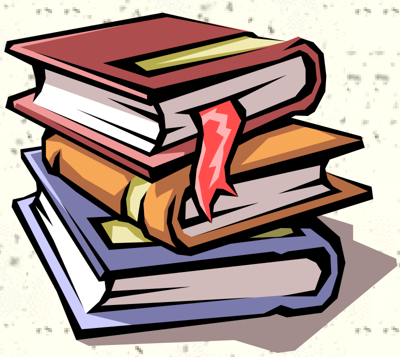
第二章 贝叶斯推断



第二章　贝叶斯推断

§2.1　条件方法

1.后验分布的特点：未知参数的后验分布是集三种信息（总体、样本和后验）于一身，它包含了所有可供利用的信息。故有关的参数估计和假设检验等统计推断都按一定方式从后验分布提取信息，其提取方法与经典统计推断相比要简单明确得多。

2.条件方法的基本思想：基于后验分布的统计推断实际上只考虑已出现的数据（样本观察值）而认为未出现的数据与推断无关，这一重要的观点被称为“条件观点”，基于这种观点提出的统计方法被称为条件方法。

3.条件方法与频率方法的区别：(以对估计的无偏性认识为例)

例如经典统计学认为参数的无偏估计应满足：

$$E\hat{\theta}(\boldsymbol{x})=\int_{\boldsymbol{x}}\hat{\theta}(x)\,p(\boldsymbol{x}\mid\theta)\,d\boldsymbol{x}=\theta$$

其中平均是对样本空间中所有可能出现的样本而求的，可实际中样本空间中绝大多数样本尚为出现过，而多数从未出现的样本也要参与平均是实际工作者难以理解的。故在贝叶斯推断中不用无偏性，而条件方法是容易被实际工作者理解和接受的。

例2.1.1

例 2.1.1 (Berger(1985))假设要对某物质进行分析,可以送纽约的实验室,也可送加利福尼亚的实验室,这两个实验室大体上一样好,故用抛硬币的方法来决定,出现"正面"送纽约,出现"反面"送加利福尼亚,抛的结果为反面,故送加利福尼亚。过一段时间,试验结果送回来之后,就要下结论和写报告,那么,结论要不要考虑硬币还要出现正面的情况呢?常识会坚决地认为:不必考虑,即只与实际进行了的试验有关,而频率派的观点认为是需要考虑的,即要对所有可能的数据,包括纽约实验室可能会得到的各种结果来求平均,这一观点常使实际工作者难以接受,关于条件方法与频率方法之间的差异及其影响将在以后诸节中逐步涉及,现转入贝叶斯推断的一些方法的讨论。

§2.2 估计

1. 贝叶斯估计

定义2.1 使后验密度$\pi(\theta|\boldsymbol{x})$达到最大的值$\theta_{MD}$ 称为最大后验估计；后验分布的中位数 $\hat{\theta}_{Me}$称为后验中位数估计；后验分布的期望值 $\hat{\theta}_E$称为 θ 的后验期望值估计，这三个估计都称为贝叶斯估计，记为 $\hat{\theta}_B$ 。

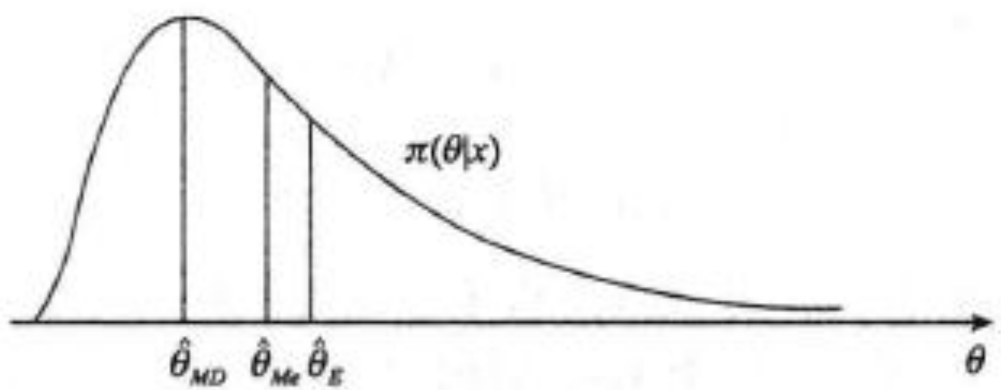


图 2.2.1 θ的三种贝叶斯估计

例 2.2 设 $x_1,\cdots,x_n$ 是来自正态总体 $N(\theta,\sigma^2)$ 的一个样本，其中 σ^2 已知，若取 θ 的共轭先验分布 $N(\mu,\tau^2)$ 作为 θ 的先验分布，其中 μ 与 τ^2 已知，求 θ 的 Bayes 估计。

解题的基本步骤：

1. 计算 θ 的后验分布：由例 1.6 知 $\theta \sim N(\mu_1,\sigma_1^2)$

2. 分析后验分布的特征：对称分布

3. 求 θ 的 Bayes 估计：

$$\hat{\theta}_B = \mu_1 = \frac{\tau^{-2}}{\sigma_0^{-2}+\tau^{-2}}\cdot\mu + \frac{\sigma_0^{-2}}{\sigma_0^{-2}+\tau^{-2}}\cdot\bar{x} = \frac{\sigma_0^2\mu+\tau^2\bar{x}}{\sigma_0^2+\tau^2}$$

作为一个数值例子，我们考虑对一个儿童做智力测验，设测验结果 $X\sim N(\theta,100)$，其中 θ 在心理学中定义为该儿童的智商，根据过去多次测验，可设 $\theta\sim N(100,225)$，应用上述方法，在 $n=1$ 时，可得在给定 $X=x$ 条件下，该儿童智商 θ 的后验分布是正态分布 $N(\mu_1,\sigma_1^2)$，其中：

$$\mu_1=\frac{100\times100+225x}{100+225}=\frac{400+9x}{13}$$

$$\sigma_1^2=\frac{100\times225}{100+225}=\frac{900}{13}=69.23=(8.32)^2$$

假如该儿童这次测验得分为 115 分，则他的智商的贝叶斯估计为

$$\hat{\theta}_B=\frac{400+9\times115}{13}=110.38$$

例2.3 为估计不合格率 θ，今从一批产品中随机抽取 n 件，其中不合格品数X服从 $B(n, p)$，一般选取 $Be(\alpha, \beta)$为 θ 的先验分布，设 α, β 已知，求 θ的Bayes估计。

解：由共轭先验分布可知，θ 的后验分布为：

$$Be(\alpha + x, \beta + n - x)$$

则得：

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

特例：选用贝叶斯假设作为先验分布，即 $\alpha = \beta = 1$

则：

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

注 意:

第一、在二项分布时，θ 的最大后验估计就是经典统计中的极大似然估计，即 θ 的极大似然估计就是取特定的先验分布下的贝叶斯估计。

第二、θ 的后验期望值估计 $\hat{\theta}_E$ 要比最大后验估计 $\hat{\theta}_{MD}$ 更合适一些。

表2.1 不合格率 θ 的二种贝叶斯估计的比较

试验号	样本量n	不合格数x	$\hat{\theta}_{MD}=\frac{x}{n}$	$\hat{\theta}_{E}=\frac{x+1}{n+2}$
1	3	0	0	0.200
2	10	0	0	0.083
3	3	3	1	0.800
4	10	10	1	0.917

例2.4 设x 是来自如下指数分布的一个观察值。

$$p(x|\theta)=e^{-(x-\theta)}, \quad x\geq\theta$$

又取柯西分布作为θ的先验分布，即：$\pi(\theta)=\dfrac{1}{\pi(1+\theta^2)}, \quad -\infty<\theta<\infty$

求θ的最大后验估计$\hat{\theta}_{MD}$。

解：由前面方法可求出 θ的后验密度：

$$\pi(\theta|x)=\frac{e^{-(x-\theta)}}{m(x)(1+\theta^2)\pi}, \quad \theta\leq x$$

为了寻找θ的最大后验估计$\hat{\theta}_{MD}$，对后验密度求导数，得：

$$\frac{d}{d\theta}\pi(\theta|x)=\frac{e^{-x}}{m(x)\pi}\left[\frac{e^{\theta}}{1+\theta^2}-\frac{2\theta e^{\theta}}{(1+\theta^2)^2}\right]=\frac{e^{-x}e^{\theta}(\theta-1)^2}{m(x)(1+\theta^2)^2\pi}\geq 0$$

由于$\pi(\theta|x)$的非减性，考虑到θ的取值不能超过x，故θ的最大后验估计应为$\hat{\theta}_{MD}=x$

2. 贝叶斯估计的误差

设 $\hat{\theta}$ 是 θ 的一个贝叶斯估计,在样本$\boldsymbol{x}$给定后，$\hat{\theta}$ 是一个数，在综合各种信息后，θ 是按 $\pi(\theta|\boldsymbol{x})$取值，所以评价一个贝叶斯估计的误差的最好而又简单的方式是用θ 对$\hat{\theta}$ 的后验均方差或平方根来度量，定义如下:

定义2.2 设参数θ 的后验分布为$\pi(\theta|\boldsymbol{x})$,贝叶斯估计为 $\hat{\theta}$，则 $(\theta-\hat{\theta})^2$的后验期望

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

称为 $\hat{\theta}$ 的后验均方差，而其平方根称为后验标准误。

注意：

(1) $\boldsymbol{E}^{\theta|x}$ 表示用条件分布 $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{x})$求期望；

(2) 当$\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_E=\boldsymbol{E}(\boldsymbol{\theta}|\boldsymbol{x})$ 时， 则$MSE(\hat{\theta}_E\big|\boldsymbol{x})=E^{\theta|x}(\theta-\hat{\theta}_E)^2=Var(\theta|\boldsymbol{x})$

称为后验方差，其平方根称为后验标准差；

(3) 后验均方差与后验方差的关系:

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

这表明，当$\hat{\theta}=\hat{\theta}_E$时，可使后验均方差达到最小，实际中常取后验均值作为θ 的贝叶斯估计值；

(4)与经典统计的两点比较：①后验方差应用的方便程度不一样；②计算的复杂程度不一样。

例2.5 设一批产品的不合格率为θ ,检查是一个一个进行,直到发现第一个不合格品为止,若X为发现第一个不合格品时已检查的产品数,则X服从几何分布,其分布列为:

$$P(X=x|\theta)=\theta(1-\theta)^{x-1}, x=1,2,\cdots$$

设 θ 的先验分布为 $P(\theta=\frac{i}{4})=\frac{1}{3}, i=1,2,3$ ，如今只获得一个样本观察值$X=3$，求 θ 的最大后验估计，后验期望估计，并计算它的误差。

解：(1)先求联合分布。因为已知θ的先验分布和在 θ 给定下，$X=3$ 的条件概率，则联合分布为：

$$\boldsymbol{P}(\boldsymbol{X}=3,\boldsymbol{\theta}=\frac{\boldsymbol{i}}{4})=\frac{1}{3}\cdot\frac{\boldsymbol{i}}{4}\cdot(1-\frac{\boldsymbol{i}}{4})^2$$

$X=3$的无条件概率为(利用全概率公式)

$$P(X=3)=\frac{1}{3}[\frac{1}{4}(\frac{3}{4})^2+\frac{2}{4}(\frac{2}{4})^2+\frac{3}{4}(\frac{1}{4})^2]=\frac{5}{48}$$

再求θ的后验分布列为：

$$P(\theta=i/4|X=3)=\frac{P(X=3,\theta=i/4)}{P(X=3)}=\frac{4i}{5}(1-\frac{i}{4})^2, i=1,2,3$$

或

θ	$1/4$	$2/4$	$3/4$
$P(\theta=i/4\vert X=3)$	$9/20$	$8/20$	$3/20$

最后得 θ 的最大后验估计：$\hat{\theta}_{MD}=1/4$

(2) $\hat{\theta}_E=E(\theta|X=3)=17/40$

(3) 因为 $E(\theta^2|X=3)=17/80$ ，所以：

$$\mathrm{Var}(\theta|x)=E(\theta^2|x)-E^2(\theta|x)=17/80-(17/40)^2=51/1600$$

$\hat{\theta}_{MD}$ 的后验均方差为

$$MSE(\hat{\theta}|x)=\mathrm{Var}(\theta|x)+(\hat{\theta}_E-\hat{\theta})^2=\frac{51}{1600}+(\frac{1}{4}-\frac{17}{40})^2=\frac{1}{16}$$

例2.6 在例2.3中，在选用共轭分布下，不合格品率θ的后验分布为贝塔分布，它的后验方差为：

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

其中n为样本量，x为样本中不合格品数，α 与β为先验分布中的两个超参数。若取$\alpha=\beta=1$ ，则其后验方差为：

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

这时θ的后验期望估计和最大后验估计$\hat{\theta}_{MD}$分别为：

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

显然，$\hat{\theta}_E$ 的后验均方差 $\mathrm{Var}(\theta\,|\,x)$ 就是上述 $\hat{\theta}_{MD}$ 的后验均方差为：

$$MSE(\hat{\theta}_{MD}\,|\,x)=\frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}+\left(\frac{x+1}{n+2}-\frac{x}{n}\right)^2$$

对若干对 (n,x) 的值算得的后验方差和后验均方差列入表2.2中。

表2.2 $\hat{\theta}_E$ 和 $\hat{\theta}_{MD}$ 的后验均方差

n	x	$\hat{\theta}_E$	Var	$\sqrt{Var}$	$\hat{\theta}_{MD}$	MSE	$\sqrt{MSE}$
3	0	1/5	0.02667	0.16	0	0.06667	0.26
10	0	1/12	0.00588	0.08	0	0.01282	0.11
10	1	2/12	0.01068	0.10	1/10	0.01512	0.12
20	1	2/22	0.00359	0.06	1/20	0.00527	0.07

§2.3 区间估计(可信区间)

一、可信区间

定义 2.3 参数θ的后验分布为$\pi(\theta|\boldsymbol{x})$，对给定的样本$\boldsymbol{x}$和概率$1-\alpha(0<\alpha<1)$，若存在这样的二个统计量$\hat{\theta}_L=\hat{\theta}_L(\boldsymbol{x})$与$\hat{\theta}_U=\hat{\theta}_U(\boldsymbol{x})$，使得：$P(\hat{\theta}_L\le\theta\le\hat{\theta}_U|\boldsymbol{x})\ge1-\alpha$

则称区间$[\hat{\theta}_L,\hat{\theta}_U]$为参数$\theta$的可信水平为 **1-α** 贝叶斯可信区间，或简称为θ的 **1-α** 可信区间。

满足$P(\theta\ge\hat{\theta}_L|\boldsymbol{x})\ge1-\alpha$的$\hat{\theta}_L$称为$\theta$的 **1-α**（单侧）可信下限；

满足$P(\theta\le\hat{\theta}_U|\boldsymbol{x})\ge1-\alpha$的$\hat{\theta}_U$称为$\theta$的 **1-α**（单侧）可信上限。

这里的可信水平和可信区间与经典统计中的置信水平与置信区间虽是同类的概念，但两者还是有本质的差别，主要表现在下面二点:

1.在条件方法下,对给定的样本***x***和可信水平1-α，通过后验分布可求得具体的可信区间，譬如，θ的可信水平为0.9的可信区间是[1.5，2.6]，这时我们可以写出

$$P(1.5 \le \theta \le 2.6 | \boldsymbol{x}) = 0.9$$

2.在经典统计中寻求置信区间有时是困难的，因为它要设法构造一个枢轴量（含有被估计参数的随机变量），使它的分布不含未知参数，这是一项技术性很强的工作。相比之下可信区间只要利用后验分布，不需要再去寻求另外的分布，可信区间的寻求要简单得多。

例2.7 设 $x_1, x_2, \cdots, x_n$ 是来自正态总体 $N(\theta, \sigma^2)$ 的一个样本观察值，其中 σ^2 已知,若正态均值的先验分布取为 $N(\mu, \tau^2)$，其中 μ 与 τ 已知，则可求得 θ 的后验分布为 $N(\mu_1, \sigma_1^2)$ ，由此很容易获得 θ 的 $1-\alpha$ 可信区间 :$P(\mu_1 - \sigma_1 \mu_{1-\alpha/2} \le \theta \le \mu_1 + \sigma_1 \mu_{1-\alpha/2}) = 1-\alpha$

其中 $\mu_{1-\alpha/2}$ 是标准正态分布 $1-\alpha/2$ 的分位数。

例2.8 80年代我国彩电平均寿命的贝叶斯估计。

经过早期筛选后的彩色电视机的寿命服从指数分布，它的密度函数为：

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

其中 $\theta>0$ 是彩电的平均寿命。

现从一批彩电中随机抽取n台进行寿命试验，试验到第 $r(r\le n)$ 台失效为止，其失效时间为 $t_1\le t_2\le\cdots\le t_r$ ，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：

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

其中 $F(t)$ 为彩电的寿命的分布函数，$s_r=t_1+\cdots+t_r+(n-r)t_r$ 称为总试验时间。

具体实施的步骤：

(1)确定参数 θ 的先验分布：倒伽玛分布 $IGa(\alpha,\beta)$

(2)利用历史资料确定两个超参数 α 和 β 的值

（用第三种方法）

(3)求出 θ 的后验分布：$IGa(\alpha+r,\beta+S_r)$

(4)用后验均值作为 θ 的贝叶斯估计：

$$\hat{\theta}=E(\theta\,|\,\boldsymbol{t})=\frac{\beta+S_r}{\alpha+r-1}$$

(5)可信下限的确定

二、最大后验密度（HPD）可信区间（Highest Posterior Density）

- **定义2.4** 设参数θ的后验密度为$\pi(\theta \mid \boldsymbol{x})$，对给定的概率$1-\alpha (0<\alpha<1)$，若在直线上存在这样一个子集$C$，满足下列二个条件：
- ① $P(C \mid \boldsymbol{x}) = 1-\alpha$
- ② 对任给$\theta_1 \in C$和$\theta_2 \notin C$，总有$\pi(\theta_1 \mid \boldsymbol{x}) \geq \pi(\theta_2 \mid \boldsymbol{x})$，则称$C$是$\theta$的可信水平为$1-\alpha$的最大后验密度可信集，简称$(1-\alpha)$ **HPD可信集**；
- 如果C是一个区间，则C又称为$(1-\alpha)$**HPD可信区间**。

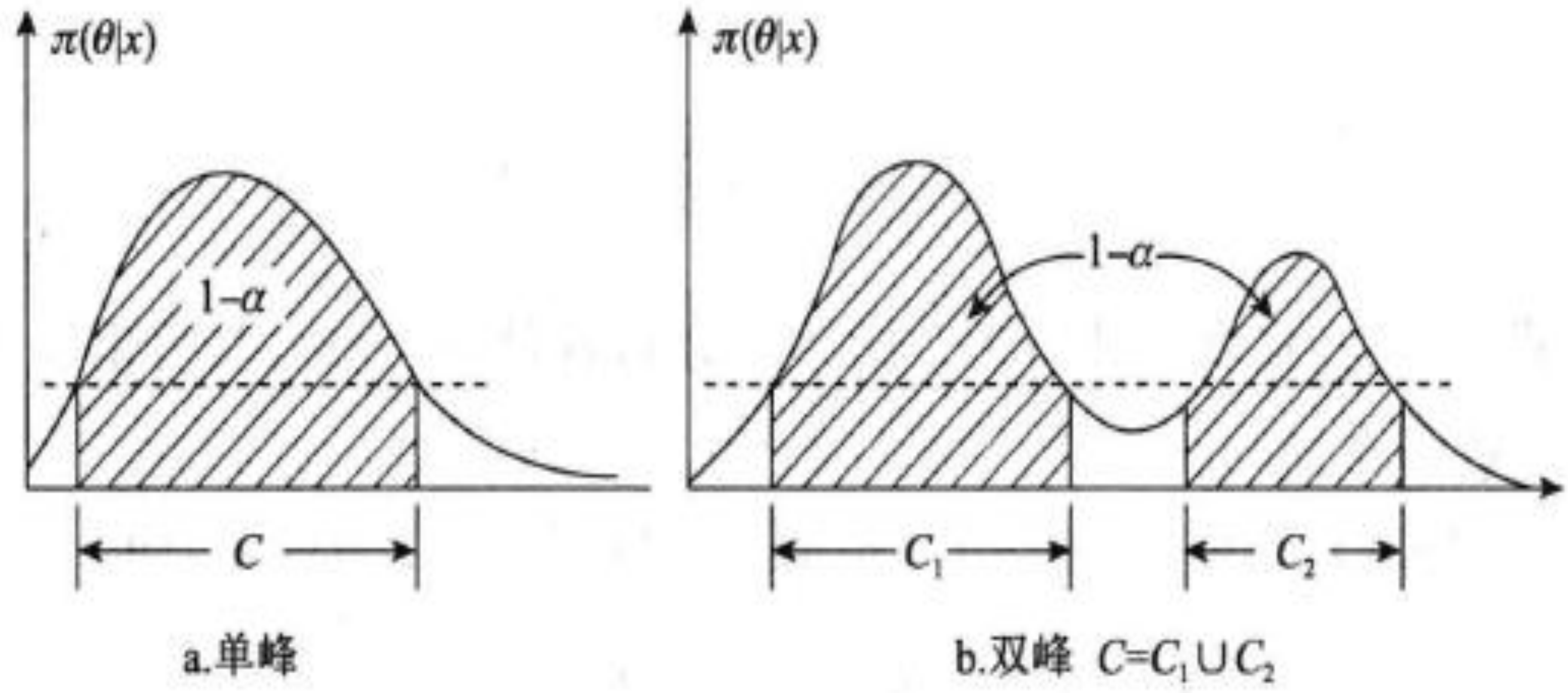


图 2.3.1　HPD 可信区间与 HPD 可信集

注：后验密度函数$\pi(\theta|\mathbf{x})$是θ的单峰连续函数时，获得θ的$(1-\alpha)$ HPD可信区间的数值计算方法：

- **Step1**：对给定的k，建立子程序；解方程$\pi(\theta|\mathbf{x})=k$,得解$\theta_1(k)$和$\theta_2(k)$,从而组成一个区间:

$$C(k)=[\theta_1(k),\theta_2(k)]=\{\theta:\pi(\theta|\mathbf{x})\ge k\}$$

- **Step2**：建立第二个子程序，用来计算概率：

$$P(\theta\in C(k)|\mathbf{x})=\int_{C(k)}\pi(\theta|\mathbf{x})d\theta$$

- **Step3**：对给定的k，若$P(\theta\in C(k)|\mathbf{x})\approx 1-\alpha$，则$C(k)$即为所求的HPD可信区间。
- 若$P(\theta\in C(k)|\mathbf{x})>1-\alpha$，则增大$k$，再转入Step1与Step2。
- 若$P(\theta\in C(k)|\mathbf{x})<1-\alpha$，则减小$k$，再转入Step1与Step2。

例2.9 在例2.8中已经确定彩电平均寿命θ的后验分布为倒伽玛分布$IGa(1.956, 42868)$，现求θ的可信水平为0.90的最大后验密度（HPD）可信区间。

- 解题的基本步骤：
- 1.确定参数θ的后验密度和分布函数：
- 后验密度：$\pi(\theta \mid \boldsymbol{x}) = \beta^2 \theta^{-3} e^{-\beta/\theta}, \quad \theta > 0$
- 分布函数：$F(\theta \mid \boldsymbol{t}) = \left(1 + \frac{\beta}{\theta}\right) e^{-\beta/\theta}, \quad \theta > 0$
- 2.确定初始值
- 3.按第一步计算初始区间
- 4.计算后验概率
- 5.验证初始区间是否满足要求，满足则停止，否则继续。

为简单起见，这里的 1.956 用近似数 2 代替，于是 θ 的后验密度为

$$\pi(\theta|x)=\beta^2\theta^{-3}e^{-\beta/\theta},\theta>0$$

其中 $\beta=42868$，它的分布函数为

$$F(\theta|t)=\left(1+\frac{\beta}{\theta}\right)e^{-\beta/\theta},\theta>0$$

这将为计算可信区间的后验概率提供方便。

另外，此后验密度是单峰函数，其众数 $\theta_{MD}=\beta/3=14289$，这就告诉我们，$\theta$ 的 HPD 可信区间的二个端点分别在此众数两侧，在这一点上的后验密度函数值为

$$\pi(\theta_{MD}|t)=\beta^2\left(\frac{3}{\beta}\right)^3e^{-3}=0.000031358$$

这个数过小，对计算不利，在以下计算中我们用 $\beta\pi(\theta|t)$来代替 $\pi(\theta|x)$，这并不会影响我们寻求 HPD 可信区间，其中

$$\beta\pi(\theta|t)=\left(\frac{\beta}{\theta}\right)^3\exp\left(-\frac{\beta}{\theta}\right)$$

我们按寻求 HPD 可信区间的程序 $1^\circ \sim 3^\circ$ 进行，经过四轮计算就获得 θ 的 0.90 的 HPD 可信区间(4735,81189)，即

$$P(4375 \leqslant \theta \leqslant 81189 \mid \boldsymbol{t}) = 0.90$$

具体计算如下：在第一轮，我们先取 $\theta_u^{(1)} = 42868$（由于它大于众数 θ_{MD}，故它是上限），代入 $\beta\pi(\theta \mid \boldsymbol{t})$，算得

$$\beta\pi(\theta_u^{(1)} \mid \boldsymbol{t}) = 0.367879$$

然后在计算机上搜索，发现当 $\theta_L^{(1)} = 6387$ 时，有

$$\beta\pi(\theta_L^{(1)} \mid \boldsymbol{t}) = 0.367867$$

这时可认为 $\beta\pi(\theta_u^{(1)} \mid \boldsymbol{t}) = \beta\pi(\theta_L^{(1)} \mid \boldsymbol{t}) = 0.3679$，$\theta$ 位于此区间的后验概率可由分布函数算出，即

$$\begin{aligned} P(\theta_L^{(1)} \leqslant \theta \leqslant \theta_u^{(1)} \mid \boldsymbol{t}) &= F(\theta_u^{(1)} \mid \boldsymbol{t}) - F(\theta_L^{(1)} \mid \boldsymbol{t}) \\ &= 0.73576 - 0.00938 = 0.72638 \end{aligned}$$

此概率比 0.90 要小，还需扩大区间。

在第二轮中，我们取 $\theta_u^{(2)}=85736$，这时

$$\beta\pi(\theta_u^{(2)}\mid t)=0.075816$$

然后在计算机上搜索，发现当 $\theta_L^{(2)}=4632$ 时，有

$$\beta\pi(\theta_L^{(2)}\mid t)=0.075811$$

可以认为 $\beta\pi(\theta_u^{(2)}\mid t)=\beta\pi(\theta_L^{(2)}\mid t)=0.0758$，而 θ 位于此区间的后验概率可类似算得。

$$P(\theta_L^{(2)}\leqslant\theta\leqslant\theta_u^{(2)}\mid t)=0.909800-0.000981=0.908819$$

此概率又比 0.90 大一点，还要缩小区间，接着进行第三轮、第四轮计算，最后获得 θ 的 0.90HPD 可信区间是(4735，81189)，全部搜索过程及中间结果列于表 2.3.1。

表2.3 可信区间的搜索过程

θ_0	β/θ_0	$\beta\pi(\theta_0)=(\beta/\theta_0)e^{-\lambda/\theta_0}$	$P(\theta\le\theta_0\mid t)=(1+\beta/\theta_0)e^{-\lambda/\theta_0}$	$P(\theta_L\le\theta\le\theta_U\mid t)$
$\theta_U^{(1)}=42868$ $\theta_L^{(1)}=6387$	1 6.71	0.367879 0.367765	0.735759 0.009383	0.726376
$\theta_U^{(2)}=85736$ $\theta_L^{(2)}=4632$	0.5 9.255	0.075816 0.075811	0.909800 0.000981	0.908819
$\theta_U^{(3)}=80883$ $\theta_L^{(3)}=4742$	0.53 9.039	0.087630 0.087654	0.900566 0.001191	0.898375
$\theta_U^{(4)}=81189$ $\theta_L^{(4)}=4735$	0.528 9.053	0.086815 0.086838	0.901189 0.001177	0.900012

§2.4 假设检验

一、假设检验

经典统计中处理假设检验问题的基本步骤：

1.建立原假设 H_0 与备择假设 H_1：

$$H_0:\theta\in\Theta_0 \leftrightarrow H_1:\theta\in\Theta_1$$

其中 Θ_0 与 Θ_1 是参数空间 Θ 中不相交的二个非空子集。

2.选择检验统计量 $T=T(\boldsymbol{x})$，使其在原假设 Θ_0 为真时概率分布是已知的。这是在经典方法中最困难的一步。

3.对给定的显著性水平 $\alpha(0<\alpha<1)$，确定拒绝域W，使犯第Ⅰ类错误（拒真错误）的概率不超过α。

4.当样本观察值 $\boldsymbol{x}$ 落入拒绝域W时，就拒绝原假设 H_0，接受备择假设 H_1；否则就保留原假设。

贝叶斯统计中处理假设检验问题的基本思想：

获得后验分布 $\pi(\theta \mid \boldsymbol{x})$后，先计算二个假设 H_0 和 H_1的后验概率：

$$\alpha_i = P(\Theta_i \mid \boldsymbol{x}), i = 0,1$$

然后比较α_0 与 α_1 的大小：

称 α_0/α_1 为后验概率比(或称后验机会比)：

当$\alpha_0/\alpha_1 > 1$时，接受 H_0；

当$\alpha_0/\alpha_1 < 1$时，接受 H_1；

当$\alpha_0/\alpha_1 \approx 1$时，不宜做判断，还需要进一步抽样或进一步收集先验信息。

由这两个学派假设检验的**基本思想**可看出贝叶斯假设检验更易理解更简单：

- 1.贝叶斯假设检验无需选择检验统计量，确定抽样分布；
- 2.无需事先给出显著性水平，确定其拒绝域；
- 3.易推广到多重假设检验的场合，检验的标准是：接受具有最大后验概率的假设。

例2.10 设 x 是从二项分布 $b(n,\theta)$ 中抽取的一个样本，现在考虑如下两个假设：

$$\Theta_0=\{\theta:\theta\le 1/2\},\qquad \Theta_1=\{\theta:\theta>1/2\}$$

若取均匀分布 $U(0,1)$ 作为 θ 的先验分布，试做出判断。

解： 因为Θ_0的后验概率为：

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

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\frac{(1/2)^{n+1}}{x+1}\left\{1+\frac{n-x}{x+2}+\frac{(n-x)(n-x-1)}{(x+2)(x+3)}+\cdots+\frac{(n-x)!\,x!}{(n+1)!}\right\}$$

在 n=5 时可计算各种 x 下的后验概率及后验机会比（见表2.4）

表2.4　θ的后验机会比

x	0	1	2	3	4	5
α_0	63/64	57/64	42/64	22/64	7/64	1/64
α_1	1/64	7/64	22/64	42/64	5764	63/64
α_0/α_1	63.0	8.14	1.91	0.52	0.12	0.016

从表中可以看出，当$x=0,1,2$时，应接受Θ_0，而在$x=3,4,5$时，应拒绝Θ_0，接受Θ_0。

二、一个重要的概念——贝叶斯因子

- **定义2.5**　设两个假设 Θ_0与 Θ_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 ，则称：

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

为贝叶斯因子。

- **说明：** 贝叶斯因子表示数据 $\boldsymbol{x}$ 支持原假设的程度。

三、简单假设对简单假设

$$\Theta_0=\{\theta_0\}\leftrightarrow\Theta_1=\{\theta_1\}$$

- 1.贝叶斯因子的计算方法及其含义。

在这种场合，两种简单假设的后验概率分别为：

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

其中$p(\mathbf{x}\,|\,\theta)$为样本的分布，这时后验机会比为：

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

- 如果要拒绝原假设 $\Theta_0=\{\theta_0\}$，则必须有 $\alpha_0/\alpha_1<1$，即：

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

即要求两密度函数值之比大于临界值，这正是著名的奈曼—皮尔逊引理的基本结果，从贝叶斯观点看，这个临界值就是两个先验概率比。

由此得到这种情形下的贝叶斯因子是：

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

它不依赖于先验分布，仅依赖于样本的似然比，这时贝叶斯因子的大小表示样本 $\boldsymbol{x}$ 支持 Θ_0 的程度。

2.例题分析（P54例2.11）

设$X \sim N(\theta,1)$，其中 θ 只有两种可能，非0即1，需要检验的假设是：

$$H_0:\theta=0 \leftrightarrow H_1:\theta=1$$

若从该总体中抽取一个容量为 n 的样本 $\boldsymbol{x}$，试计算贝叶斯因子及作出相应的决策。

- **分析**：先计算似然函数：
- $\theta=0$时，$p(\bar{x}\mid 0)=\sqrt{\dfrac{n}{2\pi}}\exp\{-\dfrac{n}{2}\bar{x}^2\}$

  $\theta=1$时，$p(\bar{x}\mid 1)=\sqrt{\dfrac{n}{2\pi}}\exp\{-\dfrac{n}{2}(\bar{x}-1)^2\}$
- 再计算贝叶斯因子：$B^{\pi}(\boldsymbol{x})=\dfrac{\alpha_0\pi_1}{\alpha_1\pi_0}=\exp\{-\dfrac{n}{2}(2\bar{x}-1)\}$
- 最后进行数值分析：假设$n=10$， $\bar{x}$ =2。则贝叶斯因子为：$B^{\pi}(\boldsymbol{x})=3.06\times10^{-7}$，这个数很小，所以应该拒绝$H_0$而接受$H_1$ 。

四、复杂假设 Θ_0 对复杂假设 Θ_1

- **1.贝叶斯因子的计算：**在这种情形下，贝叶斯因子不仅与样本有关，还依赖于参数空间 Θ 上的先验分布 $\pi(\theta)$。
- 先把先验分布 $\pi(\theta)$ 限制在 $\Theta_0 \cup \Theta_1$ 上，并令：

$$g_0(\theta) \propto \pi(\theta) I_{\Theta_0}(\theta) \qquad g_1(\theta) \propto \pi(\theta) I_{\Theta_1}(\theta)$$

于是先验分布可改写为：$\pi(\theta) = \pi_0 g_0(\theta) + \pi_1 g_1(\theta)$，$\theta \in \Theta_0 \cup \Theta_1$

$$= \begin{cases} \pi_0 g_0(\theta), & \theta \in \Theta_0 \\ \pi_1 g_1(\theta), & \theta \in \Theta_1 \end{cases}$$

其中 π_0 与 π_1 分别是 Θ_0 与 Θ_0 上的先验概率，g_0 与 g_1 分别是 Θ_0 与 Θ_1 上的概率密度函数，由此可计算出后验概率比为

$$\frac{\alpha_0}{\alpha_1} = \frac{\int_{\Theta_0} p(\boldsymbol{x}|\theta)\pi_0 g_0(\theta)d\theta}{\int_{\Theta_1} p(\boldsymbol{x}|\theta)\pi_1 g_1(\theta)d\theta}$$

则贝叶斯因子为：$B^{\pi}(\boldsymbol{x}) = \dfrac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = \dfrac{\int_{\Theta_0} p(\boldsymbol{x}|\theta) g_0(\theta)d\theta}{\int_{\Theta_1} p(\boldsymbol{x}|\theta) g_1(\theta)d\theta} = \dfrac{m_0(\boldsymbol{x})}{m_1(\boldsymbol{x})}$

2.结论分析： 由上式可看出，$B^\pi(\boldsymbol{x})$ 还依赖于 Θ_0 与 Θ_1 上的先验分布 g_0 与 g_1，这时贝叶斯因子虽已不是似然比，但仍可看作 Θ_0 与 Θ_1 上的加权似然比，它部分地消除了先验分布的影响，而强调了样本观察值的作用。

若设 $\hat{\theta}_0$ 与 $\hat{\theta}_1$ 分别是在 Θ_0 与 Θ_1 上的极大似然估计(MLE)，那么经典统计中所使用的似然比统计量

$$\lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \hat{\theta}_0)}{p(\boldsymbol{x} \mid \hat{\theta}_1)} = \frac{\sup_{\theta \in \Theta_0} p(\boldsymbol{x} \mid \theta)}{\sup_{\theta \in \Theta_1} p(\boldsymbol{x} \mid \theta)}$$

是贝叶斯因子 $B^\pi(\boldsymbol{x})$ 的特殊情况，即认为先验分布 $g_0(\theta)$ 与 $g_1(\theta)$ 的质量全部集中在 $\hat{\theta}_0$ 与 $\hat{\theta}_1$ 上。

3. **例题2.12** 设从正态总体$N(\theta,1)$中随机抽取一个容量为10的样本$\boldsymbol{x}$，算得样本均值$\bar{x}=1.5$，试对如下两个假设进行检验：

$$H_0:\theta\le 1 \leftrightarrow H_1:\theta>1$$

取θ的共轭先验分布为$N(0.5,2)$。

解：根据题意可算得θ的后验分布为$N(\mu_1,\sigma_1^2)$，其中μ_1=1.4523，σ_1^2=0.3086^2，即$\theta \sim N(1.4523,0.3086^2)$

由此可算得H_0与H_1的后验概率：

$$\alpha_0=P(\theta\le 1\,|\,\boldsymbol{x})=0.0708$$

$$\alpha_1=P(\theta>1\,|\,\boldsymbol{x})=1-\alpha_0=0.9292$$

后验机会比为：$\alpha_0/\alpha_1=0.0708/0.9292=0.0761$

可见，H_0为真的可能性较小，因此拒绝H_0，接受H_1，即认为正态均值大于1。

另外，由先验分布$N(0.5,2)$也可算得H_0与H_1的先验概率π_0=0.6368，π_1=0.3632，则先验机会比π_0/π_1=1.7533，即先验信息是支持原假设的。

再算两个机会比: $B^\pi(\boldsymbol{x}) = 0.0761/1.7533 = 0.0434$

即数据支持H_0的贝叶斯因子并不高。

两点讨论：1.样本均值对贝叶斯因子的影响(表2.4.2)；

2.先验均值对贝叶斯因子的影响(表2.4.3)。

结论：贝叶斯因子对样本信息变化的反应是灵敏的，而对先验信息变化的反应是迟钝的。

表 2.4.2　样本均值$\bar{x}$ 对贝叶斯因子的影响

$\bar{x}$	α_0	α_1	α_0/α_1	π_0/π_1	$B^{\pi}(x)$
1.5	0.0708	0.9292	0.0761	1.7533	0.0434
1.4	0.1230	0.8770	0.1403	1.7533	0.0800
1.3	0.1977	0.8023	0.2464	1.7533	0.1405
1.2	0.2946	0.7054	0.4176	1.7533	0.2382
1.1	0.4090	0.5910	0.6920	1.7533	0.3947
1.0	0.5319	0.4681	1.1363	1.7533	0.6481
0.9	0.6517	0.3483	1.8711	1.7533	1.0672
0.8	0.7549	0.2451	3.0800	1.7533	1.7567
0.7	0.8413	0.1587	5.3012	1.7533	3.0236
0.6	0.9049	0.0951	9.5152	1.7533	5.4271
0.5	0.9474	0.0526	18.0114	1.7533	10.2729

表 2.4.3　先验均值 $E(\theta)$ 对贝叶斯因子的影响

$E(\theta)$	α_0	$\alpha_0/(1-\alpha_0)$	π_0	$\pi_0/(1-\pi_0)$	$B^{\pi}(x)$
0.5	0.0708	0.0761	0.6368	1.7533	0.0434
0.6	0.0694	0.0746	0.6103	1.5782	0.0472
0.7	0.4668	0.0715	0.5832	1.3992	0.0511
0.8	0.0655	0.0701	0.5557	1.2507	0.0560
0.9	0.0630	0.0672	0.5279	1.1182	0.0601
1.0	0.0618	0.0658	0.5000	1.0000	0.0658
1.1	0.0594	0.0632	0.4721	0.8943	0.0707
1.2	0.0582	0.0618	0.4443	0.7996	0.0773
1.3	0.0559	0.0592	0.4168	0.7147	0.0828
1.4	0.0548	0.0580	0.3897	0.6336	0.0915
1.5	0.0526	0.0555	0.3632	0.5704	0.0973

五、简单原假设对复杂的备择假设

- 1.检验的基本问题
- 考察检验问题：$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta\neq\theta_0$

这是常见的一类检验问题，这里有一个对简单原假设的理解问题，当参数 θ 为连续量时，用简单假设作为原假设是不适当的，譬如，在 θ 是下雨的概率时，检验“明天下雨的概率是 0.7163891256…”是没有意义的，又如，在 θ 表示食品罐头的重量时，检验“午餐肉罐头重量是 250 克”也是不现实的，因为午餐肉罐头重量恰好是 250 克是罕见的，多数是在 250 克附近，所以在试验中接受丝毫不差的简单原假设“$\theta=\theta_0$”是不存在的，合理的原假设与备择假设应是

$$\mathrm{H}_0:\theta\in[\theta_0-\varepsilon,\theta+\varepsilon],\mathrm{H}_1:\theta\bar{\in}[\theta_0-\varepsilon,\theta_0+\varepsilon]$$

其中 ε 可选很小的数，使得 $[\theta_0-\varepsilon,\theta_0+\varepsilon]$ 与 $\theta=\theta_0$ 难以区别，譬如，ε 可选为 θ_0 的允许误差内的一个较小正数，当所选的 ε 较大时，那就不易用简单假设作为好的近似了。

对简单原假设$H_0:\theta=\theta_0$ 作贝叶斯检验时不能采用连续密函数作为先验分布，因为任何这种先验分布将给$\theta=\theta_0$ 的先验概率为零,从而后验概率也为零,所以一个有效的方法是对$\theta=\theta_0$ 给一个正概率π_0 ，而对 $\theta\neq\theta_0$ 给一个加权密度$\pi_1 g_1(\theta)$，

即θ 的先验密度为：

$$\pi(\theta)=\pi_0 \mathrm{I}_{\theta 0}(\theta)+\pi_1 g_1(\theta)$$

其中$\mathrm{I}_{\theta 0}(\theta)$为$\theta=\theta_0$ 的示性函数,$\pi_1=1-\pi_0$,$g_1(\theta)$ 为 $\theta\neq\theta_0$上的一个正常密度函数，这里可以把π_0 看作近似的实际假设

$$H_0:\theta\in[\theta_0-\varepsilon,\theta_0+\varepsilon]$$

上的先验概率，这样的先验分布是由离散和连续两部分组成。

由此得到贝叶斯因子为：

$$B^{\pi}(\boldsymbol{x})=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{p(\boldsymbol{x}\mid\theta_0)}{m_1(\boldsymbol{x})}$$

这一简单表达式要比后验概率的计算容易的多，故实际中常常是先计算贝叶斯因子$B^{\pi}(\boldsymbol{x})$，然后再计算后验概率 $\pi(\Theta_0\mid\boldsymbol{x})$：

$$\pi(\Theta_0\mid\boldsymbol{x})=\left[1+\frac{1-\pi_0}{\pi_0}\frac{1}{B^{\pi}(\boldsymbol{x})}\right]^{-1}$$

怎样推导？

例2.13 设 x 是从二项分布 $b(n,\theta)$ 中抽取的一个样本，现考察如下两个假设 $H_0:\theta=1/2 \leftrightarrow H_1:\theta\neq 1/2$ 。若设在 $\theta\neq 1/2$ 上的密度 $g_1(\theta)$ 为区间(0,1)上的均匀分布 $U(0,1)$，试做出恰当的判断。

解：由题意可求出 x 对 $g_1(\theta)$ 的边缘密度为：

$$m_1(x)=\int_0^1\binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

于是贝叶斯因子为：

$$B^\pi(x)=\frac{p(x\mid\theta_0)}{m_1(x)}=\frac{\left(\frac{1}{2}\right)^n(n+1)!}{x!(n-x)!}$$

则由上式可计算原假设$H_0: \theta = 1/2$的后验概率：

$$\pi(\theta_0 \mid x) = \left[1 + \frac{1-\pi_0}{\pi_0}\frac{1}{B^\pi(x)}\right]^{-1} = \left[1 + \frac{1-\pi_0}{\pi_0}\frac{2^n x!(n-x)!}{(n+1)!}\right]^{-1}$$

如果取$\pi_0 = 1/2, n = 5, x = 3$，则其贝叶斯因子为：

$$B^\pi(\boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \theta_0)}{m_1(\boldsymbol{x})} = \frac{\left(\frac{1}{2}\right)^n (n+1)!}{x!(n-x)!} = \frac{6!}{2^5 \cdot 3! \cdot 2!} = \frac{15}{8} \approx 2$$

由于先验机会比为1，故贝叶斯因子就是后验机会比，从而后验机会比也接近于2，应接受原假设$H_0: \theta = 1/2$。

假设的形式	Bayes因子	说　明
简单假设对简单假设	$B=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{p(x\mid\theta_0)}{p(x\mid\theta_1)}$	B 的大小表示样本支持θ_0的程度。
复杂假设对复杂假设	$B=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{m_0(x)}{m_1(x)}$	B 可看作Θ_0与Θ_1上的加权似然比，它部分地消除了先验分布的影响，而强调了样本观察值的作用。
简单假设对复杂假设	$B=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{p(x\mid\theta_0)}{m_1(x)}$	在实际应用中，可以先计算B，再计算后验概率：$\alpha_0=\left[1+\frac{1-\pi_0}{\pi_0}\frac{1}{B}\right]^{-1}$

例2.14 Berger(1995)一个临床试验有两个处理：

处理1：服药A；

处理2：同时服药A与药B。

问题：1.这两种处理方式有没有差别？

2.如果有差别，哪一种方式的疗效更好？

- 如今进行n 次对照试验，设 x_i 为第i 次对照试验中处理2与处理1的疗效之差，又设诸x_i相互独立同分布，且都服从 $N(\theta,1)$，于是前n次的样本均值 $\bar{x}_n \sim N(\theta, 1/n)$，先要考察如下二个假设：$H_0: \theta=0, H_1: \theta \neq 0$
- 由于对二个处理的疗效知之甚少，故对H_0和H_1取相等概率，即 $\pi_0=\pi_1=1/2$ ，而对 $H_1: \theta \neq 0$上的先验密度$g_1(\theta)$一般看法是：参数θ（疗效之差）接近于0比远离0更为可能，故取正态分布$N(0,2)$ 作为$g_1(\theta)$ 。

解：问题1 的解决。

由题设得到：样本分布：$p(\bar{x}\mid\theta)=\sqrt{\dfrac{n}{2\pi}}\exp\{-\dfrac{n}{2}(\bar{x}_n-\theta)^2\}$

- θ 的先验分布：$g_1(\theta)=\dfrac{1}{2\sqrt{\pi}}\exp\{-\dfrac{\theta^2}{4}\}$
- 则 $\bar{x}_n$ 对$g_1(\theta)$ 的边缘密度函数为：

$$\begin{aligned}
\boldsymbol{m}_1(\bar{\boldsymbol{x}}_n) &= \int_{-\infty}^{\infty}\boldsymbol{p}(\bar{\boldsymbol{x}}\mid\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \frac{1}{2\boldsymbol{\pi}}\sqrt{\frac{\boldsymbol{n}}{2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[\boldsymbol{n}(\bar{\boldsymbol{x}}_n-\boldsymbol{\theta})^2+\frac{\boldsymbol{\theta}^2}{2}\right]\right\}d\boldsymbol{\theta} \\
&= \frac{1}{\sqrt{2\boldsymbol{\pi}}}\frac{1}{\sqrt{2+\frac{1}{\boldsymbol{n}}}}\exp\left\{-\frac{\bar{\boldsymbol{x}}_n}{2}\Big/\left(2+\frac{1}{\boldsymbol{n}}\right)\right\}
\end{aligned}$$

这表明 $\bar{x}_n$ 对 $g_1(\theta)$ 的边际分布为正态分布 $N(0, 2+1/n)$

同时，由上述计算容易看出，在给定 $\bar{x}_n$ 下 θ (不含 $\theta=0$) 的后验分布可以算得:

$$\pi(\theta\,|\,\bar{x}_n) = p(\bar{x}_n\,|\,\theta)g_1(\theta)/m_1(\bar{x}_n)$$

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

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

即在给定 $\bar{x}_n$ 下， θ (不含 $\theta=0$)的后验分布为:

$$N(n\bar{x}_n/(n+\tfrac{1}{2}), (n+\tfrac{1}{2})^{-1})$$

则贝叶斯因子为：

$$B^{\pi}(\bar{x}_n)=\frac{p(\bar{x}\mid\theta=0)}{m_1(\bar{x}_n)}=\frac{\sqrt{\frac{n}{2\pi}}\exp\left\{-n\bar{x}_n^2/2\right\}}{\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{1+2n}}\exp\left\{-\frac{\bar{x}_n^2}{2}\Big/\left(2+\frac{1}{n}\right)\right\}}$$

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

- 由此可算得H_0和H_1的后验概率：

$$\alpha_0=P(\theta=0\mid\bar{x}_n)=\left(1+\frac{1}{B^{\pi}(\bar{x}_n)}\right)^{-1}=\frac{B^{\pi}(\bar{x}_n)}{1+B^{\pi}(\bar{x}_n)}$$

$$\alpha_1=P(\theta\neq 0\mid\bar{x}_n)=\frac{1}{1+B^{\pi}(\bar{x}_n)}$$

由于数据是逐步获得的，每获得一个新的数据后计算一次贝叶斯因子和两个后验概率，结果见下表所示。

表2.7 对照实验数据与各项后验概率

n	X_n	$\bar{x}_n$	B	α_0	α_1	α_{11}	α_{12}
1	1.63	1.63	1.006	0.417	0.583	0.054	0.529
2	1.03	1.33	0.543	0.352	0.648	0.030	0.618
3	0.19	0.95	0.829	0.453	0.547	0.035	0.512
4	1.51	1.09	0.363	0.266	0.734	0.015	0.719
5	-0.21	0.83	0.693	0.409	0.591	0.023	0.568
6	0.95	0.85	0.488	0.328	0.672	0.016	0.657
7	0.64	0.82	0.431	0.301	0.699	0.013	0.686
8	1.22	0.87	0.239	0.193	0.807	0.007	0.800
9	0.60	0.84	0.215	0.177	0.823	0.006	0.817
10	1.54	0.91	0.0888	0.082	0.918	0.003	0.915

结论：两种处理方式有差别。

问题2的解决。

- 为此我们研究下列三个假设：

$$H_0:\theta=0,\ H_{11}:\theta<0\ ,H_{12}:\theta>0$$

- 其中H_{11}表示处理2的疗效不如处理1，H_{12}表示处理2的疗效比处理1要好，同时研究这三个假设更为合理，利用 $\theta\neq 0$时θ 的后验分布

$$N(n\bar{x}_n/(n+\tfrac{1}{2}),(n+\tfrac{1}{2})^{-1})$$

- 容易算得H_{11}和H_{12}的后验概率：

$$\alpha_{11}=P(\theta<0\,|\,\bar{x}_n)=\Phi\left(\frac{-\sqrt{n}\bar{x}_n}{\sqrt{1+1/(2n)}}\right)/(1+B^{\pi}(\bar{x}_n))$$

$$\alpha_{12}=P(\theta>0\,|\,\bar{x}_n)=\Phi\left(\frac{\sqrt{n}\bar{x}_n}{\sqrt{1+1/(2n)}}\right)/(1+B^{\pi}(\bar{x}_n))$$

计算结果在上表中，可以看出应该拒绝H_{11}，而接受H_{12}，即处理2的疗效要好。

本节重点内容：

1.熟悉经典统计中假设检验的基本步骤；
2.掌握贝叶斯假设检验的基本思想与经典统计假设检验思想的区别；
3.掌握后验概率比（后验机会比）与贝叶斯因子两个重要概念；
4.熟练掌握几种特殊的假设检验情形下，贝叶斯因子的计算方法和应用。

§2.5　预 测

一、预测的基本概念与基本问题

预测：对随机变量未来观察值作出统计推断称为预测

统计预测大致有以下几种形式：

(1)设随机变量 $X \sim p(x|\theta)$，在参数θ未知情况下如何对 X 的未来的观察值作出推断？

(2)设$X_1, X_2, \cdots, X_n$是来自 $p(x|\theta)$的过去观察值，在参数θ未知情况下，如何对X 的未来的观察值作出推断？

(3)按密度函数$p(x|\theta)$得到一些数据$x_1, x_2, \cdots, x_n$后，如何对具有密度函数 $g(z|\theta)$ 的随机变量 Z 的未来的观察值作出推断，这里两个密度函数 p和g 都含有相同的未知参数θ 。

二、预测的贝叶斯方法

好多实际问题可以归结为**预测问题**，经典统计中容许区间就可以看成是解决预测问题的一种方法，但根本的困难在于参数 θ 不能被观察到。而在贝叶斯统计中可以利用θ 的先验分布 $\pi(\theta)$ 或者后验分布 $\pi(\theta \mid \boldsymbol{x})$很容易得到解决，解决的方案有两种，都是根据预测分布进行预测。

方案一：在无观测数据情形下的预测

设随机变量 $X \sim p(x|\theta)$，无 X 的观察数据，利用先验分布 $\pi(\theta)$ 容易获得未知的、但可观察的数据 x 的分布：

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

称该分布为“**先验预测分布**”，即为 X 的边缘分布。

注：这里的**先验**是指对过去的数据没有要求，**预测**是指它是可观测量的分布，有此先验预测分布就可以从中提取有用信息作出未来观测值的预测值或未来观测值的预测区间。

预测方法：用 $m(x)$ 的期望值、中位数或众数作为预测值，或者确定90%的预测区间 $[a,b]$，使得: $P^X(a \le X \le b) = 0.9$

其中 P^X 指用分布 $m(x)$ 计算概率。

方案二：有X的观测数据时的预测方法

设 X 有观察值 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$，利用后验分布 $\pi(\theta\mid x)$ 可获得未知观察值的分布。

(1) 如果要预测同一总体 $p(x\mid\theta)$ 的未来观察值，则可求出后验预测分布：

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

(2) 如果要预测另一总体 $g(z\mid\theta)$ 的未来观察值，同样可求出后验预测分布：

$$m(z\mid\boldsymbol{x})=\int_\Theta p(z\mid\theta)\pi(\theta\mid\boldsymbol{x})d\theta$$

预测方法：用 $m(z\mid\boldsymbol{x})$ 的期望值、中位数或众数作为 z 预测值，或者确定90%的预测区间 $[a,b]$ 使得：

$$P^{z|x}(a\le Z\le b\mid\boldsymbol{x})=0.90$$

其中 $P^{z|x}$ 指用后验预测分布 $m(z\mid\boldsymbol{x})$ 计算概率。

三、例题分析

例2.15 在 n 次相互独立的贝努里试验成功了x次，试对未来的 k 次相互独立的贝努里试验中成功次数 z 作出预测。

解：设成功概率为 θ ，则样本x的似然函数为:

$$L(x\mid\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

若取 θ 的共轭先验分布 $Be(\alpha,\beta)$，则其后验密度为:

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

新的样本 z 的似然函数为：$L(z\mid\theta)=\binom{k}{z}\theta^{z}(1-\theta)^{k-z}$

于是在给定 x 时，z 的后验预测分布为：

$$\pi(z\mid x)=\int_0^1\binom{k}{z}\theta^{z}(1-\theta)^{k-z}\pi(\theta\mid x)d\theta$$

$$=\binom{k}{z}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\int_0^1\theta^{z+x+\alpha-1}(1-\theta)^{k-z+n-x+\beta-1}d\theta$$

$$=\binom{k}{z}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\frac{\Gamma(z+x+\alpha)\Gamma(k-z+n-x+\beta)}{\Gamma(n+k+\alpha+\beta)}$$

实例分析：假设一赌徒在过去10次赌博中赢3次，试对未来5次赌博中他赢的次数 z 作出预测。

由以上分析可知该赌徒在未来5次赌博中他赢的次数 z 的后验预测分布为：

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

由此可计算出 z 的后验预测概率如下：

z	0	1	2	3	4	5
$m(z \mid x=3)$	0.181 3	0.302 2	0.274 7	0.164 9	0.0641	0.02128

结论: (1)区间[0，3]是 z 的92%预测区间；

(2)在未来5次赌博中能赢1到2次的可能性较大。

讨论：在无观测数据的情况下，怎样对未来的k次相互独立的贝努里试验中成功次数z作出预测。

$$m(z)=\binom{k}{z}\int_0^1 \theta^z(1-\theta)^{k-z}d\theta$$

$$=\binom{k}{z}\frac{\Gamma(z+1)\Gamma(k-z+1)}{\Gamma(k+2)}$$

$$=\frac{1}{k+1},\ z=0,1,\cdots,k$$

k=5时，m(z)=1/6

例2.16 一颗钻石在一架天平上重复称重n次，结果为$x_1,x_2,\cdots,x_n$，若把这颗钻石放在另一架天平上称重，如何对其称量值作出预测？（自学）

§2.6 似然原理

1.对似然函数的理解

若设$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$是来自密度函数 $p(x|\theta)$ 的一个样本，则其乘积：

$$p(\boldsymbol{x}|\theta)=\prod_{i=1}^{n}p(x_i|\theta)$$

有两个解释：

(1) 当 θ 给定时，$p(\boldsymbol{x}|\theta)$ 是样本 $\boldsymbol{x}$ 的联合密度函数；

(2) 当样本 $\boldsymbol{x}$ 的观察值给定时，$p(\boldsymbol{x}|\theta)$ 是未知参数θ 的函数，并称为似然函数，记为$L(\theta)$ 。

2.似然原理

(1)有了观察值$\boldsymbol{x}$之后，在做关于θ的推断和决策时，所有与试验有关的 θ 信息均被包含在似然函数$L(\theta)$ 之中。

(2)如果有两个似然函数是成比例的，比例常数与θ无关，则他们关于 θ 含有相同的信息。

3.两个学派对似然原理的不同理解而产生的影响。

似然原理是统计学规范中大家都应遵守的公理，是统计学最一般的基础原理，遵守此原理而产生的行为或行动才能认为是合理的，可经典统计学不是这样，他们在寻求最大似然估计前是承认似然原理的，但在得到最大似然估计后，他们就抛弃了似然原理，把样本观察值又看作样本，最大似然估计又被看作样本的函数，看作几个独立同分布随机变量的函数，从而用联合密度 $p(\boldsymbol{x}|\theta)$ 计算它的期望、方差，研究它的大样本性质等，依据似然原理，只有实际观测到的数据 $\boldsymbol{x}$ 才能构成 θ 的论据，只有试验中得到的有关 θ 的证据才能用于推断，因此无偏性等原理都违背了似然原理，因为它们依赖于尚未获得的观测值，由此可见，似然原理把 § 2.1 叙述的条件观点说得更清楚了，下面的例子可以进一步说明经典学派与贝叶斯学派对似然原理的不同态度而引出的问题。

例2.17 Lindley和Phillips(1976)的成果

问题的描述：设 θ 为向上抛一枚硬币时出现正面的概率，现要检验如下二个假设：

$$H_0:\theta=1/2 \leftrightarrow H_1:\theta>1/2$$

为此做了<u>一系列</u>相互独立的抛此硬币的试验，结果出现9次正面和3次反面。怎样作出合理的判断。

分析：解决该问题，关键取决于对“一系列试验”的理解。因为事先没有对它有明确的规定，因此可能有如下两种情形：

(1)事先已经决定抛12次硬币。在这种情况下，正面出现次数$X \sim b(n,\theta)$，其中n为总试验次数，即 $n=12$，于是相应的似然函数为：

$$L_1(\theta)=P_1(X=x\mid\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}=220\theta^9(1-\theta)^3$$

(2)事先规定试验进行到出现3次反面为止。则正面出现次数X服从负二项分布$Nb(k,\theta)$，其中 k为反面出现次数，即 $k=3$，于是相应的似然函数为：

$$L_2(\theta)=P_2(X=x\mid\theta)=\binom{k+x-1}{x}\theta^x(1-\theta)^{n-x}=55\theta^9(1-\theta)^3$$

经典统计作出的判断：

在二项分布模型和负二项分布模型下，犯第Ⅰ类错误的概率分别为：

$$\alpha_1 = P_1(X \geq 9 \mid \theta = 1/2) = \sum_{x=9}^{12} P_1(X = x \mid \theta = 1/2) = 0.075$$

$$\alpha_2 = P_2(X \geq 9 \mid \theta = 1/2) = \sum_{x=9}^{\infty} P_1(X = x \mid \theta = 1/2) = 0.033$$

如果取α=0.05 作为显著性水平，在二项分布模型下，$\alpha_1 > \alpha$ ，$X = 9$不包含在拒绝域内，故应接受H_0，在负二项分布模型下，$\alpha_1 < \alpha$，故 $X = 9$在拒绝域内，从而拒绝H_0，即这两个模型将得出完全不同的结论，这是与似然原理相矛盾的。

贝叶斯统计作出的结果：

(1)认清假设类型：简单假设对复杂假设

(2)确定先验分布：$\pi(\theta)=\pi_0 I_{\{0.5\}}(\theta)+\pi_1 g_1(\theta)$

其中 $\pi_0=\pi_1=1/2$， $g_1(\theta)=U(0.5,1)$

(3)计算贝叶斯因子：

$$B_i^{\pi}(x=9)=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{P_i(X=9\mid\theta=1/2)}{m_i(x=9)}$$

其中

$$P_i(X=9\mid\theta=1/2)=k_i\theta^9(1-\theta)^3=0.000244k_i$$

这里 $k_1=220,\ \ k_2=55$

$$m_i(x=9)=\int_{1/2}^{1}P_i(X=9\mid\theta=1/2)g_1(\theta)d\theta$$

$$=\int_{1/2}^{1}k_i\theta^9(1-\theta)^3\cdot 2d\theta$$

$$=2k_i\int_{1/2}^{1}(\theta^9-3\theta^{10}+3\theta^{11}-\theta^{12})d\theta$$

$$=0.000666k_i$$

由此可计算出两种情形下的贝叶斯因子：

$$B^{\pi}(x=9)=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}=\frac{P_i(X=9\mid\theta=1/2)}{m_i(x=9)}=0.3664$$

可见观测值$x=9$并不支持原假设H_0，考虑到 $\pi_0=\pi_1=1/2$

$\alpha_0+\alpha_1=1$，可得二个假设的后验概率

$$\alpha_0 = \pi(H_0 \mid x = 9) = \frac{0.3664}{1 + 0.3664} = 0.2681$$

$$\alpha_1 = \pi(H_1 \mid x = 9) = \frac{1}{1 + 0.3664} = 0.7319$$

据此我们应拒绝H_0而接受 H_1， 这个结论只与似然函数有关，而与总体是二项分布还是负二项分布无关。

练习

1. 设随机变量X的密度函数为 $p(x|\theta)=\dfrac{2x}{\theta^2},0<x<\theta<1$

(1)假如θ的先验分布为U(0,1),求θ的后验分布.
(2)假如θ的先验分布为

$$\pi(\theta)=3\theta^2,0<\theta<1$$

求θ的后验分布及后验期望估计

参考答案:

(1) $\pi(\theta|x)=\dfrac{x}{(1-x)\theta^2},x<\theta<1,\hat{\theta}_E=\dfrac{x\ln x}{x-1}$

(2) $\pi(\theta|x)=\dfrac{1}{(1-x)},x<\theta<1,\hat{\theta}_E=\dfrac{1+x}{2}$

2. 对正态分布**N(0,1)**观察**,**获得三个观察值

$$x_1 = 2, x_2 = 3, x_3 = 4$$

若**θ**的先验分布为**N(3,1),**求**θ**的**0.95**可信区间。

(答案：**[2.02,3.98])**