第三章 先验分布的确定

第三章 先验分布的确定

§3.1 主观概率

一、主观概率

1. 贝叶斯学派要研究的问题：如何用人们的经验和过去的历史资料确定概率和先验分布。

2. 经典统计确定概率的两种方法：

（1）古典方法；　（2）频率方法。

3. **主观概率**的定义：一个事件的概率是人们根据经验对该事件发生可能性所给出的个人信念。

二、确定主观概率的方法

- 1.利用对立事件的比较确定主观概率
- 2.利用专家意见确定主观概率;
- 3.向多位专家咨询确定主观概率;
- 4.充分利用历史资料，考虑现有信息加以修正,才能得到比较切合实际的主观概率。

1.利用对立事件的比较确定主观概率

例 3.1.1 一位出版商要知道一本新书畅销（事件 A）的概率是多少，以决定是否与作者签订出版合同。他在了解这本新书的内容后，根据他自己多年出书的经验认为该书畅销的可能性较大，畅销(A)比不畅销($\overline{A}$)的可能性要高出一倍，即 $P(A)=2P(\overline{A})$，由此根据概率性质 $P(A)+P(\overline{A})=1$ 可以推得 $P(A)=2/3$，即该书畅销的主观概率是 2/3。

这种用对立事件的比较来确定主观概率是最简单方法。假如对 A 与 $\overline{A}$ 说不出哪一个发生的可能性大，那就定 $P(A)=1/2$。这个方法可以推广到多个事件上这种方法虽不严格，但符合人们在实际中看待概率的直观方式。

2.利用专家意见确定主观概率

例 3.1.2 有一项带有风险的生意，欲估计成功(记为 A)的概率。为此，决策者去拜访这方面的专家(如董事长，银行家等)，向专家提这样的问题。“如果这种生意做 100 次，你认为会成功几次?”专家回答:“成功次数不会太多，大约 60 次。”这时 $P(A)=0.6$ 是专家的主观概率，可此专家还不是决策者。但决策者很熟悉这位专家，认为他的估计往往是偏保守的，过分谨慎的。决策者决定修改专家的估计，把 0.6 提高到 0.7 这样 $P(A)=0.7$ 就是决策者自己的主观概率。

这种用专家意见来确定主观概率的方法是常用的。当决策者对某事件了解甚少时，就可去征求专家意见。这里要注意两点。一是向专家提的问题要设计好，既要使专家易懂，又要使专家回答不是模棱两可；二是要对专家本人较为了解，以便作出修正，形成决策者自己的主观概率。

3. 向多位专家咨询确定主观概率

例 3.1.3 某公司在决定是否生产某种新品时，想估计该产品在未来市场上畅销(记为 A)的概率是多少，为此公司经理召集设计、财会、推销和质量管理等方面人员的座谈会，仔细分析影响新产品销路的各种因素，大家认为此新产品质量好，只要定价合理，畅销可能性很大，而影响销路的主要因素是市场竞争。据了解，还有一家工厂(简称外厂)亦有生产此新产品的想法，该厂技术和设备都比本厂强。经理在听取大家的分析后，向在座各位提出二个问题：

(i)假如外厂不生产此新产品，本公司的新产品畅销的可能性(即概率)有多大?

(ii)假如外厂要生产此新产品，本公司的新产品畅销的可能性(即概率)有多大?

在座人员根据自己的经验各写了二个数，经理在计算了二个平均值后，略加修改，提出自己看法：在上述二种情况下，本公司新产品畅销概率各为 0.9 和 0.4，这是经理在征求多位专家意见后所获得的主观概率。另据本公司情报部门报告，外厂正忙于另一项产品开发，很可能无暇顾及生产此新产品。经理据此认为，外厂将生产此新产品的概率为 0.3，不生产此新产品的概率为 0.7。

利用上述 4 个主观概率，由全概率公式可得本公司生产此新产品获畅销的概率为

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

4.充分利用历史资料，考虑现有信息加以修正

例 3.1.4 某公司经营儿童玩具多年，今设计了一种新式玩具将投入市场。现要估计此新式玩具在未来市场上的销售情况。经理查阅了本公司过去 37 种新式玩具的销售记录，得知销售状态是畅销(A_1)、一般(A_2)、滞销(A_3)分别有 29,6,2 种，于是算得过去新式玩具的三种销售状态的概率分别为

$$\frac{29}{37}=0.784,\quad \frac{6}{37}=0.162,\quad \frac{2}{37}=0.054$$

考虑到这次设计的新玩具不仅外形新颖，而且在开发儿童智力上有显著突破，经理认为此种新玩具会更畅销一些，滞销可能性更小，故对上述概率作了修改，提出自己的主观概率如下：

$$P(A_1)=0.85, P(A_2)=0.14, P(A_3)=0.01$$

这个例子表明，假如有历史数据，要尽量利用，帮助形成初步概念，然后作一些对比修正，再形成个人信念，这对给出主观概率大有好处。

注意事项：

（1）不管按照什么方法确定的主观概率必须满足概率的三条公理：

①非负性公理：对任意事件A，$0 \le P(A) \le 1$。

②正则性公理：必然事件的概率为1

③可列可加性公理：对可列个互不相容的事件A_1，$A_2, \cdots, A_n$，有

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

（2）如果发现所确定的主观概率与上述三个公理及其推出的性质相悖，必须立即修正。直到两者一致为止。（例3.1.5）

例 3.1.5 某人对事件 A, B 及其并 $A\cup B$ 分别给出主观概率如下

$$P(A)=1/3, P(B)=1/3, P(A\cup B)=3/4$$

按概率性质，应有 $P(A\cup B)\leqslant P(A)+P(B)$，然而现在

$$P(A\cup B)=3/4, P(A)+P(B)=2/3$$

这个性质不满足，这个不和谐之处是由于这组主观概率给定不恰当而引起的，必须修正。此人重新给出这三个事件的主观概率，发现并事件 $A\cup B$ 的主观概率应为 3/5。再检查就没有此种不和谐现象了。

§3.2 利用先验信息确定先验分布

在贝叶斯方法中关键的一步是确定先验分布。当总体参数θ是离散时，即参数空间Θ只含有限个或可数个点时，可对Θ中每个点确定一个主观概率，而主观概率可用§3.1所述的方法确定。

当总体参数θ是连续时，即参数空间Θ是实数轴或其上某个区间时，要构造一个先验密度$\pi(\theta)$那就有些困难了。当θ的先验信息（经验和历史数据）足够多时，下面三个方法可供使用。

一、直方图法

二、选定先验密度函数形式再估计其超参数

三、定分度法与变分度法

一、直方图法

基本步骤：

1. 把参数空间分成一些小区间；
2. 在每个小区间上决定主观概率或依据历史数据确定其频率；
3. 绘制频率直方图；
4. 在直方图上作一条光滑曲线，此曲线即为先验分布$\pi(\theta)$。

例 3.2.1 某药材店记录了吉林人参的每周销售量，现要录求每周平均销售量 θ 的概率分布。现用直方图法来确定它。直方图法的要点如下：

(1)把参数空间分成一些小区间。统计过去二年 102 个营业周的销售记录，每周平均销售是最高不超过 35 两。若以 5 两作为小区间长度，共分为 7 个小区间。

(2)在每个小区间上决定主观概率或依据历史数据确定其频率。这里是用后者，其频率见表 3.2.1。

表 3.2.1 每周平均销售量统计表

平均销售量(两)	频率
[0,5]	0.051
(5,10]	0.259
(10,15]	0.327
(15,20]	0.224
(20,25]	0.095
(25,30]	0.044
(30,35]	0.001

(3)绘制频率直方图。这里绘制的频率直方图见图 3.2.1，其中纵坐标为频率/5。

(4)在直方图上作一条光滑的曲线，此曲线就是 $\pi(\theta)$。在做光滑曲线时，尽量在每个小区间上使用得曲线下的面积与直方图的面积相等。这条曲线已在图 3.2.1 上画出，利用此曲线可求出一个单位区间上的概率，如

$$P(20<\theta\leqslant 21)=1\times\pi(20.5)=0.03$$

这样得到的先验密度常常仅限于在有限区间上，有时使用也不方便，下面一个方法更为适用些。

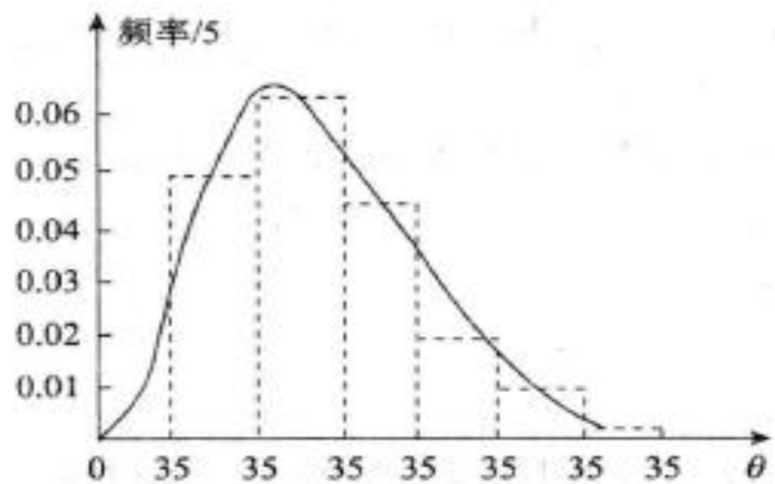

图 3.2.1　每周平均销售量的直方图

二、选定先验密度函数形式再估计其超参数

- 该方法的**要点**：
- （1）根据先验信息选定 θ 的先验密度函数 $\pi(\theta)$ 的形式，如选其共轭先验分布。
- （2）当先验分布中含有未知参数（称为超参数）时，譬如 $\pi(\theta)=\pi(\theta;\alpha,\beta)$，给出超参数 α,β 的估计值 $\hat{\alpha}$，$\hat{\beta}$ 使 $\pi(\theta;\hat{\alpha},\hat{\beta})$ 最接近先验信息。

例 3.2.2 在例 3.2.1 中对周平均销售量 θ，选用正态分布 $N(\mu,\tau^2)$ 作为先验分布，于是确定先验分布问题就转化为估计超参数 μ 与 τ^2 的问题。这可从每周平均销售量统计表上作出估计。若对 Θ 的每个小区间用其中点作代表，则可算得 μ 与 τ^2 的估计如下：

$$\hat{\mu}=2.5\times0.051+7.5\times0.259+\cdots+32.5\times0.001=13.4575$$

$$\hat{\tau}^2=(2.5-13.4574)^2\times0.051+(7.5-13.4574)^2\times0.259+\cdots$$
$$+(32.5-13.4574)^2\times0.001=36.0830=6.0069^2$$

这表明，该商店每周平均销售量 θ 的先验分布为 $N(13.4574,36.0830)$。用此先验分布可以算得

$$P(20<\theta<21)=\Phi\left(\frac{21-13.4575}{6.0069}\right)-\Phi\left(\frac{20-13.4575}{6.0069}\right)$$
$$=\Phi(1.2556)-\Phi(1.0892)$$
$$=0.8953-0.8603$$
$$=0.0350$$

例 3.2.3 设参数 θ 的取值范围是 $(-\infty,\infty)$，它的先验分布具有正态分布形式。若从先验信息可以得知：

(1)先验中位数为 0；

(2)上、下四分位数为 −1 和 1，即先验的 0.25 分位数和 0.75 分位数为 −1 和 1。要确定先验分布 $N(\mu,\tau^2)$ 中的超参数 μ 与 τ。

对正态分布，均值与中位数相等，故 $\mu=0$，另外由 0.75 分位数为 1，可列出方程

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

查标准正态分布表可知

$$1/\sigma=0.675 \text{ 或 } \sigma=1.481$$

这样就可得先验分布为 $N(0,1.481^2)$。

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

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

它的期望与方差都不存在,但其各分位数都有。由于柯西密度函数是关于α的对称函数,故其中位数是α。由已知条件知,$\alpha=0$。另外,由-1是$1/4$分位数可得方程

$$\int_{-\infty}^{-1}\frac{\beta}{\pi(\beta^2+\theta^2)}d\theta=1/4$$

由此可算得$\beta=1$。这时θ的先验分布为柯西分布$C(0,1)$。这是标准柯西分布。

这样一来，我们面临着二个先验分布都满足给定的先验信息。假如这二个先验分布差异不大，对后验分布影响也不大，那可任选一个，假如我们面临着二个差异极大的先验分布可供选择时，我们应慎重选择，因为不同的选择对后验分布影响也会很大。譬如在本例中，正态分布 $N(0,1.481^2)$ 与柯西分布 $C(0,1)$ 在形状上是很相似，都是中间高，两边低，左右对称，但在二侧的尾部的粗细相差很大，正态分布的尾部很细，柯西分布的尾部很粗，这就导致正态分布的各阶矩都存在，可柯西分布连数学期望都不存在。因此在进一步的选择前还要对先验信息进行分析，若先验信息很分散，那就不宜选用正态分布，若先验信息较为集中，那就不宜选用柯西分布。关于在一族先验分布中如何选择先验分布使后验分布波动不大。这个问题被称为“稳健性(Robustness)问题”，是贝叶斯统计近几年研究较多的问题之一。

三、定分度法与变分度法

基本概念:

(1)**定分度法**:把参数可能取值的区间逐次分为长度相等的小区间,每次在每个小区间上请专家给出主观概率.

(2)**变分度法**:该法是把参数可能取值的区间逐次分为机会相等的两个小区间,这里的分点由专家确定.

例3.9(自学)

§3.3 利用边缘分布 $m(x)$ 确定先验密度

一、边缘分布 $m(x)$

设总体 X 的密度函数为 $p(x|\theta)$，它含有未知参数θ，若θ的先验分布选用形式已知的密度函数$\pi(\theta)$，则可算得X 的边缘分布（即无条件分布）：

$$m(x)=\begin{cases}\int_{\Theta} p(x\mid\theta)\pi(\theta)d\theta, & \text{当}\theta\text{为连续时}\\ \sum_{\theta\in\Theta} p(x\mid\theta)\pi(\theta), & \text{当}\theta\text{为离散时}\end{cases}$$

当先验分布含有未知参数，譬如 $\pi(\theta)=\pi(\theta|\lambda)$，那么边缘分布 $m(x)$依赖于λ，可记为 $m(x|\lambda)$，这种边缘分布在寻求后验分布时常遇到。

例 3.3.1 设总体 $X \sim N(\theta, \sigma^2)$，其中 σ^2 已知，又设 θ 的先验分布为 $N(\mu_\pi, \sigma_\pi^2)$，则可以算得边缘分布 $m(x)$ 为 $N(\mu_\pi, \sigma_\pi^2 + \sigma^2)$。事实上，在这个例子中

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

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\pi} \exp\left\{-\frac{1}{2\sigma_\pi^2}(\theta-\mu_\pi)^2\right\}$$

于是边缘分布

$$m(x) = \frac{1}{2\pi\sigma\sigma_\pi}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_\pi)^2}{\sigma_\pi^2}\right]\right\} d\theta$$

若令

$$A = \frac{1}{\sigma^2} + \frac{1}{\sigma_\pi^2}, B = \frac{x}{\sigma^2} + \frac{\mu}{\sigma_\pi^2}, C = \frac{x^2}{\sigma^2} + \frac{\mu^2}{\sigma_\pi^2}$$

则可算得

$$m(x) = \frac{1}{2\pi\sigma\sigma_\pi}\exp\left\{-\frac{1}{2}\left(C-\frac{B^2}{A}\right)\right\}\times\sqrt{\frac{2\pi}{A}}$$

$$= \frac{1}{\sqrt{2\pi(\sigma^2+\sigma_\pi^2)}}\exp\left\{-\frac{AC-B^2}{2A}\right\}$$

由于

$$AC-B^2=(x-\mu_\pi)^2/\sigma^2\sigma_\pi^2$$

最后可得

$$m(x) = \frac{1}{\sqrt{2\pi(\sigma^2+\sigma_\pi^2)}}\exp\left\{-\frac{(x-\mu_\pi)^2}{2(\sigma^2+\sigma_\pi^2)}\right\} \qquad (3.3.2)$$

这就是我们要求的结果。除了已知的 σ^2 外，它还会有二个未知的超参数 μ 与 σ_π^2。

二、混合分布

(1) 混合分布的概念：

设随机变量 X 以概率 π 在总体 F_1 中取值，以概率 $1-\pi$ 在总体 F_2 中取值。若 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 分别是这二个总体的分布函数，则 X 的分布函数为

$$F(x)=\pi F(x|\theta_1)+(1-\pi)F(x|\theta_2) \tag{3.3.3}$$

或用密度函数(或概率函数)表示

$$p(x)=\pi p(x|\theta_1)+(1-\pi)p(x|\theta_2) \tag{3.3.4}$$

这个分布 $F(x)$ 称为 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 的**混合分布**。这里的 π 和 $1-\pi$ 可以看作一个新的随机变量 θ 的分布，即

$$P(\theta=\theta_1)=\pi=\pi(\theta_1),P(\theta=\theta_2)=1-\pi=\pi(\theta_2)$$

(2)混合样本的概念：

从上述定义容易看出：从混合分布 $F(x)$ 中抽取一个样品 x_1，相当于如下的二次抽样：

第一次，从 $\pi(\theta)$ 中抽取一个样品 θ。

第二次，若 $\theta=\theta_1$，则从 $F(x|\theta_1)$ 中再抽一个样品，这个样品就是 x_1，若 $\theta=\theta_2$；则从 $F(x|\theta_2)$ 中再抽一个样品，这个样品就是 x_1。

若从混合分布抽取一个容量为 n 的样本 $x_1,x_2\cdots,x_n$，那么其中约有 $n\pi(\theta_1)$ 个来自 $F(x|\theta_1)$，约有 $n\pi(\theta_2)$ 个来自 $F(x|\theta_2)$，这样的样本 $x_1,x_2,\cdots,x_n$ 有时也称为**混合样本**。

(3) 注意：

从上述混合分布的定义很容易看出，(3.3.1)式表示的边缘分布 $m(x)$ 是混合分布的推广，只不过是用密度函数形式表示而已。当 θ 为离散随机变量时，$m(x)$ 是由有限个或可数个密度函数混合而成，当 θ 为连续随机变量时，$m(x)$ 是由无限不可数个的密度函数混合而成。若从 $\pi(\theta)$ 抽取一个 θ，然后再从 $p(x|\theta)$ 中抽取一个 x，这个 x 可看作从 $m(x)$ 抽取的样品。按此过程抽取 n 个样品就可获得容量为 n 的混合样本，在实际中经常会遇到各种混合样本。

例 3.3.2　混合样本的例子

(1)设 $x_1, x_2, \cdots x_n$ 是 n 位考生的成绩，由于每位考生的能力 θ 是不同的，这 n 位考生的能力 $\theta_1, \theta_2, \cdots, \theta_n$ 可看作从某个分布 $\pi(\theta)$ 抽取的样本，而 x_i 是从 $p(x|\theta_i)$ 抽取的样品。这样一来，样本 $x_1, x_1, \cdots x_n$ 可看作混合样本。

(2)从一批产品中随机抽取 n 件产品，而这 n 个产品是来自三位工人之手，而这三个工人的不合格品率是不同的，故所测的产品特性 $x_1, x_2, \cdots, x_n$ 可看作一个混合样本。

(3)某厂的原料来自 p 个产地，每次改换原料都要抽一个样本检查产品质量。过去已记录了若干个样品的观察值

$$\begin{array}{cccc} x_{11} & x_{12} & \cdots & x_{1n_1} \\ x_{21} & x_{22} & \cdots & x_{2n_2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn_p} \end{array}$$

其中 $n_1, n_2, \cdots, n_p$ 分别是各自的样本容量。这 $n_1+n_2+\cdots+n_p$ 个数据可看作来自 k 个总体的混合样本，也可看作来自某混合分布的一个样本。

三、先验选择的ML-II方法

在边缘分布 $m(x)$ 的表示式(3.3.1)中，若 $p(x|\theta)$ 已知，则 $m(x)$ 的大小反映 $\pi(\theta)$ 的合理程度，这里把 $m(x)$ 记为 $m^{\pi}(x)$。当观察值 x 对二个不同的先验分布 π_1 和 π_2，有

$$m^{\pi_1}(x)>m^{\pi_2}(x)$$

时，人们可认为，数据 x 对 π_1 比对 π_2 提供更多支持。于是把 m^{π} 看作 π 的似然函数是合理的。既然 π 有似然函数可言，那么用极大似然法选取 π 就是很自然的事。这样定出的先验称为Ⅱ**型极大似然先验**，或称为 ***ML*－Ⅱ先验**。

三、先验选择的ML-II方法

假如混合样本 $x=(x_1,\cdots,x_n)$ 所涉及的先验密度函数的形式已知，未知的仅是其中的超参数，即先验密度函数族可表示如下：

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

这时寻求 ML-II 先验是较为简单的事，只要寻求这样的 $\hat{\lambda}$ 使得

$$m(\boldsymbol{x}|\hat{\lambda})=\sup_{\lambda\in\Lambda}\prod_{i=1}^{n}m(x_i|\lambda) \tag{3.3.5}$$

这可用最大化似然函数方法来实现。

例 3.3.3 设 $X \sim N(\theta,\sigma^2)$，其中 σ^2 已知，又设 $\theta \sim N(\mu_\pi,\sigma_\pi^2)$，在例 3.3.1 中已算得

$$m(x|\mu_\pi,\sigma_\pi^2)=N(\mu_\pi,\sigma_\pi^2+\sigma^2)$$

若有来自 $m(x|\mu_\pi,\sigma_\pi^2)$ 的混合样本 $x_1,x_2,\cdots,x_n$，则超参数 μ_π,σ_π^2 的似然函数为

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

$$= 2\pi(\sigma_\pi^2+\sigma^2) \mid^{-\frac{n}{2}} \exp\left\{\frac{-ns_n^2}{2(\sigma_\pi^2+\sigma^2)}\right\} \exp\left\{\frac{-n(\bar{x}-\mu_\pi)^2}{(\sigma_\pi^2+\sigma^2)}\right\}$$

其中 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, s_\pi^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$。从上式容易看出，当不考虑 σ_π^2 时，μ_π 在 $\bar{x}$ 处达到最大，所以 $\hat{\mu}_\pi=\bar{x}$ 应是 μ_π 的 $ML-$ Ⅱ 选择。将 μ_π 以 $\bar{x}$ 代入上式，只剩 σ_π^2 的函数

$$\phi(\sigma_\pi^2)=[2\pi(\sigma_\pi^2+\sigma^2)]^{-\frac{n}{2}} \exp\left\{\frac{-ns_n^2}{2(\sigma_\pi^2+\sigma^2)}\right\}$$

对 $\ln\phi(\sigma_\pi^2)$微分并令为零,可得似然方程

$$\frac{d\ln\phi(\sigma_\pi^2)}{d\sigma_\pi^2}=\frac{-n/2}{\sigma_\pi^2+\sigma^2}+\frac{-ns_n^2}{2(\sigma_\pi^2+\sigma^2)^2}=0$$

解之,可得 $\sigma_\pi^2=S_n^2-\sigma^2$。若 $S_n^2<\sigma^2$,导致 σ_n^2 为负值,这不合情理,故令 $\sigma_\pi^2=0$。于是 σ_π^2 的 ML－Ⅱ估计为

$$\hat{\sigma}_\pi^2=\begin{cases}0, & 当\ s_n^2<\sigma^2\\ s_\pi^2-\sigma^2, & 当\ s_n^2\geq\sigma^2\end{cases}$$

从而所求的 ML-Ⅱ先验为 $\pi=N(\hat{\mu}_\pi,\hat{\sigma}_\pi^2)$

四、先验选择的矩方法

当先验密度函数 $\pi(\theta|\lambda)$ 的形式已知，可利用先验矩与边缘分布矩之间的关系寻求超参数 λ 的估计。这种方法称为**先验选择的矩方法**。该方法的***具体步骤***是：

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

即

$$\mu(\theta)=E^{x|\theta}(X)$$

$$\sigma^2(\theta)=E^{x|\theta}(X-\mu(\theta))^2$$

- $E^{x|\theta}$ 表示用 θ 给定下的条件分布 $p(x|\theta)$ 求期望。

2.计算边缘密度 $m(x|\lambda)$ 的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$ ，其中:

$$\mu_m(\lambda) = E^{x|\lambda}(X) = \int_x xm(x|\lambda)dx$$

$$= \int_x x\int_\Theta p(x|\theta)\pi(\theta|\lambda)d\theta dx = \int_\Theta \int_x xp(x|\theta)dx\pi(\theta|\lambda)d\theta$$

$$= \int_\Theta \mu(\theta)\pi(\theta|\lambda)d\theta = E^{\theta|\lambda}[\mu(\theta)]$$

$$\sigma_m^2(\lambda) = E^{x|\lambda}[X - \mu_m(\lambda)]^2$$
$$= \int_x (x - \mu_m(\lambda))^2 \int_\Theta p(x\,|\,\theta)\pi(\theta\,|\,\lambda)d\theta dx$$
$$= \int_\Theta \int_x (x - \mu_m(\lambda))^2 p(x\,|\,\theta)dx\pi(\theta\,|\,\lambda)d\theta$$
$$= \int_\Theta E^{x|\theta}(x - \mu_m(\lambda))^2 \pi(\theta\,|\,\lambda)d\theta$$

其中:

$$E^{x|\theta}(x - \mu_m(\lambda))^2$$
$$= E^{x|\theta}(x - \mu(\theta) + \mu(\theta) - \mu_m(\lambda))^2$$
$$= E^{x|\theta}(x - \mu(\theta))^2 + E^{x|\theta}(\mu(\theta) - \mu_m(\lambda))^2$$
$$= \sigma^2(\theta) + (\mu(\theta) - \mu_m(\lambda))^2$$

代入上式得:

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

3. 特殊情形:当先验分布中仅含二个超参数时，即

$\lambda=(\lambda_1,\lambda_2)$, 可用混合样本 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$

计算其样本均值和样本方差，即:

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

再用样本矩代替边际分布的矩，列出如下方程

$$\hat{\mu}_m=E^{\theta|\lambda}[\mu(\theta)]$$

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

解此方程组，可得超参数 $\lambda=(\lambda_1,\lambda_2)$ 的估计 $\hat{\lambda}=(\hat{\lambda}_1,\hat{\lambda}_2)$

从而获得先验分布 $\pi(\theta\mid\hat{\lambda})$

例3.3.4 设总体X~Exp(θ)，其密度函数为

$$p(x \mid \theta)=\theta e^{-\theta x}，\ x>0,$$

参数θ的先验分布取伽玛分布$Ga(\alpha,\lambda)$，其密度函数

$$\pi(\theta \mid \alpha,\lambda)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}，\ \theta>0$$

现有混合样本的均值$\hat{\mu}_m$和方差$\hat{\sigma}_m^2$，试求超参数α，λ的矩估计。

分析:

(1)计算指数分布 Exp(θ)的期望和方差

$$\mu(\theta)=\theta^{-1} \qquad \sigma^2(\theta)=\theta^{-2}$$

(2)计算(3.3.8)中所示的三个先验矩。

$$\begin{aligned}
E^{\theta|\lambda}[\mu(\theta)] &= E^{\theta|\lambda}[\theta^{-1}] \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty \theta^{\alpha-2} c^{-\lambda\theta} d\theta \\
&= \frac{\lambda}{\alpha-1}
\end{aligned}$$

$$\begin{aligned}
E^{\theta|\lambda}[\sigma^2(\theta)] &= E^{\theta|\lambda}[\theta^{-2}] \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty \theta^{\alpha-3} c^{-\lambda\theta} d\theta \\
&= \frac{\lambda^2}{(\alpha-1)(\alpha-2)}
\end{aligned}$$

$$E^{\theta|\lambda}[\mu(\theta)-\mu_m(\lambda)]^2 = E^{\theta|\lambda}[\theta^{-1}-\frac{\lambda}{\alpha-1}]^2$$
$$= E^{\theta|\lambda}[\theta^{-2}]-\frac{2\lambda}{\alpha-1}E^{\theta|\lambda}[\theta^{-1}]+\frac{\lambda^2}{(\alpha-1)^2}$$
$$= \frac{\lambda^2}{(\alpha-1)(\alpha-2)}-\frac{2\lambda}{\alpha-1}\frac{\lambda}{\alpha-1}+\frac{\lambda^2}{(\alpha-1)^2}$$
$$= \frac{\lambda^2}{(\alpha-1)(\alpha-2)}-\frac{\lambda^2}{(\alpha-1)^2}$$

把上述三式代入(3.3.6)和(3.3.7)，即得边缘分布的期望与方差

$$\mu_m(\lambda) = \frac{\lambda}{\alpha-1}$$
$$\sigma_m^2(\lambda) = \frac{2\lambda^2}{(\alpha-1)(\alpha-2)}-\frac{\lambda^2}{(\alpha-1)^2}$$
$$= \left(\frac{\lambda}{\alpha-1}\right)^2\frac{\alpha}{\alpha-2}$$

(3)用样本矩代替边缘分布的矩，列出方程

$$\hat{\mu}_m = \frac{\lambda}{\alpha - 1}$$

$$\hat{\sigma}_m^2 = \left(\frac{\lambda}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2}$$

把第一方程代入第二方程中去，可得

$$\hat{\sigma}_m^2 = \hat{\mu}_m^2 \frac{\alpha}{\alpha - 2}$$

解之可得

$$\hat{\alpha} = \frac{2\hat{\sigma}_m^2}{\hat{\sigma}_m^2 - \hat{\mu}_m^2}$$

再代回第一方程，即得

$$\hat{\lambda} = (\alpha - 1)\hat{\mu}_m = \frac{\hat{\sigma}_m^2 + \hat{\mu}_m^2}{\hat{\sigma}_m^2 - \hat{\mu}_m^2}$$

这就是超参数和矩估计。

例3. 14 设总体$X\sim N(\theta,1)$，其中参数θ的先验分布取共轭先验 $N(\mu_\pi,\sigma_\pi^2)$ 。试估计两个参数的值。

解:

很容易算出边缘分布 $m(x|\lambda)$ 的均值与方差，其中 $\lambda=(\mu_\pi,\sigma_\pi^2)$

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

由过去的经验，决策者认为边缘分布的均值为 10，方差为 3。即 $\hat{\mu}_m=10$，$\hat{\sigma}_m^2=3$。于是所列方程为

$$
\begin{aligned}
\mu_\pi &= 10 \\
1+\sigma_\pi^2 &= 3
\end{aligned}
$$

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

§3.4　无信息先验分布

一、贝叶斯假设

二、位置—尺度参数族的无信息先验

三、用Fisher信息阵确定无信息先验

一、贝叶斯假设

1.贝叶斯假设的**基本含义**

无信息先验分布应选取在θ取值范围Θ内的均匀分布，即：

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

c为一个容易确定的常数。这种看法被称为**贝叶斯假设**，又称**拉普拉斯先验**。

说明：贝叶斯假设在很多情况下都是合理的。

例如：

（1）对一位孕妇将拥有一个男孩或是一个女孩的概率为$1/2$，这样的均匀先验显然没有提供任何信息的。

一般地，若 $\Theta=\{\theta_1,\theta_2,\cdots,\theta_n\}$为有限集，且对$\theta_i$ 发生无任何信息，那么很自然认为其上的均匀分布

$$P(\theta=\theta_i)=1/n$$

作为θ 的先验分布是合理的。（离散型）

说明：贝叶斯假设在很多情况下都是合理的。

例如：

（2）对一种新产品的市场占有率θ 没有任何信息场合下，取 $(0,1)$上的均匀分布作为其先验分布有时也是合理的。

一般地，若 $\Theta=(a,b)$ 为有限区间，且对其内任何值都没有偏爱，因此用区间(a,b) 上的均匀分布作为θ 的先验分布也是恰当表达了对θ 的一种认识。 （连续型）

2. 应用贝叶斯假设时所出现的问题

(1)当 θ 的取值范围为**无限区间**时，如为 $(0,\infty)$ 或 $(-\infty,\infty)$ 时，就无法在 Θ 上定义一个正常的均匀分布。

例 3.4.1 设总体 $X \sim N(\theta,1)$，其中 $\theta \in (-\infty,\infty)=\Theta$。若正态均值 θ 的取值无任何偏爱的话，那么 θ 的无信息先验应是 $(-\infty,\infty)$ 上的均匀分布，即

$$\pi(\theta)=c,\ -\infty<\theta<\infty.$$

但这不是一个正常的概率密度函数，因为 $\int_{-\infty}^{\infty}\pi(\theta)d\theta=\infty$。可用它并不影响后验分布的计算。按贝叶斯公式有

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x \mid \theta)\pi(\theta)d\theta}$$

$$= \frac{\frac{c}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x-\theta)^2\right\}}{\int_{-\infty}^{\infty} \frac{c}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x-\theta)^2\right\}d\theta}$$

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

可以看出，所得的后验密度（$N(x,1)$的密度）是一个正常的概率密度。且c的选择并不重要。以后常选用$\pi(\theta)=1$。若取后验均值估计θ，则$\hat{\theta}=x$。这与经典方法的结果是一样的。贝叶斯统计学家为了把这种不正常的均匀分布纳入先验分布行列，特地给出如下的广义先验分布概念。

定义3.4.1

设总体 $X \sim p(x|\theta), \theta \in \Theta$ 。若 θ 的先验分布 $\pi(\theta)$ 满足下列条件：

- ① $\pi(\theta) \geq 0$ ， 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$,
- ②由此决定的后验密度 $\pi(\theta|x)$ 是正常的密度函数，

则称 $\pi(\theta)$ 为 θ 的**广义先验密度，或广义先验分布**。

(2)贝叶斯假设不满足变换下的不变性。

例 3.4.2 考虑正态标准差 σ，它的参数空间是$(0,\infty)$。若定义一个变换

$$\eta=\sigma^2\in(0,\infty)$$

则 η 是正态方差。在$(0,\infty)$上，η 与 σ 是一对一变换，不会损失信息，也不会产生新的信息。若 σ 是无信息参数，那么 η 也是无信息参数，且它们的参数空间都是$(0,\infty)$，没有被压缩或放大。按贝叶斯假设，它们的无信息先验分布应都为常数，应该成比例。可是按概率运算法则并不是这样。若设 $\pi(\sigma)$ 为 σ 的密度函数，那么 η 的密度函数为

$$\pi^*(\eta)=\left|\frac{d\sigma}{d\eta}\right|\pi(\sqrt{\eta})=\frac{1}{2\sqrt{\eta}}\pi(\sqrt{\eta})$$

因此，若 σ 的无信息先验被选为常数，为保持数学上的逻辑推理的一致性，$\eta=\sigma^2$ 的无信息先验应与 $\eta^{-1/2}$ 成比例。这与贝叶斯假设矛盾。

二、位置-尺度参数族的无信息先验

定义：设密度函数中有两个参数μ 与σ，且密度具有下述形式：

$$p(x;\ \mu,\sigma)=\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right),\quad \mu\in(-\infty,\infty),\sigma\in(0,\infty)$$

其中$f(x)$是一个完全确定的函数，它相应于 $\mu=0$ ，$\sigma=1$时的密度，μ 称为**位置参数**，σ 称为**尺度参数**，这类分布族称为**位置-尺度**参数族。如正态分布、指数分布、均匀分布等都属于这一类。

- 特别 $\sigma=1$ 时称为**位置参数族**，而 $\mu=0$ 时称为**尺度参数族**。

（一） 位置参数的无信息先验

定理：位置参数族的先验分布可用贝叶斯假设作为无信息先验分布。

例 3.4.3 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。其充分统计量 $\bar{x}$ 的分布为 $N(\theta,\sigma^2/n)$。即

$$p(\bar{x}\mid\theta)\propto\exp\{-n(\bar{x}-\theta)^2/2\sigma^2\}$$

关于 θ 没有任何先验信息可利用时，为估计 θ 只能采用无信息先验。

$$\pi(\theta)=1$$

利用贝叶斯公式，很容易看到，在给定 $\bar{x}$ 后，θ 的后验分布为 $N(\bar{x},\sigma^2/n)$。这表明：θ 的后验期望估计 $\hat{\theta}_E=\bar{x}$，后验方差为 σ^2/n，标准差为 $\sigma/\sqrt{n}$。这些结果与经典统计中常用结果在形式上是完全一样的。

例3.4.5 设X是从正态总体$N(\theta,\sigma^2)$ 抽取的容量为1的样本，其中σ^2 已知，θ 未知，但知其为正，试求参数θ的估计。

解：这是一种带约束条件的估计问题，用贝叶斯方法解决这类问题比较容易。取参数θ 的无信息先验分布为$(0,\infty)$上的均匀分布，即：

$$\pi(\theta)=I_{(0,\infty)}(\theta)$$

由此可得后验密度：

$$\pi(\theta\,|\,x)=\frac{\exp\left\{-(\theta-x)^2/2\sigma^2\right\}I_{(0,\infty)}(\theta)}{\int_0^\infty\exp\left\{-(\theta-x)^2/2\sigma^2\right\}d\theta}$$

若取后验均值作为θ 的估计，则：

$$\hat{\theta}_E = E(\theta \mid x) = \int_0^\infty \theta \pi(\theta \mid x) d\theta$$

$$= \frac{\int_0^\infty \theta \exp\left\{-(\theta - x)^2 / 2\sigma^2\right\} d\theta}{\int_0^\infty \exp\left\{-(\theta - x)^2 / 2\sigma^2\right\} d\theta}$$

$$\overset{\eta=(\theta-x)/\sigma}{=} \frac{\int_{-\frac{x}{\sigma}}^\infty (\sigma\eta + x) e^{-\eta^2/2} \sigma d\eta}{\int_{-\frac{x}{\sigma}}^\infty e^{-\eta^2/2} \sigma d\eta}$$

$$= x + \frac{(2\pi)^{-1/2} \sigma \exp\left\{-x^2 / 2\sigma^2\right\}}{1 - \Phi(-x/\sigma)}$$

特例：

- 如设 σ=1，若 x=0.5，则

$$\hat{\theta}_E = 0.5 + \frac{(2\pi)^{-1/2}\exp(-1/8)}{1-\Phi(-0.5)} = 0.5 + \frac{0.3520}{0.6915} = 1.0090$$

- 若 x=-0.5，则

$$\hat{\theta}_E = 0.5 + \frac{(2\pi)^{-1/2}\exp(-1/8)}{1-\Phi(0.5)} = 0.5 + \frac{0.3520}{0.3085} = 0.6410$$

- 说明：无论观察值 x 为正或为负，参数 θ 的后验期望估计总为正。这是符合要求的。

(二) 尺度参数的无信息先验

定理 设总体 X 的密度函数具有形式：

$$p(x;\ \sigma)=\frac{1}{\sigma}p\left(\frac{x}{\sigma}\right),\ \ \sigma\in(0,\infty)$$

则参数 σ 的无信息先验分布为

$$\pi(\theta)={}^{1}\!/_{\sigma},\ \ \sigma>0$$

设想让 X 改变比例尺，即得 $Y=cX(c>0)$。类似地定义 $\eta=c\sigma$，即让参数 σ 同步变化，不难算出 Y 的密度函数为 $\frac{1}{\eta}p\left(\frac{y}{\eta}\right)$ 仍属尺度参数族。且若 X 的样本空间为 R^1，则 Y 的样本空间也为 R^1；若 X 的样本空间为 R^+，则 Y 的样本空间也为 R^+。此外 σ 的参数空间与 η 的参数空间都为 R^+，可见 (X,σ) 问题与 (y,η) 问题的统计结构完全相同，故 σ 的无信息先验 $\pi(\sigma)$ 与 η 的无信息先验 $\pi^*(\eta)$ 应相同，即

$$\pi(\tau)=\pi^*(\tau) \tag{3.4.5}$$

另一方面，由变换 $\eta=c\sigma$ 可以得 η 的无信息先验

$$\pi^*(\eta)=\left|\frac{d\sigma}{d\eta}\right|\pi(\sigma)=\frac{1}{c}\pi\left(\frac{\eta}{c}\right) \tag{3.4.6}$$

比较(3.4.5)和(3.4.6)可得

$$\pi(\eta)=\frac{1}{c}\pi\left(\frac{\eta}{c}\right)$$

取 $\eta=c$，则有

$$\pi(c)=\frac{1}{c}\pi(1)$$

为方便计算，令 $\pi(1)=1$，可得 σ 的无信息先验为

$$\pi(\sigma)=\sigma^{-1},\sigma>0 \qquad (3.4.7)$$

这仍是一个不正常先验，在很多场合它可成为广义先验。

例 3.4.5 设 X 服从指数分布，其密度函数为

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

其中 σ 是尺度参数。若 $x_1, \cdots, x_n$ 是来自该指数分布的一个样本，σ 的先验取无信息先验(3.4.7)，则在样本 $x=(x_1, \cdots, x_n)$ 给定下，σ 的后验密度函数为

$$\pi(\sigma \mid \boldsymbol{x}) \propto = \sigma^{-(n+1)}\exp\left\{\sum_{i=1}^{n} x_i/\sigma\right\}, \sigma > 0$$

这是倒伽玛分布 $IGa(n, \sum_{i=1}^{n} x_i)$。它的后验均值 $E(\sigma \mid x) = \sum_{i=1}^{n} x_i(n-1)$。这就是 σ 的后验期望估计，它的后验方差为 $\mathrm{Var}(\sigma \mid x) = (\sum_{i=1}^{n} x_i)^2/(n-1)^2(n-2)$。

三、用Fisher信息阵确定无信息先验

1.确定无信息先验的更一般方法(Jeffreys(1961)):

设**x**=(x_1,x_2,…,x_n)是来自密度函数p(x|θ)的一个样本，θ 为p维参数向量，则可用Fisher信息阵的平方根作为θ 的无信息先验分布。

2.寻求分布的一般步骤:

(1)写出样本的对数似然函数: $l(\theta \mid x) = \ln\left[\prod_{i=1}^{n} p(x_i \mid \theta)\right] = \sum_{i=1}^{n} \ln p(x_i \mid \theta)$

(2)求样本的信息阵: $I(\theta) = E^{x|\theta}\left(-\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\right),\ i,j = 1,2,\cdots,p$

特别在单参数的情形: $I(\theta) = E^{x|\theta}\left(-\frac{\partial^2 l}{\partial\theta^2}\right)$

(3)θ 的无信息先验密度为: $\pi(\theta) = \left[\det I(\theta)\right]^{1/2}$

其中detI(θ)表示p×p阶信息阵I(θ)的行列式。

例 3.4.8 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_m)$服从多项分布，

$$P(\boldsymbol{X}=\boldsymbol{x})=c\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_m^{x_m},x_i\geqslant 0,\sum_{i=1}^{n}x_i=n$$

其中 $c=n!\ /(x_1!\ x_2!\ \cdots x_m!)$，$\theta_i\geqslant 0$，$\sum_{i=1}^{m}\theta_i=1$。现要寻求$(\theta_1,\cdots,\theta_m)$的无信息先验分布。

首先写出对数似然函数

$$l=\sum_{i=1}^{m}x_i\ln\theta_i+\ln c=\sum_{i=1}^{m-1}x_i\ln\theta_i+(n-\sum_{i=1}^{m-1}x_i)\ln(1-\sum_{i=1}^{m-1}\theta_i)+\ln c$$

然后计算 Fisher 信息阵的每个元素

$$\frac{\partial^2 l}{\partial\theta_i^2}=-\frac{x_i}{\theta_i^2}-\frac{x_m}{\theta_m^2},i=1,2,\cdots,m-1$$

$$\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}=-\frac{x_m}{\theta_m^2},i,j=1\cdots,m-1,i\neq j$$

$$E\left(-\frac{\partial^2 l}{\partial\theta_i^2}\right)=\frac{n}{\theta_i}+\frac{n}{\theta_m},i=1,2\cdots m-1$$

$$E\left(-\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)=\frac{n}{\theta_m},i,j=1,\cdots,m-1,i\neq j$$

故

$$\det I(\boldsymbol{\theta})=\begin{vmatrix} n(\theta_1^{-1}+\theta_m^{-1}) & n\theta_m^{-1} & \cdots & n\theta_m^{-1} \\ n\theta_m^{-1} & n(\theta_2^{-1}+\theta_m^{-1}) & \cdots & n\theta_m^{-1} \\ \cdots & \cdots & \cdots & \cdots \\ n\theta_m^{-1} & n\theta_m^{-1} & \cdots & n(\theta_{m-1}^{-1}+\theta_m^{-1}) \end{vmatrix} \propto (\theta_1\theta_2\cdots\theta_m)^{-1}$$

从而$(\theta_1,\theta_2,\cdots,\theta_m)$的 Jeffreys 先验为

$$\pi(\theta_1,\theta_2,\cdots,\theta_m)\propto(\theta_1\theta_2\cdots\theta_m)^{-1/2}$$

例 3.4.9 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$是来自正态分布 $N(\mu,\sigma^2)$的一个样本。现求参数向量(μ,σ)的 Jeffreys 先验。

在例 3.4.7 已算得正态总体下的 Fisher 信息阵 $I(\mu,\sigma)$从而可得其样本分布的 Fisher 信息阵为

$$I_n(\mu,\sigma)=\begin{bmatrix} E\left(-\frac{\partial^2 l}{\partial\mu^2}\right) & E\left(-\frac{\partial^2 l}{\partial\mu\partial\sigma}\right) \\ E\left(-\frac{\partial^2 l}{\partial\mu\partial\sigma}\right) & E\left(-\frac{\partial^2 l}{\partial\sigma^2}\right) \end{bmatrix}=\begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

$$|I_n(\mu,\sigma)|=2n^2\sigma^{-4})$$

所以(μ,σ)的 Jeffreys 先验为

$$\pi(\mu,\sigma)\propto\sigma^{-2}$$

它的几个特例是：

(1)当 σ 已知时，$I(\mu)=E\left(\frac{\partial^2 l}{\partial\mu^2}\right)=\frac{n}{\sigma^2}$，故 $\pi(\mu)=1, \mu\in R^1$。

(2)当 μ 已知时，$I(\sigma)=E\left(-\frac{\partial^2 l}{\partial\sigma^2}\right)=\frac{2n}{\sigma^2}$，故 $\pi(\sigma)=1/\sigma, \sigma\in R^+$。

(3)当 μ 与 σ 独立时，$\pi(\mu,\sigma)=1/\sigma, \mu\in R^1, \sigma\in R^+$。

可见 Jeffreys 先验表明：μ 与 σ 的无信息先验分布是不独立的。在 (μ,σ) 的联合无信息先验分布的两种形式（σ^{-1} 与 σ^{-2}）中，Jeffreys 最终推荐的是 $\pi(\mu,\sigma)=1/\sigma$，从实际的使用情况看，多数人采用 Jeffreys 的最终推荐。对此后面还有进一步评论。

例 3.4.10 设 θ 为成功概率，则在 n 次独立试验中成功次数 X 服从二项分布，即

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

其对数似然函数为

$$l=x\ln\theta+(n-x)\ln(1-0)+\ln\binom{n}{x}$$

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

$$I(\theta)=E\left(-\frac{\partial^2 l}{\partial\theta^2}\right)=\frac{n}{\theta}+\frac{n}{1-\theta}=\frac{n}{\theta(1-\theta)}$$

所以，在二项分布场合，成功概率 θ 的 Jeffreys 先验为

$$\pi(\theta)\propto\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}},\theta\in(0,1)$$

关于成功概率为 θ 的无信息分布，不少统计学家从各种角度探讨，至今已有如下四种。

$\pi_1(\theta)=1$ ——Bayes(1763)和 Laplace(1812)采用过

$\pi_2(\theta)=\theta^{-1}(1-\theta)^{-1}$ ——Novick 和 Hall(1965)导出

$\pi_3(\theta)=\theta^{-1/2}(1-\theta)^{-1/2}$ ——Jeffreys(1968)导出

$\pi_4(\theta)=\theta^{\theta}(1-\theta)^{1-\theta}$ ——Zellner(1977)导出

所有这四种无信息先验都是合理的,因为它们各自从一个侧面提出自己的合理要求,然后再推导出来的。其中 π_2 是不正常密度,而 π_1 是正常密度,而 π_3,π_4 经过正则化后可成为正常密度。这四个先验虽不同但对贝叶斯统计推断的结果的影响是很小的,故都可使用。

§3.5 多层先验

一、多层先验

二、多层模型