贝叶斯统计推断

张尧庭　陈汉峰　编著

科学出版社

（京）新登字 092 号

内 容 简 介

本书重点介绍了贝叶斯统计推断的理论、方法及其基本观点，书中对贝叶斯方法和经典方法在历史上的重大分歧也予以介绍．主要内容有先验分布的选取、估计及检验、贝叶斯方法在可靠性统计分析中的应用、线性模型及多元分析、经验贝叶斯方法等．为使读者理解本书的内容，每章后面附有练习思考题．

本书可供大学高年级学生、研究生、从事应用统计工作的工作者学习参考．

贝叶斯统计推断

张尧庭　陈汉峰　编著

责任编辑　毕　颖

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1991 年 3 月第 一 版　　开本：787×1092　1/32

1994 年 6 月第二次印刷　　印张：7 5/8

印数：1801—2900　　字数：170 000

ISBN 7-03-002027-8/O·384

定价：7.20 元

序　　言

贝叶斯学派是数理统计中一个重要的学派，它有鲜明的特点和独到的处理方法，在国际上贝叶斯学派与非贝叶斯学派争论是很多的.在有些国家,争论相当激烈. 在国外，介绍贝叶斯学派理论、方法以及研究工作的书很多，但在国内还没有出版过一本这方面的书. 本书的目的就是想填补这一空缺.

要全面介绍贝叶斯学派的观点、理论和方法是有困难的，对着眼于应用、解决实际问题的人来说,更重要的是贝叶斯具有特色的一些处理方法以及相应的理论. 本书正是从这样一个角度来选材的. 因此，凡是一般教材中都有的概念、定理，我们都不加证明地予以引用，但这一部分内容是不多的. 所以本书的写法中有一个特点，经常把贝叶斯方法所得的结果与经典方法,即通常教科书中的方法所得的结果加以比较,以便更容易了解、领会贝叶斯方法的特点. 如果对经典方法的结果并不了解，遇到这样的比较,就无法理解了,所以要求读者先要具备一定的概率统计的知识.

本书的第一章先对贝叶斯方法和经典方法在总的看法上有何重大的分歧予以介绍,让读者了解观点的不同,各自在理论上的缺陷和长处，使读者有一个总的了解. 在这一章不能不涉及到对于“概率”、“统计”这些基本概念的理解以及不同的哲学的解释,我们不深入展开这一类问题，只是客观地、扼

要地介绍一下，帮助读者理解以后的内容．为了让读者更好了解贝叶斯学派与经典学派的争论，我们还简要地介绍了一下信念学派的观点，让读者自己比较．

第二、三章是贝叶斯学派基本的出发点和基本的方法，也是全书的基础．由于确定先验分布是贝叶斯学派的重要内容，引起的争议也最多，所以单写一章专门讨论．

第四章是贝叶斯方法在可靠性统计分析中的应用，这一方向现在正越来越受到国际上的重视，这一章只能是作为一个入门的材料．

数理统计涉及的内容很多，贝叶斯方法都可以从自己的理论出发去导出相应的结果，这样每一个方向都可以写成一章，篇幅就太大了，而且反映不出贝叶斯的特色．本书选了线性模型与多元分析这个方向，单独写一章．一是因为在应用上这两个内容遇到的机会多；二是两者可以合并在一起统一处理，内容本身的联系较多；三是在这方面，具有特色的贝叶斯方法也多一点．不过这一章要求读者有较好的线性代数基础，最好已看过或学过多元统计分析的内容．线性代数基础较差的读者可以跳过这一章，进入第六章的内容．贝叶斯方法在其他统计方面的应用，我们在第五章的附录内作了一点介绍．

第六章是经验贝叶斯方法．这一内容本身就可以写成一本书，我们选取了其中最实用的部分予以介绍，希望能对实际应用统计工作的读者有真正的帮助．

最好能再写一章在经济方面的应用，由于编者的水平，以及实际工作经验的限制，写这样一章是困难的，希望将来有机会修改再版时能写出这样一章．

全书写了不少说明性的数字的例子，有意识地安排了几个比较实际而数据又不算太少的例子，无非是想说明一个问

题的真正解决需要对数据作认真的分析，能从头贯彻到底．例子宁可少举，但求能真正说明一些问题．限于编者实际工作的经验，自己感到实例还是少了一些．

编者本人并不是贝叶斯派，但也不是非贝叶斯派．我们认为每一种统计方法都有它自己适用的范围，既毋需夸大，也不应该缩小．因此，编写过程中力求实事求是地、客观地介绍出贝叶斯学派的工作．当然，无论从选材、系统的选取、叙述的方法都反映了编者的一些观点．本书力求从贝叶斯学派的观点去观察统计问题，有些定义和定理就和一般书上的写法不同，如充分统计量的定义、最小二乘法的结果……，书中这样的写法我们认为也许更能体现贝叶斯方法的特点．书中有些内容是我们自己的工作，也是一些实际工作中遇到的问题．

本书的编写，前后经历了五年，修改了几次．第一稿大部分是陈汉峰同志写的，后来征求了一些专家的意见．第二稿是张尧庭同志写的，把系统作了较大的修改，写法也不同，比较注意了实用的内容．第三稿是根据第二稿的写法，按照在一些学校、讨论班上试讲的情况，又作了修改，补充了一些实际应用的例子，突出了贝叶斯方法的应用．这一不断修改的过程，反映了我们对贝叶斯方法的认识也在不断变化．魏宗舒先生曾看过前二章的内容，提出了不少宝贵的意见；茆诗松同志对全书提出了不少建议，并帮我们改正了一些错误，在这里我们表示衷心的感谢．本书的部分内容在一些学校、培训班上讲过几次，有的教师、学生提出了很好的修改意见，指出了一些问题，对我们有切实的帮助，借此机会也向这些同志一并致谢．

我们希望这本书的出版能对高年级大学生、研究生、广大的应用统计工作者有真正的帮助，诚恳地希望读者看后提出

批评和建议，使我们能够得到改进的机会.

张尧庭 陈汉峰

1989年4月

目　　录

第一章 引　论

第一节 历史概述

贝叶斯（Bayes, T. R.）学派奠基性的工作是贝叶斯的论文[1]. 可能是他自己感到了他的学说还有不完善的地方，这一论文在他生前没有发表，而是在他死后由他的朋友发表的. 著名的数学家拉普拉斯（Laplace, P. S.）用贝叶斯提出的方法，导出了重要的“相继律”，贝叶斯的方法和理论逐渐被人理解和重视起来. 尽管贝叶斯方法可以推导出一些有意义的结果，但在理论上和实际应用中也还是出现了各种各样的问题，因而在19世纪并未被大家普遍接受. 20世纪初，意大利的菲纳特（B. de Finetti），稍后一些英国的杰弗莱（Jeffreys, H.）都对贝叶斯学派的理论作出了重要的贡献. 第二次世界大战后，瓦尔德（Wald, A.）提出了统计的决策理论，在这一理论中贝叶斯解占有重要的地位；信息论的发展也对贝叶斯学派作出了新的贡献；更重要的是在一些实际应用的领域中，尤其是在社会科学、经济商业活动中，贝叶斯方法取得了成功，贝叶斯学派已成了一股不容忽视的力量. 1958年英国历史最悠久的统计杂志 Biometrika 全文重新刊登贝叶斯的论文，这就是一个明证. 20世纪50年代，以罗宾斯（Robbins, H.）为代表，提出了经验贝叶斯方法，把贝叶斯方法和经典方法相结合，引起了统计界的广泛的注意，这一方法很快就显示出它的优点，成为很活跃的一个方向.

现在，杂志上经常可以看到贝叶斯学派的论文，并已经出

版了不少专门的著作，参考文献中的[2]—[9]是比较有代表性的书。也有将贝叶斯方法系统地用于经济(如文献[10])、用于可靠性技术(如文献[11])等更专门的书。

第二节　基本观点

贝叶斯学派的起点是贝叶斯的两项工作：贝叶斯定理和贝叶斯假设。贝叶斯定理(或贝叶斯公式)在通常概率论的教科书中都有叙述，而贝叶斯假设几乎都不提及。为了方便，我们在这里再用事件的形式和随机变量的形式来叙述，有关的证明就略去了。

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 $B \subset \bigcup_{i=1}^{k} A_i$[1)]，则有

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)}, \quad i = 1, 2, \cdots, k \tag{1.1}$$

贝叶斯公式的随机变量形式：

假定随机变量 ξ, η 的联合分布密度是 $p(x, y) = p_\xi(x) f_{\eta|\xi}(y|x)$，其中 $p_\xi(x)$ 是 ξ 的边缘密度，而 $f_{\eta|\xi}(y|x)$ 是当 $\xi = x$ 时，η 对 ξ 的条件密度，于是 ξ 对 η 的条件密度 $g_{\xi|\eta}(x|y)$ 可表示为(当 $\eta = y$)

---

1) 有的书上用"$\bigcup_{i=1}^{k} A_i$ 是一必然事件"代替"$B \subset \bigcup_{i=1}^{k} A_i$"，两者略有区别，参看本章附录。

$$g_{\xi|\eta}(x|y)=\frac{p_\xi(x)f_{\eta|\xi}(y|x)}{\int_{-\infty}^{\infty}p_\xi(x)f_{\eta|\xi}(y|x)dx} \tag{1.2}$$

类似地,有

$$f_{\eta|\xi}(y|x)=\frac{q_\eta(y)g_{\xi|\eta}(x|y)}{\int_{-\infty}^{\infty}q_\eta(y)g_{\xi|\eta}(x|y)dy} \tag{1.3}$$

其中 $q_\eta(y)$ 是 y 的边缘分布密度.

值得注意的是,只要将 ξ 与 η 的位置互换一下,相应的 x 与 y 的位置也随着互换,从(1.2)立即就得到(1.3),因此这两个公式实质上是一样的.如将(1.2),(1.3)中的随机变量 ξ 和 η,改为随机向量,$p_\xi(x)$,$q_\eta(y)$ 仍然是 ξ,η 各自的边缘密度,$f_{\eta|\xi}(y|x)$,$g_{\xi|\eta}(x|y)$ 是相应的条件密度,变量 x,y 与随机向量 ξ,η,相应的也是向量,贝叶斯公式(1.2)与(1.3)仍然成立.

(1.3) 式的右端只需要知道 η 的边缘密度和 ξ 对 η 的条件密度.对于离散的随机向量,只要将分母中的积分号改为求和号,(1.3)式还是成立的.将(1.3)式与统计中的问题相联系,就可说明贝叶斯假设,下面用例子来说明这一点.

一个人打靶,打了 n 次,命中了 r 次,现在问此人打靶命中的概率 θ 应如何估计?从通常的教科书或单凭直觉,都知道用 $\frac{r}{n}$ 去估计 θ,但是这一估计有它的不合理处.例如对 $n=r=1$,估计 $\hat{\theta}=1$,而 $n=r=100$ 时还是 $\hat{\theta}=1$;对 $n=100$,$r=0$ 得 $\hat{\theta}=0$,对 $n=1$,$r=0$,$\hat{\theta}$ 也是 0.打了 100 次,每次都命中了,直觉上总感到此人命中的概率相当大;打了 1 次,命中了,这个人命中的概率不能和 100 次每次都中的一样,而 $\frac{r}{n}$ 这一估计量给出的结果是一样的.

从概率论的独立试验序列知道，已知某人打靶命中的概率是 θ，则打靶 n 次命中恰为 r 次的概率是

$$\binom{n}{r}\theta^r(1-\theta)^{n-r}$$

其中 $\binom{n}{r}$ 是二项系数(也就是组合数 C_n^r. 如果把 θ 看作一个随机变量，上面的概率就可以看成当 θ 已知时，r 对 θ 的条件概率. 要从样本值 r 去估计 θ，正好是与已知 r 时，θ 对 r 的条件密度有关. 用 $g(r|\theta)$ 表示 $\binom{n}{r}\theta^r(1-\theta)^{n-r}$，即 r 对 θ 的条件概率，如果还知道 θ 的边缘密度 $q(\theta)$，则由(1.3)的贝叶斯公式，就可求出 θ 对 r 的条件密度

$$f(\theta|r)=\frac{q(\theta)\binom{n}{r}\theta^r(1-\theta)^{n-r}}{\int_0^1 q(\theta)\binom{n}{r}\theta^r(1-\theta)^{n-r}d\theta} \tag{1.4}$$

上式中 $q(\theta)$ 与试验的结果无关. 它反映了在进行统计试验前，人们对于命中概率 θ 的知识，所以称它为先验分布(或验前分布、事先分布等). (1.4) 式的 $f(\theta|r)$ 综合了先验分布 $q(\theta)$，试验结果 r 带来的关于 θ 的信息(反映在 $g(r|\theta)$ 中)，它是与统计试验结果 r 有关的，称为后验分布(或验后分布、事后分布等). 如果我们对打靶的人不了解，他的命中概率 θ 在[0,1]中取哪个值都是同样可能的，这时先验分布 $q(\theta)$ 就是[0,1]上的均匀分布，即

$$q(\theta)=\begin{cases}1, & \text{当 } \theta\in[0,1]\\ 0, & \text{当 } \theta\in[0,1]\end{cases} \tag{1.5}$$

把(1.5)式的 $q(\theta)$ 代入(1.4)，就求出

$$f(\theta|r)=\frac{\theta^r(1-\theta)^{n-r}}{\int_0^1\theta^r(1-\theta)^{n-r}d\theta},\quad 0\leqslant\theta\leqslant 1$$

也即 $f(\theta|r)$ 是贝塔分布，上式右端的分母就是贝塔函数 $B(r+1,\ n-r+1)$. 如用后验分布的期望值即 θ 对 r 的条件期望 $E\{\theta|r\}$ 去估计 θ，就得估计量

$$\hat{\theta}=E\{\theta|r\}=\frac{1}{B(r+1,n-r+1)}\int_0^1\theta\cdot\theta^r(1-\theta)^{n-r}d\theta$$

$$=B(r+2,n-r+1)/B(r+1,n-r+1)$$

$$=\frac{r+1}{n+2}$$

当 $n=r=1$ 时，$\hat{\theta}=\frac{1+1}{1+2}=\frac{2}{3}$；当 $n=r=100$ 时，$\hat{\theta}=\frac{100+1}{100+2}=\frac{101}{102}$，显然这个估计比 $\frac{r}{n}$ 要合理. $\hat{\theta}=\frac{r+1}{n+2}$ 就是“相继律[1]”. 把上面处理这一问题的过程总结一下，就可以看出贝叶斯方法的梗概.

（1）将未知参数看成随机变量(或随机向量)，记它为 θ，于是当 θ 已知时，样本 $x_1,\cdots,x_n$ 的联合分布密度 $p(x_1,\cdots,x_n;\theta)$ 就看成是 $x_1,\cdots,x_n$ 对 θ 的条件密度，记为 $p(x_1,\cdots,x_n|\theta)$，或简写为 $p(x|\theta)$.

（2）设法确定先验分布 $\pi(\theta)$. 这是根据以往对参数 θ 的知识来确定的，是贝叶斯方法中容易引起争议的一步.

（3）利用条件分布密度 $p(x_1,\ \cdots,\ x_n|\theta)$ 和先验分布 $\pi(\theta)$，可以求出 $x_1,\cdots,x_n$ 与 θ 的联合分布和样本 $x_1,\cdots,x_n$ 的分布，于是就可用它们求得 θ 对 $x_1,\ \cdots,\ x_n$ 的条件分布密度，也就是用贝叶斯公式求得后验分布密度 $h(\theta|x_1,\cdots,$

1）从 n 次试验的结果，A 出现了 r 次，推测下一次 A 出现的可能性，即相继的“A 发生”的可能性，这样就被称作相继律.

x_n).

(4) 利用后验分布密度 $h(\theta|x_1,\cdots,x_n)$ 作出对 θ 的推断(估计 θ 或对 θ 作检验).

在(2)中,如果没有任何以往的知识来帮助我们确定先验分布 $\pi(\theta)$, 贝叶斯提出可以采用均匀分布作为 $\pi(\theta)$, 即参数在它变化的范围内,取到各个值的机会是相同的,这种确定先验分布的原则,就称为贝叶斯假设.

贝叶斯假设在直觉上易于为人们所接受, 然而进一步探讨它的确切含义时,就产生了各种不同的理解,这一方面的问题将在第二章给以较详细的讨论.

第三节 对贝叶斯学派的批评

贝叶斯方法在统计推断中出现很早,已有相当长的历史,然而从它诞生的时刻起,直到现在,不断受到反对者的种种批评.

批评集中于两点:

(1) 参数 θ 看成是随机变量是否妥当?

(2) 先验分布是否存在? 如何选取?

在有些情况下, θ 看成随机变量是合理的,它的先验分布也容易确定. 例如大批生产的某一种零件,次品率是 p, 将这种零件装箱出售, 每箱是 2000 个, 显然各箱的次品数 θ 就是一个随机变量. 一箱中恰有 θ 个次品的概率遵从二项分布 $\binom{2000}{\theta}p^{\theta}(1-p)^{2000-\theta}$, 现在从一箱中抽检 10 个, 发现有 r 件次品,要问这一箱中的次品数应如何估计, 也即要估计 θ. 这时 θ 已知时, r 的分布是

$$P(r=k)=\binom{\theta}{k}\binom{2000-\theta}{10-k}\bigg/\binom{2000}{10}$$

θ 的先验分布是 $\binom{2000}{\theta}p^{\theta}(1-p)^{2000-\theta}$，大批生产时的次品率 p 可从过去的生产记录中求得. 在这种情况下，对(1)、(2)两点是不会有疑问的.

但是如果象在第二节中的打靶问题，对某人的打靶技术事先一无所知，只能凭 n 次打靶的结果来估计. 此时把每次命中的概率 θ 看成是随机变量，似乎有些勉强. 然而进一步考虑一下，正因为对每次命中的概率没有任何知识，它在 0 与 1 之间取哪一个值的可能性全是相同的，这种可能性是由于人们对此特定事物的了解不够而形成的，这样也就可以理解 θ 这个参数是不确定的，它取各个不同的值有相同的机会，也就是 θ 可以看作是随机变量.

在讨论一些理论问题时，假定先验分布密度 $\pi(\theta)$ 是已知的，这是完全可以的，然而在实际工作中，有时对参数 θ 是没有任何过去的知识可以借鉴，而是希望通过试验结果来获得，这时的先验分布称为无信息先验分布（Non-informative Priors）. 如何去获得无信息先验分布，这是贝叶斯方法的一个重大的理论问题. 这里无信息一词的含义是确切的，它指的是没有任何信息(指在未进行试验时，或认为不考虑样本提供的信息)可以帮助我们去选用一个特定的分布作为先验分布，并非对参数 θ 其他的情况也一无所知. 我们至少知道两点：参数 θ 与样本 $x_1, \cdots, x_n$ 联合分布密度的关系，经典方法认为已知 $x_1, \cdots, x_n$ 的联合密度是 $p(x_1, \cdots, x_n; \theta)$，从贝叶斯方法的观点来看，就是已知 $x_1, \cdots, x_n$ 对 θ 的条件密度；其次我们也知道 θ 的取值范围，例如正态分布 $N(\mu, \sigma^2)$ 的两个参数，μ 的取值范围是 $(-\infty, \infty)$，σ^2 的范围是 $(0, \infty)$.

贝叶斯提出一个原则，无信息先验分布应选取在 θ 取值范围内的均匀分布，这一原则就是贝叶斯假设．在打靶问题中，每次命中的概率 θ 在 $[0, 1]$ 内均匀分布，这是可以接受的，但如象正态分布的两个参数 μ 与 σ^2，它们的变化范围都是无限的区间．在无限区间上，均匀分布是不存在的，因为均匀分布相应的密度是一个常数，即密度函数 $f(x) = c$，c 是一个常数，而 $\int_a^\infty f(x)dx = c\int_a^\infty dx = c \cdot \infty = \infty$，积分值不能为 1，这样贝叶斯假设对这一类问题就遇到了困难．另一方面，未知参数的选择是有些任意的，例如正态分布 $N(\mu, \sigma^2)$ 的参数 σ^2 既可以取 σ 为参数，也可以取 σ^2 为参数．如果 σ 在 $(0,\infty)$ 上是“均匀”分布的（假定它存在），则 σ^2 就不可能仍然是“均匀”的了；反之，如果 $\eta = \sigma^2$ 这个参数在 $(0, \infty)$ 上“均匀”分布，则 $\sigma = \sqrt{\eta}$ 在 $(0,\infty)$ 上就不可能“均匀”了．然而根据贝叶斯假设，无论对 σ 还是 $\eta = \sigma^2$，都应选“均匀”分布，这就是贝叶斯假设面临的另一个困难．

为了克服上述困难，近 200 年来，许多学者付出了辛勤的劳动，到这一世纪，有了很大的进展．虽然在理论上还没有一个完整的、一般性的方法，一举解决上述问题．但在实用的范围内，对一些常见的分布，都已得到了较好的回答，可以从几个不同的角度来说明取某些分布作为先验分布是合理的．我们将在第二章中详细讨论这些问题．

经验贝叶斯方法正是注意到了这些问题，设法从取得的样本中提取关于参数 θ 的信息，由样本值来决定先验分布，这些方法我们将在最后一章进行介绍．

第四节　对经典学派的批评

经典学派，也称为抽样学派，是指本世纪初由英国的卡尔

·皮尔逊（Pearson, K.）等人开始，经费歇（Fisher, R. A.）给以发展，到奈曼（Neyman, J.）完成理论的这一系统的成果. 在目前国内已出版的教材，这一部分——指经典学派的方法和理论——往往占有全部或绝大部分的篇幅，如点估计、假设检验、矩估计法、最大似然估计法、最小二乘法、估计的优良性、区间估计、假设检验中犯两类错误的概率、优良的假设检验方法、……等等. 实践证明经典学派的理论和方法是有意义的，它指导人们在许多领域中作出了重要的贡献. 然而这并不意味着它对任何问题都是适用的，更不能理解为它是独一无二的、完全正确的理论和方法. 正如贝叶斯学派有它的缺陷一样，经典学派也遭到别的学派的批评.

对经典学派的批评主要是下面两点:

一些问题的提法不妥;

判断统计方法好坏的标准不妥.

1. 问题的提法不妥

首先是对估计问题中的置信区间这一概念提出怀疑. 以正态总体的参数估计为例，假定样本 $x_1, \cdots, x_n$ 来自正态分布 $N(\mu, \sigma^2)$，如果方差 σ^2 是已知的，就可以求出期望值参数 μ 的置信区间. 它的推理过程是这样的:

对于 $x_1, \cdots, x_n$ 独立同分布 $N(\mu, \sigma^2)$,

取样本均值 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$,

于是 $\bar{x} \sim N(\mu, \sigma^2/n)$,

即 $\sqrt{n}\dfrac{\bar{x}-\mu}{\sigma} \sim N(0, 1)$.

可查标准正态 $N(0, 1)$ 的表，得双侧 α 分位点的值 u_α,