贝叶斯多元统计推断理论

朱慧明　韩玉启 著

科学出版社
www.sciencep.com

内 容 简 介

本书系统地研究了多元统计模型的贝叶斯推断理论及其在经济管理中的应用，主要内容包括矩阵正态分布、Wishart分布和多元 t 分布的基本定义及性质，参数先验分布的构造方法，多元线性模型和多重线性模型的贝叶斯推断理论，向量自回归VAR(p)预测模型的贝叶斯推断理论，以及多总体贝叶斯分类识别方法的构造理论.

本书可作为统计学、计量经济学和管理科学与工程等相关学科专业的高年级本科生、硕士生或博士生教材，也可作为高校教师、研究人员和科技人员的参考书.

图书在版编目(CIP)数据

贝叶斯多元统计推断理论/朱慧明，韩玉启著. —北京：科学出版社，2006
ISBN 7-03-016452-0

I. 贝… II. ①朱… ②韩… III. 贝叶斯推断-研究 IV. O212

中国版本图书馆CIP数据核字(2005)第132221号

责任编辑：陈玉琢 吕 虹 祖翠娥 / 责任校对：朱光光

责任印制：钱玉芬 / 封面设计：王 浩

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2006年1月第一版 开本：B5 (720 × 1000)
2006年1月第一次印刷 印张：9 1/2
印数：1—2 500 字数：175 000

定价：29.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

前　　言

贝叶斯统计推断理论源于英国学者贝叶斯(Thomas Bayes)于 1763 年在英国皇家学会哲学学报上发表的论文《论机会学说中一个问题的求解》(An Essay Towards Solving a Problem in the Doctrine of Chances)，该文提出了一种归纳推理的理论，后被一些统计学家发展成为一种系统的统计推断方法，简称为贝叶斯方法. 20 世纪 50 年代以后，通过 De Finetti、Savaga、Raiffa、Schlaifer、Jeffreys、Good 等统计学家大量开拓性的研究工作，贝叶斯统计推断理论获得了迅速发展和完善. 目前，在欧美等西方国家，贝叶斯统计已经成为了与经典统计学派并驾齐驱的当今两大统计学派之一；并且，英国统计学家 Lindely 认为 21 世纪将是贝叶斯统计的世界. 目前，贝叶斯统计理论在可靠性工程、风险管理工程、经济预测与决策及生物统计学等诸多领域中均获得了广泛的应用.

本书的目的是向读者介绍现代贝叶斯统计推断理论及其在经济管理领域中的应用，主要内容包括多元线性模型和多重线性模型的贝叶斯推断理论，贝叶斯统计质量控制图和贝叶斯过程能力指数评价模型的构造，多变量时间序列模型的贝叶斯推断理论及其在经济预测中的应用，以及多总体分类识别方法的贝叶斯推断理论. 对于书中的主要定理及结论，作者从数学上进行了严格的理论推导和证明. 阅读本书所要求的预备知识包括多元统计分析、矩阵代数、统计质量控制与诊断，以及时间序列分析基本理论.

本书主要研究线性模型、向量时间序列预测模型和多总体分类识别方法的贝叶斯统计推断理论及其应用，全书分为 7 章，具体内容安排如下：

第 1 章：绪论. 论述了贝叶斯思想的起源与发展、基本观点及其本质，比较贝叶斯方法和经典统计方法在统计推断模式的差异性，探讨贝叶斯推断方法及其应用现状.

第 2 章：多元统计分布. 将研究贝叶斯推断理论体系中三类随机矩阵的统计分布及其性质：多元正态分布和矩阵正态分布、Wishart 分布和逆 Wishart 分布、多元 t 分布和矩阵 t 分布.

第 3 章：参数先验分布. 将探讨扩散先验分布和共轭先验分布的构造，以及随机参数阵的贝叶斯风险决策解.

第 4 章：多元线性模型的贝叶斯推断理论. 研究多元线性模型的贝叶斯估计理论，包括模型参数的贝叶斯估计理论和线性假设检验、随机误差序列自相关的贝叶斯诊断和单位根检验；同时，将模型的贝叶斯理论应用于质量控制，构造贝

叶斯统计质量控制图.

第 5 章：多重线性模型的贝叶斯推断理论. 将通过对模型统计结构的研究，构造了参数的共轭先验分布，研究矩阵正态-Wishart 共轭先验分布下模型参数的后验分布及相应的贝叶斯估计、模型预报密度函数，据此构造均值向量控制图和多指标贝叶斯过程能力指数模型.

第 6 章：VAR(p)预测模型的贝叶斯推断理论. 将研究 Minnesota 共轭先验分布的结构，VAR(p)预测模型的贝叶斯推断理论及其在经济预测中的应用，包括非限制性 VAR(p)预测模型的贝叶斯推断理论、限制性 VAR(p)预测模型的贝叶斯推断理论.

第 7 章：多总体分类识别方法的贝叶斯推断理论. 将根据参数的充分统计量来研究各总体服从正态分布、分布参数的先验分布分别取为扩散先验分布和正态-逆 Wishart 共轭先验时，如何利用待判样品的预报密度函数，构造后验概率比和相应的分类识别规则，据此对待判样品进行分类.

其中，第 1 章、第 2 章由朱慧明和韩玉启合作完成，第 3~7 章由朱慧明完成. 在本书的撰写过程中，中国科学院研究生院陈希孺院士和武汉大学邹新堤教授提出了许多宝贵意见，这对提高本书的质量有很大的帮助. 在本书的出版过程中，得到了科学出版社和南京理工大学研究生院的大力支持，以及韦志辉教授、许鹏教授的关心和帮助.

本书的出版得到了中国博士后科学基金项目(20040350216)、江苏省博士后科研资助计划和国家社科基金项目(04CTJ03)的资助，作者在此表示衷心感谢！

由于作者水平所限，准确表达贝叶斯理论体系的各种观点并非易事，书中错缪之处在所难免，恳请专家、读者批评指正.

作者

2004 年 10 月

符号缩写及说明

tr	矩阵的迹
$\boldsymbol{0}$	零向量或零矩阵
$\boldsymbol{A}^{-}$	矩阵 $\boldsymbol{A}$ 的广义逆
$\propto$	正比符号
$\mathbf{R}^{+}$	区间 $(0,\infty)$
sup	上确界
inf	下确界
vec	矩阵的向量化算子
C_p	过程能力指数
$\mathbf{1}_n$	分量均为 1 的 n 维列向量
$\boldsymbol{J}_n$	各元素均为 1 的 $n\times n$ 矩阵
$\lvert\boldsymbol{A}\rvert_{+}$	矩阵 $\boldsymbol{A}$ 行列式的绝对值
$\pi(\theta)$	参数 θ 的先验分布
$\pi(\theta\mid\cdot)$	参数 θ 的后验分布
i.i.d	相互独立同分布
rank($\boldsymbol{A}$)	矩阵 $\boldsymbol{A}$ 的秩
$\boldsymbol{A}\otimes\boldsymbol{B}$	矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{B}$ 的 Kronecker 积
$X\overset{\text{d.f}}{=}Y$	随机变量 X 与 Y 服从相同的统计分布
$J(\boldsymbol{A}\to\boldsymbol{B})$	矩阵 $\boldsymbol{A}$ 到矩阵 $\boldsymbol{B}$ 变换的 Jacobian 行列式
$o(n^{-1/2})$	$\lim\limits_{n\to\infty} o(n^{-1/2})/n^{-1/2}=0$
$\boldsymbol{X}\sim Mt_k(\upsilon,\boldsymbol{\theta},\boldsymbol{P})$	k 维随机向量 $\boldsymbol{X}$ 服从多元 t 分布
$\boldsymbol{X}\sim Mt_{k_1\times k_2}(\upsilon,\boldsymbol{\theta},\boldsymbol{\Sigma}\otimes\boldsymbol{Q})$	$k_1\times k_2$ 随机矩阵 $\boldsymbol{X}$ 服从矩阵 t 分布
$\boldsymbol{X}\sim MN_{k_1\times k_2}(\boldsymbol{\theta},\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$	$k_1\times k_2$ 随机矩阵 $\boldsymbol{X}$ 服从矩阵正态分布
$\boldsymbol{X}\sim W_m(\upsilon,\boldsymbol{\Sigma})$	$m\times m$ 随机矩阵 $\boldsymbol{X}$ 服从 Wishart 分布
$\boldsymbol{X}\sim IW_m(\upsilon,\boldsymbol{\Sigma})$	$m\times m$ 随机矩阵 $\boldsymbol{X}$ 服从逆 Wishart 分布

目　　录

第 1 章 绪　论

1.1 引　言

在国际统计学术界中有贝叶斯统计和经典统计两大学派，这两个学派之间长期存在争论，至今也没有定论. 事实上，这两个学派的争论构成了现代数理统计发展过程的一个特色. 英国统计学家 Lindely 认为 21 世纪将是贝叶斯统计的世界，而 Efron 认为出现这个局面的可能性只有 15%. 但无论如何，这两个学派学者都认为：这场争论对现代统计理论的发展将起到积极的促进作用.

贝叶斯理论源于英国学者贝叶斯于 1763 年在皇家学会学报上发表的论文《论机会学说中一个问题的求解》(An Essay Towards Solving a Problem in the Doctrine of Chances)，该文提出了从二项分布的观察值出发对其参数进行概率推断的方法. 一般认为，贝叶斯假定参数是单位区间上的均匀分布，他所提出的有关二项分布参数推断的方法后来被称为贝叶斯定理，并且被推广到二项分布以外的应用之中以及任何统计分布.

从某种意义上讲，贝叶斯方法中的许多思想，可以追溯到 1713 年 Bernoulli 的工作，在其概率专著《推断的艺术》(Ars Conjectandi)一书中，Bernoulli 不仅发展了二项式定理，提出了交换律和结合律，而且还提出了贝叶斯逆概率问题. 但是，尽管如此，Bernoulli 并没有提出逆概率(inverse probability)的数学结构. 根据 Stigler 的观点，1774 年 Laplace 论述了一般形式下的逆概率定理. 1939 年，Jeffrey 重新发现了 Laplace 的研究工作. 在 1774~1939 年的 165 年，在科学文献中似乎没有令人信服的东西. 因此,可以说此间的贝叶斯方法是悄无声息、无人问津的. 如果把 1900 年算作是近代数理统计开始的第一年，则到 Cramer 的专著《统计数学方法》(Mathematical Methods of Statistics)出版后 50 多年中，可以说基本上是经典统计学派一统天下.

20 世纪 50 年代后，随着统计理论及方法的应用范围扩大，贝叶斯理论也受到了欢迎，并得到了迅速发展，特别是决策问题在统计应用中占有越来越重要的地位. 而对决策问题而言，先验知识的使用是不可或缺的，而且与纯理论问题比较，在这类问题中主观概率的提法往往更为自然，同时反映了决策者掌握信息的程度，因此贝叶斯观点易于接受. 20 世纪初，数学的公理化倾向影响到统计界，概率的公理化最终由 Kolmogorov 完成，并得到了普遍认同，即概率是正则化的测度. 然而，对主观概率应该用什么样的公理来描述，经过很长时间的讨论，还

没有比较一致的看法. 意大利学者 De Finetti 认为 Kolmogorov 公理中互斥事件之和的概率等于各事件概率之和的结论，对于信念而言不成立，只能是有限可加. De Finetti 的关于可换随机变量的研究成果为先验分布的客观性提供了理论基础，Fisher 的似然原理促进了贝叶斯学派的发展. 似然函数是贝叶斯学派的基点和支柱，从最大似然估计到最大后验估计，从似然比到后验比等，使贝叶斯估计、推断理论和方法获得了系统论述. 英国统计学家 Jeffreys 在无信息先验分布上的重要突破，则形成了 Jeffreys 准则，其专著《概率论》的出版则标志着贝叶斯学派的形成. 后来 Good 对主观概率的研究，Savage 对效应函数、主观概率以及它们之间相互关系的研究，把贝叶斯学派推到一个新阶段. 理论研究上的进展，如早在 1950 年 Wald 就在其著作《统计决策函数》(Statistical Decision Function)中证明了在广泛的条件下，贝叶斯决策构成本质完全类；20 世纪 70 年代 Rubin 证明了在某些一般公认的前提下，任何一种最优准则下的解必然是某个先验分布下的贝叶斯解. 所有这些都从理论的高度阐明了贝叶斯理论的合理性. Lindley 也做了不少工作，将一些经典学派的成果给出了贝叶斯学派的推导和解释，明确而尖锐地捍卫了贝叶斯学派. 由于贝叶斯方法在实际工作中取得了明显的成功，又由于经典统计学派的一些工作从理论上也需要借助贝叶斯方法才能完成，如 Minmax 解与容许解问题，因此，贝叶斯学派的队伍日趋壮大起来，成为两大统计学派之一.

1.2 贝叶斯方法的特点

1.2.1 贝叶斯方法的本质

在统计推断的基本理论和方法两个方面，贝叶斯学派与频率学派之间存在着本质性的差异，这主要表现在以下几个方面：

第一，频率学派在进行统计推断时，依据两类信息：一是模型信息，即统计总体服从何种概率分布，这是制定统计方法的基础；另一个是样本信息，即观察或试验的结果. 贝叶斯学派则除了以上两类信息外，尚利用另外一类信息，即总体分布中未知参数的分布信息.

第二，频率学派坚持概率的频率解释，并在这个基础上去理解一切统计推断的结论，如在 Neyman 的区间估计理论中，"某区间估计$[\underline{\theta},\bar{\theta}]$的置信水平为$1-\alpha$"这一推断，此处$\theta$应理解为一无随机性的未知参数，当区间估计$[\underline{\theta},\bar{\theta}]$反复大量使用时，100 次中大约平均有$(1-\alpha)\times 100$次包含了参数$\theta$. 与此相反，贝叶斯学派赞成主观概率，概率是认识主体对事件出现可能性大小的相信程度，它并不依赖事件能否重复.

第三，两个学派的具体推断理念之间存在着根本差异. 统计学奠基人 Fisher 在其 1921 年发表的论文“理论统计的数学基础”中，把统计学的任务概括为三个问题：选定模型、确定统计量和决定统计量的分布. 根据 Fisher 的观点，信息包含在样本中，但样本为数众多，因此须用少数几个统计量把信息集中起来. 而抽样分布则决定了统计量的全部性质. 目前，频率统计学派基本上都是按照这种思路来处理统计推断问题的.

与此不同，贝叶斯推断的一般模式：先验信息⊕样本信息$\Rightarrow$后验信息(参见图 1.2.1)，或$\pi(\theta)\oplus p(x|\theta)\Rightarrow\pi(\theta|x)$，此处符号“⊕”表示为贝叶斯定理的作用.

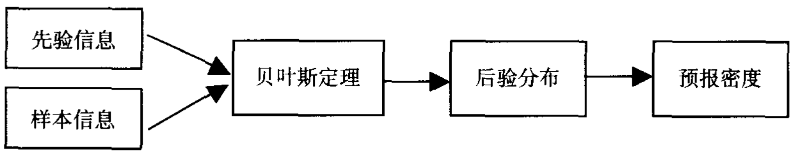

图 1.2.1　贝叶斯推断的基本模式

贝叶斯学派认为：先验分布反映了试验前对总体参数分布的认识，在获得样本信息后，人们对这个认识有了改变，其结果就反映在后验分布中，即后验分布综合了参数先验分布和样本信息. 由此可以看出，频率学派统计推断是“从无到有”的过程：在试验前，关于未知参数的情况是一无所知，而试验后则有些了解，但对了解多少并无普遍的表述方法，在实践中有赖于所使用的统计量的针对性. 贝叶斯推断则不然，它是一个“从有到有”的过程，且结果清楚自然，符合人们的思维习惯. 根据所获得的信息修正以前的看法，不一定从零开始. 从本质上说，贝叶斯推断方法概括了一般人的学习过程.

第四，贝叶斯方法只能基于参数的后验分布来分析问题. 也就是说，在获得后验分布后，如果把样本、原来的统计模型(包括总体分布和先验分布)都丢掉，一点也不会影响将来的统计推断问题，凡是符合这个准则的推断就是贝叶斯推断. 据此，频率学派中的矩估计、显著性统计检验和置信区间估计都不属于贝叶斯推断的范畴，但 MLE 估计则可视为均匀先验分布之下的贝叶斯估计. 因此，作为频率学派中一个很重要的极大似然估计，不过是在一种很特殊先验分布下的贝叶斯估计而已.

1.2.2　贝叶斯学派对经典学派的批评

我国统计学家成平教授指出，经典统计的最大缺陷在于其推断过程中过于着眼于当前数据，忽视历史的经验、人们已有的认识和知识，以及人们的主观能动性；统计推断的精度主要取决于样本大小，这对于小样本，往往是很困难甚至无

能为力的. 例如，在导弹等尖端武器的可靠性评定中，有相同条件下全弹试验难以超过 10 个，有时只有两三个试验，就必须作出决策.

张尧庭教授也认为，频率学派对一些统计问题的提法不妥，包括估计问题中的置信区间和假设检验，同时判断统计方法好坏的标准欠妥. 以方差已知的正态总体 $N(\mu,\sigma_0^2)$ 均值估计为例，假设 $X_i, i=1,2,\cdots,n$ 是来自该总体的一个样本，则可以求出均值 μ 的置信区间，其推理过程如下：

由于

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N\left(\mu,\frac{\sigma_0^2}{n}\right) \tag{1.2.1}$$

因此

$$\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1) \tag{1.2.2}$$

若 $Z_{\alpha/2}$ 为双侧 α 分位点的值，即

$$P(|\xi|>Z_{\alpha/2})=\alpha \tag{1.2.3}$$

据此

$$P\left(|\bar{X}-\mu|<Z_{\alpha/2}\sigma_0/\sqrt{n}\right)=1-\alpha \tag{1.2.4}$$

于是称 $\left[\bar{X}-Z_{\alpha/2}\sigma_0/\sqrt{n}, \bar{X}-Z_{\alpha/2}\sigma_0/\sqrt{n}\right]$ 是参数 μ 的置信水平为 $1-\alpha$ 的区间估计. 由于在经典统计的理论体系中参数 μ 是一个固定的常数，并不具有随机性，因而(1.2.4)式也就不能理解为

$$\mu\in\left[\bar{X}-Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X}-Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] \tag{1.2.5}$$

的概率等于 $1-\alpha$. 根据经典学派的基本观点，区间

$$\left[\bar{X}-Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X}-Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] \tag{1.2.6}$$

能盖住参数 μ 的概率是 $1-\alpha$，然而人们最关心的恰好是参数 μ 在该范围内的概率有多大，因此在经典统计理论中区间估计问题的提法及其解答并不令人满意.

然而，贝叶斯方法恰好不存在上述问题，因为在贝叶斯理论体系中，参数是随机变量，它本身就有统计分布. 事实上，可以从贝叶斯假设直接导出与(1.2.1)

式完全相同的等式，不过此时 μ 是随机变量，而 $\bar{X}$ 是常数. 因此，根据贝叶斯学派的观点，(1.2.4)式就是

$$\mu \in \left[\bar{X} - Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} - Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$$

这一事件发生的概率是$1-\alpha$.

1.2.3 贝叶斯方法的优点

贝叶斯理论的哲理有相当大的吸引力，而且方法简单，它在统计推断模式上与频率学派的不同之处在于：频率学派认为，似然函数概括了有关参数的全部信息，因此关于参数 θ 的统计推断只要利用似然函数就够了；而贝叶斯方法既利用了似然函数，又利用了参数先验信息. 如果先验信息很少或没有先验信息，这时贝叶斯推断方法所得到的结论与频率方法的基本相同，这一点可以从后面的讨论中看到.

与频率方法比较，贝叶斯方法具有如下几个方面的优点：①贝叶斯方法充分利用了样本信息和参数的先验信息，在进行参数估计时，通常贝叶斯估计量具有更小的方差或平方误差，能得到更精确的预测结果；②贝叶斯 HPD 置信区间比不考虑参数先验信息的频率置信区间短；③能对假设检验或估计问题所作出的判断结果进行量化评价，而不是频率统计理论中的接受、拒绝的简单判断.

1.3 贝叶斯方法的研究与应用

1.3.1 贝叶斯理论的研究

1. *先验分布理论的研究*

先验分布是贝叶斯推断理论的基础和出发点，也是贝叶斯学派研究的重点问题之一，它大体上可以分为扩散先验分布和共轭先验分布两大类，此处的扩散先验分布即一般文献中的无信息先验分布. 当然，扩散先验分布实际上包含许多信息. 根据 Bernardo 和 Ramon(1998)的观点，扩散先验分布应该满足以下几条性质：①不变性；②相合的边缘化；③相合的抽样性质；④普遍性；⑤容许性.

参数先验分布的选取方法之一是贝叶斯假设，即参数的先验分布 $\pi(\theta)$ 在 θ 的取值范围内是均匀分布的：若将 θ 的取值范围记为 Θ，并略去密度取值为 0 的部分，则参数 θ 的先验分布密度函数为

$$\pi(\theta) \propto 1, \qquad \theta \in \Theta \tag{1.3.1}$$

此处“$\propto$”表示该符号两边成正比，二者之间仅差一个常数项.

但是，贝叶斯假设中存在一个矛盾：均匀分布在参数变换时一般并不满足不变性要求，即变换后的分布不再是均匀分布. 如果参数θ选取均匀分布作为其先验分布，根据贝叶斯假设，θ的函数$h(\theta)$也应选用均匀分布作为其先验分布，然而由θ服从均匀分布这一前提，往往不能导出$h(\theta)$也服从均匀分布. 例如，标准差σ的均匀分布就不会变成σ^2的一个均匀分布.

为了克服这一矛盾，Jeffreys(1957)根据不变性的要求，提出了一种基于信息函数的扩散先验分布选择方法：若令$L(\theta)$为似然函数，Jeffreys 认为参数先验分布与$\sqrt{I(\theta)}$成比例，这里$I(\theta)$为 Fisher 信息阵. 也就是说 Jeffreys 先验分布为

$$\pi(\theta) \propto \sqrt{I(\theta)} = \sqrt{E\left[-\frac{\partial^2 L(\theta)}{\partial \theta^2}\right]} \tag{1.3.2}$$

不难证明 Jeffreys 先验分布在一对一的参数变换中保持不变. 如果$\pi(\theta)$为参数θ的先验分布，$\eta = h(\theta)$为一个一对一的参数变换，则η的 Jeffreys 先验分布为

$$\pi[h^{-1}(\eta)] \cdot \left|\frac{\mathrm{d}h^{-1}(\eta)}{\mathrm{d}\eta}\right| \tag{1.3.3}$$

Kass(1990)的研究表明：Jeffreys 先验分布能近似地保持参数后验分布的形状不变. Jeffreys 先验分布对于单参数情况很有用，但在多参数情况下就可能会出现一些问题. 在许多实际问题中，有时只需考虑部分参数的信息，此时其余参数就成为了冗余参数(nuisance parameter)，对于这类参数先验分布的构造，有两种基本方法，一是 Bernardo(1979，1994)、Sun 和 Berger(1998)提出的参照先验分布(reference prior)；另一种方法是 Stein(1985)、Tibshirani(1989)提出的概率匹配先验分布(probability matching prior). 概率匹配先验分布的基本思想：当样本容量趋于无穷大时，贝叶斯概率渐近地和相应的频率概率相匹配. 如果$X_1, X_2, \cdots, X_n$独立同分布，分布密度函数为$f(x|\theta,\omega)$，θ为所关心的参数，ω为冗余参数. 对于参数的先验分布密度$\pi(\theta,\omega)$，若有

$$P\{\theta > \theta_{1-\alpha}^{\pi}(x_1, x_2, \cdots, x_n) | \omega\} = \alpha + o(n^{-1/2}) \tag{1.3.4}$$

则称它满足一阶概率匹配准则，此处$\theta_{1-\alpha}^{\pi}(x_1, x_2, \cdots, x_n)$表示先验分布$\pi(\theta,\omega)$下参数后验分布的分位点. Peers(1965)证明了一阶概率匹配先验分布一般为微分方程组的解，Datta 和 Ghosh(1996)对其给出了更为严格和普遍的表述. 然而，由于可供选择的一阶概率匹配先验分布很多，很难决定选择哪一个更合适，因此，Mukerjee 和 Dey(1993)提出了二阶概率匹配先验分布，它和一阶概率匹配先验分

布的不同在于将(1.3.4)式中的$o(n^{-1/2})$换成$o(n^{-1})$，二阶概率匹配先验分布常常是唯一的.

最大熵(maximum entropy)原则也是确定参数先验分布的另一途径. 如果随机变量X的分布密度函数为$f(x)$，则称$-\int f(x)\log f(x)\mathrm{d}x$为随机变量$X$的熵，熵最大化作为推理的一般性原则最初由Jaynes(1957)提出的，根据Kullback的定理，先验分布应取参数θ的变化范围类熵最大的分布.

扩散先验分布构造的其他途径包括：相对似然函数法(the relative likelihood approach)、积累函数法(cumulative distribution function)、Monte-Carlo方法、Bootstrap法、随机加权法和Harr不变测度法等方法.

除了扩散先验分布外，还有一类应用广泛的先验分布，即共轭先验分布. 根据贝叶斯定理，参数后验分布与似然函数、先验分布的乘积成正比，因此似然函数的性质关系到和非正常先验分布一起是否能形成正常的后验分布，也关系到是否可以有效地应用Gibbs法进行统计推断. George(1993)研究了指数族的共轭似然先验分布，Arnold(1996)研究了指数分布族的共轭分布族，Chen(1985)研究了指数共轭分布的渐近正态性质.

目前，还没有统一的先验分布构造方法.

2. *后验分布推断技术的研究*

尽管贝叶斯推断模式简单，并且概率形式优美. 然而，在贝叶斯分析中，一般只知道后验分布密度函数的核，而难以获得具体的密度函数，也很难找到累积分布函数的数值分位点，计算边缘后验分布密度函数的困难是阻碍贝叶斯方法广泛应用的最大障碍. Dempster(1976)认为：统计推断技术的应用受概念和计算因素的阻碍，贝叶斯推断从概念上要比非贝叶斯推断直观得多，而非贝叶斯推断仅有一套复杂的特殊规则，贝叶斯方法更广泛应用的主要障碍是计算上的，这有赖于编制有效的计算程序.

在20世纪90年代以前，一些学者提出了数值和解析近似的方法来解决参数后验分布密度和后验分布各阶矩的计算问题，如Lindley数值逼近法、Naylor-Smith逼近法、Tierney-Kadane逼近法. 然而，这些方法的实现，需要依靠复杂数值和解析近似技术及相应的软件支撑.

近年来，Markov链蒙特卡罗方法(Markov Chain Monte-Carlo, MCMC)的研究对推广贝叶斯推断理论和方法的应用开辟了广阔的前景，使得贝叶斯推断理论和方法的研究得到了再度复兴，以往被认为不可能实施计算的一些统计方法变得比较容易. 目前，MCMC已经成为一种处理复杂统计问题的特别流行的工具，尤其在经常需要复杂的高维积分运算的贝叶斯分析领域更是如此. 在那里，高维积分

运算主要是用来求取普通方法无法得到的后验分布密度. 在这个背景下，一些其他方法(如线性近似、高斯积分和非递推蒙特卡罗方法等)，或者不可行，或者无法提供精确的结果；而如果合理定义和实施，MCMC 总能得到一条或几条收敛的马尔可夫链，该马尔可夫链的极限分布就是所需的后验分布，而且 MCMC 还具有一些其他重要特性. 例如，它能把一些复杂的高维问题转化为一系列简单的低维问题. 此外，它并不要求似然函数和参数先验分布具有共轭结构.

目前，在贝叶斯分析中应用最为广泛的 MCMC 方法主要有两种：Gibbs 抽样(sampler)方法和 Metroplis-Hastings 方法. Gibbs 抽样方法是由 Geman 于 1984 年提出来的，最初用于图像处理分析、人工智能和神经网络等大型复杂数据的分析，后经 Gelfand 和 Smith(1990)引入贝叶斯模型研究中，通过模拟进行积分运算，这给贝叶斯方法的实际应用产生了极其深刻的影响. Gibbs 抽样的成功在于它利用满条件分布(full conditional distribution)将多个相关参数的复杂问题降低为每次只需处理一个参数的较为简单问题.

一种比 Gibbs 抽样更一般的 MCMC 方法是 Metroplis-Hastings 方法. Metroplis (1953)提出了一种转移核的方法，Hastings(1970)随后对其加以推广，形成了 Metroplis-Hastings 方法. 有些文献研究了 Gibbs 抽样与 Metroplis 方法相结合的问题，如在 Gibbs 抽样中利用 Metroplis 方法抽取随机数，Gilks、Best 和 Tan(2000)提出了一种基于 Gibbs 抽样的调整筛选 Metroplis 抽样方法(adaptive rejection Metropolis sampling)，这种方法在贝叶斯分析中也很有应用价值.

尽管 MCMC 方法应用广泛，但很难判断何时马尔可夫链已经渐近收敛于平稳分布. 对 MCMC 方法收敛性的研究一直是个重要课题，主要集中在两个领域：第一种是理论上的，主要是分析马尔可夫转移核，预先确定抽样数目，使 Markov 链在一定容忍域内收敛于平稳分布. 例如，Polson 在 1994 年给出了离散跳跃 Metroplis-Hastings 算法的多项式时间收敛边界；而 Rosenthal(1995)则给出了谱系模型在有限样本空间下连续情形时的边界. 但这些方法都涉及复杂的数学工具，并需要进行大量的重复积分；另外，实践证明：这些边界往往过于宽松，重复次数太大，因而在实际中很难以应用. 因此，几乎所有的涉及 MCMC 的工作都转向第二种方法，即对算法产生的样本作收敛诊断. 这方面的早期工作是比较相邻递推样本的经验分布，当这两个分布差异不大时就认为近似收敛了；但这样做必须使用多条独立的平行链，以便得到简单的矩、分位数和密度估计等. 黎曼和收敛诊断(Riemann sum convergence diagnosis)方法能够较为准确地确定 MCMC 在何时收敛，得到一组收敛于平稳分布的样本，以较高精度解决一些复杂的高维积分问题.

MCMC 是目前贝叶斯推断理论体系中的重点研究问题之一，Besag(2001)、Brooks(1998)、Shao 和 Ibrhim(1989)等学者从不同的角度对贝叶斯分析中的 MCMC

方法及相关问题进行了相当广泛的研究.

1.3.2 贝叶斯方法的应用

随着贝叶斯理论和方法的不断发展和完善，以及相应的计算软件的研制，贝叶斯方法在实践中获得了日趋广泛的应用；在生物医学研究中的应用；在知识发现与数据挖掘技术中的应用.

1. *在经济研究中的应用*

贝叶斯方法在经济中的应用主要集中在三个方面：经济计量方法、商业经济和宏观经济的预测方法、经济博弈论. 在经济分析中，通常采用结构模型对经济变量之间的数量关系进行定量研究，这类模型的特点之一是规模大，它们一般由数十个，甚至一两百个方程组成，模型参数多，样本容量相对不足，因此参数估计的有效性不强，贝叶斯方法是解决这类问题的一种有效方法. 芝加哥大学的 Zellner 研究了经济计量学中的贝叶斯理论，包括回归模型、完全递归模型和分布滞后模型的贝叶斯方法研究；Banergee(1993)、Bauwens(1994，1998，1999)和 Wes(1997)研究了动态经济计量模型的贝叶斯理论；Dreze(1976)研究了联立方程模型的有限信息和完全信息的贝叶斯分析；Van Dijk(1991)、Bierens(1993)、Berger(1994)、Hoek(1995)和 Franses(1997)等人研究了单位根问题的贝叶斯理论.

同时，贝叶斯方法在经济预测中应用广泛，这主要得益于 Litterman(1986)在 Minnesota 储备银行所做的开创性的研究工作. 他利用贝叶斯多元自回归模型，对国民生产总值等七个指标进行预测，取得了很好的效果. 此后，贝叶斯方法在商业经济预测和政府宏观经济预测中获得了广泛应用，相关的研究成果逐年增多，如 Kenny、Meyler 和 Quinn(1998)利用贝叶斯方法研究了爱尔兰通货膨胀问题，与爱尔兰中央银行的预测结果比较，结果表明：前者的预测效果更好；Kasuya 和 Tanemura(2000)利用后验信息准则和 Monte Carlo 方法研制了一个小型的贝叶斯日本经济预测模型，该模型包括居民消费价格指数等 8 个经济指标；Bewley 和 Griffiths(2002)研究了对数扩散模型的贝叶斯预测方法；类似的文献还有 Spencer (1993)、McCulloch(1993)和 Amisano(1999)的研究成果.

此外，贝叶斯方法在经济博弈理论中也获得了应用，这一经济学问题可抽象成贝叶斯博弈结构. 一般博弈结构由三个部分组成，即局中人、策略空间和支付；但贝叶斯博弈结构还需要增加两个部分：特征和概率. 贝叶斯博弈论不但具有深刻的理论意义，而且有广泛的实用价值. 在市场分析研究中，一般是把研究问题转化为一个贝叶斯博弈模型，然后进行模拟，求出最优策略；在工业组织中贝叶斯博弈被大量运用，在经济发展战略决策、保险业务、金融投资方面也有重要应