



高等院校统计专业规划教材

贝叶斯统计

Bayesian Statistics

茆诗松 编著

北京)

中国统计出版社

高等院校统计学专业规划教材

贝 叶 斯 统 计

茆 诗 松 编 著

中国统计出版社

(京)新登字 041 号

图书在版编目(CIP)数据

贝叶斯统计/茆诗松编著.

—北京:中国统计出版社,1999.9

高等院校统计学规划教材

ISBN 7-5037-2930-9

I. 贝...

II. 茆...

III. 贝叶斯统计-高等学校-教材

IV. 0212

中国版本图书馆 CIP 数据核字(1999)第 10216 号

作 者/茆诗松

责任编辑/吕 军

责任校对/刘开颜

封面设计/张建民

出版发行/中国统计出版社

通信地址/北京市三里河月坛南街 75 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

电 话/(010)63459084,63266600—22500(发行部)

印 刷/科伦克三莱印务(北京)有限公司

经 销/新华书店

开 本/850×1168mm 1/32

字 数/600 千字

印 张/8.625

印 数/1—5000 册

版 别/1999 年 10 月第 1 版

版 次/1999 年 10 月第 1 次印刷

书 号/ISBN 7-5037-2930-9/O. 34

定 价/15.00 元

中国统计版图书,版权所有,侵权必究。

中国统计版图书,如有印装错误,本社发行部负责调换。

出 版 说 明

“九五”期间是我国社会主义市场经济体制逐步完善和发展的重要时期,一方面,随着高等教育体制改革和统计改革的深入发展,对统计教育模式和统计人才培养目标都提出新的要求,另一方面,科学技术的飞速发展也促使统计技术发生了重大变革,新理论、新方法和新技术不断涌现并被应用于统计实践。为了适应这种新形势的需要,全国统计教材编审委员会制定了《1996—2000年全国统计教材建设规划》,根据《规划》的要求,编委会采取招标的方式组织全国有关院校的专家、学者编写了这批统计学专业“规划教材”。

这批教材力求以邓小平理论为指导,在总结“八五”期间规划统计教材建设经验的基础上,认真贯彻以下原则:①理论紧密联系实际的原则;②解放思想、转变观念、大胆探索、努力创新的原则;③正确处理继承与发展关系的原则。通过不懈努力,把这批教材建设成为质量高、适应性强、面向 21 世纪的新教材。

相信通过这批教材的出版、发行,对推动我国统计教育改革和加快更新、改造我国统计教材体系、教材内容的步伐将起到积极的促进作用,同时对我国统计教材建设也将起到较好的示范、导向作用。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足之处,诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

1999 年 3 月

前 言

本书是按照全国统计教材编审委员会指定的《贝叶斯统计》编写大纲编写的,是供全国高等学校统计专业大学生和研究生学习用的教科书。贝叶斯统计在近50年中发展很快,内容愈来愈丰富。这里只选用其中最基本部分构成本书,相当一学期的内容,本书力图向学习过传统的概率统计(频率学派)课程的学生展示贝叶斯统计的基本面貌。也使他们能了解贝叶斯统计的基本思想,掌握贝叶斯统计的基本方法,为在实际中使用和研究贝叶斯统计打下了良好的基础。

本书共六章,可分二部分。前三章围绕先验分布介绍贝叶斯推断方法。后三章围绕损失函数介绍贝叶斯决策方法。阅读这些内容仅需要概率统计基本知识就够了。本书力图用生动有趣的例子来说明贝叶斯统计的基本思想和基本方法,尽量使读者对贝叶斯统计产生兴趣,引发读者使用贝叶斯方法去认识和解决实际问题的愿望。进而去丰富和发展贝叶斯统计。假如学生的兴趣被钓出来,愿望被引出来,那么讲授这一门课的目的也基本达到了。

贝叶斯统计是在与经典统计的争论中逐渐发展起来的。争论的问题有:未知参数是否可以看作随机变量?事件的概率是否一定要有频率解释?概率是否可用经验来确定?在这些问题的争论中,贝叶斯学派建立起自己的理论与方法。另一方面,在全球传播已有百年历史的经典统计对统计学的发展和应用起了巨大作用,但同时也暴露了一些问题。在小样本问题研究上、在区间估计的解释上、在似然原理的认识上等问题经典统计也受到贝叶斯学派的批评,在这些批评中贝叶斯学派也在不断完善贝叶斯统计。

J. O. Berger《统计决策论及贝叶斯分析》一书在1980年和1985年相继二版问世把贝叶斯统计作了较完整的叙述。在近20年中贝叶斯统计在实际中又获得广泛的应用,1991年和1995年在美国连续出版了二本《Case Studies in Bayesian Statistics》。使贝叶斯统计在理论上和实际上以及它们的结合上都得到了长足的发展。惧怕使用贝叶斯统计思想得到克服。如今贝叶斯统计也走进教室,打破经典统计独占教室的一统天下的局面,这不能不说是贝叶斯统计发展中的一些重要标志。贝叶斯统计已成为统计学中一个不可缺少的部分。相比之下,贝叶斯统计在我国的应用与发展尚属起步阶段,但我国有很好的发展贝叶斯统计的氛围。只要大家努力,贝叶斯统计在我国一定能迅速发展,跟上世界主流。

本书编写自始至终得到国家统计局教育中心的关心和帮助,没有他们的督促,本书还会延期出版。上海财经大学张尧庭教授和中国人民大学的吴喜之教授耐心细致地审阅了全书,提出许多宝贵意见,笔者都认真考虑,并作修改,这使全书增色不少。另外,何基报、顾娟、孙汉杰等阅读书稿,提出宝贵意见,还帮助打印全书,在此一并表示感谢。

由于编者水平有限,准确表达贝叶斯学派的各种观点并非易事,错谬之处在所难免,恳请国内同行和广大读者批评指正。

茆诗松

1999年1月30日

目 录

第一章 先验分布与后验分布	(1)
§ 1.1 三种信息	(1)
一、总体信息	(1)
二、样本信息	(2)
三、先验信息	(2)
§ 1.2 贝叶斯公式	(6)
一、贝叶斯公式的密度函数形式	(6)
二、后验分布是三种信息的综合	(8)
§ 1.3 共轭先验分布	(13)
一、共轭先验分布	(13)
二、后验分布的计算	(15)
三、共轭先验分布的优缺点	(16)
四、常用的共轭先验分布	(19)
§ 1.4 超参数及其确定	(20)
一、利用先验矩	(20)
二、利用先验分位数	(21)
三、利用先验矩和先验分位数	(23)
四、其它方法	(23)
§ 1.5 多参数模型	(24)
§ 1.6 充分统计量	(29)
习 题	(31)
第二章 贝叶斯推断	(35)
§ 2.1 条件方法	(35)
§ 2.2 估计	(36)
一、贝叶斯估计	(36)

二、贝叶斯估计的误差	(39)
§ 2.3 区间估计	(43)
一、可信区间	(43)
二、最大后验密度 (HPD) 可信区间	(47)
§ 2.4 假设检验	(51)
一、假设检验	(51)
二、贝叶斯因子	(53)
三、简单假设 $\Theta_0 = \{\theta_0\}$ 对简单假设 $\Theta_1 = \{\theta_1\}$	(53)
四、复杂假设 Θ_0 对复杂假设 Θ_1	(54)
五、简单原假设对复杂的备择假设	(58)
§ 2.5 预测	(63)
§ 2.6 似然原理	(67)
习 题	(72)

第三章 先验分布的确定

§ 3.1 主观概率	(75)
一、主观概率	(75)
二、确定主观概率的方法	(77)
§ 3.2 利用先验信息确定先验分布	(80)
一、直方图法	(80)
二、选定先验密度函数形式再估计其超参数	(81)
三、定分度法与变分度法	(84)
§ 3.3 利用边缘分布 $m(x)$ 确定先验密度	(88)
一、边缘分布 $m(x)$	(88)
二、混合分布	(89)
三、先验选择的 $ML-II$ 方法	(90)
四、先验选择的矩方法	(92)
§ 3.4 无信息先验分布	(95)
一、贝叶斯假设	(96)
二、位置参数的无信息先验	(98)
三、尺度参数的无信息先验	(101)
四、用 Fisher 信息阵确定无信息先验	(102)
§ 3.5 多层先验	(105)

一、多层先验	(105)
二、多层模型	(110)
习 题	(118)
第四章 决策中的收益、损失与效用	(120)
§ 4.1 决策问题的三要素	(120)
一、决策问题	(120)
二、决策问题的三要素	(122)
§ 4.2 决策准则	(125)
一、行动的容许性	(125)
二、决策准则	(127)
§ 4.3 先验期望准则	(132)
一、先验期望准则	(132)
二、两个性质	(137)
§ 4.4 损失函数	(139)
一、从收益到损失	(139)
二、损失函数	(140)
三、损失函数下的悲观准则	(142)
四、损失函数下的先验期望准则	(144)
§ 4.5 常用损失函数	(146)
§ 4.6 效用函数	(150)
一、效用和效用函数	(150)
二、效用的测定	(154)
三、效用尺度	(157)
四、常见的效用曲线	(158)
五、用效用函数作决策的例子	(161)
六、从效用到损失	(164)
习 题	(165)
第五章 贝叶斯决策	(170)
§ 5.1 贝叶斯决策问题	(170)
§ 5.2 后验风险准则	(174)
一、后验风险	(174)

二、决策函数	(177)
三、后验风险准则	(179)
§ 5.3 常用损失函数下的贝叶斯估计	(184)
一、平方损失函数下的贝叶斯估计	(184)
二、线性损失函数下的贝叶斯估计	(190)
三、有限个行动问题的假设检验	(192)
§ 5.4 抽样信息期望值	(194)
一、完全信息期望值	(194)
二、抽样信息期望值	(198)
§ 5.5 最佳样本量的确定	(202)
一、抽样净益	(202)
二、最佳样本量及其上界	(203)
三、最佳样本量的求法	(205)
§ 5.6 二行动线性决策问题的 $EVPI$	(211)
一、正态分布下二行动线性决策问题的先验 $EVPI$	(212)
二、贝塔分布下二行动线性决策问题的先验 $EVPI$	(218)
三、伽玛分布下二行动线性决策问题的先验 $EVPI$	(221)
习 题	(223)
第六章 统计决策理论	(228)
§ 6.1 风险函数	(228)
一、风险函数	(228)
二、决策函数的最优性	(230)
三、统计决策中的点估计问题	(230)
四、统计决策中的区间估计问题	(232)
五、统计决策中的假设检验问题	(233)
§ 6.2 容许性	(235)
一、决策函数的容许性	(235)
二、stein 效应	(239)
§ 6.3 最小最大准则	(242)

一、最小最大准则	(242)
二、最小最大估计的容许性	(245)
§ 6.4 贝叶斯风险	(246)
一、贝叶斯风险	(246)
二、贝叶斯风险准则与后验风险准则的 等价性	(248)
§ 6.5 贝叶斯估计的性质	(250)
习 题	(255)
附录 1 常用概率分布表	(258)
附录 2 标准正态分布函数 $\Phi(Z)$ 表	(262)
参考文献	(263)

第一章 先验分布与后验分布

§ 1.1 三种信息

统计学中有二个主要学派:频率学派与贝叶斯学派,他们之间有共同点,又有不同点,为了说清楚他们之间的异同点,我们从统计推断所使用的三种信息说起。

一、**总体信息**,即总体分布或总体所属分布族给我们的信息,譬如,“总体是正态分布”这一句话就给我们带来很多信息:它的密度函数是一条钟形曲线;它的一切阶矩都存在;有关正态变量(服从正态分布的随机变量)的一些事件的概率可以计算;有正态分布可以导出 χ^2 分布、 t 分布和 F 分布等重要分布;还有许多成熟的点估计、区间估计和假设检验方法可供我们选用。总体信息是很重要的信息,为了获取此种信息往往耗资巨大。美国军界为了获得某种新的电子元器件的寿命分布,常常购买成千上万个此种元器件,做大量寿命实验、获得大量数据后才能确认其寿命分布是什么。我国为确认国产轴承寿命分布服从两参数威布尔分布前后也花了五年时间,处理几千个数据后才定下的。又如保险费的确定与人的寿命分布密切相关,在保险业中,人的寿命分布被称为寿命表,中国人的寿命表不同于外国人的寿命表,男人的寿命表不同于女人的寿命表,北方人的寿命表不同于南方人的寿命表,当代人的寿命表与50年前人的寿命表也是不同的,而要确定这些寿命表是一项耗资费时的的工作,至今我国还缺乏此类寿命表。确定我国各类人群的寿命表是我国统计工作者的重要任务。

二、样本信息,即从总体抽取的样本给我们提供的信息。这是最“新鲜”的信息,并且愈多愈好。人们希望通过对样本的加工和处理对总体的某些特征作出较为精确的统计推断。没有样本就没有统计学可言。这是大家都理解的事实。

基于上述两种信息进行的统计推断被称为经典统计学,它的基本观点是把数据(样本)看成是来自具有一定概率分布的总体,所研究的对象是这个总体而不局限于数据本身。据现有资料看,这方面最早的工作是高斯(Gauss, C. F. 1777~1855)和勒让德(Legendre, A. M. 1752~1833)的误差分析、正态分布和最小二乘法。从十九世纪末期到二十世纪上半叶,经皮尔逊(Pearson, K. 1857~1936)、费歇(Fisher, R. A. 1890~1962)奈曼(Neyman, J.)的等人杰出的工作创立了经典统计学。如今统计学教材几乎全是叙述经典统计学的理论与方法。二十世纪下半叶,经典统计学在工业、农业、医学、经济、管理、军事等领域里获得广泛的应用。这些领域中又不断提出新的统计问题,这又促进了经典统计学的发展,随着经典统计学的持续发展与广泛应用,它本身的缺陷也逐渐暴露出来了。这些将逐步展开讨论。

现在回到我们讨论的问题上来,除上述两种信息外,在我们周围还存在第三种信息—先验信息,它也可用于统计推断。

三、先验信息,即在抽样之前有关统计问题的一些信息,一般说来,先验信息主要来源于经验和历史资料。先验信息在日常生活和工作中也经常可见,不少人在自觉地或不自觉地使用它。看下面二个例子。

例 1.1 英国统计学家 Savage(1961)曾考察如下二个统计实验:

A. 一位常饮牛奶加茶的妇女声称,她能辨别先倒进杯子里的是茶还是牛奶。对此做了十次试验,她都正确地说出了。

B. 一位音乐家声称,他能从一页乐谱辨别出是海顿(Haydn)还是莫扎特(Mozart)的作品。在十次这样的试验中,他都能正确

辨别。

在这两个统计试验中,假如认为被实验者是在猜测,每次成功概率为 0.5,那么十次都猜中的概率为 $2^{-10} = 0.0009766$,这是一个很小的概率,是几乎不可能发生的,所以“每次成功概率为 0.5”的假设应被拒绝。被实验者每次成功概率要比 0.5 大得多。这就不是猜测,而是他们的经验在帮了他们的忙。可见经验(先验信息的一种)在推断中不可忽视,应加以利用。

例 1.2 “免检产品”是怎样决定的? 某厂的产品每天都要抽检几件,获得不合格品率 θ 的估计。经过一段时间后就积累大量的资料,根据这些历史资料(先验信息的一种)对过去产品的不合格率可构造一个分布:

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i=0,1,\dots,n$$

这个对先验信息进行加工获得的分布今后称为先验分布。这个先验分布是综合了该厂过去产品的质量情况。如果这个分布的概率绝大部分集中在 $\theta=0$ 附近,那该产品可认为是“信得过产品”。假如以后的多次抽检结果与历史资料提供的先验分布是一致的。使用单位就可以对它作出“免检产品”的决定,或者每月抽检一、二次就足够了,这就省去了大量的人力与物力。可见历史资料在统计推断中应加以利用。

基于上述三种信息(总体信息、样本信息和先验信息)进行的统计推断被称为贝叶斯统计学。它与经典统计学的主要差别在于是否利用先验信息。在使用样本信息上也是有差异的。贝叶斯学派重视已出现的样本观察值,而对尚未发生的样本观察值不予考虑,贝叶斯学派很重视先验信息的收集、挖掘和加工,使它数量化,形成先验分布,参加到统计推断中来,以提高统计推断的质量。忽视先验信息的利用,有时是一种浪费,有时还会导致不合理的结论。

贝叶斯统计起源于英国学者贝叶斯(Bayes, T. R. 1702(?))~

1761)死后发表的一篇论文“论有关机遇问题的求解”。在此论文中他提出著名的贝叶斯公式和一种归纳推理方法,随后拉普拉斯(Laplace, P. C. 1749~1827)等人用贝叶斯提出的方法导出一些有意义的结果。之后虽有一些研究和应用,但由于其理论尚不完整,应用中又出现一些问题,致使贝叶斯方法长期未被普遍接受。直到二次大战后,瓦尔德(Wald, A. 1902~1950)提出统计决策函数论后又引起很多人对贝叶斯方法研究的兴趣。因为在这个理论中贝叶斯解被认为是一种最优决策函数。在 Savage, L. J. (1954)、Jeffreys, H. (1961)、Good, I. J. (1950)、Lindley, D. V. (1961)、Box, G. E. P. & Tiao, G. C. (1973)、Berger, J. O. (1985)等贝叶斯学者的努力下,对贝叶斯方法在观点、方法和理论上不断的完善。另外在这段时期贝叶斯统计在工业、经济、管理等领域内获得一批无可非议的成功应用(见 Singpurwalla 主编课题论文集, 1993~1995)。贝叶斯统计的研究论文与著作愈来愈多,贝叶斯统计的国际会议经常举行。如今贝叶斯统计已趋成熟,贝叶斯学派已发展成为一个有影响的统计学派,开始打破了经典统计学一统天下的局面。

贝叶斯学派的最基本的观点是:任一个未知量 θ 都可看作一个随机变量,应用一个概率分布去描述对 θ 的未知状况。这个概率分布是在抽样前就有的关于 θ 的先验信息的概率陈述。这个概率分布被称为先验分布。有时还简称为先验(Prior)。因为任一未知量都有不确定性,而在表述不确定性程度时,概率与概率分布是最好的语言。例 1.2 中产品的不合格品率 θ 是未知量,但每天都有一些变化,把它看作一个随机变量是合适的,用一个概率分布去描述它也是很恰当的。即使是一个几乎不变的未知量,用一个概率分布去描述它的不确定性也是十分合理的。

例 1.3 学生估计一新教师的年龄。依据学生们的生活经历,在看了新教师的照片后立即会有反应:“新教师的年龄在 30 岁到 50 岁之间,极有可能在 40 岁左右。”一位统计学家与学生们交谈,

明确这句话中“左右”可理解为 ± 3 岁,“极有可能”可理解为90%的把握。于是学生们对新教师年龄(未知量)的认识(先验信息)可综合为图1.1所示的概率分布,这也是学生们对未知量(新教师年龄)的概率表述。

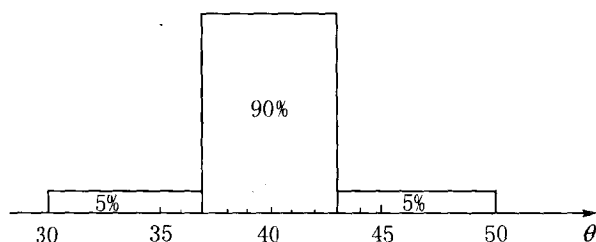


图 1.1 新教师年龄的先验分布

这里有二个问题需要进一步讨论。第一,按图1.1所示的概率分布我们可谈论未知量 θ 位于某个区间的概率。譬如, θ 位于37到43岁间的概率为0.90,即

$$P(37 \leq \theta \leq 43) = 0.90$$

可这个概率陈述在经典统计中是不允许的,因为经典统计认为 θ 是常量,它要么在37岁到43岁之间(概率为1),要么在这个区间之外(上述事件概率为零),不应有0.9的概率。可在实际中类似的说法经常可以听到。譬如:“某逃犯的年龄大约35岁左右”、“明日降水概率为0.85”、“某学生能考上大学的概率为0.95”、“这场足球赛甲队能胜的概率只有0.6左右”。这样的概率陈述能为大多数人理解、接受和采用。这种合理陈述的基础就是把未知量看作随机变量。

第二,图1.1中的概率0.90不是在大量重复试验中获得的,而是学生们根据自己的生活经验的积累对该事件发生可能性所给出的信念,这样给出的概率在贝叶斯统计中是允许的,并称为主观概率。它与古典概率和用频率确定的概率有相同的含义,只要它符合概率的三条公理即可。这一点频率学派是难以接受的,他们认为

经典统计学是用大量重复试验的频率来确定概率、是“客观的”，因此符合科学的要求，而认为贝叶斯统计是“主观的”，因而(至多)只对个人作决策有用。这是当前对贝叶斯统计的主要批评。贝叶斯学派认为引入主观概率及由此确定的先验分布至少把概率与统计的研究与应用范围扩大到不能大量重复的随机现象中来。其次，主观概率的确定不是随意的，而是要求当事人对所考察的事件有较透彻的了解和丰富的经验，甚至是这一行的专家，在这个基础上确定的主观概率就能符合实际。把这样一些有用的先验信息引入统计推断中来只会有好处，当然误用主观概率与先验分布的可能性是存在的，在这方面 Berger(1985)建议：“防止误用的最好方法是给人们在先验信息方面以适当的教育，另外在贝叶斯分析的最后报告中，应将先验分开来写，以便使其他人对主观输入的合理性作评价。”最后，贝叶斯学派也经常揭露频率学派的“客观性”，总体分布的选择对答案所产生的影响远比先验分布选择所产生的影响重大的多。而前者恰好也经常是主观的，另外评价一个统计方法好坏的标准上的选择，主观性也是很大的，都朝着对自己有利的方向选择。Good(1973)说得更直截了当：“主观主义者直述他的判断，而客观主义者以假设来掩盖其判断，并以此享受科学客观性的荣耀。”

上述的叙述对初学者可能还不能理解两学派在一些问题上的争论是多么深刻和多么激烈，但读完本章或全书后再品味这些叙述可能就有进一步的理解。

§ 1.2 贝叶斯公式

一、贝叶斯公式的密度函数形式

贝叶斯公式的事件形式在初等概率中都有叙述，这里用随机变量的密度函数再一次叙述贝叶斯公式，从中介绍贝叶斯学派的一些具体想法。