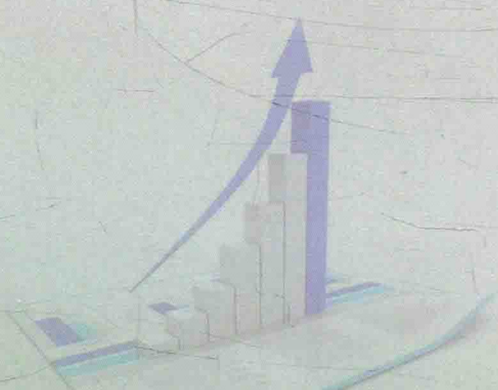
黄金超 著

贝叶斯统计分析

BEIYESI TONGJI FENXI



安徽师范大学出版社

2015年度安徽省高校自然科学研究重点项目（非参数贝叶斯和经验贝叶斯统计分析应用研究，项目编号：KJ2015A345）
2016年度滁州职业技术学院校级重点研究项目（Lomax分布和连续型单参数指数族参数的经验Bayes分析及应用，项目编号：YJZ-2016-01）
2017年度安徽省高校优秀拔尖人才培育资助项目（高校优秀青年骨干人才国内访学研修项目，项目编号：gxfx2017225）

贝叶斯统计分析

BEIYESI TONGJI FENXI

黄金超　著

安徽师范大学出版社

· 芜湖 ·

责任编辑:李　玲
装帧设计:张培树

图书在版编目(CIP)数据

贝叶斯统计分析 / 黄金超著. — 芜湖：安徽师范大学出版社, 2017.4
ISBN 978-7-5676-2690-4

Ⅰ.①贝… Ⅱ.①黄… Ⅲ.①贝叶斯统计量 Ⅳ.①O212.8

中国版本图书馆CIP数据核字(2016)第302983号

贝叶斯统计分析

黄金超　著

出版发行:安徽师范大学出版社
芜湖市九华南路189号安徽师范大学花津校区　邮政编码:241002
网　　址:http://www.ahnupress.com/
发 行 部:0553-3883578 5910327 5910310(传真)　E-mail:asdcbsfxb@126.com
印　　刷:虎彩印艺股份有限公司
版　　次:2017年4月第1版
印　　次:2017年4月第1次印刷
规　　格:710×1000　1/16
印　　张:14.25
字　　数:275千
书　　号: ISBN 978-7-5676-2690-4
定　　价:38.00元

前　言

贝叶斯统计推断理论的主要特点是使用先验分布，由样本与先验分布提供的信息得到后验分布，后验分布综合了样本与先验信息，组成较完整的后验信息，这一后验分布是贝叶斯统计推断的基础．统计学家Lindly预言：统计学的未来——一个贝叶斯统计的二十一世纪．美国统计学会的ASA和英国皇家学会的JRSS等著名杂志上，几乎每期都有贝叶斯统计方面的文章，可以说，贝叶斯统计是当今国际统计科学研究的热点．

本书是作者根据多年来对贝叶斯领域的研究成果及所积累的资料撰写而成，较为系统地介绍了贝叶斯统计学的理论基础及经验贝叶斯统计推断理论，其中相当一部分内容是作者最新的成果．第一章是贝叶斯统计分析基础，简单介绍了贝叶斯统计相关基础理论知识；第二章介绍了经验贝叶斯统计方法和密度函数核估计及部分研究成果；第三至七章是作者的研究成果，详细介绍了Weibull分布、Lomax分布、Cox分布、连续型单参数指数族分布、非指数分布、双指数分布等分布族相应参数的经验贝叶斯估计与检验问题，推广并改进了现有文献的相应结果．对贝叶斯统计分析研究工作感兴趣的读者认真研读本书后，可以较快地进入经验贝叶斯估计相关领域的研究工作．

本书的出版得到了2015年度安徽省高校自然科学研究重点项目(非参数贝叶斯和经验贝叶斯统计分析及应用研究，项目编号为KJ2015A345)、2017年度安徽省高校优秀拔尖人才培育资助项目(高校优秀青年骨干人才国内访学研修项目，项目编号为gxfx2017225)和2016年度滁州职业技术学院校级重点研究项目(Lomax分布和连续型单参数指数族参数的经验Bayes分析及应用，项目编号

为 YJZ-2016-01)的资助,同时得到了合肥工业大学数学学院凌能祥教授在科研上的悉心指导和无私奉献,在此一并表示感谢.

由于作者水平有限,书中难免有不足和谬误之处,恳请同行及广大读者批评指正.

黄金超

2016年11月

目　录

第一章

贝叶斯分析基础

§1.1 贝叶斯统计基本概念

近几十年来,统计学中的贝叶斯学派有了重大发展,如今已成为与经典学派并驾齐驱的两大统计学派之一.贝叶斯统计起源于贝叶斯(Bayes,1763)的一篇论文——机遇理论中一个问题的解,在此论文中,他提出了著名的贝叶斯公式和一种归纳推理的方法.

经典学派统计的出发点是根据样本在一定的统计模型下作出统计推断.本章假设统计模型为参数统计模型.在取得样本观测值 x 前,往往对参数 θ 统计模型中的参数有某些先验知识.在数学上,关于 θ 的先验知识的数学描述就是先验分布.贝叶斯统计的主要特点是使用先验分布,而在得到样本观测值 $\left(x_1,x_2,\cdots,x_n\right)^T$ 后,由 x 与先验分布提供信息,这是贝叶斯统计推断的基础.经典学派统计对样本量较大的样本有较好的统计推断效果.贝叶斯统计推断由于利用了先验知识,因而对小样本一般也有较好的统计推断效果.

1.1.1 三种信息

我们知道,数理统计学的任务是要通过样本推断总体.样本有两重性,当把样本视为随机变量时,它有概率分布,称为**总体分布**.如果我们已经知道总体的分布形式,这就给了我们一种信息,称为**总体信息**,即总体分布或总体所属分布族给我们的信息.

另外一种信息是**样本信息**,即从总体抽取的样本给我们提供的信息.

基于上述两种信息进行的统计推断被称为经典统计，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.

最后一种信息称为**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，不少人在自觉不自觉地使用它.

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.贝叶斯统计学与经典统计学的主要差别在于是否利用先验信息，它们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，并使它数量化，形成先验分布，参与到统计推断中来，以提高统计推断的质量.忽视先验信息的利用是一种浪费，有时还会导致不合理的结论.

贝叶斯学派的最基本观点是：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 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.

贝叶斯方法的一个主要问题是如何确定先验分布.先验分布的确定有很大的主观性、随机性.当先验分布完全未知时，如果人为地给出的先验分布与实际情形偏离较大，贝叶斯解的性质就较差.针对这一问题，Robbins(1956)首先提出经验贝叶斯(Empirical Bayes，简称EB)方法.它的实质是利用历史样本对先验分布或先验分布的某些数字特征作出直接或间接的估计，因此，EB方法是对贝叶斯方法的改进和推广，是介于经典统计学和贝叶斯统计学之间的一种统计推断方法.

1.1.2 先验分布与后验分布

定义1.1.1 参数 θ 的参数空间 Θ 上的任一概率分布，都称作**先验分布**.

本书用 $\pi(\theta)$ 表示 θ 的先验分布.这里 $\pi(\theta)$ 是随机变量 θ 的概率密度，即当 θ 为离散型随机变量时，$\pi(\theta_i)\left(i=1,2,\cdots\right)$ 表示事件 $\{\theta_i=\theta\}$ 的概率分布，即概率 $P\{\theta_i=\theta\}$；当 θ 为连续型随机变量时，$\pi(\theta)$ 表示 θ 的密度函数.

定义1.1.2 在获得样本 x 后，θ 的**后验分布**就是在给定 $X=x$ 条件下 θ 的条件分布，记为 $\pi\left(\theta|x\right)$.对于有密度的情形，它的密度函数为

$$\pi(\theta|x)=\frac{f(x|\theta)\pi(\theta)}{\int_{\Theta}f(x|\theta)\pi(\theta)\mathrm{d}\theta}. \tag{1.1.1}$$

其中 $h(x,\theta)=\pi(\theta)f(x|\theta)$ 为 X 和 θ 的联合分布，而

$$m(x)=\int_{\Theta}f(x|\theta)\pi(\theta)\mathrm{d}\theta.$$

因此

$$\pi(\theta|x)=\frac{h(x,\theta)}{m(x)}=\frac{f(x|\theta)\pi(\theta)}{m(x)}. \tag{1.1.2}$$

(1.1.1)式就是贝叶斯公式的密度函数形式，它集中了总体、样本和先验三种信息中有关 θ 的一切信息，而又排除了一切与 θ 无关的信息后所得到的结果. 从贝叶斯学派的观点来看，获取后验分布 $\pi(\theta|x)$ 后，一切统计推断都必须从 $\pi(\theta|x)$ 出发.

当 θ 是离散型随机变量时，先验分布可用先验分布列 $\{\pi(\theta_i), i=1,2,\cdots\}$ 表示，这时后验分布具有如下离散形式

$$\pi(\theta_i|x)=\frac{\pi(\theta_i)f(x|\theta_i)}{\sum_i \pi(\theta_i)f(x|\theta_i)}(i=1,2,\cdots). \tag{1.1.3}$$

假如样本来自的总体 X 也是离散的，只要把(1.1.1)或(1.1.3)式中的密度函数 $f(\theta_i|x)$ 看作事件 $\{X=x|\theta_i\}$ 的概率 $P(X=x|\theta_i)$ 即可. 此时就是**贝叶斯公式**.

贝叶斯公式把人们对 θ 的认识 $\pi(\theta)$ 调整到 $\pi(\theta|x)$. 也就是说，后验信息把先验信息和样本信息充分地结合在一起. 可以形象地表示为：样本信息加上先验信息得到后验信息，即可表示为

$$f(x|\theta)\oplus\pi(\theta)\Rightarrow\pi(\theta|x).$$

其中符号“ $\oplus$ ”应理解为贝叶斯公式的作用.

把(1.1.2)式改写为

$$\pi(\theta|x)=c(x)f(x|\theta)\pi(\theta). \tag{1.1.4}$$

其中 $c(x)=[m(x)]^{-1}$. $c(x)$ 可以看成待定常数，由下列关系式确定

$$\int_{\Theta}c(x)f(x|\theta)\pi(\theta)\mathrm{d}\theta=1.$$

在许多问题中，(1.1.4)式也常常可以略去常数而表示为

$$\pi(\theta|x)\propto f(x|\theta)\pi(\theta). \tag{1.1.5}$$

其中符号"$\propto$"表示"正比于"的意思．在(1.1.5)式中，$\pi(\theta)$ 和 $f(x|\theta)$ 也可以仅取依赖于 θ 的因子，称为分布的核．例如，二项分布 $B(n,\theta)$ 的核是 $\theta^x(1-\theta)^{n-x}$，Poisson 分布 $P(\lambda)$ 的核是 $\lambda^x e^{-\lambda}$，正态分布 $N(\mu,\sigma^2)$ 的核是 $\exp\left\{-(x-\mu)^2/2\sigma^2\right\}$，Gamma 分布 $Ga(\alpha,\lambda)$ 的核是 $x^{\alpha-1}e^{-\lambda x}$，Beta 分布 $Be(a,b)$ 的核是 $x^{a-1}(1-x)^{b-1}$．将 $\pi(\theta)$ 和 $f(x|\theta)$ 的核代入(1.1.5)式中可得到 $\pi(\theta|x)$ 的核．熟悉分布的核可以简化后验分布的计算．

例 1.1.1　设 $X_1,X_2,\cdots,X_n$ 是来自两点分布 $B(1,\theta)$ 的独立同分布(iid)样本，θ 的先验分布为 Beta 分布 $Be(a,b)$，其中 $a>0$，$b>0$．求 θ 的后验密度 $\pi(\theta|x)$．

解　样本的联合密度函数为 $f(\theta|x)=\theta^T(1-\theta)^{n-T}$，其中 $T=\sum_{i=1}^n x_i$，先验密度 $\pi(\theta)$ 的核为 $\theta^{a-1}(1-\theta)^{b-1}$．因此，由(1.1.4)式可得

$$\pi(\theta|x)=c(x)\theta^{T+a-1}(1-\theta)^{n-T+b-1}.$$

由于随机变量为 θ，所以上式为 Beta 分布 $Be(T+a,\ n-T+b)$ 的密度函数，其中 $c(x)$ 可由该分布的常数决定，即

$$c(x)=\frac{\Gamma(n+a+b)}{\Gamma(T+a)\Gamma(n-T+b)}.$$

贝叶斯估计问题中，常数因子 $c(x)$ 通常不需要求出明确的表达式．

1.1.3　点估计问题

在获得参数 θ 的后验分布后，可以用后验均值

$$\hat{\theta}_B=E(\theta|x)=\int_\Theta \theta\pi(\theta|x)\mathrm{d}\theta=\frac{\int_\Theta \theta f(x|\theta)\pi(\theta)\mathrm{d}\theta}{m(x)},$$

作为 θ 的估计量，当然也可以用后验分布的中位数或后验众数作为 θ 的估计量．

1.1.4　假设检验问题

设假设检验问题的一般形式是

$$H_0:\theta\in\Theta_H \leftrightarrow K:\theta\in\Theta_K,$$

此处 $\Theta_H\cup\Theta_K=\Theta$，$\Theta$ 是参数空间，Θ_H 是 Θ 的非空子集．

获得参数 θ 的后验分布后，计算 Θ_H 和 Θ_K 的后验概率：

$$p_H(x)=P(\theta\in\Theta_H|x),\ p_K(x)=P(\theta\in\Theta_K|x).$$

若 $p_H(x)>1/2$,则接受 H;否则,拒绝 H.

1.1.5　区间估计问题

在求得 θ 的后验密度 $\pi(\theta|x)$ 后,求统计量 $A(x)$ 和 $B(x)$,使得后验概率

$$p(A(x)\leqslant\theta\leqslant B(x)|x)=1-\alpha.$$

其中 $0<\alpha<1$ 为常数,则称 $[A(x),B(x)]$ 为 θ 的可信度为 $1-\alpha$ 的可信区间.

例1.1.2　设随机变量 X 服从二项分布 $B(n,\theta)$, θ 的先验分布为 $(0,1)$ 上的均匀分布 $U(0,1)$.求 θ 的贝叶斯点估计.

解　X的概率密度和 θ 的先验密度分别为

$$f(x|\theta)=C_n^x\theta^x(1-\theta)^{n-x}(x=0,1,2,\cdots,n),$$

$$\pi(\theta)\equiv1(0<\theta<1).$$

X和 θ 的联合分布是

$$h(x,\theta)=C_n^x\theta^x(1-\theta)^{n-x}(x=0,1,2,\cdots,n;0<\theta<1).$$

X的边缘分布是

$$m(x)=\int_0^1h(x,\theta)\mathrm{d}\theta=\int_0^1C_n^x\theta^x(1-\theta)^{n-x}\mathrm{d}\theta=\frac{1}{n+1}.$$

θ 的后验分布是

$$\pi(\theta|x)=\frac{h(x,\theta)}{m(x)}=\frac{C_n^x\theta^x(1-\theta)^{n-x}}{1/(n+1)}$$

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1}.$$

即 $\theta|x$ 服从Beta分布 $Be(x+1,n-x+1)$.

若取 θ 的贝叶斯估计为后验均值估计,则有

$$\hat{\theta}_B=E(\theta|x)=\frac{x+1}{n+2},$$

而经典统计方法下 θ 的极大似然估计(MLE)是

$$\hat{\theta}=\frac{X}{n}.$$

比较 $\hat{\theta}_B$ 与 $\hat{\theta}$,易见贝叶斯估计 $\hat{\theta}_B$ 更合理.因为当 $x=0$ 或 n 时, $\hat{\theta}=0$或1,其

值太极端了;而当 $x=0$ 或 n 时,$\hat{\theta}_B=1/(n+2)$ 或 $(n+1)/(n+2)$,既不为0,也不为1,当分别接近0或1,看上去显得更合理些.

§1.2 先验分布的选取

在贝叶斯统计中,一个重要的问题是如何利用先验信息合理地确定先验分布,它是从贝叶斯统计诞生之日起就伴随着的一个颇具争议的问题.贝叶斯学派对此做了大量的研究,获得了一些重要的成果.尽管这些成果谈不上是人们普遍接受的,但是有些原则性的考虑还是可以谈一谈的.

1.2.1 主观概率

贝叶斯统计中要使用先验信息,而先验信息主要是指经验和历史资料.

在经典统计中,概率是用非负性、正则性和可列可加性三条公理定义的.概率的确定方法主要是两种,一种是古典方法,另一种是频率方法.实际中大量使用的是频率方法,所以经典统计的研究对象是能大量重复的随机现象,而不是这类随机现象的就不能用频率方法去确定其有关事件的频率.贝叶斯学派完全同意概率的公理化定义,但认为概率也可以用经验确定,这与人们的实践活动一致,这样就可以使不能重复或不能大量重复的随机现象也可谈及概率,同时也使人们积累的丰富经验得以概括和应用.贝叶斯学派认为,一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信息,这样给出的概率称为主观概率.

确定主观概率的方法如下:

根据经验和历史资料的先验信息给出的主观概率没有固定的模式,但不管用什么方法,其所确定的主观概率都必须满足概率的三条公理,即

(1)非负性公理:对任一事件 A, $0<P(A)<1$.

(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

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

当发现所确定的主观概率与这三条公理及其推出的性质不和谐时,必须立即修正,直到和谐为止.这时给出的主观概率才能称得上概率.