张金槐　唐雪梅　编著

BAYES方法

国防科技大学出版社

（修 订 版）

Bayes 方法

（修 订 版）

● 张金槐　唐雪梅　编著

● 国防科技大学出版社

[湘]新登字009号

内 容 简 介

本书主要论述Bayes统计方法和统计决策的基本理论。内容包括Bayes统计推断、统计决策的基本理论及序贯Bayes决策方法。本书的特点是充分注意了工程技术中的应用，比较详细地讨论了验前信息的获取及其表示，并给出一些应用的例题。这些例题虽然是特殊的，但应用中的思想方法具有普遍意义。

本书可作为研究生教材和大学本科有关专业的选修课教材，也可供工程技术人员参考。

Bayes 方法
（修订版）
张金槐　唐雪梅　编著
责任编辑　何　晋
封面设计　陆荣斌
*
国防科技大学出版社 出版发行
国防科技大学印刷厂印装
*
开本：850×1168　1/32　印张：9.8125　字数：246千字
1993年9月第2版第1次印刷　印数：1—1000册
ISBN　7-81024-075-7
O·9　定价：6.50元

修订版前言

本书出版之后，我们应邀在航空航天部及有关科研、试验场地等单位，进行了多次专题讲座。使本书所涉及的内容受到越来越广泛的关注。一个共同的特点是：期望在较少的试验数之下作出试验结果的统计推断。为了适应当前的需要，并结合工程实践中提出的问题，由张金槐同志将第一版进行适当改写。在再版中，突出了应用部分，增设了Bayes 方法在飞行器试验分析中的应用以及在复杂系统可靠性评估中的应用（第4、5两章）。在应用中，注意了小子样的情况以及验前信息的获取和表示。对工程实践中的问题，提出了看法。

编著者

1993年元月

前　　言

本书讨论Bayes 方法和统计决策的基本理论。它们是统计数学中一个活跃的领域。六十年代初，在飞行器试验结果的统计推断中，一些科技工作者曾考虑应用这种方法，但限于当时的技术条件，以及关于验前信息的获取和信息的转换等方面还存在着不少有待解决的问题，没有得到应有的重视。近年来，由于科学技术的发展，Bayes 方法已在国内外得到广泛应用，而且这种统计思想在工程技术界越来越被人们所接受。因此，我们编写了这本书，希望能对读者有所帮助。

本书的初稿，最初在航天动力学与飞行试验专业举办的Bayes 方法讨论班上使用。后来，又作为选修课教材向本科生讲授。近年来，在有关的科研任务中运用了这种方法，获得了较好的效果。我们参照国内外有关文献、资料，将原有内容进行了修改和补充形成此书。

Bayes 方法和统计决策理论可广泛地应用于通信、控制、人工智能、地震和地质探矿、气象预报、经济、管理等很多领域。书中虽仅列举了在飞行器的精度和可靠性分析中的一些应用，但应用中的思想方法具有一般意义。

本书可作为本科生有关专业的选修课教材和研究生教材，也可作为工程技术人员的参考书。对于那些希望在少量试验次数之下作出统计推断的读者将有所收益。本书中有些章节的内容打了“*”号，初学者可以暂时略去，它不会影响后续内容的学习。由于作者学识水平所限，书中会有不少缺点和值得改进的地方，期待读者批评指正。

在本书写作过程中，作者参考了中国科学院系统科学研究所成平同志的“贝叶斯统计与可靠性分析”学术报告。南京大学茅宁博士对本书的原稿提出了不少中肯的意见，在此一并表示衷心感谢。

编著者

1988年12月于国防科技大学

目　　录

3 统计决策基本理论

4 Bayes方法在飞行器试验分析中的应用

5 Bayes方法在可靠性评估中的应用

1 导　　论

1.1 问题的引起

读者在学习了数理统计方法以后，可曾想过：古典的统计推断方法还会有值得商榷的地方吗？说商榷，不是指某一个具体的解题方法，而是指它的根本的学术观点。让我们从古典统计学中的假设检验方法和区间估计方法说起。因为它们可算得上是古典统计方法的代表。

例 1.1.1. 设$X \sim N(\theta, 1)$，要求检验下列统计假设

$\mathscr{H}_0$:　$\theta \leqslant 0$,

$\mathscr{H}_1$:　$\theta > 0$.

令显著性水平$\alpha = 0.05$，我们只抽样一次。此时，由Neyman检验理论，取临界区域$\{X;\ X > C\}$，使

$$P\{X > C \mid \mathscr{H}_0\} = \alpha$$

运用一致无偏有效检验（UMP检验），C满足

$$\Phi(C) = 1 - \alpha,$$

其中
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

当$\alpha = 0.05$时，$C \cong 1.6$.如果抽到的样本值$X = 10^{10}$，总可以断言$\mathscr{H}_0$是不真的，因此拒绝假设$\mathscr{H}_0$. 但按上述检验法则，犯第一种错误即弃真的概率为5%. 如果抽到的样本值$X = 2$，这时也拒绝$\mathscr{H}_0$，犯错误的概率也是5%. 按人们的常识，在第一种

情况下拒绝 $\mathscr{H}_0$ 较之第二种情况，其“信度”（可靠的程度）要高。古典检验方法却回答不了这个问题。

例 1.1.2 设 $X \sim \mathscr{U}\left(\theta-\frac{1}{2}, \theta+\frac{1}{2}\right)$，即 X 为在 $\left(\theta-\frac{1}{2}, \theta+\frac{1}{2}\right)$ 上均匀分布的变量。设有 i.i.d（即独立同分布）样本 $X_1, \cdots, X_n$. 要给出 θ 的置信区间。

已知，当 $n=25$，且置信度为95%时，θ 的置信区间为 $(\hat{\theta}-0.056, \hat{\theta}+0.056)$. 此处 $\hat{\theta}=\frac{1}{2}[\min\{X_i\}+\max\{X_i\}]$，为样本中值。今假定由试验结果，获得 $\min\{X_i\}=3.1$，$\max\{X_i\}=3.2$. 于是按古典的置信估计方法，具有置信度为95%的 θ 的置信区间是(3.096，3.206)．但是，当从另一个角度去考虑，在上述样本之下，θ 必位于 $\max\{X_i\}-\frac{1}{2}=2.7$ 和 $\min\{X_i\}+\frac{1}{2}=3.6$ 之间，如果没有其它的关于 θ 的信息，置信区间的信度应为 $(3.206-3.096)/(3.6-2.7)=0.124=12.4\%$. 这个结果是按照已经出现了样本值的最大、最小值后确定的。这与古典方法所说的 95%的置信度分庭抗礼！此外，如果在容量为25的样本中，$\min\{X_i\}=3.0$，$\max\{X_i\}=3.96$，那么 θ 必定（以100%的信度）在(3.46，3.5)之中，但古典的置信区间为（3.424，3.536），它的置信度为95%. 这样，古典的区间估计所得的结果与人们的常规认识具有较大的差异！

发生上例中情况的根本原因在于：古典统计基础建立在概率的频率意义上。统计推断方法反映了大量试验结果的统计规律性。因此假设检验中犯错误的概率和区间估计中的置信度都必须从大量重复独立试验（观察）的角度去解释。例如，在例1.1.2中，样本容量为25时，取置信度为 95% 的置信区间为

$(\hat{\theta}-0.056, \hat{\theta}+0.056)$，只能说明在获取大量的容量为25的样本之下，有95%的置信区间含有真实参数θ. 而具体得到一个样本所构成的置信区间应看作无限的可能结果中的一个。这种置信度是事先就指定了的，它与所获得的样本是无关的。但是，当人们从实用的角度看问题时，把信度和所得的样本联系在一起了。因此，产生了上述相悖的结论。

由此，在统计学上出现了另一种不同的观点，这就是Bayes学派的观点。他们认为，θ既然是未知的，而在试验中，即使能对它进行观察，也只能得到它的表现值。例如，X为Bernoulli 变量，p为其未知的分布参数，$p=P\{X=1\}$。那么，在容量为n的样本$\boldsymbol{X}=(X_1,\cdots,X_n)$被抽取之后，$p$的表现值可由$\boldsymbol{X}$确定，它为$\nu/n$. 这里$\nu$为$X_1,\cdots,X_n$中出现1的个数。因此，将$\theta$看作随机变量。此外，在试验之前，人们对它总会有某种认识。即是说，它具有验前信息。于是θ具有验前分布。再者，对于θ的统计推断，应与样本的表现值有关。在这样的思考方法之下，形成了Bayes 统计方法。这种方法在开始形成阶段，在强大的古典统计学派面前，没有引起重视。甚至一些古典统计学权威认为它不是一种科学方法，而仅仅是一种逼近方法而已。但是，本世纪以来，特别是五十年代以来，Bayes 统计学得到了新的发展。[1]-[4] 在国内，六十年代初，在国防科技部门得到了较充分的重视。因为在这些部门，总是希望能运用历史信息（如研制新产品中的继承性），而且能减少试验次数，不容许有较大的样本容量。Bayes 统计方法恰好满足了这种愿望。近年来，国内不少统计学者重视了这方面的工作，而且在不少专著中反映了Bayes 统计方法的主要内容，[5],[6]目前正在不断发展之中。

1.2 Bayes统计的基本观点及争议

考察一个随机试验。在这个试验中，有 n 个互相排斥的、竭尽可能的事件 $A_1, \cdots, A_n$。如果以 $P(A_i)$ 表示事件 A_i 发生的概率，那么 $\sum_{i=1}^{n} P(A_i)=1$. 记 B 为任一事件，则有

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}, \qquad (1.2.1)$$

$$i=1,2,\cdots,n.$$

(1.2.1)就是著名的 Bayes 公式。其中 $P(A_1)$, $\cdots$, $P(A_n)$是在做试验之前就已经知道的。这种知识，就是验前信息，而且由于 $\sum_{i=1}^{n} P(A_i)=1$，因此，这种验前信息，以概率分布$\{P(A_1)$, $\cdots$, $P(A_n)\}$的形式给出，所以常称为**验前分布**。假定在做试验中事件 B 发生了，由于这个新情况的出现，对于事件 A_1, $\cdots$, A_n发生的可能性有了新的认识。它们发生的概率由(1.2.1)给出。它是在试验之后给出来的。因此是验后的知识。而且，由于$P(A_i|B)\geqslant 0$, $\sum_{i=1}^{n} P(A_i|B)=1$，所以验后的概率 $P(A_i|B)$, $i=1,\cdots,n$ 形成了一个分布，称为**验后分布**。它综合了验前信息和试验所提供的新信息，形成了关于A_i发生可能性大小的当前认识。这个由验前信息到验后信息的转化，是 Bayes 统计的特征。

例 1.2.1 在飞机着陆时，驾驶员看不见起落架是否放下，因此装有一个报警灯。当起落架没有放下时则报警灯亮。但有时即使起落架放下了，报警灯也亮。设A_1表示“起落架放下”的事件，A_2表示“起落架没有放下”的事件。B表示“报

警灯亮”的事件。假定已知

$$P(B|A_1)=0.005,$$
$$P(B|A_2)=0.999,$$
$$P(A_1)=0.997,$$
$$P(A_2)=0.003.$$

现在要来确定当报警灯亮时起落架已放下的概率。

由公式(1.2.1)，我们得

$$P(A_1|B)=\frac{0.005\times0.997}{0.005\times0.997+0.003\times0.999}=0.62$$

这个值表示了一种统计规律性。在飞机着陆之前（验前），$P(A_1)=0.997$，因此，平均说来，在一千次飞行中，约有三次放不了起落架。但在报警灯亮了的情况下，这个平均数增加到380，也许人们要问，为什么当报警灯亮时，“起落架已放下”还会有这么大概率？这是由于验前信息告诉我们“起落架已放下”的概率很大，$P(A_1)=0.997$。因此起落架放下时，报警灯还可能会亮。

例 1.2.2 假定某工厂每日对它生产的产品进行抽样调查，以估计当日产品的废品率θ。就整批产品而言，θ当然是一个固定的未知的量，并无随机性可言。但逐日的废品率多少有些波动。从长期看，就可以把“一日的废品率”当作一个随机变量θ，而某个特定废品率则是这个随机变量的一个抽样值。由过去长期检查的数据，对θ的了解已经积累了一些知识，这就是θ的验前信息。这些知识有助于对废品率作出更切合实际的估计。

在连续随机变量X的场合，公式(1.2.1)成为

$$\pi(\theta|x)=\frac{f(X|\theta)\pi(\theta)}{\int_\Theta f(X|\theta)\pi(\theta)d\theta} \tag{1.2.2}$$

其中$f(X|\theta)$为给定分布参数θ之下的X的密度函数，$\pi(\theta)$为θ的验前密度函数。$\pi(\theta|x)$为θ的在给定X之下的密度函数，Θ为参数空间。试验之中，如果将X看作样本，$f(X|\theta)$就是θ给定后的样本的密度函数。$\pi(\theta)$反映了在试验之前对θ的认识，而$\pi(\theta|X)$则为试验之后（在获得了样本X之后）对θ取各种可能值的大小的新的认识，称$\pi(\theta|X)$为**验后分布密度**。Bayes统计推断均以$\pi(\theta|X)$作为出发点。例如，以后将会看到，当获得样本$\boldsymbol{X}=(X_1,\cdots,X_n)$之后，$\theta$的估值可取

$$\hat{\theta}=E[\theta|\boldsymbol{X}]=\int_{\Theta}\theta\pi(\theta|\boldsymbol{X})d\theta \tag{1.2.3}$$

如果要对θ作置信估计，在置信度为$1-\alpha$之下，令

$$\int_{\theta_1}^{\theta_2}\pi(\theta|\boldsymbol{X})d\theta=1-\alpha, \tag{1.2.4}$$

满足上式的$(\theta_1(\boldsymbol{X}),\ \theta_2(\boldsymbol{X}))$，就是$\theta$的置信区间。这时

$$P\{\theta_1(\boldsymbol{X})<\theta<\theta_2(\boldsymbol{X})\ |\boldsymbol{X}\}=1-\alpha$$

它表示了在给定样本$\boldsymbol{X}$之下，随机变量θ落在$(\theta_1(\boldsymbol{X}),\theta_2(\boldsymbol{X}))$中的概率为$1-\alpha$. 这种解释与古典统计是完全不同的。

上面的讨论，仅涉及Bayes方法的一些基本观点。A. Wald在本世纪四十年代末提出了统计决策观点。对统计推断方法作了新的解释，其中包含了Bayes方法。这些问题将在后续的讨论中看到。这里暂时不展开讨论了。

Bayes方法，当然不是无懈可击。古典学派恰好在Bayes统计的两个基本点上进行了抨击。这就是θ的随机性以及它存在验前分布。他们认为，θ是未知分布参数，它是客观存在的一个常量。例如光速、钢中的含碳量等，都是确定性的量。未知的东西怎么会是随机的？其次，在试验之先，对未知的θ要给出它的分布，这点也是难于办到的。

站在不同的观点，统计推断问题可能会得出截然不同的结果。不妨举一个五十年代由C. Stein[7]提出的一种现象。设

$X_i \sim N(\theta_i, 1)$，$i=1, \cdots, n$，它们为独立的随机变量，但具有不同的分布参数θ_i．可以证明

$$P\left\{\sum_{i=1}^{n} X_i^2 > \sum_{i=1}^{n} \theta_i^2 \mid \theta_1, \cdots, \theta_n\right\} > 1/2$$

上述不等式中，分布参数$\theta_1, \cdots, \theta_n$看作常量。因此这是由古典学派的观点得到的不等式。另一方面，如果假定θ_i为在$(-\infty, +\infty)$上具有验前的均匀分布的密度，即 $\pi(\theta_i) \propto 1$（这种形式的验前分布在后续讨论中将涉及），则可以证明

$$P\left\{\sum_{i=1}^{n} \theta_i^2 > \sum_{i=1}^{n} X_i^2 \mid X_1, \cdots, X_n\right\} > \frac{1}{2}$$

上述不等式中，$X_1, \cdots, X_n$是给定的，而θ_i看作随机变量。因此这是由 Bayes 学派的观点得到的不等式。由此看出，两种结果，完全不同。要去分辨哪一种正确，这是惑人的事。这里不进行这种争辩。

我们认为，由于客观世界的多样性，自然现象呈现出错综复杂的现象。人们的认识总会有差异。由此引起争议也是自然的事。双方可以在争议中丰富自己，有利于学科的发展。至于存在的一些弊端，可以在不断的研究实践中加深对一些问题的认识，并由此检验我们认识的正确性。

1.3 内容梗概

为了使读者对本书所论述的内容有一个大致的了解，这里介绍本书讨论的主要问题。

由于验前分布的获取和表示以及验后分布在 Bayes 统计中的重要作用，我们花了较多的篇幅，这就是第二章中论述的主要问题。在论述中，注意了那些验前、验后具有同一分布形式的所谓**共轭分布**。读者将会发现，有较大一类分布族具有这种

共轭性质。这给 Bayes 统计推断带来不少方便。在讨论中，我们较多地注意了正态总体中未知方差的统计推断问题。这是因为在实践中，它们是经常出现的问题。论述中，给出了较多的具有实际应用背景的例子。

在第三章中，讨论了Bayes统计决策(Statistical Decision，或称统计判决）的基本理论。读者将会发现，损失函数以及所采用的策略将是决策模型中的重要的要素。我们将看到，统计估计问题、假设检验问题等，都是人们去作出某种决策。它们都可以得到决策论的解释。在讨论中，主要介绍统计决策的两种理论——Bayes 决策及 Minimax决策理论。我们的注意力更多地集中于 Bayes 决策。为了比较决策的好坏，文中论述了决策的容许性问题。这是一个饶有兴趣的问题。我们将介绍Stein的一些结果。

第四章给出 Bayes 决策的某些应用。应用的领域是广泛的，不可能一一涉及。这一章，仅仅涉及正态总体方差检验及产品可靠性分析等问题，而且以再入飞行器的试验统计作为实际背景。因此，只能说是挂一漏万。但是我们指出，其他方面的应用途径也是相仿的。

为了考虑到科技应用中小样本的需要，专门讨论了序贯决策问题——第五章。序贯方法的特点是样本容量不是事先就确定了的，而是在每次试验之后，连同历次试验的结果，考虑是否能作出决策。如果尚不宜作出决定，则再进行下一次试验。因此，样本容量是随机的。这种决策过程较之固定样本容量之下的决策要复杂得多。主要论述两方面的内容：终止试验的规则及终止试验后的决策规则。还讨论了序贯截尾决策问题。所谓截尾，就是在一定的试验次数之后，必定要作出决定，不再进行下一次试验。我们认为，对于昂贵的消耗性试验来说，采取这种方法是十分必要的。