

学 术 著 作 丛 书

Bayesian Methods in Test Analysis

Bayes试验分析方法

张金槐 刘 琦 冯 静 编著

国防科技大学出版社

国防科技大学学术
著作专项经费资助出版

国家自然科学基金
基金资助出版

Bayes 试验分析方法

张金槐 刘 琦 冯 静 编著

国防科技大学出版社
湖南·长沙

内 容 提 要

本书系统全面论述了 Bayes 方法及其在武器装备试验分析与评估中的应用。对当前 Bayes 方法所探讨的多源融合技术和动态分布参数的估计进行了分析。适合从事武器装备试验分析与评估的科研技术人员阅读。

图书在版编目(CIP)数据

Bayes 试验分析方法/张金槐,刘琦,冯静编著. —长沙:国防科技大学出版社,2007.3

ISBN 978-7-81099-373-9

I. B… II. ①张…②刘…③冯… III. 试验分析(数学)—应用—产品质量—分析 IV. ①O212.6 ②F273.2

中国版本图书馆 CIP 数据核字(2006)第 106474 号

国防科技大学出版社出版发行

电话:(0731)4572640 邮政编码:410073

<http://www.gfkdcbs.com>

责任编辑:谷建湘 徐 飞

新华书店总店北京发行所经销

国防科技大学印刷厂印装

*

开本:850×1168 1/32 印张:8.25 字数:214 千
2007 年 3 月第 1 版第 1 次印刷 印数:1-1000 册

ISBN 978-7-81099-373-9

定价:25.00 元

序

本书论述的问题主要是运用 Bayes 方法进行试验结果分析。近年来,国防科技领域十分重视试验分析与评估的理论和方法,国内已出版了不少论著。从所论述的思想方法来说,经典方法较充分,Bayes 方法也有新的发展。然而,要适应当前发展的形势和需要,还有不少工作可做。例如,由于历史经验的积累、仿真技术的发展以及测控手段的多样化,使试验分析工作面临新的情况。一些部门提出了研制、仿真、定型试验一体化问题。此外,对于一些价格昂贵的产品,试验子样容量受到限制,但是,在试验之前有多种信息可利用,因此,在试验中采取“试试看看,看看试试”的方式,当试验发现问题时,进行改进,再进入下一次(批)试验。这样,某些被考核的性能参数是可变的。由于上述发展态势,必须在原有的试验分析和评估理论、方法的基础上不断更新内容,特别是结合工程实际背景进行理论的应用研究。本书正是为了适应这些工程需求,结合编者多年装备试验分析的实践,在

对近年来所做研究成果进行总结分析的基础上完成的。

本书撰写过程中,得到国防科技大学信息系统与管理学院系统工程系领导和同志们多方关怀,在此表示诚挚的谢意。段晓君、王华伟博士等,经常与我们一起研讨问题,为本书的编写和出版提供了许多的宝贵建议,应该说,本书也包含了他们的劳动成果。周忠宝、董豆豆、赵熠、胡超斌、高顺川等在书稿录入等方面做了不少工作,付出了艰辛劳动。在此,再次表示感谢。

本书的完成得到了国防科技大学学术著作专项经费与国家自然科学基金(“小子样条件装备可靠性试验评定方法”,编号:70571083)的资助。

阅读本书,必须具备 Bayes 统计学的基本知识。作者期望本书的出版对读者能有所裨益,并热忱欢迎提出各种不同的看法和批评意见。

作 者

目 录

第一章 概 论

- 1.1 试验分析的动态 (1)
- 1.2 对 Bayes 试验分析方法的疑虑 (3)
- 1.3 当前试验分析和评估的特点 (4)
- 1.4 本书讨论的主要内容 (5)

第二章 试验分析中的 Bayes 融合技术

- 2.1 问题的提出 (7)
- 2.2 验前信息的相容性问题 (8)
- 2.3 验前信息可信度的定义及计算 (23)
- 2.4 多源验前信息的融合验前分布 (27)
- 2.5 多源验前信息的融合验后分布及其应用
..... (35)
- 2.6 多源验前信息的 Bayes 融合鉴定方法 (43)
- 2.7 Bayes 融合技术的一些应用 (50)

第三章 Bayes 序贯分析

- 3.1 引言 (82)
- 3.2 序贯验后加权检验 (84)
- 3.3 Bayes 序贯概率比检验 (102)
- 3.4 正态总体下分布参数的 Bayes 序贯估计 (115)

3.5 多元正态总体分布参数的 Bayes 序贯验后加权 检验及估计	(125)
---	-------

第四章 动态分布参数的 Bayes 分析

4.1 一维动态分布参数的 Bayes 估计	(134)
4.2 ML - II 建模	(144)
4.3 可靠性增长的 Bayes 信息融合技术	(149)

第五章 Bayes 统计推断的稳健性

5.1 引 言	(164)
5.2 验前分布的稳健性分析	(165)
5.3 验后稳健性分析	(176)
5.4 稳健性检验	(182)
5.5 应用举例	(187)
5.6 动态分布参数 Bayes 估计中验前分布的稳健性 分析	(191)
5.7 验前分布稳健性分析应用举例	(205)

第六章 单发命中概率的估计

6.1 命中概率的表示	(213)
6.2 单发命中概率的估计	(217)
6.3 当脱靶量具有系统偏差时的命中概率估计	(224)
6.4 命中长方形域的概率的 UMVU 估计	(232)
6.5 小子样命中概率的估计	(243)
6.6 利用刀切法缩小估值偏倚的方法	(247)

参考文献	(252)
------------	-------

第一章

概 论

1.1 试验分析的动态

我国国防科技领域关于试验分析和评估的研究可以追溯到 20 世纪 60 年代初期。当时,国防部五院钱学森教授为解决工程实际中试验结果分析、鉴定和定型的理论方法问题,提出应重视试验学的研究,为此组建了“试验学专家组”。其主要工作是研究再入飞行器的精度分析与鉴定技术,这些工作涉及总体设计、各分系统的试验分析(包括可靠性)、测控系统精度分析和数据处理方法等。由于试验费用昂贵,试验的组织实施又是一个复杂的系统工程,因此优化试验程序、充分利用各种信息和小子样理论方法的应用研究等就提到了议事日程。特别是试验分析,不能完全套用经典方法(大样本下的鉴定方法),而必须研究新的理论和方法。

20 世纪 60 年代,专家组对再入飞行器战术技术指标评定进行了重点研究。由于不能进行大量试验,因此运用了 Bayes 方法。对于散布和系统性偏差的鉴定运用了 A. Wald 的序贯概率比 (SPRT) 检验和截尾序贯检验法^[1]。这种序贯法较之通常的(经典的)检验法减少了子样容量。然而,由于当时计算条件(如仿真试验)及技术条件的限制,没有对 Bayes 统计推断法进行深入研究。

当然还有认识问题,例如由于经典统计学派及习惯势力的影响,学术界对应用 Bayes 方法尚存疑虑。此外, Bayes 方法本身还存在一些关键性理论问题需进一步探索。因此,科技领域没有普遍应用 Bayes 方法。

随着 Bayes 方法理论研究的深入和科技事业的飞速发展,逐步积累了研制经验,特别是随着计算技术的发展,人们逐渐认识到运用各种信息,包括历史的、不同场合下的试验信息,将有可能使现场标准条件下的试验次数减少。而 Bayes 方法的思路恰好迎合了这种想法。此时,国外对 Bayes 方法的运用也做了种种努力和研究,如验前信息的获取及表示、统计决策理论的研究、不同信息 Bayes 融合方法等,为 Bayes 理论的运用创造了极为有利的条件。20 世纪 70 年代后,国内对信息利用的理论和方法也开展了较充分的研究, Bayes 方法受到了重视,相继开展了多方面的理论和应用研究;90 年代以来,得到了进一步的发展,国内组织了多次研讨会和技术交流。

18 世纪 60 年代, T. R. Bayes 发表了重要论文,但由于经典统计学派的强大以及 Bayes 的论文尚存在理论和应用上的问题,因此没有得到充分的发展。直到二次大战以后, A. Wald 提出了统计决策理论,在该理论中, Bayes 解被认为是一种最优决策函数,引起了不少学者对 Bayes 方法的研究兴趣。此后,在 Jeffreys^[2], Savage^[3], Robbins^[4], Raiffa & Schlaifer^[5], Box & Tiao^[6], Lindly^[7], Definetti^[8], Berger^[9] 等学者的努力下, Bayes 方法在理论、观点和方法方面得到不断完善,为 Bayes 方法的发展创造了极为有利的条件,得到了广泛的应用。在国外, Bayes 学派已发展为一个非常有影响的统计学派。特别是在国防科技领域,国外重视 Bayes 方法的运用。早在 1984 年,美国陆军部长助理就明确指出:破坏性试验必须运用序贯分析方法或 Bayes 方法来确定系统的可靠性,进行精度鉴定,最佳试验数的确定必须考虑试验耗费。运用 Bayes

方法,结合仿真、虚拟现实技术的发展,使试验次数明显减少。

1.2 对 Bayes 试验分析方法的疑虑

在国内外普遍重视 Bayes 试验分析方法的理论与应用研究状态下,国内对此方法仍存在一定的疑虑。这里,对一些不同的认识进行剖析:

(1)经典的试验统计理论和方法已应用多年,试验规程也早已成熟,另行提出 Bayes 方法有何必要和先进性?这里,涉及对经典统计的认识和 Bayes 小子样理论的理解。经典统计方法并非无懈可击。当然,在 Bayes 方法应用中,由于缺乏必要的技术条件(如仿真条件和经费支持),实现过程中也会出现困难。这不是 Bayes 理论和方法本身的问题。积极的态度是创造条件,逐步实现试验鉴定技术的革新。那种固守原有阵地,不愿研究新问题的做法是值得商榷的。至于对 Bayes 方法有一些不同看法,对有关问题进行争议,这无关紧要,科学可以在争议中发展。而工程应用中的问题,可通过不断实践来改进和完善。至于试验规程早已成熟,习惯已成自然,也不能说明不需改变。事实上,在试验中新方法的应用,试验优化程序(试验设计)的出现等,说明事物总是不断变化的,一成不变的东西是没有的。

(2)应用 Bayes 方法进行试验结果分析和鉴定,子样容量减少了,所作出的分析和推断可信吗?人们的担心可以理解。不过,所谓小子样理论,并不是“少用信息”的理论。小子样是指现场试验次数少。Bayes 方法的特点之一就是:在运用现场信息的同时,充分利用其他信息,这就是验前信息,例如,试验前可利用的历史信息、仿真信息、专家信息等。当然,不同信息源如何融合,如何使用,可信度如何等等,又都是 Bayes 方法所关心的问题。对这些发

展中的课题需认真研究。可以说,根据 Bayes 理论,在充分正确使用各类验前信息的条件下,结合现场少量的试验子样,是完全可以得到可信的试验鉴定结果的。

(3) Bayes 小子样试验鉴定方法,运用了多种验前信息,特别是仿真信息,容量可以很大。这样,现场试验为小子样,而验前信息为大容量。于是有人担心现场试验的信息将被验前信息“淹没”,而不起重要作用了,或者说,由验前信息完全决定了鉴定的结果,因此对 Bayes 方法存有疑虑。这种看法是可以理解的。然而,关键问题是对验前信息的看法及其使用。如果验前信息本身是主观的设定或者可信性不高,而使用中又是将验前信息和现场信息简单地混合使用,那么就必然会发生上述人们的担心。对于上述问题的解决在于科学地运用验前信息和现场试验信息,讲究鉴定方案中多种信息的融合方法。而这些问题正是小子样技术中要研究的根本问题。

1.3 当前试验分析和评估的特点

试验分析、评估技术必须与时俱进,适应发展的需要。在我国,试验分析与评估呈现出如下的特点:

(1) 随着高新技术装备的研制和开发,试验任务在深度和广度两方面都有了新的发展,使传统的试验分析、鉴定方法受到了挑战。

(2) 由于一些装备系统复杂、价格昂贵,对于破坏性的试验不能大量地进行,因此,提出了小子样试验下的系统鉴定技术问题。

(3) 由于测控手段的多样化,可以在试验分析中,具有多种可利用的信息,而各种信息的获取情况各异,因此获取到的信息,未必属于同一总体。因此,多传感器下的信息融合技术,是试验分析

中需考虑的重要问题。

(4)由于仿真技术的发展,对装备性能的考核可以多视角、多途径地进行。因此,仿真信息与试验鉴定一体化问题,又是当前试验分析中的一大特色。

(5)在不同状态、环境下的测控设备精度鉴定技术,已成为提高试验数据质量的重要途径。

1.4 本书讨论的主要内容

结合上述特点,本书将论述的主要内容有:

(1)多源信息的 Bayes 融合技术

在小子样 Bayes 分析中,具有多种可利用的信息,例如多源验前信息。这些信息的使用必须十分慎重和讲究。论述中,将研究多种信息的可信度问题,而在 Bayes 融合方法中,将考虑不同信源的可信度,由此给出合理的、实用的 Bayes 融合方法。

(2)Bayes 序贯方法和 Bayes 序贯截尾方法

这一主题是为了适应当前试验中实施的“试试看看,看看试试”的试验策略的需要而安排的。序贯试验分析方案中,试验数不是预先设定的,而是视试验中出现的结果而定。因此,何时终止试验(停时)以及终止试验时采取的策略和风险分析,都是十分重要的。在论述中,这些问题将受到关注。

(3)动态参数的 Bayes 试验分析

目前,一些试验往往不是同一状态下的集中试验,而是分批试验。对于批次试验中出现的进行改进后,再进入下批次试验。因此,必须运用变动统计学方法完成参数的评估。动态参数的 Bayes 估计就属于这个范畴。论述中,将讨论参数的建模策略和 Bayes 估计方法。

(4) Bayes 试验分析中的稳健性研究

这是出于工程实际的需要安排的内容。Bayes 方法的争议,往往发生在验前信息的运用。例如,由验前的多源信息,确定出一个合理的验前分布必须有所讲究,而不同的验前分布将导致不同的 Bayes 试验分析结果。因此,验前分布的稳健性以及验后 Bayes 分析的稳健性是人们关心的问题。这是一个“使人放心应用 Bayes 方法”的敏感问题。

(5) 单发命中概率的估计

书中将论述椭圆和圆散布情况下的命中概率估计问题。在单发命中概率的表示中,它是一个积分公式,而积分中含有脱靶量的均值和方差两个未知参数。在一般书籍中,总是先去估计未知参数,然后代入积分式中计算出命中概率。这种方法所计算得到的命中概率估计不是无偏的,更不具有最小方差的性质。本书详细地论述了这个问题,且给出了单发命中概率的一致无偏最小方差估计(UMVU 估计)。同时,讨论了小子样下的 Bayes 估计方法。

试验分析和评估的内容很广。书中的论述,着重于试验鉴定中的问题,注意实际应用。作者期望能引起共同的讨论,并进一步完善和发展试验分析和评估的理论和方法。

第二章

试验分析中的 Bayes 融合技术

2.1 问题的提出

当前,试验分析与鉴定中涉及的一个共性问题是在小子样、甚至是特小子样的试验条件下,如何充分地利用各种验前信息,并将这些信息和现场少量的试验信息融合,给出性能参数的评估。特别是一些设计部门和试验技术条件较好的单位,运用了仿真信息,将它和其他验前信息结合,作出试验鉴定方案。因此,小子样理论中多源信息融合技术,已成为当前试验分析与鉴定的重要问题,是国防科技领域普遍关注的问题。

然而,由于对小子样技术的认识问题以及对多源信息的处理缺乏足够的研究,使用单位的一些技术人员在谈及 Bayes 理论和方法时,持怀疑甚至否定态度,或者限制验前信息(如仿真信息)的使用,认为 Bayes 方法是一种容易引起争议的方法。我们认为应该采取科学的、实事求是的态度,分析应用中发生的问题。Bayes 理论本身是严密的、科学的具有其先进性和特色。但是在运用中,如果采用的方法不当,如验前信息和现场信息简单地混合使用,或主观地设定验前信息等,则必然产生不良后果,因而使人们产生疑虑。科学地研究验前信息的使用问题,给出严密、科学的融合理

论,并对验前信息的不同取法进行相容性(与现场试验结果比较时的相容性)检验,然后作出试验分析和鉴定方案,这是人们所期望的。本章将对这些问题进行讨论。

2.2 验前信息的相容性问题

Bayes 方法的特色是运用验前信息,而这种验前信息来源于各种不同的途径,如历次试验的数据、仿真所获得的数据等。这些信息与现场试验的信息是否属于同一总体,这就是验前信息的相容性问题或一致性问题。在某些场合,人们常常应用简单的判别方法说明相容性问题。例如,图比较说明法、Theil 不等式系数法(TIC 法)、时序模型的趋势比较法等。这些方法直观,当两个样本的差异较大时,可以较好地判断是否属于同一总体。但它仅仅是一种定性的分析(比较)方法,容易引起争议。因此,我们提倡运用定量分析方法,在一定的置信水平下,判断两个子样的相容性^[10]。

相容性问题,一般可按样本的类别,分为静态子样和动态子样的相容性问题,前者指子样是随机变量的观测值,而后者则指随机过程的观测值(现实)。

这里,还必须特别指出的是:对于小子样试验分析问题,现场子样的容量较小、甚至是特小(子样容量可小于 5);而验前子样的容量一般比较大。这时,对相容性的判别将会出现一定的困难,这也是在讨论中需注意的新问题。

2.2.1 静态数据的相容性检验

当验前信息与现场试验数据为静态数据时,验前与现场数据的相容性判别,可以应用通常的统计假设检验的方法进行。一般地,常用非参数检验方法。在工程应用中,常用 Wilcoxon-Mann-Whitney 的秩和检验法,现介绍如下:

记 $X = (X_1, \dots, X_{n_1})$ 为验前子样, $Y = (Y_1, \dots, Y_{n_2})$ 为现场子样,引入互相竞择假设:

$H_0: X$ 与 Y 属于同一总体; $H_1: X$ 与 Y 不属于同一总体。

将两个子样混合,然后由小到大排序,得次序统计量

$$Z_1 \leq Z_2 \leq \dots \leq Z_{n_1 + n_2},$$

记 $X_k = Z_j$, 即 X 中的第 k 个元 X_k 在混合排序中名列第 j , 即 X_k 的秩为 j . 记作 $r_k(X) = j$. 作统计量

$$T = \sum_{k=1}^{n_1} r_k(X),$$

它为 X 的元的秩和。则可建立如下关系

$$P(T_1 < T < T_2 | H_0) = 1 - \alpha,$$

或者

$$P(T \leq T_1; T \geq T_2 | H_0) = \alpha.$$

其中 α 为显著性水平。给定 α 之下, T_1 、 T_2 有表可查(如表 2.1 所示)。于是,当获得子样 X 和 Y 后,计算秩和 T ,在检验水平 α 之下,如果:

- (1) $T_1 < T < T_2$, 则采纳 H_0 , 即 X 与 Y 属于同一总体;
- (2) $T \leq T_1$ 或 $T \geq T_2$, 则拒绝 H_0 , 即 X 与 Y 不属于同一总体。

表 2.1 秩和检验表 $P(T_1 < T < T_2 | H_0) = 1 - \alpha$

		$\alpha = 0.025$		$\alpha = 0.05$				$\alpha = 0.025$		$\alpha = 0.05$	
n_1	n_2	T_1	T_2	T_1	T_2	n_1	n_2	T_1	T_2	T_1	T_2
2	4			3	11	5	5	18	37	19	36
	5			3	13		6	19	41	20	40
	6	3	15	4	14		7	20	45	22	43
	7	3	17	4	16		8	21	49	23	47
	8	3	19	4	18		9	22	53	25	50
	9	3	21	4	20		10	24	56	26	54
	10	4	22	5	21						
3						6	6	26	52	28	50
	3			6	15		7	28	56	30	54
	4	6	18	7	17		8	29	61	32	58
	5	6	21	7	20		9	31	65	33	63
	6	7	23	8	22	7	10	33	69	35	67
	7	8	25	9	24		7	37	68	39	66
	8	8	28	27			8	39	73	41	71
	9	9	30	10	29		9	41	78	43	76
	10	9	33	11	31		10	43	83	46	80
4	4	11	25	12	24	8	8	49	87	52	84
	5	12	28	13	27		9	51	93	54	90
	6	1	32	14	30		10	54	93	57	95
	7	13	35	15	36	9	9	63	108	66	105
	8	14	38	16	36		10	66	114	69	111
	9	15	41	17	39						
	10	16	44	18	42	10	10	79	131	83	127

在表 2.1 中, n_1, n_2 比较小, 表中 $n_1: 2 \sim 10; n_2: 3 \sim 10$. 这样, 工程应用中将遇到困难。在小子样试验中, 以容量小的子样(即现场试验子样)作为 n_1 , 而验前子样容量, 作为 n_2 的取值。而对于较小子样来说, 我们可以运用自助方法(Bootstrap 方法), 产生自助