

博士学位论文

基于贝叶斯理论的靶场试验
综合设计方法研究

**RESEARCH ON INTEGRATED DESIGN
METHODS OF RANGE TEST BASED ON
BAYESIAN THEORY**

董光玲

哈尔滨工业大学
2015 年 12 月

国内图书分类号：TB114
国际图书分类号：623

学校代码：10213
密级：公开

工学博士学位论文

基于贝叶斯理论的靶场试验 综合设计方法研究

博士研究生：董光玲

导师：姚郁教授

申请学位：工学博士

学科：控制科学与工程

所在单位：航天学院

答辩日期：2015年12月

授予学位单位：哈尔滨工业大学

Classified Index: TB114

U.D.C: 623

Dissertation for the Doctoral Degree in Engineering

**RESEARCH ON INTEGRATED DESIGN
METHODS OF RANGE TEST BASED ON
BAYESIAN THEORY**

Candidate:	Dong Guangling
Supervisor:	Prof. Yao Yu
Academic Degree Applied for:	Doctor of Engineering
Speciality:	Control Science and Engineering
Affiliation:	School of Astronautics
Date of Defence:	December, 2015
Degree-Conferring-Institution:	Harbin Institute of Technology

摘 要

靶场试验在国防科研和装备建设中有着极其重要的作用，它是验证装备的设计思想、考核技战术指标、辅助装备采办决策的基础。基于经典统计学的传统试验鉴定模式需要大样本量的现场试验数据，消耗非常高，已不能适应装备发展的要求。因此，综合应用仿真和现场试验的一体化试验与鉴定（Integrated Test and Evaluation, IT&E）模式得到了广泛认可，已成为装备试验鉴定的发展趋势。在 IT&E 模式下，如何综合考虑仿真和现场试验的不同作用，进行合理的试验设计、样本量确定和融合评估是亟待解决的三个基本问题。

本文基于贝叶斯理论系统研究了 IT&E 模式下装备靶场试验中的上述三个基本问题，主要研究内容与成果概括如下：

首先，考虑靶场试验中大量仿真试验数据为现场试验提供先验信息的特点，针对仿真和现场试验数据可能存在的分布不一致性，提出应用幂先验对仿真试验数据进行折合加权。通过分析标准幂先验和规则化幂先验在后验推断中的优良性，说明其用于靶场试验设计与评估的可行性，为后续研究奠定基础。

其次，针对靶场试验的试验点设计问题，提出一种基于贝叶斯正则化的试验设计方法，解决了因子效应分析中不适定问题的稳健求解。通过回归模型对靶场试验设计问题进行了描述。考虑正则化在求解不适定问题中的稳健性，提出了贝叶斯正则化因子效应分析方法。给出考虑因子效应信息的靶场试验设计信息矩阵，提出了贝叶斯正则化试验设计信息矩阵构造方法，给出了兼顾可能因子效应项的附加点设计方案用于现场试验设计。通过火炮射表试验的示例验证了基于贝叶斯正则化的靶场试验设计方法的有效性。

再次，针对靶场试验的样本量确定问题，提出基于修正幂先验的贝叶斯样本量确定方法，解决应用标准幂先验时存在的可信度风险与评估精度之间的矛盾。通过分析验前样本量对兴趣参数先验方差的影响，定义了试验效应指标，并提出了基于试验效应等价的修正幂先验构造方法。在修正幂先验的框架下，根据贝叶斯样本量计算准则以及兴趣参数的预后验分布特性，给出了靶场试验样本量确定方法。应用导弹精度靶场试验样本量确定的示例验证了基于修正幂先验确定靶场试验样本量的有效性。

然后，针对靶场试验的融合评估问题，给出基于规则化幂先验的靶场试验融合评估方法，改进了靶场试验评估时的多源历史数据融合结构。通过广义线性混合模型对靶场试验的融合评估问题进行了描述，考虑信息交互影响和加权

约束,给出了多源信息规则化幂先验的构造方法。针对靶场试验样本量评估,提出了基于规则化幂先验的后验方差分析方法,并以试验点上的样本量评估说明其有效性。针对靶场试验设计评估,提出了基于规则化幂先验的因子效应分析方法,并以火炮精度试验示例验证了其有效性。

最后,总结前面对靶场试验综合设计中三个基本问题的研究成果,将其用于某型空地导弹靶场试验的设计与评估。通过分析评估试验的综合设计,验证了本文提出方法的有效性。

关键词: 贝叶斯理论; 幂先验; 试验与鉴定; 试验设计; 样本量确定; 融合评估

Abstract

Range test plays an important role in national defense research and development of weapon system, which constitutes the foundation for verifying equipment design philosophy, evaluating tactical and technical specifications, and aiding decision-making of equipment acquisition. Traditional test and evaluation (T&E) mode is on the basis of classical statistics, which needs large sample size of field test data, and consumes a huge number of test resources. So it no longer satisfies the requirements of equipment development. Therefore, integrated test and evaluation (IT&E) mode with combined use of simulation and field test is getting extensive recognition, which has become the future developmental trend of equipment T&E by now. Under IT&E mode, there are three major issues need to be addressed, namely rational experimental design with comprehensive consideration on different roles of simulation test and field test, sample size determination (SSD) as well as fusion and assessment of range test.

Based on Bayesian theory, this dissertation makes systematic investigation on the above three major issues of equipment range test under IT&E mode. The primary contributions are listed as follows:

Firstly, considering the features of range tests that mass of simulation data provides prior information for field test, power prior is put forward for weighting of simulation data with probable distribution heterogeneity to field test data, whose rationality in the design and assessment of range test is illustrated by analyzing optimality properties of both standard and normalized power priors in posterior inference. Thus, it forms the foundation for subsequent researches.

Secondly, Bayesian regularization based experimental design method is proposed for design of experiment (DOE) of range test, which solves the ill-posed problem in analysis of effects of factors. The DOE problem in range test is described through regression model. Regularization method is suggested to use in factor effects analysis considering its stability in solving ill-posed problem, which lead to the proposed Bayesian regularization based factor effects analysis method. As to experimental design with certain information about underlying effects of factors on response, an optimal design with additional test points for possible factor effect items is put forward. While for experimental design with no information about underlying effects of factors on response, a space-filling design is suggested. The effectiveness of Bayesian regularization based DOE method in range test is verified through illustrative use in gun firing table test.

Thirdly, a modified power prior based SSD method is put forward for sample

sizes calculation of simulation test and field test in range test, which overcomes the contradiction between credibility risk and assessment precision while using standard power prior. The impact of prior sample size on prior variance of parameter of interest is analyzed, which lead to the index of test effect. Hence, a modified power prior elicitation method is given out based on test effect equivalence. With modified power prior framework, optimization equations for SSD of range test scheme are given out according to Bayesian SSD criteria and pre-posterior distribution features for parameter of interest. The validity of the proposed Bayesian SSD method for range test is verified with application example.

Fourthly, normalized power prior based Bayesian fusion and assessment methods are given out for range test with multi-source information, which improves fusion architecture of multi-source historical data in range test assessment. Generalized linear mixed model is used to describe fusion and assessment problems in range test. Considering mutual effects and weight constraints of multi-source information, different structural styles of normalized power prior elicitation methods are suggested. As to sample size assessment of range test, a normalized power prior based posterior variance analysis method is given out, whose effectiveness is illustrated with sample size assessment on design point. While for DOE assessment of range test, a normalized power prior based method by analysis of effects of factors is put forward with verification by gun firing precision test.

Finally, the research achievements on three fundamental problems of range test under IT&E mode are summarized and used in the integrated design and assessment of some air-to-ground guided missile range test. The effectiveness of our proposed methods are demonstrated by analysis and assessment on integrated design of test.

Keywords: Bayesian theory, power prior, test and evaluation, design of experiment, sample size determination, fusion and assessment

目 录

摘 要	I
ABSTRACT	III
第 1 章 绪 论	1
1.1 课题背景及研究目的	1
1.1.1 课题背景	1
1.1.2 研究的目的和意义	2
1.2 国内外的研究现状及分析	3
1.2.1 试验设计方法	3
1.2.2 样本量确定方法	6
1.2.3 融合评估方法	9
1.2.4 存在的主要问题	12
1.3 论文主要研究内容	13
第 2 章 幂先验及贝叶斯理论基础	16
2.1 引言	16
2.2 贝叶斯理论基础	17
2.2.1 贝叶斯公式	17
2.2.2 先验信息	18
2.3 幂先验	19
2.3.1 幂先验的定义	20
2.3.2 回归模型的幂先验	21
2.3.3 幂先验的优良性分析	24
2.4 随机幂指数的规则化幂先验	27
2.4.1 规则化幂先验的定义	27
2.4.2 正态参数的规则化幂先验	29
2.4.3 规则化幂先验的优良性分析	30
2.5 本章小结	32
第 3 章 基于贝叶斯正则化的靶场试验设计方法	33
3.1 引言	33
3.2 靶场试验设计问题的描述	34
3.3 贝叶斯正则化因子效应分析	36

3.3.1 正则化方法	36
3.3.2 贝叶斯正则化因子效应分析	37
3.4 基于贝叶斯正则化的靶场试验设计方法	39
3.4.1 考虑因子效应信息的靶场试验设计信息矩阵	39
3.4.2 贝叶斯正则化试验设计信息矩阵构造	41
3.4.3 靶场试验设计方法	43
3.5 示例分析	44
3.6 本章小结	50
第4章 基于修正幂先验的靶场试验样本量确定方法	51
4.1 引言	51
4.2 靶场试验样本量确定问题的描述	51
4.3 基于试验效应等价的修正幂先验	53
4.3.1 考虑验前样本量影响的试验效应指标	53
4.3.2 基于试验效应等价的修正幂先验	56
4.4 基于修正幂先验的靶场试验样本量确定方法	59
4.4.1 贝叶斯样本量确定准则	59
4.4.2 修正幂先验的预后验分布	59
4.4.3 靶场试验样本量确定方法	61
4.5 示例分析	62
4.6 本章小结	66
第5章 基于规则化幂先验的靶场试验融合评估方法	67
5.1 引言	67
5.2 靶场试验融合评估问题的描述	68
5.3 多源信息的规则化幂先验构造	70
5.3.1 考虑信息交互影响的规则化幂先验	70
5.3.2 考虑信息加权约束的规则化幂先验	72
5.4 基于规则化幂先验的靶场试验样本量评估方法	74
5.4.1 样本量评估的后验方差分析模型	74
5.4.2 后验方差分析模型的规则化幂先验	75
5.4.3 靶场试验样本量评估方法	76
5.5 基于规则化幂先验的靶场试验设计评估方法	78
5.5.1 试验设计评估的因子效应分析模型	79
5.5.2 因子效应分析模型的规则化幂先验	80
5.5.3 靶场试验设计评估方法	81

5.6 本章小结.....	83
第6章 某型空地导弹靶场试验的设计与评估	84
6.1 引言.....	84
6.2 空地导弹靶场试验的需求分析	85
6.3 空地导弹靶场试验设计	87
6.3.1 仿真试验设计	87
6.3.2 飞行试验设计	89
6.4 空地导弹靶场试验的样本量确定	91
6.4.1 落点均值的修正幂先验构造	91
6.4.2 靶场试验的样本量确定	92
6.5 空地导弹的靶场试验评估	94
6.5.1 因子效应系数的规则化幂先验构造	94
6.5.2 靶场试验样本量评估	97
6.5.3 靶场试验设计评估	99
6.6 本章小结.....	101
结 论.....	102
参考文献	104
攻读博士学位期间发表的论文及其它成果	116
哈尔滨工业大学学位论文原创性声明及使用授权说明	118
致 谢.....	119
个人简历.....	120

Contents

Abstract (In Chinese)	I
Abstract (In English)	III
Chapter 1 Introduction	1
1.1 Background and objective of the subject.....	1
1.1.1 Background.....	1
1.1.2 Objective and significance of the subject	2
1.2 State of the art and analysis.....	3
1.2.1 Experimental design methods.....	3
1.2.2 Sample size determination methods.....	6
1.2.3 Test fusion and assessment methods.....	9
1.2.4 Main problems.....	12
1.3 Main research contents.....	13
Chapter 2 Power prior and Bayesian theoretical basis	16
2.1 Introduction.....	16
2.2 Bayesian theory.....	17
2.2.1 Bayes formula.....	17
2.2.2 Prior information.....	18
2.3 Power prior distributions.....	19
2.3.1 Definition of power prior.....	20
2.3.2 Power prior for regression models.....	21
2.3.3 Optimality analysis of the power prior.....	24
2.4 Normalized power prior for random power exponent.....	27
2.4.1 Normalized power prior approach.....	27
2.4.2 Power prior elicitation of normal distribution parameters.....	29
2.4.3 Optimality analysis of the normalized power prior.....	30
2.5 Summary.....	32
Chapter 3 Bayesian regularization based experimental design method for range test	33
3.1 Introduction.....	33
3.2 Description of experimental design in range test.....	34
3.3 Analysis on effects of factors with Bayesian regularization.....	36
3.3.1 Regularization method.....	36
3.3.2 Analysis on effects of factors with Bayesian regularization.....	37
3.4 Range test design method based on Bayesian regularization.....	39
3.4.1 Information matrix for DOE considering effects of factors.....	39
3.4.2 Bayesian regularization-based construction of information matrix.....	41

3.4.3 Experimental design method for range test.....	43
3.5 Illustrative example.....	44
3.6 Summary.....	50
Chapter 4 Modified power prior based sample size determination method for range test.....	51
4.1 Introduction.....	51
4.2 Description of sample size determination in range test.....	51
4.3 Modified power prior based on test effect equivalence.....	53
4.3.1 Test effect index considering impact of prior sample size.....	53
4.3.2 Modeified power prior based on test effect equivalence.....	56
4.4 SSD method of range test based on modified power prior.....	59
4.4.1 Bayesian sample size determination criteria.....	59
4.4.2 Pre-posterior distribution of modified power prior.....	59
4.4.3 Sample size determination method of range test.....	61
4.5 Illustrative example.....	62
4.6 Summary.....	66
Chapter 5 Normalized power prior based fusion and assessment method for range test.....	67
5.1 Introduction.....	67
5.2 Description of fusion and assessment in range test.....	68
5.3 Normalized power prior for multi-source information.....	70
5.3.1 Normalized power prior with consideration on mutual effects.....	70
5.3.2 Normalized power prior with consideration on weight constraints.....	72
5.4 Normalized power prior-based sample size assessment of range test.....	74
5.4.1 Posterior variance analysis model for sample size assessment.....	74
5.4.2 Normalized power prior of posterior variance analysis model.....	75
5.4.3 Sample size assessment method of range test.....	76
5.5 Normalized power prior-based DOE assessment of range test.....	78
5.5.1 Factors effects analysis model for DOE assessment.....	79
5.5.2 Normalized power prior of factors effects analysis model.....	80
5.5.3 DOE assessment method of range test.....	81
5.6 Summary.....	83
Chapter 6 Designs and assessment of some air-to-ground missile range test.....	84
6.1 Introduction.....	84
6.2 Requirment analysis of air-to-ground missile range test.....	85
6.3 Experimental design of missile range test.....	87
6.3.1 Experimental design of simulation test.....	87
6.3.2 Experimental design of flight test.....	89

6.4 Sample size determination of missile range test	91
6.4.1 Modified power prior elicitation of mean droppoints.....	91
6.4.2 Sample size determination of range test	92
6.5 Assessment of air-to-ground missile range test.....	94
6.5.1 Normalized power prior elicitation for coefficients of factors effects....	94
6.5.2 Sample size assessment of range test.....	97
6.5.3 DOE assessment of range test.....	99
6.6 Summary.....	101
Conclusions.....	102
References.....	104
Papers published in the period of Ph.D. education.....	116
Statement of copyright and Letter of authorization.....	118
Acknowledgements.....	119
Resume.....	120

第1章 绪论

1.1 课题背景及研究目的

1.1.1 课题背景

装备试验(Test)是为了获取与装备性能相关的数据而进行的实践活动^[1]。鉴定(Evaluation)是通过对各种试验数据进行有条理的收集、分析并与预期性能进行比较,以支持装备采办的决策过程^[1, 2]。评估(Assessment)则是根据试验数据确定装备性能如何在某些影响因素的组合下(如自然条件、目标特性、对抗措施等)变化^[1, 3],为装备的作战使用提供支持。可见,装备试验与鉴定是实现评估的基础。靶场试验是指在靶场进行的科研、鉴定、定型、批抽检、验收或作战使用试验,通常由装备的研制方提出申请,装备研制管理部门负责安排计划,国家靶场负责实施。靶场试验的任务是给出被试验装备的准确试验结果,做出正确的试验结论,为装备定型、部队使用、承研承制单位验证设计思想和检验生产工艺等提供科学的依据。高新技术的发展使新型武器装备的自动化、信息化、综合化和智能化水平越来越高,系统日益复杂,造价也越来越高。基于经典统计学的传统试验与鉴定模式大多采用外场试验方法,需要大量的试验样本,消耗非常高,已不能满足装备发展的要求。

为适应高技术武器装备的发展,美军在20世纪90年代提出了一体化试验与鉴定(Integrated Test and Evaluation, IT&E)模式^[4]。美国国防部2008年12月颁布的DoD I 5000.2 “Operation of the defense acquisition system”中明确要求开展一体化试验与鉴定^[5];国防采办大学出版的《Test and evaluation management guide》^[2]一书中特别强调了一体化试验,要求“项目主任应协同用户及试验鉴定部门,对研制试验与鉴定(DT&E)、作战试验与鉴定(OT&E)、实弹射击试验与鉴定(LFT&E)、系统族互操作性试验、信息保障试验、以及建模与仿真(M&S)等活动进行协调,形成一个有效的连续统一体,避免进行单一试验和重复性试验,充分利用试验资源,缩短研制时间,并将试验鉴定工作紧密地与需求制定和系统的设计开发结合起来”。由于在实际应用中表现出的巨大优势,IT&E模式已成为装备试验与鉴定的发展方向^[6-8]。IT&E模式下,靶场试验综合应用仿真与现场试验等手段对装备性能进行考核,如何确定仿真和现场试验的相互关系及其作用,给出两种试验的试验点选择和样本量确定的综合设计方案,并依据得到的靶场试验数据评估综合设计方案的有效性是国家靶场目前亟需开展的研究课题。

1.1.2 研究的目的是和意义

IT&E 的基本思想是通过对各种试验项目和试验方法、各类试验资源进行统一规划和管理,实现对试验信息的综合利用,加速装备研制进程,降低研制风险,提高装备试验的效率和效益,其三维概念^[9-11]可由图 1-1 表示。

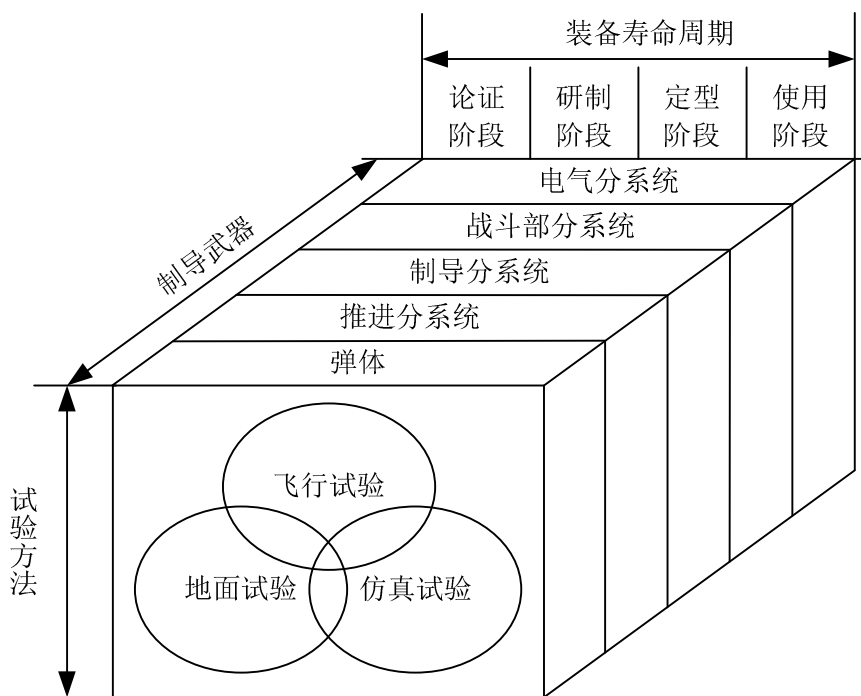


图 1-1 制导武器 IT&E 的三维概念图

Fig.1-1 Three dimensional concept map for IT&E of guided weapons

可见,IT&E 模式涵盖了靶场试验的过程、资源以及管理等多方面的内容。从狭义上说,IT&E 模式下的靶场试验综合应用仿真和现场飞行试验方法考核装备性能。美军的 IT&E 走在世界前列,已有许多文献对其相关政策^[2, 5, 12, 13]、实现思想^[7, 14-18]以及发展前景^[8, 19]进行了介绍。近年来,国内跟踪美军 IT&E 的最新发展,对我国装备采办和试验鉴定工作提出了一些对策和建议^[6, 20-23],并探索了 IT&E 在装备试验中的具体应用^[24-29]。

目前,针对 IT&E 模式下的靶场试验研究大多还是在顶层设计层面进行的探索性研究,而对于试验基础理论和综合设计方法仍缺少科学、深入的研究。基于贝叶斯理论的靶场试验综合设计方法研究的目的是构造 IT&E 模式下靶场试验综合设计的合理信息先验,为现场试验设计点数确定及优化选择、仿真和现场试验的样本量确定以及靶场试验综合设计的有效性评估提供科学合理的依据。研究的意义在于通过对相关问题的研究,构建靶场试验综合设计的理论和方法体系,为实现更加科学合理的装备试验鉴定提供支持。

1.2 国内外的研究现状及分析

IT&E 模式下的靶场试验综合设计与评估方法包括三个方面的研究内容,即试验设计方法、样本量确定方法和试验融合评估方法。下面介绍国内外相关技术的研究现状,并分析存在的主要问题。

1.2.1 试验设计方法

试验设计是以概率论和数理统计为基础,经济地、科学地安排试验的一项技术,其目的是确定合理的试验设计点,以产生良性数据并得出科学的结论。自从1935年英国著名统计学家 R. A. Fisher 的书《The Design of Experiments》出版以来,试验设计这门学科由于其应用的广泛性,获得迅速发展。根据应用对象及理论基础的不同,试验设计方法有多种分类,下面分别从经典试验设计、最优试验设计、计算机试验设计以及贝叶斯试验设计等四个方面介绍当前试验设计方法的研究现状,并分析其在靶场应用中存在的问题。

1.2.1.1 经典试验设计

经典试验设计^[30]以经典统计学为基础,研究如何合理地在因子空间中抽样,以尽可能少的试验次数来获取足够有效的资料,从而得到较可靠的结论。主要包括:析因设计、正交设计、筛选设计、响应面设计等。

析因设计^[31]是一种对多因子多水平试验进行全面交叉分组的设计方法,可用来分析全部因子效应,缺点是当因子或因子水平数过多时,所需的试验点数呈几何倍数增加。因此,当因子与水平数较多时,一般不采用析因设计,而是选择仅需因子空间部分实施的正交设计。

正交设计^[32]是田口玄一在20世纪40年代提出的一种具有里程碑意义的试验设计方法,与析因设计相比大大减少了试验点数,其布点具有“整齐可比,均匀分散”的特点,易于分析各因子的主效应。然而,其数据分析结果通常只能得到因子水平组合范围内的优选值,难以对后续试验提供明确指导。

筛选设计^[33](Plackett Burman design, PBD)是一种近似饱和的2水平试验设计方法,能够用最少的试验次数估计出因子的主效应,以从众多的考察因子中快速有效地筛选出最重要的几个因子。该设计的缺点是不能区分出因子主效应与交互效应的影响,无法给出最优因子组合方案。

响应面设计采用响应面方法描述因子与输出响应的关系,寻求用最少的试验点获得理想的响应面模型,主要用于优化因子组合以获得最好的输出^[34]。主要包括 Box-Behnken 设计(BBD)、均匀外壳设计(USD)和中心复合设计(CCD)三类。其缺点是因子水平太宽或关键因子不全时,响应面会出现吊兜

和鞍点，且响应面模型预测时仅限于因子所限的范围之内。

经典试验设计方法倾向于将试验点散布到因子空间边缘，只留下一小部分试验点在因子空间的中心，而且考虑的因子及水平数较多时，需要的试验点数急剧增加。由于因子空间边缘为边界条件，试验往往难以实施，且靶场试验的消耗极高，大多为小子样情形。因此，经典试验设计不能满足靶场试验用最少的试验点对全因子空间的装备性能进行科学评估的要求。

1.2.1.2 最优试验设计

最优试验设计是 Kiefer 在 1959 年最早提出的一类设计，也是试验设计中最重要的设计类型之一。它假定试验人员已知响应和因子间的关系可用某类回归模型表示，但模型中含有一些待估参数。最优回归设计主要研究的是如何选择试验设计点，使已知回归模型的系数能获得在一定准则下的最优估计，常用的最优准则有 D-最优，A-最优，E-最优等^[35]。

Silvey^[35]对比分析了经典试验设计与最优试验设计的优良性，指出当因子空间形状不规则，由经典试验设计方法得到的设计点数太多使得试验无法实现时，以及已知响应模型的阶次时，最优设计的优势明显。目前，最优试验设计在应用上的成功案例很多，如 Kuram 等^[36]应用 D 最优设计进行了端铣过程中植物切割液的效果评估，并实现了切割参数的优化设计。Chen 等^[37]应用 D 最优设计实现了金刚石刀具精密切割过程中的振动及表面不平度的建模与分析。刘雨等^[38]应用 D 最优试验设计解决了惯导平台系统自标定试验中的连续旋转路径设计问题。李光庆等^[39]应用最优回归设计建立越冬花椰菜氮磷钾平衡施肥效应和阶段施氮效应函数模型，为不同生产目标下的测土配方施肥提供了理论依据。José E. Bernardo 等^[40]应用中心复合设计和 D 最优设计的混合方法，对影响噪声与废气排放的多目标环境技术进行了动态编队级评估，识别出值得深入研究的关键环境技术。

对于模型已知的情形，最优试验设计不仅是一种饱和设计，而且大大减少了所需的试验点数，能够有效降低试验消耗。由于在靶场试验设计阶段掌握的装备性能信息有限，难以对性能响应模型有充分认识，无法确定合理的因子效应结构，应用最优设计存在信息矩阵构造难题。而且，当性能响应模型选择不正确时，构造的最优设计存在较大偏差，会对试验结果的有效性产生不良影响。可见，最优试验设计方法用于靶场试验设计时，存在合理确定模型结构的难题，根据验前信息确定合理的模型结构，是应用最优试验设计的关键。

1.2.1.3 计算机试验设计

20 世纪 90 年代以来，计算机试验获得飞速发展，J. Sacks 等^[41, 42]分析了计算机试验与物理试验的不同本质，指出传统物理试验包含随机误差，而计算

机试验具有确定性。针对计算机试验的这一特点,许多学者提出“空间填充”(Space Filling)的思想用于计算机试验设计。Simpson 等人^[43]证明为了得到最优的分析效果,计算机试验的选点设计应该是充满整个因子空间的。

空间填充设计方法大致可分为两类,一类为基于抽样的设计,另一类为基于准则的设计^[44]。前者对应频率学派的分析方法,后者对应贝叶斯学派的分析方法。基于抽样的设计包括:Halton 序列^[45]、Sobol 序列^[46]、格子法^[47]、拉丁超立方抽样(Latin Hypercube Sampling, LHS)^[48, 49]等。Gardner 等^[50]比较了 Monte Carlo 抽样和优化的 LHS 方法,发现在因素较多时,后者优势明显。基于准则的设计中,常用的设计准则包括:积分均方误差(IMSE)准则、最大熵(Maximum Entropy)准则、极小极大和极大极小(Minimax and Maximin)准则等。此外,方开泰教授指出,如果选均匀性作为最优拉丁方设计(Optimal Latin Hypercube Design)的准则,得到的就是均匀设计^[51]。上述方法中,LHS 和均匀设计是目前应用最广泛的计算机试验设计方法^[52],它们都是基于总体均值模型(Overall Mean Model)。LHS 是在因子空间中抽样选取一组具有相关性、同分布的试验点,具有随机性,而均匀设计是从确定性角度出发,用均匀设计表安排试验。目前,均匀设计已在军工科技、纺织、制药、化学等领域有了很多应用^[53-57],取得了较大的经济与社会效益。

空间填充设计方法对系统性能响应模型一般不做特定要求,非常适合于对潜在的因子效应项未知时的试验设计,而且需要的试验点数较少,目前已成为小子样试验设计的重要手段。空间填充设计用于靶场试验时存在的主要问题是难以确定合理的试验设计点数。设计点数过少会得到超饱和设计,即试验的次数少于欲估计的参数(主效应、交互效应、误差方差等)的数目,带来严重的评估风险;而设计点数过多又会造成试验资源浪费。

1.2.1.4 贝叶斯试验设计

贝叶斯试验设计利用验前信息和构造的效用函数,将试验设计转化为决策问题,并通过最大化一定准则下效用函数的后验期望实现。合理确定输出响应和因子间的模型结构是实现最优设计的关键,贝叶斯试验设计能够充分利用验前数据得到模型结构的先验信息^[58-60],为实现一定准则下的最优设计提供了基础。Chaloner 和 Verdinelli^[61]基于决策论将贝叶斯试验设计的相关准则归结为一种连续统一的框架,统一了线性和非线性试验设计问题,并应用效用函数给出了贝叶斯试验设计的方向。Sebastiani 等人^[62]提出了一种采用最大熵采样实现最优贝叶斯试验设计的方法,避免了在某些非线性和线性回归模型中需要采样点条件分布的难题。Huan 和 Marzouk^[63]提出一种基于仿真的最优贝叶斯试验设计方法,用于解决非线性系统的参数估计和预测问题。

在试验初始阶段,合理把握试验的设计点数是非常困难的,因此,近年来贝叶斯序贯自适应的抽样设计方法得到了广泛关注。如, Lin^[64]在其博士论文中提出了一种序贯探索的试验设计(Sequential Exploratory Experiment Design)方法; Sasena 等人^[65]利用贝叶斯方法自适应地鉴别出含有更多信息的样本点; Azadi 等^[66]针对二元响应数据提出一种贝叶斯序贯试验设计方法,并用于肌电图仪的试验中; Jin 等人^[67]对几种抽样方法进行了试验比较,显示贝叶斯序贯抽样方法能够使设计人员更好地控制采样过程; Drovandi 等^[68]针对广义非线性模型的离散数据提出一种可实现贝叶斯序贯设计的 Monte Carlo 算法。韩旭等^[69]提出一套基于正态贝叶斯序贯检验的试验设计方法,为多指标、长周期的小子样动态试验确定提供了理论依据。

贝叶斯试验设计利用验前信息得到响应模型的先验分布,对后续试验设计具有指导意义。然而,由于验前数据与真实数据往往存在分布不一致性,甚至其模型结构也存在一定差异。因此,应用贝叶斯试验设计方法面临验前信息的可信度加权融合问题,并且信息先验的构造也常常存在争议。如何由验前信息构造合理的效用函数及响应模型的信息先验是贝叶斯试验设计的难点。

在实际应用的推动下,试验设计理论和方法都取得了丰硕的成果。然而,对于 IT&E 模式下靶场试验中的仿真和现场试验综合设计,相关研究还很少,应用中大多还是采用单独设计的思想。尽管空间填充法和贝叶斯序贯最优准则为仿真试验设计提供了较好的解决方案,并且现场试验的小子样均匀设计也已发展的比较成熟^[70-72]。但这种简单融合了空间填充法和小子样均匀设计的靶场试验设计方案并不是效益最优的。如何应用贝叶斯理论分析验前数据,探索响应模型结构的先验信息,并将其用于指导后续现场试验的最优设计,使得到的现场试验设计点覆盖在需要重点考核的区域,是靶场试验设计中亟需解决的问题,也是本文对靶场试验设计研究的基本思路。

1.2.2 样本量确定方法

样本量确定(Sample size determination, SSD)是 IT&E 模式下的靶场试验技术研究中一项非常重要的内容,关系到试验消耗和评估质量,并直接影响到最终的装备鉴定结论。根据 Adcock^[73]的分类,样本量确定方法总体上可分为经典方法和贝叶斯方法两大类。其中,经典样本量确定方法源于大样本量下的经典统计学;而贝叶斯样本量确定方法由于利用验前数据构造先验信息,尤其适用于小子样试验的情形。

1.2.2.1 经典样本量确定方法

许多经典统计学教材^[74, 75]都对样本量确定问题进行了研究, Desu^[76]对经

典样本量确定方法进行了综述。总体而言，经典样本量确定方法还可以细分为两类：一类是基于参数估计精度要求的样本量确定方法，另一类是基于假设检验功效分析的样本量确定方法。

基于参数估计精度要求的样本量确定方法通过控制采样分布统计量的某些方面，得到需要的样本量大小，如在估计问题中，可能希望控制区间估计的长度或点估计的方差。此时，样本量一般由一个或多个未知参数确定，在实际应用中需要对未知的参数有初始猜测值。譬如，正态均值估计的样本量计算公式取决于方差值，而这个方差值必须用猜测值来代替。还有在著名的二项式比率区间估计的样本量选择问题中，样本量直接取决于兴趣参数的大小^[77]。对未知参数的初始猜测使得到的样本量往往只是局部最优解，并严重依赖于选取的设计值。

假设检验的统计功效分析^[78]（Statistical power analysis）提供了另一类重要的经典样本量确定方法。统计功效分析关注的四个量分别是：样本量 n 、显著性水平 α 、统计功效和效应量（Effect size, ES）。其中，统计功效是指某个检验在原假设 H_0 为假时，正确拒绝原假设的概率，用 $1-\beta$ 表示；效应量是指由于因素引起的差别，是衡量处理效应大小的指标，它不受样本容量影响，可表示不同处理下的总体均值之间差异的大小。功效分析中的四个量密切相关，给定其中任意三个，就可推算出第四个量，它们之间的关系如图 1-2 所示。基于功效分析确定样本量就是维持一个可接受的显著性水平，尽量使用较少的试验样本，实现最大化统计检验的功效。

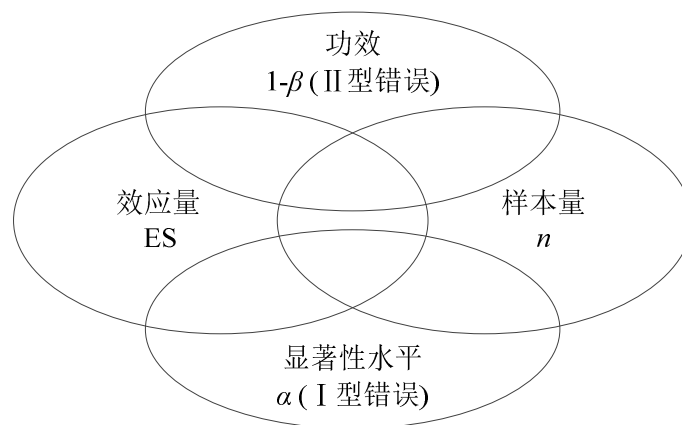


图 1-2 统计功效分析中的四个基本量

Fig.1-2 Four primary quantities in statistical power analysis

在实践中，统计功效分析大量用于医药卫生^[79]、生态环境^[80]、行为科学^[81]和试验鉴定^[82-84]等领域。国外的试验设计方案对统计功效的要求如同对显著性水平 α 一样，是不可缺少的。这是因为在统计推断中，只有同时满足对两类错误控制要求的样本量才是有效的。德国曼海姆大学的 Edgar Erdfelder 教授开

发的 G*Power 3^[85]是一款免费、通用且独立的统计功效分析软件,涵盖了 t 检验、 F 检验、 χ^2 检验族、 z 检验和其它精确检验的功效分析,为基于功效分析的样本量确定提供了一个很好的工具。

由于靶场试验综合应用仿真和现场试验对装备性能进行评估,样本量确定时不仅要给出两种试验的样本量,而且还要避免大样本量的仿真试验可能造成的数据“淹没”问题,以及现场试验样本过多带来的试验资源浪费。由前面的讨论可知,经典样本量确定方法难以确定仿真和现场试验样本的权值大小,无法解决仿真和现场试验的样本比例分配问题。

1.2.2.2 贝叶斯样本量确定方法

贝叶斯样本量确定方法用先验分布来模拟兴趣参数和多余参数的不确定性,避免了经典方法中的局部最优问题。贝叶斯推断是基于构造的后验分布进行的^[86],后验分布通过综合应用兴趣参数 θ 的验前信息(先验分布)和试验数据的似然函数得到,所以贝叶斯样本量确定是通过控制后验分布的某些概率特性来实现的^[87],如控制后验点估计和区间估计的精度。贝叶斯方法应用的数据概率分布为边缘(或称为先验预测)分布,即数据的采样分布与未知参数先验分布的混合。这样,得到的样本量就不再取决于对未知参数的假定值,而是由其先验分布决定。显然,应用这种先验信息有助于避免在试验设计阶段,对 θ 估计过于乐观而导致的样本量计算错误。

贝叶斯样本量确定的一个显著特点是需要应用两次先验分布:分别在获取兴趣参数 θ 的后验分布时和定义预后验计算的先验预测分布时。很多文献在设计阶段和后验推断阶段采用同一先验。然而,一些学者,如 Joseph 和 Belisle^[88]、Spiegelhalter 等^[89]指出在一些情况下,即使对未知的兴趣参数 θ 有充分的先验信息,这一先验也不能用于最终对 θ 的后验推断,例如医学研究的监管机构通常对此有明确要求。目前,已有多篇文献^[90, 91]提出应用这种混合方法进行样本量确定。Wang 和 Gelfand^[92]提出了一种更通用的两先验方法进行样本量确定,这种方法的最终后验分析未必基于无信息先验。De Santis^[93]应用以前类似试验的历史数据进行样本量确定,在设计阶段基于幂先验(Power prior)对历史数据折合加权,构造样本量确定的设计先验,进而根据样本量确定准则计算当前试验所需的样本量。《The Statistician》针对贝叶斯样本量确定主题还推出了两期专刊,见 volume 44, part 2 (1995)和 volume 46, part 2 (1997)。

近年来,贝叶斯样本量确定方法在国外受到了广泛关注。Adcock^[73]对样本量确定方法进行了详细回顾,Joseph 和 Belisle^[88]针对正态均值及均值差异估计的贝叶斯样本量确定方法进行了研究,并给出了求解方法。Schlaifer 和 Raiffa^[94]在基于贝叶斯决策理论的样本量确定方法研究方面做了开创性的工

作。Bernardo^[95]将统计推断也看作一个决策问题，给出了样本量选择方案。Lindley^[96]和 Walker^[97]分别基于决策的效用函数进行了样本量确定。文献[98-101]介绍了贝叶斯样本量确定方法的近期研究成果。

国内针对贝叶斯样本量确定方法的研究并不多，张湘平^[102]和张硕云等^[103]基于贝叶斯决策理论研究了小子样统计推断问题，讨论了样本量、先验信息和贝叶斯决策风险之间的关系，但该方法不适合参数估计的样本量确定。王正明等^[104]指出样本量与试验目的及可利用的先验信息相关，也与试验设计和评估方法紧密联系，介绍了应用贝叶斯决策理论和序贯分析、考虑信息可信度以及基于信息增量的样本量确定方法。然而，由于没有考虑验前样本量大小的影响，且对验前样本的权值确定具有随意性，在实际应用中常常引起争议。

贝叶斯样本量确定方法利用验前信息构造样本量确定的设计先验，能够综合考虑仿真与现场试验的样本量。尤其是 Ibrahim 等^[105]提出的幂先验不仅综合考虑了历史数据和当前数据在后验推断中的作用，而且实现了对历史数据的折合加权，为靶场试验的样本量确定提供了一种较好的解决思路。然而，当验前数据可信度很低而样本量非常大的时候，会带来可信度与评估风险之间的矛盾。如何综合考虑可信度及验前样本量大小的影响，构造贝叶斯样本量确定的合理设计先验，是靶场试验样本量确定需要解决的关键问题。

1.2.3 融合评估方法

试验融合评估是通过合适的方法把靶场试验的多源信息结合起来，对试验样本量的充分性和试验设计点的合理性进行综合评价，确保靶场试验综合设计方案对装备性能评估的费效比最优。目前，尚无研究专门针对试验融合评估，相关研究大多是针对系统性能进行的多源数据融合评估，可分为经典融合评估方法和贝叶斯融合评估方法两类。

1.2.3.1 经典融合评估方法

早期的装备试验鉴定过程中，大多采用经典统计理论进行武器性能参数的点估计、区间估计和假设检验。这些方法完全基于试验数据进行分析，具有严格的结果评价标准，对常规武器和普通弹药等试验耗费较低的装备而言，各种来源的靶场试验数据样本量较大，目前仍是主要的评估方法^[106]。

对于信息化、智能化水平较高的武器装备，由于试验成本昂贵、样本量非常有限，常常采用“试试看看、看看试试”的序贯概率比检验（SPRT: Sequential Probabilty Ratio Test）方案，以减少现场试验样本量。此时，性能参数的融合评估大多为检验型，若验前数据的样本量很大，而现场试验样本量很小，经典融合评估方法难以合理确定信息权值^[107]，结果可能存在争议。

1.2.3.2 贝叶斯融合评估方法

贝叶斯方法能够合理应用验前信息,是多源数据融合评估的首选方法^[104],其在实际应用中的三个要素概括为:信息先验的构造、参数化建模和多源数据的融合算法设计^[108]。

(1) 信息先验的构造

兴趣参数 θ 取离散值时,可根据专家经验对参数空间 Θ 中的每个点给出主观概率; θ 连续时,可采用直方图法、相对似然函数法和选定密度函数形式的方法确定其先验^[109, 110]。常用的信息先验构造方法还包括:共轭分布法^[94]、最大熵方法^[109]、Bootstrap 和随机加权法。其中,共轭分布法适用于兴趣参数的先验分布和后验分布属于同一类型的情况。最大熵法是在有一部分验前信息可以利用,而先验分布的形式未知时,希望能找到使熵最大的分布。Bootstrap 方法和随机加权法对兴趣参数的估计与给定的观测样本密切相关,能够在分布形式未知的情况下得到分布参数的估计,在小子样问题研究中受到关注^[111],其缺点在于严重依赖初始采样的结果。

幂先验是信息先验构造中一种非常重要的形式。Ibrahim 和 Chen^[105, 112]基于历史数据可用性的思想提出了幂先验的概念,其基本思路是用幂指数 δ ($0 \leq \delta \leq 1$) 来控制历史数据对当前研究的影响。Diaconis 和 Ylvisaker^[113]以及 Morris^[114]的研究最早体现了幂先验的基本思想,但是他们仅仅考虑了幂指数 δ 为预先确定的常数时的情形。Ibrahim 和 Chen 发展了这一思想,并广泛研究了幂先验的理论特性^[105, 112, 115, 116]。Duan^[117]进一步研究了幂指数服从随机分布时的情形,提出了规则化幂先验,解决了标准幂先验框架下,幂指数后验分布始终趋于 0 的问题。

(2) 参数化建模

参数化建模的目的是找到目标测量量与替代测量量之间的函数关系^[108],将不同试验因子条件下的替代试验数据折合到目标测量条件下,然后在目标测量条件下对系统性能数据进行融合。显然,参数化建模本质上是一种评估建模过程^[118],即得到能够反应输出响应和试验因子项之间的函数关系。

常用的参数化建模方法包括:线性模型法、非线性模型法、分离可交换变量模型法和序化模型法^[119]。线性模型法应用线性模型和贝叶斯方法对不同试验条件下的多源信息进行拟合^[120],该方法的最大问题在于模型选择不当会引起较大的建模误差和递推估计累积误差。非线性模型法^[121]利用非线性模型和极大似然、最小二乘或贝叶斯方法拟合不同试验条件下的多源信息。分离可交换变量模型^[122]通过分离出不同试验条件下的可交换参数实现信息融合。序化模型法^[123]利用不同试验条件下分布参数的序化关系和贝叶斯方法融合不同阶

段的多源信息,已成功应用于可靠性增长建模^[124]和精度评估^[108]等领域。

参数化建模的模型选择主要有两种思路,一种思路是应用信息准则选择模型,如基于条件 Akaike 信息准则(AIC)^[125]、广义信息准则^[126],以及围栏方法(Fence method)^[127]等。还有一种思路是基于正则化方法,如应用 SCAD 型惩罚似然方法^[128]、 L_1 型惩罚似然方法^[129-131]、两阶段惩罚拟似然方法^[132]单独选择固定效应或随机效应,应用惩罚型联合对数似然方法^[133]、极大惩罚似然估计方法^[134]、基于修正 Cholesky 分解的分层似然双惩罚方法^[135]同时选择固定效应和随机效应。

(3) 多源试验数据的融合算法

靶场试验的多源数据中,现场试验样本量一般远小于验前信息的样本量,引起人们对验前信息“淹没”现场试验信息,使现场试验信息不起作用的担忧。尤其是当验前信息与现场试验信息存在明显的分布差异时,将会使融合评估的结果出现较大的偏差^[136]。可信度描述了用户对应用验前数据进行分析、设计和评估时的信心。因此,目前常用的融合算法大多是先对验前数据的可信度进行评估,然后基于可信度加权得到系统性能的贝叶斯融合评估。

验前信息可信度评价大多是从数据的分布差异出发,如应用数据相容性检验^[136, 137]、信息散度^[138]、模糊算子理论^[139]等,由于现场试验样本量很少,基于数据分布差异的可信度评价结果具有很大不确定性^[140]。此外,针对仿真数据的信息权值确定问题,文献[141, 142]还提出了限制仿真样本量的方法,尽管这种方法能够避免数据“淹没”,却带来了如何选取仿真样本的问题,并且削弱了仿真试验的样本量优势。闫志强^[106]提出了一种改进的混合验后融合方法,使用有验前样本量约束的现场样本边缘分布替代无信息先验下的现场样本边缘分布,来避免后验权重和后验估计错误。

国外对于大样本量验前数据产生的“淹没”问题,同样采用加权的思想,其中,基于幂先验的贝叶斯融合与评估方法被普遍采用^[143-146]。该方法应用幂指数对验前异源数据进行折合加权,并将幂指数也视为一个随机变量,通过构造幂指数和兴趣参数的联合概率分布,对现场数据进行后验推断。目前,基于幂先验的多源数据融合方法在医药卫生、环境监测、道路安全、性能评估等许多领域得到了广泛应用^[147]。

由以上分析可见,针对系统性能开展的融合评估方法研究无法直接用于靶场试验评估。由于样本量主要与评估精度相关,而试验设计主要与设计模型包含的因子效应项相关。因此,如何在多源数据融合方法的基础上,应用靶场试验的多源数据分析兴趣参数评估精度及试验因子效应信息,以评价靶场试验样本量的充分性和试验点设计的合理性,是靶场试验融合评估的研究重点。

1.2.4 存在的主要问题

综上所述,对于单一试验设计、样本量确定及融合评估的研究已基本成熟。空间填充方法和贝叶斯序贯准则为仿真试验设计提供了较好的解决方案,现场试验的小子样均匀设计得到了广泛应用。根据评估精度的要求和统计功效分析确定试验样本量发展的较为成熟。而且,贝叶斯方法也广泛用于系统性能的多源数据融合评估中。然而,将上述研究成果推广到 IT&E 模式下的靶场试验综合设计与评估时,存在以下三个方面的问题:

(1) 现场试验点的设计问题

高价值的尖端武器装备使试验成本成指数级增长,为降低试验消耗,目前常用的方法就是减少试验用弹量,这种做法带来样本量与试验结果置信度相互制约的矛盾。经典试验设计、空间填充设计和贝叶斯试验设计等方法均难以用于实装实弹试验设计,无法解决小样本与高置信度的现实需求。如何利用贝叶斯理论、其它领域的成功技术,创建一种科学合理、先进实用的武器装备试验设计方法,通过对验前数据进行因子效应分析,确定合理的试验点数,构造靶场试验优化设计信息矩阵,实现现场试验最优设计,提高国家靶场试验效益,是本文需要解决的首要关键技术难题。

(2) 仿真和现场试验的样本量确定问题

目前,国家靶场提高试验鉴定样本量的常用方法是引入内场仿真试验技术,应用结果表明了该方法的科学性与有效性。但随之而来出现了三个新问题:一是如何合理给出仿真和现场试验样本的权值;二是如何有效解决仿真和现场试验的样本比例分配问题;三是能否正确解决仿真数据可信度低而样本量非常大的困境。前两个问题无法由经典样本量确定方法解决,第三个问题也对贝叶斯方法的设计先验构造提出了新的挑战,需要解决仿真可信度与评估风险之间的矛盾。因此,如何综合考虑仿真可信度与样本量的影响,构造靶场试验贝叶斯样本量确定的设计先验,是样本量确定方法研究的重点问题。

(3) 靶场试验的融合评估问题

靶场试验的有效性是试验鉴定管理部门关注的重点,关系到鉴定结论的科学性和合理性。目前的融合评估方法大多是针对系统性能进行的多源数据融合评估,无法直接用于靶场试验评估,不能给出靶场试验有效性的科学评价。靶场试验的融合评估包括对样本量的评估和对试验设计的评估,分别对应兴趣参数评估模型的预测精度是否满足要求,以及试验设计阶段考虑的因子效应信息是否完备。因此,如何在当前融合评估方法的基础上,研究应用多源验前数据构造信息先验,分析兴趣参数的预测后验方差和后验评估模型的因子效应,是解决靶场试验设计评估问题的关键。

1.3 论文主要研究内容

IT&E 模式下靶场试验综合设计与评估可分为三个阶段，如图 1-3 所示。首先，在确定试验影响因子及重点考核区域的基础上，进行靶场试验设计，得到仿真和现场试验点设计方案；其次，根据试验成本或评估精度要求，确定仿真和现场试验样本量；最后，根据获得的多源数据评估靶场试验样本量的充分性和试验点设计的有效性，完成靶场试验评估。

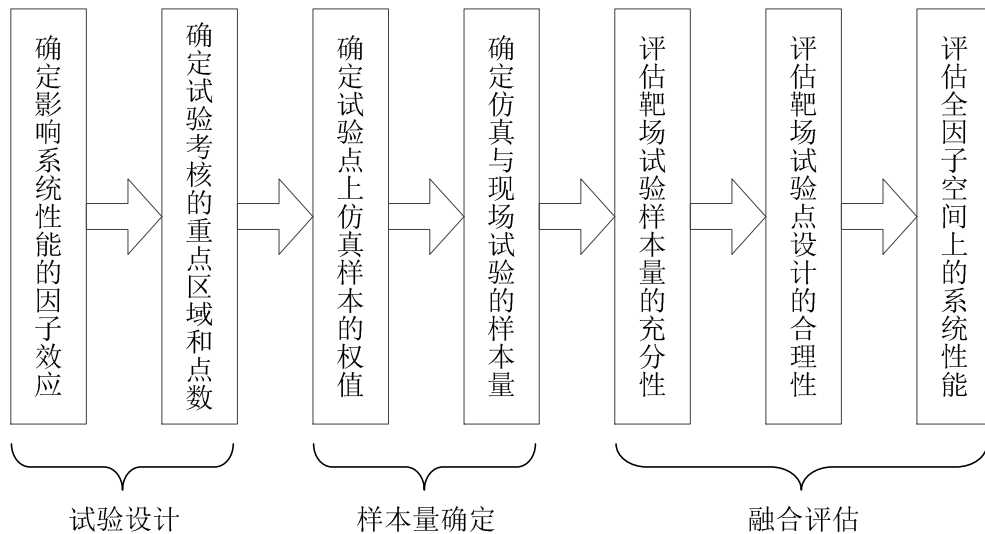


图 1-3 靶场试验综合设计与评估的过程

Fig.1-3 Process for integrated design and assessment of range test

根据以上分析，本文的主要研究内容包括 IT&E 模式下的靶场试验设计方法、靶场试验样本量确定方法和靶场试验融合评估方法，具体内容如下：

（1）靶场试验设计方法

研究的目的是解决现场试验的最优设计问题，包括试验点数确定以及试验点的优化选择。为确保试验设计的饱和性，避免参数估计风险，试验点数应大于或等于因子效应项数。贝叶斯正则化能够较稳健地实现反问题求解，为因子效应分析提供了一种良好的途径。如何将其用于分析仿真试验数据，确定影响系统性能的因子效应，并根据得到的因子效应信息构造靶场试验设计信息矩阵，实现一定准则下的最优设计，是靶场试验设计方法研究的主要内容。在此基础上，可以确定试验考核的重点区域和点数。

（2）靶场试验样本量确定方法

研究的目的是确定仿真和现场试验的样本量，给出样本量比例分配的依据。由于经典方法无法处理仿真和现场试验的样本量分配问题，而基于标准幂先验的贝叶斯样本量确定方法存在仅由大量仿真样本即可满足评估精度要求的问题。因此，综合考虑仿真可信度与样本量的影响构造合理的兴趣参数修正幂先

验成为计算仿真和现场试验样本量的关键,不仅为确定试验点上仿真样本的权值提供了科学依据,而且为准则约束下的仿真与现场试验样本量优化计算提供了目标函数,是靶场试验样本量确定方法的研究重点。

(3) 靶场试验融合评估方法

研究的目的是根据靶场试验的多源数据融合对靶场试验样本量的充分性和试验点设计的合理性进行综合评价。首先,考虑多源信息的交互影响和加权约束构造规则化幂先验,避免小子样现场试验数据的“淹没”问题。然后,由兴趣参数的预测后验方差计算式和回归系数向量的后验估计建立样本量评估的后验方差分析模型,为样本量评估提供合理依据。最后,由规则化幂先验下的模型先验及后验概率计算公式建立因子效应分析模型,为试验设计评估提供支持。由此解决靶场试验融合评估的难题。

针对以上研究内容,规划的论文研究结构如图 1-4 所示,各章节安排如下:

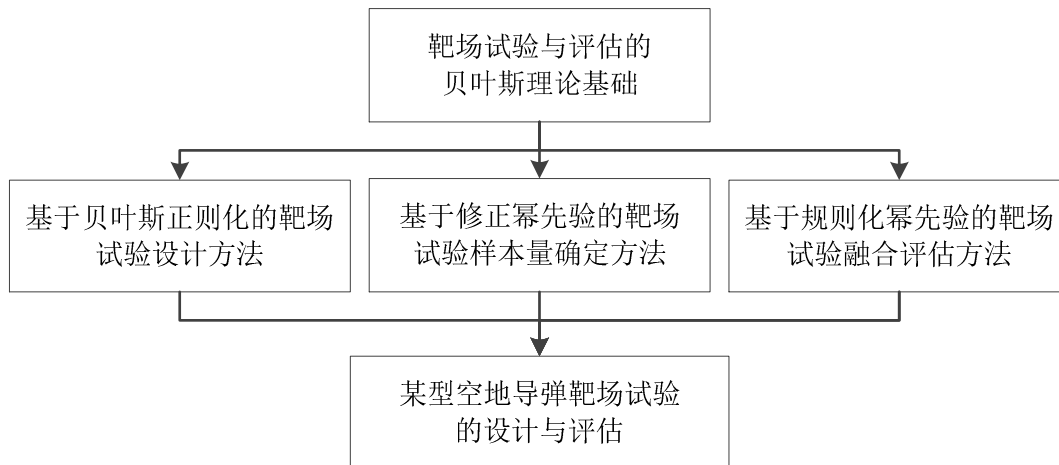


图 1-4 论文研究结构图

Fig.1-4 Research structure of the dissertation

第 1 章介绍课题的研究背景及研究的目的和意义。综述国内外在靶场试验综合设计与评估领域的主要研究成果,分别从试验设计方法、样本量确定方法、融合评估方法和存在的主要问题四个方面,阐述相关研究的现状并指出存在的主要问题和不足,引出本文研究的主要内容。

第 2 章介绍幂先验及贝叶斯理论基础。首先,回顾贝叶斯理论基础。其次,针对具有权值信息的历史数据融合问题,介绍幂先验的构造框架及其优良性。最后,分析幂指数为随机变量时的特殊性,引入规则化幂先验,并分析其优良性。本章是后续章节研究的基础。

第 3 章针对靶场试验的设计问题,提出一种基于贝叶斯正则化的靶场试验设计方法。首先,基于回归模型对靶场试验设计问题进行描述,并指出因子效应分析的重要性。其次,分析正则化方法在解不适定问题中的稳健性,提出贝

叶斯正则化因子效应分析方法。最后,给出考虑因子效应信息的靶场试验设计信息矩阵,提出贝叶斯正则化试验设计信息矩阵构造方法,给出一种兼顾可能因子效应项的备选附加点 D-最优设计用于靶场试验点设计。

第4章针对靶场试验的样本量确定问题,基于修正幂先验提出一种贝叶斯样本量确定方法。首先,对靶场试验的样本量确定问题进行描述,分析设计先验对兴趣参数后验分布的影响,说明仿真和现场试验样本量在兴趣参数后验估计中的作用。其次,分析验前样本量的影响,定义试验效应指标,给出基于试验效应等价的修正幂先验构造方法。最后,基于修正幂先验,给出准则约束下的靶场试验贝叶斯样本量确定方法。

第5章针对靶场试验的融合评估问题,基于多源信息规则化幂先验给出靶场试验的融合评估方法。首先,应用广义线性混合模型对靶场试验的融合评估问题进行描述,将其归纳为多源信息融合、试验样本量评估和试验设计评估三个基本问题。其次,考虑信息交互影响和加权约束,给出多源信息规则化幂先验构造方法。然后,针对靶场试验样本量评估,提出基于规则化幂先验的后验方差分析方法。最后,针对靶场试验设计评估,提出基于规则化幂先验的因子效应分析方法。

第6章以某型空地导弹靶场试验为背景,给出本文研究方法的应用实例。首先,对导弹靶场试验的需求进行分析。其次,应用空间填充法实现仿真试验设计,并根据仿真数据的贝叶斯正则化因子效应分析结果实现飞行试验设计。然后,根据仿真数据构造修正幂先验,实现效益最优的靶场试验样本量确定。最后,应用规则化幂先验对导弹靶场试验的样本量和试验设计进行评估,验证样本量的充分性和试验设计的合理性。

第 2 章 幂先验及贝叶斯理论基础

2.1 引言

统计推断中使用的三种信息分别为：总体信息、样本信息和先验信息。总体信息反应了总体的分布形式；样本信息是从总体中抽取的样本所提供的信息；先验信息是在抽样之前，有关统计推断问题中未知参数的一些信息，一般来自经验和历史资料。贝叶斯理论提供了一种基于上述三种信息计算兴趣参数后验概率的方法。显然，构造兴趣参数的先验分布是应用贝叶斯理论的关键。近几十年来，贝叶斯理论在试验鉴定领域得到了广泛应用，尤其是对于小子样的靶场试验综合设计与评估问题，贝叶斯方法几乎成为必然的选择。在贝叶斯方法的应用中，争议最多的是先验分布的构造，以及如何确定验前样本的信息权值。目前提出的验前样本可信度加权、限制仿真试验样本量以及信息散度加权等方法，在现场试验为小子样时得到的后验推断结果存在稳定性不足的缺点，在实际应用中仍未得到广泛认可。

靶场试验的综合设计与评估中，有三个基本问题需要解决：一是如何进行系统试验设计，确定试验应在哪些因子空间点上进行；二是怎样确定各个试验点上的样本量大小，使设计的试验方案既满足评估精度要求，又能够最大限度的节约试验成本；三是如何根据不同来源试验数据的融合结果，分析评估试验设计的有效性。应用贝叶斯理论解决上述问题的关键在于如何根据仿真试验和其他来源的验前信息，为后续现场试验构造一种适当的兴趣参数先验分布。幂先验最早由 Ibrahim 和 Chen^[105, 112]提出，提供了一种应用幂指数对验前数据进行折合加权的先验构造方法，在用于系统性能的后验推断时，现场试验数据能够同时更新兴趣参数和幂指数的后验信息，非常适合于靶场试验中验前数据和现场试验数据分布不完全一致的情形。

本章系统介绍幂先验及贝叶斯理论基础。结构安排如下：首先，介绍贝叶斯理论基础，指出在贝叶斯试验分析中构造信息先验的难点在于如何将主观信息和历史信息融入先验分布。其次，介绍标准幂先验的构造框架，给出参数估计和回归模型的幂先验构造方法，并通过兴趣参数条件后验分布的 KL 散度凸和以及 Zellner 的 IPR 效率准则，分析标准幂先验的优良性。最后，考虑幂指数为随机变量时的情形，分析标准幂先验方法存在的问题，引入随机幂指数的规则化幂先验，以正态分布为例说明规则化幂先验的构造方法，并进一步通过互信息函数分析规则化幂先验在信息传递上的优良性。

2.2 贝叶斯理论基础

2.2.1 贝叶斯公式

贝叶斯理论由英国学者贝叶斯 (T. Bayes 1701~1761) 提出的贝叶斯公式发展而来。贝叶斯公式在形式上是对条件概率的定义和全概率公式的一个简单的归纳和推理。设 A 为样本空间 Ω 中的一个事件, B_1, \dots, B_n (n 为有限的或无穷) 是样本空间 Ω 中的一个完备事件群, 且有 $P(B_i) > 0$ ($i=1, \dots, n$), $P(A) > 0$, 则事件 A 的全概率公式可写为

$$P(A) = P\left(\sum_{i=1}^n AB_i\right) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

在全概率公式的条件下, 按照条件概率的计算方法, 贝叶斯公式可表示为

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)} \quad (2-1)$$

式(2-1)给出的贝叶斯公式在实际统计推断中的应用价值在于: 当把事件 A 看作“结果”, 而把诸事件 B_1, \dots, B_n 看成导致这一结果的可能“原因”, 贝叶斯公式能够实现由“结果”推测“原因”。

统计推断中, 兴趣参数 θ 的先验分布 $\pi(\theta)$ 是指在获取样本 Y 之前对 θ 可能取值的认识。获得样本 y 后, 由于样本 y 中也包含了 θ 的信息, 因此, 对 θ 的认识也发生了变化和调整, 得到 θ 的后验分布。可见, θ 的后验分布是在给定 $Y=y$ 条件下 θ 的条件分布, 记为 $\pi(\theta|y)$, 对有密度函数的情形, 可表示为

$$\pi(\theta|y) = \frac{h(y, \theta)}{m(y)} = \frac{f(y|\theta)\pi(\theta)}{\int_{\Theta} f(y|\theta)\pi(\theta)d\theta} \quad (2-2)$$

其中, $f(y|\theta)$ 是样本 y 的条件概率密度, $h(y, \theta) = f(y|\theta)\pi(\theta)$ 为 y 和 θ 的联合分布, Θ 为参数 θ 的取值空间, 而

$$m(y) = \int_{\Theta} h(y, \theta)d\theta = \int_{\Theta} f(y|\theta)\pi(\theta)d\theta$$

称为 Y 的边缘分布。

式(2-2)是贝叶斯公式的密度函数形式, 集中了总体、样本和先验三种信息中有关 θ 的一切信息。根据贝叶斯统计学的观点, 获得后验分布 $\pi(\theta|y)$ 后, 一切统计推断 (如点估计、区间估计以及假设检验等) 都必须从 $\pi(\theta|y)$ 出发。由式(2-1)和(2-2)可知, 应用贝叶斯理论进行统计分析的关键在于构造合适的先验信息。贝叶斯统计学与经典统计学的主要区别就在于: 在考虑统计推断问题时, 除了利用抽样信息外, 还利用了参数的先验信息。由此可见, 先验信息的构造是贝叶斯统计学在实际应用中需要解决的首要问题。

2.2.2 先验信息

先验信息是指获得试验样本之前的经验和历史资料。在贝叶斯统计中通常用到的先验信息包括无信息先验 (Noninformative prior)、共轭先验 (Conjugate prior) 和多层先验 (Hierarchical prior)。

由于历史和经验信息往往难以获取,或者在实际问题中难以将这类信息融入先验信息。此时,通常的选择是应用无信息先验。无信息先验是对参数空间 Θ 上的任何一点 θ 都没有偏爱的先验^[148],如均匀先验 (Uniform prior)、Jeffreys 先验、参考先验 (Reference prior)、概率匹配先验 (Probability matching priors)、最大数据信息先验 (The Maximal Data Information Prior, MDIP) 等。鉴于无信息先验表现出的客观性,目前已越来越受到经典统计学派的认可。尽管无信息先验应用方便且更易于确定其形式,然而它们也不能适用于所有的应用背景,例如,在计算贝叶斯系数时需要有适当的 (Proper) 先验,而无信息先验往往并不是适当的。而且, Kass 和 Wasserman^[149]指出一些无信息先验 (如 Jeffreys 先验、参考先验、概率匹配先验) 是根据渐近参数确定的,对于小样本问题下的统计分析,信赖任何默认的设定先验都是危险的,可能会带来后验估计的偏倚。Berger^[150]明确指出,对于兴趣参数有历史信息或专家经验信息可以利用时,应采用信息先验。

贝叶斯计算技术获得突破之前,一般期望得到一个可以求解的后验分布。共轭先验的后验分布与先验分布属于同一个分布族,成为当时应用最广泛的信息先验。由于共轭先验可通过仅选择一两个参数来构造,非常方便灵活,目前依然得到广泛应用。然而,共轭先验有时也不准确,为了克服这一问题,可进行两方面的修正:一是构造多个共轭先验的混合形式;二是采用多维参数模型的条件共轭先验。

计算技术的发展使得应用具有近似分布形式的非共轭先验概括所有已知信息成为可能。当所给先验分布中的超参数难以确定时,可以对超参数再给出一个先验,第二个先验称为超先验;若超先验中的超参数还是难以确定,还可以再给出第三个先验……,这样由先验和超先验决定的一个新先验就称为多层先验^[151],其基本思想如下:

(1) 设 $\Gamma_1 = \{\pi_1(\theta|\lambda): \lambda \in A\}$, 其中, π_1 的函数形式已知, A 为超参数 λ 的取值范围,且 λ 未知。

(2) 设 λ 为随机变量, $\Gamma_2 = \{\pi_2(\lambda|\tau): \tau \in T\}$, 其中, π_2 的函数形式已知, T 为 τ 的取值范围。

(3) 设 τ 也是随机变量,具有先验分布 $\pi_3(\tau)$ 。

任何一个多层先验都可以写成一个一般规范的先验,以二层先验为例,该

规范的先验可表示为

$$\pi(\theta) = \int_A \pi_1(\theta|\lambda)\pi_2(\lambda) d\lambda = \int_A \pi(\theta, \lambda) d\lambda \quad (2-3)$$

式(2-3)中, $\theta \sim \pi_1(\theta|\lambda)$, $\lambda \sim \pi_2(\lambda)$, θ 和 λ 的联合密度为 $\pi(\theta, \lambda) = \pi_1(\theta|\lambda)\pi_2(\lambda)$, 故 $\pi(\theta)$ 作为联合密度 $\pi(\theta, \lambda)$ 的边缘密度就是的 θ 先验分布。对于三层的先验, 该规范的先验可表示为

$$\pi(\theta) = \int_A \int_T \pi_1(\theta|\lambda)\pi_2(\lambda|\tau)\pi_3(\tau) d\tau = \int_A \pi_1(\theta|\lambda) \left[\int_T \pi(\lambda, \tau) d\tau \right] d\lambda \quad (2-4)$$

其中, $\pi(\lambda, \tau) = \pi_2(\lambda|\tau)\pi_3(\tau)$ 。对于更多层的先验, 可由类似的方法求得 $\pi(\theta)$ 。

在实际问题中, 先验信息构建的难点在于如何在先验分布中融合个人主观信息和历史信息。第一个问题是如何将主观信息转换为先验分布。如果参数空间 Θ 为离散的, 这一问题就是简单的确定 Θ 中每个元素的主观概率。而当 Θ 为连续时, 可以应用如下几种技术: 直方图法、相对似然法、给定函数形式匹配法以及累积分布函数 (CDF) 主观构造法。Berger^[150]详细介绍了这些技术并指出, 未经训练的人员在构造先验信息时的表现并不令人满意, 对自身先验知识的过于自信往往导致构造的先验分布太过集中。

第二个问题是如何在先验分布中包含历史数据。一种很自然的想法是对历史数据应用无信息先验进行贝叶斯分析, 然后将得到的后验分布作为当前数据的信息先验。这种方法应用历史数据来更新先验信息, 由于给定的历史数据与当前数据的权值相等, 最终得到的后验分布仅仅意味着两种数据集的简单融合。然而, 尽管通常假定当前数据与历史数据服从同一分布族, 分布参数也可能随时间或配置情况而变化。如果历史数据的样本量远远大于当前数据的样本量, 且两类数据存在分布不均匀性, 这时历史数据将主导统计分析的结果, 数据融合就可能会产生误导性的结论。

2.3 幂先验

应用历史数据构造先验信息时, 如何确定其在当前研究中所占的权重比例是一个难题。由于试验成本的约束, 靶场试验能够获得的现场试验样本往往很少, 仅以这些少量的现场试验样本进行评估会导致评估结论中存在无法接受的不确定性。按照国家军用标准的要求, 装备性能评估应有足够的精度, 贝叶斯统计学利用主观信息和历史试验信息构造靶场试验综合设计与评估的先验, 有利于提高评估的精度, 已成为靶场试验与评估中不可或缺的理论基础。幂先验提供了一种对验前数据进行折合加权的信息先验构造方法, 非常适合于在靶场试验中为后续现场试验构造合理的先验信息。

2.3.1 幂先验的定义

信息更新具有序贯特性,因此,在靶场试验鉴定中采用具有参数信息先验的贝叶斯方法来融合历史数据和当前数据是一种自然的选择。为了解决传统的信息先验构造方法在融合大样本量验前数据时存在“淹没”当前数据的问题,Ibrahim 和 Chen^[105, 112]基于历史数据可用性的思想提出了幂先验,采用幂指数 δ ($0 \leq \delta \leq 1$) 来控制历史数据对当前研究的影响。

定义当前数据为 $\mathbf{D}=(n, \mathbf{y}, \mathbf{X})$, 其中, n 为样本量, \mathbf{y} 表示 $n \times 1$ 的响应向量, \mathbf{X} 表示 $n \times p$ 的协变量矩阵。给定当前数据的条件下, 兴趣参数 θ 的似然函数为 $L(\theta|\mathbf{D})$ 。令相似研究的历史数据记为 $\mathbf{D}_0=(n_0, \mathbf{y}_0, \mathbf{X}_0)$, $\pi_0(\theta|\cdot)$ 表示获得历史数据 \mathbf{D}_0 之前的 θ 的先验分布, 也称为 θ 的初始先验 (Initial prior)。假设在给定的 θ 下, 历史数据 \mathbf{D}_0 和当前数据 \mathbf{D} 为独立的随机样本。则在给定幂指数 δ 时, 用于当前研究的 θ 的幂先验定义为

$$\pi(\theta|\mathbf{D}_0, \delta) \propto (L(\theta|\mathbf{D}_0))^\delta \pi_0(\theta|c_0) \quad (2-5)$$

其中, $L(\theta|\mathbf{D}_0)$ 为基于历史数据 \mathbf{D}_0 得到的似然函数, c_0 为指定的初始先验超参数。式(2-5)给出的幂先验定义中, 参数 δ 度量了在当前研究中所需要的历史信息权值, $\delta=0$ 意味着不需要任何历史信息, 而 $\delta=1$ 说明历史数据的似然函数 $L(\theta|\mathbf{D}_0)$ 和当前研究的似然函数 $L(\theta|\mathbf{D})$ 有相等的权值, 对历史数据完全融合。因此, 式(2-5)可视为常用的先验分布贝叶斯更新的一种泛化形式。

幂先验定义中的幂指数 δ 可解释为一种度量先验信息精度或可信度的参数。如考虑方差已知时的正态分布样本, 假设 \mathbf{D}_0 包含 n_0 个观测值 y_1, y_2, \dots, y_{n_0} , 服从均值 θ 未知, 方差 σ^2 已知的正态分布。假定初始先验 $\pi_0(\theta)$ 为均匀的 (无先验信息), 则式(2-5)给出的幂先验意味着对当前数据集 \mathbf{D} , θ 的先验分布为 $\theta|\mathbf{D}_0, \delta \sim N(\bar{y}_0, \sigma^2 / \delta n_0)$, 其中, \bar{y}_0 为历史样本数据的均值。这样, 由于较小的 δ 意味着较大的幂先验方差, 而较大的 δ 意味着较小的幂先验方差, 因此 δ 可视为精度参数的一部分。

式(2-5)定义的幂先验最初是用于固定幂指数 δ 的情形, Chen 等^[112]注意到由于 δ 不一定是一个预先确定的值, 因此可将其扩展为随机情形。随机幂指数 δ 为研究者对历史数据的加权提供了更灵活的选择。由此, 通过为 δ 指定一个先验分布, Ibrahim 和 Chen 给出了 (θ, δ) 的联合幂先验定义, 其形式为

$$\pi(\theta, \delta|\mathbf{D}_0) \propto (L(\theta|\mathbf{D}_0))^\delta \pi_0(\theta) \pi(\delta|\gamma_0) \quad (2-6)$$

其中, γ_0 为指定的超参数向量。由于 $0 \leq \delta \leq 1$, Beta(a, b) 分布或均匀分布构成了 δ 的一种自然先验。在应用中, 可以通过修改超参数实现对验前数据信息权值的调整, Chen 等^[112]建议采用多个超参数集并进行灵敏度分析。

存在多个验前数据集时, 假设已有 m 个独立的研究, \mathbf{D}_{0j} 为第 j 个验前研究获取的数据集, $j=1, \dots, m$, $\mathbf{D}_0=(\mathbf{D}_{01}, \dots, \mathbf{D}_{0m})$, 对每个历史数据集选用不同的权参数 δ_j , 并取各个 δ_j 为独立且服从同一 Beta 分布的随机变量, 具有超参数 (a, b) 。令 $\underline{\delta}=(\delta_1, \dots, \delta_m)$, 则多源历史数据集的幂先验具有如下形式

$$\pi(\theta, \underline{\delta} | D_0) \propto \left(\prod_{j=1}^m [L(\theta | D_{0j})]^{\delta_j} \pi(\delta_j | a, b) \right) \pi_0(\theta) \quad (2-7)$$

式(2-7)定义的这种多源历史信息幂先验结构能够适应不同历史研究的数据集之间可能存在的分布不均匀性, 因此能够更精确的衡量不同历史数据集在用于推断兴趣参数后验分布时的作用。而且, 通过调整不同历史数据集幂指数的分布参数, 能够实现对不同历史数据集中有用信息的合理利用。

2.3.2 回归模型的幂先验

幂先验是一种通用的信息先验构造方法, 可用于构造不同模型参数的信息先验。根据响应变量的分布形式及似然函数的定义, 应用幂指数对历史数据加权, 可以得到不同模型参数的幂先验。Ibrahim^[105]和 Chen 等^[112]研究了幂先验在回归模型中的应用, 主要包括: 广义线性模型、广义线性混合模型、半参数的比例风险模型等。

2.3.2.1 广义线性模型

假设 $\{(x_i, y_i), i=1, 2, \dots, n\}$ 是当前研究的独立观测样本, 其中每个 y_i 是响应变量, $\mathbf{x}_i=(x_{i1}, x_{i2}, \dots, x_{ik})^T$ 是 $k \times 1$ 的协变量向量。给定 \mathbf{x}_i 时, 广义线性模型的响应变量 y_i 通常具有如下的指数族概率密度函数形式^[112]

$$f(y_i | x_i, \theta_i, \tau) = \exp\{\tau[y_i \theta_i - g(\theta_i)] + c(y_i, \tau)\}, \quad i=1, \dots, n, \quad (2-8)$$

式(2-8)中, 函数 g 和 c 由 y_i 的分布形式 (如二项分布、正态分布或 Poisson 分布等) 决定。分布参数可由离散参数 (Dispersion parameter) τ 和典范参数 (Canonical parameter) θ_i 表征。 $\theta_i = \theta(\eta_i)$ 为单调可微的联结函数 (Link function), $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ 为线性预报器, $\boldsymbol{\beta}=(\beta_1, \beta_2, \dots, \beta_k)^T$ 为 $k \times 1$ 回归系数向量。

令 $\mathbf{D}_0=(n_0, y_0, \mathbf{X}_0)$ 表示历史数据, 由 \mathbf{D}_0 可以得到 $\boldsymbol{\beta}$ 的似然函数

$$L(\boldsymbol{\beta} | D_0) = \prod_{i=1}^{n_0} \exp\{\tau_0 [y_{0i} \theta_{0i} - g(\theta_{0i})] + c(y_{0i}, \tau_0)\} \quad (2-9)$$

给定幂指数 δ 的条件下, 广义线性模型的幂先验形式为

$$\pi(\boldsymbol{\beta} | D_0, \delta) \propto [L(\boldsymbol{\beta} | D_0)]^\delta \quad (2-10)$$

当幂指数 δ 为服从 Beta(a, b) 分布的随机变量时, Chen^[112]给出了广义线性模型回归系数向量 $\boldsymbol{\beta}$ 和幂指数 δ 的联合幂先验分布形式

$$\pi(\boldsymbol{\beta}, \delta | D_0) \propto [L(\boldsymbol{\beta} | D_0)]^\delta \delta^{a-1} (1-\delta)^{b-1} \quad (2-11)$$

对于式(2-11)表示的联合幂先验 $\pi(\boldsymbol{\beta}, \delta | D_0)$ 的适当性 (Propriety) 条件, Chen^[112]给出了如下的引理。

引理 2.1 假定, 对于 $i=1, 2, \dots, n_0$

$$\exp\{\tau_0(y_{0i}\theta_{0i} - g(\theta_{0i}))\} \leq M \quad (2-12)$$

其中, M 为有限常值。假设存在 $y_{0i_1}, y_{0i_2}, \dots, y_{0i_k}$ ($1 \leq i_1 \leq i_2 \leq \dots \leq i_k$), 对某个 $t_0 > 0, j=1, 2, \dots, k$, 和相应的满秩为 k 的设计矩阵 $(x_{0i_1}, x_{0i_2}, \dots, x_{0i_k})^T$, 有

$$\int_{-\infty}^{\infty} e^{t_0|\eta|} \exp\left\{\tau \left[y_{0i_j} \theta(\eta) - g(\theta(\eta))\right]\right\} d\eta < \infty \quad (2-13)$$

或

$$\int_{-\infty}^{\infty} e^{t_0\eta^2} \exp\left\{\tau_0 \left[y_{0i_j} \theta(\eta) - g(\theta(\eta))\right]\right\} d\eta < \infty \quad (2-14)$$

如果满足以下条件之一

- (1). $a > k, b > 0$, 且符合式(2-13);
- (2). $a > k/2, b > 0$, 且符合式(2-14)

则, 联合幂先验 $\pi(\boldsymbol{\beta}, \delta | D_0)$ 为适当的, 即有 $\iint [L(\boldsymbol{\beta} | D_0)]^\delta \delta^{a-1} (1-\delta)^{b-1} d\delta d\boldsymbol{\beta} < \infty$ 。

2.3.2.2 广义线性混合模型

广义线性混合模型^[116] (GLMM) 是广义线性模型与线性混合效应模型的扩展, 其输出响应的概率分布形式一般可表示为

$$f(y_{it} | \boldsymbol{\beta}, \mathbf{b}_i, \tau) = \exp\{\tau[y_{it}\theta(\eta_{it}) - g(\theta(\eta_{it}))] + c(y_{it}, \tau)\} \quad (2-15)$$

其中, $\eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i$, \mathbf{b}_i 为 $q \times 1$ 的随机效应系数向量, \mathbf{x}_{it}^T 和 \mathbf{z}_{it}^T 为协变量向量。令 \mathbf{X}_i 表示第 i 行为 \mathbf{x}_{it}^T 的 $n_i \times p$ 矩阵; \mathbf{Z}_i 表示第 i 行为 \mathbf{z}_{it}^T 的 $n_i \times q$ 矩阵。定义 $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_N^T)^T$, $\mathbf{y} = (y_{11}, \dots, y_{Nn_N})^T$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_N^T)^T$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_N)$, 则基于 N 个群组的广义线性混合模型的 (\mathbf{y}, \mathbf{b}) 联合概率密度函数为

$$f(\mathbf{y}, \mathbf{b} | \boldsymbol{\beta}, T) = \prod_{i=1}^N \prod_{t=1}^{n_i} f(y_{it} | \boldsymbol{\beta}, \mathbf{b}_i) \pi(\mathbf{b}_i | T) \quad (2-16)$$

其中, $\pi(\mathbf{b}_i | T)$ 是均值为 0, 协方差矩阵 $V = T^{-1}$ 的正态分布。

假设已有历史数据 $\mathbf{D}_0 = (N_0, \mathbf{X}_0, \mathbf{y}_0, \mathbf{Z}_0)$, 其中各个符号的含义同上。在给定幂指数 δ 的条件下, $\boldsymbol{\beta}$ 的幂先验为

$$\pi(\boldsymbol{\beta} | \mathbf{D}_0, T, \delta) \propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [f(y_{0it} | \boldsymbol{\beta}, \mathbf{b}_{0i}, \tau)]^\delta \pi(\mathbf{b}_{0i} | T) d\mathbf{b}_{0i} \right) \pi_0(\boldsymbol{\beta} | c_0) \quad (2-17)$$

其中, $f(y_{0it} | \boldsymbol{\beta}, \mathbf{b}_{0i}, \tau)$ 可由式(2-15)计算。注意, 式(2-17)的结构形式是在给定

的随机效应下先取历史数据似然函数的幂, 然后对随机效应进行积分; 而不是先对随机效应进行积分, 然后求边缘历史数据似然函数的幂。显然, 采用上述结构形式其后验计算更易于实现。

由式(2-17)可见, 广义线性混合模型的完整幂先验可通过指定 δ 和 T 中参数的先验得到。假定 T 由两个参数 σ_b^2 和 ρ 决定, 并取 σ_b^2 的先验为逆 Gamma 分布 $IG(\alpha_0, \omega_0)$, ρ 的先验为尺度 Beta 分布 $\text{scbeta}(\varphi_0, \psi_0)$, 由此可以得到如下的联合幂先验分布形式

$$\begin{aligned} \pi(\boldsymbol{\beta}, \delta, \sigma_b^2, \rho | \mathbf{D}_0) \propto & \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [f(y_{0it} | \boldsymbol{\beta}, \mathbf{b}_{0i}, \tau)]^\delta \pi(\mathbf{b}_{0i} | T) d\mathbf{b}_{0i} \right) \pi_0(\boldsymbol{\beta} | c_0) \\ & \cdot \delta^{a-1} (1-\delta)^{b-1} \cdot (\sigma_b^2)^{-(\alpha_0+1)} \exp(-\sigma_b^{-2} \omega_0) \\ & \cdot (1+\rho)^{\varphi_0-1} (1-\rho)^{\psi_0-1} \end{aligned} \quad (2-18)$$

式(2-18)的适当性条件与广义线性模型的情况非常相似^[116]。由于 GLMM 的复杂性, 在贝叶斯抽样计算中可以采用重要性加权的边缘后验密度估计 (IWMDE)^[152]来得到 $\boldsymbol{\beta}$ 的边缘后验密度。由于式(2-18)给出的幂先验构造方法仅需要指定少量的超参数, 因此在变量选择中非常有吸引力。

2.3.2.3 比例风险模型

比例风险模型^[153]描述产品生存函数与环境因素之间的关系, 被广泛应用于生存分析, 一般可由如下形式的风险率函数定义

$$h(t, \mathbf{x}) = h_b(t) \exp(\mathbf{x}' \boldsymbol{\beta}) \quad (2-19)$$

其中, $h_b(t)$ 定义为 t 时刻的基准风险率函数, \mathbf{x} 表示总体中任意个体的 $p \times 1$ 协变量向量, $\boldsymbol{\beta}$ 表示 $p \times 1$ 回归系数向量。令 $0 \leq s_0 < s_1 < \dots < s_M$ 定义时间轴上的一个有限分割, 且有 $s_M > \max_i(t_i)$; $d_i = h_b(s_i) - h_b(s_{i-1})$ 表示区间 $(s_{i-1}, s_i]$, $i=1, \dots, M$ 上的基准风险率增量, $\Delta = (d_1, \dots, d_M)$ 。Ibrahim 和 Chen^[154]用一个分段常值的基准风险率模型来建立 $(\boldsymbol{\beta}, \Delta)$ 的似然函数, 并且仅用到测量的失效时间所处的区间信息。对总体中的任意个体, 比例风险模型在时刻 s 的累积分布函数为

$$\begin{aligned} F(s) &= 1 - \exp \left\{ -\exp(\eta) \int_0^s h_b(t) dt \right\} \\ &\simeq 1 - \exp \left\{ -\exp(\eta) \left((s-s_0)h_b(s_0) + \sum_{i=1}^M \delta_i (s-s_{i-1})^+ \right) \right\} \end{aligned} \quad (2-20)$$

其中, $\eta = \mathbf{x}^T \boldsymbol{\beta}$; $(t)^+ = t$ 当 $t > 0$ 时, 否则 $(t)^+ = 0$ 。令 p_i 表示区间 $(s_{i-1}, s_i]$ 内的失效概率; e_i 和 c_i 分别为第 i 个区间的失效数和右删失数据的数量, 为便于说明取第 i 个区间中前 e_i 个观测值为失效数据, 余下的 c_i 个观测值为右删失数据; \mathbf{x}_{ik} 表示第 i 个区间中第 k 个观测值的协变量向量, 定义

$$\begin{aligned}
 u_{ik}(\boldsymbol{\beta}) &= \exp\{\mathbf{x}_{ik}^T \boldsymbol{\beta}\}, \\
 a_i &= \sum_{j=i+1}^M \sum_{k=1}^{e_j} u_{jk}(\boldsymbol{\beta})(s_{j-1} - s_{i-1}), \\
 b_i &= \sum_{j=1}^M \sum_{k=e_j+1}^{e_j+c_j} u_{jk}(\boldsymbol{\beta})(s_j - s_{i-1}), \\
 T_i(\Delta) &= (s_i - s_{i-1}) \sum_{j=1}^i d_j
 \end{aligned}$$

令 $\mathbf{D}=(n, \mathbf{y}, \mathbf{X}, \mathbf{v})$ 表示当前研究的数据, 其中 $\mathbf{v}=(v_1, \dots, v_n)^T$ 为 $n \times 1$ 删失数据指示向量, 则当前研究在所有 M 个区间上的似然函数为

$$L(\boldsymbol{\beta}, \Delta | \mathbf{D}) = \left\{ \prod_{i=1}^M \exp\{-d_i(a_i + b_i)\} \right\} \cdot \left\{ \prod_{i=1}^M \prod_{k=1}^{e_i} (1 - \exp\{-u_{ik}(\boldsymbol{\beta})T_i(\Delta)\}) \right\}$$

令 $\mathbf{D}_0=(n_0, \mathbf{y}_0, \mathbf{X}_0, \mathbf{v}_0)$ 表示历史数据, $\pi_0(\boldsymbol{\beta}, \Delta)$ 为 $(\boldsymbol{\beta}, \Delta)$ 的初始先验。Ibrahim 和 Chen^[105]建议采用如下的 $(\boldsymbol{\beta}, \Delta, \delta)$ 联合幂先验

$$\pi(\boldsymbol{\beta}, \Delta, \delta | \mathbf{D}_0) \propto L(\boldsymbol{\beta}, \Delta | \mathbf{D}_0)^\delta \pi_0(\boldsymbol{\beta}, \Delta) \delta^{a-1} (1 - \delta)^{b-1} \quad (2-21)$$

其中, $L(\boldsymbol{\beta}, \Delta | \mathbf{D}_0)$ 为 $(\boldsymbol{\beta}, \Delta)$ 基于历史数据 \mathbf{D}_0 的似然函数。为了简化先验的形式, 假设 $\pi_0(\boldsymbol{\beta}, \Delta) = \pi_0(\boldsymbol{\beta})\pi_0(\Delta)$ 。Ibrahim 和 Chen^[154]指出, 如果 $\pi_0(\boldsymbol{\beta}) \propto 1$, 则当 $\pi_0(\Delta)$ 为适当的, 且 $a > p$ 时, 式(2-21)给出的联合幂先验为适当的。此外, 他们还将幂先验方法用于比例风险模型的变量选择问题。

2.3.3 幂先验的优良性分析

幂先验的优良性是指将其用于统计推断时, 能够充分、合理地利用当前数据和历史数据的信息, 得到的兴趣参数后验分布在一定信息量测度或准则下是最优的。假设 f_0 和 f_1 为满足 Lebesgue 测度的概率密度函数, 它们之间的 Kullback-Leibler (KL) 散度定义为

$$K(f_0, f_1) = \int \log \left(\frac{f_0(\theta)}{f_1(\theta)} \right) f_0(\theta) d\theta$$

令 $f_0 = \pi(\theta | \mathbf{D}, \mathbf{D}_0, \delta=0)$, $f_1 = \pi(\theta | \mathbf{D}, \mathbf{D}_0, \delta=1)$ 分别表示不融合历史数据的后验密度函数和完全融合历史与当前数据的后验密度函数。定义 θ 的任意一个密度函数 $g(\theta)$, 它与上述两种后验概率密度函数的 KL 散度凸和可表示为

$$K_g = (1 - \delta)K(g, f_0) + \delta K(g, f_1) \quad (2-22)$$

其中, δ 为 0~1 之间的一个确定的标量。KL 散度凸和反映了密度函数 g 与 f_0 和 f_1 之间的距离, 其值越小表明与 f_0 和 f_1 越接近。

定理 2.1 给定幂指数 $0 \leq \delta \leq 1$, 由幂先验得到的兴趣参数 θ 的后验分布为

$$\pi(\theta | \mathbf{D}_0, \mathbf{D}, \delta) \propto L(\theta | \mathbf{D}_0)^\delta L(\theta | \mathbf{D}) \pi_0(\theta) \quad (2-23)$$

则，式(2-23)的后验分布能够实现式(2-22)定义的 KL 散度凸和最小，即

$$g_{opt} \propto L(\theta | \mathbf{D}) L(\theta | \mathbf{D}_0)^\delta \pi_0(\theta)$$

证明：

根据式(2-22)，函数 g 与 f_0 和 f_1 之间的 KL 散度凸和为

$$\begin{aligned} & (1-\delta)K(g, f_0) + \delta K(g, f_1) \\ &= (1-\delta) \int \log(g(\theta)/f_0(\theta)) g(\theta) d\theta + \delta \int \log(g(\theta)/f_1(\theta)) g(\theta) d\theta \\ &= \int \log[g(\theta)/f_0(\theta)]^{1-\delta} g(\theta) d\theta + \int \log[g(\theta)/f_1(\theta)]^\delta g(\theta) d\theta \\ &= \int \log\left[g(\theta) / (f_0(\theta)^{1-\delta} f_1(\theta)^\delta)\right] g(\theta) d\theta \\ &= K\left(g, \frac{f_0^{1-\delta} f_1^\delta}{h(\delta)}\right) - \log(h(\delta)), \end{aligned}$$

其中， $h(\delta) = \int f_0(\theta)^{1-\delta} f_1(\theta)^\delta d\theta$ 是与 g 无关的 $f_0^{1-\delta} f_1^\delta$ 的归一化常数。显然，能够实现 KL 散度凸和最小的概率密度函数

$$g = g_{opt} = \frac{f_0^{1-\delta} f_1^\delta}{h(\delta)} \propto f_0^{1-\delta} f_1^\delta \propto L(\theta | \mathbf{D}) L(\theta | \mathbf{D}_0)^\delta \pi_0(\theta)$$

上式右侧即为式(2-23)的后验分布。

证毕

根据 Zellner 的信息处理准则^[155] (Information processing rule, IPR)，当输出与输入的信息比等于 1 时，该 IPR 的效率为 100%。考虑信息准则函数 $\Delta[g(\theta)]$ 的加权形式

$$\begin{aligned} \Delta[g(\theta)] &= I_{out} - I_{in} \\ &= \int g(\theta) \log(g(\theta)) d\theta + \int g(\theta) \log(m) d\theta \\ &\quad - \left[\int g(\theta) \log(\pi(\theta)) d\theta + \int g(\theta) \log(L(\theta)) d\theta \right] \\ &\equiv \int g(\theta) \log \left[\frac{g(\theta)}{L(\theta) \pi(\theta) / m} \right] d\theta \end{aligned} \quad (2-24)$$

其中， $g(\theta)$ 为适当的概率密度函数， $m = \int L(\theta) \pi(\theta) d\theta$ 为数据边缘密度， $L(\theta)$ 为 θ 的似然函数， $\pi(\theta)$ 为 θ 的先验分布。由于 KL 散度非负，显然最小化式(2-24)的概率密度函数为

$$g^*(\theta) = \frac{L(\theta) \pi(\theta)}{m} \quad (2-25)$$

此时，有 $\Delta[g^*(\theta)] = 0$ ，该 IPR 的效率为 100%。

定理 2.2 令 $m(\mathbf{D}, \mathbf{D}_0) = \int L(\theta | \mathbf{D}_0)L(\theta | \mathbf{D})\pi_0(\theta)d\theta$, $m(\mathbf{D}) = \int L(\theta | \mathbf{D})\pi_0(\theta)d\theta$ 。考虑式(2-24)的加权形式

$$\begin{aligned}\Delta[g(\theta)] &= \int g(\theta)\log(g(\theta))d\theta + \int g(\theta)\log(m(\mathbf{D}, \mathbf{D}_0))d\theta \\ &\quad - w_0 \int g(\theta)\log(\pi_0(\theta))d\theta \\ &\quad - w_D \int g(\theta)\log(L(\theta | \mathbf{D}))d\theta \\ &\quad - w_{D_0} \int g(\theta)\log(L(\theta | \mathbf{D}_0))d\theta\end{aligned}$$

其中, w_0 , w_D , 和 w_{D_0} 分别为初始先验 $\pi_0(\theta)$ 、当前数据和历史数据的权系数。则当取 $w_0=w_D=1$, $w_{D_0}=\delta$ 时, 有

$$K_g = \Delta[g(\theta)] + C \quad (2-26)$$

其中, $C = (1-\delta)(\log(m(\mathbf{D})) - \log(m(\mathbf{D}, \mathbf{D}_0)))$ 为与 g 无关的常值。

证明:

由于 $m(\mathbf{D}, \mathbf{D}_0)$ 与 θ 无关, 且 $\int g(\theta)d\theta = 1$, 有

$$\int g(\theta)\log(m(\mathbf{D}, \mathbf{D}_0))d\theta = \log(m(\mathbf{D}, \mathbf{D}_0))$$

由此

$$\begin{aligned}\Delta[g(\theta)] &= \int g(\theta)\log(g(\theta))d\theta + \log(m(\mathbf{D}, \mathbf{D}_0)) \\ &\quad - w_0 \int g(\theta)\log(\pi_0(\theta))d\theta \\ &\quad - w_D \int g(\theta)\log(L(\theta | \mathbf{D}))d\theta \\ &\quad - w_{D_0} \int g(\theta)\log(L(\theta | \mathbf{D}_0))d\theta\end{aligned}$$

式(2-22)的 K_g 可表示为

$$\begin{aligned}K_g &= (1-\delta) \int g(\theta)\log\left(\frac{g(\theta)m(\mathbf{D})}{L(\theta)\pi_0(\theta)}\right)d\theta + \delta \int g(\theta)\log\left(\frac{g(\theta)m(\mathbf{D}, \mathbf{D}_0)}{L(\theta)L(\theta | \mathbf{D}_0)\pi_0(\theta)}\right)d\theta \\ &= \int g(\theta)\log(g(\theta))d\theta + (1-\delta) \int \log[m(\mathbf{D})] + \delta \log[m(\mathbf{D}, \mathbf{D}_0)] \\ &\quad - \int g(\theta)\log[L(\theta | \mathbf{D})]d\theta - \delta \int g(\theta)\log[L(\theta | \mathbf{D}_0)]d\theta \\ &\quad - \int g(\theta)\log[\pi_0(\theta)]d\theta.\end{aligned}$$

显然, 当 $w_0=w_D=1$, $w_{D_0}=\delta$ 时,

$$K_g = \Delta[g(\theta)] + (1-\delta)(\log(m(\mathbf{D})) - \log(m(\mathbf{D}, \mathbf{D}_0))) = \Delta[g(\theta)] + C \quad \text{证毕}$$

定理 2.2 表明 K_g 和 $\Delta[g(\theta)]$ 具有相同的最小值, 因此幂先验是一种效率为 100% 信息处理准则。当存在多个历史数据集时, 上述等效关系同样适用。

对于幂指数 δ 为固定值的情形, 由以上 KL 散度凸和以及 Zellner 的 IPR 可以证明幂先验的优良性。然而, 在式(2-6)给出的随机幂指数的幂先验中, 如果将似然函数乘以一个常值, 幂指数 δ 的后验众数就会改变。而且, 对任意的历史数据 \mathbf{D}_0 和当前数据 \mathbf{D} , 总能找到一个常值 k_0 , 将其乘以似然函数后, 就会使 δ 的边缘后验众数变为 0。对此, Duan 给出了如下引理^[117]。

引理 2.2: 假设当前数据 \mathbf{D} 来自密度函数为 $f(x|\theta)$ 的分布, \mathbf{D}_0 为相关的历史数据集。此外, 假设 $\pi(\delta)=1$ 为 δ 的先验, θ 在 δ 条件下的后验分布对任意 δ 都是适当的。则对任何 \mathbf{D}_0 和 \mathbf{D} 应用式(2-6)给出的标准幂先验方法时, 至少存在一个正常数 k_0 满足 $\pi(\delta|\mathbf{D}_0, \mathbf{D})$ 的众数为 0, 其中 $L(\theta|x)=k_0f(x|\theta)$ 。

由引理 2.2 可见, 标准幂先验方法中这种 δ 强烈趋于 0 的现象对应用随机幂指数 δ 的灵活性造成了危害, 使历史数据的作用被低估。

2.4 随机幂指数的规则化幂先验

对于随机幂指数的情况, 应用(2-6)式的定义方法时, 幂指数的后验分布总是趋向于 0, 表明大部分历史数据集的信息没有被使用。针对这种现象, Duan^[117]提出了一种规则化幂先验 (Normalized power prior) 的贝叶斯方法, 使幂指数能够自动量化当前数据与历史数据的不均匀性, 由此实现对历史信息作用的合理控制。此外, Duan 提出的规则化幂先验对分布适当性的要求较少。

2.4.1 规则化幂先验的定义

幂指数 δ 为固定值的情况下, 式(2-5)定义的标准幂先验具有良好的适用性。然而, 对于幂指数 δ 为随机变量的情况, 式(2-6)定义的联合幂先验在实际用于 Bernoulli 模型和正态均值模型时, 无论历史数据与当前数据的相容性多好, 历史数据对后验推断的影响普遍很小, 即 δ 的分布总体趋向于 0。此时, 对 θ 的后验推断与不利用历史数据的情况没什么不同。此外, 式(2-6)定义的先验也可能是不适当的。考虑到根据可用的历史信息构造的先验最好是适当的, Duan^[117]提出了如下的 (θ, δ) 的规则化幂先验

$$\pi(\theta, \delta|\mathbf{D}_0) = \pi(\delta|\mathbf{D}_0)\pi(\theta|\delta, \mathbf{D}_0) \propto \frac{L(\theta|\mathbf{D}_0)^\delta \pi(\theta)\pi(\delta)I_A(\delta)}{\int_{\Theta} L(\theta|\mathbf{D}_0)^\delta \pi(\theta)d\theta} \quad (2-27)$$

其中, $A = \left\{ \delta: 0 < \int_{\Theta} L(\theta|\mathbf{D}_0)^\delta \pi(\theta)d\theta < \infty \right\}$, $I_A(\delta)$ 为示性函数, 表示当 $\delta \in A$ 时, $I_A(\delta)=1$, 否则 $I_A(\delta)=0$ 。式(2-6)和式(2-27)的不同之处在于: 若给定的 $\pi(\delta)$ 为适当的, 则由式(2-27)定义的 (θ, δ) 规则化幂先验是适当的, 而由式(2-6)定义的

(θ, δ) 标准幂先验未必适当。此外, 根据定理 2.1, 式(2-6)定义的随机幂指数 δ 的标准幂先验不符合似然原理。

应用当前数据 \mathbf{D} 来更新式(2-27)的先验分布 $\pi(\theta, \delta | \mathbf{D}_0)$, 可以推导出 (θ, δ) 的联合后验分布为

$$\pi(\theta, \delta | \mathbf{D}_0, \mathbf{D}) \propto L(\theta | \mathbf{D}) \pi(\theta, \delta | \mathbf{D}_0) \propto \frac{L(\theta | \mathbf{D}) L(\theta | \mathbf{D}_0)^\delta \pi(\theta) \pi(\delta)}{\int_{\Theta} L(\theta | \mathbf{D}_0)^\delta \pi(\theta) d\theta} I_A(\delta)$$

对上式在 θ 上积分, δ 的边缘后验分布可写为

$$\pi(\delta | \mathbf{D}_0, \mathbf{D}) \propto \pi(\delta) \frac{\int_{\Theta} L(\theta | \mathbf{D}) L(\theta | \mathbf{D}_0)^\delta \pi(\theta) d\theta}{\int_{\Theta} L(\theta | \mathbf{D}_0)^\delta \pi(\theta) d\theta} I_A(\delta) \quad (2-28)$$

与此类似, θ 的边缘后验分布 $\pi(\theta | \mathbf{D}_0, \mathbf{D})$ 可通过对 δ 的积分得到。如果只对 θ 感兴趣, 可在初期就对 $\pi(\theta, \delta | \mathbf{D}_0)$ 在 δ 上进行积分, 由历史信息更新得到 θ 的一个新的先验分布

$$\pi(\theta | \mathbf{D}_0) = \int_A \pi(\theta, \delta | \mathbf{D}_0) d\delta \propto \pi(\theta) \int_A \frac{L(\theta | \mathbf{D}_0)^\delta \pi(\delta) I_A(\delta)}{\int_{\Theta} L(\theta | \mathbf{D}_0)^\delta \pi(\theta) d\theta} d\delta \quad (2-29)$$

通过适当的融合历史数据, $\pi(\theta | \mathbf{D}_0)$ 可视为对当前数据进行贝叶斯分析的一种信息先验。因此, θ 的后验分布可写为

$$\begin{aligned} \pi(\theta | \mathbf{D}_0, \mathbf{D}) &\propto \pi(\theta | \mathbf{D}_0) L(\theta | \mathbf{D}_0, \mathbf{D}) \\ &\propto \pi(\theta) L(\theta | \mathbf{D}) \int_A \frac{L(\theta | \mathbf{D}_0)^\delta \pi(\delta) I_A(\delta)}{\int_{\Theta} L(\theta | \mathbf{D}_0)^\delta \pi(\theta) d\theta} d\delta \end{aligned} \quad (2-30)$$

与 Ibrahim 和 Chen 提出的扩展方法类似, 式(2-27)定义的先验很容易推广到多源历史数据集的情形。假定有 m 个与当前研究类似的历史研究, \mathbf{D}_{0j} 定义第 j 个研究的历史数据, $j=1, \dots, m$, $\mathbf{D}_0=(\mathbf{D}_{01}, \dots, \mathbf{D}_{0m})$, 对每个历史研究应用不同的权参数 δ_j 。此外, 可以假设各个 δ_j 为独立同分布(i.i.d.)的 Beta 随机变量, 具有超参数 (α, β) 。令 $\underline{\delta}=(\delta_1, \dots, \delta_m)$, 式(2-27)定义的规则化幂先验可推广为

$$\pi(\theta, \underline{\delta} | \mathbf{D}_0) \propto \frac{\left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \pi(\delta_j | \alpha, \beta) \right) \pi(\theta)}{\int_{\Theta} \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \right) \pi(\theta) d\theta} I_B(\underline{\delta})$$

其中, $B = \left\{ (\delta_1, \dots, \delta_m) : 0 < \int_{\Theta} \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \right) \pi(\theta) d\theta < \infty \right\}$ 。

不同研究之间的数据往往存在分布不均匀性, 而由一个研究得到数据一般服从同一分布。上述框架能够适应不同来源或在不同时间阶段采集的数据集可能存在的分布不均匀性。例如, 在武器系统定型试验中, 我们可能取类似型号

的试验数据作为历史数据集。而且,即使是同一武器型号,在不同试验阶段得到的试验数据也可以分成几个历史数据集以保证每个数据集内部的分布均匀性。这样,历史数据的作用可以更精确的进行评价。

2.4.2 正态参数的规则化幂先验

本节研究总体方差未知时,应用验前数据和现场试验数据对正态分布均值的估计,假设现场试验数据 $\mathbf{D}=(x_1, \cdots, x_n)$ 来自均值 μ 和方差 σ^2 均未知的正态分布, $\mathbf{D}_0=(x_{01}, \dots, x_{0n_0})$ 为验前数据。定义

$$\bar{x}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} x_{0i}, \quad \hat{\sigma}_0^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} (x_{0i} - \bar{x}_0)^2$$

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$$

分别为验前数据和现场试验数据的样本均值及方差。根据式(2-27)给出的定义,对于方差未知的正态分布的规则化幂先验可由以下的引理给出^[117]。

引理 2.3: 令 $\pi_0(\mu, \sigma^2)$ 表示 (μ, σ^2) 的初始先验分布, 假设 δ 的先验分布为 $Beta(\alpha_\delta, \beta_\delta)$, 其中 α_δ 和 β_δ 为已知的超参数。则 (μ, σ^2, δ) 的规则化幂先验为

$$\pi(\mu, \sigma^2, \delta | \mathbf{D}_0) \propto \frac{(\sigma^2)^{-\frac{\delta n_0}{2}} \exp \left\{ -\frac{\delta n_0}{2\sigma^2} [\hat{\sigma}_0^2 + (\mu - \bar{x}_0)^2] \right\} \pi_0(\mu, \sigma^2) \delta^{\alpha_\delta-1} (1-\delta)^{\beta_\delta-1}}{\int_0^1 \int_{-\infty}^{+\infty} (\sigma^2)^{-\frac{\delta n_0}{2}} \exp \left\{ -\frac{\delta n_0}{2\sigma^2} [\hat{\sigma}_0^2 + (\mu - \bar{x}_0)^2] \right\} \pi_0(\mu, \sigma^2) d\mu d\sigma^2} \quad (2-31)$$

当初始先验 $\pi_0(\mu, \sigma^2)$ 为特定的无信息先验时, Duan 给出了三个推论^[117]。

推论 2.1: 假设 (μ, σ^2) 的初始先验为 $\pi_0(\mu, \sigma^2) \propto 1/(\sigma^2)^a$, 其中 $a>0$ 为预先确定的常值。 (μ, σ^2, δ) 的联合幂先验分布可表示为

$$\pi(\mu, \sigma^2, \delta | \mathbf{D}_0) \propto \frac{\delta^{\frac{\delta n_0}{2} + a + \alpha_\delta - 2} (1-\delta)^{\beta_\delta-1}}{\left(\frac{2\sigma^2}{n_0 \hat{\sigma}_0^2} \right)^{\frac{\delta n_0}{2} + a} \Gamma \left(\frac{\delta n_0 - 3}{2} + a \right)} \exp \left\{ -\frac{\delta n_0}{2\sigma^2} [\hat{\sigma}_0^2 + (\mu - \bar{x}_0)^2] \right\}$$

其中, $b = \frac{3-2a}{n_0}$, $\delta \in \begin{cases} (b, 1], & b \geq 0 \\ [0, 1], & b < 0 \end{cases}$ 。

注意, $a=1$ 对应参考先验, 而 $a=3/2$ 为 Jeffreys 先验。因此, 如果应用 θ 的参考先验, 对于正态均值模型, (2-27)式中定义的集合 A 为 $(1/n_0, 1]$; 而如果应用 θ 的 Jeffreys 先验, 则 $A=(0, 1]$ 。

推论 2.2: 假设 $\pi_0(\mu, \sigma^2) \propto 1/(\sigma^2)^a$, δ 的边缘后验分布为

$$\pi(\delta | \mathbf{D}_0, \mathbf{D}) \propto \frac{\delta^{\frac{\delta n_0 + a + \alpha_\delta - 2}{2}} (1 - \delta)^{\beta_\delta - 1} \Gamma\left(\frac{\delta n_0 + n - 3}{2} + a\right)}{\left[\frac{\delta n}{\delta n_0 + n} \frac{(\bar{x}_0 - \bar{x})^2}{\hat{\sigma}_0^2} + \delta + \frac{n \hat{\sigma}^2}{n_0 \hat{\sigma}_0^2} \right]^{\frac{\delta n_0 + n - 3}{2} + a} \Gamma\left(\frac{\delta n_0 - 3}{2} + a\right)}$$

上式中, δ 的取值范围与推论 2.1 中的相同。

推论 2.3: 假设 $\pi_0(\mu, \sigma^2) \propto 1/(\sigma^2)^a$, 给定 δ 和数据 $(\mathbf{D}_0, \mathbf{D})$ 时, μ 的条件后验分布服从学生 t 分布, 且位置参数(location parameter)和尺度参数(scale parameter)分别为:

$$\frac{\delta n_0 \bar{x}_0 + n \bar{x}}{\delta n_0 + n},$$

$$\sqrt{\frac{2}{C(\delta) (\delta n_0 + n + 2a - 3)(\delta n_0 + n)}}$$

自由度为 $\delta n_0 + n + 2a - 3$, 其中

$$C(\delta) = \frac{2}{\frac{\delta n_0 n (\bar{x}_0 - \bar{x})^2}{\delta n_0 + n} + \delta n_0 \hat{\sigma}_0^2 + n \hat{\sigma}^2}$$

此外, 在给定幂指数 δ 和数据的条件下, σ^2 的条件后验分布服从参数为 $\frac{\delta n_0 + n + 2a - 3}{2}$, $C(\delta)^{-1}$ 的逆 Gamma (inverse-gamma) 分布。

2.4.3 规则化幂先验的优良性分析

对于随机幂指数 δ 的规则化幂先验, 其优良性除了 2.3.3 节讨论的兴趣参数条件后验分布 $\pi(\theta | \mathbf{D}_0, \mathbf{D}, \delta)$ 的 KL 散度凸和最优和 Zellner 的 IPR 效率最优外, 还能产生最优的幂指数后验分布 $\pi(\delta | \mathbf{D}_0, \mathbf{D})$ 。Shannon 应用互信息函数^[156] (Mutual information function) 来测量两个随机变量 X 和 Y 之间的依存性, 并通过期望的熵差来定义互信息

$$\mathcal{I}(Y \wedge X) \equiv H(Y) - E_x[H(Y | x)] = E_{(x,y)} \left[\ln \frac{f(x|y)}{f(x)} \right]$$

其中, $H(Y)$ 是 $f(y)$ 的熵, $H(Y|x)$ 是条件分布 $f(y|x)$ 的熵。可见, 互信息度量了一个随机变量包含另一个随机变量的信息量, 或者可以理解为由于获得另一个随机变量的信息后, 原随机变量不确定度的缩减值。

借用互信息的概念, 将 $\ln(m(\mathbf{D} | \mathbf{D}_0)/m(\mathbf{D}))$ 看作由两个样本 \mathbf{D}_0 和 \mathbf{D} 之间观测到的互信息, 其中 $m(\mathbf{D}) = \int_{\Theta} L(\theta | \mathbf{D}) \pi(\theta) d\theta$ 为 \mathbf{D} 的边缘密度, $m(\mathbf{D} | \mathbf{D}_0)$ 为给

定观测值 D_0 下 D 的密度函数。 $\ln(m(D|D_0)/m(D))$ 项度量了历史数据中可用于解释当前数据的信息量。 $m(D|D_0)$ 不仅依赖于采样分布模型,还取决于历史数据的融合方式,为了识别这种融合方式,将 $\ln m(D|D_0)$ 写成如下的形式,

$$\ln m(D|D_0) = \ln \frac{m(D|D_0, \delta)\pi(\delta|D_0)}{\pi(\delta|D_0, D)} \quad (2-32)$$

幂指数 δ 用于度量历史数据与当前数据来自同一分布的概率,因此仅由历史数据本身无法提供 δ 的任何信息。由此可见,采用 δ 的先验 $\pi(\delta)$ 来代替式(2-32)中的 $\pi(\delta|D_0)$,不会对 $\ln(m(D|D_0)/m(D))$ 的计算造成影响。这样,测量的 D 和 D_0 之间的互信息变为 $\ln(m(D|D_0, \delta)\pi(\delta)/\pi(\delta|D_0, D)) - \ln m(D)$ 。并且,我们还可以用 δ 的后验分布来计算其期望值,定义

$$\varpi(D \wedge D_0) \equiv E_{\pi(\delta|D, D_0)} \left[\ln \frac{m(D|D_0, \delta)\pi(\delta)}{\pi(\delta|D_0, D)} - \ln m(D) \right] \quad (2-33)$$

为历史数据 D_0 与当前数据 D 之间的加权互信息。

幂指数为固定值时的标准幂先验方法在融合历史数据时的优良性已在 2.3.3 节中给出了证明。当幂指数为随机变量时,规则化幂先验在融合历史数据时的优良性可以通过最大化式(2-33)的加权互信息来证明。

在幂先验的框架下, $\varpi(D_0 \wedge D)$ 测量了通过幂先验模型传递的历史数据期望信息对当前研究的贡献大小。因此,寻找最优幂先验的关键在于实现加权互信息 $\varpi(D_0 \wedge D)$ 的极大化,这一点可进一步归结为寻找最优的幂指数后验分布 $\pi^*(\delta|D_0, D)$ 。Duan^[117]用如下的引理对此进行了总结。

引理 2.4: 最大化式(2-33)给出的加权互信息 $\varpi(D_0 \wedge D)$ 的幂指数后验密度函数 $\pi(\delta|D_0, D)$ 为

$$\pi^*(\delta|D_0, D) \propto \pi(\delta) \frac{\int_{\Theta} L(\theta|D)L(\theta|D_0)^\delta \pi(\theta) d\theta}{\int_{\Theta} L(\theta|D_0)^\delta \pi(\theta) d\theta}$$

显然,上式给出的最优 $\pi^*(\delta|D_0, D)$ 恰好是式(2-28)给出的由规则化幂先验得到的幂指数 δ 的边缘后验分布。

Duan^[117]对规则化幂先验和标准幂先验进行了广泛的比较,指出规则化幂先验的构造方法更符合似然原理,且保证 (θ, δ) 联合先验适当性的条件更易满足。并且,在当前数据为小样本而历史数据的样本量非常大的情况下,规则化幂先验对兴趣参数 θ 和幂指数 δ 的后验估计性能优于标准幂先验方法,且具有较小的后验估计均方误差 (MSE)。

2.5 本章小结

本章首先回顾了贝叶斯统计学的基本思想、计算公式和先验构造方法，在贝叶斯理论的框架下对靶场试验中的统计问题进行了描述，将其归结为一类包含大量不完全可信的验前数据和小样本量的现场试验数据的试验分析与评估问题，并由此引入适用于靶场试验分析与评估的幂先验。其次，介绍了标准幂先验的构造方法及常用的几种回归模型的幂先验，并从 $\pi(\theta | \mathbf{D}_0, \mathbf{D}, \delta)$ 的 KL 散度凸和以及 Zellner 的 IPR 效率两方面分析了幂先验的优良性，说明幂指数确定时幂先验方法在贝叶斯评估中的有效性。最后，针对幂指数为随机变量的情形，分析了标准幂先验方法下幂指数分布的趋 0 特性，并通过先验分布的适当性说明规则化幂先验的合理性，给出了正态分布的规则化幂先验分布形式，并进一步通过 $\pi(\delta | \mathbf{D}_0, \mathbf{D})$ 的互信息函数特性说明规则化幂先验的优良性。

第3章 基于贝叶斯正则化的靶场试验设计方法

3.1 引言

靶场试验是武器装备鉴定/定型过程中的重要环节,根据不同试验手段(包括地面试验、仿真试验和飞行试验)获取的装备性能数据做出鉴定/定型结论。其中,靶场试验设计是一项核心工作。试验设计中,一般将试验考核的武器性能指标称为响应变量,影响武器性能指标的因素称为试验的因子,因子所处的各个状态称为因子的水平,各个因子对响应变量的影响称为因子效应^[31, 157]。现场试验是在实际应用条件下考核装备性能,数据真实可信,具有不可替代性。然而,受客观条件限制,现场试验只能获得小子样数据。因此,如何应用验前信息对现场试验进行优化设计,确定合理的试验设计点数,并使有限的现场试验样本最好的覆盖因子空间,是靶场试验设计中的重点和难点。

武器系统性能与影响因子之间的关系是一个未知的多元函数。靶场试验的目的就是通过因子空间中有限个试验点上的数据,对该未知多元函数进行估计,以得到与真实模型尽可能接近的性能评估模型,用于对系统性能的预测。均匀设计不对试验模型做任何假设,需要的设计点数较少且对模型变化具有一定的稳健性,在靶场试验中得到了广泛应用。目前,现场试验通常是按照因子水平数选择小子样均匀设计方案,并按照均匀设计表安排实施。然而,均匀设计无法回答现场试验需要的最少设计点数以及试验设计方案是否最优的问题。当存在验前信息时,为了避免试验资源浪费和系统性能的评估风险,应在合理利用验前信息的基础上对现场试验进行优化设计。

基于贝叶斯正则化的靶场试验设计方法以仿真试验数据为验前信息,应用贝叶斯正则化方法分析影响系统性能的因子效应,根据得到的因子效应信息构造靶场试验设计信息矩阵,实现现场试验的最优回归设计。本章结构安排如下:首先,基于回归模型描述试验设计中输出响应和因子之间的关系,并指出因子效应分析对靶场试验设计的重要性。其次,通过分析正则化方法在求解不适定问题中的稳健性,提出贝叶斯正则化因子效应分析方法。然后,针对靶场试验设计问题,给出考虑因子效应信息的靶场试验设计信息矩阵,提出贝叶斯正则化试验设计信息矩阵构造方法,给出一种兼顾可能因子效应项的备选附加点D-最优设计用于靶场试验设计。最后,通过火炮射表试验的示例来验证基于贝叶斯正则化的靶场试验设计方法的有效性。

3.2 靶场试验设计问题的描述

靶场试验设计的目的是确定仿真和现场试验应该在试验因子空间的哪些点上进行, 为便于问题描述, 本章对试验的模型做如下三点假设:

(1) 设 p 维因子向量表示为 $\mathbf{x}=(x_1, \cdots, x_p)^T$, \mathcal{X} 是 p 维因子空间 (也称为试验点集), 假定 \mathcal{X} 为有界闭集, 试验点只能在 \mathcal{X} 内选取。

(2) 设 $B_1(\mathbf{x}), \cdots, B_L(\mathbf{x})$ 是 \mathcal{X} 上已知的 L 个线性独立的一组基函数, 记

$$\mathbf{B}(\mathbf{x}) = (B_1(\mathbf{x}), \cdots, B_L(\mathbf{x}))^T$$

(3) 响应变量 y 与因子向量 \mathbf{x} 之间满足以下的回归模型:

$$y(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{B}(\mathbf{x}) + \varepsilon \quad (3-1)$$

其中, ε 是随机误差, 服从正态分布 $N(0, \sigma^2)$, $\boldsymbol{\beta}=(\beta_1, \cdots, \beta_L)^T$ 是未知参数向量, σ^2 是未知方差, $y(\mathbf{x})$ 表示在试验点 $\mathbf{x} \in \mathcal{X}$ 上响应变量的观测值, 而且不同试验点上的各个观测值之间相互独立。模型(3-1)也可用其均值形式表示

$$E(y) = \boldsymbol{\beta}^T \mathbf{B}(\mathbf{x}) = \mathbf{B}(\mathbf{x})^T \boldsymbol{\beta} \quad (3-2)$$

假设因子空间 \mathcal{X} 中的试验点用字母 z 表示, 而它的分量用带下标的 x 表示, 即 $z=(x_1, \cdots, x_p)^T \in \mathcal{X}$ 。设 N 是一个正整数, z_1, \cdots, z_N 是 \mathcal{X} 中的 N 个点, 称有限点集 $\xi_N=\{z_1, \cdots, z_N\}$ 为一个精确设计。记 ξ_N 中各个点上的响应变量观测值为 $\mathbf{y}=(y(z_1), \cdots, y(z_N))^T$, 由模型(3-2)可知

$$E(\mathbf{y}) = E \begin{pmatrix} y(z_1) \\ y(z_2) \\ \vdots \\ y(z_N) \end{pmatrix} = \begin{pmatrix} \mathbf{B}(z_1)^T \boldsymbol{\beta} \\ \mathbf{B}(z_2)^T \boldsymbol{\beta} \\ \vdots \\ \mathbf{B}(z_N)^T \boldsymbol{\beta} \end{pmatrix} = \begin{bmatrix} B_1(z_1) & B_2(z_1) & \cdots & B_L(z_1) \\ B_1(z_2) & B_2(z_2) & \cdots & B_L(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(z_N) & B_2(z_N) & \cdots & B_L(z_N) \end{bmatrix} \boldsymbol{\beta} = \mathbf{F}(\xi_N)^T \boldsymbol{\beta}$$

定义模型(3-1)或(3-2)的信息矩阵为

$$\mathbf{M}(\xi_N) = \mathbf{F}(\xi_N) \mathbf{F}(\xi_N)^T = \sum_{j=1}^N \mathbf{B}(z_j) \mathbf{B}(z_j)^T \quad (3-3)$$

它是一个 $L \times L$ 阶矩阵, 由式(3-1)~(3-3), $\boldsymbol{\beta}$ 的最小二乘估计的方差等于

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{M}^{-1}(\xi_N) \sigma^2 \quad (3-4)$$

由式(3-4)可见, 设计 ξ_N 的优劣可用 $\text{Var}(\hat{\boldsymbol{\beta}})$, 即用信息矩阵 $\mathbf{M}(\xi_N)$ 来衡量。

定理 3.1 信息矩阵具有下列性质:

(1) 信息矩阵 $\mathbf{M}(\xi_N)$ 是半正定的;

(2) 若设计 ξ 是离散型概率分布 $\xi = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$,

其中, z_i 和 p_i ($i=1, \cdots, n$) 分别为设计 ξ 的谱点 (即试验点) 及其概率分布。若

ξ 的谱点数 $n < L$ (L 是模型中未知参数向量 β 的维数), 则 $M(\xi)$ 是退化矩阵;

(3) 设 $\Xi = \{\xi: \xi \text{ 是 } \mathcal{X} \text{ 上的概率分布}\}$ 是设计全体, $M(\Xi) = \{M(\xi): \xi \in \Xi\}$ 是模型(3-1)的一切设计所对应的信息矩阵的全体, 则 $M(\Xi)$ 是一个闭凸集;

(4) 对任一设计 ξ , 存在离散设计

$$\tilde{\xi} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}, \quad p_i \geq 0, \quad i=1, 2, \cdots, n, \quad \sum_{i=1}^n p_i = 1$$

使得 $M(\xi) = M(\tilde{\xi}) = \sum_{i=1}^n p_i \mathbf{B}(x_i) \mathbf{B}(x_i)^T$ 。其中, $n \leq \frac{L(L+1)}{2} + 1$ 。

证明:

对性质 (1)、(2) 可由信息矩阵的定义式(3-3)直接证明。

对性质 (3), 设 ξ_1 和 ξ_2 是两个设计, 它们的信息矩阵的积分形式分别为 $M(\xi_1) = \int_{\mathcal{X}} \mathbf{B}(x) \mathbf{B}(x)^T d\xi_1$ 和 $M(\xi_2) = \int_{\mathcal{X}} \mathbf{B}(x) \mathbf{B}(x)^T d\xi_2$, 若 $0 \leq a \leq 1$, 不难证明, 设计 $a\xi_1 + (1-a)\xi_2$ 的信息矩阵为

$$\begin{aligned} M[a\xi_1 + (1-a)\xi_2] &= \int_{\mathcal{X}} \mathbf{B}(x) \mathbf{B}(x)^T d[a\xi_1 + (1-a)\xi_2] \\ &= a \int_{\mathcal{X}} \mathbf{B}(x) \mathbf{B}(x)^T d\xi_1 + (1-a) \int_{\mathcal{X}} \mathbf{B}(x) \mathbf{B}(x)^T d\xi_2 \\ &= aM(\xi_1) + (1-a)M(\xi_2) \end{aligned}$$

所以 $M(\Xi)$ 是一个凸集, 又已假定 \mathcal{X} 是闭集且 $\mathbf{B}(x)$ 是 \mathcal{X} 上的连续函数向量, 所以 $M(\Xi)$ 是一个闭凸集。

对性质 (4), $M(\xi)$ 是由主对角线及其上方的 $k(k+1)/2$ 个元素决定的对称矩阵。因此, 可将 $M(\xi)$ 看作 $k(k+1)/2$ 维空间中的一点, 由于 $M(\Xi)$ 是一个闭凸集, 根据卡拉西奥多里 (Carathéodory) 定理^[158], 性质 (4) 成立。

证毕

定理 3.1 中的性质 (4) 为确定靶场试验设计的最大谱点数提供了依据。可见, 靶场试验设计的主要问题由两部分组成:

(1) 系统性能响应的因子效应分析。目的是得到各影响因素对武器性能产生的影响, 以确定基函数向量 $\mathbf{B}(x)$, 并构造试验设计信息矩阵 $M(\xi)$;

(2) 现场试验的最优设计。根据构造的信息矩阵 $M(\xi)$, 在一定的最优准则下进行迭代计算, 获得覆盖所有重要因子效应项的最优设计方案。

试验设计的优良性准则是信息矩阵的函数 $\Phi[M(\xi)]$, 而信息矩阵可由因子效应构造的基函数向量得到。因此, 因子效应分析是靶场试验最优设计的基础。以两因子 (x_1, x_2) 试验为例, 采用二次回归模型时, 所有的因子效应项包括 $1, x_1, x_2, x_1x_2, x_1^2, x_2^2$ 。显然, 根据验前数据的因子效应分析构造合理的基函数向量 (例如 $\mathbf{B}(x) = [1, x_1, x_2, x_1x_2]^T$), 可以得到现场试验的最优设计。

3.3 贝叶斯正则化因子效应分析

靶场试验中,由于现场飞行试验成本极高,因此希望用尽量少的试验样本考核最有价值的因子空间区域,使获得的试验样本数据能够较好的反映装备在真实作战环境中的性能。根据试验设计信息矩阵的性质,确定试验设计的谱点数与优良性均需要合理的因子效应分析。可见,因子效应分析在靶场试验设计中具有非常重要的地位。正则化是解决反问题研究中不适定性的有力工具,为靶场试验设计中的因子效应分析提供了一种较好的途径。

3.3.1 正则化方法

正则化方法^[159] (Regularized methods) 主要用于解反问题中的不适定问题 (Ill-posed problem)。不适定问题是指方程的解不连续依赖于给定的数据,当数据出现微小变化时将引起解的很大变化。试验设计模型(3-1)和(3-2)中的不适定问题可以表示为:

$$\mathbf{F}^T \boldsymbol{\beta} = \mathbf{y} = \tilde{\mathbf{y}} + \boldsymbol{\varepsilon} \quad (3-5)$$

其中, $\mathbf{F}^T \in \mathbb{R}^{N \times L}$, $\mathbf{y} \in \mathbb{R}^N$, $\boldsymbol{\beta} \in \mathbb{R}^L$ 为待求解的系数向量, 向量 $\boldsymbol{\varepsilon}$ 为真值 $\tilde{\mathbf{y}}$ 的扰动误差向量。当可用的样本点数 N 较小时, 式(3-5)通常是矩阵 \mathbf{F}^T 为非满秩且条件数很大的病态矩阵。应用传统方法求解时, \mathbf{F}^T 或 \mathbf{y} 中存在的微小误差会使(3-5)的解 $\boldsymbol{\beta}$ 发生振荡, 从而与真实解相距甚远, 甚至对于实际应用来说毫无意义。基于正则化方法求解式(3-5)的系数向量的一般框架可表示为

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \{ \mathfrak{I}(\mathbf{y}, \boldsymbol{\beta}) + \lambda \cdot \mathfrak{R}(\boldsymbol{\beta}) \} \quad (3-6)$$

其中, $\mathfrak{I}(\mathbf{y}, \boldsymbol{\beta})$ 为衡量回归模型预测值与真实值之间差异的保真项; $\mathfrak{R}(\boldsymbol{\beta})$ 为对解进行约束的正则项 (惩罚项); λ 为均衡保真项和正则项的正则化参数。 $\lambda \cdot \mathfrak{R}(\boldsymbol{\beta})$ 将不适定问题(3-5)转化为适定性问题。当 $\mathfrak{I}(\mathbf{y}, \boldsymbol{\beta})$ 取最小二乘数据保真项, $\lambda \cdot \mathfrak{R}(\boldsymbol{\beta})$ 取 Tikhonov 正则项时, 式(3-6)可写为

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \left\{ \left\| \mathbf{y} - \mathbf{F}^T \boldsymbol{\beta} \right\|^2 + \lambda \left\| \boldsymbol{\beta} \right\|^2 \right\} \quad (3-7)$$

Tikhonov 正则化方法把正则化泛函的极小元作为式(3-5)的正则化近似解。该正则化解一方面使得 $\left\| \mathbf{y} - \mathbf{F}^T \boldsymbol{\beta} \right\|$ 较小, 另一方面通过惩罚项 $\lambda \left\| \boldsymbol{\beta} \right\|^2$ 来保证解的稳健性 (从而正则化解)。从优化理论的角度来看, 该二次优化问题可以看成是下述两个带约束的优化问题的惩罚函数方法:

- (1) 给定 $\rho_0 > 0$, 在解的约束条件 $\left\| \boldsymbol{\beta} \right\| < \rho_0$ 下极小化偏差函数 $\left\| \mathbf{y} - \mathbf{F}^T \boldsymbol{\beta} \right\|$;
- (2) 给定 $\Delta > 0$, 在偏差的约束条件 $\left\| \mathbf{y} - \mathbf{F}^T \boldsymbol{\beta} \right\| \leq \Delta$ 下极小化解的模 $\left\| \boldsymbol{\beta} \right\|$ 。

对于式(3-6)给出的正则化优化求解框架,通过选择不同的正则项或者惩罚函数可以得到不同的解,目前常用的正则化方法有岭回归^[160]、Lasso^[161]、弹性网^[162]、SCAD^[163]、MCP 惩罚^[164]、SICA 惩罚函数^[165]以及 Bridge 惩罚^[166]等。Fan 等^[167]对此进行了详细讨论。

正则化符合奥卡姆剃刀(Occam's razor)原理:在所有可能选择的模型中,能够很好的解释已知数据并且十分简单的才是最好的模型,也就是应该选择的模型。从贝叶斯估计的角度来看,正则化项对应模型的先验概率,即假设复杂的模型有较小的先验概率,而简单的模型有较大的先验概率。

3.3.2 贝叶斯正则化因子效应分析

Pareto 原理^[168]指出 80%的效能取决于 20%的因素。因此,合理确定重要的因子效应项对靶场试验设计具有重要意义。因子效应分析本质上是一个变量选择问题,传统的变量选取准则(如 AIC、BIC 和 RIC 等)存在稳定性低、计算负担重等缺点。尽管目前提出的一些新算法(如向前法、后退法、逐步回归法等)减少了计算的复杂度,然而,由于这些方法普遍存在求解稳定性低的缺点,当验前数据存在较大的测量噪声时,得到的变量选择结果往往很不稳定,不利于指导后续的靶场试验设计。

取试验设计模型(3-1)中的基函数向量 $\mathbf{B}(\mathbf{x})$ 包含所有可能的因子效应项,则 $\boldsymbol{\beta}$ 的真实解是稀疏的(Sparse),即 $\boldsymbol{\beta}$ 的一部分分量是 0。显然, $\boldsymbol{\beta}$ 中分量为 0 的部分对应的因子效应项是不重要的,在后续试验设计中不必太过关注;相反, $\boldsymbol{\beta}$ 中分量不为 0 的部分对应的因子效应项对系统性能有重要影响,在后续现场试验中应重点关注。应用最小二乘法求解系数向量 $\boldsymbol{\beta}$,有

$$\boldsymbol{\beta} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{F}^T \boldsymbol{\beta}\|^2 \quad (3-8)$$

上式可转换为求解 Euler-Lagrange 方程 $\mathbf{F}\mathbf{F}^T \boldsymbol{\beta} = \mathbf{F}\mathbf{y}$ 。当 $\mathbf{F}\mathbf{F}^T$ 奇异时,方程存在不唯一解。当 $\mathbf{F}\mathbf{F}^T$ 非奇异时, $\mathbf{F}\mathbf{F}^T$ 是 L 阶非负定矩阵,存在 L 个非负奇异值,记为 $s_1 \geq s_2 \geq \cdots \geq s_L \geq 0$ 。对矩阵 \mathbf{F}^T 做奇异值分解,有 $\mathbf{F}^T = \mathbf{U} \text{diag}(s_i) \mathbf{V}^T$,其中 \mathbf{U} , \mathbf{V} 分别为 N 阶和 L 阶单位正交矩阵。方程(3-8)的最小二乘解可表示为

$$\hat{\boldsymbol{\beta}} = \mathbf{V} \text{diag}(s_i^{-1}) \mathbf{U}^T \mathbf{y} = \sum_{i=1}^L \frac{\mathbf{u}_i^T \mathbf{y}}{s_i} \mathbf{v}_i$$

其中, \mathbf{u}_i , \mathbf{v}_i 分别为 \mathbf{U} , \mathbf{V} 的列向量。矩阵 \mathbf{F}^T 高度病态时,较小的奇异值趋于零,对应的特征向量具有高度震荡性,因此 \mathbf{F}^T 和 \mathbf{y} 的微小变化(如误差、噪声等)就会导致解的非稳定性变化。

正则化因子效应分析通过弱化较小的奇异值对最小二乘解的灾难性影响,

将原来最小二乘解中的不适定问题转化为适定问题,并保证与原问题真实解的保真度,其实现方程如下:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^L} \left\{ \sum_{i=1}^N \|y_i - \mathbf{B}(z_i)^T \boldsymbol{\beta}\|^2 + \sum_{j=1}^L p_{\lambda}(|\beta_j|) \right\} \quad (3-9)$$

式(3-9)中, $\lambda > 0$ 是正则化参数,用于控制数据拟合程度和解 $\boldsymbol{\beta}$ 大小之间的平衡; $p_{\lambda}(\cdot)$ 是惩罚函数,表示对解 $\boldsymbol{\beta}$ 的约束。

应用式(3-9)进行因子效应分析时,选择合理的正则化参数 λ 是进行因子效应分析的关键。当 λ 较大时,赋予惩罚函数的权较大,可以保证 $\boldsymbol{\beta}$ 较小,此时数据拟合误差会变大;反之,若 λ 取得很小,式(3-9)将非常接近原问题(3-5),此时尽管拟合误差很小,但得到解容易发生振荡而没有实际物理意义。在经典统计学框架下,通常采用交叉验证法 (Cross-Validation)、广义交叉验证法以及基于 Stein 无偏风险估计 (Unbiased Risk Estimate) 的思想选择正则化参数 λ , 实现较复杂且计算量非常大,往往难以得到一个合理值,由此得到的因子效应分析结果也并不稳定。

在贝叶斯理论框架下, Tibshirani 给出了 Lasso 的贝叶斯解释: 假设模型(3-5)中回归系数向量 $\boldsymbol{\beta}$ 的先验为 Laplace 分布 (也被称为独立双指数分布), 具有如下的结构形式,

$$\pi(\boldsymbol{\beta}) = \prod_{i=1}^p \frac{\lambda}{2} e^{-\lambda|\beta_i|} \quad (3-10)$$

则由 Lasso 得到的回归系数估计结果与其极大后验估计一致。采用式(3-10)给出的 Laplace 先验分布的优势之一在于它可以被表达为尺度混合的正态分布, 这种正态分布的方差服从独立的指数分布。由此, 通过引入潜在的尺度变量, 应用层次化贝叶斯方法可以得到回归系数的 Lasso 估计。当 σ^2 已知时, 对 $\boldsymbol{\beta}$ 可采用如下的条件先验概率

$$\pi(\boldsymbol{\beta} | \sigma^2) = \prod_{i=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_i|/\sqrt{\sigma^2}} \quad (3-11)$$

式(3-11)给出的条件先验用于贝叶斯计算时能够加快 Gibbs 抽样的收敛速度, 并且还能够保证导出单峰的后验分布。正则化参数 λ 可通过一个适当的超先验引入贝叶斯 Lasso 的 Gibbs 抽样中。将 λ 纳入 Gibbs 抽样迭代, 得到的回归系数估计不再依赖于一个固定的 λ 值, 而是将所有的 λ 边际化, 一定程度上提高了估计的稳健性。

假设 λ 服从 $\text{Gamma}(a, b)$ 的先验分布, σ^2 服从无信息先验, 即 $\pi(\sigma^2) \propto 1/\sigma^2$, 本文提出如下的贝叶斯正则化因子效应分析模型

$$\begin{cases} y|X, \beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n) \\ \beta|\sigma^2 \sim \prod_{i=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_i|/\sqrt{\sigma^2}} \\ \lambda \sim \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\ \sigma^2 \sim \frac{1}{\sigma^2} \end{cases} \quad (3-12)$$

式(3-12)中, 取 Gamma 分布参数 $(a, b)=(1, 0)$ 时, 构造的正则化参数 λ 的先验分布等价于均匀分布 $\pi(\lambda) \propto 1$ 。将其带入式(3-12)并求 σ^2 的边缘分布, 可知 σ^2 的后验估计为通常的极大似然估计, 且支持 β 后验估计的稀疏性。

对仿真数据或相似型号装备的历史试验数据, 应用上述贝叶斯正则化因子效应分析方法, 可以得到较稳定的因子效应信息, 根据因子效应信息构造试验设计信息矩阵, 可以实现一定优良性准则下的靶场试验最优设计。这样就可以用最少的谱点数覆盖需要重点考核的试验因子空间。

3.4 基于贝叶斯正则化的靶场试验设计方法

仿真与现场试验是靶场采用的两种主要试验手段。仿真一般先于现场试验进行, 为现场试验提供验前信息。现场试验一般为破坏性试验, 试验成本极高, 只能获得少量的试验样本。因此, 利用仿真和其他相似型号产品的验前数据进行因子效应分析, 构造合理的试验设计信息矩阵, 是实现靶场试验最优设计的基础。贝叶斯正则化方法提供了一种合理利用验前信息构造试验设计信息矩阵的途径, 非常适合于靶场试验设计。

3.4.1 考虑因子效应信息的靶场试验设计信息矩阵

根据仿真试验数据应用贝叶斯正则化因子效应分析模型(3-12)可得到 β 的后验稀疏解, 其对应的因子效应项线性无关。因此, 由因子效应分析结果可以得到试验设计模型(3-1)的基函数向量 $\mathbf{B}(\mathbf{x})$, 进而由式(3-3)可以得到试验设计的信息矩阵。设 x_1 和 x_2 为试验的因子, $\mathcal{X}=[a, b]$ 表示二维因子空间, 试验设计模型为二元二次回归模型

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 \quad (3-13)$$

考虑所有因子效应项 $1, x_1, x_2, x_1 x_2, x_1^2, x_2^2$ 时, $\mathbf{B}(\mathbf{x})=[1, x_1, x_2, x_1 x_2, x_1^2, x_2^2]^T$, $\beta=[\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_{11}, \beta_{22}]^T$, 有

$$B(x)B(x)^T = \begin{bmatrix} 1 & x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^2x_2 & x_1^3 & x_1x_2^2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 & x_1^2x_2 & x_2^3 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 & x_1^3x_2 & x_1x_2^3 \\ x_1^2 & x_1^3 & x_1^2x_2 & x_1^3x_2 & x_1^4 & x_1^2x_2^2 \\ x_2^2 & x_1x_2^2 & x_2^3 & x_1x_2^3 & x_1^2x_2^2 & x_2^4 \end{bmatrix} \quad (3-14)$$

只考虑因子效应项 $1, x_1, x_2, x_1x_2$ 时, $B(x)=[1, x_1, x_2, x_1x_2]^T$, $\beta=[\beta_0, \beta_1, \beta_2, \beta_{12}]^T$, 有

$$B(x)B(x)^T = \begin{bmatrix} 1 & x_1 & x_2 & x_1x_2 \\ x_1 & x_1^2 & x_1x_2 & x_1^2x_2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 \end{bmatrix} \quad (3-15)$$

将式(3-14)和(3-15)带入式(3-3)中, 可以得到考虑不同因子效应信息的试验设计信息矩阵。由信息矩阵的性质 (2) 可知, 最优设计的谱点数应不小于因子效应项的数目。由此可见, 靶场试验的最优设计点数应根据因子效应分析结果确定, 由因子效应分析结果构成的基函数向量 $B(x)$ 为靶场试验设计的信息矩阵构造提供了基础。

假设由验前数据的贝叶斯正则化因子效应分析结果得到 k 项确定的因子效应和 h 项可能的因子效应, h 项可能因子效应的存在概率为 η 。定义由 k 项确定的因子效应得到的信息矩阵为 $M_k(\xi)$, 由所有 $k+h$ 项可能的因子效应得到的信息矩阵为 $M_{k+h}(\xi)$ 。为了覆盖所有可能的因子效应项, 定义考虑因子效应信息的靶场试验设计信息矩阵

$$M(\xi) = (1-\eta)M_k(\xi) + \eta M_{k+h}(\xi) \quad (3-16)$$

其行列式 $\det(M(\xi)) = (1-\eta)\det(M_k(\xi_k)) + \eta\det(M_{k+h}(\xi_{k+h}))$ 。注意, 式(3-16)中的加号并非通常意义上的矩阵求和, 上述试验设计信息矩阵是为了得到一种带有附加点的靶场试验方案, 以兼顾可能的因子效应项。取式(3-15)对应确定的因子效应, 式(3-13)中 x_1^2 和 x_2^2 为可能因子效应时, 由式(3-16)可以得到考虑因子效应的靶场试验设计信息矩阵。其中,

$$M_k(\xi) = \sum_{j=1}^k B(z_j)B(z_j)^T = \sum_{j=1}^k \begin{bmatrix} 1 & x_1 & x_2 & x_1x_2 \\ x_1 & x_1^2 & x_1x_2 & x_1^2x_2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 \end{bmatrix}_{z_j} \quad (3-17)$$

$$M_{k+h}(\xi) = \sum_{j=1}^{k+h} B(z_j) B(z_j)^T = \sum_{j=1}^{k+h} \begin{bmatrix} 1 & x_1 & x_2 & x_1 x_2 & x_1^2 & x_2^2 \\ x_1 & x_1^2 & x_1 x_2 & x_1^2 x_2 & x_1^3 & x_1 x_2^2 \\ x_2 & x_1 x_2 & x_2^2 & x_1 x_2^2 & x_1^2 x_2 & x_2^3 \\ x_1 x_2 & x_1^2 x_2 & x_1 x_2^2 & x_1^2 x_2^2 & x_1^3 x_2 & x_1 x_2^3 \\ x_1^2 & x_1^3 & x_1^2 x_2 & x_1^3 x_2 & x_1^4 & x_1^2 x_2^2 \\ x_2^2 & x_1 x_2^2 & x_2^3 & x_1 x_2^3 & x_1^2 x_2^2 & x_2^4 \end{bmatrix}_{z_j} \quad (3-18)$$

式(3-16)~(3-18)给出了取二元二次回归模型时,考虑因子效应信息的靶场试验设计信息矩阵。在此基础上,应用一定的准则函数,可以得到一个 $k+h$ 点的最优设计方案。其中,前 k 个主要设计点对应确定的因子效应,后 h 个附加设计点对应可能的因子效应,附加设计点可根据主要设计点的试验结果进一步确定是否实施。可见,考虑因子效应信息的靶场试验设计信息矩阵不仅兼顾了可能的因子效应项,而且能够避免靶场试验资源的浪费。

确定因子效应的存在概率是构造考虑因子效应信息的靶场试验设计信息矩阵的关键,式(3-12)的贝叶斯正则化因子效应分析模型提供了因子效应系数的后验计算方法,但无法直接得到不同因子效应的存在概率排序。应用式(3-16)的信息矩阵进行靶场试验设计时,还需进一步研究不同因子效应的存在概率,给出贝叶斯正则化试验设计信息矩阵构造方法。

3.4.2 贝叶斯正则化试验设计信息矩阵构造

定义 $B_j(\mathbf{x}), j=1, 2, \dots, p$ 为全部 p 项因子效应,构成试验的基函数向量 $\mathbf{B}(\mathbf{x})$; $\boldsymbol{\gamma}=[\gamma_1, \dots, \gamma_p]^T$ 为相应的二元变量指示器; $\boldsymbol{\beta}$ 为因子效应系数向量。如果 $\boldsymbol{\gamma}$ 的分量 $\gamma_j=1$,则相应的因子效应项 $B_j(\mathbf{x})$ 包含在基函数向量 $\mathbf{B}(\mathbf{x})$ 中;如果 $\gamma_j=0$,则相应的因子效应项从 $\mathbf{B}(\mathbf{x})$ 中剔除。

根据式(3-12)给出的贝叶斯正则化因子效应分析模型,应用 MCMC 方法动态模拟 $\boldsymbol{\beta}$ 后验分布的马尔科夫链,可以有效地估计各变量指示器与所有可能模型的后验概率。后验推断时,二元变量指示器 $\boldsymbol{\gamma}$ 与十进制模型指示器 $M(\boldsymbol{\gamma})$ 的转换关系如下

$$M(\boldsymbol{\gamma}) = \sum_{j=k}^p \gamma_j 2^{j-k} \quad (3-19)$$

若式(3-19)中不包含常数项,则 $k=1$;否则 $k=0$ 。

根据 MCMC 方法得到变量指示器 $\boldsymbol{\gamma}$ 与模型指示器 $M(\boldsymbol{\gamma})$ 的后验分布抽样,可以估算每个变量指示器 $\boldsymbol{\gamma}$ 的后验概率 $f(\boldsymbol{\gamma}=1|\mathbf{y})$ 与模型指示器的后验概率 $f(M|\mathbf{y})$,其计算公式如下

$$\hat{f}(\gamma=1|\mathbf{y}) = \frac{1}{T_l - B_n} \sum_{t=B_n+1}^{T_l} I(\gamma^{(t)}=1) \quad (3-20)$$

$$\hat{f}(M|\mathbf{y}) = \frac{1}{T_l - B_n} \sum_{t=B_n+1}^{T_l} I(M^{(t)}=M) \quad (3-21)$$

式(3-20)和(3-21)中 T_l 为后验分布抽样的全部迭代次数, B_n 为抽样的燃烧(Burn-in)值; $\gamma^{(t)}$ 和 $M^{(t)}$ 分别为变量指示器与模型指示器对应的后验样本在第 t 次的迭代结果。对式(3-21)得到的后验概率最大的模型包含的因子效应项, 由式(3-20)的后验统计得到各自的存在概率排序

$$\Pr(B_1^{(b)}(\mathbf{x})) \geq \dots \geq \Pr(B_k^{(b)}(\mathbf{x})) > \eta \geq \dots \geq \Pr(B_{k+h}^{(b)}(\mathbf{x})) \quad (3-22)$$

式(3-22)中, \mathbf{x} 表示因子向量, 上标(b)对应由贝叶斯正则化分析得到的参数向量 $\boldsymbol{\beta}$ 的估计量 \mathbf{b} 。式(3-22)中存在概率大于门限 η 的前 k 项因子效应为确定因子效应, 存在概率不大于门限 η 的后 h 项为可能的因子效应。

由上述分析可见, 基于贝叶斯正则化因子效应分析模型对验前数据进行分析可以得到模型指示器的后验概率 $f(M|\mathbf{y})$ 以及变量指示器 γ 的后验概率 $f(\gamma=1|\mathbf{y})$, 确定出显著性因子并计算各因子效应项的存在概率排序。由此给出如下的贝叶斯正则化试验设计信息矩阵构造算法:

算法 3.1

- (1) 设定参数的贝叶斯先验。按照式(3-12)给出的贝叶斯正则化因子效应分析模型设定各参数的先验分布, 应用 MCMC 方法对因子效应系数向量 $\boldsymbol{\beta}$ 进行参数估计与收敛性诊断;
- (2) 确定显著性因子效应项。根据各变量指示器的后验样本, 由式(3-19)和(3-21)计算所有候选模型的后验概率, 取后验概率最大的模型包含的因子效应为显著性因子效应;
- (3) 计算因子效应项的存在概率。根据各变量指示器的后验样本, 利用式(3-20)计算各显著性因子效应项的后验概率, 并得到如式(3-22)所示的因子效应存在概率排序;
- (4) 贝叶斯正则化试验设计信息矩阵构造。由第(3)步中存在概率大于门限 η 的因子效应项得到 $M_k(\xi)$, 由所有显著性因子效应项得到 $M_{k+h}(\xi)$ 。根据式(3-16)构造信息矩阵 $M(\xi)$ 。

算法 3.1 给出了应用贝叶斯正则化构造考虑因子效应信息的靶场试验设计信息矩阵的方法。设计的优良性准则是信息矩阵的函数 $\Phi[M(\xi)]$, 因此, 对于靶场试验中的现场试验设计, 合理利用验前数据提供的因子效应信息确定试验设计信息矩阵 $M(\xi)$, 即可实现一定准则下的最优设计。

3.4.3 靶场试验设计方法

最优回归设计 (Optimum regression design) 主要讨论如何选择试验的设计点, 使已知回归模型的系数能够获得在一定准则下的最优估计。常见的最优设计准则有 D-最优, A-最优, E-最优等, 其中, D-最优准则使用范围最广且实现也较为容易。对于因子空间 \mathcal{X} 上的设计全体 $\Xi = \{\xi: \xi \text{ 是 } \mathcal{X} \text{ 上的任一设计}\}$, 若一个设计 $\xi^* \in \Xi$ 满足以下条件:

$$\det M(\xi^*) = \max_{\xi \in \Xi} \det M(\xi) \quad (3-23)$$

则称设计 ξ^* 为 D-最优设计。取 L 项因子效应构造信息矩阵 $M(\xi)$ 时, D 优良性的直观意义在于: 若在设计 ξ 下获得输出响应的观测值, 可由最小二乘法求得模型(3-1)中参数向量 β 的估计量 $b = (b_1, b_2, \dots, b_L)^T$, 由 D-最优设计得到的估计量 b 的密集椭球体积最小。

在设计 ξ 下, 由观测值用最小二乘估计确定的回归方程为 $\hat{y}(x) = \beta^T B(x)$ 。应用此回归方程进行预测时, 在信息矩阵 $M(\xi)$ 非退化的情况下, 点 x 处预测值 $\hat{y}(x)$ 的方差可表示为

$$\text{Var}[\hat{y}(x)] = B(x)^T \text{Var}(\beta) B(x) \propto B(x)^T M^{-1}(\xi) B(x) = d(x, \xi) \quad (3-24)$$

选择 D-最优准则进行试验设计时, 为了得到对应确定因子效应项的 k 点主要部分和对应可能因子效应项的 h 点附加部分, 基于贝叶斯正则化的试验设计信息矩阵式(3-16), 定义点 x 处的加权预测方差为

$$\hat{d}(x, \xi_{k+h}) = (1-\eta)d(x, \xi_k) \Big|_{M_k(\xi_k)} + \eta d(x, \xi_L) \Big|_{M_L(\xi_L)} \quad (3-25)$$

其中, ξ_L 表示包含 $L=k+h$ 个谱点的设计, ξ_k 为只包含 ξ_L 中前 k 个谱点的设计。根据式(3-25)应用如下迭代搜索算法能够得到一种兼顾所有可能因子效应的 D-最优回归设计。

算法 3.2

- (1) 设置 $j=0$, 任意给定一个非退化的初始设计 ξ_j ;
- (2) 由式(3-16)计算 ξ_j 的信息矩阵 $M(\xi_j)$;
- (3) 由式(3-25)求在设计 ξ_j 下, 最小二乘估计的预测值在因子区域中的方差最大值点 z_j , 即 $\hat{d}(z_j, \xi_j) = \max_z \hat{d}(z, \xi_j) = d$;
- (4) 构造新的设计 $\xi_{j+1} = (1-\alpha_j)\xi_j + \alpha_j\xi_{z_j}$, 其中, $\alpha_j = \delta_j L / (\delta_j + L - 1)$,
 $\delta_j = \hat{d}(z_j, \xi_j) - L$;
- (5) 计算设计 ξ_{j+1} 的信息矩阵 $M(\xi_{j+1})$;

(6) 若满足条件 $\alpha_j < \gamma_1$, $\frac{|M(\xi_{j+1}) - M(\xi_j)|}{M(\xi_{j+1})} < \gamma_2$, $\frac{|d(z_j, \xi_j) - L|}{L} < \gamma_3$ 之一,

终止; 否则, 设置 $j=j+1$, 转 (3)。

现场试验实施过程中, 首先安排由算法 3.2 得到的 D-最优设计中主要部分试验点上的试验 (即最优设计的前 k 个谱点); 然后根据获得的现场试验数据进一步分析影响系统性能的因子效应, 以判断可能的 h 项因子效应能否排除; 最后根据分析判断结果, 决定是否需要继续进行附加部分试验点上的试验。采用这种具有备选附加试验点的设计方案, 能够保证现场试验对重要因子效应的全面覆盖, 同时避免试验点过多造成的资源浪费。

基于贝叶斯正则化的靶场试验设计方法, 在仿真试验设计阶段, 发挥仿真试验消耗低、实施灵活、容易得到大量试验样本的优势, 采用空间填充设计充分挖掘各个因子对系统性能的影响, 应用贝叶斯正则化方法实现因子效应信息的稳健分析。在现场试验设计阶段, 根据由仿真试验数据获得的因子效应信息构造现场试验最优设计的信息矩阵, 由算法 3.2 得到具有备选附加试验点的设计方案来兼顾可能的因子效应项, 以提高靶场试验设计效益。

3.5 示例分析

射表是炮兵武器系统不可缺少的重要文件, 包含了射击所需的全部信息, 利用射表可以确定炮弹命中目标所需的火炮装定诸元。现代射表大多采用射弹飞行试验与仿真计算相结合的方法编制而成。目前, 射表试验通常采用固定射击点位、标准气象条件下的单因子区组设计方案, 并取射角为试验设计的单一因子。由此得到的射表在实际使用中存在问题, 如 3000m 海拔处的火炮射程可比零海拔处大 30%~50%。因此, 如何综合考虑影响射表的因子, 分析它们之间能够产生的各种因子效应, 进而得到科学合理的射弹飞行试验优化设计方案是国家靶场亟需研究的重点问题。

外弹道方程是射表编拟的基础, 典型的外弹道模型有 3D 简单质点方程、4D 修正质点弹道方程和 6D 刚体弹道方程。以 6D 刚体弹道方程为例, 在地面坐标系 (ECS) 中建立的弹体质心运动学及动力学方程为

$$\begin{aligned} \frac{dv}{dt} = \frac{\rho A_s v_r}{2m_q} \left[-C_x v_{rx3} + C_y \frac{\delta_r}{\sin \delta_r} \left(v_r \cos \delta_1 \cos \delta_2 - \frac{v_{rx1} v_{rx3}}{v_r} \right) + \right. \\ \left. C_z \frac{\delta_r}{\sin \delta_r} (v_{rz3} \cos \delta_2 \sin \delta_1 - v_{ry3} \sin \delta_2) \right] - g \cos \psi \sin \theta \end{aligned} \quad (3-26)$$

$$\frac{d\theta}{dt} = \frac{\rho A_s v_r}{2m_q v \cos \psi} \left[-C_x v_{ry3} + C'_y \frac{\delta_r}{\sin \delta_r} \left(v_r \cos \delta_1 \cos \delta_2 - \frac{v_{rx1} v_{ry3}}{v_r} \right) + \right. \quad (3-27)$$

$$\left. C'_z \frac{\delta_r}{\sin \delta_r} (v_{rx3} \sin \delta_2 - v_{rz3} \cos \delta_1 \cos \delta_2) \right] - \frac{g \cos \theta}{v \cos \psi}$$

$$\frac{d\psi}{dt} = \frac{\rho A_s v_r}{2m_q v} \left[-C_x v_{rz3} + C'_y \frac{\delta_r}{\sin \delta_r} \left(v_r \sin \delta_2 - \frac{v_{rx1} v_{rz3}}{v_r} \right) + \right. \quad (3-28)$$

$$\left. C'_z \frac{\delta_r \cos \delta_2}{\sin \delta_r} (v_{ry3} \cos \delta_1 - v_{rx3} \sin \delta_1) \right] - \frac{g \sin \psi \sin \theta}{v}$$

$$\frac{d\omega_{x1}}{dt} = -\frac{\rho A_s l^2 m'_{xz} v_r \omega_{x1}}{2I_{\xi\xi}} \quad (3-29)$$

$$\frac{d\omega_{y1}}{dt} = \frac{1}{I_{\eta\eta}} \left\{ \frac{\rho A_s l v_r}{2} \left[\frac{\delta_r}{\sin \delta_r} v_{ry1} (m'_z + m'_y) - l m'_{zz} \omega_{y1} \right] + \right. \quad (3-30)$$

$$I_{\xi\xi} \omega_{z1}^2 \tan \varphi_2 - I_{\xi\xi} \omega_{z1} \dot{\gamma} \}$$

$$\frac{d\omega_{z1}}{dt} = \frac{1}{I_{\zeta\zeta}} \left\{ \frac{\rho}{2} A_s l v_r \left[\frac{\delta_r}{\sin \delta_r} v_{rz1} (-m'_z + m'_y) - l m'_{zz} \omega_{z1} \right] - \right. \quad (3-31)$$

$$I_{\eta\eta} \omega_{y1} \omega_{z1} \tan \varphi_2 + I_{\xi\xi} \omega_{y1} \dot{\gamma} \}$$

$$\frac{d\varphi_1}{dt} = \frac{\omega_{z1}}{\cos \varphi_2} \quad (3-32)$$

$$\frac{d\varphi_2}{dt} = -\omega_{y1} \quad (3-33)$$

$$\frac{d\gamma}{dt} = \omega_{x1} - \omega_{z1} \tan \varphi_2 \quad (3-34)$$

$$\frac{dX}{dt} = v \cos \psi \cos \theta \quad (3-35)$$

$$\frac{dY}{dt} = v \cos \psi \sin \theta \quad (3-36)$$

$$\frac{dZ}{dt} = v \sin \psi \quad (3-37)$$

式(3-26)~(3-37)中各变量符号的含义见表 3-1, 且 X 、 Y 、 Z 表示弹丸在 ECS 中的三轴分量。根据以上计算模型, 结合相应的坐标几何关系式、辅助关系式、气象计算公式及边界条件, 可进行火炮射击精度仿真。根据仿真数据分析影响射击精度的因子效应, 可为射弹飞行试验设计提供依据。其中, 8 个影响弹丸落点的因子在其均值 $\mu_{(\cdot)}$ 的 $\pm 3\sigma_{(\cdot)}$ 范围内分为 24 个水平, 如表 3-2 所示。

表 3-1 6D 刚体弹道方程的变量符号

Table3-1 Variables of 6D rigid body trajectory equation

变量	物理意义	变量	物理意义
v	弹丸速度	A_s	弹丸特征面积
θ	弹道倾角	l	弹体长度
ψ	弹道偏角	δ_γ	相对攻角
ω_{x1}	弹丸滚转角速度	δ_1	纵向攻角
ω_{y1}	弹轴摆动角速度的纵向分量	δ_2	横向攻角
ω_{z1}	弹轴摆动角速度的横向分量	C_x	弹丸阻力系数
φ_1	弹轴摆动角的纵向分量	C'_y	升力系数导数
φ_2	弹轴摆动角的横向分量	C'_z	马格努斯力系数导数
γ	滚转角	m'_y	马格努斯力力矩系数导数
v_r	弹丸相对空气的运动速度	m'_z	翻转力矩系数导数
v_{rx1}	v_r 在弹体坐标系中的 x 轴分量	m'_{xz}	极阻尼力矩系数导数
v_{ry1}	v_r 在弹体坐标系中的 y 轴分量	m'_{zz}	摆动阻尼力矩系数导数
v_{rz1}	v_r 在弹体坐标系中的 z 轴分量	$I_{\zeta\zeta}$	弹丸横向摆动转动惯量
v_{rx3}	v_r 在速度坐标系中的 x 轴分量	$I_{\xi\xi}$	弹丸滚转转动惯量
v_{ry3}	v_r 在速度坐标系中的 y 轴分量	$I_{\eta\eta}$	弹丸纵向摆动转动惯量
v_{rz3}	v_r 在速度坐标系中的 z 轴分量	ρ	空气密度
m_q	弹丸质量	g	重力加速度

表 3-2 影响弹丸落点的 8 个因子及其取值范围

Table3-2 8 factors on projectile impact and their domain

序号	因子	符号	单位	均值 $\mu_{(.)}$	标准差 $\sigma_{(.)}$
1	初速	v_0	m/s	930.0	4.5
2	初始弹道倾角	θ_0	rad	0.89	0.03
3	初始弹道偏角	ψ_0	rad	0.0	0.03
4	初始纵向攻角	δ_{10}	rad	0.0	0.04
5	初始横向攻角	δ_{20}	rad	0.0	0.04
6	初始弹丸角速度 分量	ω_{x10}	rad/s	1885.0	20.0
7		ω_{y10}	rad/s	0.0	10.0
8		ω_{z10}	rad/s	0.0	10.0

靶场试验设计的主要内容包括性能响应的因子效应分析和射弹飞行试验的最优回归设计两部分。射击仿真试验数据由均匀设计 $U_{24}(24^8)$ 和式(3-26)~(3-37)给出的 6D 刚体弹道方程计算得到, 如表 3-3 所示。

表 3-3 均匀设计 $U_{24}(24^8)$ 下的弹丸落点仿真数据
Table3-3 Uniform design $U_{24}(24^8)$ and related projectile impact simulation data

No.	v_0	θ_0	ψ_0	δ_{10}	δ_{20}	ω_{x10}	ω_{y10}	ω_{z10}	X_c	Z_c
1	916.5	0.85	0.01	-0.01	0.08	1910	15	-27.5	17573	1217
2	917.63	0.91	-0.08	0.11	0.04	1875	0	-22.5	16741	-1103
3	918.75	0.97	0.03	-0.02	0	1840	-15	-17.5	14022	1697
4	919.88	0.85	-0.05	0.1	-0.04	1930	-30	-12.5	16860	-477
5	921	0.91	0.05	-0.03	-0.08	1895	17.5	-7.5	16749	-1906
6	922.13	0.97	-0.03	0.09	-0.12	1860	2.5	-2.5	16109	325
7	923.25	0.84	0.08	-0.04	0.09	1825	-12.5	2.5	16229	2681
8	924.38	0.9	-0.01	0.08	0.05	1915	-27.5	7.5	17470	626
9	925.5	0.96	-0.09	-0.05	0.01	1880	20	12.5	17888	1566
10	926.63	0.83	0.02	0.07	-0.03	1845	5	17.5	17991	1510
11	927.75	0.89	-0.07	-0.06	-0.07	1935	-10	22.5	17878	-1102
12	928.88	0.95	0.04	0.06	-0.11	1900	-25	27.5	16992	2157
13	930	0.82	-0.05	-0.07	0.1	1865	22.5	-30	17075	-347
14	931.13	0.88	0.06	0.05	0.06	1830	7.5	-25	16014	2535
15	932.25	0.94	-0.02	-0.08	0.02	1920	-7.5	-20	15923	384
16	933.38	0.82	0.08	0.04	-0.02	1885	-22.5	-15	15985	2616
17	934.5	0.88	0	-0.09	-0.06	1850	25	-10	16478	1003
18	935.63	0.94	-0.08	0.03	-0.1	1940	10	-5	17389	-1154
19	936.75	0.81	0.02	-0.1	0.11	1905	-5	0	17632	1237
20	937.88	0.87	-0.06	0.02	0.07	1870	-20	5	14256	-633
21	939	0.93	0.05	-0.11	0.03	1835	27.5	10	17333	2582
22	940.13	0.8	-0.04	0.01	-0.01	1925	12.5	15	19425	-142
23	941.25	0.86	0.07	-0.12	-0.05	1890	-2.5	20	17310	2655
24	942.38	0.92	-0.02	0	-0.09	1855	-17.5	25	16505	343

表 3-3 给出的数据中,弹丸落点的纵向坐标 X_c 和横向坐标 Z_c 为性能响应。曹宁^[169]根据表 3-3 的数据,取一阶线性模型为完整回归模型,应用逐步回归法研究了各因子对弹丸落点坐标(X_c, Z_c)的影响,建立了如下的回归方程

$$\begin{cases} X_c = -6518.9389 - 8185.7362\theta_0 + 4221.6885\delta_{10} + 16.3254\omega_{x10} \\ \quad + 45.9343\omega_{y10} + 35.3075\omega_{z10} \\ Z_c = 831.3748 + 18453.9056\psi_0 \end{cases} \quad (3-38)$$

正则化因子效应分析同样采用一阶线性完整回归模型,取 Gamma 分布参数(a, b)=(1, 0),应用式(3-12)给出的层次化贝叶斯 Lasso 模型可得到因子效应项系数向量的稀疏解,如表 3-4 所示。

表 3-4 贝叶斯 Lasso 的因子效应分析结果
Table3-4 Analysis of effects of factors with Bayesian Lasso

X_c		Z_c	
因子	系数	因子	系数
常数项	30575	常数项	830.45
v_0	-38.822	ψ_0	18454
θ_0	-9090.5		
ω_{x10}	16.204		
ω_{y10}	41.22		
ω_{z10}	40.695		

对比表 3-4 和式(3-38)中的因子效应项可见,贝叶斯 Lasso 因子效应分析结果与逐步回归法的分析结果基本上一致,均识别出初始弹道倾角 θ_0 、初始弹丸角速度 $\omega_{x10}, \omega_{y10}, \omega_{z10}$ 为纵向落点 X_c 的主要影响因子。由算法 3.1 可进一步得到 X_c 的显著性因子效应项 $v_0, \theta_0, \delta_{10}, \omega_{x10}, \omega_{y10}, \omega_{z10}$ 。其中,初始纵向攻角 δ_{10} 为存在概率 $\eta=0.0063$ 的可能因子效应项。显然,贝叶斯正则化因子效应分析结果不仅准确,而且能够区分确定性因子效应和可能的因子效应,对射弹飞行试验设计具有更好的指导意义。

根据贝叶斯正则化因子效应分析结果构造射弹飞行试验设计信息矩阵,由式(3-25)定义的加权预测方差,应用算法 3.2 可以得到兼顾所有可能因子效应项的 6+1 点 D-最优回归设计,据此安排射弹飞行试验,得到的现场飞行试验数据如表 3-5 所示。

表 3-5 D-最优设计的现场试验数据

Table3-5 D-optimal design and related output response

No.	v_0	θ_0	ψ_0	δ_{10}	δ_{20}	ω_{x10}	ω_{y10}	ω_{z10}	X_c	Z_c
1	943.5	0.98	0.09	0.12	0.12	1945	-30	30	17060	2491.3
2	916.5	0.98	0.054	0.12	0.12	1945	30	-30	17420	1827
3	916.5	0.8	0.09	-0.072	-0.024	1825	-30	-30	14538	2495.7
4	943.5	0.8	-0.054	-0.12	0.072	1945	30	-30	18781	-166.07
5	943.5	0.98	-0.09	0.024	-0.072	1825	30	30	17364	-830.41
6	916.5	0.8	-0.018	-0.024	0.024	1849	30	30	19119	498.28
7*	916.5	0.8	0.018	0.072	-0.12	1945	-30	30	-	-

由表 3-3 和表 3-5 中前 6 个试验点上的数据应用式(3-12)进一步分析因子效应信息,可以发现因子效应项 δ_{10} 对 X_c 和 Z_c 无显著影响。因此,现场试验实施过程中可以排除因子效应项 δ_{10} 对应的第 7 个附加设计点。根据表 3-3 中的仿真试验数据和表 3-5 中的现场试验数据,重新基于因子效应分析结果进行回归建模,得到如下的弹丸落点(X_c, Z_c)的回归方程

$$\begin{cases} X_c = -7268.3 - 7560.2\theta_0 + 16.41\omega_{x10} + 37.892\omega_{y10} + 31.891\omega_{z10} \\ Z_c = 830.6 + 18458\psi_0 \end{cases} \quad (3-39)$$

式(3-39)给出了基于贝叶斯正则化的靶场试验设计方法得到的弹丸落点响应模型,为了说明基于贝叶斯正则化的靶场试验设计方法的有效性,另外选择传统的小子样均匀设计 $U_6(6^8)$ 进行对比分析,均匀设计 $U_6(6^8)$ 下相应的现场试验数据如表 3-6 所示。

表 3-6 均匀设计 $U_6(6^8)$ 下的现场试验数据Table3-6 Uniform design $U_6(6^8)$ and related output response

No.	v_0	θ_0	ψ_0	δ_{10}	δ_{20}	ω_{x10}	ω_{y10}	ω_{z10}	X_c	Z_c
1	916.5	0.98	-0.018	-0.12	0.12	1825	18	30	16915	503.78
2	921.9	0.872	-0.09	-0.072	-0.12	1921	-30	6	16695	-827.82
3	927.3	0.836	0.054	0.072	0.024	1849	6	-6	16793	1823
4	932.7	0.8	0.09	-0.024	-0.072	1897	30	18	19519	2492
5	943.5	0.908	-0.054	0.12	0.072	1945	-18	-18	16564	-157.76
6	938.1	0.944	0.018	0.024	-0.024	1873	-6	-30	15133	1172

根据表 3-3 和表 3-6 中的试验数据重新进行因子效应分析,并基于分析结果进行回归建模,可以得到弹丸落点(X_c, Z_c)的响应模型如下:

$$\begin{cases} X_c = -7343.4 - 7543.8\theta_0 + 16.442\omega_{x10} + 37.921\omega_{y10} + 31.86\omega_{z10} \\ Z_c = 831.18 + 18446\psi_0 \end{cases} \quad (3-40)$$

D-最优准则对应试验设计信息矩阵行列式最大,取表 3-4 的因子效应分析结果构造信息矩阵,可得表 3-5 中 D-最优设计的信息矩阵行列式为 1.6421e07,表 3-6 中均匀设计的行列式为 4.5177e05。显然,基于贝叶斯正则化的靶场试验设计的 D 优良性更好,能够更好的覆盖重要因子空间。

为了对比由不同设计得到的性能响应模型的泛化性能,在因子空间中随机抽取 20 个点的测试点集。分别计算模型(3-39)和(3-40)的均方根误差(Root mean square error, RMSE)和最大绝对误差(Maximal absolute error, MAE),结果如表 3-7 所示。

表 3-7 两种试验设计方案的拟合性能对比
Table3-7 Fitting performance comparison of two design schemes

试验设计方案	RMSE		MAE	
	X_c	Z_c	X_c	Z_c
D-最优设计	51.78	16.82	194.89	36.59
均匀设计	57.58	18.66	198.41	39.53

可见,基于贝叶斯正则化的靶场试验设计方法得到的响应模型(3-39)拟合性能优于由传统的小子样均匀设计得到的响应模型(3-40)。这是由于贝叶斯正则化因子效应分析充分利用了验前信息获取影响系统性能响应的重要因子效应项,并由此在后续现场试验设计中应用最优回归法覆盖需要重点关注的因子空间。因此,基于贝叶斯正则化的靶场试验设计方法能够极大地提高试验的效益,用最少的试验样本获取最多的试验信息。

3.6 本章小结

本章研究了基于贝叶斯正则化的靶场试验设计方法。首先,考虑系统性能响应与因子之间的关系,给出了基于回归模型的靶场试验设计问题的描述,并说明了因子效应信息对靶场试验设计的重要性。其次,通过分析正则化方法求解不适定问题的稳健性,提出了贝叶斯正则化因子效应分析方法。然后,给出了考虑因子效应信息的靶场试验设计信息矩阵,提出了贝叶斯正则化试验设计信息矩阵构造方法,得到兼顾可能因子效应项的备选附加点 D-最优设计用于有因子效应信息的靶场试验设计,并给出了无因子效应信息试验的空间填充设计。最后,应用火炮射表试验的示例验证了基于贝叶斯正则化的靶场试验设计方法的有效性。

第4章 基于修正幂先验的靶场试验样本量确定方法

4.1 引言

靶场试验样本量研究的目的是确定仿真和现场试验的样本量大小,使设计的样本量分配方案既满足评估精度要求,同时又消耗最少的试验资源。样本量确定方法总体上可分为经典方法和贝叶斯方法两大类^[73]。经典样本量确定方法一般是通过参数估计的精度要求或假设检验的功效分析来确定一个具体的样本量值,在将其用于 IT&E 模式下的靶场试验样本量确定时,无法解决仿真试验和现场试验的样本量比例分配问题。因此,贝叶斯样本量确定方法就成为靶场试验样本量确定的必然选择。

贝叶斯样本量确定方法的基本思想是根据兴趣参数的先验信息及其预后验估计性能准则,给出确定样本量的具体方法,在用于计算靶场试验的样本量时,存在难以确定验前样本信息权值的问题。并且在仿真试验样本量非常大的情况下,基于标准幂先验构造的设计先验,其分布特性往往已经满足贝叶斯样本量确定准则的要求,此时,对兴趣参数的估计似乎已不需要现场试验样本的参与,但是由于仿真试验数据存在可信度问题,完全应用仿真试验样本进行武器系统性能评估存在较大的风险。

针对靶场试验样本量确定时存在的上述问题和矛盾,本章在考虑验前样本量影响的基础上,提出一种基于修正幂先验的贝叶斯样本量确定方法,避免存在大量仿真样本时的后验评估风险。具体结构安排如下:首先,通过分析设计先验对兴趣参数后验分布的影响,说明仿真和现场试验样本量在兴趣参数后验估计中的作用,由此给出靶场试验样本量确定问题的描述。其次,分析验前样本量对兴趣参数先验方差的影响,定义一种同时考虑可信度与样本量影响的试验效应指标,并基于试验效应等价构造贝叶斯样本量确定的修正幂先验。然后,在修正幂先验框架下,根据贝叶斯样本量计算准则以及兴趣参数的预后验分布特性,给出准则约束下的靶场试验样本量确定方法。最后,分析靶场试验样本量的优化设计目标,以导弹精度试验的样本量确定为例,给出在成本和评估精度约束下的样本量确定方法,通过分析计算结果验证基于修正幂先验的靶场试验样本量确定方法的有效性。

4.2 靶场试验样本量确定问题的描述

靶场试验中的样本量确定是根据试验与鉴定的任务要求,确定合理的仿真

和现场试验的样本量。由于经典样本量确定方法无法处理仿真试验和现场试验的样本量分配比例和权值，不能得到一个适当的样本量分配方案。因此，贝叶斯样本量确定方法就成为确定靶场试验样本量的必然选择。

贝叶斯样本量确定方法与经典样本量确定方法的根本区别在于对兴趣参数先验信息的应用。假设我们在因子空间的某个试验点上已有 n_0 个仿真试验样本 $\mathbf{x}_0 = (x_{01}, x_{02}, \dots, x_{0n_0})$ ，根据仿真试验样本 \mathbf{x}_0 得到的兴趣参数先验分布为 $\pi(\theta|\mathbf{x}_0)$ 。为保证该试验点上的评估精度，还需要一定样本量的现场试验，假设现场试验样本 $\mathbf{X}_n = (X_1, \dots, X_n)$ 的容量为 n ，其联合概率密度 $f_n(\bullet|\theta)$ 取决于待估兴趣参数 θ 。给定样本数据 $\mathbf{x}_n = (x_1, \dots, x_n)$ ，似然函数 $L(\theta; \mathbf{x}_n) \propto f_n(\mathbf{x}_n|\theta)$ ，以及兴趣参数先验分布 $\pi(\theta|\mathbf{x}_0)$ ，兴趣参数 θ 的后验分布可表示为

$$\pi(\theta|\mathbf{x}_n) = \frac{f_n(\mathbf{x}_n|\theta)\pi(\theta|\mathbf{x}_0)}{\int_{\Theta} f_n(\mathbf{x}_n|\theta)\pi(\theta|\mathbf{x}_0)d\theta} \quad (4-1)$$

定义 $T(\mathbf{x}_n)$ 为 θ 后验分布的一般函数，其性能可由现场试验的样本量 n 控制。例如 $T(\mathbf{x}_n)$ 可以为后验方差，或最大后验密度（Highest posterior density, HPD）区间的宽度，或某个假设的后验概率。在观测到现场试验样本数据 \mathbf{x}_n 之前， $T(\mathbf{X}_n)$ 可视为随机变量。贝叶斯样本量计算的基本思想是通过选择 n 使观测值 $T(\mathbf{x}_n)$ 提供兴趣参数 θ 的准确信息，样本量确定由以下边缘密度函数实现

$$m_n(\mathbf{x}_n; \pi) = \int_{\Theta} f_n(\mathbf{x}_n|\theta)\pi(\theta|\mathbf{x}_0)d\theta \quad (4-2)$$

上式为采样分布 $f_n(\mathbf{x}_n|\theta)$ 与 θ 先验分布 $\pi(\theta|\mathbf{x}_0)$ 的混合。大多数贝叶斯样本量计算准则都是通过选择最小的 n ，对给定的 $\varepsilon > 0$ 或 $\varepsilon' \in (0, 1)$ ，使以下两式之一成立

$$E_m[T(\mathbf{X}_n)] \leq \varepsilon \quad \text{or} \quad P_m[T(\mathbf{X}_n) \in A] \leq \varepsilon' \quad (4-3)$$

其中， $E_m[\bullet]$ 定义由边缘分布 m_n 计算的期望值。 $P_m[\bullet]$ 表示与密度函数 m_n 相关的概率测量。 A 为 $T(\mathbf{X}_n)$ 取值空间的一个子集。边缘密度函数 $m_n(\mathbf{x}_n; \pi)$ 与样本量 n 密切相关，总体来说，随着样本量 n 的增大，密度函数 $m_n(\mathbf{x}_n; \pi)$ 的形状变的越来越尖锐，由其计算得到的 $E_m[\bullet]$ 和 $P_m[\bullet]$ 也越来越小。

由式(4-2)可知，对于分布形式已知的试验样本，边缘密度函数 $m_n(\mathbf{x}_n, \pi)$ 由兴趣参数 θ 的先验分布 $\pi(\theta|\mathbf{x}_0)$ 决定。因此，靶场试验的贝叶斯样本量确定方法中如何由验前信息构造合适的兴趣参数先验是研究的重点内容。

靶场试验过程中，仿真为现场试验提供先验信息，而幂先验能够实现对验前数据的折合加权。在应用贝叶斯样本量确定方法计算现场试验所需的样本量时，幂先验提供了一种可行的设计先验构造方法。本章将进一步考虑验前样本量对兴趣参数先验分布的影响，提出试验效应指标，以构造适用于靶场试验贝

叶斯样本量确定的修正幂先验。

4.3 基于试验效应等价的修正幂先验

4.3.1 考虑验前样本量影响的试验效应指标

假设 \mathbf{D}_0 代表样本量为 n_0 的仿真试验数据, $L(\theta|\mathbf{D}_0)$ 是在给定的 \mathbf{D}_0 下, 兴趣参数 θ 的似然函数, $\pi_0(\theta)$ 为无信息先验。根据式(2-5), 给定固定幂指数 δ 时, 基于标准幂先验构造的样本量计算设计先验可表示为

$$\pi^p(\theta|\mathbf{D}_0) \propto \pi_0(\theta)L(\theta|\mathbf{D}_0)^\delta, \delta \in (0, 1) \quad (4-4)$$

如果幂指数 δ 为服从一定分布的随机变量, 由式(2-27)可以得到基于规则化幂先验的样本量计算设计先验

$$\pi^p(\theta|\mathbf{D}_0) = \pi_0(\theta) \int_{\mathcal{B}} \frac{L(\theta|\mathbf{D}_0)^\delta \pi(\delta) I_{\mathcal{B}}(\delta)}{\int_{\Theta} L(\theta|\mathbf{D}_0)^\delta \pi(\theta) d\theta} d\delta \quad (4-5)$$

其中, $\mathcal{B} = \{\delta: 0 < \int_{\Theta} L(\theta|\mathbf{D}_0)^\delta \pi(\theta) d\theta < \infty\}$, $I_{\mathcal{B}}(\delta)$ 为示性函数。式(4-4)和(4-5)分别给出了在固定幂指数和随机幂指数的情况下, 应用幂先验构造贝叶斯样本量确定设计先验的方法。由此, 可以得到后续现场试验数据 \mathbf{X}_n 的边缘分布

$$m_n(\mathbf{x}_n|\mathbf{D}_0) = \int_{\Theta} f_n(\mathbf{x}_n|\theta) \pi^p(\theta|\mathbf{D}_0) d\theta \quad (4-6)$$

根据式(4-6)给出的边缘分布 m_n , 基于不同的贝叶斯样本量计算准则, 可以确定出后续试验所需的样本量。显然, 幂先验指数 δ 在贝叶斯样本量确定中具有非常重要的作用。由于缺少可操作性的说明, 且没有给出如何估计 δ 的方法, 幂先验在后验推断中的应用往往存在较大的随意性, Spiegelhalter 等^[89]对此提出了批评。De Santis 针对幂先验用于独立同分布 (IID) 数据兴趣参数 θ 的后验推断问题, 在文献[170, 171]中给出了以下两个引理及其证明:

引理 4.1: 当兴趣参数 θ 的最大似然估计存在且唯一时, 由 \mathbf{D}_0 得到的后验分布 π^p 和由样本量 $r=\delta n_0$ 的标准试验数据得到的后验分布具有相等的估计值。

引理 4.2: 当 $f_n(\cdot|\theta)$ 为指数族时, 在均值相等的情况下, 由 \mathbf{D}_0 构造的幂先验分布 π^p 与样本量为 $r=\delta n_0$ 的验前样本数据的标准后验分布一致。

由式(4-4)基于标准幂先验构造的设计先验可见, 幂指数 δ 的含义与仿真可信度基本类似, 因此在靶场试验样本量确定过程中, 一种直观的思路是取仿真可信度 C_0 作为幂指数 δ 来构造设计先验。假设历史试验数据 $\mathbf{D}_0 = (x_{01}, \dots, x_{0n_0})$ 和当前试验数据 $\mathbf{D} = (x_1, \dots, x_n)$ 都服从 IID 正态分布 $N(\mu, \sigma^2)$, 均值 μ 和方差 σ^2 均未知, μ 为待估的兴趣参数, 定义

$$\begin{cases} \bar{x}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} x_{0i}, s_0^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} (x_{0i} - \bar{x}_0)^2 \\ \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \end{cases}$$

取无信息先验 $\pi(\mu, \sigma^2) \propto (1/\sigma^2)^a$ 作为总体参数 (μ, σ^2) 的初始先验, 其中, $a=1$ 对应参考先验, $a=1.5$ 对应 Jeffreys 先验。根据 2.4.2 节中的推论 2.3, 在给定幂指数 δ 和仿真数据 \mathbf{D}_0 时, 由式(4-4)可以得到在幂先验框架下, 总体均值 μ 的条件后验分布 $\pi(\mu|\mathbf{D}_0, \delta)$ 服从学生 t 分布, 且其位置参数为 \bar{x}_0 , 尺度参数为 $\sqrt{s_0^2/(\delta n_0 + 2a - 3)}$, 自由度为 $\delta n_0 + 2a - 3$ 。根据学生 t 分布的性质, 可以得到总体均值 μ 的先验方差为

$$\text{var}(\mu|\delta, \mathbf{D}_0) = \frac{s_0^2}{\delta n_0 + 2a - 5} \quad (4-7)$$

当幂指数 δ 为具有概率密度函数 $\pi(\delta|\mathbf{D}_0)$ 的随机变量时, 兴趣参数 θ 的幂先验方差可由下式得到

$$\begin{aligned} \text{var}(\theta|\mathbf{D}_0) &= \int_{\Theta} (\theta - \bar{\theta})^2 \cdot \pi(\theta|\mathbf{D}_0) d\theta = \int_{\Theta} (\theta - \bar{\theta})^2 \int_A \pi(\theta, \delta|\mathbf{D}_0) d\delta d\theta \\ &= \int_{\Theta} (\theta - \bar{\theta})^2 \int_A \pi(\delta|\mathbf{D}_0) \cdot \pi(\theta|\delta, \mathbf{D}_0) d\delta d\theta \\ &= \int_A \int_{\Theta} (\theta - \bar{\theta})^2 \cdot \pi(\theta|\delta, \mathbf{D}_0) d\theta \pi(\delta|\mathbf{D}_0) d\delta \\ &= \int_A \text{var}(\theta|\delta, \mathbf{D}_0) \cdot \pi(\delta|\mathbf{D}_0) d\delta \end{aligned} \quad (4-8)$$

当不能确定仿真可信度的具体取值时, 可以根据仿真系统 VV&A 的结果, 构造区间 0~1 上服从概率分布 $\pi(\delta|\mathbf{D}_0) = \text{Beta}(\delta; \alpha_0, \beta_0)$ 的一个随机变量来描述仿真可信度。直接取仿真可信度作为幂先验指数时, 根据式(4-7)和(4-8), 有

$$\begin{aligned} \text{var}(\mu|\mathbf{D}_0) &= \int_b^1 \frac{\hat{\sigma}_0^2}{\delta n_0 + 2a - 3} \text{Beta}(\delta; \alpha_0, \beta_0) d\delta \\ &= \frac{\hat{\sigma}_0^2 \cdot \Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0) \cdot \Gamma(\beta_0)} \cdot \int_b^1 \frac{\delta^{\alpha_0-1} (1-\delta)^{\beta_0-1}}{\delta n_0 + 2a - 3} d\delta \end{aligned} \quad (4-9)$$

其中, $b=(3-2a)/n_0$ 。

假设仿真样本的标准差 $s_0 = 1.2$, 分别取固定幂指数 $\delta_f=0.5$ 和随机幂指数 $\delta \sim \text{Beta}(2, 2)$, 图 4-1 给出了均值 μ 的先验方差 $\text{var}(\mu|\mathbf{D}_0)$ 与验前样本量 n_0 的关系。其中, 实线对应固定幂指数 $\delta_f=0.5$, 虚线对应随机幂指数 $\delta \sim \text{Beta}(2, 2)$ 。

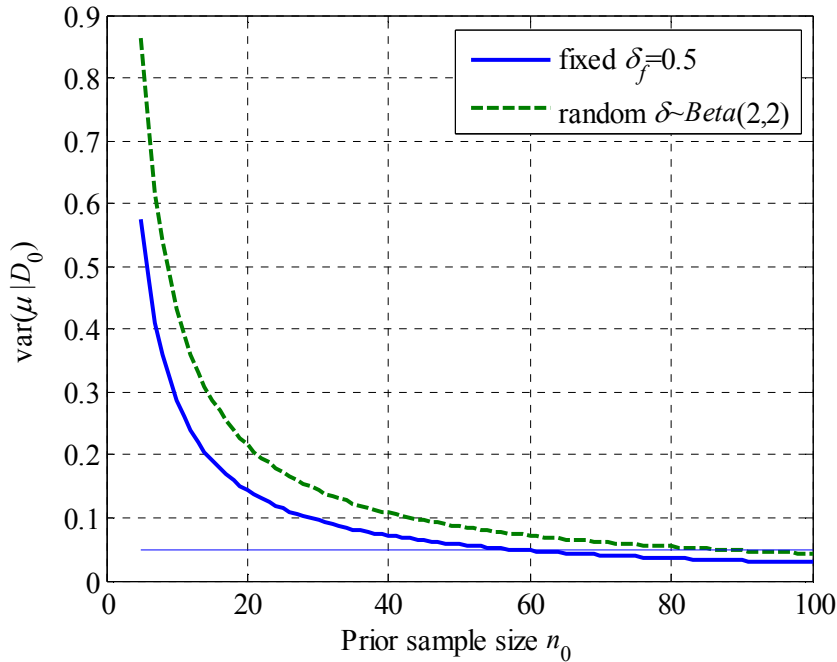


图 4-1 先验方差与验前样本量的关系

Fig.4-1 Prior variance vs. prior sample size

可见,随着 n_0 的增大, $\text{var}(\mu|D_0)$ 趋于 0。也就是说,由大样本量仿真数据构造的设计先验就已经满足式(4-3)给出的贝叶斯样本量计算准则的要求,不需要现场试验数据的参与就能对兴趣参数进行推断。然而,在仿真可信度很低的情况下,即使仿真试验样本量非常大,评估的风险也很高。显然,直接取仿真可信度作为幂指数构造设计先验时,无法合理均衡兴趣参数的评估精度与评估可信度风险之间的矛盾。这是由于应用标准幂先验构造设计先验时,仅考虑了仿真可信度,而没有考虑验前样本量的影响。根据引理 4.1 和引理 4.2,仿真试验样本能够等效的标准样本量 $r=\delta n_0$ 会很大,造成设计先验不合理。

为了解决上述矛盾,一种思路是构造一个综合考虑仿真可信度和样本量的试验效应指标,该指标与仿真样本量和可信度均成正比,并且在一定的可信度下随仿真样本量的增大趋于一个受可信度约束的有限值。而在仿真样本量一定的条件下,可信度越高的仿真试验具有的试验效应指标也越大。基于上述思想,取兴趣参数的后验方差作为评估精度要求,定义如下的试验效应指标。

定义 4.1: 假设某种试验方法具有可信度 $c \in [0, 1]$, 应用该试验方法获得样本量为 n_0 的数据 D_0 , 取待估兴趣参数 θ 的后验方差作为后验估计性能, 则根据 D_0 对兴趣参数 θ 进行后验估计时, 试验效应指标定义为

$$D_E = c \cdot \exp(-\text{var}(\theta|D_0)), c \in [0, 1] \quad (4-10)$$

其中, $\text{var}(\theta|D_0)$ 表示由无信息先验和试验数据 D_0 得到的兴趣参数 θ 的后验方差。

可见, $c=1$ 时, 若 $n_0 \rightarrow \infty$, 有 $\text{var}(\theta|\mathbf{D}_0) \rightarrow 0$, $D_E \rightarrow 1$, 此时 \mathbf{D}_0 等效于真实的样本数据。 c 较低时, 即使 $n_0 \rightarrow \infty$, $\text{var}(\theta|\mathbf{D}_0) \rightarrow 0$, 试验效应指标 $D_E \rightarrow c$ 。

由试验效应指标的定义可见, 该指标反应了将一定样本量的试验数据用于兴趣参数后验估计时的效应力大小。IT&E 模式下, 对用于靶场试验与鉴定的仿真试验应提出试验效应指标要求。由式(4-10)可见, 只有当仿真可信度高于试验效应指标, 且仿真样本量足够时, 仿真试验才能满足试验效应指标要求。为了便于衡量综合应用仿真和现场试验时的试验效应指标, 根据试验效应指标等价的含义, 提出仿真试验数据 \mathbf{D}_0 的等效样本量定义。

定义 4.2: 假设仿真试验的可信度为 C_0 , 在其样本量为 n_0 时, 与样本量为 n_r 的现场试验 (可信度为 1) 的试验效应指标相等, 则定义 n_r 为仿真试验数据 \mathbf{D}_0 的等效样本量 (Equivalent sample size, ESS)。

由式(4-10)可知, 随着仿真样本量 n_0 的增加, 根据无信息先验和 \mathbf{D}_0 得到的兴趣参数 θ 的后验方差 $\text{var}(\theta|\mathbf{D}_0) \rightarrow 0$, 试验效应 $D_E \rightarrow C_0$ 。因此, \mathbf{D}_0 的等效样本量 n_r 趋于一个由 C_0 决定的有限值, 并且可信度 C_0 越高, 等效样本量 n_r 越大, 与实际情况相符。根据 4.3.1 节介绍的引理 4.1 和引理 4.2, 用于构造幂先验的幂指数 δ 可以看作是等效样本量与验前样本量之比。这样, 基于试验效应指标等价得到仿真数据 \mathbf{D}_0 的等效样本量 n_r 以后, 可以计算出考虑验前样本量影响的修正幂指数, 进而得到基于试验效应等价的修正幂先验构造方法。

4.3.2 基于试验效应等价的修正幂先验

靶场试验的测试数据一般服从 IID 的正态分布, 根据等效样本量的定义, 在试验效应指标相等的假设条件下, 由式(4-10)可以推导出仿真数据 \mathbf{D}_0 的等效样本量计算公式。取样本总体均值 μ 为兴趣参数 θ , 在给定验前样本 \mathbf{D}_0 时, Bernardo^[148]给出了均值 μ 的后验分布为学生 t 分布

$$\pi(\mu|\mathbf{D}_0) = \text{St}(\mu|\bar{x}_0, (n_0-1)s_0^{-2}, n_0-1)$$

由此, 可以得到 μ 的先验方差

$$\text{var}(\mu|\mathbf{D}_0) = \frac{s_0^2}{n_0-3}, n_0 > 3$$

根据等效样本量的定义, 有

$$c \cdot \exp(-\text{var}(\mu|\mathbf{D}_0)) = c \cdot \exp\left(-\frac{s_0^2}{n_0-3}\right) = \exp\left(-\frac{s_r^2}{n_r-3}\right)$$

对上式进行推导, 不难得到等效样本量的计算公式

$$n_r = \frac{(n_0 - 3)s_r^2}{s_0^2 - (n_0 - 3)\ln(c)} + 3, n_0 > 3 \quad (4-11)$$

其中, s_r 为未知的现场试验样本的标准差, 仿真数据和现场试验数据满足方差齐性假设时, 可以取 $s_r = s_0$ 。显然, 当 $n_0 \rightarrow \infty$ 时, $n_r \rightarrow 3 - s_r^2 / \ln(c)$; 而当 $c=1$ 时, $n_r = n_0$, 与实际的情况相符。定义修正幂指数 $\delta_m = n_r / n_0$, 用于构造贝叶斯样本量计算的设计先验, 有

$$\delta_m = \frac{(n_0 - 3)s_r^2}{n_0 [s_0^2 - (n_0 - 3)\ln(c)]} + \frac{3}{n_0}, n_0 > 3 \quad (4-12)$$

根据式(4-12)计算的修正幂指数 δ_m 和式(4-4)、(4-5)定义的样本量计算设计先验, 可以构造出基于试验效应等价的修正幂先验

$$\pi_m^p(\theta | \mathbf{D}_0) = \pi_0(\theta) L(\theta | \mathbf{D}_0)^{\delta_m} \quad (4-13)$$

式(4-12)和(4-13)给出了在仿真可信度 c 下, 由样本量为 n_0 的仿真数据构造的兴趣参数 θ 的修正幂先验。仿真和现场试验数据服从方差齐性假设时, 给定样本标准差 $s_r = s_0 = 1.2$, 图 4-2 给出了不同仿真可信度下的修正幂指数 δ_m 与仿真样本量 n_0 的关系。可见, 修正幂指数 δ_m 随仿真样本量 n_0 的增大而减小, 式(4-11)表明其等效样本量随 n_0 的增加趋于有限值。而且, 同样的仿真样本量下, 可信度 C_0 越高, 修正幂指数 δ_m 越大。

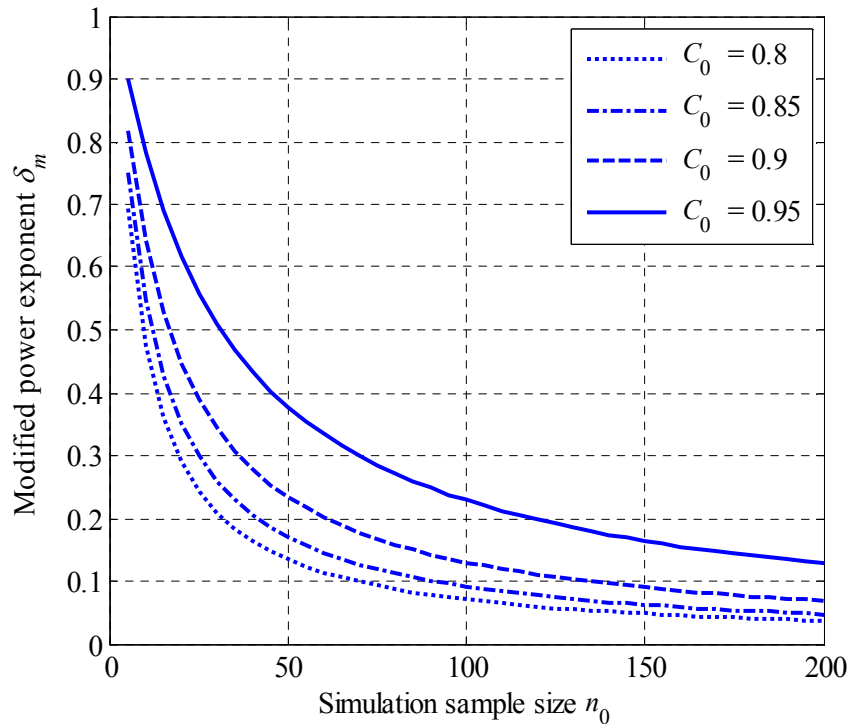


图 4-2 修正幂指数与仿真样本量的关系

Fig.4-2 Modified power exponent vs. simulation sample size

为了进一步分析修正幂先验框架下验前样本量 n_0 与兴趣参数 μ 的先验方差 $\text{var}(\mu|D_0)$ 之间的关系,应用式(4-12)和(4-8)分别计算在固定可信度 $C_f=0.5$ 和随机可信度 $C \sim \text{Beta}(2, 2)$ 时,由修正幂先验得到的兴趣参数先验方差。图 4-3 给出了它们之间的关系曲线。可见,修正幂先验方差随验前样本量的增加趋于有限值,不会出现仅由大样本量的验前数据就能满足评估精度要求的矛盾。

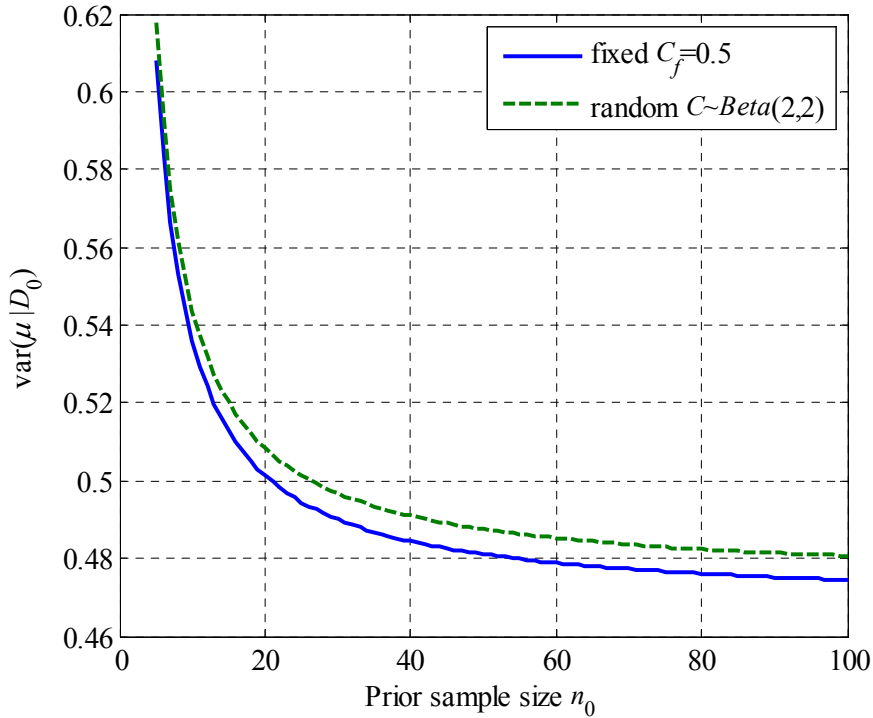


图 4-3 修正幂先验方差与验前样本量的关系

Fig.4-3 Modified power prior variance vs. prior sample size

靶场试验数据的分布特性由相应的统计检验确定,通常情况下,绝大多数靶场试验数据都服从正态分布。在通过正态性检验后,对于服从 IID 正态分布的试验数据,基于修正幂指数 δ_m 和历史试验数据 D_0 ,可以构造用于后续现场试验贝叶斯样本量计算的设计先验,其概率分布函数可表示为

$$p(\mu, \lambda | D_0) = N(\mu | \mu_0, n_r \lambda) \cdot \text{Ga}(\lambda | \alpha_0, \beta_0) \quad (4-14)$$

其中, $N(\mu | \mu_0, n_r \lambda)$ 表示均值为 μ_0 , 精度为 $n_r \lambda$ 的正态分布, n_r 为由式(4-11)计算得到的等效样本量; $\text{Ga}(\lambda | \alpha_0, \beta_0)$ 表示参数为 α_0 和 β_0 的 Gamma 分布。且有 $\lambda = 1/\sigma^2$, $\mu_0 = \bar{x}_0$, $\alpha_0 = (n_r - 1)/2$, $\beta_0 = n_r s_0^2 / 2$ 。

式(4-14)给出了基于修正幂先验构造的用于现场试验贝叶斯样本量确定的设计先验。由此,根据式(4-6)可以计算得到后续现场试验数据 X_n 的边缘分布,然后,就可以根据样本量的贝叶斯确定准则,计算出现场试验所需的样本量,得到准则约束下的最优靶场试验样本量分配方案。

4.4 基于修正幂先验的靶场试验样本量确定方法

4.4.1 贝叶斯样本量确定准则

贝叶斯样本量确定准则可以看作是对式(4-3)给出的兴趣参数后验评估精度要求的具体细化。目前常用的贝叶斯样本量确定准则有以下三种^[93]：

a) 平均后验方差准则 (Average posterior variance criterion, APVC)：对于给定的 $\varepsilon > 0$ ，选择最小的样本量 n ，使之满足

$$E[\text{var}(\theta|X_n)] \leq \varepsilon \quad (4-15)$$

式(4-15)中， $\text{var}(\theta|X_n)$ 为兴趣参数 θ 的后验方差。此时， $T(x_n) = \text{var}(\theta|x_n)$ ，这一准则通过后验分布的散布大小来实现对估计精度的控制。

b) 平均长度准则 (Average length criterion, ALC)：对于给定的 $l > 0$ ，寻找最小的样本量 n ，使以下不等式成立

$$E[L_\alpha(X_n)] \leq l \quad (4-16)$$

式(4-16)中，对于固定的显著性水平 $\alpha \in (0, 1)$ ， $L_\alpha(X_n)$ 为置信水平 $(1-\alpha)$ 下 θ 后验区间估计的随机长度。此时的 $T(x_n) = L_\alpha(x_n)$ 。这一准则由 Joseph 等人^[172]提出，用于控制最大后验密度区间的平均长度。

c) 长度概率准则 (Length probability criterion, LPC)：对于给定的 $l > 0$ 和 $\varepsilon' \in (0, 1)$ ，选择最小的样本量 n ，使之满足

$$\Pr[L_\alpha(X_n) \geq l] \leq \varepsilon' \quad (4-17)$$

与平均长度准则相同，此时 $T(x_n) = L_\alpha(x_n)$ ，长度概率准则也可以写为(4-3)式中的后一种形式。 $A = (l, U)$ ，其中 U 定义了可信区间长度的上限。Joseph 和 Belisle^[88]推导了 LPC 为最差输出准则 (Worst outcome criterion, WOC)^[172]的一种特殊情况，关于 WOC 的详细介绍还可以参考文献[172, 173]。

上述三种贝叶斯样本量确定准则中，准则 a) 是一种点估计类型方法，基于控制点估计的精度；准则 b) 和准则 c) 为区间估计类型的方法，是一种基于控制置信区间估计长度的概率分布法。

4.4.2 修正幂先验的预后验分布

应用贝叶斯样本量确定准则 APVC、ALC 和 LPC 计算后续现场试验所需的样本量时，兴趣参数的预后验分布是一个关键问题。对 4.3.1 节给出的历史数据 D_0 和当前数据 D 的假设条件，为了获得由修正幂先验构造样本量确定的设计先验时，兴趣参数 μ 的平均后验方差、平均置信区间估计长度和置信区间估计长度概率等预后验分布情况，有如下定理。

定理 4.1: 对于服从 IID 正态分布的试验数据, 取均值 μ 和精度 λ 的先验为正态伽马分布 $\pi(\mu, \lambda | \mathbf{D}_0) = N(\mu | \mu_0, n_r \lambda) \cdot \text{Ga}(\lambda | \alpha_0, \beta_0)$, 其中, \mathbf{D}_0 为历史数据, n_r 为等效样本量。若获得样本量为 n 的现场试验数据 \mathbf{x}_n , 则兴趣参数 μ 的平均后验方差 $E[\text{var}(\mu | \mathbf{x}_n)]$ 、后验区间估计的平均长度 $E[L_\alpha(\mathbf{x}_n)]$ 以及区间估计长度大于 l 的预后验计算概率 $\Pr[L_\alpha(\mathbf{x}_n) > l]$ 分别为

$$E[\text{var}(\mu | \mathbf{x}_n)] = \frac{n_r s_0^2}{(n + n_r)(n_r - 3)}, n_r > 3 \quad (4-18)$$

$$E[L_\alpha(\mathbf{x}_n)] = 2t_{1-\alpha/2}(n-1) \sqrt{\frac{n_r s_0^2}{n(n-1)}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \frac{\Gamma(n_r/2-1)}{\Gamma((n_r-1)/2)} \quad (4-19)$$

$$\Pr[L_\alpha(\mathbf{x}_n) > l] = 1 - F_{n_r} \left[\frac{n(n-1)}{4t_{1-\alpha/2}^2(n-1)} l^2 \right] \quad (4-20)$$

其中, $t_{1-\alpha/2}(n-1)$ 是自由度为 $n-1$ 的学生 t 分布的上 $\alpha/2$ 分位点, $F_{n_r}(\cdot)$ 表示参数为 $[(n_r-1)/2, n_r s_0^2/2, (n-1)/2]$ 的 Gamma-Gamma 分布的累积分布函数。

证明:

对均值和方差参数均未知的正态分布, 根据 Bernardo 和 Smith^[148] 给出的均值 μ 的条件后验分布

$$\begin{aligned} p(\mu | \mathbf{x}_n) &= \text{St}(\mu | \mu_n, (n + n_r)(\alpha_0 + n/2)\beta_n^{-1}, 2\alpha_0 + n) \\ \mu_n &= (n_r + n)^{-1}(n_r \mu_0 + n\bar{x}) \\ \beta_n &= \beta_0 + \frac{ns^2}{2} + \frac{1}{2}(n_r + n)^{-1} n_r n (\mu_0 - \bar{x})^2 \end{aligned} \quad (4-21)$$

可以得到 μ 的预后验方差为

$$\text{var}(\mu | \mathbf{x}_n) = \frac{\beta_n}{(n + n_r)(\alpha_0 + n/2)} \cdot \frac{2\alpha_0 + n}{2\alpha_0 + n - 2} \quad (4-22)$$

可见, 均值 μ 的平均预后验方差为 $E(ns^2)$ 和 $E((\mu_0 - \bar{x})^2)$ 的函数。其中, ns^2 为 Gamma-Gamma 分布, \bar{x} 为 t 分布, 其概率分布函数分别为

$$\begin{aligned} p(ns^2) &= \text{Gg}(ns^2 | \alpha_0, 2\beta_0, (n-1)/2) \\ p(\bar{x}) &= \text{St}(\bar{x} | \mu_0, n_r n (n_r + n)^{-1} \alpha_0 \beta_0^{-1}, 2\alpha_0) \end{aligned}$$

根据以上两种分布函数的性质, 可以得到

$$E((\mu_0 - \bar{x})^2) = \text{var}(\bar{x}) = \frac{(n_r + n)\beta_0}{n_r n \alpha_0} \cdot \frac{2\alpha_0}{2\alpha_0 - 2} = \frac{(n_r + n)s_0^2}{(n_r - 3)n} \quad (4-23)$$

$$E(ns^2) = \frac{(n-1)\beta_0}{\alpha_0 - 1} = \frac{n_r(n-1)s_0^2}{n_r - 3} \quad (4-24)$$

由式(4-21)~(4-24)容易推导得到兴趣参数 μ 的平均后验方差计算式(4-18)。

兴趣参数 μ 的区间估计长度的概率分布为

$$L_{\alpha}(x_n) = \frac{2s}{\sqrt{n-1}} t_{1-\alpha/2}(n-1) \quad (4-25)$$

其中, $t_{1-\alpha/2}(n-1)$ 是自由度为 $n-1$ 的学生 t 分布的上 $\alpha/2$ 分位点。容易验证, 兴趣参数 μ 的后验区间估计的平均长度为式(4-19)。

对于给定的长度 $l > 0$, 由式(4-25)给出的区间估计长度的概率分布可以得到兴趣参数 μ 的区间估计长度大于 l 的预后验计算概率为式(4-20)。

证毕

定理 4.1 给出了在贝叶斯样本量确定准则 APVC、ALC 和 LPC 下, 兴趣参数的预后验分布特性。由此, 根据给定的评估精度要求 $\varepsilon > 0$ 或 $\varepsilon' \in (0, 1)$, 由式(4-3)可以计算得到满足准则要求的最小后续试验样本量 n^* 。需要注意的是, 当历史数据的可信度很低的时候, 可能会对评估带来不良影响。因此, 应对靶场试验样本量确定方案的优良性条件进行研究。单独采用现场试验样本 \mathbf{D} 进行评估时, Bernardo 给出了均值 μ 的后验分布为 $\pi(\mu | \mathbf{D}) = \text{St}(\mu | \bar{x}, (n-1)s^{-2}, n-1)$ 。假设仿真和现场试验样本具有方差齐性, 为确保靶场试验综合设计方案(仿真+现场试验)相对于单纯应用现场试验方案时的优良性, 有

$$\frac{n_r s_0^2}{(n+n_r)(n_r-3)} < \frac{s_0^2}{n-3} \Rightarrow n_r > \sqrt{3n} \quad (4-26)$$

上式即为采用综合设计方案的优良性条件。由此可见, 当历史数据的等效样本量 n_r 与当前样本量 n 相比较小时, 综合设计方案具有优良性。而当现场试验样本量足够大时, 由于仿真数据的可信度会对评估带来不良影响, 此时单独应用现场试验数据进行评估会更为优良。

式(4-18)~(4-20)给出了在贝叶斯样本量确定准则 APVC、ALC 和 LPC 下, 由修正幂先验得到的兴趣参数预后验分布, 结合式(4-26)给出的综合设计方案的优良性条件, 就可以计算出满足一定准则要求的最优靶场样本量方案。

4.4.3 靶场试验样本量确定方法

IT&E 模式下, 靶场试验样本量分配的关键问题是如何在满足贝叶斯样本量确定准则的约束下, 合理分配仿真和现场试验的样本量, 使设计的样本量分配方案既满足准则要求的评估精度, 又能最大限度的节省试验成本。以导弹精度半实物仿真试验为例, 其单次试验成本仅包括设备损耗、电力消耗以及一次性消耗部件(如气源、引信触发装置、蓄电池等)的花费, 绝大部分弹上设备都可重复使用, 相比现场实弹飞行试验能够极大的节约试验成本。

假设仿真和现场试验的单位样本消耗分别为 u_0, u_1 ，准则约束下的靶场试验样本量确定就是根据贝叶斯样本量确定准则要求（如兴趣参数的平均后验方差、区间估计的平均长度或一定长度区间估计的概率）来确定所需要的仿真和现场试验样本量分配方案，使得到的样本量分配方案成本最低。此时，样本量分配方案的优化设计可表示为

$$\begin{aligned} \min T_c(N) &= n_0 u_0 + n u_1 \\ \text{s.t.} \quad &\begin{cases} R_B(N, C_0, s_0) \leq \varepsilon \\ n_0 > 3 \\ n \geq 1 \end{cases} \end{aligned} \quad (4-27)$$

其中， $T_c(N)$ 表示试验成本消耗，为优化的目标函数； $R_B(N, C_0, s_0)$ 为贝叶斯样本量确定准则的约束条件，计算中可根据不同的样本量确定准则由式(4-18)~(4-20)得到； $N=[n_0, n]$ 为包含仿真和现场试验样本量的优化向量； C_0 为仿真可信度； s_0 为仿真试验样本的标准差。

式(4-27)给出了准则约束下的靶场试验样本量确定方法，它是一个非线性整数规划方程，难以得到仿真和现场试验样本量的解析解，设计过程中一般只能采用数值计算的方法得到具体的靶场试验样本量分配方案。由式(4-27)确定靶场试验综合设计方案的样本量时，应非常注意式(4-26)给出的优良性条件。尤其是当仿真可信度较低，而要求的评估精度非常高时，必须根据式(4-26)确定综合设计方案的优良性条件，避免不良先验信息带来的负面影响。

4.5 示例分析

IT&E 模式下靶场试验样本量优化设计的目标主要有两点：一是在试验成本约束下确定靶场试验所需的两种样本量大小，使评估精度最优；二是根据要求的评估精度，综合规划仿真和现场试验的样本量，使试验资源消耗最少。本节将以导弹精度靶场试验样本量确定为例，给出上述两种情况下的贝叶斯样本量确定方案。其中，仿真和现场试验的基本情况如表 4-1 所示。

表 4-1 仿真和现场试验的基本情况

Table4-1 Basic conditions of simulation and range tests

类型	单位样本消耗 (\$)	预估的落点标准差 s_0	试验可信度 (C_0)
仿真试验	50	1.2	0.80~0.95
现场试验	2000	-	1.0

例 4.1：对于表 4-1 给出的仿真和现场试验基本情况，设靶场试验的总经费预算不超过 $T_c=20,000\$$ ，样本量优化设计目标是使兴趣参数（落点均值 μ ）的平均

后验方差最小，求仿真和现场试验的最优样本量分配方案。

本例中，样本量分配方案的设计目标是使兴趣参数的平均后验方差最小，约束条件是试验总成本不超过 T_c ，且仿真和现场试验的样本量分别为大于 3 和不少于 1 的整数，由此可以建立下列非线性整数规划方程

$$\begin{aligned} \min & \begin{cases} R_{APV}(N, C_0, s_0) = E[\text{var}(\mu | \mathbf{x}_n)] \\ N = [n_0, n] \end{cases} \\ \text{s.t.} & \begin{cases} n_0 u_0 + n u_1 \leq T_c \\ n_0 > 3 \\ n \geq 1 \end{cases} \end{aligned} \quad (4-28)$$

其中， $R_{APV}(N, C_0, s_0)$ 为优化的目标函数，由(4-18)式给出，表示兴趣参数估计的平均后验方差； N 为包含仿真和现场试验样本量的优化向量； $u_0=50$, $u_1=2000$ 分别为表 4-1 中给出的仿真和现场试验的单位样本消耗； C_0 为仿真可信度； s_0 为仿真样本的标准差； n_r 为根据(4-11)式得到的等效样本量。

根据式(4-28)编写计算程序，应用遍历求解算法可以得到在试验总成本 T_C 的约束下，落点均值 μ 的评估精度最优的靶场试验样本量分配方案。表 4-2 中给出了不同仿真可信度 (SC) 对应的最优样本量分配方案 (SSA)、仿真试验的等效样本量 (ESS)、本文样本量确定方法的平均后验方差 (APV_i) 和传统纯现场试验样本量确定方法的平均后验方差 (APV_f)。

表 4-2 成本约束下的最优靶场试验样本量方案
Table 4-2 Optimal sample size schemes under cost constraint

SC	SSA		ESS	APV_i	APV_f
C_0	n_0	n	n_r	$E[\text{var}(\mu \mathbf{D}_0, \mathbf{D})]$	$E[\text{var}(\mu \mathbf{D})]$
0.80	80	8	6.9659	0.1690	0.16
0.85	80	8	8.9669	0.1275	0.16
0.90	80	8	12.6516	0.0914	0.16
0.95	160	6	24.8609	0.0531	0.16

对比 APV_i 和 APV_f 可见，当 $C_0 \geq 0.85$ 时， APV_i 小于 APV_f ，说明 C_0 较大时，本文提出的样本量确定方法具有更好的评估精度。而且，随着 C_0 的增加，最优 SSA 方案中仿真样本量 n_0 所占的比例和 ESS 越来越大， APV_i 越来越小，说明兴趣参数的评估精度越来越高。

图 4-4 是仿真可信度为 0.85 时，在试验总经费约束下，平均后验方差与现场试验样本量之间的关系曲线。其中，仿真样本量由分配的试验经费计算得到。显然，综合设计方案比传统的纯现场试验方案的评估效果更为优良。

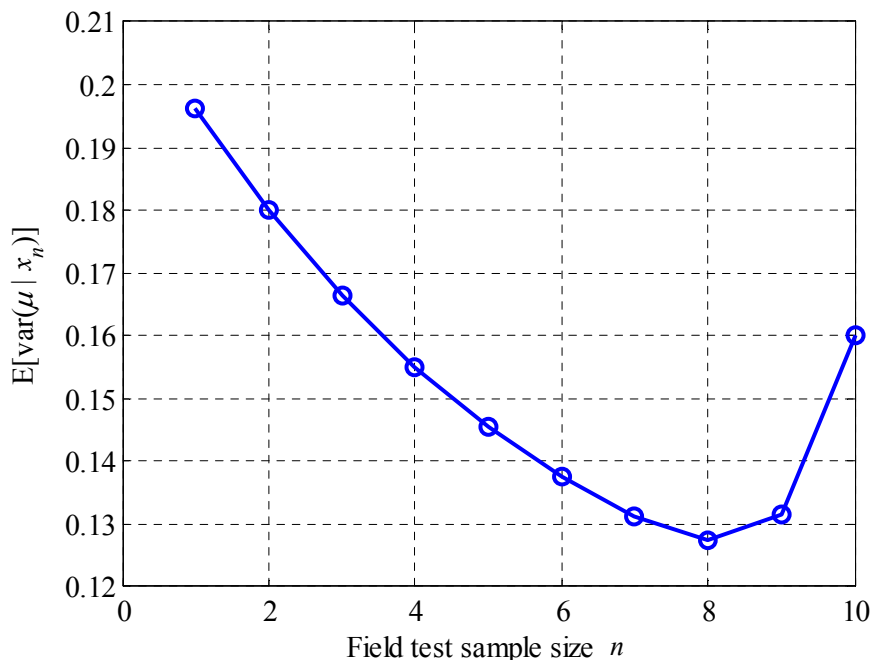


图 4-4 平均后验方差与现场样本量的关系

Fig.4-4 Average posterior variance vs. field test sample size

需要注意的是，当仿真试验的可信度 C_0 很低（如表 4-2 中 $C_0=0.8$ 时）而且要求的评估精度很高时，对由式(4-28)得到的样本量分配方案应按照式(4-26)给出的优良性条件进行检验。如果现场试验的单位成本并不是很高，而且得到的靶场试验样本量分配方案中 $n > n_t^2/3$ ，说明的仿真试验样本对兴趣参数的评估会产生不良影响，此时应考虑只用现场试验对兴趣参数进行评估。

例 4.2：对于表 4-1 给出的仿真和现场试验基本情况，要求落点均值 μ 的评估精度（取 APVC）不大于 $\varepsilon=0.1$ ，样本量优化设计目标是试验总成本最低，求评估精度约束下仿真和现场试验的最优样本量分配方案。

考虑样本量分配方案的设计目标及约束条件，以及仿真和现场试验样本量分别为大于 3 和不小于 1 的整数要求，可以建立下列非线性整数规划方程

$$\begin{aligned} \min T_c(N) &= n_0 u_0 + n u_1 \\ \text{s.t. } &\begin{cases} R_{\text{APV}}(N, C_0, s_0) \leq \varepsilon \\ n_0 > 3 \\ n \geq 1 \end{cases} \end{aligned} \quad (4-29)$$

其中， $T_c(N)$ 为优化的目标函数，表示试验总成本； $\varepsilon=0.1$ 为落点均值 μ 的评估精度约束。其它各变量的含义同例 4.1。

根据式(4-29)编写计算程序，应用遍历求解算法可以得到在落点均值的评估精度约束下，试验总成本最低的靶场试验样本量分配方案。表 4-3 中给出了不

同仿真可信度 (SC) 对应的最优样本量分配方案 (SSA)、修正幂指数 (MPE)、本文样本量确定方法的平均后验方差 (APV_i)、传统纯现场试验样本量确定方法的平均后验方差 (APV_f)，以及试验总成本 (TC)。

表 4-3 APVC 约束下的最优靶场试验样本量方案
Table 4-3 Optimal sample size schemes under required APVC

SC	SSA		MPE	APV_i	APV_f	TC
C_0	n_0	n	δ_m	$E[\text{var}(\mu \mathbf{D}_0, \mathbf{D})]$	$E[\text{var}(\mu \mathbf{D})]$	T_C
0.80	100	18	0.0706	0.0999	0.0758	41000
0.85	68	13	0.1298	0.1000	0.1108	29400
0.90	88	6	0.1456	0.1000	0.2057	16400
0.95	33	2	0.4835	0.0988	1.4400	5650

由表 4-3 可见,随着 C_0 增加,最优 SSA 方案中现场试验样本量 n 越来越小,试验成本也越来越小。同时,修正幂指数 δ_m 随 C_0 的提高越来越大,说明仿真数据在后验评估中所占的权重越来越大,降低了需要的现场试验样本量,使试验成本大大降低。对比 APV_i 与 APV_f 可见,当 $C_0 \geq 0.85$ 时, APV_i 小于 APV_f ,说明 C_0 较大时,本文提出的样本量确定方法具有更好的评估精度。

应注意,综合设计方案的优良性与 C_0 和评估精度要求相关。仿真可信度 $C_0=0.8$ 时,给定较高的评估精度 $\varepsilon=0.1$,传统纯现场试验需要的样本量为 16,此时综合设计方案没有优势。当给定较低的评估精度 $\varepsilon=0.2$ 时,传统方法计算的现场试验样本量为 9,综合设计方案的样本量 $n_0=62$, $n=6$,显然,综合设计方案在试验成本上同样具有较大优势。此外,当仿真试验的可信度 C_0 很高而要求的评估精度不高时,可能仅由仿真样本即可满足兴趣参数的评估精度要求,不需要现场试验样本的参与,即式(4-27)中约束条件 $n \geq 1$ 不再成立。

以上分析说明,在进行导弹精度靶场试验的样本量设计时,应根据式(4-26)给出的靶场试验方案优良性条件和兴趣参数的后验评估精度要求,确定选用何种试验模式。如果仿真试验的可信度较低,兴趣参数的评估精度要求较高,此时对仿真试验样本的应用必须非常小心,避免可信度较低的仿真试验样本对兴趣参数评估的负面影响。而当仿真试验的可信度较高时,采用综合设计方案不仅能够极大的节省试验经费,而且通过合理利用仿真试验数据还能够提高评估精度,带来良好的试验效益。表 4-4 中给出了仿真可信度 $C_0=0.9$ 时,设计的几种满足后验评估精度要求 (APVC) $\varepsilon=0.1$ 的样本量设计方案。

表 4-4 $C_0=0.9$ 时满足 APVC 准则的设计方案
Table 4-4 Test schemes for required APVC with $C_0=0.9$

SSA		MPE	ESS	APV_i	APV_f	TC
n_0	n	δ_m	n_r	$E[\text{var}(\mu \mathbf{D}_0, \mathbf{D})]$	$E[\text{var}(\mu \mathbf{D})]$	T_C
428	4	0.0333	14.2435	0.1000	0.4800	29400
149	5	0.0907	13.5119	0.1000	0.3600	17450
88	6	0.1456	12.8118	0.1000	0.2880	16400
61	7	0.1989	12.1317	0.1000	0.2400	17050
47	8	0.2455	11.5367	0.0996	0.2057	18350
37	9	0.2948	10.9064	0.0998	0.1800	19850
31	10	0.3351	10.3896	0.0993	0.1600	21550

由表 4-4 可见，随着仿真试验样本量的增加，用于构造贝叶斯样本量计算设计先验的修正幂指数 δ_m 减小，仿真试验的等效样本量 n_r 趋于一个有限值。因此，即使仿真试验的样本量再大，也不会出现仅由大量仿真试验样本得到的设计先验就已满足贝叶斯样本量计算准则的情况，避免了兴趣参数后验评估的可信度风险。另一方面，应用修正幂先验进行融合评估时，综合设计方案的兴趣参数评估精度 (APV_i) 较相同样本量下单纯应用现场试验的评估精度 (APV_f) 明显提高，说明了基于修正幂先验进行靶场试验融合评估的优越性。

4.6 本章小结

本章针对靶场试验的样本量确定问题，研究了基于修正幂先验的贝叶斯样本量确定方法。首先，通过分析设计先验对兴趣参数后验分布的影响，说明了仿真和现场试验样本量在兴趣参数后验估计中的作用，由此给出了靶场试验样本量确定问题的描述。其次，在分析验前样本量对兴趣参数先验分布影响的基础上，定义了一种考虑验前样本量影响的试验效应指标，并基于试验效应等价提出了修正幂先验的构造方法。然后，根据贝叶斯样本量确定准则以及兴趣参数的预后验分布特性，给出了准则约束下的靶场试验样本量确定方法。最后，应用示例说明了成本约束和评估精度约束下的靶场试验样本量确定方法，验证了基于修正幂先验的靶场试验样本量确定方法的有效性。

第5章 基于规则化幂先验的靶场试验融合评估方法

5.1 引言

在试验与鉴定领域,融合评估是对试验获取的数据进行整理、归纳、融合与推断的过程,是实现对系统性能客观评价的基础。一般而言,数据融合是一个把多源信息通过合适的方法结合起来得到一个更满意结果的过程,确定不同来源信息的权值是数据融合的关键;而评估是对受多种因素影响的事物做出全面评价的一种多因素决策过程,选择一个合适的模型是实现科学评估的基础。可见,融合是评估的基础,而评估是融合的泛化与推广。靶场试验的融合评估就是通过对多源信息的合理加权,建立起评估靶场试验样本量和试验设计点的方法体系,使靶场试验的综合设计方案满足既能满足系统性能后验评估精度的要求,又能最大限度的节约现场试验的样本量。

在多源数据的融合评估研究中,由于经典统计学难以确定不同来源信息的权值,无法避免少量有用信息被“淹没”的问题。因此,目前一般都是基于贝叶斯理论进行多源试验数据的融合评估。孙锐^[174]通过对已有多源数据融合方法的研究分析,指出贝叶斯融合方法应用的难点在于获取合理的先验分布。此外,评估问题本质上是基于试验数据确定系统性能响应与其影响因子之间函数关系的一个过程,回归分析是实现对系统性能在整个因子空间上的评估,建立指标响应函数的常用方法。在融合评估的过程中,如何刻画不同来源的群组数据之间存在的信息差异,构造适当的模型先验并由此选出性能优良的评估模型,是应用贝叶斯理论进行靶场试验融合评估的难点。

鉴于此,本章进一步研究如何在幂先验的框架下综合应用多源信息对靶场试验进行融合评估。具体安排如下:首先,应用广义线性混合模型对靶场试验的融合评估问题进行描述,将其归结为多源信息融合、靶场试验样本量评估和试验设计评估等三个基本问题。其次,分别考虑多源信息的交互影响和加权约束条件,给出多源信息规则化幂先验构造方法。然后,针对靶场试验的样本量评估,给出样本量评估的后验方差分析模型及其规则化幂先验构造方法,应用Gibbs抽样算法实现兴趣参数的后验方差分析,并通过评估第4章给出的样本量方案说明基于规则化幂先验的靶场试验样本量评估的合理性。最后,考虑因子效应分析在靶场试验设计中的重要性,给出试验设计评估的因子效应分析模型,提出基于规则化幂先验的靶场试验设计评估方法,并通过评估第3章的试验设计方案验证基于规则化幂先验的靶场试验设计评估方法的有效性。

5.2 靶场试验融合评估问题的描述

基于靶场多源数据的融合结果, 来分析兴趣参数评估精度以及试验设计时考虑的因子效应完备性, 是评估靶场试验综合设计有效性的基本思路。广义线性混合模型能够采用线性方法来描述响应变量与因子之间的非线性关系, 并通过引入随机效应来刻画群组数据间的相关、过度离散 (Over-dispersion) 以及分布异质性 (Heterogeneity) 等问题, 适用于分析和处理各类重复测量的群组数据, 突破了传统线性模型要求观测值是彼此独立且等方差的条件限制, 目前已受到越来越多的关注。

广义线性混合模型的上述特性使其非常适合于靶场试验中对异质多源试验数据的融合评估。令 y_{it} , $i=1, \dots, N$, $t=1, \dots, n_i$ 表示第 i 个群组中输出响应的第 t 个测量值, 假设 y_{it} 服从指数族的采样分布, 有

$$p(y_{it}|\theta_{it}, \tau) = \exp\{\tau[y_{it}\theta_{it} - g(\theta_{it})] + c(y_{it}, \tau)\}$$

其中, τ 为离散参数 (Dispersion parameter); θ_{it} 为典范参数 (Canonical parameter), 通过函数 $\theta_{it} = \theta(\eta_{it})$ 与协变量相关, 其中 $\eta_{it} = x_{it}^T \boldsymbol{\beta} + z_{it}^T \mathbf{b}_i$, \mathbf{b}_i 为 $q \times 1$ 随机效应向量, x_{it}^T 和 z_{it}^T 分别为矩阵 X_i 和 Z_i 的行, X_i 为 $n_i \times p$ 矩阵, Z_i 为 $n_i \times q$ 矩阵, $\mathbf{b}_i \sim N_q(0, V)$, V 是 $q \times q$ 协方差矩阵, $\theta(\cdot)$ 是单调可微的联结函数, η_{it} 为线性预测器。这样, $p(y_{it}|\theta_{it}, \tau) \equiv p(y_{it}|\boldsymbol{\beta}, \mathbf{b}_i, \tau)$, 有

$$p(y_{it}|\boldsymbol{\beta}, \mathbf{b}_i, \tau) = \exp\{\tau[y_{it}\theta(\eta_{it}) - g(\theta(\eta_{it}))] + c(y_{it}, \tau)\} \quad (5-1)$$

取 $V = \sigma_b^2 \Sigma_b$, Σ_b 的第 (j, j^*) 个元素具有形式 $\sigma_{jj^*} = \rho^{|j-j^*|}$, $\rho^{|j-j^*|}$ 为 (b_{ij}, b_{ij^*}) 之间的相关系数, 且有 $-1 \leq \rho \leq 1$, 实际计算中取 $T=V^{-1}$ 可能更易于实现。为了便于说明, 假设 $\tau=\tau_0$, 用于 Logistic 和 Poisson 回归时 $\tau_0=1$, 定义式(5-1)中 $c(y)=c(y, \tau_0)$, $p(y_{it}|\boldsymbol{\beta}, \mathbf{b}_i) = p(y_{it}|\boldsymbol{\beta}, \mathbf{b}_i, \tau_0)$ 。

令 $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_N^T)^T$, $\mathbf{y} = (y_{11}, \dots, y_{Nn_N})^T$, $\mathbf{X} = (X_1^T, \dots, X_N^T)^T$, $\mathbf{Z} = \text{diag}(Z_1, \dots, Z_N)$, 则基于 N 个群组数据的 (\mathbf{y}, \mathbf{b}) 联合概率密度为

$$f(\mathbf{y}, \mathbf{b}|\boldsymbol{\beta}, T) = \prod_{i=1}^N \prod_{t=1}^{n_i} f(y_{it}|\boldsymbol{\beta}, \mathbf{b}_i) \pi(\mathbf{b}_i|T) \quad (5-2)$$

其中, $\pi(\mathbf{b}_i|T)$ 是均值为 0, 协方差矩阵 $V=T^{-1}$ 的多元正态分布。令 $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$, 则式(5-2)可写为向量形式

$$\begin{aligned} p(\mathbf{y}, \mathbf{b}|\boldsymbol{\beta}, T) &= p(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}) p(\mathbf{b}|T) \\ &= \exp\left\{\tau_0 \left[\mathbf{y}^T \boldsymbol{\theta}(\boldsymbol{\eta}) - J^T g(\boldsymbol{\theta}(\boldsymbol{\eta})) \right] + J^T c(\mathbf{y}) \right\} p(\mathbf{b}|T) \end{aligned} \quad (5-3)$$

其中, J 为全 1 的列向量, $\boldsymbol{\theta}(\boldsymbol{\eta})$ 为式(5-2)中各元素的向量化形式, 有

$$p(\mathbf{b}|T) = \prod_{i=1}^N (2\pi)^{-q/2} |T|^{1/2} \exp\left\{-(1/2)\mathbf{b}_i^T T \mathbf{b}_i\right\}$$

式(5-3)有许多优良特性。首先,在允许个体群组独特相关结构并保证群组间相互独立性的同时,考虑了组内数据的相关性。其次,模型适用于不平衡数据,即各群组响应向量的长度不一定相同。最后,当对个体均值函数的趋势感兴趣时,随机效应的后验分布或估计具有实际意义。为了导出响应的相关结构,对随机效应部分积分,可以得到似然函数

$$p(\mathbf{y}|\boldsymbol{\beta}, T) = \int_{R^{Nq}} p(\mathbf{y}, \mathbf{b}|\boldsymbol{\beta}, T) d\mathbf{b} \quad (5-4)$$

其中, R^{Nq} 表示 Nq 维的欧氏空间。显然,式(5-4)刻画了输出响应的分布特性,非常适合于在靶场试验融合评估中构造兴趣参数的信息先验,对多源数据进行合理的折合加权,为靶场试验融合评估提供基础。

对比式(3-2)给出的试验设计模型和式(5-4)给出的融合评估模型可以发现,广义线性混合模型中的联结函数 $\theta(\cdot)$ 以及线性预测器 η 对应试验设计模型中的基函数向量 $\mathbf{B}(\mathbf{x})$ 。靶场试验的融合评估应该在获取的多源数据基础上,考虑样本量确定时的评估精度要求和试验设计时的因子效应信息实现。

靶场试验的样本量确定是根据样本量的贝叶斯确定准则,按照一定的评估精度要求得到的。因此,对于靶场试验样本量的评估应通过对兴趣参数后验估计的精度来实现。根据获得的多源试验数据,建立系统性能回归模型,如果由回归模型得到的兴趣参数预测方差满足评估精度要求,则可以认为靶场试验样本量是合理的,可以进一步根据样本数据对试验设计和系统性能进行评估。当兴趣参数的后验方差不满足精度要求时,可以根据与精度要求的差异补充一定的样本量,使新的样本量方案满足评估精度要求。

靶场试验设计是根据仿真试验数据对影响输出响应的因子效应进行分析,并以此作为现场试验最优设计的基础。可见,对于试验设计的评估应根据试验设计阶段考虑的因子效应是否全面进行。根据获得的多源试验数据重新进行因子效应分析,如果得到的因子效应均包含在试验设计阶段所考虑的范围之内,则可以认为靶场试验设计是合理的,能够满足考核系统性能的要求。当因子效应分析结果与试验设计时考虑的不一致时,可以根据因子效应信息的差异,补充一定的试验设计点,使新的试验设计方案满足系统性能考核要求。

综上所述,靶场试验的融合评估应首先根据兴趣参数的后验方差评估样本量是否足够。在此基础上,应用仿真和现场试验数据进行因子效应分析,通过试验设计对因子效应的覆盖性进行试验设计评估。显然,采用上述“评估-设计”序贯循环的思想,能够提高靶场试验综合设计的效益。

5.3 多源信息的规则化幂先验构造

靶场试验融合评估关注的不仅仅是用模型来描述数据，更重要的是通过多源数据融合提高对兴趣参数的评估精度。因此，通过对多源历史数据构造合理的信息先验实现融合，避免少量有用数据被“淹没”是研究的重点。本节考虑多源验前信息的交互影响和加权约束，进一步研究靶场试验融合评估中多源信息的规则化幂先验构造方法。

5.3.1 考虑信息交互影响的规则化幂先验

幂指数反映了历史数据集在兴趣参数后验估计中的作用，当幂指数的后验分布受不同历史数据集共同影响时，验前信息之间存在交互影响。假设有 m 项历史研究，定义 \mathbf{D}_{0j} 为第 j 个历史研究的数据， $j=1, 2, \dots, m$ ， $\underline{\mathbf{D}}_0 = (\mathbf{D}_{01}, \dots, \mathbf{D}_{0m})$ 。对不同的历史数据集选用不同的加权幂指数 δ_j ，令 $\underline{\delta} = (\delta_1, \dots, \delta_m)$ ，并取各个 δ_j 为独立同分布且超参数为 (α_0, β_0) 的 Beta 随机变量时，式(2-27)定义的规则化幂先验可以推广到存在多源历史数据集的情形

$$\pi(\theta, \underline{\delta} | \underline{\mathbf{D}}_0) \propto \frac{\left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \pi(\delta_j | \alpha_0, \beta_0) \right) \pi(\theta)}{\int \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \right) \pi(\theta) d\theta} I_B(\underline{\delta}) \quad (5-5)$$

其中，

$$B = \{(\delta_1, \dots, \delta_m) : 0 < \int \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \right) \pi(\theta) d\theta < \infty\}$$

式(5-5)给出了多源历史数据的幂指数 $\underline{\delta}$ 与兴趣参数 θ 的联合幂先验，显然，上述多源信息规则化幂先验由各个历史数据集的似然函数乘积得到，历史数据集之间存在交互影响。融合当前数据 \mathbf{D} 后，兴趣参数 θ 对 $\underline{\delta}$ 的条件后验分布

$$\pi(\theta | \underline{\delta}, \underline{\mathbf{D}}_0, \mathbf{D}) \propto \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \pi(\delta_j | \alpha_0, \beta_0) \right) L(\mathbf{D}) \pi(\theta) \quad (5-6)$$

$\underline{\delta}$ 的边缘后验分布为

$$\pi(\underline{\delta} | \underline{\mathbf{D}}_0, \mathbf{D}) \propto \frac{\int \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \pi(\delta_j | \alpha_0, \beta_0) \right) L(\mathbf{D}) \pi(\theta) d\theta}{\int \left(\prod_{j=1}^m L(\theta | \mathbf{D}_{0j})^{\delta_j} \right) \pi(\theta) d\theta} I_B(\underline{\delta}) \quad (5-7)$$

式(5-5)~(5-7)给出了考虑信息交互影响的规则化幂先验构造方法及其后验推断形式。可见，每个历史数据集的幂指数 δ_j 不仅与 \mathbf{D}_{0j} 和 \mathbf{D} 的不一致性程度有关，而且还与 \mathbf{D} 和所有历史数据集 $\underline{\mathbf{D}}_0$ 的不一致性相关。而且，式(5-7)还表明，所有幂指数 δ_s 的后验分布之间存在相互影响。

假设当前数据 $\mathbf{D}=(y_1, \cdots, y_n)$ 来自均值 μ 和方差 σ^2 均未知的正态分布, 且有历史数据集 $\mathbf{D}_{01}=(y_{011}, \cdots, y_{01m_1})$ 和 $\mathbf{D}_{02}=(y_{021}, \cdots, y_{02m_2})$ 。令,

$$\begin{aligned}\bar{y}_{01} &= \frac{1}{m_1} \sum_{i=1}^{m_1} y_{01i}, \quad \hat{\sigma}_{01}^2 = \frac{1}{m_1} \sum_{i=1}^{m_1} (y_{01i} - \bar{y}_{01})^2 \\ \bar{y}_{02} &= \frac{1}{m_2} \sum_{i=1}^{m_2} y_{02i}, \quad \hat{\sigma}_{02}^2 = \frac{1}{m_2} \sum_{i=1}^{m_2} (y_{02i} - \bar{y}_{02})^2 \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2\end{aligned}$$

取 (μ, σ^2) 的初始先验为 Jeffreys 先验, 即 $\pi(\mu, \sigma^2) \propto (\sigma^2)^{-1.5}$, 幂指数 δ_j 的初始先验为 $(0, 1)$ 上的均匀分布, 则 (μ, σ^2) 的后验分布为

$$\begin{aligned}\mu | \sigma^2, \underline{\delta}, \mathbf{D}_0, \mathbf{D} &\sim N\left(\frac{\delta_1 m_1 \bar{y}_{01} + \delta_2 m_2 \bar{y}_{02} + n \bar{y}}{\delta_1 m_1 + \delta_2 m_2 + n}, \frac{\sigma^2}{\delta_1 m_1 + \delta_2 m_2 + n}\right) \\ \sigma^2 | \underline{\delta}, \mathbf{D}_0, \mathbf{D} &\sim \text{Inverse-Gamma}\left(\frac{\delta_1 m_1 + \delta_2 m_2 + n}{2}, \beta_1\right)\end{aligned}\quad (5-8)$$

其中,

$$\beta_1 = \frac{2}{n \hat{\sigma}^2 + \delta_1 m_1 \hat{\sigma}_{01}^2 + \delta_2 m_2 \hat{\sigma}_{02}^2 + \frac{n \delta_1 m_1 (\bar{y} - \bar{y}_{01})^2 + n \delta_2 m_2 (\bar{y} - \bar{y}_{02})^2 + \delta_1 m_1 \delta_2 m_2 (\bar{y}_{01} - \bar{y}_{02})^2}{\delta_1 m_1 + \delta_2 m_2 + n}}$$

分别将 (μ, σ^2) 和幂指数 δ_j 的初始先验带入式(5-6)和(5-7), 整理后可以得到幂指数 $\underline{\delta}$ 的边缘后验分布为

$$\pi(\delta_1, \delta_2 | \mathbf{D}_0, \mathbf{D}) \propto \frac{\Gamma\left(\frac{\delta_1 m_1 + \delta_2 m_2 + n}{2}\right) (\delta_1 m_1 + \delta_2 m_2 + n)^{-0.5} \beta_1^{\frac{\delta_1 m_1 + \delta_2 m_2 + n}{2}}}{\Gamma\left(\frac{\delta_1 m_1 + \delta_2 m_2}{2}\right) (\delta_1 m_1 + \delta_2 m_2)^{-0.5} \beta_2^{\frac{\delta_1 m_1 + \delta_2 m_2}{2}}}\quad (5-9)$$

$$\text{其中, } \beta_2 = \frac{2}{\delta_1 m_1 \hat{\sigma}_{01}^2 + \delta_2 m_2 \hat{\sigma}_{02}^2 + \frac{\delta_1 m_1 \delta_2 m_2 (\bar{y}_{01} - \bar{y}_{02})^2}{\delta_1 m_1 + \delta_2 m_2}}。$$

靶场试验数据大多服从正态分布, 不同研究的数据集分布往往并不一致。式(5-5)给出的规则化幂先验框架能够适应不同历史数据集之间可能存在的分布不均匀性。例如, 在导弹精度评估中, 可取相似型号导弹的试验数据和仿真试验数据作为“历史”数据集。并且, 在研制和定型试验的不同阶段获取的数据可以分成几个历史数据集, 以确保每个数据集内部的分布均匀性, 这样就可以更加精确的评价不同历史数据集的作用, 避免了兴趣参数先验分布由样本量较大的验前数据集主导的情形。

5.3.2 考虑信息加权约束的规则化幂先验

试验过程中不同阶段的历史数据均可用于评估系统的性能，由于各阶段的历史数据往往具有分布不均匀性，因此可将其看作具有变动特性的历史数据集，以保证各个数据集内部的分布均匀性。此时，可对不同历史数据集赋予不同的幂指数进行加权，用式(5-5)给出的考虑信息交互影响的规则化幂先验进行融合评估。然而，对于在不同阶段获取的历史数据，其幂指数可能是相互依赖的。因此，如何确定验前数据的信息权值特性，使需要重点关注的数据集有更大的信息权值是值得深入研究的问题。

考虑信息加权约束的规则化幂先验是在考虑信息交互影响的规则化幂先验基础上，增加对不同历史数据集幂指数 δ_j 的约束条件，使可信度较高的数据集具有较大的信息权值。定义所有验前数据中可信度最高的历史数据集的幂指数为 δ_1 ，其他历史数据集的幂指数依据可信度的高低排序分别为 $\delta_2, \dots, \delta_m$ ，给定如下的信息加权约束

$$0 \leq \delta_m \leq \dots \leq \delta_2 \leq \delta_1 \leq 1 \quad (5-10)$$

则由式(5-5)和(5-10)即可构造出考虑信息加权约束的规则化幂先验。

例 5.1：文献[117]中给出了 1991 年至 2000 年间，弗吉尼亚州 Chopawamsic 溪上的一个水质观测站记录的水质观测数据（pH 数据）。数据分析的兴趣参数是 1999-2000 年间水质 pH 数据的均值。分别考虑信息加权约束和交互影响构造规则化幂先验，说明两种规则化幂先验在实际应用中的不同。

本例中，1999 年至 2000 年采集的 pH 数据作为当前数据 D ，而 1991 至 1998 年采集的 pH 数据被分成 4 组历史数据样本。 D_{01} 表示 1997-1998 年的历史数据， D_{02} 表示 1995-1996 年的历史数据， D_{03} 为 1993-1994 年的历史数据， D_{04} 为 1991-1992 年的历史数据。表 5-1 给出了当前数据和历史数据的概括统计量。

表 5-1 1991-2000 年度某水质观测站采集的 pH 数据

Table5-1 pH data collected during 1991 to 2000 at a water monitoring station

	当前数据	历史数据			
	D	D_{01}	D_{02}	D_{03}	D_{04}
样本量	20	19	18	21	21
样本均值	6.37	6.95	7.03	7.15	6.74
样本方差	0.90	0.25	0.22	0.15	0.28

假设当前数据 D 来自正态分布 $N(\mu, \sigma^2)$ ，均值 μ 和方差 σ^2 均未知。应用 Jeffreys 先验作为 (μ, σ^2) 的初始先验，均匀分布作为幂指数的先验 $\pi(\delta_j), j=1, \dots, 4$ 。取近期数据的可信度较高时，考虑信息加权约束的规则化幂先验可表示为

$$\pi(\mu, \sigma^2, \underline{\delta} | \underline{D}_0) \propto \frac{\left(\prod_{j=1}^4 L(\mu, \sigma^2 | \underline{D}_{0j})^{\delta_j} \pi(\delta_j) \right) \pi(\mu, \sigma^2)}{\int \left(\prod_{j=1}^4 L(\mu, \sigma^2 | \underline{D}_{0j})^{\delta_j} \right) \pi(\mu, \sigma^2) d\mu d\sigma^2} \quad (5-11)$$

其中, $\pi(\delta_1)$ 是 $[\delta_2, 1]$ 上的均匀分布, $\pi(\delta_2)$ 是 $[\delta_3, \delta_1]$ 上的均匀分布, $\pi(\delta_3)$ 是 $[\delta_4, \delta_2]$ 上的均匀分布, $\pi(\delta_4)$ 是 $[0, \delta_3]$ 上的均匀分布。

根据式(5-11)定义的规则化幂先验和幂指数的约束条件, 建立水质 pH 数据的贝叶斯统计模型, 由 MCMC 方法可以得到水质 pH 均值 μ 的后验估计。实际计算过程中, 可以应用 OpenBUGS^[110]构造兴趣参数的规则化幂先验, 进而实现 Gibbs 抽样计算, 得到兴趣参数的后验推断。表 5-2 给出了考虑信息加权约束的规则化幂先验、交互影响的规则化幂先验、应用全部历史数据和不用历史数据时的 pH 均值估计结果。可见, 如果没有加权约束条件 $0 \leq \delta_4 \leq \delta_3 \leq \delta_2 \leq \delta_1 \leq 1$, 历史数据 \underline{D}_{04} 在估计中的作用将更大。假定各幂指数独立同分布时, 历史数据在估计中的作用仅由其与当前数据的分布一致性决定。信息加权约束本质上是一种主观信息, 基于评估人员偏好的测试数据在性能估计中更有用这一思想。

表 5-2 不同幂指数特性下 pH 均值的规则化幂先验评估结果
Table5-2 Assessment of pH means using normalized power prior methods with different specifications on power parameters

先验构造方法	$\hat{\mu}$	$\hat{\mu}$ 的标准差	后验均值			
			δ_1	δ_2	δ_3	δ_4
加权约束的规则化幂先验	6.632	0.172	0.423	0.219	0.121	0.059
交互影响的规则化幂先验	6.619	0.158	0.234	0.203	0.134	0.389
全部历史数据(all δ 's=1)	6.847	0.072				
不用历史数据(all δ 's=0)	6.366	0.224				

实践中经常会遇到存在不止一个历史数据集的情形。这些历史数据集可能是由以前的试验收集的, 或者是从类似研究中得到的。“历史数据集”是一个模糊的概念, 并没有严格的定义。因此, 历史数据集在融合评估中的作用较大程度上取决于评估人员的选择。由于不同评估人员对历史数据可用性的认识不一定相同, 如何确定这种主观信息未必能仅仅依靠调整幂指数的分布 $\pi(\underline{\delta})$ 就能实现, 应根据实际情况给出不同历史数据集的加权约束条件。在靶场试验融合评估中, 根据研究人员对每个历史数据集在当前研究中相对重要性的认识, 对不同的历史数据集定义合理的加权约束形式, 对规则化幂先验在实际应用中的推广至关重要。

5.4 基于规则化幂先验的靶场试验样本量评估方法

靶场试验的样本量评估是根据获得的试验数据应用统计分析方法得到兴趣参数的后验评估精度，以判断样本量是否合理的一个过程。根据 5.2 节式(5-1)的描述，取输出响应的均值作为兴趣参数时，后验评估精度可由回归系数向量 β 的后验估计决定。由此，本文提出样本量评估的后验方差分析模型，由 β 后验估计的协方差矩阵计算兴趣参数的后验评估精度，实现对样本量的评估。

5.4.1 样本量评估的后验方差分析模型

第 4 章指出靶场试验的样本量是根据兴趣参数的后验评估精度确定的。获得试验数据后，对样本量的评估也应该从兴趣参数的后验估计精度出发。第 3 章式(3-4)给出了在因子效应确定时，回归系数 β 的最小二乘估计的方差为

$$\text{var}(\hat{\beta}) = M^{-1}(\xi_N) \sigma^2 \quad (5-12)$$

其中， $M(\xi_N)$ 为信息矩阵， ξ_N 是包含所有试验点的精确设计， N 为设计中的总样本点数。显然，回归系数 β 的最小二乘估计方差与不计重复点的设计矩阵和样本量大小均有关。在一定的设计下，各个试验设计点上的样本量越大，回归系数 β 的估计方差越小。可见，上式反映了样本量对评估精度的作用。

靶场试验融合评估阶段，对样本量的评估也应该通过计算兴趣参数的后验方差实现。一般情况下，靶场试验的兴趣参数为输出响应的均值，根据设计阶段的因子效应信息确定基函数向量 $B(x)$ ，则由式(3-2)可以得到因子空间点 x 上的兴趣参数后验估计

$$E(y) = \beta^T B(x) = B(x)^T \beta \quad (5-13)$$

上式中基函数向量 $B(x)$ 是由因子效应信息和空间点 x 确定的一个固定值。可见，因子空间点 x 上的兴趣参数预测后验方差可表示为

$$\text{var}(E(y)) = B(x)^T \text{var}(\beta) B(x) \quad (5-14)$$

由式(5-14)可知，样本量评估关键在于确定回归系数向量 β 的协方差矩阵。靶场试验综合采用仿真和现场试验，仿真试验的样本权值与现场试验样本不同，因此，回归系数向量 β 的协方差矩阵无法根据式(5-12)计算得到。

假设由多源验前数据构造的 β 规则化幂先验为 $\pi(\beta, \delta | D_0)$ ，则根据现场试验数据 D 得到回归系数向量 β 的后验估计为

$$\pi(\beta, \delta | D_0, D) \propto \pi(\beta, \delta | D_0) L(\beta | D) \quad (5-15)$$

由式(5-15)可以得到回归系数向量 β 的后验协方差矩阵估计 $\text{var}(\hat{\beta})$ 。

式(5-14)和(5-15)给出了样本量评估的后验方差分析模型。靶场试验样本量评估时，应基于仿真样本数据 D_0 和可信度 C_0 ，由式(4-12)计算修正幂指数构造

回归系数 β 的规则化幂先验, 然后根据现场试验数据得到回归系数 β 的后验估计及协方差矩阵。这样, 就可以通过式(5-14)计算兴趣参数的后验方差, 并判断兴趣参数的后验精度是否满足要求, 实现对靶场试验样本量的评估。

5.4.2 后验方差分析模型的规则化幂先验

由于靶场试验即存在大样本量低可信度的仿真试验数据, 又有少量真实可信的现场试验数据, 对回归系数 β 的估计应考虑对仿真样本的加权, 5.2 节给出的规则化幂先验是实现了对仿真样本合理加权并避免大量仿真信息“淹没”少量现场试验样本的一种合理途径。

对于靶场试验融合评估的广义线性混合模型, 基于似然函数的经典频率论方法在计算模型参数时需要随机效应进行高维积分, 实现难度极大。因此, 贝叶斯计算方法就成为一种合理的选择。假设靶场试验获取的历史数据 $D_0 = (N_0, X_0, y_0, Z_0)$ 包含 N_0 个群组, 其中, 第 i 个群组有 $n_{0i} \times 1$ 个响应向量 y_{0i} , X_{0i} 和 Z_{0i} 分别为 $n_{0i} \times p$ 和 $n_{0i} \times q$ 的固定效应和随机效应协变量矩阵, $i=1, 2, \dots, N_0$, 令 $b_0 = (b_{01}^T, \dots, b_{0N_0}^T)^T$, $y_0 = (y_{01}^T, \dots, y_{0N_0}^T)^T$, $X_0 = (X_{01}^T, \dots, X_{0N_0}^T)^T$ 。在给定幂指数 δ 的条件下, 系数向量 β 的幂先验可表示为

$$\pi(\beta | D_0, T, \delta) \propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} | \beta, b_{0i})]^{\delta} \pi(b_{0i} | T) db_{0i} \right) \pi_0(\beta) \quad (5-16)$$

其中, $p(y_{0it} | \beta, b_{0i})$ 为历史数据 y_{0it} 的概率密度, $\pi(b_{0i} | T)$ 是均值为 0, 协方差矩阵为 T^{-1} 的正态分布。 $\delta \in [0, 1]$ 为度量历史数据在当前研究中权值大小的先验参数。尽管式(5-16)没有解析的闭合式, 但却有一些有益的数值计算特性, 其向量形式可写为

$$\begin{aligned} \pi(\beta | D_0, T, \delta) &\propto \int_{R^{N_0 q}} [p(y_0 | \beta, b_0)]^{\delta} \pi(b_0 | T) db_0 \\ &= \int_{R^{N_0 q}} \exp \left\{ \delta \left\{ \tau_0 \left[y_0^T \theta(\eta_0) - J_0^T g(\theta(\eta_0)) \right] + J_0^T c(y_0) \right\} \right\} \pi(b_0 | T) db_0 \end{aligned} \quad (5-17)$$

其中, J_0 是全 1 的列向量, $\eta_0 = X_0 \beta + Z_0 b_0$ 为式(5-16)中逐个元素的向量化形式。对于多源历史数据, 在指定超参数 $(\delta, \sigma_b^2, \rho)$ 的先验后, 根据式(5-5)~(5-7), 可以得到后验方差分析模型中系数向量 β 的规则化幂先验

$$\pi(\beta, \underline{\delta} | D_0, T) \propto \frac{\prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} | \beta, b_{0i})]^{\delta_i} \pi(b_{0i} | T) db_{0i} \right) \pi(\beta) \pi(\underline{\delta})}{\int \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} | \beta, b_{0i})]^{\delta_i} \pi(b_{0i} | T) db_{0i} \right) \pi(\beta) \pi(\underline{\delta}) d\beta} I_B(\underline{\delta}) \quad (5-18)$$

其中,

$$\mathcal{B} = \left\{ (\delta_1, \dots, \delta_k) : 0 < \int \prod_{i=1}^k \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} | \boldsymbol{\beta}, b_{0i})]^{\delta_i} \pi(b_{0i} | \mathbf{T}) db_{0i} \right) \pi(\boldsymbol{\beta}) \pi(\underline{\delta}) d\boldsymbol{\beta} < \infty \right\}$$

获取当前数据后, 根据式(5-18)给出的规则化幂先验, 应用 MCMC 方法, 可以得到广义线性混合模型中回归系数向量 $\boldsymbol{\beta}$ 的 Gibbs 抽样。根据 $\boldsymbol{\beta}$ 中各个元素的 Gibbs 抽样值计算其协方差矩阵, 则由式(5-14)可以得到因子空间点 \mathbf{x} 上的兴趣参数后验方差。

5.4.3 靶场试验样本量评估方法

Gibbs 抽样由 Geman S 和 Geman D 提出^[175], 能够通过从一些一元分布中抽样, 产生高维多元联合分布的平稳、不可约且非周期性的马尔科夫链, 非常适用于兴趣参数的贝叶斯计算。令模型参数向量为 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$, 则在其 MCMC 计算中, $\boldsymbol{\beta}$ 的联合分布 $\pi(\boldsymbol{\beta})$ 为目标抽样分布。定义 $p-1$ 维的随机变量

$$\boldsymbol{\beta}_{-j} = (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p),$$

记 $\beta_j | \boldsymbol{\beta}_{-j}$ 的条件概率密度为 $\pi(\beta_j | \boldsymbol{\beta}_{-j})$, $j=1, \dots, p$ 。Gibbs 抽样算法是从这 p 个条件分布中产生候选点, 以解决直接从 $\pi(\boldsymbol{\beta})$ 中抽样的难题, 算法流程如图 5-1 所示。

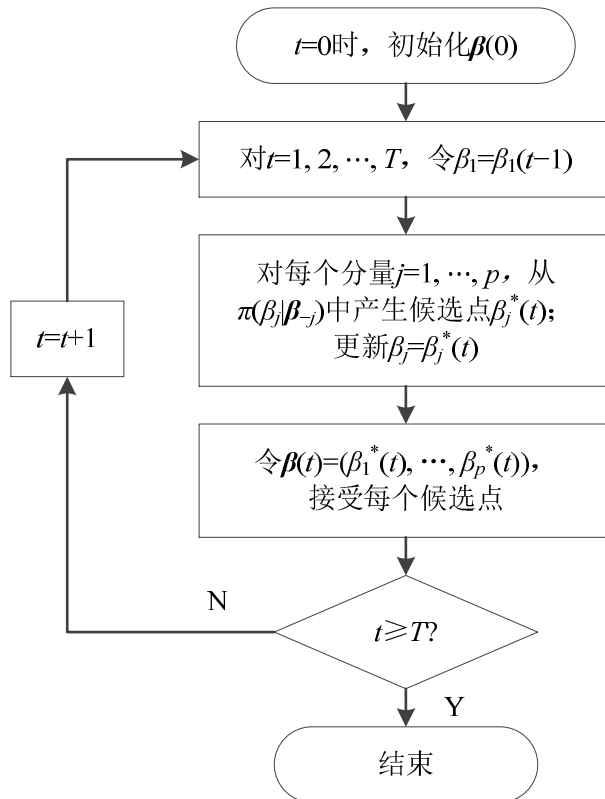


图 5-1 Gibbs 抽样算法流程图

Fig.5-1 Gibbs sampling algorithm flow chart

假设模型参数向量 $\beta=(\beta_1, \cdots, \beta_p)$ 的联合概率密度可表示为 $f(\beta_1, \cdots, \beta_p)$, β_i 的边缘密度为 $f_i(\beta_i)$, $i=1, \cdots, p$ 。如果 $f_i(\beta_i)>0$, 意味着 $f(\beta_1, \cdots, \beta_p)>0$, 则联合密度 f 满足正性条件。Hammersley-Clifford 定理^[176]证明在正性条件下, Gibbs 抽样算法可由完全条件分布唯一的确定联合分布。因此, 通过模型参数向量 $\beta=(\beta_1, \cdots, \beta_p)$ 的 Gibbs 抽样能够正确地估计 β 的后验分布情况, 并得到 β 的后验协方差矩阵。获得靶场试验数据后, 在 Gibbs 抽样算法的支持下, 本文提出如下的靶场试验样本量评估方法:

- (1) 对多源验前数据 D_0 进行融合, 根据式(5-18)构造回归系数向量 β 的规则化幂先验;
- (2) 根据现场试验数据 D , 应用式(5-15)得到回归系数向量 β 的后验概率密度函数 $\pi(\beta, \delta|D_0, D)$;
- (3) 应用 Gibbs 抽样算法得到回归系数向量 β 的后验协方差矩阵, 由式(5-14)计算因子空间点 x 上的兴趣参数后验方差;
- (4) 对比计算的兴趣参数后验方差与要求的评估精度的一致性, 完成对靶场试验样本量的评估。

例 5.2: 以 4.5 节给出的评估精度约束下的靶场试验样本量分配方案为例, 说明基于规则化幂先验的靶场试验样本量评估方法。假设由表 4-3 中可信度为 0.9 时的样本量方案 ($n_0=88, n=6$) 得到的仿真和现场试验数据如表 5-3 所示。

表 5-3 试验设计点上的仿真和现场试验样本数据

Table5-3 Simulation and field test samples on experimental design point

Simulation data								Field data
(1-11)	(12-22)	(23-33)	(34-44)	(45-55)	(56-66)	(67-77)	(78-88)	(1-6)
1.244	2.129	0.692	0.478	0.640	-1.526	1.106	2.292	-2.354
0.226	-1.173	2.004	-0.768	0.941	0.794	-3.035	-1.184	-0.132
-1.554	0.394	1.055	0.150	-0.431	1.014	-1.851	0.016	1.092
0.962	-0.058	2.006	-0.378	0.250	-1.472	1.762	-0.762	-0.674
1.607	1.738	2.912	0.125	-2.819	0.440	-1.222	1.567	-0.345
0.602	-0.808	-0.533	0.648	1.500	-2.008	0.353	-1.422	-0.850
0.317	1.388	0.523	-1.112	-1.905	2.322	-0.851	-0.497	
1.462	-0.812	1.070	-2.817	-0.040	1.708	0.476	0.870	
-0.517	2.575	0.562	1.221	-0.561	-1.413	-1.217	-0.813	
1.481	0.221	-0.307	-0.430	-0.614	-2.126	-0.419	-0.385	
-0.336	0.645	2.620	-0.179	-0.216	-0.593	1.259	1.964	

根据表 4-3 给出的修正幂指数 0.1456，基于修正幂指数构造仿真数据的规则化幂先验，应用现场试验数据进行后验推断，可以得到均值 μ 和标准差 σ 的后验分布如图 5-2 所示。

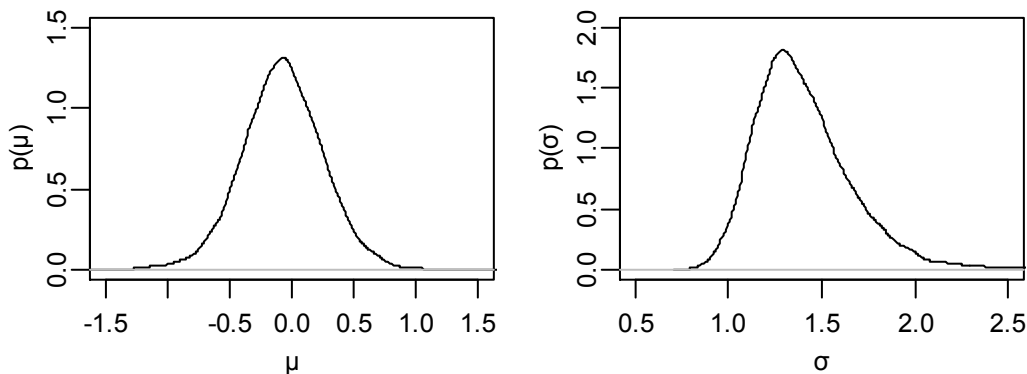


图 5-2 兴趣参数的后验分布

Fig.5-2 Posterior distribution for parameters of interest

表 5-4 中给出了兴趣参数的后验评估结果，其中，根据规则化幂先验估计的 μ 的后验标准差为 0.3153，即后验方差为 0.0994，满足样本量确定时提出的平均后验方差不大于 0.1 的要求。

表 5-4 兴趣参数的后验评估结果

Table5-4 Posterior assessment for parameter of interest

Node	mean	sd	MC_error	val2.5pc	median	val97.5pc
μ	-0.05322	0.3153	0.003142	-0.6804	-0.05023	0.5641
σ^2	1.3992	0.2397	0.002895	1.016	1.368	1.964

为了验证基于规则化幂先验的靶场试验样本量评估方法的合理性，取现场试验数据和仿真数据存在较大分布差异时的样本量评估结果进行对比。例如由表 5-3 中给出的现场试验数据加 2 模拟样本均值的分布差异，应用本文的样本量评估方法可得到兴趣参数 μ 的后验标准差为 0.3545，即后验方差为 0.1256，样本量方案不满足评估精度要求。可见，现场试验数据和仿真数据存在较大的分布差异时，本文提出的样本量评估方法能够正确识别由仿真数据构造的设计先验不合理造成的样本量计算错误。此时，可根据样本量评估结果由现场试验数据修正仿真可信度，并重新计算靶场试验的样本量。

5.5 基于规则化幂先验的靶场试验设计评估方法

靶场试验设计评估是根据得到的靶场试验数据，重新评估试验设计的合理性。鉴于现场试验的成本极高，试验设计阶段的重点是根据仿真数据进行因子效应分析，实现现场试验的最优回归设计。基于规则化幂先验的靶场试验设计

评估通过综合应用各种试验信息，评估试验设计时考虑的因子效应是否完备。

5.5.1 试验设计评估的因子效应分析模型

试验设计评估本质上是基于获取的靶场试验数据重新进行模型选择，通过与设计阶段考虑的因子效应对比，来评估试验设计是否充分。通常的模型选择方法有两类，一类是应用信息准则选择模型，还有一类方法是基于正则化选择模型。由于靶场试验获取的多源数据具有不同的信息权值，而上述方法难以实现对不同来源信息的合理加权。直接基于多源数据融合的模型选择方法进行因子效应分析时，少量的现场试验数据会被“淹没”，使因子效应分析结果由仿真试验数据主导，对靶场试验设计评估带来风险。

定义 \mathcal{M} 为模型空间，其所有的模型可由 $m=1, 2, \dots, k$ 列举出来， k 为 \mathcal{M} 的尺度。模型 k 是包含了所有可用协变量的完整模型，取所有可能的协变量个数为 p 时，则有 $k=2^p$ 。令 $\beta^{(k)} = (\beta_1, \dots, \beta_p)'$ 表示模型 k 的回归系数向量； $\beta^{(m)}$ 表示模型 m 的 $k_m \times 1$ 回归系数向量，包含了 k_m 个特定的协变量；且有 $\beta^{(k)} = (\beta^{(m)'}, \beta^{(-m)'})'$ ，其中 $\beta^{(-m)}$ 是 $\beta^{(k)}$ 删除 $\beta^{(m)}$ 得到的向量。此外，令 $\mathbf{D}^{(m)}$ 和 $\mathbf{D}_0^{(m)}$ 表示在模型 m 下的当前数据与历史数据。由模型的后验概率分析靶场试验的因子效应信息时，取模型 m 的先验概率

$$p(m) \equiv p(m | \mathbf{D}_0^{(m)}) = \frac{\int p_0^*(\beta^{(m)} | \mathbf{D}_0^{(m)}) d\beta^{(m)}}{\sum_{m \in \mathcal{M}} \int p_0^*(\beta^{(m)} | \mathbf{D}_0^{(m)}) d\beta^{(m)}} \quad (5-19)$$

其中， $p_0^*(\beta^{(m)} | \mathbf{D}_0^{(m)})$ 为模型 m 下，由历史数据得到的因子效应系数的概率密度。可见， $p(m)$ 中的分子项是在模型 m 下各因子效应系数联合先验的归一化常数。因此，式(5-19)定义的先验模型概率是一种很自然的选择，反应了贝叶斯信息更新的序贯性。式(5-19)定义的 $p(m)$ 对应由数据 $\mathbf{D}_0^{(m)}$ 应用均匀模型先验 $p_0(m) = 2^{-p}$ 得到的模型 m 的后验概率。

根据式(5-19)给出的模型先验和贝叶斯公式，在获得当前数据 $\mathbf{D}^{(m)}$ 后，可以得到模型 m 的后验概率

$$p(m | \mathbf{D}^{(m)}) = \frac{p(\mathbf{D}^{(m)} | m) p(m)}{\sum_{j=1}^k p(\mathbf{D}^{(j)} | j) p(j)} \quad (5-20)$$

其中， $p(\mathbf{D}^{(m)} | m)$ 表示在模型 m 下当前数据 $\mathbf{D}^{(m)}$ 的边缘分布。这样，根据式(5-20)选择后验概率最大的模型 m^* 作为因子效应分析结果，通过与设计阶段考虑的因子效应信息对比，就可以实现试验设计评估。

式(5-19)和(5-20)给出了试验设计评估的因子效应分析模型。由于靶场试验

数据来源广泛，既包含大量可信度较低的仿真试验数据，又有少量真实可信的现场试验数据。直接采用式(5-19)和(5-20)计算模型的后验概率时，如果不同数据的真实统计模型差异较大，会使模型选择结果出现较大的偏差，造成评估风险。因此，应用因子效应分析模型评估靶场试验设计时，应对获得的靶场试验数据进行折合加权，以构造合理的模型先验。

对比因子效应分析模型得到的结果与试验设计阶段考虑的因子效应信息，可以评估试验设计是否充分。若因子效应分析结果与设计阶段一致时，则认为靶场试验设计合理。如果得到的因子效应信息存在偏差，可根据新的因子效应信息构造试验设计的信息矩阵，并判断是否需要补充试验设计点。

5.5.2 因子效应分析模型的规则化幂先验

根据式(5-19)，基于历史数据 \mathbf{D}_0 构造模型先验 $p(m|\mathbf{D}_0^{(m)})$ 时，确定因子效应系数的先验分布 $p_0^*(\beta^{(m)}|\mathbf{D}_0^{(m)})$ 是最关键的一步。考虑到多源历史存在的不同权值特性，为了得到模型空间 \mathcal{M} 上的合理的模型先验，对模型 m 选取如下的回归系数规则化幂先验

$$p_0^*(\beta^{(m)}|\mathbf{D}_0^{(m)}) = \frac{\int_{-1}^1 \int_0^\infty \int_0^1 \pi(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}_0^{(m)}) d\delta d\sigma_b^2 d\rho}{\int_{R^{(m)}} \int_{-1}^1 \int_0^\infty \int_0^1 \pi(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}_0^{(m)}) d\delta d\sigma_b^2 d\rho d\beta^{(m)}} \quad (5-21)$$

其中，

$$\begin{aligned} \pi(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}_0^{(m)}) &\propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [f(y_{0it} | \beta^{(m)}, b_{0i})]^\delta \pi(b_{0i} | T) db_{0i} \right) \pi_0(\beta^{(m)} | c_0) \\ &\quad \times \delta^{a-1} (1-\delta)^{b-1} \cdot (\sigma_b^2)^{-(\alpha_0+1)} \exp(-\sigma_b^{-2} \omega_0) \\ &\quad \times (1+\rho)^{\phi-1} (1-\rho)^{\psi_0-1} \end{aligned}$$

由式(2-18)给出。式(5-21)定义的回归系数规则化幂先验可以根据式(5-18)对幂指数和随机效应 b_{0i} 的超参数积分得到。

式(5-19)和(5-21)构成了因子效应分析模型的规则化幂先验。靶场试验中，仿真试验成本较低，可以大量实施，通过采用足够点数的空间填充设计，能够得到较稳健的设计，对其进行评估的价值不大。而现场试验由于成本极高，试验难度和消耗都非常大，能够获得的试验点数很少，在设计时应尽量覆盖需要重点考核的因子区间。因此，靶场试验设计评估主要是针对现场试验设计的评估。根据靶场试验得到的多源数据，应用因子效应分析模型，通过判断设计时考虑的因子效应的完备性评估靶场试验设计的合理性。

5.5.3 靶场试验设计评估方法

由式(5-21)可见, $p_0^*(\beta^{(m)}|\mathbf{D}_0^{(m)})$ 与 $\beta^{(m)}$ 的边缘先验分布成比例。令 $p(\mathbf{D}^{(m)}|m)$ 表示模型 m 下, 当前数据 $\mathbf{D}^{(m)}$ 的边缘分布, 具有如下的形式

$$p(\mathbf{D}^{(m)}|m) = \int_{-1}^1 \int_0^1 \int_0^1 \int_{R^{k_m}} p(y|\beta^{(m)}, T) \pi(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}_0^{(m)}) d\beta^{(m)} d\delta d\sigma_b^2 d\rho \quad (5-22)$$

其中, $p(y|\beta^{(m)}, T)$ 可由式(2-16)对随机效应 \mathbf{b} 积分得到, $\pi(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}_0^{(m)})$ 由式(2-18)给出。可见, $p(\mathbf{D}^{(m)}|m)$ 为模型 m 下联合后验分布 $p(\beta^{(m)}, \delta, \sigma_b^2, \rho | \mathbf{D}, \mathbf{D}_0)$ 的归一化常数。这样, 式(5-20)给出的模型 m 的后验概率 $p(m|\mathbf{D}^{(m)})$ 可写为

$$p(m|\mathbf{D}^{(m)}) = \frac{p(\beta^{(-m)} = 0 | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)})}{\sum_{j=1}^k p(\beta^{(-j)} = 0 | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)})} \quad (5-23)$$

上式中, 为了标记方便, 取 $p(\beta^{(-k)} = 0 | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)}) = 1$ 。显然, 由于推导过程中的相约, 式(5-23)在代数上不依赖于先验模型概率 $p(m)$ 。这一特性使我们可以直接计算后验模型概率, 而不必在数值上计算先验模型概率。

由于模型的复杂性极高, 直接求解 $p(\beta^{(-m)} = 0 | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)})$ 的解析解很难得到。在实际计算中, 我们可以采用Chen提出的重要性加权边缘后验密度估计(IWMDE)方法^[152]来估计这些边缘后验密度。IWMDE是一种蒙特卡洛方法, 非常适合于存在联合后验密度归一化常数的边缘后验密度估计。IWMDE方法仅需要一组完整模型后验分布的MCMC样本来计算复杂的后验模型概率。根据IWMDE方法, 由MCMC仿真得到的 $p(\beta^{(-m)} = 0 | \mathbf{D}^{(k)})$ 的相容估计为

$$\hat{p}(\beta^{(-m)} = 0 | \mathbf{D}^{(k)}) = \frac{1}{B} \sum_{l=1}^B w(\beta_{(l)}^{(-m)} | \beta_{(l)}^{(m)}, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)}) \times \frac{p^*(\beta_{(l)}^{(m)}, \beta^{(-m)} = 0, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)} | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)})}{p^*(\beta_{(l)}^{(k)}, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)} | \mathbf{D}^{(k)}, \mathbf{D}_0^{(k)})}$$

其中, $w(\beta^{(-m)} | \beta^{(m)}, \delta, \sigma_b^2, \rho, b, b_0)$ 是在给定 $\beta^{(m)}, \delta, \sigma_b^2, \rho, b$ 和 b_0 的条件下, $\beta^{(-m)}$ 的已知条件密度; $p^*(\beta, \delta, \sigma_b^2, \rho, b, b_0 | \mathbf{D}, \mathbf{D}_0)$ 是完整模型的非归一化联合后验; $\{(\beta_{(l)}^{(k)}, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)}), l=1, 2, \dots, B\}$ 是联合后验分布 $p(\beta, \delta, \sigma_b^2, \rho, b, b_0 | \mathbf{D}, \mathbf{D}_0)$ 的一个MCMC样本。 w 的最优选择是在给定 $\beta^{(m)}, \delta, \sigma_b^2, \rho, b, b_0$ 下 $\beta^{(-m)}$ 的条件后验密度。对于完全分层的中心化参数情形, 在给定其他参数的条件下, $\beta^{(k)}$ 的条件后验分布是正态的, 因此能够得到最优 w 的闭合式。

获得靶场试验数据后，在上述 IWMDE 方法的支持下，本文提出如下的靶场试验设计评估方法：

- (1) 对多源验前数据 D_0 进行融合，根据式(5-18)和(5-21)构造模型 m 下的回归系数向量 $\beta^{(m)}$ 的规则化幂先验；
- (2) 根据式(5-19)和(5-21)计算模型 m 的先验概率 $p(m|D_0^{(m)})$ ；
- (3) 根据靶场试验数据，由 Gibbs 抽样算法得到 $p(\beta, \delta, \sigma_b^2, \rho, b, b_0 | D, D_0)$ 的 MCMC 样本 $\{(\beta_{(l)}^{(k)}, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)}), l=1, 2, \dots, B\}$ ；
- (4) 对 MCMC 样本 $\{(\beta_{(l)}^{(k)}, \delta_{(l)}, \sigma_{b(l)}^2, \rho_{(l)}, b_{(l)}, b_{0(l)}), l=1, 2, \dots, B\}$ 应用 IWMDE 方法计算后验模型概率，对比后验模型概率最大的模型 m^* 与试验设计时考虑的因子效应信息的一致性，完成靶场试验设计评估。

例 5.3 以 3.5 节给出的火炮射表试验为例，取表 3-3 给出的数据为仿真试验数据，表 3-5 给出的数据为现场试验数据，基于规则化幂先验对现场试验设计的有效性进行评估。

为了验证基于规则化幂先验的靶场试验设计评估方法的合理性，取仿真数据幂指数分布不同时的试验设计评估结果进行对比。应用幂指数 δ 的先验均值和方差直观地描述其分布情况，有 $\mu_\delta = a/(a+b)$ ， $\sigma_\delta^2 = \mu_\delta(1-\mu_\delta)/(a+b+1)$ 。分别取式(5-21)中幂指数 δ 的先验分布超参数为 $(a, b)=(10, 10)$ （即 $(\mu_\delta, \sigma_\delta)=(0.5, 0.11)$ ）、 $(a, b)=(20, 20)$ （即 $(\mu_\delta, \sigma_\delta)=(0.5, 0.078)$ ）和 $(a, b)=(100, 10)$ （即 $(\mu_\delta, \sigma_\delta)=(0.91, 0.027)$ ）。 σ_b^2 和 ρ 选择模糊先验，取其先验分布超参数为 $(\alpha_0, \omega_0)=(3.51, 0.005)$ 和 $\varphi_0=\psi_0=1$ 。应用 OpenBUGS 编程计算产生 MCMC 样本对后验模型概率进行估计，得到的几种幂指数分布下的后验模型概率如表 5-5 所示。

表 5-5 几种幂指数分布下的后验模型概率
Table 5-5 Posterior model probabilities for several values of $(\mu_\delta, \sigma_\delta)$

$(\mu_\delta, \sigma_\delta)$	m	$p(m D)$
(0.5, 0.11)	$(\theta_0, \omega_{x10}, \omega_{y10}, \omega_{z10})$	0.7257
(0.5, 0.078)	$(\theta_0, \omega_{x10}, \omega_{y10}, \omega_{z10})$	0.7445
(0.91, 0.027)	$(v_0, \theta_0, \omega_{x10}, \omega_{y10}, \omega_{z10})$	0.8929

由表 5-5 可见，当仿真数据的幂指数 δ 分布趋于 1 时，给予仿真数据较大的信息权值（如表 5-5 最后一行所示），模型选择结果对幂指数分布参数 $(\mu_\delta, \sigma_\delta)$ 敏感，后验模型概率较大，表示选择的模型明显占优，试验设计评估结果受仿真数据影响较大。当幂指数 δ 的分布中心为 0.5 时（表 5-5 前两行），模型选择

结果对幂指数分布参数($\mu_\delta, \sigma_\delta$)不敏感, 后验模型概率较小, 试验设计评估结果受仿真数据的影响变小。对比表 5-5 与表 3-4 的因子效应分析结果可见, 在获得足够数据的融合评估阶段, 因子效应分析结果一般都包含在设计阶段考虑的范围之内。随着获取的试验信息越多, 因子效应分析结果越明确, 即后验模型概率越大, 这也说明基于规则化幂先验的靶场试验设计评估方法的合理性。

5.6 本章小结

本章研究基于规则化幂先验的靶场试验融合评估问题。首先, 分析了靶场试验数据的分布特性, 基于 GLMM 对靶场试验的融合评估问题进行了描述, 将其归结为多源信息融合、靶场试验样本量评估和试验设计评估三个基本问题。其次, 分别考虑多源信息的交互影响和加权约束, 给出了多源信息的规则化幂先验融合方法。然后, 针对靶场试验的样本量评估, 建立了靶场试验样本量评估的后验方差分析模型, 并给出了其规则化幂先验构造方法, 应用 Gibbs 抽样算法实现了基于规则化幂先验的样本量评估。最后, 针对靶场试验设计评估, 建立了靶场试验设计评估的因子效应分析模型, 并给出了其规则化幂先验构造方法, 应用 IWMDE 方法实现了基于规则化幂先验的靶场试验设计评估。

第 6 章 某型空地导弹靶场试验的设计与评估

6.1 引言

精度是导弹武器系统的核心指标之一，也是导弹武器系统试验与鉴定中考核的重点。在实际作战过程中，精度受多种因素影响，而且制导控制系统本身也表现出较强的非线性和时变的特点，因此仅依靠小样本量的现场试验数据对其进行考核，显然不能满足装备试验鉴定的精度要求。另外，由于导弹武器系统的造价非常高，而现场飞行试验大多为破坏性试验，因此，采用大样本的现场试验与鉴定模式的消耗极大，已不能满足装备发展的要求。鉴于上述原因，综合应用仿真和现场飞行试验的导弹武器系统靶场试验与鉴定模式受到各国的重视，已成为导弹靶场试验的必然选择。

制导武器飞行试验大多为小子样情形，而均匀设计能够以较少的试验次数解决多因素、多水平数下的试验设计难题，因此，近年来国内的制导武器系统飞行试验普遍采用了小子样均匀设计方法。尽管均匀设计是一种稳健的试验设计方法，但由于这种设计方法没有充分利用仿真试验数据提供的先验信息，只是强调了试验设计点的“均匀分散”特性，不能实现对系统性能重要因子效应项的最优覆盖。因此，根据贝叶斯理论的基本观点，均匀设计并不是最优的。如何合理利用仿真试验提供的验前信息优化飞行试验设计，得到合理的靶场试验样本量分配方案，并根据获得的多源试验数据对导弹靶场试验进行融合评估，是导弹靶场试验设计与评估需要解决的关键问题。

本章将以某型空地导弹靶场试验的设计与评估为背景，基于贝叶斯正则化方法进行靶场试验设计，基于修正幂先验进行靶场试验样本量确定，基于规则化幂先验对靶场试验的样本量和试验设计方案进行评估。首先，对空地导弹靶场试验的概况进行描述，说明仿真试验与飞行试验的关联及差异，并给出靶场试验设计的基本需求信息。其次，针对靶场试验设计，分别应用空间填充设计和第 3 章给出的基于贝叶斯正则化的最优回归设计确定仿真试验和飞行试验的试验点。然后，根据仿真试验得到的先验信息，应用第 4 章给出的方法构造修正幂先验，确定在成本和评估精度约束下的靶场试验样本量分配方案。最后，根据在空地导弹靶场试验中获得的多源试验数据，应用第 5 章给出的方法构造多源信息规则化幂先验，通过后验方差分析评估空地导弹靶场试验的样本量，通过因子效应分析评估空地导弹的靶场试验设计。

6.2 空地导弹靶场试验的需求分析

空地导弹是指从飞行器上发射攻击地（水）面目标的导弹，也称为空对面导弹，其作战任务是杀伤或摧毁敌方目标，包括舰艇、运输船、海上作业平台、机场、码头、桥梁、重兵集结地、大型军事设施及工业中心等。导弹能否准确可靠地命中预定目标是完成作战任务的关键，空地导弹靶场试验主要就是对其射击精度进行考核。由于受自然环境、技术参数以及目标特性等多种误差因素影响，空地导弹即使在相同的发射条件下对同一目标进行重复射击，弹着点也不可能完全重合，图 6-1 给出了导弹落点散步的分布规律。导弹精度按其性质可分为系统误差和随机误差。其中，系统误差是指导弹实际的落点均值相对于理论落点的偏差，可由直角坐标系中的 (y_0, z_0) 或极坐标系中的 (r_0, η_0) 表示；随机误差是指导弹实际落点相对于落点均值的偏差，可用标准偏差 (σ_y, σ_z) 、圆概率偏差（CEP）或概率偏差（PE）表示。

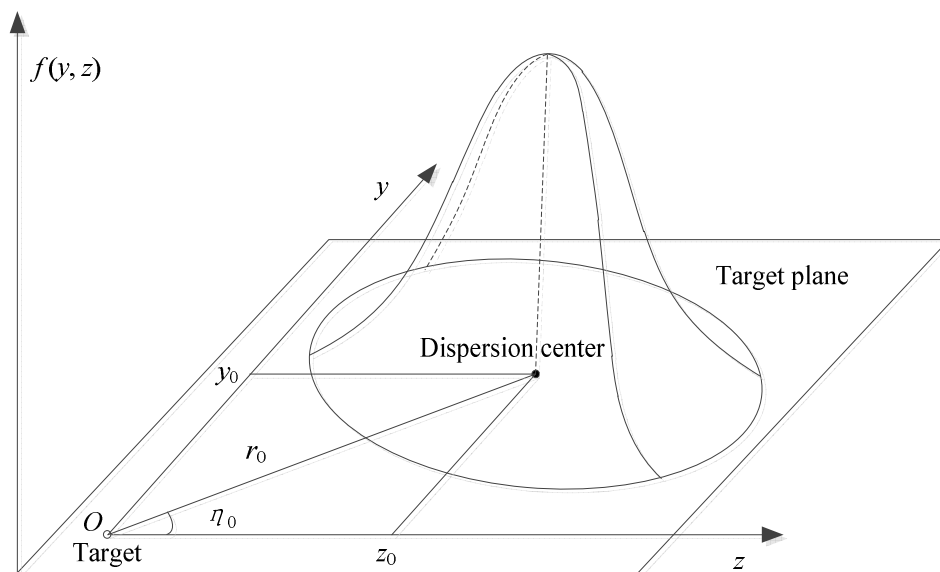


图 6-1 落点散步的分布规律

Fig.6-1 Distribution for dispersion of fall points

在已知系统误差和随机误差的情况下，导弹的纵向和横向落点可用二维正态随机分布表示

$$f(y, z) = \frac{1}{2\pi\sigma_y\sigma_z\sqrt{1-\rho_{yz}^2}} \exp \left\{ \frac{1}{2(1-\rho_{yz}^2)} \left[\frac{(y-y_0)^2}{\sigma_y^2} - \frac{2\rho_{yz}(y-y_0)(z-z_0)}{\sigma_y\sigma_z} + \frac{(z-z_0)^2}{\sigma_z^2} \right] \right\}$$

$$\rho_{yz} = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y-y_0)(z-z_0)f(y, z)dydz}{\sigma_y\sigma_z} = \frac{\text{Cov}(y, z)}{\sigma_y\sigma_z}$$

其中, y 和 z 分别表示纵向和横向的随机落点坐标; σ_y, σ_z 分别为随机变量 y 和 z 的标准差; y_0, z_0 分别为随机变量 y 和 z 的期望值; ρ_{yz} 是随机变量 y 和 z 的相关系数; $\text{Cov}(y, z)$ 为协方差。

空地导弹射击精度的影响因素主要包括: 发射距离、发射高度、发射速度、导弹平飞高度、末段自导距离、风速、风向、大气温度以及目标特性等。根据理论计算、计算机仿真、以及研制阶段的挂飞试验等分析手段, 在靶场试验阶段可以确定出对空地导弹命中精度影响较大的因子有: 发射距离 D 、发射高度 H 、发射速度 V 和末段自导距离 d 等, 将其作为靶场试验中需要重点考核的影响因子, 而将其它因素 (如导弹平飞高度、风速、风向、大气温度等) 对导弹精度的影响看作随机干扰。根据实际需要和可能的作战使用条件确定的各因子取值范围及水平数如表 6-1 所示。

表 6-1 空地导弹命中精度的影响因子及取值范围

Table 6-1 Influencing factors and their range for air-to-ground missile hit accuracy

影响因子	变量符号	取值范围	水平数
发射距离	D	30~100 km	8
发射高度	H	5~9 km	5
发射速度	V	600~800 km/h	3
末段自导距离	d	8~15 km	8

空地导弹武器系统是一个采用了多通道控制技术的非线性时变系统, 导弹精度试验是为了考核这一复杂系统在各种随机干扰下的性能指标。根据表 6-1 给出的因子水平数, 如果进行全面试验需要在 960 个因子空间点上安排试验。显然, 单纯采用现场飞行试验来鉴定导弹精度耗资巨大, 无法实现。IT&E 模式下的空地导弹精度试验通过综合应用仿真和现场试验进行导弹精度评估。试验综合设计就是在这 960 个因子空间点中选出合适的点安排仿真和现场试验, 并确定每个试验点上需要的样本量。仿真试验采用某型空地导弹仿真系统进行, 该系统由非形式化方法、静态方法、动态方法和形式化方法等 VV&A 技术综合得到的仿真可信度为 0.9。另外, 根据试验任务要求, 由靶场试验数据融合评估得到的性能评估模型对落点均值预测的平均后验方差不大于 0.1。

综上所述, 空地导弹靶场试验综合设计就是根据表 6-1 给出的试验因子取值范围及水平数、仿真系统的可信度以及融合评估的后验精度要求设计仿真和现场飞行试验的试验点并确定各个试验点上的样本量大小。靶场试验的评估是根据获得的多源数据, 对试验设计选择的试验点和样本量是否合理进行评价, 以判断靶场试验综合设计的有效性。

6.3 空地导弹靶场试验设计

空地导弹靶场试验设计包括仿真试验设计和飞行试验设计两部分。仿真试验先于飞行试验,为飞行试验提供验前信息。在仿真试验设计阶段,对系统性能的了解较少,采用空间填充设计具有较好的稳健性。飞行试验的样本量很少,应尽量覆盖重要的因子区域,因此根据由仿真数据分析得到的因子效应信息进行最优设计是一种合理的方法。

6.3.1 仿真试验设计

仿真试验设计时,确定适当的空间填充点数是一个重要的问题。实践中,一般采用多项式模型分析试验设计的因子效应,鉴于高阶模型往往会带来过拟合等不良影响,因此一般只考虑到因子的二阶项。以纵向落点为例,应用二阶多项式模型分析试验设计的因子效应时,根据表 6-1 给出的影响因子,试验设计模型的完整结构可表示为

$$y = \beta_{y0} + \beta_{y1}D + \beta_{y2}H + \beta_{y3}V + \beta_{y4}d + \beta_{y5}DH + \beta_{y6}DV + \beta_{y7}Dd + \beta_{y8}HV + \beta_{y9}Hd + \beta_{y10}Vd + \beta_{y11}D^2 + \beta_{y12}H^2 + \beta_{y13}V^2 + \beta_{y14}d^2 + \varepsilon \quad (6-1)$$

靶场试验设计的因子效应分析就是通过试验数据确定式(6-1)中对系统性能有重要影响的因子效应项。考虑二阶多项式模型时,式(6-1)的因子效应项为 15。根据理论分析及历史经验,仿真试验的空间填充设计选择 20 点即可较好的涵盖式(6-1)中所有的因子效应项。由于表 6-1 中因子间的水平数各不相同,属于混合水平试验设计,而分层拉丁超立方体抽样(Stratified Latin Hypercube Sampling, SLHS)方法非常适合于混合水平的空间填充设计,因此,基于 SLHS 方法实现空地导弹精度的仿真试验设计是一种合理的选择,具体流程如下:

(1) 设置输入因子及其水平数。根据表 6-1 设定各个因子的范围,并按照水平数划分因子区间。区间划分完成后,可以实现对各个因子区间的编号排序。

(2) 选择试验设计点数及准则。给定仿真试验的设计点数 20,选择最小距离准则进行迭代计算,可以得到仿真试验的 SLHS 设计方案。

(3) 生成并导出仿真试验的设计方案。得到 SLHS 设计方案之后,根据表 6-1 拟定好的水平对应地安排试验,并相应地把水平与因子取值范围对应好,即可实现仿真试验点设计。

按照上述设计流程,应用 Matlab 2014 中的 Model-Based Calibration Toolbox (MBC) 进行编程,可以得到空地导弹仿真试验点设计方案。应用某型空地导弹仿真系统在每个试验设计点上进行 20 次仿真试验,得到的初步仿真试验数据如表 6-2 所示。

表 6-2 初步的仿真试验数据 (样本量 $n_0 = 20$)
Table 6-2 Results of initial simulation test (sample size $n_0 = 20$)

No.	D	H	V	d	\bar{y}_i	$\hat{\sigma}_{yi}^2$	\bar{z}_i	$\hat{\sigma}_{zi}^2$
1	80	5	600	12	2.8010	1.0595	4.3245	1.0867
2	30	7	600	8	0.3667	1.0982	-0.1977	1.0057
3	100	8	700	13	5.0558	1.2655	6.3325	0.9278
4	70	6	700	9	6.5702	0.8291	7.1813	1.2261
5	40	8	700	15	-13.8480	1.0828	-13.1952	1.1026
6	90	9	700	14	0.9718	0.7933	2.2425	1.0617
7	50	9	800	15	-12.8444	0.7895	-11.6499	0.9605
8	80	8	800	13	-0.6954	1.6250	0.2660	0.8027
9	30	9	600	10	-3.6964	1.1171	-3.6830	0.6511
10	60	6	800	12	-3.2764	1.2446	-2.2439	0.8331
11	100	6	600	11	9.7355	0.9827	10.9203	0.6131
12	70	9	600	8	9.3436	0.9284	10.0397	0.6714
13	50	7	800	15	-12.9501	1.0569	-12.1504	1.2724
14	100	7	800	10	9.7678	0.7318	11.2390	1.3014
15	40	8	600	13	-7.3866	0.9830	-6.9436	0.8240
16	60	5	700	14	-6.6962	1.2327	-6.0171	0.8143
17	90	6	800	11	5.1505	1.1039	6.5133	1.4472
18	60	5	700	9	3.9828	0.7333	4.6747	1.6592
19	40	5	700	11	-4.2377	1.5053	-3.5695	0.9866
20	80	7	800	9	7.8283	0.7450	8.3784	1.1089

表 6-2 中给出的仿真试验数据为因子效应分析提供了基础。显然, 在仿真系统具有一定可信度的条件下, 由仿真试验数据获得的因子效应信息能够指导现场飞行试验设计。现场飞行试验由于消耗极高, 只能获得少量的试验样本, 根据仿真试验数据得到的因子效应信息指导现场试验设计, 能够使飞行试验的设计点位于需要重点考核的因子空间区域, 提高靶场试验的效益。

6.3.2 飞行试验设计

飞行试验在仿真试验之后进行，对其进行最优设计依据的因子效应信息是根据仿真试验数据分析得到的。因此，合理的因子效应分析是实现飞行试验最优设计的关键。以纵向落点的因子效应分析为例，基于 SLHS 设计得到初步的仿真试验数据后，飞行试验的设计流程可概括如下：

(1) 根据式(6-1)给出的试验设计模型的完整结构，对表 6-2 中的数据建立如下的统计模型

$$\begin{cases} y_{ij} \sim N(\mu_{yi}, \sigma^2) \\ \mu_{yi} = \beta_{y0} + \beta_{y1}D_i + \beta_{y2}H_i + \beta_{y3}V_i + \beta_{y4}d_i + \beta_{y5}D_iH_i + \beta_{y6}D_iV_i + \beta_{y7}D_id_i \\ \quad + \beta_{y8}H_iV_i + \beta_{y9}H_id_i + \beta_{y10}V_id_i + \beta_{y11}D_i^2 + \beta_{y12}H_i^2 + \beta_{y13}V_i^2 + \beta_{y14}d_i^2 \end{cases} \quad (6-2)$$

其中， $i=1, 2, \dots, 20$ 表示仿真试验的各个设计点， $j=1, 2, \dots, 20$ 表示各个试验设计点上的样本数据。

(2) 对式(6-2)中的参数向量 β_y 构造如下的条件 Laplace 先验

$$\pi(\beta_y | \sigma^2) = \prod_{i=0}^{14} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_{yi}|/\sqrt{\sigma^2}} \quad (6-3)$$

取正则化参数 λ 服从 Gamma(1, 0) 分布；方差 σ^2 服从无信息先验 $\pi(\sigma^2) \propto 1/\sigma^2$ 。由式(6-2)和(6-3)建立表 6-2 中仿真数据的层次化贝叶斯 Lasso 模型。

(3) 根据上述层次化贝叶斯 Lasso 模型，应用 OpenBUGS 进行编程计算，可以得到各个因子效应系数 $\beta_{yi}, i=0, 1, \dots, 14$ 的后验稀疏解。显然，不为 0 的系数 β_{yi} 对应的因子效应项对导弹精度有重要影响，飞行试验设计时应重点覆盖对导弹精度有重要影响的因子效应项。

按照上述流程对表 6-2 中的仿真数据进行因子效应分析，应用 OpenBUGS 编程进行 Gibbs 抽样计算，可以得到纵向和横向落点的贝叶斯正则化因子效应分析结果，如表 6-3 和表 6-4 所示。

表 6-3 纵向落点的贝叶斯正则化因子效应分析

Table 6-3 Analysis of effects of factors for vertical droppoints with Bayesian regularization method

变量	系数	变量	系数
(Intercept)	3.5858	H*d	-0.0149
D	0.1825	V*d	-0.0010
d	-0.1405	d^2	-0.0541
D*H	0.0054		

表 6-4 横向落点的贝叶斯正则化因子效应分析

Table 6-4 Analysis of effects of factors for lateral droppoints with Bayesian regularization method

变量	系数	变量	系数
(Intercept)	1.6215	H*d	-0.0008
D	0.2101	V*d	-0.0008
D*H	0.0038	d^2	-0.0649
H*V	-0.0001		

根据表 6-3 的分析结果，影响纵向落点的因子效应包含 7 项，因此应用最优回归设计时，可以用 7 个因子空间点实现对因子效应的覆盖。而由表 6-4 给出的横向落点的因子效应信息，同样可以用 7 个因子空间点实现对因子效应的覆盖。然而，对比表 6-3 和 6-4 可以发现因子效应信息的分析结果有一项不同。这样在现场试验设计时，可以采用的方案有两种：一种方案是选择 7 个试验点尽量实现纵向和横向落点的因子效应信息覆盖，即在试验设计的最优计算中，取两种设计的信息矩阵的折合加权，尽量使试验设计方案达到综合最优；另一种方案是取因子效应分析结果的综合，即考虑全部 8 个因子效应项实现最优设计。显然，后一种方案具有更大的稳健性。

鉴于后一种方案有更大的稳健性且试验点数仅增加一个，本文取后一种设计方案，考虑所有 8 个因子效应项“(Intercept)、D、d、D*H、H*V、H*d、V*d 和 d^2”构造试验设计模型中的基函数向量 $\mathbf{B}(\mathbf{x})$ ，由式(3-3)建立现场试验最优设计的信息矩阵 $M(\xi)$ ，应用 Matlab 2014a 中的 MBC 编程可以得到现场试验的 8 点 D-最优设计方案如表 6-5 所示。

表 6-5 现场试验的 D-最优设计

Table 6-5 D-optimal design for field test

No	D	H	V	d
1	100	9	600	15
2	30	5	800	8
3	100	9	800	11
4	30	5	700	15
5	30	9	800	8
6	30	9	600	11
7	100	9	600	8
8	30	9	800	15

由此，我们就得到了空地导弹靶场试验的试验点设计方案。其中，仿真试验设计采用 SLHS 充分挖掘影响系统性能的因子效应信息，为现场试验的最优设计提供合理依据。现场试验设计时，首先基于贝叶斯正则化因子效应分析确定试验设计信息矩阵；然后根据试验设计信息矩阵实现最优回归设计。

6.4 空地导弹靶场试验的样本量确定

修正幂先验的构造是确定靶场试验样本量最关键的内容。由表 6-2 给出的初步仿真试验结果可以得到试验设计点上的仿真样本方差均值。结合试验需求分析中给出的仿真系统可信度，由式(4-11)和式(4-12)可以计算仿真试验的等效样本量并得到构造修正幂先验的幂指数。在修正幂先验框架下考虑试验成本和评估精度约束，可以确定出靶场试验需要的样本量。

6.4.1 落点均值的修正幂先验构造

为了构造靶场试验样本量确定的修正幂先验，首先根据仿真可信度和样本量确定仿真试验的最大等效样本量，然后根据仿真试验的最大等效样本量，在综合考虑试验成本和评估精度约束的条件下确定仿真和现场试验的样本量分配方案。根据式(4-11)给出的等效样本量计算公式

$$n_r = \frac{(n_0 - 3)s_r^2}{s_0^2 - (n_0 - 3)\ln(c)} + 3, n_0 > 3 \quad (6-4)$$

由表 6-2 给出的初步仿真试验数据，给定仿真可信度 $c=0.9$ 的条件下，由式(6-4)可以得到等效样本量 n_r 与仿真样本量 n_0 的关系，如图 6-2 所示。

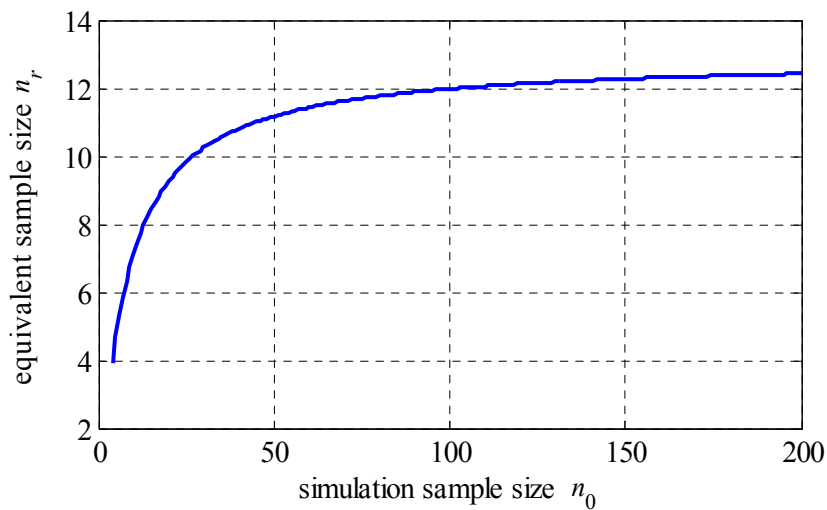


图 6-2 仿真试验的等效样本量

Fig.6-2 Equivalent sample size of simulation test

由图 6-2 可见，仿真试验的等效样本量随仿真样本量增加趋于一个有限值 12.8957。式(4-12)定义的修正幂指数为等效样本量与验前样本量之比，因此，修正幂指数将随着仿真样本量的增大趋于 0，不会出现仅由大量仿真试验数据构造的设计先验就已满足评估精度要求的情况。

在现场试验数据和仿真试验数据满足方差齐性的假设下，有 $s_r^2 = s_0^2 = 1.0426$ 。对表 6-2 中第一个设计点上的纵向落点数据，给定仿真可信度 $c=0.9$, $n_0=20$ 时，由式(6-4)计算的等效样本量 $n_r=9.2547$ ，由式(4-12)可知修正幂指数 $\delta_m=0.4627$ 。图 6-3 中给出了应用 OpenBUGS 构造的试验设计点上纵向落点均值的修正幂先验、标准幂先验和完全应用仿真数据得到的共轭先验。显然，修正幂先验与标准幂先验和共轭先验相比具有更大的尾部，避免了单独由仿真数据即可满足评估精度要求的风险。

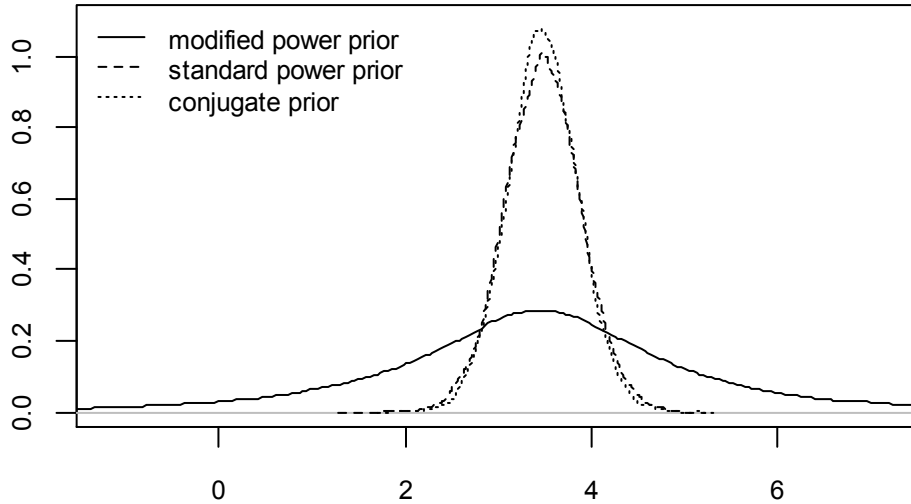


图 6-3 纵向落点均值的修正幂先验

Fig.6-3 Modified power prior of mean vertical droppoints

由第 4 章的引理 4.1 和引理 4.2 可知，根据仿真试验数据构造的修正幂先验等价于由等效样本量大小的现场试验数据和无信息先验得到的标准贝叶斯后验分布。根据构造的落点均值的修正幂先验，在试验成本和评估精度要求约束下，可以得到导弹精度靶场试验的样本量方案。

6.4.2 靶场试验的样本量确定

根据式(4-18)，由验前样本的等效样本量和验前样本方差可以计算均值参数的平均后验方差

$$E[\text{var}(\mu | \mathbf{x}_n)] = \frac{n_r s_0^2}{(n + n_r)(n_r - 3)}, n_r > 3$$

可见，单纯应用仿真数据能够得到的落点均值的最小平均后验方差（APV）为 0.1054。当要求的评估精度（取 APVC）高于 0.1054 时，单纯应用仿真试验已不能满足要求，此时应考虑综合应用仿真和现场试验的方案。当要求的评估精度较高时，由于仿真可信度的限制，在试验效应指标等价的约束下，即使仿真试验的样本量再大，也不能无限制的提高评估精度。而且，如果评估精度的要求非常高，较低的仿真可信度甚至会对评估带来不良影响。

根据式(4-26)给出的采用综合设计方案的优良性条件，

$$\frac{n_r s_0^2}{(n+n_r)(n_r-3)} < \frac{s_0^2}{n-3}$$

当 $n_r > \sqrt{3n}$ 时，采用综合设计方案具有优良性；当 $n_r \leq \sqrt{3n}$ 时，综合设计方案不如单纯采用现场试验进行评估。由于仿真试验的等效样本量上限为 12.8957，因此，当现场试验样本量 $n > 55$ 时，综合设计方案将不再具有优良性。此时，落点均值的平均后验方差为 0.0197。

由以上分析可见，当落点均值的平均后验方差精度要求低于 0.1054 时，仅由仿真样本即可实现评估。当评估的后验方差精度要求高于 0.1054 时而低于 0.0197 时，应综合应用仿真和飞行试验样本。而当评估的后验方差精度要求高于 0.0197 时，根据综合设计方案的优良性条件，此时先验样本的可信度会对评估产生不良影响，应仅使用现场试验样本进行评估。

6.2 节给出的试验需求中要求评估模型对落点均值预测的平均后验方差不大于 0.1。仿真试验样本的方差 $s_0^2 = 1.0426$ ，在方差齐性假设下，单纯采用现场试验进行评估时每个试验点上需要的样本量为 14。由于仿真试验的最大等效样本量为 12.8957，为了满足评估精度要求，需要的现场试验最小样本量为 2。表 6-6 给出了考虑成本及评估精度约束时，由式(4-27)得到的现场试验样本量为 2 和 3 时的综合设计方案，以及单纯采用现场试验时的样本量方案。

表 6-6 靶场试验的样本量方案

Table 6-6 Sample size plan of missile hit accuracy test

样本量设计 方案	仿真样本量 n_0	等效样本量 n_r	现场试验样本量 n	均值 μ 的 平均后验方差
仿真+现场	95	12.2768	2	0.0999
仿真+现场	60	11.7359	3	0.0979
纯现场试验	-	-	14	0.0948

仿真可信度是根据仿真试验结果与历史数据、地面测试数据和专家经验的对比分析给出的一个综合估计量，难免存在度量上的偏差。鉴于此，为了保证试验评估结果的稳健性，在成本允许的范围内，不妨适当增加现场试验样本量，

取试验设计点上仿真样本量为 60，飞行试验样本量为 3 的样本量方案。此时，落点均值的平均后验方差为 0.0979，满足评估精度要求。

6.5 空地导弹的靶场试验评估

空地导弹靶场试验的融合评估是为了验证试验综合设计的有效性，对试验样本量以及试验设计点进行的评估。根据前两节给出的空地导弹仿真试验和现场飞行试验的试验设计方案和样本量确定方案，实施试验并获取数据后，可以应用第 5 章给出的方法对空地导弹靶场试验进行融合评估。

6.5.1 因子效应系数的规则化幂先验构造

根据 6.4.2 节的靶场试验样本量确定方案，仿真试验样本量取 60。因此，需要在表 6-2 给出的初步仿真试验数据（样本量为 20）的基础上，继续在每个试验设计点上进行 40 次独立试验，表 6-7 给出了仿真试验在各个试验设计点上的样本均值及样本方差。现场试验的样本量为 3，按照表 6-5 给出的试验点设计方案进行现场试验，得到的现场试验数据如表 6-8 所示。

表 6-7 完整的仿真试验数据（样本量 $n_0 = 60$ ）

Table 6-7 Results of total simulation test (sample size $n_0 = 60$)

No.	D	H	V	d	\bar{y}_i	$\hat{\sigma}_y^2$	\bar{z}_i	$\hat{\sigma}_{zi}^2$
1	80	5	600	12	3.3814	0.8802	4.2948	1.0734
2	30	7	600	8	0.8718	0.9433	0.2534	1.3879
3	100	8	700	13	5.4448	1.2776	6.4674	0.9011
4	70	6	700	9	6.6039	0.7506	6.8934	1.2704
5	40	8	700	15	-12.9826	0.8635	-13.0448	0.9330
6	90	9	700	14	1.2077	1.0100	2.1180	1.2200
7	50	9	800	15	-12.2621	0.8312	-12.2913	1.1471
8	80	8	800	13	-0.0678	0.9405	0.2585	0.9772
9	30	9	600	10	-3.0315	0.9119	-3.9978	0.9518
10	60	6	800	12	-2.7434	0.9064	-2.5009	1.2039
11	100	6	600	11	10.1834	0.9144	10.8237	0.8087
12	70	9	600	8	9.6918	1.0124	10.2369	0.9306
13	50	7	800	15	-12.0721	1.2708	-12.1999	1.0424

表 6-7（续表）

No.	D	H	V	d	\bar{y}_i	$\hat{\sigma}_{yi}^2$	\bar{z}_i	$\hat{\sigma}_{zi}^2$
14	100	7	800	10	10.5304	1.0395	11.4758	0.8595
15	40	8	600	13	-6.9863	0.6344	-7.1851	1.0704
16	60	5	700	14	-6.5266	1.0849	-5.8986	0.7430
17	90	6	800	11	5.9227	1.0412	6.3561	0.8831
18	60	5	700	9	4.7039	1.1446	4.6987	1.0878
19	40	5	700	11	-3.6786	1.0547	-3.4867	1.1542
20	80	7	800	9	8.2733	1.2549	8.7928	0.8420

表 6-8 现场试验结果（样本量 $n = 3$ ）
Table 6-8 Results of field test (sample size $n = 3$)

No.	D	H	V	d	\bar{y}_i	$\hat{\sigma}_{yi}^2$	\bar{z}_i	$\hat{\sigma}_{zi}^2$
1	100	9	600	15	4.2374	1.5089	1.4343	0.5566
2	30	5	800	8	1.4908	0.8947	-1.3451	0.8028
3	100	9	800	11	8.7754	2.2837	7.5697	2.6441
4	30	5	700	15	-12.8445	1.2469	-14.9922	0.7021
5	30	9	800	8	-0.6935	0.4390	-2.6177	0.4140
6	30	9	600	11	-5.9443	0.8303	-6.9272	0.6150
7	100	9	600	8	17.4797	0.9965	16.5221	1.5548
8	30	9	800	15	-16.2659	0.7989	-18.8555	1.3411

靶场试验样本量评估是通过靶场试验数据得到的后验评估精度来判断样本量是否满足要求。根据表 6-3 和表 6-4 给出的贝叶斯正则化因子效应分析结果，空地导弹的精度评估模型可表示为如下的结构形式

$$\begin{cases} E(y) = \beta_{y0} + \beta_{y1}D + \beta_{y2}d + \beta_{y3}DH + \beta_{y4}Hd + \beta_{y5}Vd + \beta_{y6}d^2 \\ E(z) = \beta_{z0} + \beta_{z1}D + \beta_{z2}DH + \beta_{z3}HV + \beta_{z4}Hd + \beta_{z5}Vd + \beta_{z6}d^2 \end{cases} \quad (6-5)$$

根据式(5-14)，应用落点均值的后验方差描述评估精度时，可由式(6-5)中系数向量 $\beta_y=(\beta_{y0}, \cdots, \beta_{y6})^T$ 和 $\beta_z=(\beta_{z0}, \cdots, \beta_{z6})^T$ 的后验估计协方差矩阵计算落点均值的后验估计方差 $\text{var}[E(y|\mathbf{D}_0, \mathbf{D})]$ 和 $\text{var}[E(z|\mathbf{D}_0, \mathbf{D})]$ ，通过与后验评估的精度要求对比实现对导弹精度试验样本量的评估。

空地导弹试验设计评估本质上是根据靶场试验数据重新评价设计阶段考虑的因子效应项是否完备。由于评估综合应用了仿真和飞行试验数据,因此需要应用规则化幂先验对仿真数据进行加权。根据式(5-21)构造因子效应分析的规则化幂先验时,考虑式(6-1)给出的完整模型结构确定系数向量 β_y 和 β_z 。

由上述分析可见,构造系数向量 β_y 和 β_z 的规则化幂先验是实现导弹精度试验融合评估的关键。根据试验需求分析中给出的仿真系统可信度 0.9,由表 6-7 可知仿真样本方差的估计值 $\hat{s}_0^2 = 1.1018$,由表 6-8 可知现场样本方差的估计值 $\hat{s}_r^2 = 1.0064$,则根据式(4-12)可以得到修正幂指数 $\delta_m = 0.1993$ 。由于得到的修正幂指数 δ_m 为固定值,因此根据 5.4.2 节的式(5-18)构造导弹精度试验融合评估的规则化幂先验时可以消去幂指数的先验分布 $\pi(\underline{\delta})$,有

$$\pi(\beta | D_0, T, \delta) \propto \prod_{i=1}^{N_0} \left(\int_{R^q} \prod_{t=1}^{n_{0i}} [p(y_{0it} | \beta, b_{0i})]^\delta \pi(b_{0i} | T) db_{0i} \right) \pi(\beta) \quad (6-6)$$

其中, b_{0i} 为随机效应系数, $N_0 = 20$ 为仿真试验的设计点数, $n_{0i} = 60$ 为各个设计点上的样本量, T 为随机效应的协方差矩阵, δ 为幂指数。 y_{0it} 表示仿真试验在第 i 个设计点上的第 t 次测量值。

根据表 6-7 和表 6-8 给出的靶场试验数据,导弹精度试验融合评估的规则化幂先验可通过指定式(6-6)中 T 的参数先验得到。假定 T 由参数 σ_b^2 和 ρ 决定,取 σ_b^2 的先验为逆 Gamma 分布,超参数 $(\alpha_0, \omega_0) = (3.51, 0.005)$;取 ρ 为 $[-1, 1]$ 上的均匀分布时,其尺度 Beta 先验的超参数 $\varphi_0 = \psi_0 = 1$ 。图 6-4 给出了应用 OpenBUGS 构造的因子效应系数 β_{y1} 的规则化幂先验及其后验分布。

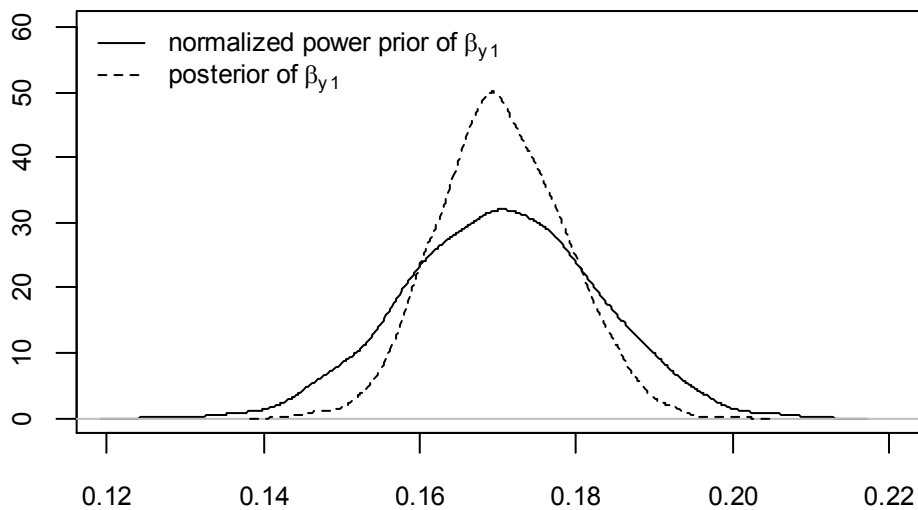


图 6-4 试验融合评估中系数 β_{y1} 的规则化幂先验及其后验分布

Fig.6-4 Normalized power prior and posterior of β_{y1} in test fusion and assessment

6.5.2 靶场试验样本量评估

根据上一节构造的回归系数向量 β_y 和 β_z 的规则化幂先验,应用 Gibbs 抽样计算可以得到回归系数的后验估计及其协方差矩阵,用于靶场试验样本量评估。图 6-5 给出了导弹精度试验评估的 BUGS 语言描述,其中, **beta** 表示模型参数向量, **X** 表示由表 6-3 或 6-4 中的因子效应项构成的协变量向量, **b** 为随机效应系数, rou、b.tau 和 tau.c 为超参数, mu 为各试验设计点上的落点均值, N 表示试验设计点数, n 表示各个试验设计点上的样本量。

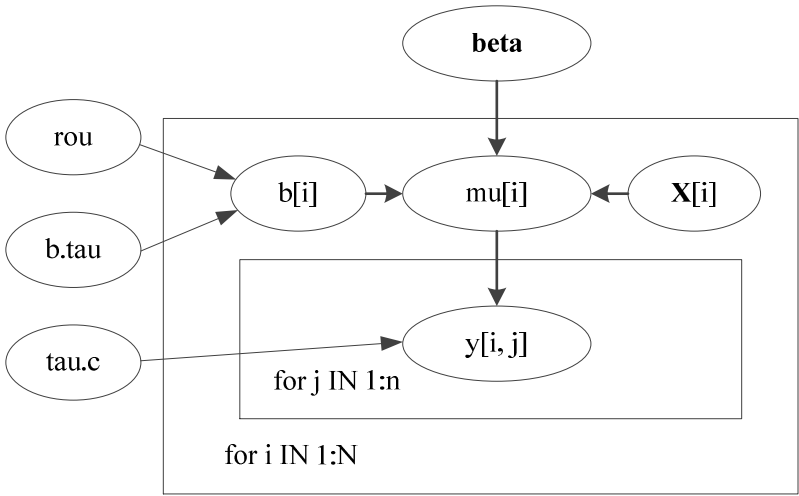


图 6-5 导弹精度试验评估的 BUGS 语言描述

Fig.6-5 BUGS language for missile accuracy test assessment

根据图 6-5 给出的 BUGS 语言描述,应用 OpenBUGS 进行 Gibbs 抽样计算,得到落点因子效应系数的后验推断如表 6-9~表 6-12 所示。

表 6-9 纵向落点因子效应系数的后验估计

Table 6-9 Posterior estimation of factor effects coefficients for vertical dropoints

Fixed effects coefficients (95% CIs):					Res Std		
Name	Estimate	SE	Lower	Upper	Estimate	Lower	Upper
(Intercept)	4.9379	1.2583	2.6716	7.2041	0.1974	0.1519	0.2565
D	0.1708	0.0086	0.1552	0.1863			
d	-0.24	0.2470	-0.6848	0.2047			
D*H	0.0073	0.0011	0.0054	0.0092			
H*d	-0.0284	0.0055	-0.0382	-0.0185			
V*d	-9.4865e-4	4.61E-05	-0.0010	-8.66E-04			
d^2	-0.0452	0.0102	-0.0636	-0.0268			

表 6-10 纵向落点因子效应系数的协方差矩阵

Table 6-10 Covariance matrix of factor effects coefficients for vertical droppoints

	β_{y0}	β_{y1}	β_{y2}	β_{y3}	β_{y4}	β_{y5}	β_{y6}
β_{y0}	1.1876	0.0034	-0.2267	-3.81E-04	0.0012	-4.58E-06	0.0095
β_{y1}	0.0034	5.60E-05	-8.95E-04	-6.78E-06	3.05E-05	-3.67E-08	3.03E-05
β_{y2}	-0.2267	-8.95E-04	0.0457	1.01E-04	-3.94E-04	-4.34E-08	-0.0018
β_{y3}	-3.81E-04	-6.78E-06	1.01E-04	8.56E-07	-3.92E-06	4.62E-09	-3.27E-06
β_{y4}	0.0012	3.05E-05	-3.94E-04	-3.92E-06	2.24E-05	-1.76E-08	1.04E-05
β_{y5}	-4.58E-06	-3.67E-08	-4.34E-08	4.62E-09	-1.76E-08	1.59E-09	-4.51E-08
β_{y6}	0.0095	3.03E-05	-0.0018	-3.27E-06	1.04E-05	-4.51E-08	7.83E-05

表 6-11 横向落点因子效应系数的后验估计

Table 6-11 Posterior estimation of factor effects coefficients for lateral droppoints

Fixed effects coefficients (95% CIs):					Res Std		
Name	Estimate	SE	Lower	Upper	Estimate	Lower	Upper
(Intercept)	3.5299	0.2546	3.0003	4.0594	0.1712	0.1318	0.2225
D	0.1651	0.0062	0.1522	0.1779			
D*H	0.0104	7.83E-04	0.0088	0.0120			
H*V	-3.93E-04	1.20E-04	-6.42E-04	-1.44E-04			
H*d	-0.0256	0.0064	-0.0390	-0.0123			
V*d	-7.26E-04	8.24E-05	-8.98E-04	-5.55E-04			
d^2	-0.0613	0.0043	-0.0702	-0.0524			

表 6-12 横向落点因子效应系数的协方差矩阵

Table 6-12 Covariance matrix of factor effects coefficients for lateral droppoints

	β_{z0}	β_{z1}	β_{z2}	β_{z3}	β_{z4}	β_{z5}	β_{z6}
β_{z0}	0.0648	-0.0011	1.36E-04	-1.55E-05	1.39E-04	6.06E-06	-2.62E-04
β_{z1}	-0.0011	3.79E-05	-4.73E-06	3.59E-07	1.54E-06	-2.53E-07	7.42E-06
β_{z2}	1.36E-04	-4.73E-06	6.13E-07	-4.42E-08	-3.74E-07	3.12E-08	-8.51E-07
β_{z3}	-1.55E-05	3.59E-07	-4.42E-08	1.43E-08	-6.22E-07	-8.95E-09	4.70E-07
β_{z4}	1.39E-04	1.54E-06	-3.74E-07	-6.22E-07	4.13E-05	3.75E-07	-2.46E-05
β_{z5}	6.06E-06	-2.53E-07	3.12E-08	-8.95E-09	3.75E-07	6.79E-09	-3.29E-07
β_{z6}	-2.62E-04	7.42E-06	-8.51E-07	4.70E-07	-2.46E-05	-3.29E-07	1.82E-05

根据表 6-10 和表 6-12 给出的因子效应系数协方差矩阵,应用式(5-14)计算表 6-2 和表 6-5 给出的仿真和现场试验设计点上的纵向和横向落点均值的预测方差,得到的纵向落点均值的后验方差不超过 0.0204,横向落点均值的后验方差不超过 0.0208。显然满足空地导弹精度试验需求中提出的后验方差不大于 0.1 的要求。由式(5-14)和因子效应系数协方差矩阵可以进一步计算整个因子空间中预测点上的兴趣参数后验方差,通过计算可发现由靶场试验数据得到的落点均值预测方差不大于 0.076。可见,6.4 节给出的靶场试验样本量确定方案能够满足评估精度要求。

单独应用表 6-8 的实测数据进行 Gibbs 抽样计算,得到因子效应系数的后验估计及其协方差矩阵。由式(5-14)可以计算出整个因子空间中纵向和横向落点均值的预测后验方差分别为 0.1734 和 0.2316,不能满足 6.1 节中落点均值预测的平均后验方差不大于 0.1 的要求。可见,采用仿真与现场试验结合的一体化试验模式,能够提高评估精度,减少试验消耗。

在对靶场试验样本量进行评估的基础上,可确认靶场试验样本量的充分性,对试验设计进行评估。试验设计评估是根据获得的靶场试验数据重新分析试验的因子效应信息,以判断设计阶段考虑的因子效应信息是否充分,评估的关键是应用规则化幂先验对仿真数据进行加权约束,避免大量的仿真试验数据主导因子效应分析结果。

6.5.3 靶场试验设计评估

空地导弹靶场试验设计评估是在试验样本量评估的基础上,根据靶场试验数据进一步分析飞行试验设计阶段考虑的因子效应信息是否完备,对靶场试验设计的有效性做出评价,能够确保靶场试验数据用于导弹精度评估时的合理性。应用表 6-7 和表 6-8 给出的靶场试验数据进行试验设计评估时,模型选择是需要解决的首要问题。靶场试验设计阶段考虑的空地导弹精度纵向落点的完整二次结构模型为

$$y = \beta_{y0} + \beta_{y1}D + \beta_{y2}H + \beta_{y3}V + \beta_{y4}d + \beta_{y5}DH + \beta_{y6}DV + \beta_{y7}Dd + \beta_{y8}HV + \beta_{y9}Hd + \beta_{y10}Vd + \beta_{y11}D^2 + \beta_{y12}H^2 + \beta_{y13}V^2 + \beta_{y14}d^2 + \varepsilon_y \quad (6-7)$$

$$z = \beta_{z0} + \beta_{z1}D + \beta_{z2}H + \beta_{z3}V + \beta_{z4}d + \beta_{z5}DH + \beta_{z6}DV + \beta_{z7}Dd + \beta_{z8}HV + \beta_{z9}Hd + \beta_{z10}Vd + \beta_{z11}D^2 + \beta_{z12}H^2 + \beta_{z13}V^2 + \beta_{z14}d^2 + \varepsilon_z \quad (6-8)$$

可见,完整二次结构模型中协变量个数 $p=14$,由此可以计算出评估模型空间的尺度 $k=2^p=16384$,也就是说共有 16384 个可选的模型。取表 6-7 给出的仿真试验数据为 D_0 ,应用式(5-19)和(5-21)可以得到任一模型 m 的先验概率

$$p(m) \equiv p(m | \mathbf{D}_0^{(m)}) = \frac{\int p_0^*(\beta^{(m)} | \mathbf{D}_0^{(m)}) d\beta^{(m)}}{\sum_{m \in \mathcal{M}} \int p_0^*(\beta^{(m)} | \mathbf{D}_0^{(m)}) d\beta^{(m)}} \quad (6-9)$$

取系数向量 β_y 和 β_z 的规则化幂先验计算式(6-9)给出的模型先验, 结合表 6-8 的现场试验数据 \mathbf{D} , 根据式(5-19)~(5-23)给出的因子效应分析模型及其规则化幂先验应用 R 和 OpenBUGS 进行编程计算, 可以得到导弹精度试验的后验模型概率, 如表 6-13 和表 6-14 所示。

表 6-13 导弹精度试验纵向落点的后验模型概率
Table 6-13 Posterior model probabilities of vertical droppoints

No	model	prob
1	(D, D*H, H*d, V*d, d^2)	0.7186
2	(D, H, D*H, V*d, d^2)	0.1740
3	(D, d, D*H, H*V, V*d, H^2, d^2)	0.0356

表 6-14 导弹精度试验横向落点的后验模型概率
Table 6-14 Posterior model probabilities of lateral droppoints

No	model	prob
1	(D, D*H, H*V, V*d, d^2)	0.7313
2	(D, H, d, D*H, V*d, d^2)	0.1226
3	(D, H, D*H, H*V, V*d, d^2)	0.0224

由表 6-13 可见, 根据靶场试验数据得到的最可能的后验模型中因子效应项主要包括 D、D*H、H*d、V*d 和 d^2, 与表 6-3 中设计阶段考虑的因子效应相比少了一项 d。因此, 现场飞行试验设计时考虑的因子效应是足够的, 能够涵盖实际的因子效应信息。表 6-14 中横向落点的因子效应分析结果同样说明了这一结论。应注意的是, 当仿真可信度较低时, 在设计阶段基于仿真数据得到的因子效应信息可能与实际情况并不一致, 此时, 基于规则化幂先验的试验设计评估能够判断靶场试验设计方案对重要因子效应项的覆盖性。

单独应用表 6-8 给出的现场试验数据进行分析时, 取幂指数为 1, $\mathbf{D}_0^{(m)} = \mathbf{D}$, 根据式(6-9)应用 R 和 OpenBUGS 进行编程计算, 可以得到导弹精度试验的后验模型概率如表 6-15 和表 6-16 所示。

表 6-15 导弹精度试验纵向落点的后验模型概率

Table 6-15 Posterior model probabilities of vertical droppoints

No	model	prob
1	$(D, d, D^*H, H^*d, V^*d, d^2)$	0.3385
2	$(D, d, H, D^*H, V^*d, d^2)$	0.2761
3	$(D, D^*H, H^*V, V^*d, H^2)$	0.1564

表 6-16 导弹精度试验横向落点的后验模型概率

Table 6-16 Posterior model probabilities of lateral droppoints

No	model	prob
1	$(D, D^*H, H^*V, H^*d, V^*d, d^2)$	0.3173
2	$(D, H, d, D^*H, H^*V, V^*d, d^2)$	0.2956
3	$(D, H, D^*H, H^*V, H^*d, V^*d, d^2)$	0.2621

表 6-15 和表 6-16 中的现场试验数据分析结果与试验设计阶段考虑的因子效应信息一致,说明了基于贝叶斯正则化的靶场试验设计方法的合理性。此外,表 6-15 和表 6-16 中的后验模型概率明显低于表 6-13 和表 6-14 中的结果,这是由于缺少了仿真数据的信息,降低了因子效应分析结果的确性。

在靶场试验的融合评估阶段,根据获得的仿真和现场试验数据对空地导弹靶场试验的样本量和试验设计进行融合评估,可以判断靶场试验的样本量是否足够、试验设计点是否合理。在样本量和试验设计均已通过评估的基础上,可以确保应用靶场试验数据对装备性能进行评估的科学性和合理性。

6.6 本章小结

本章以空地导弹靶场试验的设计与评估为背景,给出了前文提出的试验设计方法、样本量确定方法和融合评估方法的应用示例。首先,介绍了空地导弹靶场试验的概况及需求。其次,应用分层拉丁超立方体抽样实现了仿真试验设计,根据贝叶斯正则化因子效应分析结果构造试验设计信息矩阵,给出了现场试验的 D-最优设计。再次,考虑仿真可信度和样本量的影响,构造了落点均值的修正幂先验,给出了考虑成本和评估精度约束的样本量确定方案。然后,根据空地导弹试验的多源数据构造了因子效应系数的规则化幂先验,通过后验方差和因子效应分析实现了对样本量和试验设计的评估。

结 论

IT&E 模式下的靶场试验综合设计与评估方法是当前试验与鉴定领域中的研究热点。本文基于贝叶斯理论,系统深入地研究了 IT&E 模式下的靶场试验设计、靶场试验样本量确定和靶场试验融合评估等三个基本问题。取得的主要研究成果如下:

(1) 在靶场试验综合设计与评估的贝叶斯理论基础方面,讨论了靶场试验中多源验前信息的不同权值特性。针对权值固定的验前信息,给出了标准幂先验构造方法,根据兴趣参数后验分布的 KL 散度凸性和 Zellner 信息处理准则证明了其优良性;针对权值服从随机分布的验前信息,给出了规则化幂先验构造方法,并根据幂指数对历史数据和当前数据的加权互信息证明了其优良性。

(2) 在靶场试验设计方面,通过对试验设计问题的描述,指出了因子效应信息在试验设计中的重要作用。分析了正则化方法在因子效应分析中的合理性,提出了贝叶斯正则化因子效应分析方法。对于有因子效应信息试验,基于贝叶斯正则化因子效应分析结果构造试验设计信息矩阵,提出了一种兼顾可能因子效应项的附加点最优设计方法。通过火炮射表试验的示例验证了基于贝叶斯正则化的靶场试验设计方法的有效性。

(3) 在靶场试验样本量确定方面,通过分析设计先验对兴趣参数后验分布的影响,以及仿真和现场试验样本量在后验估计中的作用,给出了靶场试验样本量确定问题的描述。考虑验前样本量对兴趣参数先验分布的影响,定义了试验效应指标,构造了基于试验效应等价的修正幂先验。根据贝叶斯样本量计算准则以及兴趣参数的预后验分布特性,提出了基于修正幂先验的靶场试验样本量确定方法,解决了靶场试验综合方案设计中的样本量确定难题。

(4) 在靶场试验融合评估方面,通过对靶场试验融合评估问题的描述,将其归结为多源信息融合、样本量评估和试验设计评估三个基本问题。考虑多源信息的交互影响和加权约束,给出了多源信息的规则化幂先验融合方法。对于靶场试验的样本量评估,提出了基于规则化幂先验的后验方差分析方法。对于靶场试验设计评估,提出了基于规则化幂先验的因子效应分析方法。

(5) 将上述研究成果应用到某型空地导弹靶场试验的综合设计与评估中,分别进行了导弹靶场试验设计和样本量确定,通过对比分析试验设计阶段和融合评估阶段的兴趣参数后验方差及因子效应信息,对空地导弹靶场试验综合设计的有效性进行了评估,由此验证了本文提出方法的合理性和有效性。

对于以上研究结果，论文的主要创新点总结如下：

(1) 针对靶场试验设计问题，通过分析正则化在实现稳健因子效应分析中的优势，提出了基于贝叶斯正则化的靶场试验设计方法，解决了靶场试验设计中合理进行因子效应分析的难题，给出了考虑因子效应信息的附加点最优回归设计方法，实现了效益最优的靶场试验点综合设计。

(2) 针对靶场试验的样本量确定问题，通过分析验前样本量对标准幂先验构造设计先验的影响，给出了修正幂先验的定义，提出了基于修正幂先验的贝叶斯样本量确定方法，解决了应用标准幂先验确定样本量时存在的可信度风险与评估精度之间的矛盾，实现了试验成本约束和评估精度约束下的靶场试验综合方案样本量优化设计。

(3) 针对靶场试验的融合评估问题，给出了基于多源信息规则化幂先验的贝叶斯融合评估方法。构造了考虑多源信息交互影响和加权约束的规则化幂，提出了用于样本量评估的后验方差分析方法和用于试验设计评估的因子效应分析方法，改进了靶场试验评估时的多源历史数据融合结构。

后续研究工作的展望与设想：

(1) 本文研究的试验设计方法主要是针对参数估计型的靶场试验，且考虑的分布形式为正态分布，在后续研究中还将考虑试验数据具有其它分布形式的靶场试验设计问题。

(2) 本文提出的样本量确定方法主要针对性能参数试验，在以后的研究中将探讨寿命试验和可靠性试验的贝叶斯样本量确定问题，并基于历史试验数据构造考虑数据动态分布差异的修正幂先验。

(3) 本文研究的靶场试验融合评估问题，针对的是靶场试验综合设计方案样本量的充分性和试验设计点的合理性，以后的研究中将针对装备系统效能和作战效能在全因子空间中的动态分布情况进行评估。

参考文献

- [1] 武小悦, 刘琦. 装备试验与评价[M]. 北京: 国防工业出版社, 2008: 500.
- [2] Claxton J D, Cavoli C, Johnson C. Test and Evaluation Management Guide[M]. 6th Edition ed. Fort Belvoir, VA: The defense acquisition university press, 2012: 318.
- [3] Hagan G. Glossary of defense acquisition acronyms & terms[M]. 13th Edition ed. Fort Belvoir, VA: The defense acquisition university press, 2009: 242.
- [4] Kraft E M. Integrated Test and Evaluation- A knowledge-based approach to system development: 1st AIAA Aircraft Engineering, Technology, and Operations Congress[C]. Los Angeles, AIAA 95-3982.
- [5] DOD Instruction 5000.2 Operation of the defense acquisition system[R]. Washington, US. Department of Defense, 2008.
- [6] 雷帅, 丁士民. 基于一体化试验鉴定发展的试验鉴定发展规律[J]. 国防科技. 2011, 32(4): 36-39.
- [7] Milton D, Alan H, Jason K, et al. Demonstration of an Integrated Test and Evaluation (IT&E) Process for Airframe-Propulsion Systems as applied to a Current Weapon System Program[C]. 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Orlando, Florida, AIAA 2010-1039.
- [8] Skelley M L, Langham T F, Peters W L. Integrated test and evaluation for the 21st century[C]. USAF Developmental Test and Evaluation Summit, Woodland Hills, California, AIAA 2004-6873.
- [9] 赵世明, 王江云, 费惠佳. 导弹综合试验与评估方法研究[J]. 战术导弹技术. 2012(2): 50-54.
- [10] 杨磊, 武小悦. 美军装备一体化试验与评价技术发展[J]. 国防科技. 2010, 31(02): 8-14.
- [11] 张宝珍. 综合试验与评价策略及其在美军武器装备研制中的应用[J]. 测控技术. 2007, 26(03): 8-10.
- [12] DoD 5000.2- R, Mandatory procedures for major defense acquisition programs and major automated information system acquisition programs[R]. Washington, Department of Defense, 2002.
- [13] Mosser-Kerner D. Test and Evaluation Policies and Practices: A New Emphasis[R]. DTIC Document, 2008.
- [14] Douglas K, Michael W. Planning for the KC-X Integrated Test[C]. U.S. Air Force T&E Days 2009, Albuquerque, New Mexico. AIAA 2009-1759.

- [15] David K, Donald M. Integration of Turbine Engine Ground and Flight Test Using an Integrated Approach and Common Analysis Tools[C]. U.S. Air Force T&E Days, Los Angeles, California, AIAA 2008-1616.
- [16] Engelund W C. Hyper-X Aerodynamics: The X-43A Airframe-Integrated Scramjet Propulsion Flight-Test Experiments[J]. Journal of Spacecraft and Rockets. 2001, 38(6): 801-802.
- [17] Reagan W. Applications of Statistically Defensible Test and Evaluation Methods to Aircraft Performance Flight Testing[C]. 28th Aerodynamic Measurement Technology, Ground Testing, and Flight Testing Conference, New Orleans, Louisiana, AIAA 2012-2723.
- [18] David W, Richard S. Lessons Learned in Combined Developmental and Operational Test: Ten Years of F-15 Testing by the OFP CTF[C]. 28th Aerodynamic Measurement Technology, Ground Testing, and Flight Testing Conference, New Orleans, Louisiana, AIAA 2012-3163.
- [19] Bjorkman C E A, Zeis Jr C J E. Testing 2020: Evaluating Weapons Systems and Capabilities in a Joint Environment[C]. U.S. Air Force T&E Days, Destin, Florida, AIAA 2007-1602.
- [20] 王国盛, 洛刚. 美军一体化试验鉴定分析及启示[J]. 装备指挥技术学院学报. 2010, 21(02): 95-98.
- [21] 杨磊, 武小悦. 美军装备一体化试验与评价技术发展[J]. 国防科技, 2010, 31(02): 8-14.
- [22] 张连仲, 李进, 薄云蛟. 一体化联合试验体系内涵和特征研究[J]. 装备学院学报. 2014, 25(05): 113-116.
- [23] 孟冲, 宋华文. 关于推进一体化装备质量管理体系建设的战略思考[J]. 装备指挥技术学院学报. 2012, 23(1): 34-39.
- [24] 池建军, 罗小明, 李尚鸿等. 复杂电磁环境下装备试验模式[J]. 装甲兵工程学院学报, 2012, 26(1): 10-13.
- [25] 赵喜春, 时维科. 导弹定型鉴定中仿真与飞行试验一体化研究[J]. 现代防御技术. 2012, 40(06): 150-154.
- [26] 陈行勇, 雷国强. 雷达对抗装备一体化试验设计框架[J]. 电子信息对抗技术, 2012, 27(4): 61-64.
- [27] 汪连栋, 戚宗锋, 符淑芹, 等. 复杂电磁环境下电子信息系统试验的思考[J]. 航天电子对抗. 2013, 29(01): 31-34.
- [28] 李洪涛, 高顺林, 谢君红等. 水中兵器一体化试验研究[J]. 国防科技, 2013, 34(1): 1-4.
- [29] 刘保根, 张召奎. 论水雷一体化试验[J]. 水雷战与舰船防护, 2013, 21(02): 82-84.

- [30] 程敬丽, 郑敏, 楼建晴. 常见的试验优化设计方法对比[J]. 实验室研究与探索, 2012, 31(7): 7-11.
- [31] 王万中. 试验的设计与分析[M]. 北京: 高等教育出版社, 2004: 432.
- [32] 方开泰, 刘民千, 周永道. 试验设计与建模[M]. 北京: 高等教育出版社, 2011: 307.
- [33] Plackett R L, Burman J P. The design of optimum multifactorial experiments[J]. Biometrika, 1946, 33(4): 305-325.
- [34] 李国帅, 魏志, 李巍等. 现代试验设计方法在高速风洞试验中的应用[J]. 航空学报, 2015, 36(3): 782-788.
- [35] Silvey S D. Optimal design: an introduction to the theory for parameter estimation[M]. New York: Springer Science & Business Media, 2013.
- [36] Kuram E, Ozcelik B, Bayramoglu Met al. Optimization of cutting fluids and cutting parameters during end milling by using D-optimal design of experiments[J]. Journal of Cleaner Production, 2013, 42: 159-166.
- [37] Chen C, Chiang K, Chou Cet al. The use of D-optimal design for modeling and analyzing the vibration and surface roughness in the precision turning with a diamond cutting tool[J]. The International Journal of Advanced Manufacturing Technology, 2011, 54(5-8): 465-478.
- [38] 刘雨, 郭刚, 苏宝库. 惯导平台系统自标定试验的 D 最优设计方法[J]. 航空精密制造技术, 2011, 47(05): 14-17.
- [39] 李光庆, 谢祝捷, 姚雪琴. 花椰菜氮磷钾平衡施肥和阶段施氮效应函数的建立及在配方施肥中的应用[J]. 中国农业科学, 2011, 44(10): 2070-2080.
- [40] Jose E B, Emmanuel L, Michelle Ket al. Application of Mixture Design of Experiments for Dynamic Fleet-Level Evaluation of Multi-Objective Environmental Technology Trade-offs[C]. AIAA/3AF Aircraft Noise and Emissions Reduction Symposium, Atlanta, GA, AIAA 2014-2734.
- [41] Sacks J, Welch W J, Mitchell T Jet al. Design and analysis of computer experiments[J]. Statistical science, 1989, 4(4): 409-423.
- [42] Sacks J, Schiller S B, Welch W J. Designs for computer experiments[J]. Technometrics, 1989, 31(1): 41-47.
- [43] Simpson T W, Lin D K J, Chen W. Sampling strategies for computer experiments: Design and analysis[J]. International Journal of Reliability and Applications, 2001, 2(3): 209-240.
- [44] Santner T J, Williams B J, Notz W I. The Design and Analysis of Computer Experiments[M]. Springer Science & Business Media, 2013.
- [45] Bhat C R. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model[J]. Transportation Research Part B:

- Methodological, 2001, 35(7): 677-693.
- [46] Sobol I M. On quasi-monte carlo integrations[J]. Mathematics and Computers in Simulation, 1998, 47(2): 103-112.
- [47] Fang K, Li R, Sudjianto A. Design and modeling for computer experiments[M]. CRC Press, 2010.
- [48] Iman R L. Latin hypercube sampling[M]. Wiley Online Library, 2008.
- [49] Helton J C, Davis F J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems[J]. Reliability Engineering & System Safety, 2003, 81(1): 23-69.
- [50] Gardner M M, Ramanath V, Ayyalasomayajula Pet al. From Small X to Large X: Assessment of Space-Filling Criteria for the Design and Analysis of Computer Experiments[C]. 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, Rhode Island, AIAA 2006-1714.
- [51] 方开泰. 均匀试验设计的理论、方法和应用——历史回顾[J]. 数理统计与管理, 2004, 23(03): 69-80.
- [52] 邓海松. 基于稀疏先验的计算机试验元建模方法研究[D]. 南京: 南京理工大学, 系统工程, 2011.
- [53] 费惠佳, 徐海, 徐光耀. 均匀设计在反舰导弹抗干扰仿真试验中的应用[J]. 火箭与制导学报, 2011, 31(6):66-68.
- [54] 张国秋, 王文璇. 均匀试验设计方法应用综述[J]. 数理统计与管理, 2013, 32(01):89-99.
- [55] 曹煜彤, 高向华, 胡祖明等. 基于均匀设计优化 PPTA 纤维的热处理工艺[J]. 纺织学报, 2010, 31(12): 1-3.
- [56] 全欣, 陈高峰, 陆雁等. 基于均匀设计分析黄芪汤活性组分抗二甲基亚硝胺大鼠肝纤维化的配伍作用[J]. 中国中西医结合杂志, 2011, 31(10): 1389-1393.
- [57] Wang J, Chen Y, Wang Yet al. Optimization of the coagulation-flocculation process for pulp mill wastewater treatment using a combination of uniform design and response surface methodology[J]. Water Research, 2011, 45(17): 5633-5640.
- [58] Chaloner K. Optimal Bayesian Experimental Design for Linear Models[J]. The Annals of Statistics, 1984, 12(1): 283-300.
- [59] Pilz J U R. Bayesian estimation and experimental design in linear regression models[M]. John Wiley & Sons Inc, 1991.
- [60] Han C, Chaloner K. Bayesian Experimental Design for Nonlinear Mixed-Effects Models with Application to HIV Dynamics[J]. Biometrics, 2004,

- 60(1): 25-33.
- [61] Chaloner K, Verdinelli I. Bayesian experimental design: A review[J]. *Statistical Science*, 1995, 10(3): 273-304.
- [62] Sebastiani P, Wynn H P. Maximum entropy sampling and optimal Bayesian experimental design[J]. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 2000, 62(Part 1): 145-157.
- [63] Huan X, Marzouk Y M. Simulation-based optimal Bayesian experimental design for nonlinear systems[J]. *Journal of Computational Physics*, 2013, 232(1): 288-317.
- [64] Lin Y. An efficient robust concept exploration method and sequential exploratory experimental design[D]. Atlanta: Georgia Institute of Technology, 2004.
- [65] Sasena M, Parkinson M, Goovaerts Pet al. Adaptive experimental design applied to an ergonomics testing procedure[C]. *ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada, DETC2002/DAC-34091.
- [66] Azadi N A, Fearnhead P, Ridall Get al. Bayesian Sequential Experimental Design for Binary Response Data with Application to Electromyographic Experiments[J]. *Bayesian Analysis*, 2014, 9(2): 287-306.
- [67] Jin R, Chen W, Sudjianto A. On sequential sampling for global metamodeling in engineering design[C]. *ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada, DETC2002/DAC-34092.
- [68] Drovandi C C, McGree J M, Pettitt A N. Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data[J]. *Computational Statistics & Data Analysis*, 2013, 57(1): 320-335.
- [69] 韩旭, 王建宇, 祖先锋. 基于 Bayes 序贯检验的长周期系统动态试验方法[J]. *系统工程与电子技术*, 2012, 34(9):1957-1965.
- [70] 唐雪梅, 张金槐, 邵凤昌等. 武器装备小子样试验分析与评估[M]. 北京: 国防工业出版社, 2001: 321.
- [71] 段晓君, 王正明. 基于特征量的小子样试验鉴定精度研究[J]. *兵工学报*, 2003, 24(03): 367-372.
- [72] 孙锦, 李国林, 许诚. 基于仿真信息的反舰导弹小子样靶场试验综合评定方法[J]. *系统工程理论与实践*, 2014, 34(04): 1077-1088.
- [73] Adcock C J. Sample size determination: a review[J]. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1997, 46(2): 261-283.
- [74] Cochran W G. Sampling techniques[M]. New York: John Wiley & Sons, 2007.

- [75] Ayyub B M, McCuen R H. Probability, statistics, and reliability for engineers and scientists[M]. 3rd. Boca Raton: CRC press, 2011.
- [76] Desu M M. Sample Size Methodology[M]. Boston: Elsevier Science, 2012.
- [77] Machin D, Campbell M J, Tan Set al. Sample size tables for clinical studies[M]. John Wiley & Sons, 2011.
- [78] Murphy K R, Myors B, Wolach A. Statistical power analysis: A simple and general model for traditional and modern hypothesis tests[M]. Routledge, 2014.
- [79] Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies[J]. Journal of human reproductive sciences, 2012, 5(1): 7-13.
- [80] Guillera-Aroita G, Lahoz-Monfort J J. Designing studies to detect differences in species occupancy: power analysis under imperfect detection[J]. Methods in Ecology and Evolution, 2012, 3(5): 860-869.
- [81] Cohen J. Statistical power analysis for the behavioral sciences[M]. 2nd ed. New York: Routledge, 1988: 590.
- [82] Lewis J R. Sample sizes for usability tests: mostly math, not magic[J]. interactions, 2006, 13(6): 29-33.
- [83] Hancock P A, Ross J M, Szalma J L. A meta-analysis of performance response under thermal stressors[J]. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2007, 49(5): 851-877.
- [84] Freeman L J, Johnson T H, Simpson J R. Power Analysis Tutorial for Experimental Design Software[R]. Alexandria, VA: DTIC Document, 2014.
- [85] Faul F, Erdfelder E, Buchner A et al. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses[J]. Behavior research methods, 2009, 41(4): 1149-1160.
- [86] Box G E, Tiao G C. Bayesian inference in statistical analysis[M]. John Wiley & Sons, 2011.
- [87] Nassar M M, Khamis S M, Radwan S S. On Bayesian sample size determination[J]. Journal of Applied Statistics, 2011, 38(5): 1045-1054.
- [88] Joseph L, Belisle P. Bayesian sample size determination for normal means and differences between normal means[J]. Journal of the Royal Statistical Society: Series D (The Statistician), 1997, 46(2): 209-226.
- [89] Spiegelhalter D J, Abrams K R, Myles J P. Bayesian approaches to clinical trials and health-care evaluation[M]. Chichester, West Sussex, England: John Wiley & Sons, Ltd, 2004.
- [90] Spiegelhalter D J, Freedman L S. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion[J]. Statistics in medicine, 1986, 5(1): 1-13.

- [91] Joseph L, Du Berger R, Belisle Pet al. Bayesian and mixed Bayesian/likelihood criteria for sample size determination[J]. Statistics in Medicine, 1997, 16(7): 769-781.
- [92] Wang F, Gelfand A E. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models[J]. Statistical Science, 2002, 17(2): 193-208.
- [93] De Santis F. Using historical data for Bayesian sample size determination[J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2007, 170(1): 95-113.
- [94] Schlaifer R, Raiffa H. Applied statistical decision theory[M]. Boston: MIT Press, 1961.
- [95] Bernardo J M. Statistical Inference as a Decision Problem: The Choice of Sample Size[J]. Journal of the Royal Statistical Society. Series D (The Statistician), 1997, 46(2): 151-153.
- [96] Lindley D V. The choice of sample size[J]. Journal of the Royal Statistical Society: Series D (The Statistician), 1997, 46(2): 129-138.
- [97] Walker S G. How many samples?: a Bayesian nonparametric approach[J]. Journal of the Royal Statistical Society: Series D (The Statistician), 2003, 52(4): 475-482.
- [98] Beavers D P, Stamey J D. Bayesian sample size determination for binary regression with a misclassified covariate and no gold standard[J]. Computational Statistics & Data Analysis, 2012, 56(8): 2574-2582.
- [99] Khalifa M, Khan F, Haddara M. Bayesian sample size determination for inspection of general corrosion of process components[J]. Journal of Loss Prevention in the Process Industries, 2012, 25(1): 218-223.
- [100] Islam A F M S, Pettit L I. Bayesian Sample Size Determination Using Linex Loss and Linear Cost[J]. Communications in Statistics - Theory and Methods, 2012, 41(2): 223-240.
- [101] Brutti P, De Santis F, Gubbiotti S. Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials[J]. Statistics in medicine, 2009, 28(17): 2185-2201.
- [102] 张湘平. 小子样统计推断与融合理论在武器系统评估中的应用研究[D]. 长沙: 国防科学技术大学, 控制科学与工程, 2003.
- [103] 张硕云, 武小悦, 刘琦. 基于两类风险的正态分布 Bayes 试验样本量[J]. 航空计算技术, 2008, 38(5): 11-13, 17.
- [104] 王正明, 卢芳云, 段晓君等. 导弹试验的设计与评估[M]. 北京: 科学出版社, 2010: 663.
- [105] Ibrahim J G, Chen M H. Power prior distributions for regression models[J].

- Statistical Science, 2000, 15(1): 46-60.
- [106] 闫志强. 装备试验评估中的变动统计方法研究[D]. 长沙: 国防科学技术大学, 控制科学与工程, 2010.
- [107] 王飞, 张梅仓, 文艳等. 多因素条件下制导武器命中精度综合试验鉴定方法[J]. 装备指挥技术学院学报, 2011, 22(01): 112-115.
- [108] 陈璇. 复合制导武器系统战技指标的融合评估方法研究[D]. 长沙: 国防科学技术大学, 应用数学, 2011.
- [109] 茆诗松, 汤银才. 贝叶斯统计(第2版)[M]. 北京: 中国统计出版社, 2012: 302.
- [110] Cowles M K. Applied Bayesian statistics: with R and OpenBUGS examples[M]. Springer, 2013.
- [111] 袁修开, 吕震宙, 岳珠峰. 小样本下分位数函数的Bootstrap置信区间估计[J]. 航空学报, 2012, 33(10): 1842-1849.
- [112] Chen M, Ibrahim J G, Shao Q. Power prior distributions for generalized linear models[J]. Journal of Statistical Planning and Inference, 2000, 84(1): 121-137.
- [113] Diaconis P, Ylvisaker D, Others. Conjugate priors for exponential families[J]. The Annals of statistics, 1979, 7(2): 269-281.
- [114] Morris C N, Others. Natural exponential families with quadratic variance functions[J]. The Annals of Statistics, 1982, 10(1): 65-80.
- [115] Ibrahim J G, Chen M H, Sinha D. On optimality properties of the power prior[J]. Journal of the American Statistical Association, 2003, 98(461): 204-213.
- [116] Chen M H, Ibrahim J G, Shao Q Met al. Prior elicitation for model selection and estimation in generalized linear mixed models[J]. Journal of statistical planning and inference, 2003, 111(1-2): 57-76.
- [117] Duan Y. A modified Bayesian power prior approach with applications in water quality evaluation[D]. Blacksburg, VA: Virginia Polytechnic Institute and State UniversityDepartment of Statistics, 2005.
- [118] Fahrmeir L, Kneib T. Bayesian smoothing and regression for longitudinal, spatial and event history data[J]. OUP Catalogue, 2011.
- [119] 明志茂. 动态分布参数的 Bayes 可靠性综合试验与评估方法研究[D]. 长沙: 国防科学技术大学, 机械工程, 2009.
- [120] Fei X, Lu C, Liu K. A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction[J]. Transportation Research Part C: Emerging Technologies, 2011, 19(6): 1306-1318.
- [121] Zhang Y, Dukic V, Guszczka J. A Bayesian non-linear model for forecasting insurance loss payments[J]. Journal of the Royal Statistical Society: Series A

- (Statistics in Society), 2012, 175(2): 637-656.
- [122] Efron B. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction[M]. Taylor & Francis Ltd, 2012: 25, 59.
- [123] Li D, Chang F M, Chen K. Building reliability growth model using sequential experiments and the Bayesian theorem for small datasets[J]. Expert Systems with Applications, 2010, 37(4): 3434-3443.
- [124] 黄至辉, 严良文. 基于贝叶斯信息融合的制造执行系统可靠性增长分析[J]. 现代制造工程, 2014(05): 21-26.
- [125] Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models[J]. Biometrika, 2005, 92(2): 351-370.
- [126] Pu W, Niu X. Selecting mixed-effects models based on a generalized information criterion[J]. Journal of Multivariate Analysis, 2006, 97(3): 733-758.
- [127] Jiang J, Rao J S, Gu Z et al. Fence methods for mixed model selection[J]. The Annals of Statistics, 2008, 36(4): 1669-1692.
- [128] Lan L. Variable selection in linear mixed model for longitudinal data[D]. ProQuest, UMI Dissertations Publishing, 2006.
- [129] Schelldorfer J, Bühlmann P, De Geer S V. Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ_1 -Penalization[J]. Scandinavian Journal of Statistics, 2011, 38(2): 197-214.
- [130] Schelldorfer J U R, Meier L, B U Hlmann P. GLMMLasso: An Algorithm for High-Dimensional Generalized Linear Mixed Models Using ℓ_1 -Penalization[J]. Journal of Computational and Graphical Statistics, 2014, 23(2): 460-477.
- [131] Foster S D, Verbyla A P, Pitchford W S. Estimation, prediction and inference for the LASSO random effects model[J]. Australian & New Zealand Journal of Statistics, 2009, 51(1): 43-61.
- [132] Yang H. Variable selection procedures for generalized linear mixed models in longitudinal data analysis[D]. North Carolina State University, 2007.
- [133] Bondell H D, Krishna A, Ghosh S K. Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models[J]. Biometrics, 2010, 66(4): 1069-1077.
- [134] Ibrahim J G, Zhu H, Garcia R I et al. Fixed and Random Effects Selection in Mixed Effects Models[J]. Biometrics, 2011, 67(2): 495-503.
- [135] 许佩蓉. 高维纵向数据中边际模型和混合效应模型的若干研究[D]. 上海: 华东师范大学, 概率论与数理统计, 2013.
- [136] 张金槐, 张士峰. 验前大容量仿真信息“淹没”现场小子样试验信息问题[J]. 飞行器测控学报, 2003, 22(3): 1-6.
- [137] 焦鹏. 导弹制导仿真系统 VV&A 理论和方法研究[D]. 长沙: 国防科学技

- 术大学, 控制科学与工程, 2010.
- [138] 段晓君, 黄寒砚. 基于信息散度的补充样本加权融合评估[J]. 兵工学报, 2007, 28(10): 1276-1280.
- [139] 莫家庆, 胡忠望, 叶雪琳. 基于模糊理论的可信计算信任评估方法研究[J]. 计算机应用, 2013, 33(01): 142-145.
- [140] Dong G, He F, Yao Yet al. Sample Size Determination Method for Credibility Evaluation Based on Statistical Power Analysis[C]. The 11th World Congress on Intelligent Control and Automation (WCICA 2014), Shenyang, China, 2014.
- [141] 徐德坤. 弹道导弹命中精度评定方法及其应用研究[D]. 长沙: 国防科学技术大学, 航天与材料工程, 2007.
- [142] 张士峰, 蔡洪. Bayes 分析中的多源信息融合问题[J]. 系统仿真学报, 2000, 12(1): 54-57.
- [143] Matteucci M, Veldkamp B P. The approach of power priors for ability estimation in IRT models[J]. Quality & Quantity, 2015, 49(3): 917-926.
- [144] Hobbs B P, Carlin B P, Mandrekar S Jet al. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials[J]. Biometrics, 2011, 67(3): 1047-1056.
- [145] Lee S, Choi J, Kim S W. Bayesian approach with the power prior for road safety analysis[J]. Transportmetrica, 2010, 6(1): 39-51.
- [146] Duan Y, Ye K, Smith E P. Evaluating water quality using power priors to incorporate historical information[J]. Environmetrics, 2006, 17(1): 95-106.
- [147] Ibrahim J G, Chen M, Gwon Yet al. The power prior: theory and applications[J]. Statistics in Medicine, 2015, 34(28): 3724-3749.
- [148] Bernardo J M, Smith A F M. Bayesian theory[M]. New York: John Wiley & Sons, 2000.
- [149] Kass R E, Wasserman L. The selection of prior distributions by formal rules[J]. Journal of the American Statistical Association, 1996, 91(435): 1343-1370.
- [150] Berger J O. Statistical decision theory and Bayesian analysis[M]. 2nd. New York: Springer, 2004.
- [151] 韦来生, 张伟平. 贝叶斯分析[M]. 第 1 版. 安徽 合肥: 中国科学技术大学出版社, 2013: 369.
- [152] Chen M. Importance-Weighted Marginal Bayesian Posterior Density Estimation[J]. Journal of the American Statistical Association, 1994, 89(427): 818-824.
- [153] 洪东跑, 马小兵, 赵宇. 基于比例风险模型的可靠性综合评估[J]. 系统工

- 程与电子技术, 2010, 32(10): 2132-2135.
- [154] Ibrahim J G, Chen M. Prior distributions and Bayesian computation for proportional hazards models[J]. Sankhyā: The Indian Journal of Statistics, Series B, 1998, 60(1): 48-64.
- [155] Zellner A. Optimal information processing and Bayes's theorem[J]. The American Statistician, 1988, 42(4): 278-280.
- [156] Shannon C E. A mathematical theory of communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2001, 5(1): 3-55.
- [157] 吴密霞. 线性混合效应模型引论[M]. 北京: 科学出版社, 2013: 217.
- [158] Titchmarsh E C. The theory of functions[M]. London: Oxford University Press, 1939.
- [159] Tikhonov A N. Solution of incorrectly formulated problems and the regularization method[J]. Soviet Math. Dokl, 1963, 4: 1035-1038.
- [160] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning[M]. Springer New York, 2009.
- [161] Tibshirani R. Regression shrinkage and selection via the Lasso[J]. Journal of the Royal statistical society, Serues B: Methodological, 1996, 58(1): 267-288.
- [162] Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net[J]. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 2005, 67(2): 301-320.
- [163] Fan J, Li R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties[J]. Journal of the American Statistical Association, 2001, 96(456): 1348-1360.
- [164] Zhang C. Nearly unbiased variable selection under minimax concave penalty[J]. The Annals of Statistics, 2010, 38(2): 894-942.
- [165] Lv J, Fan Y. A unified approach to model selection and sparse recovery using regularized least squares[J]. The Annals of Statistics, 2009, 37(6A): 3498-3528.
- [166] Frank I E, Friedman J H. A Statistical View of Some Chemometrics Regression Tools[J]. Technometrics, 1993, 35(2): 109-135.
- [167] Fan J, Lv J. A Selective Overview of Variable Selection in High Dimensional Feature Space[J]. Stat Sin, 2010, 20(1): 101-148.
- [168] Newman M E J. Power laws, Pareto distributions and Zipf's law[J]. Contemporary physics, 2005, 46(5): 323-351.
- [169] 曹宁. 车载炮射击精度分析[D]. 南京: 南京理工大学, 火炮、自动武器与弹药工程, 2012.
- [170] De Santis F, Fulvio S. Alternative Bayes Factors for Model Selection[J]. The

- Canadian Journal of Statistics, 1997, 25(4): 503-515.
- [171] De Santis F. Power priors and their use in clinical trials[J]. The American Statistician, 2006, 60(2): 122-129.
- [172] Joseph L, Wolfson D B, Du Berger R. Sample size calculations for binomial proportions via highest posterior density intervals[J]. The Statistician, 1995, 44(2): 143-154.
- [173] Di Bacco M, Amore G D, Scalfari Fet al. Two Experimental Settings in Clinical Trials: Predictive Criteria for Choosing the Sample Size in Interval Estimation[M]. Applied Bayesian Statistical Studies in Biology and Medicine. Norwell, MA, USA: Springer US, 2004: 109-130.
- [174] 孙锐. 基于 D-S 证据理论的信息融合及在可靠性数据处理中的应用研究[D]. 成都: 电子科技大学, 机械电子工程, 2012.
- [175] Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images[J]. IEEE Trans Pattern Anal Mach Intell, 1984, 6(6): 721-741.
- [176] Besag J. Spatial interaction and the statistical analysis of lattice systems[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1974: 192-236.

攻读博士学位期间发表的论文及其它成果

（一）发表的学术论文

- [1] Guangling Dong, Yu Yao, Xingdan Wang, Chi He. Credibility analysis of flight motion table in simulation with frequency domain norm[C]. 2012 9th IEEE International Conference on Mechatronics and Automation (ICMA 2012), Chengdu, China, 2012: 168-173. (EI 收录号: 20124315597624)
- [2] 董光玲, 姚郁, 赫赤, 王杏丹. 制导武器仿真转台动态频响指标研究[C]. 第 31 届中国控制会议, 安徽 合肥, 2012: 7669-7673. (EI 收录号: 20130716024802)
- [3] 董光玲, 姚郁, 贺风华, 赫赤. 制导精度一体化试验的 Bayesian 样本量计算方法[J]. 航空学报, 2015. 36(02): 第 575-584 页. (EI 收录号: 20152100877508)
- [4] Guangling Dong, Fenghua He, Yu Yao, Hui Zhao. Sample size determination method for credibility evaluation based on statistical power analysis[C]. The 11th World Congress on Intelligent Control and Automation (WCICA 2014), Shenyang, China, 2014. (EI 收录号: 20152600977515)
- [5] Guangling Dong, Chaofeng Huang, Chi He, etc. Study on identification method of rapidity performance parameters DEWE 2010 based[C]. 2013 2nd International Conference on Measurement, Instrumentation and Automation (ICMIA 2013), Guilin, China, 2013: 2433-2436. (ISTP 检索, WOS: 000329080300486, EI 收录号: 20133516681339)
- [6] Chi He, Guangling Dong, Qiang Li, Zhengguo Dai. Design of artillery electrical performance parameters measurement and control system based on dual-bus[C]. 2012 9th IEEE International Conference on Mechatronics and Automation (ICMA 2012), Chengdu, China, 2012: 2048-2053. (EI 收录号: 20124315597943)
- [7] Chi He, Guangling Dong, Qiang Li, etc. Design of automatic force application system and outlier detection for force sensor[C]. The 7th International Conference on Sensing Technology (ICST 2013), Wellington, New Zealand, 2013: 766-770. (EI 收录号: 20141317521843)
- [8] Xueying Huang, Guangling Dong, Chi He, Yun Bo. Study on metamodel based assessment method for integrated test[C]. 2013 4th IEEE International

Conference on Grey Systems and Intelligent Services (GSIS 2013), Macau, China, 2013: 219-224. (EI 收录号: 20141317520325)

- [9] Chi He, Guangling Dong, Chaoyuan Cai, etc. Design of shock simulation system for shipborne weapons based on electrohydraulic servomotor[C]. 2011 2nd International Conference on Mechanic Automation and Control Engineering (MACE 2011), Inner Mongolia, China, 2011: 192-195. (EI 收录号: 20114314452963)

(二) 参与的科研项目及获奖情况

- [1] 装甲车辆炮控系统试验测试技术. 军队科技进步三等奖, 排名第 4, 2010.
- [2] 通用数字式正弦机系统. 军队科技进步三等奖, 排名第 2, 2011.
- [3] 火炮随动系统射击性能半实物仿真试验技术. 军队科技进步二等奖, 排名第 4, 2012.

致 谢

本文是在我的导师姚郁教授的精心指导和严格要求下完成的。导师在课题研究的各个阶段都给予了认真的关注和悉心的指导，提出了许多具有启发性和建设性的建议。导师渊博的学识、敏锐的洞察力、严谨求实的科学态度、高度的责任意识、执着的钻研精神、孜孜不倦的工作态度、不断求新的科研思维以及雷厉风行的作风给我留下了深刻的印象，这些都将使我受益终生。

感谢课题组的马杰老师、杨宝庆老师、贺风华老师、陈松林老师、马克茂老师、赵辉老师、周宣圻老师以及其他老师对我学习和生活上的关心和帮助，他们踏实勤奋的工作作风，积极乐观的人生态度深深地感染着我！

感谢霍鑫博士、王晓晨博士、刘凯博士、张鹏博士、郭洋博士、郭健博士、孟桂芝博士、张佳为博士以及博士生王杏丹、李兴龙、陶瑞超、刘峰、史维佳、王龙、刘晓坤、马龙彪、汪洋、张斯伦、张曦、李博文、范红霞等给予我的关心和帮助，与他们多年的共同生活和学习给我留下了美好的回忆。

特别感谢我的父母、爱人和孩子，他们为我的求学生涯提供了最无私的支持与帮助，让我能够全身心地投入到科研和学术研究中，正是他们不断的关心和鼓励使我得以顺利完成学业。

最后，再次向所有给过我帮助的人们表示深深地感谢！

