

最优设计的 计算机证明与构造

朱伟勇 胡晨江 等编著

Zuiyou sheji de CAD

$$\eta(x) = \dots + \sum_{i=1}^n \nu \ln x_i$$

东北工学院出版社

最优设计的计算机证明与构造

朱伟勇 胡晨江
陈惠森 刘坤芳 编著
傅连魁 王金玉

东北工学院出版社

内 容 提 要

本书共十章, 前四章中提出对数项混料回归新模型及其应用, 对它的 D -最优设计进行了CAD证明与构造。第五章利用计算机构造常用的正交设计表, 并对特殊的正交设计直和法和裂区法等编制了专门的程序与方差分析计算。后五章中利用FORTRAN等语言编出自动构造和计算“多元回归分析”、“二次回归正交设计”、“Box的旋转设计”、“Kiefer的 D -最优设计”和“Scheffé的混料设计”的程序, 以及它们的应用实例的统计分析、极值判别和寻找最佳工艺条件的优化计算程序。

此外, 在附录中除列出Scheffé的单纯形格子点设计的Pascal语言, 以及由三至七个分量的混料设计表格和计算机绘制的三元相图的程序与实例外, 还列出二至八个因素的二次回归正交设计表格。

本书可供从事采矿、冶金、机械、自控、材料、轻化工、制药和管理的科技人员使用, 可作为理工科高等院校教师、研究生和高年级学生讲授和学习应用统计、应用软件的教学参考书, 也可供学生完成学位论文、毕业设计使用。

最优设计的计算机证明与构造

朱伟勇 胡晨江 等编著

东北工学院出版社出版 辽宁省新华书店发行
(沈阳·南湖) 东北工学院印刷厂印刷

开本 787×1092 毫米 1/16 印张: 50 字数: 1200千字
1987年6月第1版 1987年6月第1次印刷

印数 1~4000册

责任编辑 车荫昌 封面设计 赫林

ISBN 7-81006-003-1/O·1
统一书号: 13476·6 定价: 8.50元

前 言

本书应用计算机辅助设计(CAD)的思想和方法,证明和构造《最优设计理论与应用》中许多问题;这是近廿年迅速发展起来的应用数理统计的一个崭新的分支。

本书共分十章,其中包括作者近年来取得的成果“具有对数项的混料模型及其 D -最优测度设计和 D_n -最优确切设计”、“混料设计中构造单纯形格子点的一个算法”(分别参加1984和1985年美国第144届和145届统计数学会及第二届日中统计讨论会,并发表于“美国统计计算进展”等上)、以及“Kiefer的 D -最优设计、Box的旋转设计、Scheffé的混料设计以及Draper的倒数项设计”等等,上述这些设计都是目前国内外应用效果较好的新统计设计方法。

上述的这些设计无论在工程陶瓷新材料的开发研究,或是在工程管理、采矿工程、钢铁冶金、有色冶炼、机械制造、自动控制、电子仪器、轻化工、医药卫生、农业饲料等领域内建立数学模型、配方配比科学试验获取最佳工艺条件和质量控制等问题上,都取得了一系列的较好成果。

为了方便读者应用和理论研究,本书列出在不同领域内的应用实例卅多个,并附有上述各种方法的程序包,包括自动构造各种回归设计表格,计算回归系数,方差分析,极值判别,区间估计,获取最佳工艺条件的有关线性与非线性规划的优化方法等等,分别用FORTRAN, BASIC和PASCAL等语言编写,并在东北工学院计算中心ACOS-400机及CCS-400微机上经过多次算题考验。其中大部分实例曾分别在《应用数学学报》、《应用概率统计》、《数学的实践与认识》和《东北工学院学报》等杂志上发表过。

本书可供工业、农业和国防系统从事科学试验的研究人员、工程师、技术员和工程技术干部使用,可作为工科高等院校教师、研究生和高年级学生讲授和学习应用数理统计课程的教学工具书和参考书。同时,也可供工科和应用统计研究生,高年级学生上机实习和完成学位论文、毕业设计时使用。本书主要内容曾给研究生和进修教师讲授多次,是在拙著《正交与回归正交试验法的应用》及《最优设计理论与应用》的基础上,修改、补充而发展起来的。

本书前四章给出了一个完整的应用理论研究过程,它包括新模型的提出,应用CAD方法证明和构造模型的 D -最优设计,和构造一大批 D_n -最优设计表格,以及新模型在工程陶瓷材料研究中的应用和某些统计分析,同时把它与国外文献上的实例作了详细的分析比较。

第一章首先综述了到目前为止人们提出的各种混料模型;接着,类似于具有倒数项的混料模型,引出了用以描述边界附近混料响应极剧变化的一个新模型,即具有对数项的多项式混料模型。利用McLean和Anderson的实例以及按第二章表2—4中的确切设计所应用的一个实例,说明了我们的新模型是一个较好的实用模型,同时,还通过此两例对新模型与倒数项混料模型进行了比较。为了检验模型中对数项的存在,我们给出了几个 F -统计量所定义的拒绝域。最后,给出了一个对数项混料模型的简化过程。

在第二章,首先叙述了 D -最优性的概念以及几种常见模型的 D -或 D_n -最优设计。目前,为了寻找 D -最优(测度)设计和 D_n -最优(确切)设计,国外已形成了专门的CAD方法,

DAA03/03

即试验的计算机辅助设计方法。自 1969 年以来,苏联莫斯科大学教授 Fedorov,英国皇家学院教授 Wynn 和美国的 Mitchell 等人相继给出了为寻求 D -最优设计和 D_n -最优设计的几个迭代算法。以后,也不时有人提出了一些类似或修正的算法。这些算法一般是建立在 Kiefer-Wolfowitz 等价定理的基础上的。在第二节列出了本书所用到的其中几个算法。Goldfarb 提出的共轭梯度投影算法是求解线性约束最优化的问题的最有效方法之一。在这个算法的基础上,我们提出和使用了一个组合的搜索方法来求方差函数和 Δ 函数的最大值点,这是我们在使用 CAD 方法时所进行的一个改进。因为以前人们在遇到这个求最大值问题时,常限定一个候选点集。在第四、五节里,我们使用第二节列出的几个算法及其修改,构造了具有对数项的混料模型在三分量和四分量时的近似 D -最优测度设计以及一大批可能是 D_n -最优的确切设计表格。同时,对这些设计作了详细的讨论并与倒数项模型的确切设计进行了比较。

第三章是作为一个理论基础来写的。这章概括了 CAD 方法所需的基本引理,以及重要定理和公式,并给出了它们的详细推导和证明。还列出了本书所需用的几个重要算法以及其中两个的收敛性证明。对于一些一般的计算公式,还给出了本书所采用到的特殊形式。

第四章给出了证明具有对数项混料模型的 D -最优性的“组合搜索法”完整 FORTRAN 程序。分别证明和构造了在三分量和四分量的情况下,一次和二次的对数项混料模型的近似 D -最优测度设计和 D_n -最优确切设计。并给出了计算实例和程序。

上述四章内容是经过三年多的研究,先后得到国内外许多著名统计学家和教授的热心指教、严格评审和批评修改后,所完成的新结果。在此谨向华东师范大学数理统计系魏宗舒教授和王万中副教授,中国现场统计研究会理事长张里千教授和副理事长林少宫教授,中国科技大学数学系陈希孺教授,武汉大学统计系主任张尧庭教授,中国科学院系统研究所所长成平研究员和项可风研究员,美国佛罗里达大学 J. Cornell 教授表示衷心的感谢!

本书的后六章根据正交设计、回归正交设计、Box 回归旋转设计、Kiefer D -最优设计、Scheffé 回归混料设计(单纯形格子点设计)及多元回归等内容,结合实际例子编写了一批计算程序及表格,可供广大读者具体应用。有关这些内容的详细数学原理和推导,请参考华东师范大学教授茆诗松、周纪芑、丁元和吕乃刚等编著的《回归分析及其试验设计》或拙著《最优设计理论与应用》等书。

本书的出版得到中国科学院和冶金工业部科学基金的资助。在编写过程中,得到我的老师东北工学院数学系赵惠元教授、自控系潘德惠教授以及王淑琴、关颖男等同志的热情鼓励!特别是,东北工学院计算中心主任王泳嘉教授,副主任魏朋三、杨兴义及程序操作人员为本书大量计算证明提供的方便和大力支持,在此一并致谢!

全书由朱伟勇、胡晨江主编,各章的执笔人是:胡晨江(第一、二、三、四章),刘坤芳(第五章),傅连魁(第六、八、九章),王金玉(第七、十章)。附录三的程序是由陈惠森、朱伟勇提供的,附录五由研究生段晓东提供。

由于编者水平有限,错误难免,请广大读者批评指正。

朱伟勇

1986年3月于东北工学院计算中心

目 录

第一章 具有对数项的混料模型

- 1.1 现有的各种混料模型 (1)
- 1.2 具有对数项的混料模型 (4)
- 1.3 检验模型中是否含有对数项 (9)
- 1.4 对数项混料模型的序列简化 (15)

第二章 具有对数项混料模型的 D -最优测度设计和 D_n -最优确切设计

- 2.1 D -最优性以及几种常见模型的 D -最优设计 (17)
- 2.2 寻求 D -最优测度设计或 D_n -最优确切设计的几个迭代算法 (19)
- 2.3 求函数极值的一个组合算法 (22)
- 2.4 三分量对数项混料模型的测度设计和确切设计 (24)
- 2.5 四分量对数项混料模型的测度设计和确切设计 (31)
- 2.6 总结与后记 (39)

第三章 “CAD”方法的主要结果

- 3.1 Kiefer-Wolfowitz 等价定理和 Fedorov 的求 D -最优设计的迭代算法 (41)
- 3.2 求 D_n -最优设计的单点交换法 (47)
- 3.3 Wynn 的序列发生方法 (50)

第四章 回归模型 D -最优设计的构造程序

- 4.1 计算机构造 D -最优设计的 FORTRAN 程序 (53)
- 4.2 计算机构造 D_n -最优设计的 FORTRAN 程序 (61)
- 4.3 具有对数项混料回归的 FORTRAN 计算程序 (66)

第五章 正交试验设计与应用程序

- 5.1 正交表的构造 (114)
 - 5.1.1 以素数 (3, 5, 7, 11 等) 为水平数的正交表 (114)
 - 5.1.2 二水平正交表 (115)
 - 5.1.3 拟水平正交表 (115)
 - 5.1.4 拟因子正交表 (116)
- 5.2 正交试验设计的分析 (117)
 - 5.2.1 正交试验方案的设计 (117)

5.2.2 直观分析·····	(118)
5.2.3 方差分析·····	(121)
5.3 设计方法 ·····	(125)
5.3.1 拟水平设计·····	(125)
5.3.2 拟因子设计·····	(126)
5.3.3 裂区法设计·····	(128)
5.3.4 部分追加法设计·····	(132)
5.3.5 直和法设计·····	(135)
5.3.6 组合法设计·····	(140)
5.3.7 直积法设计·····	(141)
5.3.8 可计算性项目的三次设计(正交优化设计) ·····	(144)
5.4 正交试验设计实例与程序 ·····	(148)
5.4.1 打印正交表程序·····	(149)
5.4.2 炼铁应用程序·····	(156)
5.4.3 三甲胺合成应用程序·····	(172)
5.4.4 钒钛合金重轨钢成分对性能的影响应用程序·····	(177)
5.4.5 改善薄钢板性能的试验应用程序·····	(188)
5.4.6 氯丁胶降低成本试验应用程序·····	(203)
5.4.7 泡沫聚乙烯电线的制造应用程序·····	(210)

第六章 多元回归正交设计程序及应用

6.1 多元线性回归程序及应用 ·····	(356)
6.2 二次回归正交设计程序及应用 ·····	(383)

第七章 Box 旋转设计程序与应用

7.1 基本概念·····	(459)
7.2 二次正交旋转组合设计·····	(461)
7.3 二次几乎正交旋转组合设计 ·····	(467)
7.4 二次通用旋转组合设计·····	(472)

第八章 用数值方法构造 D -最优设计程序

8.1 基本数学原理 ·····	(509)
8.2 程序功能及使用说明·····	(515)
8.3 计算举例·····	(519)

第九章 具有边效应混料模型的 D -最优设计的构造

9.1 构造具有边效应混料模型 D -最优设计的程序 ·····	(537)
9.2 构造具有边效应混料模型 D_n -最优设计的程序 ·····	(552)

第十章 多元线性回归应用程序及其它

10.1 多元线性回归分析程序及应用	(607)
10.2 混料回归设计的 XVERT 法程序及应用	(618)
10.3 用步长加速法求带有约束条件的函数极值程序及应用	(623)

附 录

(一) 常用正交表	(638)
(二) 二次回归正交表	(692)
(三) 一个构造单纯形格子点的 PASCAL 程序及格子点设计	(703)
(四) F 分布表	(786)
(五) 混料回归模型的计算机作图程序及实例	(789)

第一章 具有对数项的混料模型

1.1 现有的各种混料模型

在混料问题中, 响应是每种成分在混料总量中所占比例的函数, 而与混料的总量无关。每种成分所占的比例只能在 0 与 1 之间变化, 其相加总和必须是 1。设 $x_i (i=1, 2, \dots, q)$ 表示 q 分量系统中第 i 种成分所占的比例, 则混料问题要受约束条件

$$\begin{cases} \sum_{i=1}^q x_i = 1 \\ x_i \geq 0 \quad (i=1, 2, \dots, q) \end{cases} \quad (1-1)$$

的限制。令 $x = (x_1, x_2, \dots, x_q)'$, 它是混料的自变量向量。式(1-1)定义了一个 $(q-1)$ 维正规范单纯形, 通常称其为 q 分量混料试验的可行区域或利益区域。令 $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ 代表由 p 个未知参数组成的向量, 它是需要进行试验来估计的。一般所考虑的响应 η 与 x 和 θ 的函数关系为

$$\eta = f(x, \theta) \quad (1-2)$$

假定在观察响应 η 时存在试验误差, 并以 ξ 记之, 因此, 对响应 η 的每个观察 $Y_\mu, \mu=1, 2, \dots, n$

$$Y_\mu = \eta_\mu + \xi_\mu = f(x_\mu, \theta) + \xi_\mu \quad (1-3)$$

我们总假定随机变量 $\xi_\mu (\mu=1, 2, \dots, n)$ 是独立同分布的, 其均值为零, 协方差矩阵为 $\sigma^2 I$, 其中 σ^2 是未知参数。

混料问题是近二、三十年新发展起来的十分活跃的统计分支, 混料模型的研究一直是混料问题的重要课题之一。自 1955 年 Claringbold^[3] 到现在止, 许多统计学家和统计工作者提出了各种各样的适合于各种数据结构或适用于各种统计目的所需的混料回归模型。

Claringbold^[3] 利用了约束条件 (1-1), 把 q 个混料分量 $x_i, i=1, 2, \dots, q$, 变换到 $(q-1)$ 个独立变量 $z_j, j=1, 2, \dots, q-1$, 然后使用标准的二次多项式模型

$$\eta(z) = \beta_0 + \sum_{j=1}^{q-1} \beta_j z_j + \sum_{i < j}^{q-1} \beta_{ij} z_i z_j \quad (1-4)$$

Scheffé^[11] 不采用变换分量, 而通过约束条件(1-1)在不失去分量的前提下修改关于 x_i 的通常多项式模型。从一般 q 个变量的 r 次多项式

$$\begin{aligned} \eta(x) = & \beta_0 + \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \dots \\ & + \sum_{i_1 < i_2 < \dots < i_r}^q \beta_{i_1 i_2 \dots i_r} x_{i_1} x_{i_2} \dots x_{i_r} \end{aligned} \quad (1-5)$$

Scheffé 得到: 线性典型多项式

$$\eta(x) = \sum_{i=1}^q \beta_i x_i \quad (1-6)$$

二次典型多项式

$$\eta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (1-7)$$

和立方典型多项式

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i < j}^q \gamma_{ij} x_i x_j (x_i - x_j) \\ & + \sum_{i < j < k}^q \beta_{ijk} x_i x_j x_k \end{aligned} \quad (1-8)$$

这些典型多项式, 当使用 Scheffé 的单纯形格子点设计时, 其回归系数最小二乘估计的计算将大大地简化。

Scheffé 还定义了特殊的典型多项式

$$\begin{aligned} \eta_q(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \dots \\ & + \beta_{12\dots q} x_1 x_2 \dots x_q \end{aligned} \quad (1-9)$$

它是从 q 次典型多项式删去某些交叉乘积而得到的。当模型(1-9)使用单纯形重心设计时, Scheffé 提出了其系数估计的直接计算公式。这个模型的次数还可进一步去掉某些高次项而降低。例如, 常用的特殊立方多项式

$$\eta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i < j < k}^q \beta_{ijk} x_i x_j x_k \quad (1-10)$$

当模型的次数降到 r 时 ($1 \leq r < q$), 一般地可得到如下多项式

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \dots \\ & + \sum_{i_1 < \dots < i_r}^r \beta_{i_1 \dots i_r} x_{i_1} \dots x_{i_r} \end{aligned} \quad (1-11)$$

Cox^[5] 提出了表达式与式(1-5)相同的模型, 但其中要求选择一个参考混料和使用含有这个参考混料的一些约束来估计参数。Cox 模型具有与 Scheffé 模型不同的解释, 但却得出同样的预测响应。进一步的讨论见文献[5]。

Kenworthy^[8] 和 Snee^[12] 等人使用了关于混料分量之比的一般多项式模型。例如, 当 $q=3$ 时, 我们就能使用二次模型

$$\eta(x) = \beta_0 + \sum_{i=1}^2 \beta_i z_i + \beta_{12} z_1 z_2$$

其中变量 z_1, z_2 是如下比例:

$$z_1 = x_2/x_1, \quad z_2 = x_3/x_1$$

或

$$z_1 = x_1/x_2, \quad z_2 = x_2/x_3 \quad (1-12)$$

或

$$z_1 = \frac{x_1}{x_1 + x_2}, z_2 = x_2/x_3$$

上面这个模型仅仅是比模型中的一个,事实上比模型是一类很广的模型簇。

Becker^[2]定义了一个可加分量,即不活泼分量。假设

$$\eta(x) = f(x_1, x_2, \dots, x_q)$$

表示一个 q 分量混料系统的响应,并且另一个分量 x_{q+1} 要加到这个混料系统中去。进一步假定 x_{q+1} 关于其它 q 个分量是可加的,那么这 $(q+1)$ 个分量的混料系统的响应可以表示为

$$\eta(x) = \beta_{q+1}x_{q+1} + (1 - x_{q+1})f\left(\frac{x_1}{1 - x_{q+1}}, \dots, \frac{x_q}{1 - x_{q+1}}\right) \quad (1-13)$$

Becker 规定,在可加分量 x_{q+1} 加入混料系统的情况下,系数对分量 $x_1, x_2, \dots, x_q$ 的解释应是不变的。一次齐次模型满足这个不变性准则。Becker 提出了三种这样的模型:

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} \min(x_i, x_j) + \dots \\ & + \beta_{12\dots q} \min(x_1, x_2, \dots, x_q) \end{aligned} \quad (1-14)$$

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j / (x_i + x_j) + \dots \\ & + \beta_{12\dots q} x_1 x_2 \dots x_q / (x_1 + x_2 + \dots + x_q) \end{aligned} \quad (1-15)$$

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} (x_i x_j)^{1/2} + \dots \\ & + \beta_{12\dots q} (x_1 x_2 \dots x_q)^{1/q} \end{aligned} \quad (1-16)$$

如果式(1-15)中的某项的分子等于零,那么这项的值就定义为零。

为了拟合当一个或多个分量的值趋于零(即趋于利益区域的边界)时响应值产生的急剧变化,Draper 和 St. John^[7]提出了具有倒数项的多项式混料模型:

$$\eta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{-i} x_i^{-1} \quad (1-17)$$

$$\eta(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^q \beta_{-i} x_i^{-1} \quad (1-18)$$

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i < j < k} \beta_{ijk} x_i x_j x_k \\ & + \sum_{i=1}^q \beta_{-i} x_i^{-1} \end{aligned} \quad (1-19)$$

$$\begin{aligned} \eta(x) = & \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i < j < k} \beta_{ijk} x_i x_j x_k \\ & + \sum_{i < j} \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{i=1}^q \beta_{-i} x_i^{-1} \end{aligned} \quad (1-20)$$

这些模型(1-17)~(1-20)都是 Scheffé 典型多项式加倒数项 x_i^{-1} 而得到的。附加的项 x_i^{-1}

是用来解释 $x_i \rightarrow 0$ 时响应可能产生的急剧变化, Draper 和 St. John 称这些倒数项或它们的系数为边效应。

注意到 x_i^{-1} 的泰勒展开式是

$$1/x_i = 1 - (x_i - 1) + (x_i - 1)^2 - (x_i - 1)^3 + (x_i - 1)^4 - \dots \quad (1-21)$$

它含有项 $x_i, x_i^2, x_i^3, \dots$, 因此 x_i^{-1} 能看作多项式的特殊形式。所以, 具有倒数项的混料模型是 Scheffé 典型模型的推广, 同样注意到 x_i^{-1} 能表示为

$$1/x_i = \sum_{j=1}^q x_j/x_i = 1 + \sum_{j \neq i} x_j/x_i \quad (1-22)$$

因此, 具有倒数项的混料模型是同某些比模型相联系的。

Aitchison 和 Bacon-Shone^[1] 引进了对数比混料模型:

$$\eta_1(x) = \beta_0 + \sum_{i=1}^{q-1} \beta_i z_i \quad (1-23)$$

$$\eta_2(x) = \eta_1(x) + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \gamma_{ij} z_i z_j \quad (1-24)$$

其中

$$z_i = \log(x_i/x_q), \quad (i = 1, 2, \dots, q-1) \quad (1-25)$$

用式(1-25)代入式(1-23)和(1-24), 就可以得到

$$\eta_1(x) = \beta_0 + \sum_{i=1}^q \beta_i \log x_i \quad (\beta_1 + \dots + \beta_q = 0) \quad (1-26)$$

$$\eta_2(x) = \eta_1(x) + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} (\log x_i - \log x_j)^2 \quad (1-27)$$

这里, 新参数 $\beta_{ij} (i = 1, \dots, q-1, j = i+1, \dots, q)$ 与 $\gamma_{ij} (i = 1, \dots, q-1, j = i, \dots, q-1)$ 的关系为

$$\beta_{ij} = -\frac{1}{2} \gamma_{ij} \quad (i = 1, \dots, q-1, j = i+1, \dots, q-1)$$

$$\beta_{iq} = \gamma_{ii} + \frac{1}{2} \sum_{k < i} \gamma_{ki} + \frac{1}{2} \sum_{k > i} \gamma_{ik} \quad (i = 1, \dots, q-1)$$

Aitchison 和 Bacon-Shone 讨论了这个模型的某些有关混料的假设, 例如, 子分量混料的不活泼性的假设, 混料剖分可加性假设, 混料的完全可加性假设等; 他们并且指出了这些假设同模型的参数假设是等价的, 利用参数假设的检验来进行有关混料的假设检验。

1.2 具有对数项的混料模型

当某个混料系统的响应在区域边界附近产生急剧变化时, Draper 和 St. John 使用了带倒数项的典型多项式来描述之。同样, 这种急剧变化也可以用对数项来描述。为此, 我们提出带有对数项的多项式混料模型:

$$\eta_1(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \gamma_i \ln x_i \quad (1-28)$$

$$\eta_2(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i=1}^q \gamma_i \ln x_i \quad (1-29)$$

$$\eta_3(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i < j < k}^q \beta_{ijk} x_i x_j x_k + \sum_{i=1}^q \gamma_i \ln x_i \quad (1-30)$$

$$\eta_4(x) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i < j < k}^q \beta_{ijk} x_i x_j x_k + \sum_{i < j}^q \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{i=1}^q \gamma_i \ln x_i \quad (1-31)$$

由于当 $0 < x_i < 1$ 时 $\ln x_i < 0$, 我们可以在模型(1-28)-(1-31)中把 $\ln x_i$ 换成 $-\ln x_i = \ln \frac{1}{x_i}$, 相应地 γ_i 换成 $-\gamma_i$, 这样, 对数项中的负号就可以反映到系数中去。当 $x_i \rightarrow 0$, $\ln x_i / (\frac{1}{x_i}) = x_i \ln x_i \rightarrow 0$, 因此, $\frac{1}{x_i}$ 趋于无穷的速度要比 $\ln x_i$ 来得快。同时, 由关系 $\ln x_i = -\ln \frac{1}{x_i}$ 可以看出, 取对数起着减缓混料响应急剧变化的作用。所以, 从某种意义上来说, 我们的对数项模型所描述的是混料响应在区域边界附近极端变化不如倒数项所描述的那么剧烈的情况。

对数项混料模型是 Scheffé 多项式加对数项而得到的。既然 $\ln x_i$ 也是用来解释响应的急剧变化的, 我们如同 Draper 和 St. John^[7] 那样也称这些对数项或它们的系数为边效应。注意 $\ln x_i$ 在 1 处的泰勒级数展开为

$$\begin{aligned} \ln x_i &= (x_i - 1) - \frac{1}{2}(x_i - 1)^2 + \frac{1}{3}(x_i - 1)^3 - \dots \\ &+ \frac{(-1)^{k-1}}{k}(x_i - 1)^k + \dots \end{aligned} \quad (1-32)$$

因此, 对数项混料模型是 Scheffé 多项式模型的推广。由于 $\ln x_i$ 的导数为 $\frac{1}{x_i}$, 从多项式次数的角度上, 就可说 $1/x_i$ 的次数要比 $\ln x_i$ 低。所以, 在进行向前序列模型拟合时可以先用倒数项模型, 如不适宜, 再用对数项模型。

令 $z_i = \ln(x_i/x_q)$, $i = 1, 2, \dots, q-1$. 如果 $\gamma_1 + \gamma_2 + \dots + \gamma_q = 0$, 则 $\sum_{i=1}^q \gamma_i \ln x_i = \sum_{i=1}^{q-1} \gamma_i \ln \left(\frac{x_i}{x_q} \right) = \sum_{i=1}^{q-1} \gamma_i z_i$. 更一般地有 $\sum_{i=1}^q \gamma_i \ln x_i = \sum_{i=1}^{q-1} \gamma_i \ln(x_i/x_q) + \left(\sum_{i=1}^q \gamma_i \right) \ln x_q$. 因此, 对数项混料模型还同对数比混料模型有些联系。

由于 $x_i \rightarrow 0$ 时, $\ln x_i \rightarrow -\infty$, 为数学上研究方便, 我们总是假定实际的试验区域不包括这个边界, 且对 x_i 限制为

$$x_i \geq \varepsilon_i > 0, i = 1, 2, \dots, q \quad (1-33)$$

其中 ε_i 是一些由实际问题所给出的充分小量。换言之，试验允许在靠近边界的地方，但不能在边界上进行。从式(1-28)至(1-31)看到，尽管上面每个模型都含所有的 $\ln x_i (i = 1, 2, \dots, q)$ ，但如果试验者能推断出仅有一个或几个分量具有边界效应，那么也可在模型中仅使用有关这些分量的对数项。

如果下列形式的混料约束存在时，

$$x_i \geq a_i > 0, i = 1, 2, \dots, q, \sum_{i=1}^q a_i < 1$$

我们能按照 Kurotori^[9] 提出的方法把这些分量变换成拟分量。但对数项中的混料要求 $x_i > 0, i = 1, 2, \dots, q$ ，为此，实行下列变换：

$$\text{令} \quad a'_i = a_i - \varepsilon_i, i = 1, 2, \dots, q \quad (1-34)$$

这里 ε_i 是一些充分小的正数，使 $a_i \gg \varepsilon_i > 0$ 。

$$\text{再令} \quad x'_i = \frac{x_i - a'_i}{1 - \sum_{i=1}^q a'_i}, i = 1, 2, \dots, q \quad (1-35)$$

当 $x_i = a_i$ 时，就有 $x'_i = \varepsilon_i / (1 - \sum a_i + \sum \varepsilon_i) > 0$ 。因此，边界 $x_i = a_i$ 不是变换到零边界，而是移动到零边界附近。这样一来，就使得对数项在边界 $x_i = a_i$ 上也有意义。

当设计的混料含有 $x_i = 0$ ，并且试验者想使用对数项混料模型来拟合观察数据，可以加一个小的正数到每个 x_i （比如说加 c_i ）上。为使新产生的变量也为混料分量，考虑变换

$$x'_i = (1 - \sum_{i=1}^q c_i) x_i + c_i \quad (1-36)$$

它满足 $x'_i \geq c_i, c_i > 0$ ，且 $\sum_{i=1}^q x'_i = 1$ 。这样，试验者就可以使用关于 x'_i 的对数项混料模型来拟合这些数据。特别地，我们可以令 $c_i = c > 0, i = 1, 2, \dots, q$ 。对于倒数项混料模型，Draper 和 St. John 认为 c_i 取 0.02 至 0.05 之间的值对大部分情况是适宜的，对于对数项混料模型，我们认为 c_i 一般可取 0.01 至 0.05 之间的值。

下面，举两例来说明所提出的新模型的应用，同时把它同倒数项混料模型进行分析和比较。

例 1. Mclean 和 Anderson 的数据。Mclean 和 Anderson^[10] 给出了一个用来说明他们的极端顶点设计的实例。一个闪光装置是混合下列四种化学物质：锰(x_1)，硝酸钠(x_2)，硝酸铈(x_3)和粘合剂(x_4)制成的。这四种化学物质的用量受到如下约束

$$\begin{aligned} 0.40 \leq x_1 \leq 0.60, 0.10 \leq x_2 \leq 0.50 \\ 0.10 \leq x_3 \leq 0.50, 0.03 \leq x_4 \leq 0.08 \end{aligned} \quad (1-37)$$

试验所考虑的观察值(Y)是用 kcd (千坎德拉) 来计算的闪光装置的发光强度。

试验点及所观察的数据如表 1 所示，我们使用一次多项式加对数项模型(1-28)来拟合这些数据。

由于这四个分量有非零下界，因此用公式 (1-35) 把分量 $x_i (i = 1, \dots, 4)$ 变换到拟分量 $x'_i (i = 1, \dots, 4)$ ，其中取 $\varepsilon_i = \varepsilon = 0.02$ 。利用通常的最小二乘方法，就得到关于拟分量 x'_i 模

表 1-1 来自 Mclean 和 Anderson^[10]的数据

试 验 号	Y	x_1	x_2	x_3	x_4
1	75	.4	.1000	.4700	.030
2	180	.4	.1000	.4700	.080
3	195	.6	.1000	.2700	.030
4	300	.6	.1000	.2200	.080
5	145	.4	.4700	.1000	.030
6	230	.4	.4200	.1000	.080
7	220	.6	.2700	.1000	.030
8	350	.6	.2200	.1000	.080
9	220	.5	.1000	.3450	.055
10	260	.5	.3450	.1000	.055
11	190	.4	.2725	.2725	.055
12	310	.6	.1725	.1725	.055
13	260	.5	.2350	.2350	.030
14	410	.5	.2100	.2100	.080
15	425	.5	.2225	.2225	.055

型(1-28)的回归方程

$$\hat{Y}(x') = 96.17x'_1 + 332.56x'_2 + 303.86x'_3 + 2271.94x'_4 + 109.65\ln x'_1 + 63.04\ln x'_2 + 50.71\ln x'_3 - 70.41\ln x'_4 \quad (1-38)$$

以及相应的方差分析表 (表 1-2)。

表 1-2 闪光装置数据对于式(1-38)的方差分析表

来 源	平 方 和	自 由 度	平均平方和	F-比
回归偏差	116102.5	7	16586.07	9.86
剩余偏差	11770.875	7	1681.55	
总 偏 差	127873.375	14		

由于 $9.86 > F(7, 7, 0.01) = 6.99$, 故式(1-38)是显著的。

同样, 对于拟分量 $x'_i (i=1, 2, \dots, 4)$, 用 Draper 和 St. John 的一次加倒数项模型 (1-17), 得到如下回归方程

$$\hat{Y}(x') = 150.178x'_1 + 262.374x'_2 + 218.499x'_3 + 1756.6x'_4 - 7.88/x'_1 - 5.047/x'_2 - 3.928/x'_3 + 3.15/x'_4 \quad (1-39)$$

及相应的方差分析表 (表 1-3)。

表 1-3 闪光装置数据对于式(1-39)的方差分析表

来 源	平 方 和	自 由 度	平均平方和	F-比
回归偏差	116400.75	7	16628.68	10.146
剩余偏差	11472.625	7	1638.95	
总 偏 差	127873.375	14		

这个设计没有重复试验, 不存在 σ^2 的估计, 要判定所用模型的拟合优良性是不可能的。可是剩余标准差 s 可作为判定模型从直观上来讲是否好的一个指标, 另外, 还可使用修改的 R^2 统计量 R_A^2 ,

$$R_A^2 = 1 - \frac{(\text{剩余平方和})/(n-p)}{(\text{总偏差平方和})/(n-1)} \quad (1-40)$$

对于回归方程(1-38), $s = 41$, $R_A^2 = 0.816$; 而对回归方程(1-39), $s = 40.5$, $R_A^2 = 0.82$, 由此可以看出, 用一次加对数项模型来拟合这个闪光装置数据同一次加倒数项模型几乎是同样好的。Snee^[12]使用了许多模型来拟合这些数据, 其中有典型多项式模型, 比模型和 Becker 的齐次模型, Draper 和 St. John^[7]用表列出了这些模型拟合的 s 值和 R_A^2 值, 其中最大的 R_A^2 值才达到 0.75, 最小的 s 值只为 0.48。因此, 可以说, 对数项模型比二次典型多项式模型、比模型和 Becker 的齐次模型对于拟合这个闪光装置数据要好得多。

如果使用向后减法回归过程来精简一次加对数项回归方程(1-38), 就要剔除 x_1' , 并且可得回归方程

$$\begin{aligned} \hat{Y}(x') = & 259.31x_2' + 230.65x_3' + 2475.29x_4' + 113.75\ln x_1' \\ & + 63.1\ln x_2' + 50.78\ln x_3' - 94.6\ln x_4' \end{aligned} \quad (1-41)$$

其 s 值为 39.7, R_A^2 值为 0.83。

例 2. 材料系工程陶瓷研究室应用本书第二章一个确切设计表所进行的试验实例。一种重稀土氧化物坩埚材料是由混合如下三种化学物质: 稀土氧化物 $RE_2O_3(x_1)$, 氧化锆 $ZrO_2(x_2)$, 钼 $Mo(x_3)$ 而制成的, 这三个分量受到如下的约束

$$\begin{aligned} 0.4 & \leq x_1 \leq 0.8 \\ 0.05 & \leq x_2 \leq 0.2 \\ 0.1 & \leq x_3 \leq 0.6 \end{aligned} \quad (1-42)$$

试验要进行的观察指标(Y)是这种坩埚材料的开口孔隙率。

试验点及所得到的观察数据列于表 1-4 中。表 1-4 中有几个点的观察值很小, 第 3 和第 8 个点处, 观察值分别只有 0.67, 1.71, 这些点正是位于利益区域的边界。因此, 响应在边界附近似乎有急剧变化。尽管它们是急剧变化小的, 但只要对每个观察值用一个比较大的数(比如 20)去减, 则得到的响应就是急剧变大的。所以, 对一般问题来说, 无论响应观察值

表 1-4 坩埚材料的试验数据

试 验 号	Y	x_1	x_2	x_3
1	6.61	0.4	0.05	0.55
2	12.30	0.4	0.2	0.4
3	0.67	0.8	0.05	0.15
4	13.95	0.4	0.08	0.52
5	18.985	0.4	0.17	0.43
6	5.82	0.48	0.05	0.47
7	16.115	0.48	0.17	0.35
8	1.71	0.72	0.05	0.23
9	8.36	0.72	0.08	0.20
10	15.28	0.53	0.10	0.37

是急剧变大还是变小, 其实是一样的, 都可以看成是急剧变大的。基于上面的考虑, 我们使用一次加对数项模型(1-28)来拟合这些数据。

在本例中, 尽管分量受到下界约束, 但我们并不作拟分量变换, 而是直接使用原分量。利用最小二乘法, 立刻得到模型(1-28)的回归方程

$$\begin{aligned} \hat{Y}(x) = & 103.91x_1 - 115.13x_2 + 144.92x_3 + 11.11\ln x_1 \\ & + 33.24\ln x_2 - 1.58\ln x_3 \end{aligned} \quad (1-43)$$

以及相应的方差分析表(表 1-5)

表 1-5

坭埭材料数据对于式(1-43)的方差分析表

来 源	平 方 和	自 由 度	平均平方和	F-比
回归偏差	341.86	5	68.372	21.95
剩余偏差	12.46	4	3.115	
总 偏 差	354.32	9		

由于 $21.95 > F(5, 4, 0.01) = 15.5$, 所以, 回归方程(1-43)在 0.01 的水平上是显著的。对于式 (1-43), 我们计算出 $s = 1.765$, $R_A^2 = 0.92$ 。

再用 Draper 和 St. John 的一次加倒数项模型(1-17)来模拟坭埭材料数据, 得到如下回归方程

$$\hat{Y}(x) = 26.73x_1 - 14.24x_2 + 76.30x_3 - 7.48x_1^{-1} - 1.40x_2^{-1} + 0.83x_3^{-1} \quad (1-44)$$

和相应的方差分析表 (表 1-6)

表 1-6

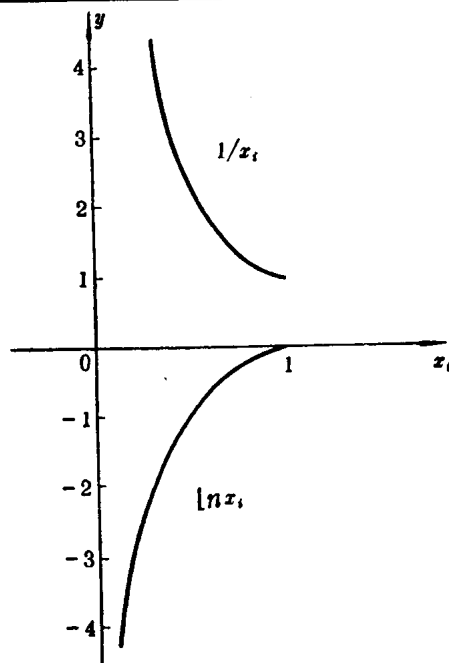
坭埭材料数据对于回归方程(1-44)的方差分析表

来 源	平 方 和	自 由 度	平均平方和	F-比
回归偏差	337.79	5	67.56	16.35
剩余偏差	16.53	4	4.13	
总 偏 差	354.32	9		

从表中的 $16.35 > F(5, 4, 0.01) = 15.5$ 可以看到, 式(1-44)也是显著的。对于方程 (1-44), 同样计算出 $s = 2.033$, $R_A^2 = 0.895$ 。

将式(1-43)的 s 值和 R_A^2 值同式(1-44)的 s 值和 R_A^2 值相比较可以看出, 前者的 s 值小于后者的 s 值, 前者的 R_A^2 值大于后者 R_A^2 的值。因此, 对于本例的坭埭材料数据, 在剩余标准差的意义上, 一次加对数项模型要比一次加倒数项模型拟合的更好。

对于上面两个实例, 用对数项模型和倒数项模型去拟合并无太大的差别, 这一事实可以从 $\ln x_i$ 与 $\frac{1}{x_i}$ 在区间 $(0, 1)$ 中的形状和变化趋势得到解释, 对于自变量 x_i 不太靠近于零的时候, 只要在两个函数的前面配与适当的系数, 其曲线就可比较接近。

图 1-1 函数 $\ln x_i$ 与 $\frac{1}{x_i}$ 的图象

1.3 检验模型中是否含有对数项

首先考虑一般线性回归模型

$$Y = X\beta + \varepsilon \quad (1-45)$$