FXXMXCSGJLLYYY

王新洲　著

非线性模型参数估计理论与应用

全国优秀出版社
武汉大学出版社

非线性模型参数估计理论与应用

王新洲　著

■ 国家自然科学基金项目(49474204;49874002)资助
■ 国家高技术研究发展计划(863 计划)项目(2001AA135081)资助
■ 国家测绘局测绘科技发展基金项目(99010)资助
■ 测绘遥感信息工程国家重点实验室开放研究基金(测绘基础研究)项目((00)0204)资助

武汉大学出版社

图书在版编目(CIP)数据

非线性模型参数估计理论与应用/王新洲著. —武汉：武汉大学出版社，2002.6
ISBN 7-307-03515-4

Ⅰ.非… Ⅱ.王… Ⅲ.非线性—数学模型—参数估计 Ⅳ.O211.67

中国版本图书馆 CIP 数据核字(2002)第 018702 号

责任编辑：夏炽元　　责任校对：张　昕　　版式设计：支　笛

出版：武汉大学出版社　(430072　武昌　珞珈山)
(电子邮件：wdp4@whu.edu.cn　网址：www.wdp.whu.edu.cn)
发行：新华书店湖北发行所
印刷：湖北恒吉印务有限公司
开本：787×1092　1/8　印张：9.125　字数：215 千字
版次：2002 年 6 月第 1 版　　2002 年 6 月第 1 次印刷
ISBN 7-307-03515-4/O・262　　定价：15.00 元

序

我们知道，所谓参数估计，它是数理统计中的一个名词，其含义是根据含有误差的观测向量，依一定的数学模型和准则求解未知参数.实际上在测量工作中这就是平差问题。通常无论是在数理统计，还是测量平差中，依据线性数学模型的参数估计理论已是非常成熟而系统的理论，是其主要研究和应用的对象.可是一般情况下，包括我们测量工作，非线性数学模型的出现要比线性模型频繁得多，对此，传统的方法是将数学模型进行线性化近似，即将非线性问题转化成线性问题来处理.在过去测量精度不高的情况下，可以忽略由此线性化近似而引起的数学模型的误差.可是这种线性化近似已不能满足现代高精度测量工作的要求，它完全可能导致参数估计结果的精度扭曲.为此本书作者于1994～2001年期间在国家自然科学基金项目“非线性随机模型估计理论并用于全球定位系统”等项目的资助下，深入系统地研究了非线性模型的参数估计理论及其在测量数据处理中的应用，取得了一系列高水平的成果.本书就是在这些研究成果的基础上，通过整理扩充而成的.其主要特点，一是直接对非线性模型进行处理，从而精确地揭示了非线性模型中估计量的统计性质，从理论上解决了由于非线性模型的线性化近似所引起的各种问题；二是紧密结合测量数据处理实际，抓住非线性模型参数估计的算法和精度评定两个基本问题进行深入浅出的论述，在严谨的理论分析的基础上给出了便于实际应用的公式.所以本书既有较高的理论水平，又有较大的实用价值。

本书作者王新洲教授从事测量数据处理理论与应用的教学与科研工作十余年，具有较强的科研能力和较高的业务水平，在这一学科领域既取得了一些高水平成果，也积累了较丰富的实践经验，同时具有很强的写作能力.本书不仅文字流畅、层次分明，而且论证严谨、逻辑性强.我相信这会是一本深受读者欢迎的好书！

宁津生

2002年5月5日

前 言

非线性模型参数估计理论是国家自然科学基金委员会于1994年在自然科学学科发展战略调整调查报告《大地测量学》一书中提出的大地测量学学科发展面临的重大基础理论问题之一，也是国际大地测量协会大地测量数学与物理基础“统计学”专题研究组1991～1995年的重点研究内容之一.因为测量上大量的数学模型都是非线性模型，很多观测方程都具有很强的非线性性，对级数展开的初值点十分敏感，所以需要研究模型空间的容许曲率问题，以及在非线性函数空间的平差理论和可普及的实用方法.

对于大地测量中大量的非线性模型，传统的方法是进行线性近似，即将其展开为泰勒级数，取至一次项，而略去二次以上各项.如此线性近似，必然会引起模型误差.由于过去测量精度不高，线性近似所引起的模型误差一般小于观测误差，故线性近似所引起的模型误差可忽略不计.随着科学技术的不断发展，现在的测量精度已大大提高，致使线性近似所引起的模型误差与观测误差相当，甚至还会大于观测误差.因此，用近似的理论、模型、方法去处理具有很高精度的观测结果，从而导致精度损失，显然是不合理的.现代科学技术要求估计结果的精度尽可能提高.这样，传统的线性近似的方法就不能满足当今科学技术的要求.另外，有些非线性模型对参数的近似值十分敏感，若近似值的精度较差，线性近似时就会产生较大的模型误差.此时用线性模型的精度评定理论去评定估计结果的精度，会得到一些虚假的优良统计性质，人为地拔高了估计结果的精度.

数理统计中对非线性参数估计的研究始于20世纪60年代初期.但在开始的20年中进展并不快.直到20世纪80年代初，加拿大统计学家Bates和Watts引进曲率度量后此理论才得到较快的发展.

本书的内容主要来自我们的研究成果.同时参考了国内外一系列资料和最新研究成果.本书的主要内容包括：线性模型参数估计理论的回顾；非线性模型非线性强度的度量理论；非线性最小二乘估计的各种算法及其效率比较；非线性最小二乘估计量的统计性质及精度评定；非线性模型的其他估计方法等.本书从理论上解决了线性近似所引起的问题，是将非线性科学引入测绘领域的一个开端.

本书附录A中介绍的立体阵，在本书中经常用到，建议读者事先熟练掌握它.

由于作者水平有限，谬误之处在所难免，恳请各位专家和读者不吝赐教.

作 者

2002年2月22日于武昌

目　录

前　言 ………………………………………………………… (1)

第一章　参数估计概述 ………………………………………… (1)
　第一节　线性模型参数估计理论的回顾 ……………………… (1)
　第二节　非线性模型参数估计问题的提出及其进展 ………… (28)

第二章　非线性模型非线性强度的度量 ……………………… (31)
　第一节　非线性强度的概念 …………………………………… (31)
　第二节　非线性强度的度量 …………………………………… (34)
　第三节　曲率立体阵 …………………………………………… (41)
　第四节　非线性模型线性近似的容许曲率 …………………… (44)

第三章　非线性最小二乘估计 ………………………………… (51)
　第一节　非线性最小二乘估计的定义及存在性定理 ………… (51)
　第二节　非线性最小二乘估计的近似解法 …………………… (52)
　第三节　非线性最小二乘估计的迭代解法 …………………… (53)
　第四节　几种迭代算法的效率比较 …………………………… (66)
　第五节　非线性最小二乘估计的直接解法 …………………… (71)
　第六节　非线性最小二乘估计的其他算法 …………………… (75)

第四章　非线性最小二乘估计的统计性质与精度评定 ……… (82)
　第一节　非线性最小二乘估计的统计性质 …………………… (82)
　第二节　单位权方差 σ^2 的估计 …………………………… (89)
　第三节　非线性函数的误差传播 ……………………………… (95)

第五章　非线性模型参数的其他估计 ………………………… (99)
　第一节　带约束的非线性模型参数估计 ……………………… (99)
　第二节　非线性模型参数的稳健估计 ………………………… (101)
　第三节　非线性模型参数的拟似然估计 ……………………… (103)
　第四节　非线性模型参数的贝叶斯估计 ……………………… (104)
　第五节　非线性模型非参数小波估计 ………………………… (107)
　第六节　非线性模型半参数估计 ……………………………… (108)

第六章　非线性模型参数估计在测量上的应用……（110）
第一节　在导线网平差中的应用……（110）
第二节　在工程测量中的应用……（112）
第三节　在摄影测量中的应用……（113）
第四节　在GPS定位中的应用……（114）

附录A　立体阵的定义、运算及其性质……（116）
附录B　蒙特卡罗积分……（123）
附录C　解非线性方程组的一类离散的Newton算法……（127）

第一章　参数估计概述

参数估计(Parameter Estimation)是一种基本的统计推断形式,也是数理统计学的一个重要分支[1](成平,陈希孺等,1985),更是测量数据处理(Surveying Data Processing)理论的重要组成部分.由于迄今为止,参数估计的一系列成果主要集中在线性模型(Line Model),而且线性模型参数估计理论是非线性模型(Nonlinear Model)参数估计的基础,它们之间有很多联系.所以,本章首先简要回顾线性模型参数估计中的基本理论和方法.

第一节　线性模型参数估计理论的回顾

线性模型是数理统计学中发展比较早的分支之一.关于它的参数估计问题,可以追溯到18世纪初[2][3](王松桂,1987).1806年著名数学家A.M.Legendre在《决定彗星轨道新方法》中从代数观点提出了最小二乘法(Least Square).而早在1794年,年仅17岁的高斯(C.F.Gauss)就提出用最小二乘法从带有误差的观测值中找出待定量的最优值.但高斯只到1809年才在《天体运动的理论》中正式发表他的方法[4](武汉测绘科技大学测量平差教研室,1996).后来,马尔可夫(A.A.Markov)于1900年证明了最小二乘估计的方差(Variance)最小的性质.形成了著名的Gauss-Markov定理,从而奠定了最小二乘法在线性模型参数估计中的地位.

1944年,R.C.Bose引入了可估函数的概念,加之广义逆矩阵的应用,使得设计矩阵为列降秩的线性模型参数估计理论表述得更加严密而简洁[2][3](王松桂,1987).20世纪60年代中期开始研究观测误差的方差-协方差矩阵为奇异矩阵的线性模型参数估计问题.Goldman和Zelen率先提出了用满秩线性变换把估计模型化为方差-协方差矩阵为$\sigma^2 I$,且带有线性约束的情形.后来C.R.Rao采用推广最小二乘法的途径,提出了所谓的"最小二乘法统一理论".该统一理论既适合于设计矩阵列满秩或列降秩,又适合于观测误差的方差-协方差矩阵奇异的情况.这些结果构成了线性模型最小二乘估计理论的基本内容[2][3](王松桂,1987).

一、线性模型参数的最小二乘估计

设线性模型为

$$\begin{cases} L = BX + \Delta \\ \mathrm{E}(\Delta) = 0 \\ \mathrm{Var}(\Delta) = \sigma^2 Q_{LL} \end{cases} \tag{1-1-1}$$

式中:L为$n\times n$的观测向量;B为$n\times t$的设计矩阵;X为$t\times 1$的未知参数向量;Δ为$n\times 1$的观测误差向量;σ^2为单位权方差因子;Q_{LL}为$n\times n$的协因数矩阵,且$Q_{LL}>0$.

线性模型(1-1-1)式是测量平差(Surveying Adjustment)中最常用的数学模型.其中第一式通常称为观测方程(Observation Equation),测量平差基础中称之为函数模型(Function Model).而将(1-1-1)式中的第二、第三式称之为随机模型[4](Stochastic Model)(武汉测绘科技大学测量平差教研室,1996).

记未知参数 X 的最小二乘估计为$\hat{X}_{\mathrm{LS}}$;残差(Residual)向量为 V.用 $\hat{X}_{\mathrm{LS}}$和 V 代替X 和 Δ,则观测方程变为

$$V = B\hat{X}_{\mathrm{LS}} - L \tag{1-1-2}$$

测量上通常称(1-1-2)式为误差方程(Error Equation).

再记观测值的权矩阵为

$$P = P' = Q_{LL}^{-1} \tag{1-1-3}$$

线性模型(1-1-1)式中未知参数 X 的最小二乘估计,就是寻求 X 的一个估值,使

$$V'PV = (B\hat{X}_{\mathrm{LS}} - L)'P(B\hat{X}_{\mathrm{LS}} - L) = \min \tag{1-1-4}$$

为此,将(1-1-4)式对 $\hat{X}_{\mathrm{LS}}$求一阶导数,并令其为零,得

$$\frac{\mathrm{d}V'PV}{\mathrm{d}\hat{X}_{\mathrm{LS}}} = V'PB = 0$$

即

$$B'PV = 0 \tag{1-1-5}$$

将(1-1-2)式代入(1-1-5)式,得

$$B'PB\hat{X}_{\mathrm{LS}} - B'PL = 0 \tag{1-1-6}$$

测量平差中称(1-1-6)式为法方程(Normal Equation).

法方程在最小二乘估计中起着非常重要的作用.

当设计矩阵 B 列满秩时,由法方程(1-1-6)式可解得未知参数 X 的最小二乘解为

$$\hat{X}_{\mathrm{LS}} = (B'PB)^{-1}B'PL \tag{1-1-7}$$

当设计矩阵 B 列降秩时,法方程(1-1-6)的系数矩阵 $B'PB$ 也降秩.在这种情况下,法方程(1-1-6)式的惟一解为

$$\hat{X}_{\mathrm{LS}} = (B'PB)^{+}B'PL \tag{1-1-8}$$

测量平差中称这种情况为秩亏自由网平差(Adjustment of Free-Network with Rank Deficiency).

二、线性模型参数最小二乘估计的统计性质

1.无偏性

对(1-1-7)式所确定的最小二乘估计取数学期望,得

$$\mathrm{E}(\hat{X}_{\mathrm{LS}}) = (B'PB)^{-1}B'P\mathrm{E}(L) \tag{1-1-9}$$

由(1-1-1)式可得

$$\mathrm{E}(L) = BX + \mathrm{E}(\Delta) = BX \tag{1-1-10}$$

将(1-1-10)式代入(1-1-9)式,得

$$\mathrm{E}(\hat{X}_{\mathrm{LS}}) = (B'PB)^{-1}B'PBX = X \tag{1-1-11}$$

(1-1-11)式表明最小二乘估计 $\hat{X}_{\mathrm{LS}}$是未知参数 X 的无偏估计.

2.方差最小性

设线性模型(1-1-1)式中参数 X 的最优线性无偏估计(BLUE)为

$$\hat{X}=ML \tag{1-1-12}$$

式中:M 为$t\times n$ 的待定系数矩阵.

对(1-1-12)式取数学期望,得

$$E(\hat{X})=ME(L)=MBX$$

因为 $\hat{X}=ML$ 为最优线性无偏估计,所以有

$$MB=I \tag{1-1-13}$$

式中:I 为$t\times t$ 的单位阵.

由(1-1-12)式应用协因数传播律,得

$$MQ_{LL}M'=\min \tag{1-1-14}$$

下面来求满足条件(1-1-13)式,又使(1-1-14)式成立的矩阵 M. 为此,组成函数

$$\Phi=MQ_{LL}M'+2(MB-I)K \tag{1-1-15}$$

(1-1-15)式对 M 求偏导数,并令其为零,得

$$\frac{\partial\Phi}{\partial M}=2Q_{LL}M'+2BK=0$$

即

$$MQ_{LL}+K'B'=0$$

于是有

$$M=-K'B'Q_{LL}=-K'B'P \tag{1-1-16}$$

将(1-1-16)式代入(1-1-13)式,得

$$-K'B'PB=I$$

所以有

$$K=-(B'PB)^{-1} \tag{1-1-17}$$

将(1-1-17)式代入(1-1-16)式得最优线性无偏估计的系数矩阵为

$$M=(B'PB)^{-1}BP \tag{1-1-18}$$

将(1-1-18)式代入(1-1-12)式,得线性模型中未知参数 X 的最优线性无偏估计为

$$\hat{X}=(B'PB)^{-1}B'PL=\hat{X}_{LS} \tag{1-1-19}$$

所以 $\hat{X}_{LS}$为线性模型参数的最优线性无偏估计. 由此证明线性模型最小二乘估计 $\hat{X}_{LS}$ 具有最小方差.

三、单位权方差因子 σ^2 的惟一非负最优二次无偏估计

线性模型(1-1-1)式中单位权方差因子 σ^2 由下式估计:

$$\hat{\sigma}^2=\frac{V'PV}{n-t} \tag{1-1-20}$$

式中:n 为观测值的个数;t 为必要观测的个数.

可以证明,$\hat{\sigma}^2$ 是单位权方差因子 σ^2 的惟一最小方差非负二次无偏估计. 下面仅证 $\hat{\sigma}^2$ 的无偏性. 至于 $\hat{\sigma}^2$ 的其他性质的证明,有兴趣的读者可参阅《线性牛型的理论及其应用》[2][3](王松桂,1987).

为了证明 $\hat{\sigma}^2$ 的无偏性,我们先证明一个定理.

定理 1-1-1 设 y 为$n\times 1$ 的随机向量,$E(y)=\mu$,$Var(y)=D$,M 为$n\times n$ 的对称方

阵,则

$$E(Y'MY)=\mathrm{tr}(MD)+\mu'M\mu \tag{1-1-21}$$

式中:tr(Z)表示求方阵 Z 的迹.

证明:

$$\begin{aligned}
&E(Y'MY)\\
&=E[(y-\mu)'M(y-\mu)+\mu'My+y'M\mu-\mu'M\mu]\\
&=E[(y-\mu)'M(y-\mu)]+2E(y'M\mu)-\mu'M\mu\\
&=E\{\mathrm{tr}[(y-\mu)'M(y-\mu)]\}+\mu'M\mu\\
&=E\{\mathrm{tr}[M(y-\mu)(y-\mu)']\}+\mu'M\mu\\
&=\mathrm{tr}\{ME[(y-\mu)(y-\mu)']\}+\mu'M\mu\\
&=\mathrm{tr}(MD)+\mu'M\mu
\end{aligned}$$

定理 1-1-1 得证.

下面再来证明 $\hat{\sigma}^2$ 的无偏性.

因为由(1-1-2)式,顾及(1-1-1)式,有

$$\begin{aligned}
E(V)&=B(\hat{X}_{\mathrm{LS}})-E(L)\\
&=BX-BX=0
\end{aligned} \tag{1-1-22}$$

$$\mathrm{Var}(V)=\sigma^2(Q_{LL}-B(B'PB)^{-1}B') \tag{1-1-23}$$

于是,由定理 1-1-1,顾及(1-1-3)式,有

$$\begin{aligned}
E(V'PBV)&=\sigma^2\mathrm{tr}\{P(Q_{LL}-B(B'PB)^{-1}B')\}\\
&=\sigma^2[\mathrm{tr}(PQ_{LL})-\mathrm{tr}(PB(B'PB)^{-1}B')]\\
&=\sigma^2[\mathrm{tr}(\underset{n\times n}{I})-\mathrm{tr}(\underset{t\times t}{I})]\\
&=\sigma^2(n-t)
\end{aligned}$$

所以

$$E(\hat{\sigma}^2)=\frac{E(V'PV)}{n-t}=\sigma^2 \tag{1-1-24}$$

因为 $\hat{\sigma}^2$ 是基于 X 的最小二乘估计 $\hat{X}_{\mathrm{LS}}$导出的,所以有文献称 $\hat{\sigma}^2$ 为 σ^2 的最小二乘估计.

四、线性模型最小二乘估计的序贯解法

当我们已经得到线性模型的最小二乘估计后,又因某种原因增加了一个或多个观测值时,为了利用以前的估计结果,不再额外增加估计工作量,就可以采用最小二乘估计的序贯解法.为此,将线性模型(1-1-1)式中的观测向量 L 分为两组,记为 L_{k-1}和 L_k.它们的权矩阵分别记为 P_{k-1}和 P_k,并设 L_{k-1}与 L_k 不相关.则误差方程(1-1-2)式可写为

$$\begin{cases}V_{k-1}=B_{k-1}\hat{X}_{\mathrm{LS}}-L_{k-1}\\ V_k=B_k\hat{X}_{\mathrm{LS}}-L_k\end{cases} \tag{1-1-25}$$

式中:L_{k-1}为 $n_{k-1}\times 1$ 的观测向量;L_k 为 $n_k\times 1$ 的观测向量;V_{k-1}为 $n_{k-1}\times 1$ 的残差向量;V_k 为 $n_k\times 1$ 的残差向量;B_{k-1}为 $n_{k-1}\times 1$ 的设计矩阵,$n_{k-1}>t$,且有 $\mathrm{rk}(B_{k-1})=t$(注:rk(B)表示矩阵 B 的秩),即 B_{k-1}列满秩;B_k 为 $n_k\times 1$ 的设计矩阵.与(1-1-25)式相应

的权矩阵为 $P=\begin{pmatrix} P_{k-1} & 0 \\ 0 & P_k \end{pmatrix}$.

单独用(1-1-25)式中的第一式估计 $\hat{X}_{\mathrm{LS}}$，得

$$\hat{X}_{\mathrm{LS}}^{k-1}=(B_{k-1}'P_{k-1}B_{k-1})^{-1}B_{k-1}'P_{k-1}L_{k-1} \tag{1-1-26}$$

而用(1-1-25)式整体估计 $\hat{X}_{\mathrm{LS}}$，得

$$\begin{aligned}\hat{X}_{\mathrm{LS}}&=(B_{k-1}'P_{k-1}B_{k-1}+B_k'P_kB_k)^{-1}(B_{k-1}'P_{k-1}L_{k-1}+B_k'P_kL_k)\\&=\hat{X}_{\mathrm{LS}}^{k-1}+(B_{k-1}'P_{k-1}B_{k-1}+B_k'P_kB_k)^{-1}B_k'P_k(L_k-B_k\hat{X}_{\mathrm{LS}}^{k-1})\end{aligned} \tag{1-1-27}$$

令

$$\begin{cases} Q_{X_{k-1}}=(B_{k-1}'P_{k-1}B_{k-1})^{-1} \\ Q_X=(Q_{X_{k-1}}^{-1}+B_k'P_kB_k)^{-1} \end{cases} \tag{1-1-28}$$

则(1-1-27)式可写为

$$\begin{aligned}\hat{X}_{\mathrm{LS}}&=\hat{X}_{\mathrm{LS}}^{k-1}+Q_XB_k'P_k(L_k-B_k\hat{X}_{\mathrm{LS}}^{k-1})\\&=\hat{X}_{\mathrm{LS}}^{k-1}+J(L_k-B_k\hat{X}_{\mathrm{LS}}^{k-1})\end{aligned} \tag{1-1-29}$$

式中:

$$J=Q_XB_k'P_k \tag{1-1-30}$$

由(1-1-28)式的第二式可得

$$Q_{X_{k-1}}^{-1}=Q_X^{-1}-B_k'P_kB_k \tag{1-1-31}$$

将(1-1-31)式两边左乘 Q_X，得

$$Q_XQ_{X_{k-1}}^{-1}=I-JB_k \tag{1-1-32}$$

将(1-1-32)式两边右乘 $Q_{X_{k-1}}$，得

$$Q_X=Q_{X_{k-1}}-JB_kQ_{X_{k-1}} \tag{1-1-33}$$

由(1-1-28)式的第二式，应用矩阵反演公式，可得

$$Q_X=Q_{X_{k-1}}-Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1}B_kQ_{X_{k-1}} \tag{1-1-34}$$

比较(1-1-33)式与(1-1-34)式，得

$$J=Q_XB_k'P_k=Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1} \tag{1-1-35}$$

将(1-1-35)式代入(1-1-29)式，得最小二乘序贯解的递推公式为

$$\hat{X}_{\mathrm{LS}}=\hat{X}_{\mathrm{LS}}^{k-1}+Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1}(L_k-B_k\hat{X}_{\mathrm{LS}}^{k-1}) \tag{1-1-36}$$

当 $n_k=1$ 时，$(P_k^{-1}+B_kQ_{X_{k-1}}B_k')$ 和 $L_k-B_k\hat{X}_{\mathrm{LS}}^{k-1}$ 都是一个数. 因此在这种情况下，最小二乘序贯解计算非常简单，所以总是假定 $n_k=1$，逐次递推，直到求出整体估计时的最后结果[4]（武汉测绘科技大学测量平差教研室，1996）. 最小二乘序贯解法可应用于测量控制网平差后又因某种原因增加观测值的情况.

五、带有线性约束的线性模型参数估计

以上讨论的线性模型参数估计，都是假定未知参数之间是相互独立的，故对未知参数 X 没有附加任何约束条件，它是自由参数. 但在实际工作中，往往在未知参数 X 之间并非相互独立，而是存在线性关系（当然也可能是存在非线性关系，在此仅讨论线性关系）. 此时，线

性模型(1-1-1)式就变为

$$\begin{cases} L = BX + \Delta \\ CX + C_0 = 0 \\ \mathrm{E}(\Delta) = 0 \\ \mathrm{Var}(\Delta) = \sigma^2 Q_{LL} \end{cases} \tag{1-1-37}$$

用 V 和 $\hat{X}_{\mathrm{LS}}$代替上式中的 Δ 和 X 可得

$$\begin{cases} V = B\hat{X}_{\mathrm{LS}} - L \\ C\hat{X}_{\mathrm{LS}} + C_0 = 0 \end{cases} \tag{1-1-38}$$

求(1-1-38)式的最小二乘解,就是在满足约束条件 $C\hat{X}_{\mathrm{LS}} + C_0 = 0$ 的情况下,求 $V'PV$ 的极小值.为此,按求条件极值的方法组成函数

$$\varphi = V'PV = 2K'(C\hat{X}_{\mathrm{LS}} + C_0) \tag{1-1-39}$$

将上式对 $\hat{X}_{\mathrm{LS}}$求偏导数,并令其为零,得

$$\frac{\partial \varphi}{\partial \hat{X}_{\mathrm{LS}}} = 2V'P\frac{\partial V}{\partial \hat{X}_{\mathrm{LS}}} + 2K'C = 0$$

由此得

$$B'PV + C'K = 0 \tag{1-1-40}$$

将(1-1-40)式与(1-1-38)式联立求解,消去 V 后,得

$$\left.\begin{array}{l} B'PB\hat{X}_{\mathrm{LS}} + C'K - B'PL = 0 \\ C\hat{X}_{\mathrm{LS}} + C_0 = 0 \end{array}\right\} \tag{1-1-41}$$

将(1-1-41)式的第一式左乘 $C(B'PB)^{-1}$后,减去第二式,得

$$C(B'PB)^{-1}C'K - (C(B'PB)^{-1}B'PL + C_0) = 0$$

解得

$$K = (C(B'PB)^{-1}C')^{-1}(C(B'PB)^{-1}B'PL + C_0) \tag{1-1-42}$$

将(1-1-42)式代入(1-1-41)式的第一式,得参数 X 的最小二乘估计为

$$\begin{aligned} \hat{X}_{\mathrm{LS}} = & (B'PB)^{-1}\{I - C'(C(B'PB)^{-1}C')^{-1}C(B'PB)^{-1}\}B'PL \\ & - (B'PB)^{-1}C'(C(B'PB)^{-1}C')^{-1}C_0 \end{aligned} \tag{1-1-43}$$

六、线性模型参数最小二乘估计的统一理论

在线性模型(1-1-1)式中,若 Q_{LL} 为已知的奇异方阵,即 $|Q_{LL}| = 0$,则称之为奇异线性模型,即

$$L = BX + \Delta, \mathrm{E}(\Delta) = 0, \mathrm{Var}(\Delta) = \sigma^2 Q_{LL}, |Q_{LL}| = 0 \tag{1-1-44}$$

由前述知,线性模型(1-1-1)式的最小二乘估计就是寻求使

$$V'PV = V'Q^{-1}V = \min \tag{1-1-45}$$

的未知参数的估值.但在奇异线性模型(1-1-44)中,Q_{LL} 的凯利逆 Q_{LL}^{-1}不存在.这就使前述的一般最小二乘理论不能应用.此时,人们自然地会想到用 Q_{LL} 的广义逆 Q_{LL}^{-}来代替 Q_{LL} 的凯利逆 Q_{LL}^{-1}.这样就变成求解 $V'Q_{LL}^{-}V = \min$ 的问题.但由于 Q_{LL}^{-} 不惟一,$V'Q_{LL}^{-}V$ 就不惟一,即极小值不惟一.为了解决这个问题,印度统计学家 C. R. Rao 提出了一种有效方法,从而建立了线性模型参数最小二乘估计的统一理论.这一统一理论的基本思想是,寻求一个矩

阵 M 代替观测值的协因数阵 Q_{LL}，然后解优化问题：

$$V'M^{-}V=\min \tag{1-1-46}$$

在求解优化问题(1-1-46)式之前，人们自然会问，M 矩阵是否存在？若 M 矩阵存在，它的形式是怎样的？Rao 已证明[3]（陈希孺，王松桂，1987）可以取 $M=Q_{LL}+BGB'$，其中 G 为对称矩阵，且满足 $\mathrm{rk}(G)=\mathrm{rk}(Q_{LL}\mid B)$. 这样的 M 适于 $Q_{LL}\geqslant 0$ 的情况. 当 $Q_{LL}>0$ 时，取 $G=0$. 当 $|Q_{LL}|=0$ 时，G 的一个简单的选择是 $G=kI, k>0$. 当 $M=Q_{LL}+BGB'$，G 为对称矩阵，且 $\mathrm{rk}(G)=\mathrm{rk}(Q_{LL}\mid B)$ 时，$V'M^{-}V$、$B'M^{-}B$ 和 $B'M^{-}L$ 都与 M^{-} 的选择无关[3]（证明见陈希孺，王松桂，1987）. 于是求解优化问题(1-1-46)式，得

$$\hat{X}_{\mathrm{LSR}}=(B'M^{-}B)^{-}B'M^{-}L \tag{1-1-47}$$

$\hat{X}_{\mathrm{LSR}}$表示由 C. R. Rao 给出的最小二乘估计.

七、线性模型参数的稳健估计

线性模型最小二乘估计是线性最优无偏估计，这一结论是在假设 $\mathrm{E}(\Delta)=0$，$\mathrm{Var}(\Delta)=\sigma^2Q_{LL}$下得到的. 但在实际中，由于观测向量 L 中往往不可避免地存在粗差(Gross Error)，这时，线性模型中的假设就得不到满足. 当观测值不满足线性模型的假设条件，即观测值遭到粗差污染时，最小二乘估计具有明显的负面影响，即最小二乘估计不具有抵抗粗差干扰的特性，单个观测值的偏差也可能导致最小二乘估计面目全非[5]（杨元喜，1993）. 为此，人们就试图寻求一类参数估计方法，使其具有下述特点：

(1) 在假定的观测分布模型下，估值应是最优或接近最优的；

(2) 假设的分布模型与实际的分布模型有较小差异时，估值受粗差的影响较小；

(3) 当假设分布模型与实际分布模型有较大偏差时，估值不受破坏性的影响[6][7]（罗永光，王海云，1987；周江文，黄幼才等，1997）.

前两个特点是要求稳健估计在假定模型正确或假定模型略有偏差时，对估计结果影响较小. 这就只需要模型近似正确. 这两个特点的代价是在模型正确时，估计结果并非最优，只是接近最优. 第三个特点保证了在相当坏的情况下，估计结果也不会变得太坏. 在线性模型参数估计中，具有如上特点的估计称为线性模型参数的稳健估计，简称稳健估计(Robust Estimation).

稳健估计是一种能抵抗粗差干扰的估计. 其实质就是牺牲最小二乘估计的最优性，达到抵抗粗差污染的目的. 下面扼要介绍稳健估计. 为推导公式方便，将误差方程(1-1-2)式改写为

$$V=B\hat{X}_{\mathrm{R}}-L=\begin{pmatrix} b_1\\ b_2\\ \vdots\\ b_n\end{pmatrix}\hat{X}_{\mathrm{R}}-\begin{pmatrix} L_1\\ L_2\\ \vdots\\ L_n\end{pmatrix} \tag{1-1-48}$$

式中，b_i 为设计矩阵B 的第 i 行；$\hat{X}_{\mathrm{R}}$ 表示未知参数 X 的稳健估计.

设第 i 个观测值的权为 p_i，则按 M 估计原理，未知参数 X 的稳健估计就是求解下列优化问题：

$$\sum_{i=1}^{n}p_i\rho(v_i)=\sum_{i=1}^{n}p_i\rho(b_i\hat{X}_{\mathrm{R}}-L_i)=\min \tag{1-1-49}$$

上式对 $\hat{X}_R$ 求导数，并令其为零．同时记 $\varphi(v_i)=\dfrac{\partial\rho}{\partial v_i}$，则有

$$\sum_{i=1}^{n} p_i\varphi(v_i)b_i = 0 \tag{1-1-50}$$

令
$$\frac{\varphi(v_i)}{v_i}=W_i,\bar{P}_{ii}=P_iW_i \tag{1-1-51}$$

式中：W_i 称为权因子，$\bar{P}_{ii}$称为等价权．

于是(1-1-50)式可写为

$$B'\bar{P}V=0 \tag{1-1-52}$$

将(1-1-48)式代入(1-1-52)式，得

$$B'\bar{P}B\hat{X}_R-B'\bar{P}L=0 \tag{1-1-53}$$

由此可得

$$\hat{X}_R=(B'\bar{P}B)^{-1}B'\bar{P}L \tag{1-1-54}$$

由于等价权 $\bar{P}$ 的引入，使得(1-1-54)式既能抵抗粗差的污染，又保留了最小二乘估计的形式．使最小二乘估计公式简单，计算方便的优点得到充分体现．所以周江文等[7]称(1-1-54)式为抗差最小二乘估计．

由(1-1-51)式可以看出，权因子 W_i 是残差 v_i 的非线性函数．为了使等价权更切合实际，需要通过迭代计算，以改善权因子．

八、线性模型参数的信息扩散估计

由前述知，当观测值服从正态分布假设成立时，线性模型参数的最小二乘估计是最优线性无偏估计．但当观测值遭到粗差污染时，最小二乘估计具有明显的负面影响．单个粗差都可使最小二乘估计面目全非．为了解决这个问题，数理统计学家们提出了具有抵抗粗差能力的稳健估计．稳健估计是在假设观测值服从污染分布的前提下，通过牺牲最优性来达到抵抗粗差的目的．在实际的参数估计过程中，人们一般不是选用最小二乘估计，就是选用稳健估计．由于不知道观测值究竟服从什么分布，所以不论选用哪种估计，都要担很大的风险[8]（王新洲，1999）．例如，当采用最小二乘估计时，要担观测值不服从正态分布，估计结果受粗差污染的风险；当采用稳健估计时，要担观测值服从正态分布，而估计结果并非最优的风险．根据担此风险的原因知，如果对任何一组观测值，在参数估计之前，就能估计出它所服从的具体分布，然后再根据此分布进行参数估计，就可避免上述风险．信息扩散估计（Information Diffusion Estimation）就是这样避免上述风险的一种参数估计方法．“信息扩散估计是一种智能估计，即当观测值中不含粗差（错误或异常值）时，参数估计的结果与最小二乘估计一样，是最优无偏估计；当观测值中含有粗差时，该估计不仅能很好地抵御粗差的影响，而且不需要迭代，比现有的稳健估计方法都好”[9]（黄崇福，2002）．

1．信息扩散原理与扩散估计

设 $W=\{w_1,w_2,\cdots,w_n\}$ 是知识样本，L 是基础论域，记 w_i 的观测值为 l_i．再设 $x=\varphi(l-l_i)$，则当 W 非完备时，存在函数 $\mu(x)$，使 l_i 点获得的量值为 1 的信息可按 $\mu(x)$ 的量值扩散到 l 上去．且扩散所得到的原始信息分布 $Q(l)=\sum_{j=1}^{n}\mu(x)=\sum_{j=1}^{n}\mu(\varphi(l-l_i))$ 能

更好地反映 W 在总体的规律.这一原理称为信息扩散原理[10](黄崇福,王家鼎,1995).

根据这一原理对母体概率密度函数的估计称为扩散估计.扩散估计的确切定义如下:

定义 1-1-1 设 $\mu(x)$为定义在$(-\infty,\infty)$上的一个波雷尔可测函数,$\Delta_n>0$ 为常数,则称

$$\hat{f}(l)=\frac{1}{n\Delta_n}\sum_{j=1}^{n}\mu\left(\frac{l-l_j}{\Delta_n}\right) \tag{1-1-55}$$

为母体概率密度函数 $f(l)$的一个扩散估计[10](黄崇福,王家鼎,1995).式中:$\mu(x)$称为扩散函数;Δ_n 称为窗宽.

$$x=\frac{l-l_j}{\Delta_n} \tag{1-1-56}$$

2.扩散函数 $\mu(x)$的确定

由(1-1-55)式知,扩散估计的关键是扩散函数 $\mu(x)$的具体形式.对于不同的 $\mu(x)$,可得到不同的扩散估计结果.在《模糊信息优化处理技术及其应用》[10](黄崇福,王家鼎,1995)中根据分子扩散理论,导出的正态扩散函数为

$$\mu(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{1-1-57}$$

将(1-1-57)式代入(1-1-55)式,顾及(1-1-56)式,得母体概率密度函数 $f(l)$的正态扩散估计为

$$\hat{f}(l)=\frac{1}{n\Delta_n}\sum_{j=1}^{n}\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\left(\frac{l-l_j}{\Delta_n}\right)^2}{2\sigma^2}\right]\right\}=\frac{1}{nh\sqrt{2\pi}}\sum_{j=1}^{n}\exp\left[-\frac{(l-l_j)^2}{2h^2}\right] \tag{1-1-58}$$

式中:

$$h=\sigma\Delta_n \tag{1-1-59}$$

(1-1-59)式所确定的 h 称为标准正态扩散的窗宽.

3.经验窗宽

由(1-1-58)式知,母体概率密度函数 $f(l)$的正态扩散估计 $\hat{f}(l)$不仅与观测值 l_i、子样容量(观测值的个数)n 有关,而且还与标准正态扩散的窗宽 h 有关.当观测完成后,观测值 l_i 和观测值的个数n 都是已知量,此时只有窗宽 h 未知.因此,要根据观测值 l_i 来估计母体概率密度函数$f(l)$,首先要确定标准正态扩散的窗宽 h.在《模糊信息优化处理技术及其应用》[10](黄崇福,王家鼎,1995)中根据择近原则导出的窗宽 h 的经验公式为

$$h=\frac{\alpha(b-a)}{n-1} \tag{1-1-60}$$

式中:$a=\min(l_i)$, $b=\max(l_i)$, $i=1,2,\cdots,n$;α 是n 的函数,α 与n 的关系[8](王新洲,1999)列于表 1-1-1.

表 1-1-1

n	3	4	5	6	7
α	0.849321800	1.273982782	1.698643675	1.336252561	1.445461208
n	8	9	10	11	12
α	1.395189816	1.422962345	1.416278786	1.420835443	1.420269570

续表

n	13	14	15	16	17
α	1.420698795	1.420669671	1.420693321	1.420692226	1.420693101
n	18	19	20	21	22
α	1.420693101	1.420693101	1.420693101	1.420693101	1.420693101

由表1-1-1知,当 $n \geqslant 17$ 时,$\alpha \equiv 1.420693101$.

4.最优窗宽

经验窗宽具有计算简单的优点.但研究表明,经验窗宽不适用于所有分布[11](游扬声,2001).而且采用不同的窗宽 h,母体概率密度函数 $f(l)$ 的估计精度是不同的.因此,需要在某种最优准则下求得最优窗宽.我们在母体概率密度函数估计的均方误差最小这一准则下导出了最优窗宽的迭代公式[12](王新洲,游扬声,2001):

$$\begin{cases} h^0 = b - \dfrac{a}{n-1} \\ h_i^{k+1} = \dfrac{(\hat{f}^k(l_i)(h^k)^4}{2n\sqrt{\pi}(\hat{f}^k(l_i+h^k) - 2\hat{f}^k(l_i) + \hat{f}^k(l_i - h^k)^2}, \quad h_i^{k+1} \leqslant \dfrac{b-a}{3} \\ h^{k+1} = \dfrac{1}{n}\sum\limits_i h_i^{k+1} \\ \hat{f}^{k+1}(l) = \dfrac{1}{\sqrt{2\pi}nh^k}\sum\limits_i \exp\dfrac{-(l-l_i)^2}{2(h^k)^2} \end{cases} \tag{1-1-61}$$

为了考察最优窗宽的效果,我们模拟了100个服从标准正态的观测值,分别采用经验窗宽(1-1-60)式和最优窗宽(1-1-61)式,对母体概率密度函数 $f(l)$ 进行了估计;同时还模拟了40个服从卡方(chi-squared)分布的观测值,分别采用经验窗宽(1-1-60)式和最优窗宽(1-1-61)式,对母体概率密度函数 $f(l)$ 进行估计.估计结果见图1-1-1和图1-1-2所示.

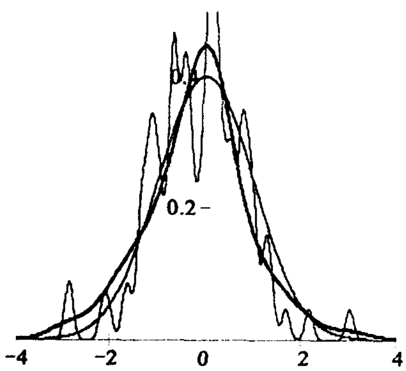

图1-1-1　粗实线为最优窗宽估计结果;细实线为标准正态分布曲线;虚线为一般窗宽估计结果.

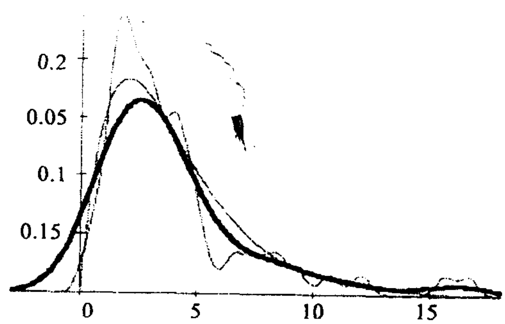

图1-1-2　粗实线为最优窗宽估计结果;细实线为chi-squared分布曲线;虚线为一般窗宽估计结果.

由图1-1-1和图1-1-2可以看出,采用最优窗宽估计母体概率密度函数,其精度比采用