●研究生教材 ●研究生教材

（第二版）

线性模型参数估计及其改进

张金槐 编著

■研究生教材■

线性模型参数估计及其改进

张金槐　编著

国防科技大学出版社

湖南·长沙

图书在版编目(CIP)数据

EA01/21

线性模型参数估计及其改进/张金槐编著. —长沙:国防科技大学出版社,1999.3

ISBN 7-81024-526-0

Ⅰ.线性模型参数估计及其改进

Ⅱ.张金槐

Ⅲ.①线性模型②参数估计及其改进

Ⅳ.O211

国防科技大学出版社出版发行

电话:(0731)4555681　　邮政编码:410073

E-mail:gfkdcbs@public.cs.hn.cn

责任编辑:文　慧　　责任校对:黄　煌

新华书店总店北京发行所经销

长沙交通学院印刷厂印装

*

850×1168　1/32　印张:13.125 字数:329千

1999年3月第2版第1次印刷　印数:1—1000册

*

定价:17.00元

内 容 提 要

本书论述线性模型参数估计及其改进方法。主要讨论了一般最小二乘方理论、Bayes 估计、压缩估计、主成份估计，同时讨论了线性时变系统状态向量的估计方法及自适应估计技术，论述了运用正交多项式的轨道拟合理论和信息融合估计技术。考虑到工程技术中的实用性，列举了常被应用的示例。

本书可作为有关专业的研究生教材，也可供从事飞行器试验结果分析的工程技术人员参考。

前　言

线性模型参数估计是统计数学的一个分支。它在国防科技领域占有重要的地位。从20世纪60年代以来，关于飞行器的惯性制导误差分析、动力装置系统性能参数的分析、飞行器运动参数的精度分析以及与之相关的导航、定轨、预报等问题，广泛地应用了线性模型参数估计的方法。当然，在工业技术、农业生产、经济规划等领域，它同样也有重要的应用。在工程实践中，涌现出不少应用中的理论问题，例如线性模型中的病态问题，需要我们予以充分注意。因此，改进估计的方法，使之能更好地应用于实际，这是目前要解决的重要问题。

为了适应国防工业部门及部队研究单位的需要，我们从80年代以来，在研究生中开设了《线性模型参数估计及其改进》课程，受到了多方面的关注。为此，我们将有关内容整理，并补充了科研中的一些成果，形成了这本书。

本书可作为有关专业研究生教材，也可作为工程技术人员的参考书。

本书在完稿过程中，金治明教授审阅了全部内容，提出了很好的修改意见。特此致谢。

作者　　1991年

修订版前言

本书出版后，作为自动控制理论及应用专业的研究生教材使用了数年，并受到国防科学技术部门的重视。在飞行器试验的结果分析、轨道确定、预报、系统辨识等方面得到了广泛的应用。一些专家建议，为了更好地适应当前试验分析、精度评估、自适应滤波等应用的需要，将内容适当更新，并增添近期的研究成果出修订版本，为此，出版了这个修订本。在新版中，基本上保持原书的结构。在内容上，新添了运用正交多项式的轨道拟合理论，加强了自适应滤波的论述，并增加了信息融合估计技术。

作者　　1998 年 9 月

●目　　录●

1 导 论

本书内容分两部分:线性模型参数估计以及参数估计的改进。由于它们互相联系,因此论述中互相渗透。关于参数估计,目前已有不少著作,特别是从应用角度编写的书籍就更多,这些都是十分需要的。在参数估计方法中,如最小二乘方估计、线性无偏最小方差估计、模型参数的检验等内容,都见诸一般书籍。为了照顾内容的系统及后续内容论述上的方便,本书只作一般性讨论。不过,本书在参数估计方面也有它的特点。例如最小二乘方理论,它并不是在所有的场合都能得到满意的解答。我们将讨论线性模型$(\boldsymbol{Y},\boldsymbol{X}\beta,\sigma^2\boldsymbol{I})$未知参数向量$\beta$的估计,其中矩阵$\boldsymbol{X}$不一定具有满列秩。此外,如果$\boldsymbol{X}$的列向量间存在近似的线性相关,则最小二乘方估计将产生较大的方差,致使估值是不可信的。这是人们关心的问题。本书将充分注意到这种情况,并提出判断这种复共线性对估计的严重程度和应采取的处置方法。我们还将讨论β的线性函数$\boldsymbol{C}^{\mathrm{T}}\beta$的无偏估计问题,这就是$\beta$的可估性问题。这里,Gauss-Markov 定理是重要的。它指出:在线性模型$(\boldsymbol{Y},\boldsymbol{X}\beta,\sigma^2\boldsymbol{I})$中,如果$\boldsymbol{C}^{\mathrm{T}}\beta$的无偏估计存在,则$\boldsymbol{C}^{\mathrm{T}}\hat{\beta}_{\mathrm{LS}}$为唯一的线性无偏最小方差估计。此处$\hat{\beta}_{\mathrm{LS}}$是$\beta$的最小二乘估计。由此可知,可估性问题是重要的,在本书中将花一定篇幅论述这个问题。本书还将引入 Bayes 估计,它是与经典的估值观点不同的一种估计。它将未知参数看成随机的,且具有验前分布。因此,增加了被估参数的信息,它对于提高估计的性能是有益的。它与参数估计的改进有密切的联系。目前,这种方法正受到国防科技部门的重视。我们还将讨论变参数线性模型状态向量的估计方法。由于这种方法应用广泛,例如飞行器的轨道跟踪、拦截、

导航、GPS(全球卫星定位系统)中的数据处理等等。另外,在应用中还有不少需要进一步研究的问题。我们注意到这些特点,除论述估值的方法外,还专门讨论了自适应估计技术。

就作者写作的本意来说,重点是线性模型参数估计的改进。这是由于近几年来的科研工作的实践使我们感到有这种需要。本书将力图反映这方面的最新成果,但由于涉及面广,而且一些方法正在发展之中,还不能说已经很成熟。因此,只能择其中之主要者,特别是我们在科研实践中已应用且证明其效果是比较好的内容。

下面就来谈谈线性模型参数估计改进的有关问题。

考虑线性模型

$$Y = X\beta + e, \tag{1.1}$$

其中,Y 为 $n\times 1$ 观测向量,β 为 $p\times 1$ 向量,X 为 $n\times p$ 矩阵,e 的均值为 0,方差阵为 $\mathrm{Var}(e)=\sigma^2 I$,这里 I 为单位阵(如果 $\mathrm{Var}(e)=\sigma^2 G>0$,那么只要令 $Z=G^{-\frac{1}{2}}Y$,则原来模型将转化为$(Z,U\beta,\sigma^2 I)$,这里 $U\beta=G^{-\frac{1}{2}}X\beta$. 对于模型(1.1),可知(见第 2 章)$\beta$ 的最小二乘方估计为

$$\hat{\beta}_{\mathrm{LS}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y, \tag{1.2}$$

它是 β 的无偏估计,且

$$\mathrm{Var}[\hat{\beta}_{\mathrm{LS}}] = \sigma^2(X^{\mathrm{T}}X)^{-1}.$$

我们指出,作为 Y 的预报,当给定 X 时,预报值 $\hat{Y}=X\hat{\beta}_{\mathrm{LS}}$,它使残差平方和 $\mathrm{SSE}\triangleq\|Y-X\hat{\beta}_{\mathrm{LS}}\|^2=\min$,但是,作为 β 估计的 $\hat{\beta}_{\mathrm{LS}}$,它是否总是“好”的?注意下列关系式

$$\begin{aligned}\mathrm{E}[\|\hat{\beta}_{\mathrm{LS}}\|^2] &= \mathrm{E}[\hat{\beta}_{\mathrm{LS}}^{\mathrm{T}}\hat{\beta}_{\mathrm{LS}}] \\ &= \mathrm{tr}[\mathrm{Var}(\hat{\beta}_{\mathrm{LS}})] + \|\beta\|^2.\end{aligned} \tag{1.3}$$

由此可知,$\mathrm{E}[\|\hat{\beta}_{\mathrm{LS}}\|^2]$总是偏离真实参数 β 的长度的平方 $\|\beta\|^2$. 此外,由于 $X^{\mathrm{T}}X\triangleq C$ 为实对称阵,于是存在正交阵 Q,使

$$Q^{\mathrm{T}}CQ = \Lambda = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_p \end{bmatrix},$$

其中，$\lambda_1, \cdots, \lambda_p$ 为 $X^{\mathrm{T}}X$ 的非零特征值。于是

$$\begin{aligned} C &= Q\Lambda Q^{\mathrm{T}}, \\ C^{-1} &= Q\Lambda^{-1}Q^{\mathrm{T}}, \operatorname{tr}(X^{\mathrm{T}}X)^{-1} \\ &= \operatorname{tr}C^{-1} = \operatorname{tr}(Q\Lambda^{-1}Q^{\mathrm{T}}) \\ &= \operatorname{tr}(\Lambda^{-1}QQ) = \operatorname{tr}\Lambda^{-1} \\ &= \sum_{i=1}^{p} \frac{1}{\lambda_i}. \end{aligned}$$

这样，

$$\mathrm{E}[\, \| \hat{\beta}_{\mathrm{LS}} \|^2] = \sigma^2 \sum_{i=1}^{P} \frac{1}{\lambda_i} + \| \beta \|^2. \tag{1.4}$$

因此，如果 $\mathrm{rk}(X^{\mathrm{T}}X) < p$，则$\min\limits_i \lambda_i = 0$，此时通常的 LS 估计不存在。如果 X 的列间具有近似的线性相关关系，则$\min\limits_i \lambda_i$ 将甚小，于是 $\mathrm{E}[\, \| \hat{\beta}_{\mathrm{LS}} \|^2] \gg \| \beta \|^2$，因此 LS 估计是不可信赖的。这种情况就是人们常说的 X 具有近似复共线性关系。或者说，矩阵 $X^{\mathrm{T}}X$ 具有病态性。

为了克服这种估值“长度”过大的弊病，一个最直接的方法是引入关于 β 的约束。例如 $\| \beta \| \leqslant C$. 这样，控制了 $\hat{\beta}_{\mathrm{LS}}$的长度不会无限制地扩大。另外，如果引入关于 β 的验前信息，它将有利于改善估计的性质，这就是所谓 Bayes 估计方法。

现在再来看看当 $\mathrm{rk}X = r < p$，或者 X 的列具有近似线性相关时，如何处置 β 的估计问题。将 X 分解为两部分：

$$\underset{n\times p}{X} = [\underset{n\times r}{X_1} \;\vdots\; \underset{r\times(p-r)}{X_2}],$$

不妨假定 X_1 是 X 的前 r 列，且它们是 X 的线性无关的列，X_2 的列与 X_1 的列线性相关。于是 X_2 总可以表示为

$$X_2 = \underset{(n\times r)}{X_1} \ \underset{r\times(p-r)}{F},$$

其中 F 是某矩阵。这样

$$X = \underset{n\times r}{X_1}[\underset{r\times r}{I_r} \vdots \underset{r\times(p-r)}{F}] \triangleq \underset{n\times r}{X_1} \ \underset{r\times p}{G},$$

其中 $G=[I_r \vdots F]$，且 $\mathrm{rk}G=r$，$\mathrm{rk}X_1=r$，而

$$X\beta = X_1 G\beta \triangleq X_1\alpha,$$

此处 $\alpha=\underset{r\times p}{G} \ \underset{p\times 1}{\beta}$ 为 r 维向量，它是未知的。这样，原模型 $Y=X\beta+e=X_1\alpha+e$，而 α 的 LS 估计为

$$\hat{\alpha} = (X_1^{\mathrm{T}}X_1)^{-1}X_1^{\mathrm{T}}Y.$$

由于 $\mathrm{rk}(X^{\mathrm{T}}X)=r$，$\hat{\alpha}$ 是可以计算的。因此，将原来 LS 估计不存在的问题归化为可估计问题（估计 α）。这种方法最早由 Bock(1963)提出，称之为重新参数化方法。Johnson(1971)讨论了这种方法的一些优良性质。这里注意 $\alpha=G\beta$，它是 β 的一个线性函数。当 β 原来的无偏估计不存在时，能否找出 β 的分量的一个线性组合，使它的无偏估计存在，且具有某种优良的性质，这就是线性估计中受到人们重视的可估性问题。

前已述及，如果坚守线性无偏最小方差估计，那么可估性分析是重要的，且由著名的 Markov 定理，一个可估函数 $G\beta$ 具有线性无偏最小方差估计。此外，如果我们的估计不必限于无偏类，且估计的优良性指标也不必限于 MV，而我们关心的问题是估计量和真值之间的距离（在统计决策理论中，它可以用平方损失函数来定义）。这样，线性模型参数估计不必限于 LS 估计，也不必限于最小方差估计。这就是 70 年代发展起来的线性估计的改进问题。

在估计的改进问题讨论中，容许性估计历来是理论界和工程实践中的重要问题。这方面的研究参见 Vinod 和 Ullah 等人的工作[5],[6]。本书将给予足够的注意。

必须指出，估计的好坏与多种因素有关。例如线性模型的建立、模型中 β 分量的选择、随机误差 e 的统计特性以及观测数据的

质量等都直接影响估值的性能。因此模型的检测、回归因子的选取、e 的统计特性的鉴定或估计、数据质量分析等都是重要的。

在论述中,本书仍从一般的最小二乘方理论开始,这就是第 2 章的主要内容。这一章讨论 Gauss-Markov 的二乘方理论以及具有线性约束之下的参数估计方法,同时讨论可估性问题。第 3 章讨论线性模型参数估计的 Bayes 方法,给出了当具有先验信息之下的参数估计方法。目的在于利用历次试验的信息,改进估计的性能。第 4、第 5 两章论述 James-Stein 估计及岭估计方法,这些方法直接改进了最小二乘方估计的性能。这方面的内容极其丰富,文献中有很多讨论,我们只能择其中之常用者作较深入的讨论。第 6 章为主成份估计方法。其基本思想是:当人们在形成线性模型时,总是想尽量多考虑一些因素(自变量),但过多的因素,使它们之间产生相关关系。因此,如何选择主要因素而又不致于使信息过多地损失,且能较好地估计出未知参数。这一章给出了估计方法及其主要性质,同时讨论了变量的选择问题。第 7 章讨论 Webster-Mason 的特征值估计方法。从某种意义上说,它是主成份估计方法的改进。第 8 章讨论估计的容许性。从估计的容许性考虑估计的优良性问题,使我们的视野扩大,不致囿于某种特定的方法。当然,具体地找出容许估计并不是一件容易的事。为了适应国防科技领域的需要,第 9 章增设了线性时变系统状态的自适应估计方法,它适用于飞行器的定轨、预报及控制、导航等领域。这一章讨论了自适应估计的最常用方法。第 10 章给出了若干线性模型未知参数估计的实例。由于应用的多样性,讨论只能说是挂一漏万的。这一章给出了国防工业技术部门广为关注的一些个别的应用例子。

由于在论述中运用了较多的向量代数和矩阵的理论和方法,特编写了一个附录,作为运用这方面知识的参考。

2 线性模型参数估计的一般最小二乘方理论

关于线性模型参数的最小二乘方估计方法，在数理统计教科书以及回归分析的书籍中均有讨论。我们这里不作系统论述。只是讨论常被运用的和在后续内容中常涉及的有关内容。有些问题，我们指出参考文献以备查考。

2.1 线性模型及未知参数的最小二乘方估计

假设 Y 是一个随机变量，它与变量 $X_1,\cdots,X_P$ 之间有如下关系

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_pX_p+e, \tag{2.1.1}$$

其中，$\beta_0,\cdots,\beta_p$ 为未知参数，e 是随机误差。一般地说，$X_1,\cdots,X_p$ 的取值有两种情况需加以区别：

1） $X_1,\cdots,X_p$ 完全可以预先确定。即是说，它们是可控制的量；

2） $X_1,\cdots,X_p$ 由观测或其他手段获得，在获取过程中具有随机误差，因此，它们的取值是随机的，应将它们看作随机变量。

我们的问题是当 $X_1,\cdots,X_p$ 给定时，确定出 $\beta_0,\cdots,\beta_p$ 的估计和确定 e 的方差特性。在讨论中，总是假定 $\mathrm{E}[e]=0$. 这样，在 1)之下，

$$\mathrm{E}[Y]=\beta_0+\beta_1X+\cdots+\beta_pX_p, \tag{2.1.2}$$

在 2)之下，

$$\mathrm{E}[Y|X_1,\cdots,X_p]=\beta_0+\beta_1X_1+\cdots+\beta_pX_p. \tag{2.1.3}$$

当 $X_1,\cdots,X_p$ 和 Y 均为随机变量，关系式(2.1.3)称为线性回归。在工程应用中，当 $X_i(i=1,\cdots,p)$为非随机时也称为回归。在下面的讨论中，主要讨论 $X_1,\cdots,X_p$ 为非随机的情形。习惯上，常称 X_i $(i=1,\cdots,p)$为自变量，Y 为因变量。按统计术语，称 X_i 为回归因子，Y 为响应变量。

以后，总是假定进行了 n 次观测，而在第 i 次观测中，$X_1,\cdots,X_p$ 和 Y 分别取值 $X_{1i},\cdots,X_{pi}$和 Y_i，于是

$$Y_i=\beta_0+\beta_1X_{1i}+\cdots+\beta_pX_{pi}+e_i,\quad i=1,\cdots,n.$$

且假定

$$\mathrm{E}[e_i]=0,\quad \forall\ i,$$
$$\mathrm{Cov}(e_i,e_j)=\sigma^2\delta_{i,j}.$$

这里 $\delta_{i,j}$为 Kronecker δ 函数。上述模型用矩阵和向量形式表示时，记

$$\boldsymbol{Y}=\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{bmatrix},\quad \mathbf{1}=\begin{bmatrix}1\\ 1\\ \vdots\\ 1\end{bmatrix},$$

$$\boldsymbol{X}=\begin{bmatrix}X_{11} & X_{21}\cdots X_{P1}\\ X_{12} & X_{22}\cdots X_{p2}\\ \vdots & \\ X_{1n} & X_{2n}\cdots X_{pn}\end{bmatrix},\quad \boldsymbol{e}=\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix},$$

则

$$\boldsymbol{Y}=\mathbf{1}\beta_0+\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e}. \tag{2.1.4}$$

其中

$$\mathrm{E}[\boldsymbol{e}]=0,$$
$$\mathrm{Var}[\boldsymbol{e}]=\mathrm{E}[\boldsymbol{e}\cdot\boldsymbol{e}^{\mathrm{T}}]=\sigma^2\boldsymbol{I}_n.$$

把 $\mathbf{1}\beta_0$ 从模型中分离出来的这种写法是为了应用的需要。在理论

论述中并没有这种必要，而是将线性模型写成

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{2.1.4'}$$

常记此模型为$(\boldsymbol{Y}, X\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$。

线性模型在工程实践中屡见不鲜。下面的例子可见其一斑。

[**例 2.1.1**] 设

$$\mathrm{E}[Y(t)] = \beta_0 + \beta_1 t + \cdots + \beta_m t^m,$$

其中，$\beta_0, \beta_1, \cdots, \beta_m$ 是未知参数，而在 $t=t_i(i=1,\cdots,t_n)$所获得的观测量为

$$Y_i = Y(t_i) = \beta_0 + \beta_1 t_i + \cdots + \beta_m t_i^m + \varepsilon(t_i).$$

记

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_m]^{\mathrm{T}},$$

$$\boldsymbol{e} = [\varepsilon(t_1), \cdots, \varepsilon(t_n)]^{\mathrm{T}},$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & t_1 & t_1^2 \cdots t_1^m \\ 1 & t_2 & t_2^2 \cdots t_2^m \\ \vdots & & \\ 1 & t_n & t_n^2 \cdots t_n^m \end{bmatrix},$$

则

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e},$$

这是熟知的多项式曲线拟合模型。它在飞行器试验结果的轨道特性分析中是常被应用的。

[**例 2.1.2**] 设有线性离散系统

$$\boldsymbol{X}(K+1) = \boldsymbol{\Phi}(K+1, K)\boldsymbol{X}(K),$$

其中，$\boldsymbol{X}(K)$为 t_k 时刻系统所处状态向量，$\boldsymbol{\Phi}(K+1,K)$为状态转移矩阵。设观测量与状态之间的关系为

$$\boldsymbol{Z}(K) = \boldsymbol{H}(K)\boldsymbol{X}(K) + \boldsymbol{V}(K), \quad K = 1, \cdots, n.$$

于是由 $\boldsymbol{X}(K) = \boldsymbol{\Phi}(K, 0)\boldsymbol{X}(0)$，可将关于初始状态 $X(0)$的估计问题归结为如下线性模型的参数估计问题：

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e},$$

其中 $\boldsymbol{\beta} = \boldsymbol{X}(0)$， $\boldsymbol{e} = [\boldsymbol{V}_1, \cdots, \boldsymbol{V}_n]^{\mathrm{T}}$， $\boldsymbol{Y} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n]^{\mathrm{T}}$，