摘要

线性模型是一类很重要的数学模型，它在经济、国防等许多领域都有着广泛而重要的应用。线性模型理论的研究已成为国际统计界的研究重点之一。本文基于一般的线性模型，对其中的参数估计问题进行了研究。

我们对线性模型的参数估计理论进行了比较系统的研究，主要结论有：

1 在 LSE 估计的稳健性方面，我们用不同于 Zyskind 的方法，得到了一些新结论。首先，我们对设计矩阵列满秩的情形，给出了 GLSE 和 LSE 相等的两个充要条件和一个充分条件。揭示了 GLSE 和 LSE 相等与它们方差相等之间的关系。对于设计矩阵一般情形，我们给出了三个充分条件，较好地解决了 LSE 估计稳健性方面的问题。

2 我们讨论了 M-估计的渐近正态性，改进了这方面的一些结论。目前的结论中，关于误差的分布函数假定为在 0 点处有正的导数，我们将该条件推广到一个更一般的条件，证明了 M-估计的渐近正态性。

3 我们研究了不可估函数的线性估计的可容许性，首先我们对于一类特殊的设计矩阵和误差，给出了可容许性的几个充要条件。然后，我们证明了一般情形均可化为该特殊情形进行分析，得到不可估函数线性估计的可容许性的充要条件。基本上解决了这方面的问题。

4 我们利用可估函数估计的可容许性的性质以及一些矩阵不等式，给出了可估函数全体可容许估计的显式表达式。

5 在一个很一般的条件下，证明了回归系数相合估计的存在性。

关键词： 线性模型，参数估计，最小二乘估计，M-估计，相合性，容许估计

Abstract

Linear model is a very important mathematical model. It applies to many fields such as economy and national defence. The study of linear model has become focus in international statistics group. Based on the normal linear model, we study parameter estimate.

We give a systematic account on the theory of parameter estimate. The main results are as follows:

1 On the robustness of LSE, using the methods other than Zyskind's, we gained some new results. First, we discuss the case that the design matrix is row full order, get two s.n conditions and one sufficient condition in which GLSE is equivalent to LSE. We point out the meaning that the variance of GLSE is the same as the variance of LSE. In common case, presenting three sufficient conditions, we solve the problem nicely.

2 We discuss the asymptotic theory of M-estimate, and improve some conclusions. In prior conclusion, It is supposed that the distributive function of error has positive derivative at zero. Generalizing the hypothesis, we study the asymptotic theory of M-estimate.

3 we study the admissible estimate of the function which can not be estimated. First, for a kind of special design matrix and error, we bring forward some s.n condition on admissible estimate. Then, we prove that any case can convert into the special case.

4 Using some matrix inequalities, we get the expression of admissible estimate.

5 We prove the existence of regression coefficient consistent estimate in the generalizing condition.

Keywords: linear model, parameter estimate, LSE, M-estimate, consistent, admissible estimate

第一章　　预备知识

§ 1.1 线性模型简介

线性模型是一类很重要的统计模型。它包括通常的线性回归模型、方差分析模型、协方差分析模型和方差分量模型等，这些统计模型在国民经济的发展中都有着广泛而重要的应用；另一方面,线性模型的基本理论与方法也为其它统计模型的理论与方法提供了基本的工具。

现在我们给出线性模型一般的数学定义,从某种意义上讲,它能把各种形式的线性模型都包罗在内。

假设我们对变量 Y 作 n 次观测,得到 $y_1, y_2, \ldots, y_n$,它们可以表为如下线性组合形式:

$$
\begin{aligned}
y_1 &= x_{11}\beta_1 + x_{12}\beta_2 + \cdots + x_{1p}\beta_p + e_1 \ , \\
y_2 &= x_{21}\beta_1 + x_{22}\beta_2 + \cdots + x_{2p}\beta_p + e_2 \ , \\
&\cdots \qquad \cdots \qquad \cdots \\
y_n &= x_{n1}\beta_1 + x_{n2}\beta_2 + \cdots + x_{np}\beta_p + e_n \ ,
\end{aligned}
\tag{1.1}
$$

此处 x_{ij} (i=1, 2, . . . , n; j=1, 2, . . . , p)均为已知常数,对于不同类型的特殊模型, x_{ij} 的取值具有不同的特点。例如,在线性回归模型中, x_{ij} 可以是随机变量的一组给定值,它既可取一些离散值,也可取连续值。而在方差分析模型中,它只能取 0, 1 两个值。$\beta_1, \beta_2, \ldots, \beta_p$ 为模型的非随机的未知参数。$e_i (i=1,2,\ldots,n)$ 为模型的随机误差，也称为模型误差，它反映了 Y 值构成中由大量偶然性因素的影响所形成的那一部分，或更确切地说，没被诸 X 因素所反映的那部分。要降低 $e_i (i=1,2,\ldots,n)$ 的影响，就需要尽量找出对 Y 有影响的各种系统性因

素，把它从$e_i(i=1,2,...,n)$中分离出来。但这样做，受科学水平、实验条件以及人、财、物等各种条件的制约，而且也不见得一定有利。$e_i(i=1,2,...,n)$满足$Ee_i=0(i=1,2,...,n)$，记$Cov(e_i,e_j)=\sigma_{ij}$,对一切 i，j.又记

$$y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

$$\beta=\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \qquad e=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}, \Sigma=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

则(1.1)变为$y=X\beta+e$,其中e满足$Ee=0,Cove=\Sigma$ (1.2)

定义 1 称模型$y=X\beta+e$，$Ee=0,Cove=\Sigma$，为线性统计模型，简称线性模型，其中$y_{n\times 1}$为观测向量，$X_{n\times p}$称为设计矩阵，$\beta_{p\times 1}$称为未知参数向量，e称为随机误差向量。

对于模型(1.2)，通常假定$\Sigma=\sigma^2 I_n$或$\Sigma=\sigma^2 V$或其它适当假设. 其中$\sigma^2>0$未知，$V\geq 0$已知。更一般地，我们可将 X 理解成一个线性变换(线性算子)，而将Σ理解成对称，非负定映射(称为协差算子)。换言之，我们有如下更一般的

定义 2 设（Ω, F, P），p∈P，是一概率空间族，y 是$\Omega\to$H（一个向量空间）的一个可测变换（随机元）。若存在线性子空间 L⊆H 及 H→H 的对称非负定映射的集合 D，使得

i＞ $E_p y\in L$, $\forall p\in P$;

ii＞ Span{ $E_p y$: p∈D}=L;

iii＞ $Cov_p(y)$ ∈D, $\forall p\in P$.

则称 y 服从线性模型 M(L, D)。

注：设（Ω, F, P）是一概率空间，H 是带有内积<· ，·>的一个向量空间，A 是 H 上的一个σ-代数，则称 Ω→H 的一个可测变换 y（即 $y^{-1}(A)\subseteq F$）为 H 值随机元（r. e.）。

若存在 μ ∈H, 使得∀A∈H，有

E〈A，y〉=〈A，μ〉

则称 μ 为随机元 y 的数学期望，记作 Ey=μ。

若存在 H→H 的对称、非负定映射Σ，使得∀A、B∈H，有 Cov(<A, y>, <A, y>)=<A, ΣB>，则称Σ为随机元 y 的协方差算子，记作 Cov(y)=Σ或 Dy=Σ. 其中 H→H 的映射Σ是对称的，记作Σ=Σ′，意指：<A, ΣB>=<ΣA，B>，∀A、B∈H。

而 H→H 的对称映射Σ是非负定的，记作Σ≥0，意指：<A，ΣA>≥0，∀A∈H。

线性模型在工程实践中是屡见不鲜的。

例 1 设 $Y(t)=\beta_0+\beta_1 t+...+\beta_m t^m+e(t)$，其中$\beta_0$，$\beta_1$，...，$\beta_m$是未知参数，在 $t=t_i(i=1,2,...,n)$所获得的观测量为

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

记$\beta=(\beta_0,\beta_1,...,\beta_m)^T$, $e=(e(t_1),...,e(t_n))^T$,

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

则 $Y=X\beta+e$

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常被用到。

例 2 考虑线性系统输入滑动和

$$y(k)=\sum_{i=1}^{N} w_i x(k-i)$$

其中 x(k-1),...,x(k-N)为线性系统的输入系列, $w_i(i=1,...,N)$ 为权系数。输出信号 y(k)的测量值为

$Z(j)=y(j)+e(j)$, $j=k-L,k-L+1,...,k$

由观测量 z(k-L),...,Z(k)去估计权系数 $w_1,...,w_N$ 的线性模型可以写为 $Z=X\beta+e$

其中 $Z=(z(k-L),...,Z(k))^T$,

$$X=\begin{pmatrix} x(k-L-1) & x(k-L-2) & \cdots & x(k-L-N) \\ \cdots & & \cdots & \\ x(k-2) & x(k-3) & \cdots & x(k-N-1) \\ x(k-1) & x(k-2) & \cdots & x(k-N) \end{pmatrix},$$

$$\beta=(w_1,...,w_N)^T, \quad e=(e(k-L),...,e(k))^T$$

这是熟知的线性平滑问题,它在数据处理中应用比较广泛.

§1.2 矩阵知识

矩阵知识在线性模型中有重要的应用，为使行文的完整，我们将一些后文所需的结论列于此。首先提出本文中常用的一些记号。m 行 n 列的矩阵 A 常称为 m×n 矩阵 A,或 $A_{m\times n}$。当 m=n 时称为 n 阶方阵,方阵 A 的行列式记为|A|。矩阵 A 的转置记为 A' 或 A^T。一列矩阵称为列向量,而一行矩阵称为行向量.我们以不加“′”或“T”的向量表示列向量,例如 a 为一列向量,而 a' 为行向量。向量 $a=(a_1,...,a_n)'$ 的

长为$\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}$，记为‖a‖。矩阵A的秩记为rk(A)或rank(A)。若A为方阵，则其迹(trace)，即主对角线元素之和，记为 tr(A)。本文中，若非特别指出，一般都考虑实矩阵。

若A为一线性子空间，则其正交补空间记为$A^{\perp}$。设矩阵$A=(a_1\vdots\cdots\vdots a_n)$，则由A的列向量$a_1,...,a_n$生成的线性子空间记为$\mu(A)$，其正交补空间记为$\mu^{\perp}(A)$。显然$\mu(A)=\{A\alpha \mid \alpha\in R^n\}$。

正定矩阵A记为A>0，半正定矩阵A记为A≥0。本文中，正定矩阵与半正定矩阵皆指对称矩阵。若 A-B≥0(A-B>0)，则记为 A≥B(A>B)。n阶单位矩阵记为I_n。

§1.2.1 广义逆矩阵

定义1 对于矩阵$A_{m\times n}$，一切满足方程组AXA=A的矩阵X称为矩阵A的广义逆，记为A^-。

定理1 对于任意给定的矩阵$A_{m\times n}$，A^-必存在。

证明：若$A_{m\times n}=0$，则任意n×m阶矩阵都是A的广义逆。当A≠0时，设rk(A)=r，则必存在非奇异矩阵P与Q使得

$$A=P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q$$

设X为A的广义逆，则有

$$AXA=A\Leftrightarrow P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}QXP\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q=P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q$$

$$\Leftrightarrow \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} QXP \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

若记 $QXP = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$ ，B_{11} 为 r 阶方阵，则上式

$$\Leftrightarrow \begin{pmatrix} B_{11} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Leftrightarrow B_{11} = I_r$$

于是 $AXA = A \Leftrightarrow X = Q^{-1} \begin{pmatrix} I_r & B_{12} \\ B_{21} & B_{22} \end{pmatrix} P^{-1}$

其中 B_{12}, B_{21}, B_{22} 任意。

推论 1.1（1）对于任意给定的矩阵 $A_{m\times n}$，若 A 对称，则在所有 A^- 中至少有一个是对称的。若 A≥0，则至少有一个 A^-≥0；

（2）A^-唯一 $\Leftrightarrow$ A 为可逆矩阵，此时 $A^- = A^{-1}$；

（3）A^-A 与 A A^-均为幂等矩阵，且

$$rk(A^-) \geq rk(A) = rk(A^-A) = rk(AA^-) \ ;$$

（4）若 $\mu(B) \subset \mu(A)$，$\mu(C) \subset \mu(A')$，则 $C'A^-B$ 与 A^- 的选择无关。

上述事实不难从定理 1 以及广义逆的定义得到。

定理 2 A^- 有下列基本性质：

（1）$(A^-)'$ 为 A' 的一个广义逆，特别地，若 $A = A'$，则 $(A^-)'$ 为 A 的一个广义逆。

（2）若 P, Q 均为可逆方阵，则 $(PBQ)^- = Q^{-1}B^-P^{-1}$，从而有

$$B^- = Q(PBQ)^- P$$

证明：（1）显然。只证（2）

$$PBQ(PBQ)^- PBQ = PBQ \Leftrightarrow BQ(PBQ)^- PB = B$$

$$\Leftrightarrow Q(PBQ)^- P = B^-$$

$$\Leftrightarrow (PBQ)^- = Q^{-1}B^-P^{-1}$$

定理 3 $A(A^{'}A)^{-}A^{'}A=A$，　　$A^{'}A(A^{'}A)^{-}A^{'}=A^{'}$

证明：令 $B=A(A^{'}A)^{-}A^{'}A-A$，则

$$\begin{aligned}B^{'}B&=\{A[(A^{'}A)^{-}A^{'}A-I]\}^{'}\{A[(A^{'}A)^{-}A^{'}A-I]\}\\&=[(A^{'}A)^{-}A^{'}A-I]^{'}A^{'}A[(A^{'}A)^{-}A^{'}A-I]\\&=[(A^{'}A)^{-}A^{'}A-I]^{'}[A^{'}A-A^{'}A]\\&=0\end{aligned}$$

所以 $B=0$，即 $A(A^{'}A)^{-}A^{'}A=A$。同理可证第二式。

定理 4 若记 $P_A=A(A^{'}A)^{-}A^{'}$，则 P_A 与 $(A^{'}A)^{-}$ 的选取无关且 P_A 对称、幂等。

证明：任取 $A^{'}A$ 的两个广义逆 $(A^{'}A)_1^{-}$ 与 $(A^{'}A)_2^{-}$，令

$B=A(A^{'}A)_1^{-}A^{'}-A(A^{'}A)_2^{-}A^{'}$，　　则

$$\begin{aligned}B^{'}B&=\{A[(A^{'}A)_1^{-}-(A^{'}A)_2^{-}]A^{'}\}^{'}\{A[(A^{'}A)_1^{-}-(A^{'}A)_2^{-}]A^{'}\}\\&=A[(A^{'}A)_1^{-}-(A^{'}A)_2^{-}]^{'}A^{'}A[(A^{'}A)_1^{-}-(A^{'}A)_2^{-}]^{'}A^{'}\\&=A[(A^{'}A)_1^{-}-(A^{'}A)_2^{-}]^{'}[A^{'}-A^{'}]\\&=0\end{aligned}$$

所以 $B=0$，即 $A(A^{'}A)_1^{-}A^{'}=A(A^{'}A)_2^{-}A^{'}$，亦即 P_A 与 $(A^{'}A)^{-}$ 的选择无关。

注意到 $A^{'}A$ 对称，故至少有一对称的 $(A^{'}A)^{-}$（见推论 1.1），选此对称的 $(A^{'}A)^{-}$，即知 $P_A=A(A^{'}A)^{-}A^{'}$ 对称，又

$$\begin{aligned}P_A^2&=A(A^{'}A)^{-}A^{'}\cdot A(A^{'}A)^{-}A^{'}\\&=A(A^{'}A)^{-}A^{'}=P_A,\end{aligned}$$

即 P_A 幂等。

注：1. 若 A 幂等，则（ i ）A 的特征值是 0 或 1；

（ii）tr(A)=rk(A).

2. 若 A 对称、幂等，则 A≥0.

下面的定理用矩阵的广义逆给出了相容线性方程组的通解。

定理 5 设 $AX=b$ 为一相容线性方程组，则

（1） 对于任一广义逆A^-，$x=A^-b$必为方程组的解；

（2） 齐次线性方程组$AX=0$的通解为$x=(I-A^-A)z$，此处z为任意的向量，A^-为任意固定的一个广义逆；

(3)线性方程组$AX=b$的通解为$x=A^-b+(I-A^-A)z$，其中z为任意的向量，A^-为任意固定的一个广义逆。

证明：（1）由相容性假设知，存在x_0，使得$Ax_0=b$，故对任一A^-，$A(A^-b)=AA^-Ax_0=Ax_0=b$，即A^-b为解。

（2）设x_0为$AX=0$的任一解，即$Ax_0=0$，那么$x_0=(I-A^-A)x_0+A^-Ax_0=(I-A^-A)x_0$，即任一解都取$(I-A^-A)z$的形式。反过来，对于任意的$z$，因$A(I-A^-A)z=(A-AA^-A)z=0$，故$(I-A^-A)z$必为解。

（3）任意取定一个广义逆A^-，由（1）知$x_1=A^-b$为方程组的一个特解，由（2）知$x_2=(I-A^-A)z$为齐次方程组$AX=0$的通解。依据非齐次线性方程组的解的结构定理知x_1+x_2为$AX=b$的通解。

定理 6 设$AX=b$为相容线性方程组，且$b\neq 0$，那么当A^-取遍A的所有广义逆时，$x=A^-b$构成了该方程组的全部解。

证明：首先由定理 5 知，A^-b必为相容线性方程组$AX=b$的解，故只需再证$AX=b$的任一解必有A^-b的形式。事实上，若x_0是$AX=b$的解，由定理 5 知$x_0=A^-b+(I-A^-A)z_0$，对某个z_0，因为$b\neq 0$，所以存在矩阵u，使得$z_0=ub$（例如可取$u=z_0(b'b)^{-1}b'$），故

$$x_0=A^-b+(I-A^-A)ub=[A^-+(I-A^-A)u]b\hat{=}Gb$$

易验证$G=A^-+(I-A^-A)u$为A的一个广义逆。

§1.2.2 矩阵不等式

下面的矩阵不等式在本文后面有重要的应用，

定理 1 设 $D=diag(d_1,\cdots,d_p)$ 和 $G=diag(g_1,\cdots,g_p)$ 皆为 p 阶对角阵，且 $0\le D\le I_n$，若存在正交矩阵 Q，使得

$tr(I-G)^2\le trD^2$，

$QG^2Q^{'}\le(I-D)^2$，

则此二式必为等式。

证明：由于 $0\le D\le I_n$，则 $0\le d_i\le 1,\ i=1,\dots,p$。

记 $Q^{'}=(q_1,\cdots,q_p)$，$f_i^2=q_i^{'}G^2q_i$, i=1,...,p，则 f_i^2 为 $QG^2Q^{'}$ 的第 i 个对角元。由 $QG^2Q^{'}\le(I-D)^2$ 知，$f_i^2\le(1-d_i)^2$，i=1，. . . , p

等价地，$d_i\le 1-f_i$，i=1，. . . , p

由 Cauchy 不等式以及 Q 的正交性，得到

$(u_i^{'}Gu_i)^2\le u_i^{'}u_i\cdot u_i^{'}G^2u_i=f_i^2$，i=1，. . . , p

于是 $f_i\ge u_i^{'}Gu_i$，i=1，. . . , p

故 $\sum_{i=1}^{p}f_i\ge\sum_{i=1}^{p}u_i^{'}Gu_i=tr(QGQ^{'})=tr(G)=\sum_{i=1}^{p}g_i$

从而 $tr(D^2)=\sum_{i=1}^{p}d_i^2\le\sum_{i=1}^{p}(1-f_i)^2=p-2\sum_{i=1}^{p}f_i+\sum_{i=1}^{p}f_i^2$

$$=p-2\sum_{i=1}^{p}f_i+tr(QG^2Q^{'})$$

$$=p-2\sum_{i=1}^{p}f_i+tr(G^2)$$

$$=p-2\sum_{i=1}^{p}f_i+\sum_{i=1}^{p}g_i^2$$

$$=\sum_{i=1}^{p}(1-g_i)^2+2(\sum_{i=1}^{p}g_i-\sum_{i=1}^{p}f_i)$$

$$\leq\sum_{i=1}^{p}(1-g_i)^2$$

$$=tr(I-G)^2$$

于是，第一式成为等式。下证第二式也为等式，若 $0\leq(I-D)^2-QG^2Q'$，但是不为 0，于是它的对角元 $(1-d_i)^2-f_i^2$，i=1,...,p 中至少有一个为正，亦即至少对某个 i 有 $d_i<1-f_i$，这和第一式为等式相矛盾。故第二式也为等式。

定理 2 若 A 为非对称的方阵，则存在正交矩阵 P，使得 $tr(PA)>tr(A)$。

证明：对 A 的阶数进行数学归纳法。

当 n=2 时，设 $A=\begin{pmatrix}a & c\\ d & b\end{pmatrix}$，不妨假定 $c>d$。取正交矩阵

$P(\varepsilon)=\begin{pmatrix}1-\varepsilon & -t\\ t & 1-\varepsilon\end{pmatrix}$，此处 $\varepsilon>0$ 充分小，且 $t=\sqrt{2\varepsilon-\varepsilon^2}$。此时，我们有 $tr(P(\varepsilon)A)-tr(A)=-\varepsilon(a+b)+t(c-d)$，由 $t>0$ 和 $c>d$ 知，只要 $\varepsilon>0$ 充分小，就有 $tr(P(\varepsilon)A)-tr(A)>0$。

假设定理结论对一切 n-1 阶非对称的方阵都成立，现设 A 为 n 阶非对称方阵，则 A 至少有一个非对称的主子阵。

若 A 的左上角子阵 A_{n-1} 非对称，此时 A 有形式

$$A=\begin{pmatrix}A_{n-1} & c\\ d & b\end{pmatrix},\qquad A_{n-1}\neq A_{n-1}'$$

根据归纳假设，存在 n-1 阶正交矩阵 P_{n-1}，使得 $tr(P_{n-1}A_{n-1})>tr(A_{n-1})$，

取正交矩阵 $P=\begin{pmatrix} P_{n-1} & 0 \\ 0 & 1 \end{pmatrix}$，则有

$$tr(PA)=tr(P_{n-1}A_{n-1}+b)>tr(A_{n-1})+b=tr(A).$$

若 A 的左上角子阵 A_{n-1} 对称，则由 A 的非对称性知，存在 $a_{kn}\neq a_{nk}$，$1\leq k\leq n-1$，此时将 A 的第 n 行与第 1 行互换，第 n 列与第 1 列互换，亦即用初等矩阵

$$\Delta=\begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

去左、右乘 A，记得到的矩阵为 A^*，$A^*=CAC$。显然 A^* 是非对称的，存在正交矩阵 P^*，使 $tr(P^*A^*)>tr(A^*)$，但 $tr(A^*)=tr(A)$，

$$tr(P^*A^*)=tr(P^*CAC)=tr(CP^*CA)\hat{=}tr(PA),$$

其中 $P\hat{=}CP^*C$ 仍为正交矩阵。从而结论成立。

§1.3 参数估计

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的研究可以追溯到上世纪初，著名数学家 A.M.Legendre 和 C.F.Gauss 先后于 1806 年和 1809 年独立地把最小二乘法应用于观测数据的误差分析。后来，A.A.Markov 于 1900 年证明了最小二乘估计的方差最小性质，即著名的 Gauss-Markov 定理，奠定了最小二乘法在参数估计理论中的地位。R.C.Bose 在 1944 年引入的可估计函数的

概念以及广义逆矩阵的应用，使得设计矩阵为列降秩的线性模型的估计理论表述更加严格而简洁。

线性模型参数估计是统计数学中的一个活跃分支。它在很多领域都有重要的应用。比如在国防科技领域，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。参数估计的重要性质以及如何改进参数估计是目前工程实践中迫切需要解决的问题。

第二章 参数估计的一般最小二乘估计理论

§2.1 最小二乘估计的背景

在实际问题中，有些量$(\theta_1,\ldots,\theta_k)$不能观测，或不容易观测，而另外的一些量$(x_0,x_1,\ldots,x_k)$容易观测，根据其专业理论，它们有线性关系：$x_0+x_1\theta_1+\cdots+x_k\theta_k=0$

对可观测量$(x_0,x_1,\ldots,x_k)$观测若干次，得观测数据$(x_{i0},x_{i1},\ldots,x_{ik})$，i=1,...,n。根据这n次观测数据去估计$(\theta_1,\ldots,\theta_k)$，得到方程组

$x_{i0}+x_{i1}\theta_1+\cdots+x_{ik}\theta_k=0$　　i=1,...,n

如果依据该方程组来求$(\theta_1,\ldots,\theta_k)$，便存在一些问题，首先，此方程组可能无解，其次，即使在该方程组中选出部分方程求解，那么应当如何选择，哪一组更精确，另外，选出的部分方程求出的解和真值的差别有多大。

由于各种可知或不可知、可控或不可控因素在观测过程中的影响和干扰，使得$(x_0,x_1,\ldots,x_k)$的观测值产生误差，所以，更合理的模型是$x_{i0}+x_{i1}\theta_1+\cdots+x_{ik}\theta_k=e_i$　　i=1,...,n。其中e_i为随机误差。

如何充分利用全部的观测结果，从而得到一个好的估计$(\hat{\theta}_1,\cdots,\hat{\theta}_k)$，这个问题困扰了18世纪的许多数学家，包括欧拉、拉普拉斯等人，他们的思路一直囿于构造部分方程组来求解。

勒让德在研究天文学计算彗星的轨道时，提出了与前人相异的思想，他认识到寻找的$(\hat{\theta}_1,\cdots,\hat{\theta}_k)$应该使各次观测产生的误差

e_i（i=1，...，n）的平方和最小，这在当时数据处理上是一个很大的进步，开辟了工程计算的新道路。

在勒让德的方法中，最小二乘估计（LSE）完全是一种计算方法，缺少误差分析，对方程组的近似解的性质好坏无法评价。实际上，误差的性质决定了参数估计的统计性质。Gauss 的研究给出了误差的正态理论，他依据这个分布，独立地导出了 LSE。

LSE 方法与误差理论相结合，使 LSE 得以极大的成功，在实际应用中，LSE 是 19 世纪数理统计学中占统治地位的主题。1815 年，LSE 已成为法国、普鲁士在天文和测地学中的标准工具，1825 年在英国广泛应用。目前，相关分析、回归分析、方差分析以及其它一些与 LSE 有密切关系的统计方法仍在应用中占重要的地位。

§2.2 未知参数的最小二乘估计

考虑一般线性模型

$$\begin{cases} y = X\beta + e \\ e \sim (0, \sigma^2 I_n) \end{cases} \text{（即 Ee=0，De=}\sigma^2 I_n\text{）} \tag{1}$$

我们的问题是当 $x_1,...,x_p$ 给定时，如何确定 β 的估计。获得回归系数 $\beta = (\beta_1,...,\beta_p)^T$ 的估计的一种方法是最小二乘法，它是将误差平方和

$$e^T e = (y - x\beta)^T (y - x\beta) \hat{=} \|y - x\beta\|^2$$

关于 β 极小化的方法。

定义 1 若 $\hat{\beta} \in R^p$ 满足 $\left\|y - x\hat{\beta}\right\|^2 = \min_{\beta \in R^p} \|y - x\beta\|^2$　　（2）

则称$\hat{\beta}$为β的最小二乘估计（Least Squares Estimate，简记为 LSE)。

注意到，$\left\|y-x\hat{\beta}\right\|^2=(y-x\beta)^T(y-x\beta)$

$$=\sum_{i=1}^{n}[Y_i-(\beta_1x_{i1}+\cdots+\beta_px_{ip})]^2$$

为求其极小值点$\hat{\beta}=(\hat{\beta}_1,\cdots,\hat{\beta}_p)^T$，可将$(y-x\beta)^T(y-x\beta)$ 对

β_j（j=1,...p）分别求偏导数，并令其为零得：

$$2\sum_{i=1}^{n}[Y_i-(\beta_1x_{i1}+\cdots+\beta_px_{ip})](-x_{ij})=0，\ j=1,...,p$$

整理后得：

$$(\sum_{i=1}^{n}x_{i1}x_{ij})\beta_1+\cdots+(\sum_{i=1}^{n}x_{ip}x_{ij})\beta_p=\sum_{i=1}^{n}x_{ij}Y_i\quad ,\ j=1,...,p$$

将上面的 p 个方程合写成一个矩阵方程，即

$$X'X\beta=X'y \tag{3}$$

称此方程为关于回归系数β的正规方程，它在线性模型理论中起着重要作用。

定理 1（1）正规方程必有解（相容）；

（2）β的 LSE $\hat{\beta}$必为正规方程的解；

（3）正规方程的任一解必为β的 LSE。

证明：（2）如上所证。由线性方程组的理论知，正规方程有解的充要条件为

rk$(X'X\vdots X'y)$=rk$(X'X)$

该等式是成立的，首先，rk$(X'X\vdots X'y)\geq$rk$(X'X)$，又

$$\mathrm{rk}(X'X \vdots X'y)=\mathrm{rk}[X'(X \vdots y)]\leqslant \mathrm{rk}(X')=\mathrm{rk}(X'X)$$

所以 $\mathrm{rk}(X'X \vdots X'y)=\mathrm{rk}(X'X)$，从而正规方程必有解。(1) 成立。

设 $\hat{\beta}$ 为正规方程的任一解，即 $X'X\hat{\beta}=X'y$，则 $\forall\beta\in R^p$ 有

$$(y-x\beta)^T(y-x\beta)=[(y-x\hat{\beta})+x(\hat{\beta}-\beta)]^T[(y-x\hat{\beta})+x(\hat{\beta}-\beta)]$$

$$=(y-x\hat{\beta})^T(y-x\hat{\beta})+2(\hat{\beta}-\beta)^T X'(y-X\hat{\beta})+(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)$$

$$\geqslant(y-x\hat{\beta})^T(y-x\hat{\beta})$$

其中用到 $(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)\geqslant 0$，且

$$(\hat{\beta}-\beta)^T X'(y-X\hat{\beta})=(\hat{\beta}-\beta)^T(X'y-X'X\hat{\beta})=0$$

所以 $(y-x\hat{\beta})^T(y-x\hat{\beta})=\min\limits_{\beta\in R^p}(y-x\beta)^T(y-x\beta)$

即 $\left\|y-x\hat{\beta}\right\|^2=\min\limits_{\beta\in R^p}\left\|y-x\beta\right\|^2$

从而正规方程的任一解 $\hat{\beta}$ 必为 β 的 LSE。(3) 成立。

推论 1.1 正规方程 $X'X\beta=X'y$ 的通解为 $\hat{\beta}=(X'X)^-X'y$，此处 $(X'X)^-$ 为 $X'X$ 的任一广义逆。

证明：由定理 1 和第一章 §1.2.1 节定理 5 即得。

推论 1.2 若 rk(X)=p, 则 β 的 LSE 由公式 $\hat{\beta}=(X'X)^{-1}X'y$ 唯一确

定，且此时$\hat{\beta}$为β的无偏估计，而$D\hat{\beta}=\sigma^2(X'X)^{-1}$。

推论 1.3 假设改变回归因子的尺度，使对一切 i、j 有 $x_{ij}=k_j w_{ij}(k_j\neq 0)$，则在此种变换下$\hat{y}=X\hat{\beta}$保持不变。

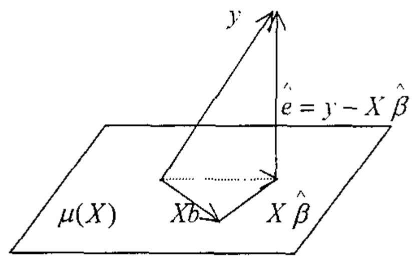

图 1

注：若在$e^Te=(y-x\beta)^T(y-x\beta)\hat{=}\|y-x\beta\|^2$中令$x\beta=\theta$，则$\theta\in\mu(X)$，从而最小二乘法就是将$\|y-\theta\|^2$关于$\theta\in\mu(X)$极小化的一种方法。

由于 y 与$\mu(X)$诸点之间的距离以垂直距离为最短（参见图 1），故当$y-\hat{\theta}\perp\mu(X)$时，即$(y-\hat{\theta})'X\alpha=0,\forall\alpha\in R^p$，亦即$X'(y-\hat{\theta})=0$时

$$\left\|y-\hat{\theta}\right\|^2=\min_{\theta\in R(X)}\|y-\theta\|^2$$

此时$X'\hat{\theta}=X'y$，又注意到$\hat{\theta}=X\beta$，对某个$\beta\in R^p$，故有

$$X'X\beta=X'y$$

§2.3 可估计函数与 Gauss-Markov 定理

当设计矩阵 X 为列满秩时，正规方程 $X'X\beta = X'y$ 有唯一解 $\hat{\beta} = (X'X)^{-1}X'y$（见 §2.2 推论 1.2）。但当 rk(X)<p 时，广义逆 $(X'X)^{-}$ 的不唯一性导致了 LS 解 $\hat{\beta}$ 的不唯一。从而用 $\hat{\beta}$ 去估计 β 时，就会遇到两个问题，第一，$\hat{\beta}$ 不唯一；第二，这许多 $\hat{\beta}$ 中没有一个是 β 的无偏估计，因此在 rk(x)<p 时，根本不存在 β 的线性无偏估计。

事实上，设 A 为 $p\times n$ 矩阵，使得 Ay 为 β 的无偏估计，则 $\mathrm{E}Ay=\beta$，$\forall\beta\in R^{p}$，即 $AX\beta=\beta$，$\forall\beta\in R^{p}$，从而 $AX=I_p$，但是我们有 $p=rk(I_p)=rk(AX)\le rk(X)<p$，故这样的 A 根本不存在。此时，我们称 β 是不可估计的。

说的明白一些，假定我们要称重量分别为 β_1 和 β_2 的两个物体，可是每次将它们放在重量为 β_3 的容器里去称，各称两次，记 y_{ij} 为第 i 件物体第 j 次称出的重量，误差记为 e_{ij}，则

$y_{ij}=\beta_3+\beta_i+e_{ij}$，$e_{ij}\sim(0,\ \sigma^2)$，i=1,2; j=1,2。

由于我们每次把容器和一件物体放在一起称，又事先不知道容器重量，我们不能估算出两个物体以及容器的重量 β_1、β_2 和 β_3，但是两个物体重量之差却可估计，因为将两次称出的值 y_{1j} 和 y_{2j} 相减就是 $\beta_1-\beta_2$ 的无偏估计。如此看来，我们应放弃估计整个 β 的想法，考查能否估计 β 的某些线性函数。

定义 1 若存在 $\alpha'\beta$ 的线性无偏估计 $b'y$，其中 $\alpha=(\alpha_1,\dots,\alpha_p)'$，

$b=(b_1,\ldots,b_n)'$，则称β的线性函数$\alpha'\beta=\sum_{i=1}^{p}\alpha_i\beta_i$是可估的。

定理 1 $\alpha'\beta$是可估计函数的充要条件是$\alpha\in\mu(X')$。

证明：必要性 若$\alpha'\beta$可估，则存在$b'y$，使得 $Eb'y=\alpha'\beta$，$\forall\ \beta\in R^p$，即 $b'X\beta=\alpha'\beta$，$\forall\ \beta\in R^p$，这等价于$b'X=\alpha'$，亦即$\alpha=X'b\in\mu(X')$。

充分性 若$\alpha\in\mu(X')$，则$\exists b\in R^n$，使得$\alpha=X'b$，显然$b'y$就是$\alpha'\beta$的线性无偏估计，所以$\alpha'\beta$可估。

注：定理 1 说明，使$\alpha'\beta$可估的全体$p\times1$向量α构成子空间$\mu(X')$。于是，若$\alpha_{(1)}$、$\alpha_{(2)}$为$p\times1$向量，使得$\alpha_{(1)}'\beta$和$\alpha_{(2)}'\beta$均可估，那么，对任意两个数 m, n, 线性组合 $m\alpha_{(1)}'\beta+n\alpha_{(2)}'\beta$都可估。

例 1 对线性模型$\begin{cases}y_1=\beta_1+\beta_2+e_1\\ y_2=\beta_1+\beta_3+e_2\\ y_3=\beta_1+\beta_2+e_3\end{cases}$，证明：$\sum_{i=1}^{3}C_i\beta_i$可估$\Leftrightarrow C_1=C_2+C_3$。

证明：$$\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}1&1&0\\1&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\\\beta_3\end{pmatrix}+\begin{pmatrix}e_1\\e_2\\e_3\end{pmatrix}$$

由定理 1 知$\sum_{i=1}^{3}C_i\beta_i$可估$\Leftrightarrow\begin{pmatrix}C_1\\C_2\\C_3\end{pmatrix}\in\mu\left(\begin{pmatrix}1&1&1\\1&0&1\\0&1&0\end{pmatrix}\right)$

$$\Leftrightarrow\exists t=\begin{pmatrix}t_1\\t_2\\t_3\end{pmatrix},\quad\begin{pmatrix}C_1\\C_2\\C_3\end{pmatrix}=\begin{pmatrix}1&1&1\\1&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}t_1\\t_2\\t_3\end{pmatrix}$$

$$\Leftrightarrow C_1=C_2+C_3$$

例 2 对线性模型 $y=X\beta+e,\ \ e\sim(0,\sigma^2I_n)$， 证明线性函数

$\beta_1-\beta_2$， $\beta_1-\beta_3$， . . . ， $\beta_1-\beta_p$ 可估 $\Leftrightarrow$ 对一切满足 $\sum_{i=1}^{p}C_i=0$ 的 $C_1,\dots,C_p$ 来说， $\sum_{i=1}^{p}C_i\beta_i$ 可估。

证明：由定理 1 知，

$$\beta_1-\beta_2,\dots,\beta_1-\beta_p \text{可估} \Leftrightarrow \begin{pmatrix}1\\-1\\0\\\vdots\\0\end{pmatrix}\in\mu(X'),\begin{pmatrix}1\\0\\-1\\\vdots\\0\end{pmatrix}\in\mu(X'),\dots,\begin{pmatrix}1\\0\\0\\\vdots\\-1\end{pmatrix}\in\mu(X')$$

$$\Leftrightarrow \forall C_i\in R,(i=2,\dots,p),-C_2\begin{pmatrix}1\\-1\\0\\\vdots\\0\end{pmatrix}-C_3\begin{pmatrix}1\\0\\-1\\\vdots\\0\end{pmatrix}-\cdots-C_p\begin{pmatrix}1\\0\\\vdots\\0\\-1\end{pmatrix}\in\mu(X')$$

$$\Leftrightarrow \forall C_i\in R,i=2,3,\cdots,p,\begin{pmatrix}-C_2-C_3-\cdots-C_p\\C_2\\C_3\\\vdots\\C_p\end{pmatrix}\in\mu(X')$$

$$\Leftrightarrow \text{令 } C_1=-C_2-C_3-\cdots-C_p,\quad \begin{pmatrix}C_1\\C_2\\\vdots\\C_p\end{pmatrix}\in\mu(X')$$

$\Leftrightarrow$ 对一切满足 $\sum_{i=1}^{p}C_i=0$ 的 $C_1,\dots,C_p$ 来说， $\sum_{i=1}^{p}C_i\beta_i$ 可估。

推论 1.1 $\alpha'\beta$ 是可估计函数的充要条件是 $\alpha\in\mu(X'X)$。

此因 $\mu(X'X)=\mu(X')$。

推论 1.2 若$\alpha'\beta$可估，则$\alpha'\hat{\beta}$唯一，且 $E\alpha'\hat{\beta}=\alpha'\beta$，其中$\hat{\beta}$为正规方程的任一解。

证明：$\alpha'\beta$可估，则$\alpha\in\mu(X'X)$，$\exists b\in R^p$，使$\alpha=X'Xb$，

$\alpha'\hat{\beta}=b'X'X(X'X)^-X'y=b'X'y$与$(X'X)^-$的选择无关。

$E\alpha'\hat{\beta}=E(b'X'y)=b'X'X\beta=\alpha'\beta$，即$\alpha'\hat{\beta}$为$\alpha'\beta$的无偏估计。

定义 2 对可估函数$\alpha'\beta$，称$\alpha'\hat{\beta}$为$\alpha'\beta$的 LS 估计。

注：当 rk(X)=r=p 时，一切线性函数$\alpha'\beta$都是可估的，β也是可估的，其无偏估计为$\hat{\beta}=(X'X)^{-1}X'y$，此时称$\hat{\beta}$为$\beta$的 LS 估计。只有当$\beta$可估时，我们才称$\hat{\beta}$为$\beta$的 LS 估计，否则，称其为$\beta$的 LS 解。

对于可估函数$\alpha'\beta$，其无偏估计有无穷多个。实际上，设$b'y$为$\alpha'\beta$的一个无偏估计，那么$\forall c\in\mu(X)^{\perp}$，$(b+c)'y$也是$\alpha'\beta$的一个无偏估计，这是因为 $E(b+c)'y=Eb'y+Ec'y=\alpha'\beta$。

定义 3 称$b'y$是$\alpha'\beta$的最佳线性无偏估计（BLUE）或 Gauss-Markov 估计（GME），如果（1）$b'y$是$\alpha'\beta$的无偏估计；
（2）对$\alpha'\beta$的任一线性无偏估计$c'y$，都有 $D(b'y)\leqslant D(c'y)$。

定理 2（Gauss-Markov 定理）若$\alpha'\beta$可估，则 LS 估计$\alpha'\hat{\beta}$是$\alpha'\beta$的唯一 BLUE 估计。

证明：定理 1 推论 1.2 已证$\alpha'\hat{\beta}$是$\alpha'\beta$的无偏估计，而线性性显

然成立，故$\alpha'\hat{\beta}$是$\alpha'\beta$的线性无偏估计。

$$
\begin{aligned}
D(\alpha'\hat{\beta}) &= D(\alpha'(X'X)^{-}X'y) \\
&= \sigma^2\alpha'(X'X)^{-}X'X(X'X)^{-}\alpha
\end{aligned}
$$

由$\alpha'\beta$可估知，$\alpha \in \mu(X')$，$\exists b \in R^n$，$\alpha = X'b$，利用

$X'X(X'X)^{-}X' = X'$ 得到

$$
\begin{aligned}
D(\alpha'\hat{\beta}) &= \sigma^2\alpha'(X'X)^{-}X'X(X'X)^{-}X'b \\
&= \sigma^2\alpha'(X'X)^{-}X'b \\
&= \sigma^2\alpha'(X'X)^{-}\alpha
\end{aligned}
$$

又若$c'y$是$\alpha'\beta$的任一线性无偏估计，则$X'c=\alpha$，

$$
\begin{aligned}
D(c'y)-D(\alpha'\hat{\beta}) &= \sigma^2[c'c-\alpha'(X'X)^{-}\alpha] \\
&= \sigma^2[c'-\alpha'(X'X)^{-}X'][c'-\alpha'(X'X)^{-}X']' \\
&\geqslant 0
\end{aligned}
$$

等号成立当且仅当$c' = \alpha'(X'X)^{-}X'$，亦即$c'y=\alpha'\hat{\beta}$。

Gauss-Markov 定理提供了可估函数$\alpha'\beta$的 BLUE 的非常方便的求法。该定理奠定了 LS 估计在线性模型参数估计理论中的地位。出它所刻画的 LS 估计在线性无偏估计类中的最优性，使得人们长期以来把 LS 估计当作线性模型的唯一最好的估计。但是，到了本世纪六七十年代，许多研究标明，在许多情况下，LS 估计的性质并不很好，若采用另外一些更好的度量估计优劣的标准，LS 估计不一定最优，本文后面将进行论述。

推论 2.1 $X\beta$ 必可估，即每一 $e_i'X\beta$ 均可估，i=1,...,n。且 $X\hat{\beta}$ 是 $X\beta$ 的 BLUE，其中 $\hat{\beta}=(X'X)^{-}X'y$。

推论 2.2 若 $\hat{\theta_i}=\alpha_i'\beta$ 可估，i=1,...,k，则它们的线性组合 $\sum_{i=1}^{k}c_i\theta_i$ 也可估，且 $\sum_{i=1}^{k}c_i\hat{\theta}_i=\sum_{i=1}^{k}c_i\alpha_i'\hat{\beta}$ 是 $\sum_{i=1}^{k}c_i\theta_i$ 的 BLUE。

推论 2.3 若 $rk(X)_{n\times p}=p$，则 $\forall\alpha\in R^p$，$\alpha'\beta$ 可估。

定理 3 对于线性模型 $y=X\beta+\varepsilon$，$\varepsilon\sim(0,\sigma^2 I_n)$，$A\beta$ 可估的充要条件是下列条件之一成立：

（1） 存在矩阵 B，使 $A=BX$；

（2） $rk\begin{bmatrix}A\\X\end{bmatrix}=rk[X]$；

（3） $rk[X(I-A^{-}A)]=rk(X)-rk(A)$；

（4） $AX^{-}X=A$。

证明： $A\beta$ 可估，即 $e_i'A\beta$ 均可估，i=1,...,n。

$\Leftrightarrow A'e_i\in\mu(X')$，i=1,...,n.

$\Leftrightarrow \exists b_i\in R^n$，$A'e_i=X'b_i$，i=1,...,n.

$\Leftrightarrow A'(e_1e_2\cdots e_n)=X'(b_1b_2\cdots b_n)$

$\Leftrightarrow A'=X'(b_1b_2\cdots b_n)$

$\Leftrightarrow$ 令 $B=(b_1b_2\cdots b_n)'$，$A=BX$，即（1）成立。

（1）$\Rightarrow$（2）：

$rk\begin{bmatrix}A\\X\end{bmatrix}\geq rk[X]$ 显然成立，又 $rk\begin{bmatrix}A\\X\end{bmatrix}=rk\begin{bmatrix}BX\\X\end{bmatrix}=rk(\begin{bmatrix}B\\I\end{bmatrix}X)\leq rk[X]$，

故 $rk\begin{bmatrix} A \\ X \end{bmatrix} = rk[X]$，（2）成立。

（2）$\Rightarrow$（3）：

设 $rk(A)=r$，则存在可逆矩阵 Q，使得 $AQ=(I_r \vdots 0)$，于是

$$\begin{aligned} rk[X(I-A^{-}A)] &= \dim[Xt : At=0] \\ &= \dim[\begin{pmatrix} A \\ X \end{pmatrix} t : At=0] \\ &= \dim[\begin{pmatrix} A \\ X \end{pmatrix} Qt : AQt=0] \\ &= \dim\left\{\begin{pmatrix} I_r & 0 \\ u_1 & u_2 \end{pmatrix} t : (I_r \vdots 0)t=0\right\} \end{aligned}$$

（其中 $(u_1 \vdots u_2)=XQ$，$t=\begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix}_{(m-r)\times 1}^{r\times 1}$）

$$\begin{aligned} &= \dim\{u_2 t_{(2)} : t_{(2)}\text{任意}\} \\ &= rk(u_2) \\ &= rk\begin{pmatrix} I_r & 0 \\ u_1 & u_2 \end{pmatrix} - rk(I_r) \\ &= rk\left[\begin{pmatrix} A \\ X \end{pmatrix} Q\right] - rk(A) \\ &= rk\left[\begin{pmatrix} A \\ X \end{pmatrix}\right] - rk(A) \\ &= rk(X) - rk(A) \end{aligned}$$

(3) $\Rightarrow$ (2)： 由（2）$\Rightarrow$（3）证明过程知，

$$rk[X(I-A^{-}A)] = rk\left[\begin{pmatrix} A \\ X \end{pmatrix}\right] - rk(A)$$

而 $rk[X(I-A^{-}A)] = rk(X) - rk(A)$

故 $rk\begin{bmatrix} A \\ X \end{bmatrix} = rk[X]$, (2) 成立。

（2）$\Rightarrow$（4）：证明方法类似（2）$\Rightarrow$（3）

$$rk(A - AX^{-}X) = rk\left[\begin{pmatrix} A \\ A \end{pmatrix}\right] - rk(A) = 0$$

$$A - AX^{-}X = 0$$

$$AX^{-}X = A$$

（4）$\Rightarrow$（1）：令 $B = AX^{-}$ 即可。

总之，我们证明了：

$$\begin{tikzcd}
A\beta\text{可估} \arrow[r, leftrightarrow] & (1) \arrow[r, leftrightarrow] & (2) \arrow[r, leftrightarrow] \arrow[d] & (3) \\
 & & (4) \arrow[ul] &
\end{tikzcd}$$

§2.4 β的广义最小二乘估计

到目前为止，我们的讨论都假定误差协方差矩阵为 $\sigma^2 I$，但是，客观上存在着许多线性模型，其误差协方差矩阵为 $\sigma^2 V$，此时 β 的估计又是怎样的呢？

在 $V>0$ 且已知的情况下，存在着唯一的正定对称矩阵 $V^{\frac{1}{2}}$，使 $(V^{\frac{1}{2}})^2 = V$ 用 $V^{-\frac{1}{2}} = (V^{\frac{1}{2}})^{-1}$ 左乘 $y = X\beta + e$，并记 $\tilde{y} = V^{-\frac{1}{2}}y$，$\tilde{X} = V^{-\frac{1}{2}}X$，

$\tilde{e}=V^{-\frac{1}{2}}e$ 则得到 $\tilde{y}=\tilde{X}\beta+\tilde{e}$ $\qquad E\tilde{e}=0, Cov(\tilde{e})=\sigma^2 I$

对此变换后的模型来说，β 的 LSE 为

$$\beta^{*}=(\tilde{X}'\tilde{X})^{-}\tilde{X}'\tilde{y}=(X'V^{-1}X)^{-}X'V^{-1}y$$

定义 1 称上述 β^{*} 为线性模型 $\begin{cases} y=X\beta+e \\ e\sim(0,\sigma^2 V) \end{cases}$（V>0 已知）中 β 的广义最小二乘估计，特别地，当 $V=diag(\sigma_1^2,\cdots,\sigma_n^2),\sigma_i^2$，i=1,...,n 已知时，称 β^{*} 为加权最小二乘估计。

若我们对 $\tilde{e}'\tilde{e}=e'V^{-1}e$

$$\begin{aligned} &=(y-X\beta)'V^{-1}(y-X\beta) \\ &=y'V^{-1}y-2\beta'X'V^{-1}y+\beta'X'V^{-1}X\beta \end{aligned}$$

关于 β 求导，并令其为零得：

$$-2X'V^{-1}y+2X'V^{-1}X\beta=0$$

从而 $\beta=(X'V^{-1}X)^{-}X'V^{-1}y\hat{=}\beta^{*}$

特别地，当 X 为列满秩时，β 的广义 LSE 唯一地为

$\beta^{*}=(X'V^{-1}X)^{-1}X'V^{-1}y$

由于上述模型和以前讨论的模型只是误差协方差矩阵不同，而线性函数 $c'\beta$ 的可估性又与协方差矩阵无关，于是，对上述模型 $c'\beta$ 可估的充要条件仍为 $c\in\mu(X')$。

定义 2 称 $c'\beta^{*}$ 为可估函数 $c'\beta$ 的广义最小二乘估计，简记为 GLS 估计。对应地，当 V 为对角矩阵时，$c'\beta^{*}$ 称为 $c'\beta$ 的加权最小二乘估计，简记为 WLS 估计。

定理 1 设 V>0 且已知，若 $c'\beta$ 可估，则 $c'\beta^{*}$ 是 $c'\beta$ 的唯一 BLUE。

证明：因为$c'\beta$可估，所以$c=X'b$，$b\in R^n$。从而

$c'\beta^*=b'X(X'V^{-1}X)^-X'V^{-1}y$唯一（与广义逆的选择无关）

且 $Ec'\beta^*=b'X(X'V^{-1}X)^-X'V^{-1}X\beta=b'X\beta=c'\beta$

即 $c'\beta^*$是$c'\beta$的线性无偏估计。

$$D(c'\beta^*)=\sigma^2c(X'V^{-1}X)^-X'V^{-1}X(X'V^{-1}X)^-c$$
$$=\sigma^2c(X'V^{-1}X)^-c$$

设$a'y$为$c'\beta$的任一线性无偏估计，则$c=X'a$，故

$$D(a'y)-D(c'\beta^*)=\sigma^2[a'Va-c'(X'V^{-1}X)^-c]$$
$$=\sigma^2[a'Va-a'X(X'V^{-1}X)^-X'a]$$

（令$b=V^{\frac{1}{2}}a$，$Q=V^{-\frac{1}{2}}X$，$P_Q=Q(Q'Q)^-Q'$）

$$=\sigma^2[b'b-b'Q(Q'Q)^-Q'b]$$
$$=\sigma^2b'(I-P_Q)b$$
$$\geq 0$$

此因P_Q为对称、幂等矩阵，$I-P_Q\geq 0$，这就证明了$c'\beta^*$的方差最小性。

上式等号成立$\Leftrightarrow (I-P_Q)b=0\Leftrightarrow b=P_Qb$

$\Leftrightarrow a=V^{-1}X(X'V^{-1}X)^-c\Leftrightarrow a'y=c'\beta^*$

唯一性得证。

推论 1.1 若 V>0 已知，则当 X 为列满秩时，$c'\beta^*$是$c'\beta$的唯一 BLUE。

§2.5 LSE 估计的稳健性

虽然稳健性（robustbess）这种思想在统计学中由来已久，但直

到本世纪中期才受到统计学家的重视，1953 年 G.E.Box 第一次明确提出“稳健性”一词。直观上讲，稳健性是指统计推断关于统计模型即假设条件具有相对稳定性。也就是说，当模型假设发生某种微小变化时，相应的统计推断也只有微小的改变。此时，我们就说统计推断关于这种微小变化具有稳健性。在前面，关于线性模型有一个重要的假设 $Cov(e)=\sigma^2 I$。在此条件下，我们证明了可估函数 $c'\beta$ 的 LS 估计 $c'\hat{\beta}$ 是 BLUE。但在应用上我们不可能要求一个实际问题完全满足该假设。事实上，我们也根本无法知道，它确实满足这一假设，只能通过分析或检验，判断假设 $Cov(e)=\sigma^2 I$ 是否大致上可以接受。因此，我们总是希望当实际的 $Cov(e)$ 与 $\sigma^2 I$ 相差不太远时，LS 估计 $c'\hat{\beta}$ 仍然保持原来的最优性或即使不是最优的，也不要变得太坏。若是这样的话，我们说 LS 估计关于协方差矩阵是稳健的。相反，若出现失之毫厘，谬之千里的情况，这个估计就不具有稳健性。

考查线性模型 $\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ （V>0），

可估函数 $c'\beta$ 的最佳线性无偏估计（BLUE）为 $c'\beta^* = c'(X'V^{-1}X)^- X'V^{-1}y$ 且 $c'\beta^*$ 与广义逆的选择无关。如果我们无视 $Cov(e)=\sigma^2 V \neq \sigma^2 I$，而按以前的 $Cov(e)=\sigma^2 I$ 情形来处理，这就导致了 LS 解 $\hat{\beta}=(X'X)^- X'y$。如此看来，对任一可估函数 $c'\beta$，我们就有了两个估计：LSE 估计 $c'\hat{\beta}$ 和 GLSE 估计 $c'\beta^*$，两者都是无偏估计，而后者为 BLUE 估计（§2.4 定理 1）。一般来说，$c'\hat{\beta} \neq c'\beta^*$，即 LSE 估计和 BLUE 估计不一定相

等，这是和 $Cov(e)=\sigma^2 I$ 情形所不同的。什么条件下二者相等呢？Zyskind 曾给出了一个充要条件，我们用不同于 Zyskind 的方法，得到一些新的结论，解决了这个问题。在某些情况下，比 Zyskind 的结果更好。先介绍 Zyskind 的结果。

定理 1(Zyskind) 对于线性模型 $\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ （V>0），则

$c'\hat{\beta} = c'\beta^*$，$\forall c \in \mu(X') \Leftrightarrow a'Vb = 0, \forall a \in \mu^{\perp}(X), b \in \mu(X)$

$\Leftrightarrow \mu(VX) = \mu(X)$。特别当 X 列满秩时，$\hat{\beta} = \beta^* \Leftrightarrow \mu(VX) = \mu(X)$。

定理 2 对于线性模型 $\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ （V>0），若 $rk(X_{n\times p}) = p$，则

$\hat{\beta} = \beta^*$（从而对任意的 c 有 $c'\hat{\beta} = c'\beta^*$）的充要条件为 $D(\hat{\beta}) = D(\beta^*)$。

注：由§2.4 定理 1 知，一般来说，$D(\hat{\beta}) \geq D(\beta^*)$。

证明：必要性显然成立。下证充分性，

$$D(\hat{\beta}) = \sigma^2 (X'X)^{-1} X'VX(X'X)^{-1}$$

$$D(\beta^*) = \sigma^2 (X'V^{-1}X)^{-1}$$

$$D(\hat{\beta}) = D(\beta^*) \Leftrightarrow (X'X)^{-1} X'VX(X'X)^{-1} = (X'V^{-1}X)^{-1}$$

$$\Leftrightarrow X'VX(X'X)^{-1} = X'X(X'V^{-1}X)^{-1}$$

$$Cov\{(X'X)^{-1}X'y, [V^{-1}X(X'V^{-1}X)^{-1} - X(X'X)^{-1}]'y\}$$

$$= \sigma^2 (X'X)^{-1} X'V[V^{-1}X(X'V^{-1}X)^{-1} - X(X'X)^{-1}]$$

$$= \sigma^2 (X'X)^{-1}[X'VV^{-1}X(X'V^{-1}X)^{-1} - X'VX(X'X)^{-1}]$$

$=\sigma^2(X'X)^{-1}[X'X(X'V^{-1}X)^{-1}-X'VX(X'X)^{-1}]$

$=0$

故 $D(\beta^*)=D\{\hat{\beta}+[V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}]'y\}$

$$=D(\hat{\beta})+D\{[V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}]'y\}$$

从而 $D(\beta^*)\geqslant D(\hat{\beta})$，又因 β^* 为 BLUE，故 $D(\hat{\beta})\geq D(\beta^*)$。

所以 $D(\beta^*)=D(\hat{\beta})$，因而

$D\{[V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}]'y\}=0$，即

$\sigma^2[V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}]'V[V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}]=0$

由于 V 为正定矩阵，故

$V^{-1}X(X'V^{-1}X)^{-1}-X(X'X)^{-1}=0$

$(X'X)^{-1}X'=(X'V^{-1}X)^{-1}X'V^{-1}$

从而 $\hat{\beta}=\beta^*$。

推论 2.1 对于线性模型$\begin{cases} y=X\beta+e \\ e\sim(0,\sigma^2V) \end{cases}$ （V>0），若 $rk(X_{n\times p})=p$，

则 $\hat{\beta}=\beta^*$（从而对任意的 c 有 $c'\hat{\beta}=c'\beta^*$）的充要条件为

$\mu(VX)=\mu(X)$。

证明：由定理 2 的证明过程以及 $rk(X_{n\times p})=p$ 知，

$D(\hat{\beta})=D(\beta^*)\Leftrightarrow X'VX(X'X)^{-1}=X'X(X'V^{-1}X)^{-1}$

$$\Leftrightarrow VX=X(X'V^{-1}X)^{-1}X'X$$

$$\Leftrightarrow VX(X'X)^{-1}X'V^{-1}X = X$$

$$\Leftrightarrow \mu(VX) = \mu(X)$$

推论 2.2 对于线性模型$\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ (V>0)，若 X 为 n 阶可逆方阵，矩阵 V 和 X 以及X'乘积可换，则$\hat{\beta} = \beta^*$（从而对任意的 c 有$c'\hat{\beta} = c'\beta^*$）。

上面我们对 X 为列满秩的情形，给出了 GLSE 和 LSE 相等的一个充要条件和一个充分条件，对于一般情形，我们在下面定理 3 及其推论中给出一些充分条件。为证明定理 3，先提出一些引理。

引理 1 若$a \in \mu(X)$，$b \in \mu^{\perp}(X)$且$\{Vx : x \in \mu(X)\} \subseteq \mu(X)$，则对于线性模型$\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ (V>0)，有$Cov(a'y, b'y) = 0$。

证明：$Cov(a'y, b'y) = a'D(y)b = \sigma^2 a'Vb = \sigma^2 (Va)'b = 0$

引理 2 对于线性模型$\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ (V>0)，若$c'y$是 0 的无偏估计，则有$c \in \mu^{\perp}(X)$。

证明：由于对一切β，有$E(c'y) = c'X\beta$，所以$c'X = 0$，即$c \in \mu^{\perp}(X)$。

定理 3 对于线性模型$\begin{cases} y = X\beta + e \\ e \sim (0,\sigma^2 V) \end{cases}$ (V>0)，若$VX(X'X)^{-}c \in \mu(X)$，则对任意可估函数$c'\beta$，有 $c'\hat{\beta} = c'\beta^*$。

证明：$c'\hat{\beta} = c'(X'X)^{-}X'y$

$c'\beta^* = c'(X'V^{-1}X)^- X'V^{-1}y$

$c'\beta$ 可估，由§2.3 中定理 1 知，$c \in \mu(X)$，又

$c'\beta^* - c'\hat{\beta} = [V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'y$ 是 0 的无偏估计，由引理知，$Cov\{c'(X'X)^- Xy, [X(X'X)^- c - V^{-1}X(X'V^{-1}X)^- c]'y\} = 0$，

由于 $c'\beta^*$ 是 $c'\beta$ 的 BLUE，故 $D(c'\hat{\beta}) \ge D(c'\beta^*)$。另一方面，我们有

$$
\begin{aligned}
D(c'\beta^*) &= D\{c'(X'X)^- X'y + [V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'y\} \\
&= D[c'(X'X)^- X'y] + D\{[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'y\} \\
&\geqslant D[c'(X'X)^- X'y] \\
&= D(c'\hat{\beta})
\end{aligned}
$$

故 $D(c'\hat{\beta}) = D(c'\beta^*)$，且 $D\{[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'y\} = 0$，又

$D\{[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'y\}$
$= \sigma^2 [V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'V[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]$

所以

$[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c]'V[V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c] = 0$

由 V 的正定性知，$V^{-1}X(X'V^{-1}X)^- c - X(X'X)^- c = 0$

从而 $c'\hat{\beta} = c'\beta^*$。

推论 3.1 对于线性模型 $\begin{cases} y = X\beta + e \\ e \sim (0, \sigma^2 V) \end{cases}$ （V>0），若存在矩阵 $C_{p\times n}$，使得 $V = XC$，则对任意可估函数 $c'\beta$，有 $c'\hat{\beta} = c'\beta^*$。

推论 3.2 对于线性模型$\begin{cases} y = X\beta + e \\ e \sim (0, \sigma^2 V) \end{cases}$ （V>0）， 若存在矩阵 $D_{p\times p}$，使得$VX = XD$，则对任意可估函数$c'\beta$，有 $c'\hat{\beta} = c'\beta^*$。

§2.6 最小二乘估计的缺陷

近几十年来，LSE 在数理统计中的统治地位开始动摇。在处理许多回归问题时，发现 LSE 并不理想，LSE 的性质与设计矩阵 X 有着紧密关系。注意下列关系式

$$E\left\|\hat{\beta}\right\|^2 = E(\hat{\beta}'\hat{\beta}) = tr[D(\hat{\beta})] + \|\beta\|^2$$

出此可知，$E\left\|\hat{\beta}\right\|^2$总是偏离真实参数$\beta$的长度的平方$\|\beta\|^2$。此外，由于$X'X$为实对称矩阵，于是存在正交矩阵 Q，使得

$$Q'X'XQ = \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix},$$

其中$\lambda_1,\ldots,\lambda_p$为$X'X$的非零特征值。于是，我们有

$$X'X = Q\Lambda Q',$$

$$(X'X)^{-1} = Q\Lambda^{-1}Q',$$

$$tr(X'X)^{-1} = tr(Q\Lambda^{-1}Q') = tr(\Lambda^{-1}QQ') = tr\Lambda^{-1} = \sum_{i=1}^{p}\frac{1}{\lambda_i}$$

从而 $E\left\|\hat{\beta}\right\|^2=\|\beta\|^2+\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i}$。

因此，如果 $rk(X'X)<p$，则 $\min_i \lambda_i=0$，此时 LS 估计不存在。如果 X 的列向量之间具有近似的线性相关关系，则 $\min_i \lambda_i$ 将很小，于是 $E\left\|\hat{\beta}\right\|^2 >> \|\beta\|^2$，此时 LS 估计是不可信赖的，估计的精度很差。针对这种情况，统计学者提出了一些改进 LSE 的方法。

这些方法主要从三个方面改进 LSE，有的直接从减小估计的均方误差出发，从而降低 LSE 的病态性，如岭估计和 Stein 估计，有的从消除多重共线性出发进而改进 LSE 估计，如主成份估计，笔者曾探讨了该方法的一些应用问题，见[54]。还有的从消除异常值的影响出发改进 LSE，造成 LSE 不好的一种原因是数据中可能混有异常值，如果能从数据中找出异常值，则可将异常值去掉，重新建立模型。但问题是要发现这种异常值往往很困难，因而人们就想方设法在保留异常值的同时减少它们对回归系数估计的影响。LSE 受异常值影响较大的主要原因是选用了平方函数 X^2，如果选用随 $|X|$ 增长要比 X^2 来得慢的函数 $p(X)$ 去代替 X^2，以限制异常值对回归系数估计的影响，这就是目前用得较多的 M-估计，这方面的研究还不十分成熟，M-估计的研究已成为统计学研究的前沿和热点。

第三章 线性模型中的M-方法

§3.1 基本概念和性质

理论研究和实践经验表明，线性回归中最常用的方法---最小二乘法，在一些情况下表现不理想。近几十年来，统计学家提出了许多替代方法供选择使用。M-方法就是其中之一，也是最受重视的一种。

定义 1 对于线性模型 $y = X\beta + e$，即 $y_i = x_i'\beta + e_i$（1≤i≤n），选定一个定义于 R^1 上的函数 ρ，令

$$H_\rho(\beta) = \sum_{i=1}^{n} \rho(y_i - x_i'\beta) \qquad (\beta \in R^p)$$

$H_\rho(\beta)$ 的一个最小值点 $\hat{\beta}_n$ 称为 β 的 M-估计，

$$H_\rho(\hat{\beta}_n) = \min\{H_\rho(\beta) : \beta \in R^p\}$$

注：此处并为要求 $\hat{\beta}_n$ 有唯一性，事实上，$\hat{\beta}_n$ 可以不唯一。

这种估计是 Huber 最先关于位置参数模型

$$y_i = \alpha_0 + e_i \qquad i=1, \ldots, n$$

于 1964 年引进的。1973 年，Huber 又将这种估计拓展到一般的线性模型 $y = X\beta + e$。自那以后，这一方向逐渐成为统计学的热点。

M-估计的命名，与它同极大似然估计（Maximum Likelihood Estimate,缩写为 MLE）的联系有关。设在模型 $y_i = x_i'\beta + e_i$（1≤i≤n）中，误差 $e_1, e_2, \ldots, e_n$ 独立同分布（简记为 iid.）有已知的公共密度 f，则似然函数为

$$L(\beta)=\prod_{i=1}^{n} f(y_i - x_i'\beta)$$

β 的 MLE β_n^* 使 $L(\beta)$ 达到最大。如果记 $\rho=-\log f$，则 β_n^* 就是 $\sum_{i=1}^{n}\rho(y_i - x_i'\beta)$ 的最小值点。所以 MLE 是 M-估计的一个特例，这个 M 也是取自 MLE 的首位字母。

若 f(x) 为 Laplace 分布的密度：

$$f(x)=(2\sigma)^{-1}\exp(-|x|/\sigma)$$

则由 $\rho(u)=|u|/\sigma+const$

引出的估计是最小一乘估计或称为最小绝对偏差估计（简记为 LADE）。

LADE 是 M-估计的一个重要的特例，在相当长的时间内，由于在计算及理论上的困难，进展不大。近几十年来，随着计算问题的解决，在应用上越来越受到人们的重视，在理论上也有了比较深入的进展。

在 M-估计定义中的函数 ρ 在一定限度内可以自由选择，以适应不同的要求。例如，当 $|u|\to\infty$，$\rho(u)$ 增长较慢时，$\hat{\beta}_n$ 有较好的样本稳健性，即对样本中可能混入的少量异常值有较强的抵抗力。这种对稳健性的追求正是引进 M-估计的动力。

上面我们定义了线性模型中回归系数的 M-估计，M-估计是否存在是需要回答的问题，亦即在什么条件下

$H_\rho(\hat{\beta}_n)=\min\{H_\rho(\beta):\beta\in R^p\}$ 右端的极值能够达到。

关于这个问题，我们有

定理 1 设下面三个条件成立：

（1）ρ在R^1上处处连续，

（2）存在 a,使$\rho(u)$当$u \ge a$时，非降且不恒等于一个常数；$u \le a$时，非增且不恒等于一个常数，

（3）设计矩阵$X=(x_1 \cdots x_n)$的秩为p。

则存在$\hat{\beta}_n \in R^p$，使得$H_\rho(\hat{\beta}_n)=\min\{H_\rho(\beta):\beta \in R^p\}$。

注：条件（3）是必要的，因为若 X 的秩小于p，则线性模型$y = X\beta + e$ 中β不可估，此时估计β的问题毫无意义。条件（1）和（2）只涉及到函数ρ，$y_i - x_i'\beta$可视为实际数据和估计之间的偏离，而$\rho(y_i - x_i'\beta)$可视为由这种偏离而引起的某种损失的度量。条件（1）实际上是说，损失随偏离连续变化。对于条件（2），设 a=0，则它意味着同一个方向的偏离，绝对值越大，所导致的损失也应越大。

证明：记$H_0=\inf\{H_\rho(\beta):\beta \in R^p\}$，

根据条件（2），对一切$\beta \in R^p$，有$H_\rho(\beta) \ge n\rho(a)$，故$-\infty < H_0 < \infty$。在$R^p$中找一点列$\{b_m\}$，使得

$$\lim H_\rho(b_m)=H_0$$

若 $\lim\|b_m\| \ne \infty$，则可以找到$\{b_m\}$的子列$\{b_{m'}\}$，使得$b_{m'} \to b \in R^p$，利用ρ的连续性，有$H_\rho(b)=H_0$，于是 b 达到了$H_\rho(\beta)$的最小值。因此，我们只需证明$\lim\|b_m\| \ne \infty$。

假设$\lim\|b_m\|=\infty$，则可以找出$\{b_m\}$的一个子列，不妨设此子列为$\{b_m\}$本身，使得当$m \to \infty$时，有

i $b_m/\|b_m\| \to$某个单位向量 h.

ii $x_i'b_m$有极限g_i（i=1，. . .，n），且

$g_i=-\infty$（$1 \le i \le n_1$）

$g_i=\infty$ （$n_1+1 \le i \le n_2$）

$|g_i|<\infty$ （$n_2+1 \le i \le n$）

（为满足 ii，可以适当更改 $x_1,\dots,x_n$ 下标的编号）

从而必有 $x_i'h=0 \quad (n_2+1\le i\le n)$

这是因为若对此范围内的某个 i，上式不成立，则对于该 i 有：$\|x_i'b_m\|=\|b_m\|\|x_i'b_m/\|b_m\|\|\to\infty\cdot|x_i'h|=\infty$，这与 ii 矛盾。

根据 $x_i'h=0 \quad (n_2+1\le i\le n)$ 以及 X 的秩为 p 的假定，有 $x_i'h \quad (1\le i\le n_2)$ 不能全为 0。记

$$q=\min(\rho(\infty),\rho(-\infty))-\rho(a)$$

依条件（2），有 $q>0$，因此根据 ii，可找到充分大的 m, 使得

$$\sum_{i=n_2+1}^{n}\rho(y_i-x_i'b_m)<\sum_{i=n_2+1}^{n}\rho(y_i-g_i)+\frac{q}{2}$$

$$y_i-x_i'b_m>a \qquad (1\le i\le n_1)$$

$$y_i-x_i'b_m<a \qquad (n_1+1\le i\le n_2)$$

记 $d_i=|x_i'h|/|y_i-x_i'b_m-a| \qquad (1\le i\le n_2)$

因 $x_i'h \quad (1\le i\le n_2)$ 不全为 0，故 $d\equiv\max(d_1,\dots,d_{n_2})>0$

不妨设 $d=d_1$，且 $x_1'h>0$。令 $\tilde{b}_m=b_m+\dfrac{h}{d}$，则

$$x_i'\tilde{b}_m=x_i'b_m \qquad (n_2+1\le i\le n),$$

$$y_1-x_1'\tilde{b}_m=y_1-x_1'b_m-x_1'h/d=a,$$

$$\begin{aligned}y_i-x_i'\tilde{b}_m&=y_i-x_i'b_m-x_i'h/d\\&\ge y_i-x_i'b_m-|x_i'h|/d_i\\&=y_i-x_i'b_m-|y_i-x_i'b_m-a|\\&=y_i-x_i'b_m-(y_i-x_i'b_m-a)\\&=a \qquad (2\le i\le n_1)\end{aligned}$$

同理，可以证明

$$y_i - x_i' \tilde{b}_m \le a \qquad (n_1 + 1 \le i \le n_2)$$

综合上述各式，并利用条件（2）得到：

$$H_\rho(\tilde{b}_m) \le \rho(a) + (n_1 - 1)\rho(\infty) + (n_2 - n_1)\rho(-\infty) + \sum_{i=n_2+1}^{n} \rho(y_i - x_i' b_m)$$

$$\le n_1 \rho(\infty) + (n_2 - n_1)\rho(-\infty) + \sum_{i=n_2+1}^{n} \rho(y_i - g_i) - \frac{q}{2}$$

但有 ρ 的连续性和 $\{b_m\}$ 所满足的条件 ii，以及 $\lim H_\rho(b_m) = H_0$ 知：上式右端前三项之和即为 H_0，于是由上式得到矛盾的结果 $H_\rho(\tilde{b}_m) < H_0$，这证明了 $\lim\|b_m\| \ne \infty$，从而定理成立。

§3.2 M-估计的渐近正态性

考查线性模型 $y = X\beta + e$，亦即

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

其中 e_i 独立同分布，$Ee_i = 0$，(i=1,...,n)。本节我们重点讨论 M-估计中 LADE 的渐近正态性。

$$\beta=\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_p\end{pmatrix}$$ 的 LADE 估计定义为使

$$H_p(\beta)=\sum_{i=1}^{n}\left|y_i-x_{i1}\beta_1-\cdots-x_{ip}\beta_p\right|$$ 取最小值的点 $\hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_p\end{pmatrix}$，

对于 LADE 的渐近正态性，目前已经有了比较好的结果，如

定理 1 对于线性模型 $y=X\beta+e$ 亦即 $y_i=x_i'\beta+e_i$，$Ee_i=0$，e_i 独立同分布，其分布函数为 F(x)，F 在 0 点处有导数 $f(0)>0$，且 $\exists n_0$，

$n\geq n_0$ 时，$S_n\hat{=}\sum_{i=1}^{n}x_ix_i'$ 为 p 阶正定矩阵，当 $n\to\infty$ 时，

$$d_n\hat{=}\max_i(x_i'S_n^{-1}x_i)\to 0$$

则 $n\to\infty$ 时，$2f(0)S_n^{\frac{1}{2}}(\hat{\beta}_n-\beta)$ 依分布收敛于 $N(0,I_p)$。

在讨论 LADE 的渐近正态性时，我们一般假定误差的分布函数在 0 点处有正的导数。如果该分布函数在 0 点处没有正的导数，LADE 的渐近正态性又是怎样的呢？下面我们将讨论这一问题。

设 $\{a_n\}$ 为正的单调趋于无穷大的常数列，

$$\Psi_n(t)\hat{=}\int_0^t a_n(F(\frac{s}{a_n})-F(0))ds,$$

易见对任意的 n，$\Psi_n(t)$ 为凸函数。若对任意的 t，$\Psi_n(t)$ 的极限存在，则定义 $\Psi(t)=\lim_{n\to\infty}\Psi_n(t)$，显然它在 $[0,\infty]$ 中取值。

若 $\lim_{n\to\infty} a_n (F(\frac{t}{a_n}) - F(0)) = \varphi(t)$， (*)

则 $\Psi(t) = \int_0^t \varphi(s)ds$。

若 F(x) 在 0 点处有正的导数，可取 $a_n = \sqrt{n}$，则 $\varphi(t) = f(0)t$，

$\Psi(t) = \frac{f(0)t^2}{2}$。可见条件（*）比 F(x) 在 0 点处有正的导数更弱。

下面我们讨论 $a_n(\hat{\beta}_n - \beta)$ 的收敛性，首先定义函数

$$Z_n(u) = \sum_{i=1}^{n} [\left| e_i - \frac{x_i' u}{a_n} \right| - |e_i|]$$

由 $H_p(\beta) = \sum_{i=1}^{n} \left| y_i - x_{i1}\beta_1 - \cdots - x_{ip}\beta_p \right|$ 知，$Z_n(u)$ 的最小值点为

$a_n(\hat{\beta}_n - \beta)$。

引理 1 若 $\{Z_n\}$ 是 R^p 上凸函数列（如上定义），$Z_n(u)$ 依分布收敛于 $Z(u)$，$Z(u)$ 有唯一的最小值点 U，则当 $n \to \infty$ 时，

$U_n = a_n(\hat{\beta}_n - \beta)$ 依分布收敛于 U。（参见[18]）

为给出 LADE 的渐近分布，我们先提出一些基本条件。

（1）e_i 独立同分布，其分布函数为 F(x) 在 0 点处连续；

（2）存在正定矩阵 C，使得 $\lim_{n\to\infty} \frac{1}{a_n^2} X_n' X_n = C$；

（3）对任意的 u，$\lim_{n\to\infty} \frac{1}{a_n^2} \sum_{i=1}^{n} \Psi_n(u' x_i) = \zeta(u)$，其中 $\zeta(u)$ 是取值于 $[0,\infty]$ 的凸函数。

定理 2 对于线性模型 $y = X\beta + e$ 亦即 $y_i = x_i'\beta + e_i$，若条件（1）、（2）、（3）成立，则 $Z_n(u) \xrightarrow{d} Z(u) = -u'W + 2\zeta(u)$（依分布收敛），其中 W 是一个均值为 0，协方差为矩阵 C 的 p 维正态随机变量。

证明：易知当 $x \neq 0$ 时，有等式

$$|x-y| - |x| = -y[I(x>0) - I(x<0)] + 2\int_0^y [I(x \le s) - I(x \le 0)]ds$$

其中 $I(A)$ 表示集合 A 的示性函数。从而我们可将 $Z_n(u)$ 写成两部分，

$$Z_n(u) = Z_n^{(1)}(u) + Z_n^{(2)}(u)$$

其中 $Z_n^{(1)}(u) = -\frac{1}{a_n}\sum_{i=1}^{n} x_i'u[I(e_i>0) - I(e_i<0)]$

$$Z_n^{(2)} = 2\sum_{i=1}^{n}\int_0^{x_i'u/a_n}[I(e_i \le s) - I(e_i \le 0)]ds$$

$$= 2\sum_{i=1}^{n}\int_0^{v_{ni}}[I(e_i \le s) - I(e_i \le 0)]ds$$

$$\hat{=} \sum_{i=1}^{n} Z_{ni}^{(2)}(u)$$

其中 $v_{ni} = x_i'u/a_n$

利用 Lindeberg-Feller 中心极限定理，根据条件（2），则对于任意的 u，$Z_n^{(1)}(u) \xrightarrow{d} -u'W$。

另外，我们有

$$Z_n^{(2)}(u) = \sum_{i=1}^{n} E(Z_{ni}^{(2)}(u)) + \sum_{i=1}^{n}(Z_{ni}^{(2)}(u) - E(Z_{ni}^{(2)}(u)))$$

令 $v_i = x_i'u = a_n v_{ni}$，则有

$$\sum_{i=1}^{n} E(Z_{ni}^{(2)}(u)) = 2\sum_{i=1}^{n}\int_0^{v_{ni}}(F(s)-F(0))ds$$

$$= \frac{2}{a_n^2}\sum_{i=1}^{n}\int_0^{v_i} a_n(F(\frac{s}{a_n})-F(0))ds$$

$$= \frac{2}{a_n^2}\sum_{i=1}^{n}\Psi_n(u' x_i)$$

$$\to 2\zeta(u)$$

下面考查 $Z_n^{(2)}(u)$ 的第二部分，

$$D(Z_n^{(2)}(u)) = \sum_{i=1}^{n} E[(Z_{ni}^{(2)}(u)-E(Z_{ni}^{(2)}(u)))^2]$$

$$= \sum_{i=1}^{n}\{E[Z_{ni}^{(2)}(u)]^2 - [E(Z_{ni}^{(2)}(u))]^2\}$$

$$\le \sum_{i=1}^{n} E[Z_{ni}^{(2)}(u)]^2$$

$$\le \frac{2}{a_n}\max_i |x_i' u| \sum_{i=1}^{n} E(Z_{ni}^{(2)}(u))$$

$$= \frac{2}{a_n}\max_i |x_i' u| E(Z_n^{(2)}(u))$$

因此，当 $\zeta(u) < \infty$ 时，$Z_n^{(2)}(u) - E(Z_n^{(2)}(u)) \xrightarrow{p} 0, (n \to \infty)$

从而 $Z_n^{(2)}(u) \xrightarrow{p} 2\zeta(u), (n \to \infty)$。

当 $\zeta(u) = \infty$ 时，我们有

$$P(|Z_n^{(2)}(u) - E(Z_n^{(2)}(u))| > \varepsilon E(Z_n^{(2)}(u))) \le \frac{D(Z_n^{(2)}(u))}{\varepsilon^2 (EZ_n^{(2)}(u))^2}$$

$$\le \frac{2\max_i |x_i'u|/a_n}{\varepsilon^2 E(Z_n^{(2)}(u))}$$

$$\to 0$$

故 $Z_n^{(2)}(u) \xrightarrow{p} \infty = \zeta(u),(n \to \infty)$，

因此，$Z_n(u) \xrightarrow{d} -u'W + 2\zeta(u) = Z(u)$。

推论 2.1 在定理 2 的假设下，若 $Z(u)$ 以概率 1 有唯一最值点 U，

则 $a_n(\hat{\beta}_n - \beta) \xrightarrow{d} U$。

第四章 参数估计的可容许性

§4.1.基本概念和性质

对于一般模型的参数估计问题，设θ为待估参数向量，根据某种方法如最小二乘估计等导出θ的一个估计$\tilde{\theta}$。由于$\tilde{\theta}$一般和θ有一定的偏差，基于$\tilde{\theta}$的统计决策会产生一定的损失。用$L(\tilde{\theta},\theta)$记这个损失，它是θ的函数，称为**损失函数**。其平均损失$EL(\tilde{\theta},\theta)$称为$\tilde{\theta}$的风险函数，记为$R(\tilde{\theta},\theta)$。一般采用的损失函数多为θ的二次函数，如

$$L(\tilde{\theta},\theta)=(\tilde{\theta}-\theta)'D(\tilde{\theta}-\theta)\hat{=}\left\|\tilde{\theta}-\theta\right\|_D^2\quad, D\geqslant 0, \tag{4.1}$$

$$L(\tilde{\theta},\theta)=(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)' \tag{4.2}$$

后者称为矩阵损失。对（4.1），其风险函数为

$$R(\tilde{\theta},\theta)=E(\tilde{\theta}-\theta)'D(\tilde{\theta}-\theta) \tag{4.3}$$

称上式为$\tilde{\theta}$的广义均方误差（记作 GMSE($\tilde{\theta}$)）。特别地，当 D=I 时，风险函数为$R(\tilde{\theta},\theta)=E\left\|\tilde{\theta}-\theta\right\|_I^2=E\left\|\theta-\tilde{\theta}\right\|_I^2$，称为$\tilde{\theta}$的均方误差（记作 MSE($\tilde{\theta}$)）。对矩阵损失（4.2），风险函数为

$$R(\tilde{\theta},\theta)=E(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)' \tag{4.4}$$

称为$\tilde{\theta}$的均方误差矩阵（记为 MSEM($\tilde{\theta}$)）。

定义 1 设$\tilde{\theta}_1$和$\tilde{\theta}_2$为 θ 的两个估计，如果对于风险函数$R(\cdot,\cdot)$，有

（1）$R(\tilde{\theta}_1,\theta)\leqslant R(\tilde{\theta}_2,\theta)$，对一切 θ 成立；

（2）至少存在一个θ_0使不等号成立。

则称$\tilde{\theta}_1$关于风险函数$R(\cdot,\cdot)$一致优于$\tilde{\theta}_2$。若在某个估计类中，不存在一致优于$\tilde{\theta}$的估计，则称$\tilde{\theta}$在该估计类中关于风险函数$R(\cdot,\cdot)$为 θ 的可容许估计，简称为$\tilde{\theta}$为 θ 的可容许估计。否则，称$\tilde{\theta}$为 θ 的不可容许估计。

直观上看，可容许性是对一个估计的最起码的要求。因为如果一个估计是不可容许的，那么我们就能够找到另外一个更好的估计去代替它。一般说来，对一个未知参数向量，可容许估计是很多的，构成一个很大的估计类。我们可以根据各种不同的优良性指标，如无偏性、方差最小性、线性估计等等，从估计类中选择一个特殊的估计。

为记号简单，如果关于风险函数（4.3）和（4.4），$\tilde{\theta}$分别为 θ 的可容许估计，则分别记为$\tilde{\theta}\overset{D}{\sim}\theta$和$\tilde{\theta}\underset{M}{\sim}\theta$。特别地，当 D=I 时，简记为$\tilde{\theta}\sim\theta$。

定理 1 若$\tilde{\theta}\sim\theta$，则

（1）$\tilde{\theta}\underset{M}{\sim}\theta$，

（2）对一切 D≥0，$\tilde{\theta}\overset{D}{\sim}\theta$。

证明：(1)设$\tilde{\theta}\sim\theta$，但$\tilde{\theta}\underset{M}{\sim}\theta$不成立，即存在另外一个估计$\theta^*$，使得对一切θ有$E(\theta^*-\theta)(\theta^*-\theta)'\le E(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)'$，且对于某个$\theta_0$不等号成立。由于 A≥B⇒tr(A) ≥tr(B),A>B⇒ tr(A) >tr(B),在上式两端取 tr,得$E\|\theta^*-\theta\|^2\le E\|\tilde{\theta}-\theta\|^2$，且在$\theta_0$不等号成立，即$\theta^*$一致优于$\tilde{\theta}$，与$\tilde{\theta}\sim\theta$矛盾。故（1）成立。

（2）设对某个 D≥0, $\tilde{\theta}\overset{D}{\sim}\theta$不成立.则存在一个估计$\theta^*$,对一切θ,有$E\|\theta^*-\theta\|_D^2\le E\|\tilde{\theta}-\theta\|_D^2$，且在某个$\theta_0$处不等号成立.设$\lambda$为 D 的最大特征值，记$F=\lambda^{-1}D$，则有$E\|\theta^*-\theta\|_F^2\le E\|\tilde{\theta}-\theta\|_F^2$，对一切θ成立，且在$\theta_0$处不等号成立.作新的估计 $\theta^{**}=\tilde{\theta}+F(\theta^*-\tilde{\theta})$

其风险函数为

$$E\|\theta^{**}-\theta\|^2=E\|\tilde{\theta}-\theta\|^2+E\|\theta^*-\tilde{\theta}\|_{F^2}^2+E(\tilde{\theta}-\theta)'F(\theta^*-\tilde{\theta})$$

$$+E(\theta^*-\tilde{\theta})'F(\tilde{\theta}-\theta)$$

因$F\le I,F^2\le F$，所以$E\|\theta^*-\tilde{\theta}\|_{F^2}^2\le E\|\theta^*-\tilde{\theta}\|_F^2$。

注意到

$$E\|\theta^*-\tilde{\theta}\|_F^2+E(\tilde{\theta}-\theta)'F(\theta^*-\tilde{\theta})$$

$$=E(\theta^*-\tilde{\theta})'F(\theta^*-\tilde{\theta})+E(\tilde{\theta}-\theta)'F(\theta^*-\tilde{\theta})$$

$$=E(\theta^*-\theta)'F(\theta^*-\tilde{\theta})$$

从而风险函数

$$E\|\theta^{**}-\theta\|^2\leq E\|\tilde{\theta}-\theta\|^2+E(\theta^*-\theta)'F(\theta^*-\tilde{\theta})$$

$$+E(\theta^*-\tilde{\theta})'F(\tilde{\theta}-\theta)$$

$$=E\|\tilde{\theta}-\theta\|^2+E(\theta^*-\theta)'F(\theta^*-\theta)-E(\tilde{\theta}-\theta)'F(\tilde{\theta}-\theta)$$

$$=E\|\tilde{\theta}-\theta\|^2+(E\|\theta^*-\theta\|_F^2-E\|\tilde{\theta}-\theta\|_F^2)$$

$$\leq E\|\tilde{\theta}-\theta\|^2,$$

且在 θ_0 处不等号成立，即 θ^{**} 一致优于 $\tilde{\theta}$，这与 $\tilde{\theta}\sim\theta$ 矛盾，从而(2)成立。

另外，若对一切 D≥0， $\tilde{\theta}\overset{D}{\sim}\theta$，则必有 $\tilde{\theta}\sim\theta$. 所以我们重点讨论均方误差意义下的可容许估计。

定理 2 对于矩阵 A， $\tilde{\theta}\sim\theta\Rightarrow A\tilde{\theta}\sim A\theta$。当 A 可逆时，其逆亦真。

证明： 设 $\tilde{\theta}\sim\theta$，由定理 1，有 $\tilde{\theta}\overset{A'A}{\sim}\theta$，下面证明，$\tilde{\theta}\overset{A'A}{\sim}\theta\Rightarrow A\tilde{\theta}\sim A\theta$。用反证法，若 $A\tilde{\theta}\sim A\theta$ 不成立，则存在估计 θ^* 一致优于 $A\tilde{\theta}$。于是，对一切 θ 有

$$E\|\theta^*-A\theta\|^2\leq E\|A\tilde{\theta}-A\theta\|^2$$

且对某个θ_0处不等号成立。记$P_A = A(A'A)^- A'$，注意到

$(I-P_A)P_A = 0, (I-P_A)A = 0$，所以

$$\|\theta^* - A\theta\|^2 = \|\theta^* - P_A\theta^* + P_A\theta^* - A\theta\|^2$$

$$= \|\theta^* - P_A\theta^*\|^2 + \|P_A\theta^* - A\theta\|^2$$

从而我们有，

$$E\|P_A\theta^* - A\theta\|^2 \le E\|\theta^* - A\theta\|^2$$

$$\le E\left\|A\tilde{\theta} - A\theta\right\|^2$$

$$= E\left\|\tilde{\theta} - \theta\right\|^2_{A'A}$$

对一切θ成立，且在θ_0处不等号成立。若记$\theta^{**} = (A'A)^- A'\theta^*$，

则有$P_A\theta^* = A\theta^{**}$，于是，上式就变为

$$E\|A\theta^{**} - A\theta\|^2 \le E\left\|\tilde{\theta} - \theta\right\|^2_{A'A}$$

即对一切θ有，$$E\|\theta^{**} - \theta\|^2_{A'A} \le E\left\|\tilde{\theta} - \theta\right\|^2_{A'A},$$

且不等号在θ_0处成立，这表明θ^{**}一致优于$\tilde{\theta}$，与$\tilde{\theta} \overset{A'A}{\sim} \theta$相矛盾，从而

证明了$A\tilde{\theta} \sim A\theta$。

现在证明另外一个结论，A 可逆，依据定理 1，我们有

$$A\tilde{\theta} \sim A\theta \Rightarrow A\tilde{\theta} \overset{(A^{-1})'A^{-1}}{\sim} A\theta$$

根据前一结论证明过程知，

$A\tilde{\theta} \overset{(A^{-1})'A^{-1}}{\sim} A\theta \Rightarrow \tilde{\theta} \sim \theta$，结论成立。

考虑线性模型 $y = X\beta + e$，记 $\mu = X\beta$，对于 $\mu = X\beta$ 在线性估计类中关于均方误差的可容许估计，我们有下述主要结果：

定理 3 对于线性模型 $y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 I$，$Ay \sim \mu = X\beta$ 的充要条件是 A 对称，且其特征值都在[0, 1]中。

为证明定理 3，先给出一个重要的引理

引理 1 若 y 为 n 维随机变量，A 为 n 阶对称矩阵，则有

$$Ey'Ay = tr[ACov(y)] + Ey'AEy。$$

证明：

$$\begin{aligned} Ey'Ay &= Etr(y'Ay) = Etr(Ayy') = tr(EAyy') \\ &= tr(AEyy') = tr\{A[Cov(y) + EyEy']\} \\ &= tr[ACov(y)] + tr(AEyEy') \\ &= tr[ACov(y)] + tr(Ey'AEy) \\ &= tr[ACov(y)] + Ey'AEy \end{aligned}$$

定理 3 的证明：必要性：

设 $Ay \sim \mu$，对 $I - A$ 作奇异值分解，

$$I - A = P\Lambda Q', \qquad \Lambda = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$$

其中 $\Lambda_r = diag(\lambda_1, \cdots, \lambda_r)$， $\lambda_i > 0, i = 1, \ldots, r$， P 和 Q 为正交矩阵。

利用引理 1 有，

$$\begin{aligned} E\|Ay - \mu\|^2 &= E(Ay - \mu)'(Ay - \mu) \\ &= \mu'(I - A)'(I - A)\mu + \sigma^2 tr(AA') \\ &= \sigma^2 tr[(I - P\Lambda Q')'(I - P\Lambda Q')] + \mu' Q\Lambda^2 Q'\mu \\ &\hat{=} \sigma^2 (n - 2\sum_{i=1}^{r} \lambda_i m_i + \sum_{i=1}^{r} \lambda_i^2) + \sum_{i=1}^{r} \lambda_i^2 \varphi_i^2 \end{aligned}$$

其中 m_i 为矩阵 $P'Q$ 的 (i,i) 元， $\varphi_1, \ldots, \varphi_r$ 为 $Q'\mu$ 的前 r 个分量。由 P, Q 的

正交性，得到$|m_i| \le 1$。因$\lambda_i > 0, i=1,\ldots,r$，故从$Ay \sim \mu$知，必有$m_i = 1,\ i=1,\ldots,r$。否则，取$Ay^* = (I - Q\Lambda Q')y$作为$\mu$的估计，可以使均方误差更小，与$Ay \sim \mu$矛盾。由 Cauchy 不等式知，此时 P, Q 的前r列相同，又因它们后$n-r$列的选取不改变 A，故可以取$P = Q$，
从而$A = I - P\Lambda P'$对称。

A 的特征值为$1-\lambda_1,\ldots,1-\lambda_r,1,\ldots,1$，问题归结为证明
$\lambda_i \le 1$，$i=1,\ldots,r$，将$m_i = 1,\ i=1,\ldots,r$代入Ay的均方误差公式得到

$$E\|Ay-\mu\|^2 = \sigma^2[n - r + \sum_{i=1}^{r}(\lambda_i - 1)^2] + \sum_{i=1}^{r}\lambda_i^2\varphi_i^2$$

从而$\lambda_i \le 1$，$i=1,\ldots,r$，否则，将λ_i中大于 1 的λ_i换成 1 得到Λ^*，用
$A^*y = (I - P\Lambda^* P')y$去估计$\mu$，有更小的均方误差，这和$Ay \sim \mu$矛盾。

充分性：

设 A 对称，则存在正交矩阵 Q，使得
$A = Q\Lambda Q'$，$\Lambda = diag(\lambda_1,\ldots,\lambda_r,0,\ldots,0)$，$\lambda_i \in [0,1], i=1,\ldots,r$
由引理 1 知，$E\|Ay-\mu\|^2 = \sigma^2 tr\Lambda^2 + \mu' Q(I-\Lambda)^2 Q'\mu$

$$\hat{=} \sigma^2 tr\Lambda^2 + \theta'(I-\Lambda)^2\theta$$

其中$\theta = Q'\mu$。设By为μ的任一线性估计，由必要性知，不妨设 B 对称且特征值都在[0, 1]中，故$B = P\Delta P'$，其中P为正交矩阵，Δ为对称矩阵，且$0 \le \Delta \le I$。

$$E\|By-\mu\|^2 = \sigma^2 tr\Delta^2 + \mu' P(I-\Delta)^2 P'\mu$$
$$= \sigma^2 tr\Delta^2 + \theta' Q' P(I-\Delta)^2 P' Q\theta$$

如果$E\|By-\mu\|^2 \le E\|Ay-\mu\|^2$，则

$$\sigma^2 tr\Delta^2 + \theta' Q' P(I-\Delta)^2 P' Q\theta \le \sigma^2 tr\Lambda^2 + \theta'(I-\Lambda)^2\theta$$

特别地，首先取$\theta = 0, \sigma^2 = 1$，得到

$$tr\Delta^2 \le tr\Lambda^2$$

又令$\sigma^2 \to 0$，并由θ的任意性得到

$$Q'P(I-\Delta)^2P'Q \le (I-\Lambda)^2$$

根据第一章§1.2.2中的定理1知，上面二不等式均为等式，从而$\sigma^2 tr\Delta^2 + \theta'Q'P(I-\Delta)^2P'Q\theta = \sigma^2 tr\Lambda^2 + \theta'(I-\Lambda)^2\theta$，

即By不优于Ay。

定理4 对于线性模型$y = X\beta + e$，$Ee = 0, \text{cov}(e) = \sigma^2 I$，

$$l'y \sim c'\mu \Leftrightarrow l'l \le l'c。$$

证明：充分性：设$l'l \le l'c$，取$M = ll'/l'c$，显然它是一个对称方阵，满足$l' = c'M$，且唯一的特征值满足$0 < \frac{l'l}{l'c} \le 1$。根据定理3有$My \sim \mu$，再由定理2知，$c'My = l'y \sim c'\mu$。

必要性：

设$l'y \sim c'\mu$，$E(l'y - c'\mu)^2 = \sigma^2 l'l + [(c-l)'\mu]^2$

令$b = c + \alpha(l-c), 0 < \alpha < 1$

$$E(b'y - c'\mu)^2 = \sigma^2[(1-\alpha)^2 c'c + \alpha^2 l'l + 2\alpha(1-\alpha)l'c] + \alpha^2[(c-l)'\mu]^2$$

由$0 < \alpha < 1$和$l'y \sim c'\mu$知，

$$l'l \le (1-\alpha)^2 c'c + \alpha^2 l'l + 2\alpha(1-\alpha)l'c$$

即$(1+\alpha)l'l \le (1-\alpha)c'c + 2\alpha l'c$，

令$\alpha \to 1$，即得结论。

定理5 对于线性模型$y = X\beta + e$，$Ee = 0, \text{cov}(e) = \sigma^2 I$，

$$Ay \sim B\mu \Leftrightarrow AA' \le AB'。$$

证明：必要性：设$Ay \sim B\mu$，根据定理2知，对于任意的向量l，$l'Ay \sim l'B\mu$，再由定理4知，$l'AA'l \le l'AB'l$，故只需证明AB'的对称性。若AB'不对称，则$(B-A)B'$也不对称。根据第一章§1.2.2中的定理2知，存在正交矩阵P，使得

$$tr(P(B-A)B') > tr(B-A)B'$$

取 $C = B - P(B - A)$，易证 $E\|Cy - B\mu\|^2 < E\|Ay - B\mu\|^2$，矛盾。从而 AB' 对称。

充分性：

对矩阵 A 作奇异值分解

$$A = P\Lambda Q' \quad , \qquad \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$$

其中 P, Q 为正交矩阵，$\Lambda_1 = diag(\lambda_1, \dots, \lambda_r) > 0$。

由 $AA' \le AB'$ 得到 $P\Lambda^2 P' \le P\Lambda Q' B' PP'$

亦即 $\Lambda^2 \le \Lambda Q' B' P \hat{=} \Lambda T \hat{=} \Lambda\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$，$T_{11}$ 为 r 阶方阵，$B = PT'Q'$。

由 AB' 对称知 ΛT 对称，从而 $T_{12} = 0$，于是

$$\Lambda T = \begin{pmatrix} \Lambda_1 T_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

从而 $\Lambda_1 T_{11}$ 对称，且 $0 < \Lambda_1^2 \le \Lambda_1 T_{11}$，于是 T_{11} 可逆，

$$M_{11} \hat{=} (T_{11}^{-1})' \Lambda_1 = \Lambda_1 (T_{11}' \Lambda_1)^{-1} \Lambda_1 \le I$$

即 M_{11} 对称且特征值都在[0, 1]中。取

$$M = Q\begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix} Q'$$

显然，M 的特征值都在[0, 1]中，根据定理 3 知，$My \sim \mu$，又易证 $A = BM$，根据定理 2 知，$Ay = BMy \sim B\mu$。

推论 5.1 对于线性模型 $y = X\beta + e$，$Ee = 0, \mathrm{cov}(e) = \sigma^2 I$，

$Ay \sim A\mu$。

定理 6 对于线性模型 $y = X\beta + e$，$Ee = 0, \mathrm{cov}(e) = \sigma^2 V$，有

（1）$Ay \sim \mu \Leftrightarrow AV$ 对称，且 A 的特征根都在[0, 1]中；

（2）$l'y \sim c'\mu \Leftrightarrow l'Vl \le l'Vc$；

（3） $Ay \sim B\mu \Leftrightarrow AVA' \le AVB'$

只需用$V^{-\frac{1}{2}}$去乘模型即化为上面已讨论的情形。

§4.2 可估函数估计的可容许性

考查线性模型 $y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 V$， $V > 0$的若干个可估函数的可容许估计。设 C 为$m \times p$矩阵，$\mu(C') \subset \mu(X')$，即$C\beta$为 m 个可估函数。存在$m \times n$矩阵 D，使得$C = DX$。

引理 1 对于线性模型$y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 V$， $V > 0$，设Ay为可估函数$C\beta$的任一估计，则

$$E\|Ay - C\beta\|^2 \ge E\|AX\beta^* - C\beta\|^2$$

其中$\beta^* = (X'V^{-1}X)^- X'V^{-1}y$。

证明： $E\|Ay - C\beta\|^2 = E[(Ay - C\beta)'(Ay - C\beta)]$

$= E[(Ay - AX\beta^* + AX\beta^* - C\beta)'(Ay - AX\beta^* + AX\beta^* - C\beta)]$

$= E\|AX\beta^* - C\beta\|^2 + E\|Ay - AX\beta^*\|^2 + 2E[(AX\beta^* - C\beta)'(Ay - AX\beta^*)]$

由于$E(Ay - AX\beta^*) = A[Ey - E(X\beta^*)] = 0$

$E[(AX\beta^* - C\beta)'(Ay - AX\beta^*)] = E[(AX\beta^*)'A(y - X\beta^*)]$

$= tr\{A'AE[(y - X\beta^*)(X\beta^*)']\}$

注意到$X\beta^*$与残差向量$y - X\beta^*$不相关，故上式为 0。从而

$$E\|Ay-C\beta\|^2=E\|AX\beta^*-C\beta\|^2+E\|Ay-AX\beta^*\|^2 \qquad (**)$$

$$\geq E\|AX\beta^*-C\beta\|^2$$

引理 1 说明，在研究可估函数$C\beta$的可容许估计时，只需考虑形如$AX\beta^*$的线性估计类就可以了。

定理 1 对于线性模型$y=X\beta+e$，$Ee=0,\mathrm{cov}(e)=\sigma^2V$，$V>0$，设$C\beta$可估，则$Ay\sim C\beta \Leftrightarrow$（1）$AV=AX(X'V^{-1}X)^-X'$，

（2）$AVA'\leq AX(X'V^{-1}X)^-C'$。

证明：先证条件（1）的必要性。由（**）知，若$Ay\sim C\beta$，则$E\|AX\beta^*-C\beta\|^2=0$。此时有$Ay=AX\beta^*$，即$Ay=AX(X'V^{-1}X)^-X'V^{-1}y$对一切 y 成立。故$A=AX(X'V^{-1}X)^-X'V^{-1}$，（1）的必要性得证。因此我们可以仅考虑满足（1）的线性估计Ay。

因$C\beta$可估，存在矩阵 D，使得$C=DX$，于是$C\beta=DX\beta=D\mu$

对于满足（1）的 A，根据§4.1 定理 6，我们有

$$Ay\sim C\beta \Leftrightarrow Ay=AX(X'V^{-1}X)^-X'V^{-1}y\sim D\mu$$

$$\Leftrightarrow AX(X'V^{-1}X)^-X'V^{-1}X(X'V^{-1}X)^-X'A'\leq AX(X'V^{-1}X)^-C'$$

因为$X'V^{-1}X(X'V^{-1}X)^-X'=X'$，并利用(1)，所以，上式

$$\Leftrightarrow AX(X'V^{-1}X)^-X'A'\leq AX(X'V^{-1}X)^-C'$$

$$\Leftrightarrow AVA'\leq AX(X'V^{-1}X)^-C'。$$

特别地，当$R(X_{n\times p})=p$时，我们有

推论 1.1 对于线性模型$y=X\beta+e$，$Ee=0,\mathrm{cov}(e)=\sigma^2V$，$V>0$，若$R(X_{n\times p})=p$，则

(1) $A\beta^*\sim C\beta \Leftrightarrow A(X'V^{-1}X)^{-1}A'\leq A(X'V^{-1}X)^{-1}C'$，

(2) $C\beta^*\sim C\beta$，

(3) $\beta^* \sim \beta$。

更进一步，若假设 V=I，记 $\hat{\beta} = (X'X)^{-1}X'y$，则有

推论 1.2 对于线性模型 $y = X\beta + e$， $Ee = 0, \operatorname{cov}(e) = \sigma^2 I$，若 $R(X_{n\times p}) = p$，则有：

(1) $A\hat{\beta} \sim C\beta \Leftrightarrow A(X'X)^{-1}A' \le A(X'X)^{-1}C'$，

(2) $C\hat{\beta} \sim C\beta$，

(3) $\hat{\beta} \sim \beta$。

§4.3 不可估函数的线性估计的可容许性

考查线性模型

$$y = X\beta + e, \quad Ee = 0, \operatorname{cov}(e) = \sigma^2 V, \quad V > 0, \tag{1}$$

设 $C\beta$ 为待估的参数函数，C 为 $k \times p$ 矩阵。本节我们在线性估计类中考查不可估函数 $C\beta$ 估计的可容许性。

引理 1 $C\beta$ 可估的充要条件是 $\mu(C') \subset \mu(X')$，即存在矩阵 $B_{n\times k}$，使得 $C' = X'B$，亦即 $C = B'X$。

引理 2 线性模型(1)存在不可估函数 $C\beta$ 的充要条件是 rk(X)<P。

引理 3 若 y 为 n 维随机变量，A 为 n 阶对称矩阵，则有

$Ey'Ay = tr[ACov(y)] + Ey'AEy$。（参见 § 4.1 中引理 1）

引理 4 若矩阵$\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix}$是非负定的，则 $B=0$。

证明：矩阵$\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix}$是非负定，故存在矩阵 $F=\begin{pmatrix} F_1 & F_2 \\ F_3 & F_4 \end{pmatrix}$，使得

$$\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix} = F'F$$

从而 $F_2 = F_4 = 0$， $B=0$。

注：也可如下证明，若 $B \neq 0$，则存在 $x_1 \neq 0$，使得 $B'x_1 \neq 0$。令 $x_2 = -cB'x_1$， $c>0$ 充分大，则

$(x_1', x_2')\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1'Ax_2 - 2c(B'x_1)'B'x_1 < 0$，这和$\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix}$是非负定相矛盾。

我们首先考虑一种特殊情形，$X=\begin{pmatrix} \Lambda_m & 0 \\ 0 & 0 \end{pmatrix}_{n\times p}$ (m<p), V=I, 其中 $\Lambda_m = diag(\lambda_1, ..., \lambda_m) > 0$ 。 并 记 $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{m\times 1}_{(p-m)\times 1}$ ， $A=(A_1 \quad A_2)$ ， $C=(C_1 \quad C_2)$，其中 A_1, C_1 为 $k\times m$ 矩阵， A_2 为 $k\times(n-m)$ 矩阵， C_2 为 $k\times(p-m)$ 矩阵。风险函数为

$$\begin{aligned} R(Ay, C\beta) &= E(Ay - C\beta)'(Ay - C\beta) \\ &= \sigma^2 tr(A'A) + \beta'(AX-C)'(AX-C)\beta \\ &= \sigma^2[tr(A_1'A_1) + tr(A_2'A_2)] + \beta'(AX-C)'(AX-C)\beta \\ &= \sigma^2[tr(A_1'A_1) + tr(A_2'A_2)] + \beta'\begin{pmatrix} (A_1\Lambda_m - C_1)' \\ -C_2' \end{pmatrix}(A_1\Lambda_m - C_1, \quad -C_2)\beta \end{aligned}$$

命题 1 $C\beta$ 可估的充要条件是 $C_2 = 0$。

命题 2 对$\forall\beta\in R^p$，$R[(A_1,0)y,C\beta]\le R(Ay,C\beta)$，并且等号成立当且仅当$A_2=0$。

证明：$R[(A_1,0)y,C\beta]\le R(Ay,C\beta)$

$$\Leftrightarrow tr(A_2'A_2)=0\Leftrightarrow A_2=0$$

命题 3 若$C_2'(B_1-A_1)\ne 0$，则$(B_1,0)y$不可能优于$(A_1,0)y$，若$C_2'(B_1-A_1)=0$，则$(B_1,0)y$优于$(A_1,0)y$的充要条件是

$tr(B_1'B_1)\le tr(A_1'A_1)$，

$(B_1\Lambda_m-C_1)'(B_1\Lambda_m-C_1)\le(A_1\Lambda_m-C_1)'(A_1\Lambda_m-C_1)$，

且至少有一个在某β_0处不等号成立。

证明：$R[(A_1,0)y,C\beta]$

$$=\sigma^2 tr(A_1'A_1)+\beta'\begin{pmatrix}(A_1\Lambda_m-C_1)'\\-C_2'\end{pmatrix}(A_1\Lambda_m-C_1,\quad -C_2)\beta$$

$$=\sigma^2 tr(A_1'A_1)+\beta'\begin{pmatrix}(A_1\Lambda_m-C_1)'(A_1\Lambda_m-C_1) & -(A_1\Lambda_m-C_1)'C_2\\-C_2'(A_1\Lambda_m-C_1) & C_2'C_2\end{pmatrix}\beta$$

若$C_2'(B_1-A_1)\ne 0$，则$C_2'(B_1-C_1)\ne C_2'(A_1-C_1)$，根据引理 4 知，$R[(B_1,0)y,C\beta]\le R[(A_1,0)y,C\beta]$不可能对任意的$\beta$成立，即$(B_1,0)y$不可能优于$(A_1,0)y$。反之，若$C_2'(B_1-A_1)=0$，则

$C_2'(B_1-C_1)=C_2'(A_1-C_1)$，从而$(B_1,0)y$优于$(A_1,0)y$等价于

$tr(B_1'B_1)\le tr(A_1'A_1)$，

$(B_1\Lambda_m-C_1)'(B_1\Lambda_m-C_1)\le(A_1\Lambda_m-C_1)'(A_1\Lambda_m-C_1)$，

且至少有一个在某β_0处不等号成立。

定理 1 若C_2行满秩，则$Ay\sim C\beta$的充要条件是$A_2=0$。

证明：必要性有命题 2 即得。

充分性：若$(A_1,0)y\sim C\beta$不成立，则存在By优于$(A_1,0)y$，由命题 2 知，$(B_1,0)y$必优于$(A_1,0)y$，再由命题 3 知，$C_2'(B_1-A_1)=0$，因C_2行满秩，故$B_1=A_1$，这和$(B_1,0)y$必优于$(A_1,0)y$矛盾。从而

$$(A_1,0)y \sim C\beta 。$$

推论 1.1 若 $a=\begin{pmatrix}a_1\\a_2\end{pmatrix}^{m\times1}_{(p-m)\times1}$，$b=\begin{pmatrix}b_1\\b_2\end{pmatrix}^{m\times1}_{(n-m)\times1}$，$b'y\sim a'\beta(a_2\neq0,a'\beta$ 不可估）的充要条件是 $b_2=0$。

定理 2 $Ay\sim\beta$ 的充要条件是（1）$A_2=0$；（2）记 A 的前 m 行 m 列元素构成的矩阵为 A_{11}，对任意的 m 阶方阵 B_{11}，下述结论不成立：

$tr(B_{11}B_{11}')\le tr(A_{11}A_{11}')$ 和

$(B_{11}\Lambda_m-I_m)'(B_{11}\Lambda_m-I_m)\le(A_{11}\Lambda_m-I_m)'(A_{11}\Lambda_m-I_m)$

且上述两式至少有一对某参数 β_0 不等号成立。

证明： 由命题 2 知，$Ay\sim\beta$ 当且仅当 $A_2=0$，此时 $C=I_p$，$C_1=\begin{pmatrix}I_m\\0\end{pmatrix}$，$C_2=\begin{pmatrix}0\\I_{p-m}\end{pmatrix}$，记 $A_1=\begin{pmatrix}A_{11}\\A_{21}\end{pmatrix}$，$B_1=\begin{pmatrix}B_{11}\\B_{21}\end{pmatrix}$。由命题 3 知，若 $B_{21}\neq A_{21}$，则 $(B_1,0)y$ 不可能优于 $(A_1,0)y$；若 $B_{21}=A_{21}$，则 $(B_1,0)y$ 优于 $(A_1,0)y$ 的充要条件是

$tr(B_{11}B_{11}')\le tr(A_{11}A_{11}')$ 和

$(B_{11}\Lambda_m-I_m)'(B_{11}\Lambda_m-I_m)\le(A_{11}\Lambda_m-I_m)'(A_{11}\Lambda_m-I_m)$

且上述两式至少有一对某参数 β_0 不等号成立。

从而 $(A_1,0)y\sim\beta$ 等价于对任意的 B_{11}，上述结论不成立。

上面我们讨论了一类特殊的设计矩阵，下面我们考虑一般的情形的不可估函数的可容许性问题。

对线性模型 $y=X\beta+e$，$Ee=0,\mathrm{cov}(e)=\sigma^2V$，$V>0$，作线性变换，得到 $V^{-\frac{1}{2}}y=V^{-\frac{1}{2}}X\beta+V^{-\frac{1}{2}}e$

由矩阵的奇异值分解知，存在正交矩阵 P, Q, 使得

$V^{-\frac{1}{2}}y=P\begin{pmatrix}\Lambda_m & 0\\ 0 & 0\end{pmatrix}Q\beta+V^{-\frac{1}{2}}e$，此处 $\Lambda_m=diag(\lambda_1,\dots,\lambda_m)>0$，

m=rk(X)<p， $\lambda_1,\dots,\lambda_m$ 为 $V^{-\frac{1}{2}}X$ 的奇异值。

记 $Z=P'V^{-\frac{1}{2}}y$， $\beta^*=Q\beta$， $e^*=P'V^{-\frac{1}{2}}e$，则

$y=X\beta+e$， $Ee=0,\mathrm{cov}(e)=\sigma^2V$， $V>0$，　　变为

$Z=\begin{pmatrix}\Lambda_m & 0\\ 0 & 0\end{pmatrix}\beta^*+e^*$， $Ee^*=0,\mathrm{cov}(e^*)=\sigma^2I$

此即上面我们所考虑的情形。

定理 3 设 $CQ'=(C_1^*,C_2^*)$， C_1^* 为 $k\times m$ 矩阵， C_2^* 为 $k\times(p-m)$ 矩阵，$C\beta$ 可估的充要条件是 $C=(C_1^*,0)Q$。

定理 4 设 $CQ'=(C_1^*,C_2^*)$， C_1^* 为 $k\times m$ 矩阵， C_2^* 为 $k\times(p-m)$ 矩阵，$AV^{\frac{1}{2}}P=(A_1^*,A_2^*)$， A_1^* 为 $k\times m$ 矩阵， A_2^* 为 $k\times(n-m)$ 矩阵，若 $rk(C_2^*)=k$，则 $Ay\sim C\beta$ 的充要条件是 $A_2^*=0$。

定理 5 设 $Qa=\begin{pmatrix}a_1^*\\ a_2^*\end{pmatrix}^{m\times 1}_{(p-m)\times 1}$， $PV^{1/2}b=\begin{pmatrix}b_1^*\\ b_2^*\end{pmatrix}^{m\times 1}_{(n-m)\times 1}$， $(a_2^*\neq 0)$，则 $b'y\sim a'\beta\Leftrightarrow b_2^*=0$。

§4.4 回归系数的可容许估计类

考查线性模型 $y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 V$， $V > 0$，对于该模型中参数的线性估计的可容许性，我们在§4.2 节进行了讨论。

定理 1 对于线性模型 $y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 V$， $V > 0$，设 $C\beta$ 可估，则 $Ay \sim C\beta \Leftrightarrow$ （1） $AV = AX(X'V^{-1}X)^{-}X'$，

（2） $AVA' \le AX(X'V^{-1}X)^{-}C'$。

本节我们讨论可估函数 $C\beta$ 的全体线性可容许估计的显式表达式。

引理 1 设 L 和 S 都是 $m \times n$ 矩阵，则 $LL' \le LS'$（此式包含要求 LS' 为对称方阵）的充要条件是存在方阵 $M \ge 0$，M 的特征值都在[0，1]中，使得 $L = SM$，而 $rk(L) = rk(M)$。

引理 2 设 S 是 $m \times n$ 矩阵，则有

$$\{A : A \in R^{m \times n}, AA' \le AS'\} = \{SM : 0 \le M \le I_n\}$$

证明：$\{A : A \in R^{m \times n}, AA' \le AS'\} \supset \{SM : 0 \le M \le I_n\}$ 显然成立；
又由引理 1 知，$\{A : A \in R^{m \times n}, AA' \le AS'\} \subset \{SM : 0 \le M \le I_n\}$，
从而结论正确。

定理 2 对于线性模型 $y = X\beta + e$， $Ee = 0, \mathrm{cov}(e) = \sigma^2 V$， $V > 0$，设计矩阵 X 列满秩，设 $C\beta$ 可估，则 $C\beta$ 的全体线性可容许估计构成集合 $D = \{C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1}y : 0 \le M \le I_p\}$。

证明：$Ay \sim C\beta \Leftrightarrow$ （1） $AV = AX(X'V^{-1}X)^{-1}X'$，

（2） $AVA' \le AX(X'V^{-1}X)^{-1}C'$

$$\Leftrightarrow AV = AX(X'V^{-1}X)^{-1}X' \text{ 且}$$

$$AX(X'V^{-1}X)^{-1}X'A' \le AX(X'V^{-1}X)^{-1}C'$$

$$\Rightarrow \exists M, 0 \le M \le I_p,\quad AX[(X'V^{-1}X)^{\frac{1}{2}}]^{-1} = C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M$$

记 $B = AX(X'V^{-1}X)^{-1}$

$$\Rightarrow \exists M, 0 \le M \le I_p,\quad B(X'V^{-1}X)^{\frac{1}{2}} = C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M$$

$$\Rightarrow \exists M, 0 \le M \le I_p$$

$$B(X'V^{-1}X)^{\frac{1}{2}}[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1} = C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1}$$

显然 $B(X'V^{-1}X)^{\frac{1}{2}}[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1} = BX'V^{-1} = A$

从而 $Ay \sim C\beta \Rightarrow \exists M, 0 \le M \le I_p$，

$$A = C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1}$$

$\Rightarrow Ay \in D$

反之，若 $Ay \in D$，则$\exists M, 0 \le M \le I_p$，

$$A = C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1}$$

$$AX(X'V^{-1}X)^{-}X'A' - AX(X'V^{-1}X)^{-}C'$$

$$= C[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}(M^2 - M)[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}C'$$

$$\le 0$$

而显然有 $AV = AX(X'V^{-1}X)^{-1}X'$，从而 $Ay \sim C\beta$。

推论 2.1 对于线性模型 $y = X\beta + e$，$Ee = 0, \mathrm{cov}(e) = \sigma^2 V$，$V > 0$，设计矩阵 X 列满秩，若$\beta$可估，则$\beta$的全体线性可容许估计构成集合

$$D = \{[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}M[(X'V^{-1}X)^{\frac{1}{2}}]^{-1}X'V^{-1}y : 0 \le M \le I_p\}$$

第五章 参数估计的相合性

§5.1 基本概念

考虑线性模型 $y = X\beta + e$，亦即 $y_i = x_i'\beta + e_i$。

定义 1 设 $f(\beta)$ 为 β 的一个已知函数，$T_n \hat{=} T_n(y_1,\dots,y_n)$ 为 $f(\beta)$ 的一个估计。若在某种意义下，当样本量 $n \to \infty$ 时，T_n 收敛于 $f(\beta)$，则称 T_n 为 $f(\beta)$ 的相合估计，或称 T_n 有相合性。

在数理统计中，收敛意义主要有依概率收敛（记作 $T_n \xrightarrow{p} f(\beta)$）、几乎必然收敛（记作 $T_n \to f(\beta) a.s.$）、和 r 阶矩收敛等。在这些意义下的相合性分别称作弱相合、强相合以及 r 阶矩相合。由概率论知，后两种相合性可以推出弱相合性，但后两者之间无蕴含关系。由于弱相合是一种基本类型的相合性，通常把弱相合简称为相合。

一般来说，我们称 T_n 为 $f(\beta)$ 的相合估计，都是指收敛性要对一切参数以及误差分布都成立。

估计的相合性，是大样本理论中讨论最多、最受重视的一个问题，其原因主要是，第一，有相合性虽不能说明该估计怎样好，但是如果一个估计没有相合性，则总是不好的，因为如果不论样本量多么大，估计值与真值仍有很大的可能出现显著的偏差，则很难相信在样本量不很大时，该估计会有良好的表现。故通常在提出一个估计量时，总是要把有无相合性作为一个考查的对象。第二，与其它深层的大样本性质相比，相合性的要求最底，更有可能在较弱的条件下，获得深入的结果。

关于 M-估计的相合性问题，除了 LSE 这个特例研究得较早并得出较完整的结果外，对于 LADE 研究得也较多，但起步较晚。至于一般的 M-估计的相合性，Huber 在 1973 年进行了研究，随后陆续发表了一些论文。但综观这些工作，所提的条件都过于繁多，因而这方面还只能认为是一个初步的研究。1993 年赵林城对 ρ 为凸时 M-估计的相合性提出了一个充分条件。

考虑 $y = X\beta + e$，亦即 $y_i = x_i'\beta + e_i$，设 ρ 为 R 上非单调的凸函数，β 的 M-估计记作 $\hat{\beta}_n$，以 Ψ_- 和 Ψ_+ 分别记 ρ 的左右导函数。

定理 1 设在线性模型 $y = X\beta + e$ 中随机误差 $e_1, e_2, \ldots$ 独立同分布，又存在函数 Ψ，满足 $\Psi_-(u) \le \Psi(u) \le \Psi_+(u)$ $(u \in R)$ 以及下面诸条件：

（1） $E\Psi(e_1) = 0$；

（2）存在常数 $c > 0, \Delta > 0$，使得

$$|E\Psi(e_1 + u)| \ge c|u|, |u| < \Delta$$

（3）存在常数 $\Delta > 0$，使得

$$E\Psi^2(e_1 \pm \Delta) < c < \infty$$

记 $S_n = \sum_{i=1}^{n} x_i x_i'$，则当 $\lim_n S_n^{-1} = 0$ 时，$\hat{\beta}_n$ 为 β 的相合估计。

注：$\lim_n S_n^{-1} = 0$ 隐含了当 n 充分大时 S_n 满秩的要求，这相当于要求设计矩阵 X 列满秩，这是必要的，否则，β 不可估计。

对于 LADE，我们有

定理 2 设在线性模型 $y = X\beta + e$ 中，随机误差 $e_1, e_2, \ldots$ 独立同分布，有中位数 0。e_1 的分布函数 F 满足条件：

当 $|u| < \Delta$ 时，　$|1 - 2F(u)| \ge c|u|$

其中 $c > 0, \Delta > 0$ 都是常数，记 $S_n = \sum_{i=1}^{n} x_i x_i'$，则当 $\lim_n S_n^{-1} = 0$ 时，β 的

LADE $\hat{\beta}_n$ 为 β 的相合估计。

该结果是 LADE 相合性研究中的一个突出成就，因为自七十年代以来，有不少统计学家都在这方面下过功夫，发表了一些结果，但是其中个别的在证明中存在缺陷，而一般的又要求条件过强，例如在误差的分布上，有的假定在 0 的邻域内有非 0 的连续导数，而该定理甚至不要求分布函数在 0 点连续。

考查线性函数 M-估计的相合性，设 $c \neq 0$ 为 p 维向量，$c'\beta$ 的 M-估计定义为 $c'\hat{\beta}_n$，在定理 1 的条件下，$\hat{\beta}_n$ 为相合，故 $c'\hat{\beta}_n$ 相合。事实上，可以将条件变得弱一些。

定理 3 在线性模型 $y = X\beta + e$ 中，随机误差 $e_1, e_2, \ldots$ 独立同分布，$\hat{\beta}_n$ 为对应于凸函数 ρ 的 M-估计，且满足定理 1 中条件（1）、（2）、（3），则当 $\lim c'S_n^{-1}c = 0$ 时，$c'\hat{\beta}_n$ 为 $c'\beta$ 的相合估计。

通常在应用中，线性回归函数常包含一常数项，这当然是 $y = X\beta + e$ 的特例，因为只需取 x_i 的第一个分量为 1 即可。在有些情况下，有必要将该常数项分开出来，此时线性模型 $y = X\beta + e$ 可以写成：$y_i = \alpha + x_i'\beta + e_i \qquad 1 \le i \le n$

很明显，上述模型中 α, β 的 M-估计的相合性问题也适用于定理 1,2,3，但我们有必要单独讨论该模型的相合性问题。

引理 1 设 G 是一个由某些无穷实数列 $a = (a_1, a_2, \ldots)$ 组成的集合。

记 $\overline{a_n} = \dfrac{\sum_{i=1}^{n} a_i}{n}$，$\sigma_n^2(a) = \sum_{i=1}^{n}(a_i - \overline{a_n})^2$，

设 G 满足下列条件：

（1）$\sup\{|a_i|: a\in G\}<\infty \quad (i=1,2,\ldots)$；

（2）$\inf\{\sigma_n^2(a): a\in G\}\to\infty \qquad$ 当 $n\to\infty$

则对 G 中的 a 一致地有

$$\overline{a_n}^2/\sigma_n^2(a)\to 0$$

证明：对于任意的 $\varepsilon>0$，存在自然数 m，使得 $\frac{4}{m}\le\varepsilon$，其次找充分大的 n_0，使得 $\sigma_{n0}(a)\sqrt{\varepsilon}>2\sup\{|a_i|: a\in G, 1\le i\le m\}$。对于任意的 $a\in G$，用 $n\ge\max(n_0,m)$ 计算 $\overline{a_n}^2/\sigma_n^2(a)$，若 $\overline{a_n}^2\le\sigma_n^2(a)\varepsilon$，则 $\overline{a_n}^2/\sigma_n^2(a)\le\varepsilon$；若 $|\overline{a_n}|>\sigma_n(a)\sqrt{\varepsilon}$，则依据 n_0 的取法以及当 $n\ge n_0$ 时，$\sigma_n(a)\ge\sigma_{n0}(a)$，我们有

$$\sigma_n^2(a)\ge\sum_{i=1}^{m}(a_i-\overline{a_n})^2\ge\sum_{i=1}^{m}(|\overline{a_n}|-\overline{a_n}/2)^2=\frac{m\overline{a_n}^2}{4}$$

因此 $\overline{a_n}^2/\sigma_n^2(a)\le\frac{m}{4}\le\varepsilon$，由 $\varepsilon>0$ 的任意性以及 m,n_0 与 a 无关，得到 $\overline{a_n}^2/\sigma_n^2(a)\to 0$ 对 a 一致成立。证毕。

显然，$S_n=\sum_{i=1}^{n}\begin{pmatrix}1\\x_i\end{pmatrix}(1\quad x_i')=\begin{pmatrix}n & n\overline{x_n}'\\ n\overline{x_n} & S_{n0}\end{pmatrix}$

其中 $S_{n0}=\sum_{i=1}^{n}x_ix_i'$，$\overline{x_n}=\sum_{i=1}^{n}\frac{x_i}{n}$。

若记 $T_n=\sum_{i=1}^{n}(x_i-\overline{x_n})(x_i-\overline{x_n})'$，则

$$S_n^{-1}=\begin{pmatrix}(n-n^2\overline{x_n}S_{n0}^{-1}\overline{x_n}')^{-1} & *\\ * & T_n^{-1}\end{pmatrix}$$

因此，$\lim_n S_n^{-1}=0$ 等价于下列两式同时成立，

$\lim(n-n^2\overline{x_n}S_{n0}^{-1}\overline{x_n}^{'})^{-1}=0$

$\lim T_n^{-1}=0$

事实上，$\lim T_n^{-1}=0$可以推出另一式，

由于$S_{n0}=T_n+n\overline{x_n x_n}^{'}$，故

$$S_{n0}^{-1}=T_n^{-1}-nT_n^{-1}\overline{x_n}\,\overline{x_n}^{'}T_n^{-1}/(1+n\overline{x_n}^{'}T_n^{-1}\overline{x_n})$$

由此得到

$$(n-n^2\overline{x_n}S_{n0}^{-1}\overline{x_n}^{'})^{-1}=n^{-1}+\overline{x_n}^{'}T_n^{-1}\overline{x_n}$$

因此要证明$\lim(n-n^2\overline{x_n}S_{n0}^{-1}\overline{x_n}^{'})^{-1}=0$，只需证明

$$\lim\overline{x_n}^{'}T_n^{-1}\overline{x_n}=0$$

为证明上式，利用引理 1，设$G=\{(a^{'}x_1,a^{'}x_2,\ldots):\|a\|=1\}$

引理 1 中条件（1）显然成立，又由于

$$\lim T_n^{-1}=0,\quad \sum_{i=1}^{n}(a^{'}x_i-a^{'}\overline{x_n})^2=a^{'}T_n a,$$

故上式在$\|a\|=1$上的最小值（即T_n的最小特征值）随着$n\to\infty$而趋向无穷大，条件（2）满足。依引理 1，我们有，对于满足条件$\|a_n\|=1$的任一列$\{a_n\}$，有 $\dfrac{(a_n^{'}\overline{x_n})^2}{a_n^{'}T_n a_n}\to 0$。

特别地，我们取$a_n=\dfrac{T_n^{-1}\overline{x_n}}{\left\|T_n^{-1}\overline{x_n}\right\|}$，则得到$\lim\overline{x_n}^{'}T_n^{-1}\overline{x_n}=0$。

定理 4 对于线性模型$y_i=\alpha+x_i^{'}\beta+e_i$　$1\le i\le n$，若它满足定理 1 的 3 个条件，则由$\lim T_n^{-1}=0$可推出回归系数的相合性。

下面我们讨论一下$\lim T_n^{-1}=0$的统计意义，若$p=1$，则有

$$\lim\sum_{i=1}^{n}(x_i-\overline{x_n})^2=\infty$$

该条件的含义是，x_i不能太聚集在某点 a（如$\overline{x_n}$）的附近。这一要求在统计上比较清楚，若x_i都聚集在某点 a 附近，则回归直线就如

同支在$(a,\alpha+a\beta)$处的杠杆。在 a 点附近的x_i的观测值只要稍有误差，就等同于在支点旁拨动杠杆，这必将引起回归直线显著变动。此时，回归参数很难有精确的估计，从而也就难以达到相合的要求。

§5.2 回归系数相合估计的存在性

对于线性模型$y=X\beta+e$，亦即$y_i=x_i'\beta+e_i$，$1\le i\le n$，在什么条件下回归系数存在相合估计，一直受到许多统计学家的关心。

当$Ee_i=0$，且$\{e_i\}$满足 Gauss-Markov 条件时，Drygas 在文[20]中给出了 LSE 相合的充要条件。K.C.Li 进一步证明了

定理 1 对于线性模型$y=X\beta+e$，若$e_1,e_2,\ldots$独立同分布，$Ee_1=0$，$0<De_1<\infty$，e_1的密度函数为f，$f>0$，在任一有界区间上绝对连续，且f的 Fisher 信息量$\int_{-\infty}^{\infty}(f'(x)/f(x))^2 f(x)dx<\infty$，记$S_n=\sum_{i=1}^{n}x_ix_i'$

则回归系数β的相合估计存在的充要条件是$\lim S_n^{-1}=0$。

一般来说，为β的相合估计存在而施加于$\{x_i\}$上的条件，与对$\{e_i\}$的假定有关。对$\{e_i\}$要求苛刻，则施加于$\{x_i\}$上的条件变弱，反之亦然。

一个自然的问题是当误差密度函数不满足定理 1 条件，β的相合估计存在的条件又是怎样的呢？陈希孺在文[19]中对一特殊情况给出了

定理 2 对于线性模型$y_i=\alpha+\beta x_i+e_i$，$(1\le i\le n)$，若$e_1,e_2,\ldots$独立同分布，$Ee_1=0$，e_1的密度函数为f，f在一有界区间$[\sigma_1,\sigma_2]$外为 0，

而在该区间内处处大于 0，且满足 Lipshitz 条件，则β的相合估计存在的充要条件是$H_n=\sum_{i=1}^{n}\left|x_i-\overline{x_n}\right|\to\infty$，其中$\overline{x_n}=\frac{1}{n}\sum_{i=1}^{n}x_i$，而$\alpha$的相合估计存在的充要条件是$\overline{x_n}/H_n\to 0$。

显然，线性模型$y_i=\alpha+\beta x_i+e_i$，$(1\le i\le n)$，当$\alpha=0$时，就成为$y_i=\beta x_i+e_i$，$(1\le i\le n)$，对此模型我们有

定理 3 在线性模型$y_i=\beta x_i+e_i$，$(1\le i\le n)$中，$e_1,e_2,\ldots$独立同分布，$Ee_1=0$，e_1的密度函数为f，f在一有界区间$[\sigma_1,\sigma_2]$外为 0，而在该区间$[\sigma_1,\sigma_2]$内满足：存在常数$C>0$，$s\ge 0$，$k\ge 0$，在σ_1和σ_2的半邻域内$f>C$。对于充分小的$\delta>0$以及任意的$x\in[\sigma_1,\sigma_2]$，有

$$\sup\{\left|f(t)-f(x)\right|:\left|t-x\right|\le\delta,t\in[\sigma_1,\sigma_2]\}\le kx^s\delta,$$

若$\lim_{n}\sum_{i=1}^{n}\left|x_i\right|=\infty$，则回归系数$\beta$的相合估计存在。

证明：若$\lim_{i}x_i\ne 0$，则必有$\lim_{n}\sum_{i=1}^{n}x_i^2=\infty$，由§5.1 定理 1 知$\beta$的最小二乘估计为相合估计。

若$\lim_{i}x_i=0$，则$\lim_{n}\overline{x_n}=\lim_{n}\frac{1}{n}\sum_{i=1}^{n}x_i=0$，

固定自然数 m，当$n\ge m$时，

$$H_n\hat{=}\sum_{i=1}^{n}\left|x_i-\overline{x_n}\right|$$

$$\ge\sum_{i=1}^{m}\left|x_i-\overline{x_n}\right|$$

$$\ge\sum_{i=1}^{m}(\left|x_i\right|-\left|\overline{x_n}\right|)$$

$$= \sum_{i=1}^{m} |x_i| - m|\overline{x_n}|$$

由 $\lim_n \overline{x_n} = \lim_n \frac{1}{n}\sum_{i=1}^{n} x_i = 0$ 知，当 n 充分大时，$|\overline{x_n}| < \frac{1}{m}$，从而

$$H_n \geq \sum_{i=1}^{m} |x_i| - 1$$

再由 $\lim_n \sum_{i=1}^{n} |x_i| = \infty$ 知，$H_n \to \infty (n \to \infty)$，于是由定理 2 知，此时回归系数 β 的相合估计存在。

注：定理 3 关于 $\{e_i\}$ 的条件，比定理 2 中关于 $\{e_i\}$ 的条件弱，但是定理 2 的充分性的证明（见文[19]），只要在 σ_1 和 σ_2 的半邻域内 e_1 的密度 f 有非 0 下界，该证明就有效。

第六章 本文结论及进一步发展

线性模型在数理统计中占有重要地位。一些富有实用意义的统计分支，诸如回归分析、方差分析和多元分析等，都以这种模型理论为基础，或与之有密切联系。有关线性模型的一些较为古典的内容，在一般统计类教科书中有不同程度的介绍。近几十年来，线性模型理论无论在广度和深度上都有不少新发展。参数估计是线性模型研究中的主要内容，自从 Gauss 和 Legendre 创立最小二乘法以来，参数估计一直是一个十分吸引人的活跃领域。本文是在导师的指导下,对线性模型中的参数估计进行了初步的研究，阐述了线性模型参数的一些重要估计及其性质，在其中做了一点工作，但由于这方面内容博大精深，本人水平有限，加之时间有限，文中不当之处，敬请批评指正。

一 本文的工作

我们对线性模型的参数估计理论进行了研究，主要工作有：

1 在 LSE 估计的稳健性方面，我们用不同于 Zyskind 的方法，得到了一些新结论。首先，我们对设计矩阵列满秩的情形，给出了 GLSE 和 LSE 相等的两个充要条件和一个充分条件（参见§2.5 定理 2 以及推论 2.1、推论 2.2）。揭示了 GLSE 和 LSE 相等与它们方差相等之间的关系。对于设计矩阵一般情形，我们给出了三个充分条件（参见§2.5 定理 3、推论 3.1、推论 3.2），较好地解决了 LSE 估计稳健性方面的问题。

2 我们讨论了 M-估计的渐近正态性，改进了这方面的一些结论。

目前的结论中，关于误差的分布函数假定为在 0 点处有正的导数，我们将该条件推广到一个更一般的条件，证明了 M-估计的渐近正态性。（参见 § 3.2）

3 我们研究了不可估函数的线性估计的可容许性，首先我们对于一类特殊的设计矩阵和误差，给出了可容许性的几个充要条件。然后，我们证明了一般情形均可化为该特殊情形进行分析，得到不可估函数线性估计的可容许性的充要条件。基本上解决了这方面的问题。（参见 § 4.3）

4 我们利用可估函数估计的可容许性的性质以及一些矩阵不等式，给出了可估函数全体可容许估计的显式表达式。（参见 § 4.4）

5 在一个很一般的条件下，证明了回归系数相合估计的存在性。

二 进一步发展

本文在第二章讨论了 LS 估计的稳健性问题，但只是就它关于协方差的稳健性进行讨论。与此类似的是 BLUE 估计关于设计矩阵的稳健性。这个问题的实际背景也很明显，但是关于设计矩阵的稳健性要比关于协方差的稳健性复杂。在关于设计矩阵的稳健性研究中，在一个模型中可估的函数，到了另一个模型可能根本不可估。

本文第四章讨论了参数估计的容许性，对于不可估计函数方面的讨论，只局限于线性估计类，对于所有估计类的情形还没有进一步讨论。在回归系数的可容许估计类中，给出了显式表达式，但只是对于协方差已知且正定的情况进行了讨论。其它情形还有待于进一步研究。

本文第五章给出了回归系数相合估计存在性的一个充分条件，它是否还是必要的，或在什么条件下是必要的，本文没有进行讨论。

参考文献

[1]陈希孺，王松桂. 近代回归分析—原理、方法及应用. 合肥：安徽人民出版社，1987

[2]王松桂. 线性模型的理论及其应用. 合肥：安徽教育出版社，1987

[3]陈希孺，白志东，赵林城，吴月华. 线性模型中最小一乘估计的渐近正态性. 中国科学 A 辑, 1990(5)

[4]陈希孺. 数理统计引论. 北京：科学出版社，1981

[5]陈希孺. 高等数理统计. 合肥：中国科学技术大学出版社，1999

[6]樊家琨. 应用多元分析. 开封：河南大学出版社，1993

[7]Rao, C. R., Estimation of Parameters in a Linear Model, Ann. Statist., 1976(4)

[8]吴启光. 一般的 Gauss-Markoff 模型中回归系数的线性估计的可容许性. 应用数学学报，1986，9（2）

[9]朱显海，鹿长余. 线性模型中参数的线性估计的可容许性. 数学年刊，1987，8A（2）

[10]朱显海，鹿长余. 在二次损失下回归系数的非齐次线性估计的可容许性. 应用概率统计，1986，2（2）

[11]陈希孺，吴启光，陈桂景，赵林城. 线性模型参数的估计理论. 北京：科学出版社，1985

[12]Akahira, M. and Takeuchi, K. Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency. Springer, New York. 1981

[13]Akahira, M. and Takeuchi, K. Non-regular statistical

estimation. Springer, New York. 1995

[14]Arcones, M. A. Distributional convergence of M-estimators under unusual rates. Statist. Probab. Lett. 21. 1994

[15]Arcones, M. A. The Bahadur-Kiefer representation of L_p regression estimators. Econometric theory. 12. 1996

[16]Babu, G. J. Strong representation for LAD estimators in linear models. Probab. Theory Related Fields. 83. 1989

[17]He, X. and Shao, Q.-M. General Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. Ann. Statist. 24. 1996

[18]Hjort, N. L. and Pollard, D. Asymptotics for minimisers of convex processes. Statistical Research report, Univ. Oslo. 1993

[19]Chen Xiru. On a problem of existence of consistent estimate, in the development of statistics: Recent contributions from China. Pitman Research Notes in Mathematical Series, Longman Scientific & Technical, London, 1992

[20]Drygas H. Weak and strong consistency of the least squares estimators in regression model. Z. Wahrsch. Verw, Gebiete, 1976

[21]Lai T L, Robbins H and Wei C Z. Strong consistencyof least squares estimates in multiple regression II. J. Multivariate Anal., 1979

[22]Li K C. Regression models with infinitely many

parameters:consistency of bounded linear functionals. Ann.Statist.,1984

[23]Zhao L C,Rao C R and Chen X r. A note on the consistency of M-estimators in linear models,in Stochastic Process.A Festschrift in honor of Gopinath Kallianpur,Springer-Verlag,1993

[24]Chen,X.R. A problem on weak consistency of linear estimates,Chinese Ann.of Math.,2,1981

[25]王正明，易东云.测量数据建模与参数估计，长沙：国防科技大学出版社，1996

[26]齐全跃，陈敏等.跟踪雷达测量误差统计模型（II）：参数估计强相合性，北京：应用数学学报，1997（2）

[27]王正明，周海银.关于回归模型的参数估计效率，北京：数学的实践与认识，1999（4）

[28]孙淑珍，王石青.回归系数的线性有偏估计，吉林大学学报，1998（4）

[29]王松桂，贾忠贞.矩阵论中的不等式.合肥：安徽教育出版社，1994

[30]陈宝明，詹金龙.论线性模型中最小二乘估计的有效性，昆明理工大学学报，1999（2）

[31]Chen J B,Chen T. The error tatio efficiency of the mean square in the general Gauss-markov model,The China-Japan Symposium Kunming,1991(4)

[32]王松桂.线性模型中均值向量的LSE和BLUE之差的欧氏范数界.应用数学学报，1994（5）

[33]Wang S G,Yan H.Kantorovich type inequality and

measures of inefficiency of the GLSE. Acta Math. Appl Sinica, 1989(5)

[34]王学仁，詹金龙，陈建宝. 方差分量模型中回归系数和参数的可容许估计. 数学学报，1994（5）

[35]Zhan J L, Chen J B. The inefficiency of least squares in Gauss-markov and variance component models. Acta Math Appl Sinica, 1998(2)

[36]倪国熙. 常用的矩阵理论和方法，上海科学技术出版社，1984

[37]陈希孺，赵林城. 线性模型中的 M 方法，上海科学技术出版社，1996

[38]陈希孺. 线性估计弱相合的一个问题. 数学年刊，1981

[39]陈希孺. 线性模型参数 M 估计的线性表示. 中国科学 A 辑，1993（23）

[40]Bai Z D, Chen X R, Miao B Q and Rao C R. Asymptotic theory of least distances estimate in multivariate linear models. Statistics, 1990(21)

[41]Bai Z D, Rao C R and Wu Y. M-estimation of multivariate linear regression parameters under a convex discrepancy function, Statistica Sinica, 1992(2)

[42]Bai Z D, Rao C R and Zhao L C. Weak representation of regression estimates in multivariate linear models obtained by minimizing a convex function of residuals. Tech Report Penn state Univ, 1992

[43]Bassett G and koenker R. Asymptotic theory of least absolute error regression, J Amer statist Assoc, 1978(73)

[44]Bickel P J. One step Huber estimates in the linear model. J Amer statist Assoc, 1975(70)

[45]Chen X R and Wu Y H. On a necessary condition for the consistency of L_1 estimates in linear models. Comm Statist, Theory and methods, 1993(22)

[46]Chen X R, Zhao L C and Wu Y H. On conditions of consistency of ML_1N estimates. Statistica Sinica, 1993(3)

[47]Heiler S and Willers R. Asymptotic normality of M-estimates in the linear model. Statistics, 1988(19)

[48]Huber P J. Robust regression. Ann Statist, 1973(1)

[49]Huber P J. Robust estimation of a location parameter. Ann Math Statist, 1964(35)

[50]Koenker R W and Portnoy S. M-estimation of multivariate regressions. J Amer Statist Assoc, 1990(85)

[51]Loeve M. Probability Theory. 4-th ed., Springer-Verlag, 1977

[52]Singer J M and Sen P K. M-methods in multivariate linear models. J Multivariate Analysis, 1985(17)

[53]成平，陈希孺，陈桂景，吴传义. 参数估计. 上海科学技术出版社，1985

[54] . 主成分分析在研究生中期考核中的应用. 海峡两岸统计研讨会论文集，1999

[55]严士健，刘秀芳. 测度与概率. 北京师范大学出版社，1994

[56]王松桂. 线性模型参数估计的新进展. 数学进展，1985（14）

[57]Niemiro W. Asymptotics for M-estimators defined by

convex minimization, Ann Statist, 1992(20)

[58]Pollard D. Asymptotics for least absolute deviation regression estimators. Econom Theory, 1991(7)

攻读学位期间发表的论文

1 Carleman 不等式的加强，苏州铁道师院学报，Vol.16，1999（4）

2 Feller 公式的一个构造性证明，北京成人教育学院学报，1999(4)

3 主成分分析在研究生中期考核中的应用，海峡两岸统计研讨会论文集，1999

4 射表中的小子样理论研究，数理统计与管理，Vol.18 增刊

5 矩阵连根式序列的收敛性，北方交通大学学报 No.2，2000

6 随机系数代数方程平均实根个数的估计，北方交通大学学报 No.2，2000

7 中位数、数学期望与众数的大小关系，高等数学研究，No.1，2000

致谢

本文是在我的导师　　教授的悉心指导下完成的。从论文的选题、文献的分析、理论的研究直到论文的撰写都得到了导师的谆谆教诲。导师学识渊博、治学严谨、为人谦和。他不仅在学术上严格要求，注重能力的培养，使我增长才干，而且在为人上以身作则，言传身教，使我受益颇深。所有这些，都为我今后的学习和工作打下了坚实的基础。在此，我谨向导师　　教授致以学生最衷心的感谢和最诚挚的敬意。

本人在北方交通大学数学系求学六年多，一直得到数学系老师们的关心和帮助。许多老师都曾给我上过课，他们兢兢业业，把自己的一切献给了伟大的科学事业，值得我永远学习。刘彦佩教授、刘坤会教授、季文铎教授、陈治中教授、吴发恩副教授、汪成咏副教授、张后扬副教授等都曾给予我大力的支持和鼓励，本人的不断进步也是和他们的指导分不开的。

最后，向所有关心和支持我的人们表示衷心的感谢！