数学卷

中国科学技术

经·典·文·库

线性模型参数的估计理论

陈希孺 陈桂景 吴启光 赵林城 著

科学出版社
www.sciencep.com

中国科学技术经典文库·数学卷

线性模型参数的估计理论

陈希孺　陈桂景
吴启光　赵林城　著

科学出版社
北京

内　容　简　介

本书为作者近几年在数理统计线性模型参数估计理论方面所做的研究工作的总结.

全书共分四章. 第一章是预备知识, 第二章讨论线性模型回归系数的最小二乘估计及一般线性估计的相合性问题, 第三章介绍误差方差估计的大样本性质, 第四章讨论小样本理论, 即回归系数的线性估计与误差方差的二次型估计的容许性问题.

本书读者对象为高等院校数学系高年级大学生、研究生、教师和数理统计科学研究工作者.

图书在版编目(CIP)数据

线性模型参数的估计理论/陈希孺等著. —北京: 科学出版社, 2010
(中国科学技术经典文库 · 数学卷)
ISBN 978-7-03-028792-2

Ⅰ. 线…　Ⅱ. 陈…　Ⅲ. 线性模型-参数估计　Ⅳ. O212

中国版本图书馆 CIP 数据核字(2010) 第 166909 号

责任编辑: 刘嘉善　陈玉琢 / 责任校对: 朱光兰
责任印制: 钱玉芬 / 封面设计: 王　浩

科学出版社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

北京佳信达欣艺术印刷有限公司 印刷
科学出版社发行　各地新华书店经销
*

2010 年 9 月第　一　版　开本: B5 (720 × 1000)
2010 年 9 月第一次印刷　印张: 14
印数: 1—2 000　字数: 272 000

定价: 58.00 元
(如有印装质量问题, 我社负责调换)

序　言

本书的目的是介绍线性模型理论的若干新发展. 对数理统计知识有过一点接触的人, 都了解线性模型的重要地位. 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. 因此, 有关这种统计模型的一些较为古典的内容, 在一般数理统计教科书中都有不同程度的介绍. 近几十年来, 特别是 20 世纪 60 年代以来, 线性模型理论无论在广度和深度上都有不少新发展, 像大样本理论、可容许的线性与二次型估计、非参数和 Robust 估计、序贯和 Bayes 方法以及自变量也带随机误差的所谓 "Error In Variables" 模型等等. 这些发展大都有实用上的意义: 有的改进了传统的估计方法而提供了较好的估计, 有的扩大了模型的应用范围, 有的在误差的正态性不成立的情况下提供了可用的大样本检验和区间估计等. 另一些发展的主要意义则在于纯理论方面, 它加深了我们对这个重要模型的性质的认识.

本书作者近年来在这个领域里做了一点研究工作, 对其现状作了一些了解, 写作这本专著的念头就是由此而起. 但由于篇幅所限而且由于不少新的发展目前还远未达到比较成熟和定型, 所以要写一本详尽的、包括到目前为止的所有主要成果的专著是不现实的. 我们希望本书内容以我们自己的工作为基础, 这样, 对所涉及的课题能作较深入的论述. 因此, 我们挑选了线性模型参数的线性和二次型估计的大样本理论和容许性这些题材, 并把书名定为《线性模型参数的估计理论》.

本书共分四章. 第一章是预备知识. 在这一章开头列举了为阅读本书所需的各种预备知识. 由于本书使用的方法没有超出古典分析、矩阵及一般分析概率论的范围, 即使只读过少数基本课程的概率统计专业学生, 也不难看懂本书的绝大部分内容. 第二章讨论线性模型回归系数的最小二乘估计及一般线性估计相合性问题, 介绍了在各种意义下的相合性条件. 从概率论角度, 可以把本章内容看作是古典大数定律的某种推广, 因为它的主题无非是关于线性型 $\sum_{i=1}^{n} a_{ni}e_i$ (这里 a_{ni} 是常数, e_i 为满足一定性质的随机变量) 在各种意义下收敛到 0 的问题. 只是 $\{a_{ni}\}$ 是由试验点列 $\{x_i\}$ 所决定, 而条件必须加在 $\{x_i\}$ 上, 而不能直接涉及 $\{a_{ni}\}$, 因而增加了复杂性. 第三章讨论线性模型的另一重要参数 —— 误差方差估计的大样本性质, 主要是讨论基于残差平方和的二次型估计. 在相合性、渐近于正态分布的一致和非一致性速度等问题上, 都得到了较理想的结果. 在一个特殊情况下, 这方面的工作早在 20 世纪 40 年代已由许宝騄教授开其端. 本章的工作可以看作是他的工作的继续

和发展. 第四章讨论回归系数的线性估计与误差方差的二次型估计的容许性问题. 前两节主要是介绍 Rao, Cohen, Stein, James 和 Brown 等人的工作, 其中包括了近代参数估计理论中的若干重大成果, 过去在中文文献中还很少介绍. 以后几节包括在矩阵损失下线性估计的容许性及误差方差二次型估计的容许性, 则主要是本书作者及其合作者的工作.

作者的一个希望是使一些初进入研究工作的青年读者相信, 使用初等工具也可以解决数理统计学中比较困难的理论问题, 并达到比较深入的结果. 总观近四十年来数理统计学发展的状况, 给人的印象是: 虽然新的结果大量涌现, 但这个数学分支仍保持了这样一个特点: 其多数重大结果依赖于熟练和深入地使用古典方法的技巧, 而不是更新的数学工具.

本书写作分工情况如下: 第一章至第四章的初稿分别由陈希孺、陈桂景、赵林城和吴启光执笔, 最后由陈希孺写成定稿. 由于作者水平所限, 书中不妥和错误之处肯定不少, 希望同行专家和广大读者不吝赐教.

作 者

1981 年 11 月 24 日

目　　录

序言

第一章　预备知识 …… 1

§1.1　矩阵与线性模型 …… 1

§1.2　判决函数与容许性 …… 10

§1.3　概率论中的若干极限定理 …… 18

参考文献 …… 29

第二章　回归系数最小二乘估计的相合性 …… 30

§2.1　LS 估计弱相合的条件 …… 31

§2.2　一般线性弱相合估计的存在问题 …… 43

§2.3　LS 估计的 r 阶平均相合性 …… 54

§2.4　LS 估计的强相合性 …… 64

参考文献 …… 88

第三章　误差方差估计的大样本性质 …… 90

§3.1　σ_n^2 的相合性 …… 91

§3.2　一致性收敛速度（Ⅰ） …… 104

§3.3　一致性收敛速度（Ⅱ） …… 119

§3.4　非一致性收敛速度 …… 135

§3.5　σ_n^2 的分布的渐近展开 …… 151

参考文献 …… 155

第四章　线性模型参数估计的容许性问题 …… 156

§4.1　回归系数的线性估计的可容许性Ⅰ (在线性估计类中) …… 157

§4.2　回归系数的线性估计的可容许性Ⅱ (在一般估计类中) …… 171

§4.3　矩阵损失下回归系数线性估计的可容许性 …… 182

§4.4　误差方差的二次型估计的可容许性Ⅰ (在二次型估计类中) …… 189

§4.5　误差方差的二次型估计的可容许性Ⅱ (在一般估计类中) …… 213

参考文献 …… 216

第一章　预 备 知 识

阅读本书所需的预备知识有以下三方面：一是相当于大学二年级程度的数学分析与线性代数，二是数理统计. 除了初等教本中包含的一般性内容外，还需要一点线性模型的估计理论知识、判决函数的基本概念、Bayes 方法初步和估计的容许性理论中若干较不常见的结果. 三是概率论. 需要测度论、强弱极限理论及鞅论的初步知识，程度大体上相当于 Loève 的专著 Y[1]，个别地方还需用到 Petrov 专著 [2] 和 Stout 的专著 [3] 中的材料.

在本章中，我们打算对本书中常用的一部分知识 (如线性模型的最小二乘估计理论和矩阵的广义逆等) 以及某些在文献中不易查阅的事实给以较仔细的叙述. 对其他内容，主要是概率论中极限理论方面，因涉及面太广，自无法在此详细叙述. 但准备将一些常用结果不加证明地汇集一下，以便于查阅. 了解这些结果的确切意义而不必涉及其证明细节，就可以读懂本书的有关部分. 当然，如果要进一步做这方面的研究工作，则必须去钻研上面提到的有关著作，或与之相当的著作.

§1.1　矩阵与线性模型

先提出本书中常用的一些记号. m 行 n 列的矩阵 A 常称为 $m\times n$ 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'. 一列矩阵称为列向量，而一行矩阵称为行向量. 我们总是以不加 “$'$” 的向量表列向量. 例如，a 为一列向量，而 a' 则为行向量. 向量 $a=(a_1,\cdots,a_n)'$ 的长为 $\left(\sum\limits_1^n a_i^2\right)^{1/2}$，记为 $\|a\|$. 矩阵 A 的秩记为 $\mathrm{rk}(A)$. 若 A 为方阵，则其迹 (trace)，即主对角线元之和，记为 $\mathrm{tr}(A)$. 本书中涉及的矩阵都是实的.

若 $\mathscr{A}$ 为一线性子空间，则其正交补空间记为 $\mathscr{A}^{\perp}$. 设 $A=(a_1\vdots\cdots\vdots a_n)$，则由 A 的列向量 $a_1,\cdots,a_n$ 张成 (或生成) 的线性子空间记为 $\mathscr{M}(A)$. 其正交补空间记为 $\mathscr{M}^{\perp}(A)$.

正定方阵 A 记为 $A>0$. 半正定 (又称非负定) 方阵 A 记为 $A\geqslant 0$. 在本书中，正定与半正定方阵必为对称. 若 $A-B\geqslant 0(A-B>0)$，则记为 $A\geqslant B(A>B)$. n 阶单位阵记为 I_n，或简记为 I.

设 $A:m\times n$ 的 (i,j) 元为 a_{ij}，则记为 $A=(a_{ij})_{i=1,\cdots,m,j=1,\cdots,n}$. 设 $A_n=$

$(a_{ij}^{(n)})_{i=1,\cdots,u,j=1,\cdots,v}$, $n=1,2,\cdots$. 若

$$\lim_{n\to\infty} a_{ij}^{(n)}=0, \quad i=1,\cdots,u, \quad j=1,\cdots,v,$$

则称 $\lim\limits_{n\to\infty} A_n=0$ 或 $A_n\to 0$. 若 $\{b_n\}$ 为一列正数, 且当 $n\to\infty$ 时, $a_{ij}^{(n)}=O(b_n)(o(b_n))$ 对 $i=1,\cdots,u,j=1,\cdots,v$, 则称 $A_n=O(b_n)(o(b_n))$.

(一) 广义逆　设 A 为任一矩阵. 若矩阵 B 满足关系

$$ABA=A, \tag{1.1}$$

则称 B 为 A 的一个 "减号广义逆", 记为 $B=A^-$. 在本书中, 我们只用这种方式定义的广义逆, 因此在以后, "减号" 两字常省去.

若 A 为一满秩 (非异) 方阵, 则 A^- 惟一且等于 A 的逆矩阵 A^{-1}. 以下将看到, 这事实之逆亦真. 由这个性质可知, A^- 是 A^{-1} 的某种推广. 这在以下的性质 1 中看得更为明显.

广义逆的性质 (以下只涉及在后面有用的):

1. $B=A^-$ 的充要条件为, 若 $Ax=c$ 有解, 则 $x=Bc$ 为其一解.

证　设 $B=A^-$, 则式 (1.1) 成立. 记 $A=(a_1\vdots\cdots\vdots a_n)$, 有 $ABa_i=a_i$, $i=1,\cdots,n$. 因为 $Ax=c$ 有解, 任取其一解 $x^*=(x_1^*,\cdots,x_m^*)'$, 故有 $c=\sum\limits_1^m x_i^*a_i$, 因而

$$ABc=A\sum_1^m x_i^*\ Ba_i=\sum_1^m x_i^*ABa_i=\sum_1^m x_i^*a_i=c,$$

即 Bc 为一解. 反过来, 若所设条件成立, 则因 $Ax=a_i$ 有解 ($(0,\cdots,0,1,0,\cdots,0)'$ 即为一解), 有 $ABa_i=a_i\,(i=1,\cdots,n)$, 因而式 (1.1) 成立, 故 $B=A^-$.

2. 对任意 A, A^- 必存在. 更进一步: 若通过初等变换将 A 变为

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

的形状, 则 $B=A^-$ 的充要条件为: B 有形式

$$B=Q^{-1}\begin{pmatrix} I_r & D\\ E & F\end{pmatrix}P^{-1},$$

此处 D,E,F 任意 (当然, 若 A 为 $m\times n$, 则 $\begin{pmatrix} I_r & D\\ E & F\end{pmatrix}$ 必须为 $n\times m$).

证　设 A 为 $m\times n$. 任取一个 $n\times m$ 的矩阵 B, 表之为 $\begin{pmatrix} G & D\\ E & F\end{pmatrix}P^{-1}$. 则

$$
\begin{aligned}
ABA = A \Leftrightarrow P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} G & D \\ E & F \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q \\
\Leftrightarrow \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} G & D \\ E & F \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \\
\Leftrightarrow \begin{pmatrix} G & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \\
\Leftrightarrow G = I_r,
\end{aligned}
$$

明所欲证. 由这个结果推出以下几点:

a. A^- 必存在.

b. $\mathrm{rk}(A^-) \geqslant \mathrm{rk}(A)$. 更进一步: 若 $A: m\times n$, 则对任意 t, $\mathrm{rk}(A) \leqslant t \leqslant \min(m,n)$, 存在 A^-, 致 $\mathrm{rk}(A^-) = t$.

c. 当且仅当 A 为满秩方阵时, A^- 才惟一.

3. 对任意矩阵 A, 有

$$AA'(AA')^- A = A, \tag{1.2}$$

$$A'(AA')^- AA' = A'. \tag{1.3}$$

证 因为任一实矩阵 B=0 的充要条件是 BB'=0, 记 $B = AA'(AA')^- A - A$. 有 $B' = (CA)' = A'C'$, 其中 $C = AA'(AA')^- - I$, 故由式 (1.1) (改其中的 A 为 AA', B 为 $(AA')^-$)

$$BB' = [AA'(AA')^- AA' - AA']C' = 0,$$

因而 B=0. 这证明了式 (1.2). 式 (1.3) 类似证明.

4. 设 A 为对称方阵, 将 A 表为

$$A = P\mathrm{diag}(\lambda_1, \cdots, \lambda_n)P',$$

此处 P 为正交阵, $\lambda_1, \cdots, \lambda_n$ 为 A 的全部特征根. 又 $\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ 记一对角阵, 其主对角线元依次为 $\lambda_1, \cdots, \lambda_n$. 则 A 之一广义逆为

$$A^- = P\mathrm{diag}(\lambda_1^{-1}, \cdots, \lambda_n^{-1})P',$$

此处约定 $O^{-1} = O$. 由此可知, 若 A 对称 (半正定), 则必存在对称 (半正定) 的 A^-.

(二) 投影矩阵与幂等矩阵 设 $\mathscr{A}$ 为 n 维实向量空间中之一线性子空间. 如所周知, 任一向量 x 可分解为 $x = a_x + b_x$, 其中 $a_x \in \mathscr{A}$, 而 $b_x \in \mathscr{A}^\perp$, 且这种分解是惟一的. 变换 $x \to a_x$ 显然是一线性变换, 称为 "向 $\mathscr{A}$ 的投影变换". 将此变换用矩阵表为 $a_x = B_x$, 则 B 称为 "向 $\mathscr{A}$ 的投影 (变换) 阵". 一般地, 若 B 为向某一线性子空间的投影阵, 则称 $\mathscr{B}$ 为投影阵.

有如下的重要结果.

定理 1.1　设 A 为任一矩阵, 则向 $\mathscr{M}(A)$ 的投影阵为 $P_A = A(A'A)^- A'$(由投影阵的惟一性, 这结果也表示 $A(A'A)^- A'$ 与 $(A'A)^-$ 的取法无关).

证　只需验证以下三条: a. $P_A x \in \mathscr{M}(A)$, 对任意 x. b. $P_A x = x$, 当 $x \in \mathscr{M}(A)$. c. $x - P_A x \perp \mathscr{M}(A)$, 对任意 x.

a 是显然的. 对 b, 注意若 $x \in \mathscr{M}(A)$, 则存在向量 c, 使得 $x = Ac$. 于是由式 (1.2),

$$P_A x = P_A A c = A(A'A)^- A' A c = Ac = x.$$

为证 c, 任取 $y = Ad \in \mathscr{M}(A)$, 则由式 (1.3)

$$\begin{aligned} y'(x - P_A x) &= d'A'(I - A(A'A)^- A')x \\ &= d'[A' - A'A(A'A)^- A']x = 0. \end{aligned}$$

明所欲证.

这个结果给投影阵以一明确的表达式, 是广义逆的一重要应用. 易见, 向 $\mathscr{M}^{\perp}(A)$ 的投影阵为 $I - P_A$. 又 P_A 之秩等于 A 之秩. 事实上, 由 $P_A = A(A'A)^- A'$ 知 $\mathrm{rk}(P_A) \leqslant \mathrm{rk}(A)$. 又在式 (1.2) 中改 A 为 A', 得 $A'P_A = A'$, 因而

$$\mathrm{rk}(P_A) \geqslant \mathrm{rk}(A') = \mathrm{rk}(A).$$

若一方阵 A 满足 $A^2 = A$, 则称 A 为幂等阵. 以后我们只考虑对称的幂等阵. 易见这种方阵的特征根只能为 1 或 0, 故得

1. A 为对称幂等阵的充要条件为: 存在正交阵 P, 致

$$A = P' \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} P, \quad r = \mathrm{rk}(A). \tag{1.4}$$

设 $P = (p_{ij})_{i,j=1,\cdots,n}$, 则由式 (1.4) 得

$$x'Ax = \sum_{i=1}^{r} \left(\sum_{j=1}^{n} p_{ij} x_j \right)^2, \quad x = (x_1, \cdots, x_n)'. \tag{1.5}$$

又利用关系式 $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, 由式 (1.4) 立得

2. 对称幂等阵的迹等于其秩 (此性质对一般幂等阵也成立).

幂等阵之一重要性质为其与投影阵的联系:

3. A 为投影阵的充要条件为: A 对称幂等.

证　设 A 为向线性子空间 $\mathscr{A}$ 的投影阵. 在 $\mathscr{A}$ 中找向量 $b_1, \cdots, b_n$, 使得 $\mathscr{A} = \mathscr{M}(B)$, $B = (b_1 \vdots \cdots \vdots b_n)$, 则 $A = P_B = B(B'B)^- B'$. 因 $B'B$ 对称, 故由广义逆

的性质 4, 知存在对称的 $(B'B)^-$. 取此作为 P_B 中的 $(B'B)^-$(前已指出, 由 P_B 的惟一性, P_B 与 $(B'B)^-$ 的取法无关), 知 $A = P_B$ 为对称阵. 又由式 (1.2)(改 A 为 B') 有

$$A^2 = B(B'B)^- B'B(B'B)^- B' = B(B'B)^- B' = A,$$

知 A 为幂等的. 反过来, 若 A 为对称幂等, 则易见, A 即为向 $\mathscr{M}(A)$ 的投影阵. 因 $Ax \in \mathscr{M}(A)$, 又

$$\begin{aligned} x \in \mathscr{M}(A) &\Rightarrow x = Ac \Rightarrow Ax = AAc = Ac = x; \\ y \in \mathscr{M}(A) &\Rightarrow y = Ac \Rightarrow y'(x - Ax) = c'A'(I - A)x \\ &\qquad\qquad\qquad\qquad\quad = c'A(I - A)x \\ &\qquad\qquad\qquad\qquad\quad = c'(A - A^2)x = 0. \end{aligned}$$

明所欲证.

(三) 线性模型与最小二乘估计 设 $\beta = (\beta_1, \cdots, \beta_p)'$ 为未知的 p 维向量, $x_i = (x_{1i}, \cdots, x_{pi})'$ $(i = 1, \cdots, n)$ 为已知的 p 维向量. $e_1, \cdots, e_n$ 为随机变量, 则称

$$Y_i = x_i'\beta + e_i, \quad i = 1, \cdots, n \tag{1.6}$$

为一线性 (回归) 模型. 事实上, 这表示一个包含 p 个自变量 $X_1, \cdots, X_p$ 和一个因变量 Y 的结构. 在第 i 次试验或观察时, 自变量 $X_1, \cdots, X_p$ 分别取值 $x_{1i}, \cdots, x_{pi}$, 而因变量则取值 Y_i. 在一些问题中, $x_{1i}, \cdots, x_{pi}$ 的值可事先指定. 这时称 $x_1, \cdots, x_n$ 为 "试验点列", 而 $X = (x_1 \vdots \cdots \vdots x_n)'$ 为 "设计矩阵". 即使在 x_i 之值不能由试验者自由选择的场合, 我们为方便计也沿用以上术语. $\beta_1, \cdots, \beta_p$ 称为回归系数. $x_i'\beta$ 可视为因变量 Y_i 的值中, 依赖于自变量的部分, 而 e_i 则被视为第 i 次试验的随机性误差. 因此, 常假定

$$E_{ei} = 0, \quad i = 1, \cdots, n. \tag{1.7}$$

引进 $Y = (Y_1, \cdots, Y_n)'$, $e = (e_1, \cdots, e_n)'$, 可将线性模型 (1.6) 写为矩阵形式

$$Y = X\beta + e. \tag{1.8}$$

在大样本理论中, 式 (1.6) 中的 n 是不固定的, 这时写成 $Y_i = x_i'\beta + e_i$, $i = 1, \cdots, n, \cdots$. 相应于式 (1.8) 的形式, 则上式为 $Y_{(n)} = X_{(n)}\beta + e_{(n)}$, $n = 1, 2, \cdots$.

线性模型理论中首要问题之一, 就是利用 x_i 和 $Y_i\,(i = 1, \cdots, n)$, 以对 β 及其线性函数作出估计, 及检验关于它的假设. 在此我们只讨论估计问题. 称 $\hat{\beta}$ 为 β 的最小二乘 (least squares, 简记为 LS) 估计, 若 $\hat{\beta}$ 满足条件

$$\|Y - X\hat{\beta}\|^2 = \min\{\|Y - X\beta\|^2 : \beta \in \mathbb{R}^p\}. \tag{1.9}$$

用微分法, 得出必要条件为：$\hat{\beta}$ 是方程

$$S\beta = X'Y \quad (S = X'X) \tag{1.10}$$

之解. 方程 (1.10) 称为正则方程.

定理 1.2　1° 方程 (1.10) 必有解. 2° 方程 (1.10) 之任一解必满足条件 (1.9). 3° 反之, 若 β^* 满足条件 (1.9), 则 β^* 为 (1.10) 之一解.

证　1° 显然. 因由 (1.2) 易知：

$$\hat{\beta} = S^{-}X'Y \tag{1.11}$$

就是 (1.10) 之一解. 若 $\hat{\beta}$ 为 (1.10) 之任一解, 则对任意 $\beta \in \mathbb{R}^p$ 有

$$\begin{aligned}\|Y - X\beta\|^2 &= \|Y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2 \\ &\quad + 2(\hat{\beta} - \beta)'X'(Y - X\hat{\beta}) \\ &= \|Y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2 \geqslant \|Y - X\hat{\beta}\|^2. \end{aligned} \tag{1.12}$$

这证明了 2°. 若 β^* 满足条件 (1.9), 则以 $\beta = \beta^*$ 代入式 (1.12), 应有 $\|X(\hat{\beta} - \beta^*)\|^2 = 0$, 因而 $X\hat{\beta} = X\beta^*$, 故

$$S\beta^* = X'X\beta^* = X'X\hat{\beta} = S\hat{\beta} = X'Y,$$

即 β^* 确为方程 (1.10) 之一解. 证毕.

若 $\mathrm{rk}(X) = p$, 则 $|S| \neq 0$, 这时称为 “满秩情形”. 在满秩情形下, 式 (1.10) 的解, 即 β 的 LS 估计是惟一的：

$$\hat{\beta} = S^{-1}X'Y. \tag{1.13}$$

在这种情形下称 β 为可估的. 一般, 只有在 β 可估时, 才谈到其 LS 估计. 但我们以后不坚持这一点, 而称可表为式 (1.11) 的 $\hat{\beta}$ 为 β 的 LS 估计.

容易证明, 若 β 可估而条件 (1.7) 满足, 则 β 的 LS 估计 $\hat{\beta}$ 为无偏的. 又若 $e = (e_1, \cdots, e_n)'$ 的协差阵为 Σ, 则 $\hat{\beta}$ 的协差阵为

$$\mathrm{cov}(\hat{\beta}) = S^{-1}X'\Sigma XS^{-1}. \tag{1.14}$$

事实上, 有 $\hat{\beta} = S^{-1}X'Y = S^{-1}X'(X\beta + e) = \beta + S^{-1}X'e$, 故由 $Ee = 0$ 立知 $\hat{\beta}$ 为无偏的. 再注意到 S 对称, 即得

$$\mathrm{cov}(\hat{\beta}) = S^{-1}X'\Sigma(S^{-1}X')' = S^{-1}X'\Sigma XS^{-1}.$$

一个重要的特例是

$$\Sigma = \sigma^2 I_n, \quad 0 < \sigma^2 < \infty, \tag{1.15}$$

这时由式 (1.14) 得

$$\mathrm{cov}(\hat{\beta}) = \sigma^2 S^{-1}. \tag{1.16}$$

如果线性模型 (1.8) 满足条件 (1.7) 和 (1.15), 则称它为 Gauss-Markov 模型, 简称为 GM 模型 (在此并未要求 X 满秩). 条件 (1.7) 和 (1.15) 也称为 GM 条件或 GM 假定. 在讨论线性模型时, 一般常假定条件 (1.7), 故 GM 条件一般特指式 (1.15). 在式 (1.15) 中, σ^2 为随机误差的 (公共) 方差. 它一般假定为未知, 是线性模型的一个重要参数.

在证明定理 1.3 时曾得到表达式

$$\hat{\beta} - \beta = S^{-1} X' e, \tag{1.17}$$

它把 $\hat{\beta} - \beta$ 表为 $e_1, \cdots, e_n$ 的线性组合, 这是一个很有用的关系式.

(四) 可估函数与 Gauss-Markov(GM) 定理, BLUE 设有线性模型 (1.8), 并设条件 (1.7) 成立. 设 $c'\beta$ 为 β 之一线性函数. 若存在 $c'\beta$ 之一无偏估计量 $\varphi(Y)$, 则称 $c'\beta$ 为可估的 ($\varphi(Y)$ 自然还可依赖于 $x_1, \cdots, x_n$. 因 $x_1, \cdots, x_n$ 已知且无随机性, 这个依赖关系不必指出). 若 C 为 $r \times p$ 矩阵, 其各行向量为 $c_1', \cdots, c_r'$, 而 $c_i'\beta$ 皆可估 ($i = 1, \cdots, r$), 则称 $C\beta$ 为可估的.

定理 1.3 设 $c'\beta$ 可估, 则必存在 $c'\beta$ 的形如 $a'Y$ 的线性无偏估计. 又 $c'\beta$ 可估的充要条件为 $c \in \mathscr{M}(X')$.

证 首先, 由于 $c'\beta = E_\beta \varphi(Y)$, 而 Y 的分布只通过 $X\beta$ 而依赖于 β①, 有

$$X\beta = X\beta^* \Rightarrow c'\beta = c'\beta^*.$$

换句话说, 若 $\beta - \beta^*$(它可取 $\mathbb{R}^p$ 中任意向量为值) 与 $\mathscr{M}(X')$ 正交, 则必与 C 正交, 因而 $c \in \mathscr{M}(X')$. 反过来, 若 $c \in \mathscr{M}(X')$, 则存在 a, 使 $c = X'a$. 这时 $E_\beta(a'Y) = a'X\beta = c'\beta$, 因而 $c'\beta$ 可估. 这一举证明了定理中的两个结论.

以 $\hat{\beta}$ 记 β 的任一个 LS 估计. 当 $c'\beta$ 可估时, 称 $c'\hat{\beta}$ 为 $c'\beta$ 的 LS 估计. 易见此时 $c'\hat{\beta}$ 与 $\hat{\beta}$ 的取法无关. 事实上, 若 $\hat{\beta}_{(1)}$ 和 $\hat{\beta}_{(2)}$ 为两个 LS 估计, 则 $X'Y = S\hat{\beta}_{(1)} = S\hat{\beta}_{(2)}$. 因 $c'\beta$ 可估, 有 $c \in \mathscr{M}(X') = \mathscr{M}(X'X) = \mathscr{M}(S) = \mathscr{M}(S')$, 故存在 d, 使得 $c = S'd$, 而

$$c'\hat{\beta}_{(1)} = d'S\hat{\beta}_{(1)} = d'S\hat{\beta}_{(2)} = c'\hat{\beta}_{(2)}.$$

明所欲证.

---

① 此处隐含了这样的假定: 在模型 (1.8) 中, e 的分布与 β 无关. 若给线性模型以更一般的定义: $EY = X\beta$, $\mathrm{COV}(Y) = \sigma^2 I$, 则定理 1.3 的前一结论不必成立.

定理 1.4 (GM 定理)　设 GM 条件满足. 若 $c'\beta$ 可估, 则在 $c'\beta$ 的一切线性无偏估计类中, 其 LS 估计 $c'\hat{\beta}$ 是惟一的方差一致最小的估计.

证　先明确一点: 从字面上看, $c'\beta$ 的线性估计可以包括形如 $a'Y+a_0$ 的非齐次估计. 但要这估计为无偏, 必须 $c'\beta=E_\beta(a'Y+a_0)=a'X\beta+a_0$, 对一切 $\beta\in\mathbb{R}^p$. 特别, 取 $\beta=0$ 得 $a_0=0$. 因此, 可限于考虑齐次线性估计.

现证 $c'\hat{\beta}$ 无偏. 因 $c'\beta$ 可估, 由定理 1.3 及其证明, 知存在 a, 使得 $c=Sa$, 故 $c'\hat{\beta}=a'S\hat{\beta}$. 据式 (1.10), 有 $c'\hat{\beta}=a'X'Y$, 故 $E_\beta(c'\hat{\beta})=a'X'X\beta=a'S\beta=c'\beta$. 因而 $c'\hat{\beta}$ 为 $c'\beta$ 之一无偏估计.

现任取 $c'\beta$ 之一无偏估计 $a'Y$. 由无偏性有 $c'\beta=E_\beta(a'Y)=a'X\beta$ 对一切 β, 故 $a'X=c'$. 而 $c'\hat{\beta}=a'X\hat{\beta}$. 依式 (1.15), 有

$$
\begin{aligned}
\operatorname{var}_{\beta,\sigma}(c'\hat{\beta}) &= a'XS^-X'a\sigma^2=a'P_Xa\sigma^2 \\
&= a'P_XP_Xa\cdot\sigma^2=\sigma^2\left\|P_Xa\right\|^2.
\end{aligned}
$$

另一方面, 有 $\operatorname{var}_{\beta,\sigma}(a'Y)=\sigma^2\left\|a\right\|^2$. 因为对任意 a 有 $\|a\|\geqslant\|P_Xa\|$, 故 $c'\hat{\beta}$ 的方差总不会超过 $a'Y$ 的方差. 而要 $\|a\|=\|P_Xa\|$, 充要条件为 $a\in\mathscr{M}(X)$, 即存在 d, 使得 $a=Xd$. 这时有

$$
c'\hat{\beta}=a'X\hat{\beta}=d'X'X\hat{\beta}=d'S\hat{\beta}=d'X'Y=a'Y.
$$

这证明了 $c'\hat{\beta}$ 为使方差最小的惟一的线性无偏估计. 定理证毕.

这个重要定理奠定了在 GM 模型下, LS 估计的重要地位. 如果有线性模型 (1.8) 且仍假定条件 (1.7) 成立, 但代替 GM 条件 (1.15), 假定

$$
\operatorname{cov}(e)=\sigma^2V,\quad V>0\text{ 已知},\quad 0<\sigma^2<\infty,\sigma^2\text{ 未知},\tag{1.18}
$$

则关于线性函数 $c'\beta$ 的可估性定义、LS 估计等, 并无改变. 因这些只依赖于 Y 的均值向量, 而不依赖于其协差阵. 但是, 在 $V\neq I$ 的场合, 可估函数 $c'\beta$ 的 LS 估计 $c'\hat{\beta}$ 一般已不再具有定理 1.4 中所描述的性质. 事实上, 上述模型不难转化到 GM 模型: 作变换 $Z=V^{-1/2}Y$, 则有

$$
Z=\tilde{X}\beta+\tilde{e},
$$

其中 $\tilde{X}=V^{-1/2}X$, $\tilde{e}=V^{-1/2}e$. 而

$$
\operatorname{cov}(\tilde{e})=V^{-1/2}\sigma^2VV^{-1/2}=\sigma^2I,
$$

故对 Z 而言, GM 条件满足. 若 $c'\beta$ 可估, 在 Z 模型求 β 之任一 LS 估计, 为

$$
\tilde{\beta}=(X'V^{-1}X)^-X'V^{-1/2}Z=(X'V^{-1}X)^-X'V^{-1}Y.\tag{1.19}
$$

因而得到 (据 GM 定理) 在一切线性无偏估计类中, 惟一的无偏方差最小估计为

$$c'\tilde{\beta} = c'(X'V^{-1}X)^{-}X'V^{-1}Y. \tag{1.20}$$

这个估计在文献中常称为 “最佳线性无偏估计” (best linear unbiased estimate, 简记为 BLUE). 它与 $c'\beta$ 的在 Y 模型下的 LS 估计 $c'(X'X)^{-}X'Y$ 一般不同. 有时, 也称式 (1.20) 为 GM 估计. 于是, 在 GM 条件下, GM 估计重合于 LS 估计.

(五) σ^2 的估计 假定线性模型 (1.8) 满足 GM 条件 (1.7) 与 (1.15), 于是有估计 σ^2 的问题. 任取 β 之一 LS 估计 $\hat{\beta}$(在估计 σ^2 时, 不必要求 β 可估). 称

$$\delta_i = Y_i - x_i'\hat{\beta}\,(i = 1, \cdots, n)$$

为残差. 易见它们不依赖于 $\hat{\beta}$ 的选择 (因为 $x_i'\beta$ 为可估, 而 $x_i'\hat{\beta}$ 为其 LS 估计. 前已指出, 它与 $\hat{\beta}$ 的选择无关). 通常, 以 $\delta_1, \cdots, \delta_n$ 的平方和除以适当常数 (使之成为无偏的) 作为 σ^2 的估计, 称之为 “基于残差平方和的估计”. 记 $\delta = (\delta_1, \cdots, \delta_n)'$, 有 $\delta = Y - X\hat{\beta} = Y - XS^{-}X'Y = (I - P_X)Y$, 而 P_X, 因而 $I - P_X$ 为对称幂等阵, 故

$$\|\delta\|^2 = Y'(I - P_X)'(I - P_X)Y = Y'(I - P_X)Y. \tag{1.21}$$

注意到 $(I - P_X)X = O$(据式 (1.3)), 有

$$(I - P_X)Y = (I - P_X)X\beta + (I - P_X)e = (I - P_X)e. \tag{1.22}$$

因此由式 (1.21) 又有

$$\|\delta\|^2 = e'(I - P_X)e. \tag{1.23}$$

利用式 (1.7) 及式 (1.15), 由式 (1.23) 得

$$\begin{aligned} E\|\delta\|^2 &= \sigma^2\mathrm{tr}(I - P_X) = \sigma^2\mathrm{tr}(I) - \sigma^2\mathrm{tr}(P_X) \\ &= n\sigma^2 - \sigma^2\mathrm{rk}(P_X) = (n - r)\sigma^2, \quad r = \mathrm{rk}(X), \end{aligned}$$

故知

$$\hat{\sigma}^2 = \|\delta\|^2 / (n - r) \tag{1.24}$$

为 σ^2 之一无偏估计. 据式 (1.23) 及 (1.5), 有

$$\|\delta\|^2 = \sum_1^n e_i^2 - \sum_{i=1}^r \left(\sum_{j=1}^n p_{ij}e_j\right)^2, \tag{1.25}$$

其中

$$\sum_{k=1}^n p_{ik}p_{jk} = \delta_{ij}, \quad 当\ i, j = 1, \cdots, r. \tag{1.26}$$

表达式 (1.25) 对研究 $\hat{\sigma}^2$ 的大样本性质很重要, 见第三章.

§1.2 判决函数与容许性

(一) 统计判决理论的基本概念 设 X 为随机元, 取值于空间 $\mathscr{X}$. 通常, X 是 n 维 ($n \geqslant 1$) 的随机向量, 而 $\mathscr{X}$ 是 $\mathbb{R}^n$ 或其一 Borel 子集. 取定一个由 $\mathscr{X}$ 的某些子集构成的 σ 域 $\mathscr{B}$, 则可测空间 $(\mathscr{X}, \mathscr{B})$ 称为 X 的样本空间. 当 $\mathscr{X}$ 为 $\mathbb{R}^n$ 或其 Borel 子集时, $\mathscr{B}$ 总取为 $\mathscr{X}$ 的一切 Borel 子集构成的 σ 域. 这时样本空间 $(\mathscr{X}, \mathscr{B})$ 称为是欧氏的.

X 的概率分布就是 $\mathscr{B}$ 上的概率测度 P. 在统计问题中, P 不是完全已知, 而是依赖于参数 θ, θ 取值于某集合 Θ, 称为参数空间. 通常, Θ 为 m 维 ($m \geqslant 1$) 欧氏空间或其 Borel 子集. X 的分布族记为 $(P_\theta, \theta \in \Theta)$.

设对 X 进行观察而得到样本 x. 统计判决问题, 就是要依据 x 作出某种决定或采取某种行动. 例如, 设问题是要估计 θ 的某一函数 $g(\theta)$ 之值, 则采取一个决定或行动, 意味着依据样本 x 而给出一个值 $T(x)$, 作为 $g(\theta)$ 之估计值. 在一个确定的问题中, 所能采取的行动的全体 A 一般是确定的. 以上面估计 $g(\theta)$ 为例, 可取 $A = \{g(\theta) : \theta \in \Theta\}$, 或取 A 为包含 $\{g(\theta) : \theta \in \Theta\}$ 之适当集合. A 称为行动空间. 在大多数情形下, A 总是欧氏空间或其 Borel 子集.

为评价和比较所采取的行动的后果, 引进损失函数 $L(\theta, a)$, 定义于 $\theta \in \Theta$ 和 $a \in A$. 它表示当参数值为 θ 而采取行动 a 时所遭受的损失, 损失函数是非负的, 且要满足一定的可测性条件. 在 A 中引进由其某些子集构成的 σ 域 $\mathscr{B}_1$, 要求对任意固定的 $\theta \in \Theta$, $L(\theta, a)$ 作为 $a \in A$ 的函数, 是 Borel 可测的. 在有些问题中还需在 Θ 中引进由其某些子集构成的 σ 域 $\mathscr{B}_2$. 这时要求 $L(\theta, a)$ 作为 $(\Theta \times A, \mathscr{B}_2 \times \mathscr{B}_1)$ 上的函数为 Borel 可测.

样本空间和分布族 $\{\mathscr{X}, \mathscr{B}, P_\theta, \theta \in \Theta\}$、行动空间 A 及损失函数 $L(\theta, a)$, 称为统计判决问题的三要素. 给定一统计判决问题就等于给定这三个要素.

定义于 $(\mathscr{X}, \mathscr{B})$ 而取值于 $(A, \mathscr{B}_1)$ 的可测映射 $\delta(x)$, 称为判决函数, 或更确切地称为非随机化的判决函数. 本书将只涉及这种判决函数. 判决函数是统计判决问题的 "解" 的形式: 因为根据 $\delta(x)$, 一旦有了样本 x, 就从 A 中挑出 $\delta(x)$ 作为所采取的行动, 当参数值为 θ 时, 使用判决函数 δ 所遭受的平均损失为

$$R(\theta, \delta) = E_\theta L(\theta, \delta(X)) = \int_{\mathscr{X}} L(\theta, \delta(x)) dP_\theta(x), \quad \theta \in \Theta,$$

称为 δ 的风险函数. 在目前尚称流行的理论中, 两个判决函数的优劣比较, 全基于其风险函数. 因此, 自然地称两判决函数 δ_1 和 δ_2 等价, 若 $R(\theta, \delta_1) \equiv R(\theta, \delta_2)$ 于 Θ 上. 称 δ_1 一致优于 δ_2, 若 $R(\theta_1 \delta_1) \leqslant R(\theta, \delta_2)$ 对任意 $\theta \in \Theta$, 且 δ_1 与 δ_2 不等价. 又若存在一判决函数 δ^*, 使得对任意判决函数 δ, 或者有 δ 等价于 δ^*, 或者有 δ^* 一致

优于 δ, 则称 δ^* 为所给判决问题的一致最优解. 在这种情形下, 采用 δ^* 有足够的理由. 然而, 在一切有实用意义的问题中, 这种一致最优解几乎不存在, 因此必须考虑其他较为松一些的标准, 以便在这种标准下可挑出最优的判决函数. 下段中谈到的 Bayes 解就是一个重要的例子. 不过, 一致最优准则虽没有实用价值, 但有一个与之联带的概念, 却是极为重要且构成本书第四章的主题, 那就是判决函数的可容许性. 简而言之, 若 δ 为一判决函数, 而不存在一致优于 δ 的判决函数, 则称 δ 为可容许的, 否则为不可容许的. 显然, 可容许性与样本分布族及损失函数的选择有关. 又, 可能在问题中已对所允许使用的判决函数作了限制, 例如, 在点估计中, 限制使用无偏估计、线性估计之类. 一般, 设 $\mathscr{D}$ 为允许使用的判决函数类, 而 $\delta \in \mathscr{D}$. 称 δ 为 $\mathscr{D}$ 可容许的, 若不存在 $\delta_1 \in \mathscr{D}$, 使得 δ_1 一致优于 δ. 显然, 判决函数 δ 是否可容许, 与所限制的类 $\mathscr{D}$ 有关.

以上就是 A. Wald 在 20 世纪 40 年代末期所发表的统计判决理论的极简略的纲要. Wald 创立这个理论的目的, 是为了要把形形色色的统计问题归纳到一个统一的模式内. 30 年来的统计发展史大体上证明了这个想法在实质上并未实现. 但是, 这个理论对近代统计的发展却产生了重大的影响. 从理论的角度看, 它提出了不少新问题以至新的研究方向, 容许性就是其中之一. 从实际的角度看, 它把统计问题的解看成一种行动, 通过分析行动的后果 —— 损失, 使问题的提法及其解更能适合特定的应用. 而在此以前, 人们只考虑统计推断, 在这里, 问题带有一般的性质, 而不考虑所作推断在种种特定情形下可能招致的后果.

(二) Bayes 解 沿用前面的记号. 在 $(\Theta, \mathscr{B}_2)$ 上引进概率测度 H, 它称为 θ 的"先验分布". 设判决函数 δ 的风险函数为 $R(\theta, \delta)$, 称

$$R_H(\delta) = \int_{\Theta} R(\theta, \delta) dH(\theta) \tag{1.27}$$

为 δ 的 Bayes 风险.

若判决函数 δ_H 满足条件: $R_H(\delta_H) \leqslant R_H(\delta)$ 对任意判决函数 δ 成立, 则称 δ_H 为在先验分布 H 之下的一个 Bayes 解.

从判决函数的优良性准则的观点看, 先验分布 H 不过是提供了一个"权", 借以对风险函数 $R(\theta, \delta)$ 进行加权平均, 而获得用以比较的标准 $R_H(\delta)$. 然而, 从实用的观点看, H 表达了在观察到样本 X 之前对 θ 的了解. 因而它的选择应与这一了解一致. 这正是 H 获得"先验分布"这名称的由来. 从统计学中很有影响的"Bayes 学派"的观点看, 在任意统计问题中, 先验分布都是一个必不可少的要素. 如果没有客观的依据去确定这个分布, 则也可以用主观的考虑去确定之. 这种观点和实践在统计学界引起了争论, 涉及许多带哲学性的统计基础问题, 在此我们没有必要涉及这些问题, 因为在本书中, Bayes 方法只是作为一种理论论证的工具, 先验分布的选择是根据论证的需要而不是与实际是否符合的考虑.