

单位代码: 10183

学 号: 20010311016

吉林大学博士学位论文

赵世舜

分类号: 0212

矩阵加权估计及 James-Stein 估计的再研究

Matrix-Weighted Combined Estimator and Further
Study of James-Stein Estimator

赵世舜

指导教师姓名 史宁中 教授 吉林大学

专业名称 概率论与数理统计

论文答辩日期 2006年12月6日

授予学位日期 _____

答辩委员会主席 王兆军

论文评阅人 张忠占 杨晓云 寇新

二零零六年十月

吉林大学博士学位论文原创性声明

本人郑重声明：所呈交学位论文，是本人在指导教师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：赵世舜

日期：二零零六年十月八日

致 谢

将最诚挚的感谢献给我的导师——史宁中教授。导师身兼数职，百忙之中却始终关注着我的论文写作，总是在我最需要的时候给予我悉心的指导和热情的鼓励。他广阔的视野和理性的穿透力使我一次次拨开思索的迷雾。他严谨的治学态度、高屋建瓴的学者风范以及宽宏坦荡的高洁人品无不是我的人生典范，而导师的包容力更是令我折服并受用终生。寥寥数语无法表达对导师的感谢和崇敬之情，我将继续努力，不负导师的期望。

衷心感谢宋立新教授，感谢他多年来在学业上、工作中和生活对我的关怀、鼓励与指导。在本人撰写论文期间，宋老师非常关心论文进展情况，还不时予以指导。作为我的硕士生指导教师及学术启蒙人，宋老师以其渊博的学识、宽广的胸襟自始至终深深地感染着我，并时刻激励着我克服各种困难、不断前进。

本文写作过程中，陶建副教授给予了无私的帮助和指导，在这里我要向他表示诚挚的谢意。

衷心感谢杨晓云教授和王德辉副教授，感谢他们对我各方面的指导与鼓励。

衷心感谢郭建华教授、张宝学教授、高魏教授对本论文提出的宝贵建议和热情帮助。

衷心感谢院领导以及赖民、董志山、郑术蓉、宋海燕、胡国荣、朱复康等各位老师和同窗对我的热情帮助和鼓励。

衷心感谢我的父母和家人，他们为我倾注了太多的心血，没有他们的栽培与教导就没有我的成长。

在此，我要特别感谢我的爱人，感谢她在生活中乃至学业上给予我的极大支持与帮助。每当我在学业上受到挫折时，总能得到她的安慰和鼓励，使我克服困难，完成学业。

最后，向所有关心我、爱护我和给予我帮助的人再一次致以诚挚的谢意！

提 要

本文主要讨论了矩阵加权形式的估计并进一步研究了 James-Stein 估计和 Stein 提出的估计的性质。论文大致分为三部分：第一部分将 Green and Strawderman(1991) (独立条件下) 以及 Judge and Mittelhammer(2004) (相关条件下) 提出的加权组合估计量推广到了更一般的矩阵加权形式的估计量, 讨论了矩阵加权组合估计和原来的组合估计的风险之间的关系, 论述了矩阵加权估计的优良性与实用性。第二部分, 根据第一部分的思想 and 形式, 我们把 James-Stein 估计推广到矩阵形式, 同时比较了它和 James-Stein 估计的风险之间的关系。第三部分, 我们初步的讨论了 Stein(1956) 提出的估计与 James-Stein 估计之间的关系, 并论述了 Stein(1956) 提出的估计的优良性。

目 录

前 言	1
第一章 独立情况的矩阵加权组合估计量.....	6
§1.1 引 言	6
§1.2 矩阵的导出.....	6
§1.3 未知参数的估计.....	14
§1.4 风险的数值模拟比较.....	15
第二章 相关情况的矩阵加权组合估计量.....	19
§2.1 引 言	19
§2.2 矩阵的导出.....	19
§2.3 未知参数的估计.....	26
§2.4 风险的数值模拟比较.....	27
§2.5 实例比较.....	33
2.5.1 回归模型以及估计量的给出.....	33
2.5.2 数值模拟比较.....	34
第三章 矩阵形式的 JAMES-STEIN 估计.....	40
§3.1 引 言	40
§3.2 研究的动机和矩阵的导出.....	40
§3.3 风险的数值模拟比较.....	45
第四章 对 STEIN 估计的进一步研究.....	47
§4.1 引 言	47
§4.2 $\hat{\delta}^S(a, b)$ 和 $\hat{\delta}(a)$ 关系的讨论.....	47
§4.3 数值模拟比较.....	51
结论与下一步的研究问题.....	54
参考文献	56

攻博期间发表的学术论文.....	63
中文摘要	1
ABSTRACT.....	9

前 言

对于 p 维正态分布, 当 $p = 1$ 时, Hodges and Lehmann (1950), Girshick and Savage (1951) 及 Blyth (1951) 分别用不同的方法证明了均值 θ 的通常的估计量, 即样本均值, 在平方损失下是可容许的。当 $p = 2$ 时, Stein (1956) 证明了样本均值在平方损失下是可容许的, 同时, 也得出了令统计学家非常惊讶的结果: 设 X 表示 p 维正态随机向量, $E(X) = \theta$, $\text{cov}(X, X) = I$ 。则当 $p \geq 3$ 时, X 不再是可容许的了。当 a 充分小, b 充分大时, 估计量 $(1 - \frac{a}{b + X'X})X$ 虽然是有偏的, 但它的风险小于 X 的风险。随后, James and Stein (1961) 给出了完全控制估计量 X 的 James-Stein 估计 $\hat{\delta}^{JS} = (1 - \frac{p-2}{X'X})X$, 并讨论了估计量 $\hat{\delta}(a) = (1 - \frac{a}{X'X})X$, 指出当 $a = p - 2$ 时, $\hat{\delta}(a) = \hat{\delta}^{JS}$ 的风险达到最小。

自 Stein (1956) 以来, 掀起了各国统计学家研究容许性的热潮。Stein (1964) 讨论了均值未知的条件下, 正态分布的方差估计量的不可容许性。Brown (1966) 证明了当维数 $p \geq 3$ 时, 在凸损失意义下, 位置参数的最优同变估计是不可容许的。

同时, 许多统计学家对 James-Stein 估计 $\hat{\delta}^{JS} = (1 - \frac{p-2}{X'X})X$ 进行了讨论。Srivastava and Bilodeau (1989) 讨论了椭球分布的 James-Stein 估计。Brandwein and Strawderman (1990, 1991) 讨论了球对称情况下的 James-Stein 估计及其一般形式。Stein (1981),

Brandwein and Strawderman(1990)给出了 James-Stein 估计的风险一致优于 X 的风险的较为简单的证明。Ralescu, Brandwein and Strawderman(1992)讨论了非正态球对称分布的位置参数族的 James-Stein 估计。还有很多工作也可以参看 Susarla(1976), Moore and Brook(1978), Alam and Hawkes(1979), Phillips(1984), Ohtani(1999)等。

统计学家也对 James-Stein 估计进行了很多的修正与改进。Branchik(1970)关于 James-Stein 估计给出了非常好的改进,得到了

风险比 $\hat{\delta}^{JS}$ 的风险更小的统计量 $\hat{\delta}_+^{JS} = (1 - \frac{(p-2)\sigma^2}{X'X})_+ X$, 这里

$(a)_+ = \begin{cases} a, & a \geq 0 \\ 0, & a < 0 \end{cases}$ 。虽然,它还是不可容许的,但 Efron and Morris

(1973a)证明了 $\hat{\delta}_+^{JS}$ 是接近于 Bayes 规则的,同时提出难以寻出控制

$\hat{\delta}_+^{JS}$ 的估计量。直到 20 年后, Shao and Strawderman(1994)提出了

风险比 $\hat{\delta}_+^{JS}$ 的风险更小的估计量。另外 Berry(1994)用 stein 形式

的方差估计来改进多元正态均值的 James-Stein 估计。相应的工作还可参见 Oman(1983), Li and Bhoj(1988), Kubokawa(1991)

人们也把 James-Stein 估计用于很多领域,例如经常用于回归分析领域。Ullah and Ullah(1973)在回归模型中引入了所谓的重 K 类估计方法改进最小二乘估计量。Ohtani(1993)在错误指定 (misspecified) 线性回归模型中比较了 James-Stein 估计和 James-Stein 估计的正部,导出了 James-Stein 估计的正部的预测风险的精确表达式,并给出了在错误指定条件下, James-Stein 估计的正部优于 James-Stein 估计的充要条件。Srivastava and

Srivastava(1993)讨论了回归模型系数的 James-Stein 估计在 Pitman 准则下的性质。Ohtani(1996)讨论了在错误指定模型中 James-Stein 估计的改进。其它的工作还可参见 Ohtani(1998), Namba(2000)。

同时,许多统计学家探讨了 James-Stein 估计与 Bayes 估计及 minimax 估计之间的关系。最近, Maruyama(2003, 2004)利用混合先验给出了正态均值向量的广义 Bayes 估计量。在一定条件下,这些估计是可以控制 James-Stein 估计的。早期的工作还有, Efron and Morris (1972, 1973, 1975)利用 James-Stein 估计提出了经验 Bayes 估计,并比较了它们之间的关系。Baranchik(1973)研究了当变量个数多于 3 时,多元回归问题中极大似然估计的不可容许性,并获得了参数的 James-Stein 估计形式的 minimax 估计。Takada(1979)研究了方差混合的正态分布的一个先验分布族,给出了 James-Stein 估计的正部形式的后验众数,并证明了该后验众数的风险小于极大似然估计的风险。Groenewald and van der Merwe(1979)导出多元回归模型中参数的 James-Stein 估计形式的 minimax 估计的渐近分布,并讨论了这种 minimax 估计和极大似然估计的关系。Li(1982)得到了风险优于 James-Stein 估计的经验 Bayes 估计,这种经验 Bayes 估计是可容许的,并且是渐近最优的。Fan and Fang (1985)讨论了椭球等高分布的位置参数的两阶段 Stein 估计及 Minimax 估计,并给出了一些关于样本均值和 Stein 估计的序贯 Minimax 性质。Maruyama(1999)得到了一类广义 Bayes 估计,证明了它优于 James-Stein 估计,并说明了该类 Bayes 估计中的一个序列弱收敛到 James-Stein 估计的正部。

James-Stein 估计也用到其它的一些地方, Taniguchi and

Hirukawa(2005)给出了基于高斯过程的 James-Stein 估计。Kumar, Tripathi and Misra(2005)讨论了序约束情况下, 正态均值的 James-Stein 形式的估计。Robert, Gene Hwang and Strawderman (1993) 在 Pitman 准则下讨论了 James-Stein 估计, 证明了在该准则下 James-Stein 估计优于估计量 X 。

在众多关于 James-Stein 估计的改进和应用的工作中, 令我们非常感兴趣的是 Green and Strawderman(1991)利用 Stein 估计的思想和形式给出了新的加权组合估计量的方法。与以前的方法不同的是这里的辅助变量是可能有偏的。该文章中, 作者讨论了正态分布下, 这种新的组合估计的优良性。Judge and Mittelhammer (2004) 将 Green and Strawderman(1991) 的结果推广到加权估计中变量之间相关的情形。作者通过数值模拟的方法讨论了加权估计的优良性。同时, 还考虑了估计量的相合性和渐近正态性, 以及估计量的渐近风险。

本文将 Green and Strawderman(1991) (独立条件下) 以及 Judge and Mittelhammer (2004) (相关条件下) 提出的加权组合估计量推广到了更一般的矩阵加权形式。而后, 运用类似的思想 and 形式给出了矩阵形式的 James-Stein 估计量。最后, 我们讨论了 Stein(1956)提出的估计量(Stein 估计)与 James-Stein 估计的关系。

全文安排如下:

第一章 给出了独立情况下, 使得矩阵加权组合估计量的风险达到最小的矩阵, 并且讨论了相应估计量的优良性;

第二章 给出了相关情况下, 使得矩阵加权组合估计量的风险达到最小的矩阵, 并且讨论了相应估计量的优良性;

第三章 讨论了矩阵形式的 James-Stein 估计以及该估计的风险和优良性;

第四章 进一步研究了 Stein(1956)提出的估计量。并比较了该估计量与 James-Stein 估计的关系, 讨论了该估计量的优良性。

第一章 独立情况的矩阵加权组合估计量

§ 1.1 引言

Green and Strawderman(1991)利用 Stein 估计的思想和形式给出了的加权组合估计量的新方法。与以前的方法不同的是这里的辅助变量是可能有偏的,并讨论了这种加权估计量的优良性。这一章中,我将 Green and Strawderman(1991)提出的加权组合估计 $\delta_a^{lc}(X, Y) = aX + (1-a)Y$ 推广到更一般的矩阵加权形式 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 。

本章安排如下: 1.2 节给出使得加权估计量 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 ; 1.3 节讨论了未知参数的估计问题; 1.4 节应用数值模拟比较了两种加权组合估计的风险之间的关系。这里我们需要强调的是,如果没有特殊指明,本论文所讨论的问题都是在平方损失意义下进行研究的。

§ 1.2 矩阵的导出

设 p 维随机向量 $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, 并且 X 与 Y 独立, 其中 σ^2, τ^2 已知, θ, ξ 未知。Green and Strawderman(1991)

给出了使得 $\delta_a^{lc}(X, Y) = aX + (1-a)Y$ 的风险达到最小的 a_0 为

$$\frac{p\tau^2 + \xi'\xi}{p\tau^2 + p\sigma^2 + \xi'\xi}, \quad \text{这时,} \quad \delta_{a_0}^{lc}(X, Y) = X - \frac{p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi} \times$$

$$(X - Y), \quad \text{它的风险为 } MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{(p\tau^2 + \xi'\xi)p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi}.$$

如果把 $\delta_a^{lc}(X, Y)$ 写成 $aIX + (1-a)IY$, 则 $\delta_a^{lc}(X, Y)$ 成为了一个矩阵加权组合估计量。那么, 在矩阵加权估计中, $\delta_{a_0}^{lc}(X, Y)$ 的风险是否还能达到最小, 是我们感兴趣的。下面定理 1.1 给出了使得矩阵加权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 及 $\delta_{A_0}^{mc}(X, Y)$ 的风险。

定理 1.1 设 p 维随机向量 $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, 并且 X 与 Y 独立, 其中 σ^2 , τ^2 已知, θ , ξ 未知。则使得矩阵加权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 为 $(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi')$ 。这时, $\delta_{A_0}^{mc}(X, Y)$ 的风险为

$$MSE(\delta_{A_0}^{mc}(X, Y)) = tr(\sigma^2(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}).$$

为证明定理 1.1, 我们不加证明的给出下面三个引理。

引理 1.1 矩阵 $(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}$ 和 $\tau^2 I + \xi\xi'$ 是可交换的, 即

$$(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1} = (\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi').$$

引理 1.2 设方阵 $A \geq B$ (即 $A - B \geq 0$, $A - B$ 为非负定矩阵),

则 $trA \geq trB$ ，且等号成立 $\Leftrightarrow A = B$ 。

引理 1.3 若方阵 A 使得 $A'A = 0$ ，则 $A = 0$ 。

下面来证明定理 1.1

由 $X \sim N_p(\theta, \sigma^2 I)$ ， $Y \sim N_p(\theta + \xi, \tau^2 I)$ ，并且 X ， Y 独立，我们可以得到

$$\begin{aligned} &MSE(\delta_A^{mc}(X, Y)) \\ &= E[(\delta_A^{mc}(X, Y) - \theta)'(\delta_A^{mc}(X, Y) - \theta)] \\ &= E[(A(X - \theta) + (I - A)(Y - \theta))'(A(X - \theta) + (I - A)(Y - \theta))] \\ &= tr(A'A \cdot \sigma^2) + tr[(I - A)'(I - A)(\tau^2 I + \xi\xi')] \\ &= tr[(\tau^2 I + \sigma^2 I + \xi\xi')A'A - A'(\tau^2 I + \xi\xi') - (\tau^2 I + \xi\xi')A + \tau^2 I + \xi\xi'] \\ &= tr[(\tau^2 I + \sigma^2 I + \xi\xi')A'A] - tr[(\tau^2 I + \sigma^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1} \times \\ &(\tau^2 I + \xi\xi')A] - tr[(\tau^2 I + \sigma^2 I + \xi\xi')A'(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}] + \\ &p\tau^2 + \xi'\xi). \end{aligned}$$

令 $B = \tau^2 I + \sigma^2 I + \xi\xi'$ ， $C = \tau^2 I + \xi\xi'$ ，由引理 2.1 知 B^{-1} 与 C 是可交换的，从而

$$\begin{aligned} &MSE(\delta_A^{mc}(X, Y)) \\ &= tr[B(A'A - A'B^{-1}C - CB^{-1}A + B^{-1}CCB^{-1})] + tr[C(I - CB^{-1})] \\ &= tr[B(A - B^{-1}C)'(A - B^{-1}C)] + tr(\sigma^2 B^{-1}C). \end{aligned}$$

由 B 为正定矩阵，可知存在非奇异矩阵 Q 使得 $B = QQ'$ 。因此

有

$$MSE(\delta_A^{mc}(X, Y)) = \text{tr}[(A - B^{-1}C)Q]'(A - B^{-1}C)Q + \text{tr}(\sigma^2 B^{-1}C).$$

又 $\text{tr}[(A - B^{-1}C)Q]'(A - B^{-1}C)Q \geq 0$, 由引理 1.2, 可知当且仅当 $((A - B^{-1}C)Q)'(A - B^{-1}C)Q = 0$ 时, $\delta_A^{mc}(X, Y)$ 风险才能达到最小。

又由引理 1.3, 可知由 $((A - B^{-1}C)Q)'(A - B^{-1}C)Q = 0$ 即可得到 $(A - B^{-1}C)Q' = 0$, 也就是 $A - B^{-1}C = 0$ 。

因此, 当 $A = A_0 = (\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi')$ 时, $\delta_A^{mc}(X, Y)$ 的风险达到最小, 这时 $\delta_{A_0}^{mc}(X, Y)$ 的风险为

$$MSE(\delta_{A_0}^{mc}(X, Y)) = \text{tr}(\sigma^2 (\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}).$$

定理 1.1 得证。

下面我们将讨论 $\delta_{a_0}^{lc}(X, Y)$ 和 $\delta_{A_0}^{mc}(X, Y)$ 的关系, 通过引理 1.4、定理 1.2 和推论 1.1 来比较 $\delta_{A_0}^{mc}(X, Y)$ 的风险与 $\delta_{a_0}^{lc}(X, Y)$ 的风险之间的关系。

引理 1.4 设 p 阶对称方阵 $A > 0$, $B = B'$, 则存在非奇异矩阵 C , 使

$$CAC' = I,$$

且

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

其中 $\lambda_1, \dots, \lambda_p$ 为 $|B - \lambda A| = 0$ 的根。

该引理的证明可以参见张尧庭、方开泰(1997)。

定理 1.2 设 p 阶对称方阵 $A > 0$, $B > 0$, 则

$$\frac{\text{tr}A^{-1} \cdot \text{tr}B^{-1}}{\text{tr}A^{-1} + \text{tr}B^{-1}} \geq \text{tr}(A + B)^{-1}.$$

证明: 由 $A > 0$, $B > 0$, 及引理 1.4 可知

$$A^{-1} = (C^{-1}(C')^{-1})^{-1} = C'C, \quad B^{-1} = (C^{-1}\Lambda(C')^{-1})^{-1} = C'\Lambda^{-1}C,$$

且

$$(A + B)^{-1} = (C^{-1} \cdot (C')^{-1} + C^{-1}\Lambda(C')^{-1})^{-1} = C'(I + \Lambda)^{-1}C.$$

因此,

$$h_p = \frac{\text{tr}A^{-1} \cdot \text{tr}B^{-1}}{\text{tr}A^{-1} + \text{tr}B^{-1}} = \frac{\text{tr}(C'C) \cdot \text{tr}(C'\Lambda^{-1}C)}{\text{tr}(C'C) + \text{tr}(C'\Lambda^{-1}C)}.$$

令 $T = CC'$, $t_i > 0$ 表示 T 的对角线元素, $\lambda_i > 0$ 为 Λ 的对角线元素,

$i = 1, 2, \dots, p$, 则有

$$h_p = \frac{\text{tr}T \cdot \text{tr}(\Lambda^{-1}T)}{\text{tr}T + \text{tr}(\Lambda^{-1}T)} = \frac{\sum_{i=1}^p t_i \cdot \sum_{i=1}^p \frac{t_i}{\lambda_i}}{\sum_{i=1}^p t_i + \sum_{i=1}^p \frac{t_i}{\lambda_i}},$$

且

$$l_p = \text{tr}(A + B)^{-1} = \text{tr}(C'(I + \Lambda)^{-1}C) = \text{tr}((I + \Lambda)^{-1}T)$$

$$= \sum_{i=1}^p \frac{t_i}{1 + \lambda_i}.$$

故要证明定理，只需要证明：

$$\frac{\sum_{i=1}^p t_i \cdot \sum_{i=1}^p \frac{t_i}{\lambda_i}}{\sum_{i=1}^p t_i + \sum_{i=1}^p \frac{t_i}{\lambda_i}} \geq \sum_{i=1}^p \frac{t_i}{1 + \lambda_i}.$$

下面我们用数学归纳法证明该不等式。

易知，当 $p=1$ 时，等号成立。当 $p=2$ 时，

$$h_2 = \frac{(t_1 + t_2) \left(\frac{t_1}{\lambda_1} + \frac{t_2}{\lambda_2} \right)}{t_1 + t_2 + \frac{t_1}{\lambda_1} + \frac{t_2}{\lambda_2}} = \frac{(t_1 + t_2)(\lambda_2 t_1 + \lambda_1 t_2)}{\lambda_1 \lambda_2 (t_1 + t_2) + \lambda_2 t_1 + \lambda_1 t_2},$$

$$l_2 = \frac{t_1}{1 + \lambda_1} + \frac{t_2}{1 + \lambda_2} = \frac{t_1 + t_2 + \lambda_2 t_1 + \lambda_1 t_2}{(1 + \lambda_1)(1 + \lambda_2)}.$$

令 $a = t_1 + t_2$ ， $b = \lambda_2 t_1 + \lambda_1 t_2$ ，则

$$h_2 = \frac{ab}{\lambda_1 \lambda_2 a + b}, \quad l_2 = \frac{a + b}{(1 + \lambda_1)(1 + \lambda_2)}.$$

从而有

$$h_2 - l_2 = \frac{(\lambda_1 + \lambda_2)ab - \lambda_1 \lambda_2 a^2 - b^2}{(\lambda_1 \lambda_2 a + b)(1 + \lambda_1)(1 + \lambda_2)}.$$

由于

$$\begin{aligned} & (\lambda_1 + \lambda_2)ab - \lambda_1 \lambda_2 a^2 - b^2 \\ &= -(\lambda_1 a - b)(\lambda_2 a - b) \\ &= -(\lambda_1(t_1 + t_2) - \lambda_2 t_1 - \lambda_1 t)(\lambda_2(t_1 + t_2) - \lambda_2 t_1 - \lambda_1 t) \\ &= (\lambda_1 - \lambda_2)^2 t_1 t_2 \geq 0, \end{aligned}$$

因此, 当 $p = 2$ 时, 不等式成立.

假设 $p = k$ 时, 不等式成立, 即

$$h_k = \frac{\sum_{i=1}^k t_i \cdot \sum_{i=1}^k \frac{t_i}{\lambda_i}}{\sum_{i=1}^k t_i + \sum_{i=1}^k \frac{t_i}{\lambda_i}} \geq l_k = \sum_{i=1}^k \frac{t_i}{1 + \lambda_i}.$$

当 $p = k + 1$ 时, 令 $a = \sum_{i=1}^k t_i$, $b = \sum_{i=1}^k \frac{t_i}{\lambda_i}$, $c = \frac{a}{b}$, 则有

$$\begin{aligned} h_{k+1} &= \frac{(\sum_{i=1}^k t_i + t_{k+1}) \cdot (\sum_{i=1}^k \frac{t_i}{\lambda_i} + \frac{t_{k+1}}{\lambda_{k+1}})}{\sum_{i=1}^k t_i + t_{k+1} + \sum_{i=1}^k \frac{t_i}{\lambda_i} + \frac{t_{k+1}}{\lambda_{k+1}}} = \frac{(a + t_{k+1}) \cdot (b + \frac{t_{k+1}}{\lambda_{k+1}})}{a + t_{k+1} + b + \frac{t_{k+1}}{\lambda_{k+1}}} \\ &= \frac{(a + t_{k+1}) \cdot (\frac{a}{c} + \frac{t_{k+1}}{\lambda_{k+1}})}{a + t_{k+1} + \frac{a}{c} + \frac{t_{k+1}}{\lambda_{k+1}}}, \\ l_{k+1} &= \sum_{i=1}^k \frac{t_i}{1 + \lambda_i} + \frac{t_{k+1}}{1 + \lambda_{k+1}} \\ &\leq \frac{ab}{a + b} + \frac{t_{k+1}}{1 + \lambda_{k+1}} = \frac{a}{1 + c} + \frac{t_{k+1}}{1 + \lambda_{k+1}}. \end{aligned}$$

因此, 由 $p = 2$ 时的证明知, 当 $p = k + 1$ 时, 不等式成立.

定理 1.2 得证.

推论 1.1

1. 当 $p = 1$ 时, $MSE(\delta_{A_0}^{mc}(X, Y)) = MSE(\delta_{a_0}^{lc}(X, Y))$;
2. 当 $p \geq 2$ 时, $MSE(\delta_{A_0}^{mc}(X, Y)) \leq MSE(\delta_{a_0}^{lc}(X, Y))$.

证明:

$$\begin{aligned} MSE(\delta_{A_0}^{mc}(X, Y)) &= tr(\sigma^2(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}) \\ &= tr[(\tau^2 I + \xi\xi')^{-1} + (\sigma^2 I)^{-1}]^{-1} \\ &= tr(A + B)^{-1}, \end{aligned}$$

且 $MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{(p\tau^2 + \xi'\xi)p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi} = \frac{trA^{-1} \cdot trB^{-1}}{trA^{-1} + trB^{-1}}$.

又 $\tau^2 I + \xi\xi' > 0$, $\sigma^2 I > 0$, 因此, 由定理 1.2 知推论 1.1 成立。

注释 1.1 注意到

$$(\tau^2 I + \sigma^2 I + \xi\xi')^{-1} = (\tau^2 I + \sigma^2 I)^{-1} - \frac{(\tau^2 I + \sigma^2 I)^{-1} \xi\xi' (\tau^2 I + \sigma^2 I)^{-1}}{1 + \xi' (\tau^2 I + \sigma^2 I)^{-1} \xi},$$

参见方开泰(1989)。于是

$$\begin{aligned} A_0 &= (\tau^2 I + \sigma^2 I + \xi\xi')^{-1} (\tau^2 I + \xi\xi') \\ &= (\tau^2 + \sigma^2)^{-1} (I - \frac{\xi\xi'}{\tau^2 + \sigma^2 + \xi'\xi}) (\tau^2 I + \xi\xi'), \end{aligned}$$

从而

$$\begin{aligned} A_0 \xi &= (\tau^2 + \sigma^2)^{-1} (I - \frac{\xi\xi'}{\tau^2 + \sigma^2 + \xi'\xi}) (\tau^2 I + \xi\xi') \xi \\ &= \frac{\tau^2 + \xi'\xi}{\tau^2 + \sigma^2 + \xi'\xi} \xi, \end{aligned}$$

即, $\frac{\tau^2 + \xi'\xi}{\tau^2 + \sigma^2 + \xi'\xi}$ 和 ξ 分别为 A_0 的特征根和特征向量。

§ 1.3 未知参数的估计

在 § 1.2 我们给出了使得 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 。但是，注意到这里 ξ 的未知性，因此需要对它进行估计。

令 $X_i \stackrel{i.i.d.}{\sim} N_p(\theta, \sigma^2 I)$, $Y_i \stackrel{i.i.d.}{\sim} N_p(\theta + \xi, \tau^2 I)$, $i = 1, 2, \dots, n$, 并且, X_i 与 Y_i 相互独立。这时,

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)\right] = p\tau^2 + p\sigma^2 + \xi'\xi,$$

同时,

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)'\right] = \tau^2 I + \sigma^2 I + \xi\xi'.$$

因此，我们可以用下面两个估计量来估计 $\delta_{A_0}^{lc}(X, Y)$ 和 $\delta_{A_0}^{mc}(X, Y)$,

$$\begin{aligned} \hat{\delta}^{lc}(X, Y) &= \frac{1}{n} \sum_{i=1}^n \hat{\delta}_i^{lc}(X, Y) \\ &= \frac{1}{n} \sum_{i=1}^n \left[X_i - \frac{p\sigma^2}{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (X_i - Y_i) \right] \\ &= \bar{X} - \frac{p\sigma^2}{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}), \end{aligned}$$

$$\begin{aligned} \hat{\delta}^{mc}(X, Y) &= \frac{1}{n} \sum_{i=1}^n \hat{\delta}_i^{mc}(X, Y) \\ &= \frac{1}{n} \sum_{i=1}^n \{X_i - \sigma^2 [\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)']^{-1} (X_i - Y_i)\} \\ &= \bar{X} - \sigma^2 [\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)']^{-1} (\bar{X} - \bar{Y}). \end{aligned}$$

我们将在 § 1.4 利用数值模拟比较 $\hat{\delta}^{lc}(X, Y)$ 和 $\hat{\delta}^{mc}(X, Y)$ 的风险之间的关系。

§ 1.4 风险的数值模拟比较

在这一节中，我们将考虑估计 $\hat{\delta}^{lc}(X, Y)$ 和 $\hat{\delta}^{mc}(X, Y)$ 的风险的数值模拟的比较。观察维数 p 、 $\xi'\xi$ 、 σ^2 和 τ^2 的变化对两种估计的风险的比值 $r = MSE(\hat{\delta}^{mc}(X, Y)) / MSE(\hat{\delta}^{lc}(X, Y))$ 的影响。

表 1.1 给出了各个变量的不同取值，表 1.2-表 1.5 分别给出了 σ^2 、 τ^2 、 p 和 $\xi'\xi$ 的变化对比值 r 的影响。

表 1.1

变量	取 值													
p	4 8													
$\xi'\xi$	0	0.04	0.16	0.36	0.64	1	2	4	10	16	50	110	400	10000
σ^2	0.01 0.05 0.1 0.5 1 5 10 25 50 100													
τ^2	0.1 0.5 1 5 10 25 50 100 1000 5000													

表 1.2

$$\tau^2 = 1 \quad p = 4$$

$\begin{matrix} \sigma^2 \\ \xi\xi \end{matrix}$	0.01	0.05	0.1	0.5	1	5	10	25	100
0	1.0003	1.0041	1.0051	1.0270	1.0485	1.2299	1.4691	>1	>1
0.04	0.9999	1.0021	1.0022	1.0131	1.0388	1.1416	1.2715	>1	>1
0.16	0.9942	0.9799	0.9709	0.9612	0.9768	1.0591	1.1263	>1	>1
0.36	0.9793	0.9249	0.8872	0.8468	0.8689	0.9955	1.0557	>1	>1
0.64	0.9507	0.8345	0.7768	0.7134	0.7571	0.9406	0.9973	1.0829	>1
1	0.9181	0.7592	0.6013	0.5972	0.6528	0.8519	0.9455	1.0353	>1
4	0.8349	0.5522	0.4296	0.2956	0.3180	0.5477	0.7074	0.8688	1.0353
10	0.8548	0.5639	0.4152	0.2220	0.2081	0.3221	0.4572	0.6785	0.9278
16	0.8742	0.6042	0.4468	0.2206	0.1865	0.2446	0.3497	0.5588	0.8671
50	0.9078	0.6688	0.5029	0.2362	0.1830	0.1957	0.2598	0.4314	0.7751
110	0.9725	0.8690	0.7625	0.4273	0.2977	0.1556	0.1487	0.1901	0.4150
400	0.9908	0.9572	0.9185	0.7051	0.5410	0.2390	0.1746	0.1413	0.1949
10000	0.9998	0.9977	0.9969	0.9828	0.9660	0.8471	0.7291	0.5377	0.2692

表 1.3

$$p = 4 \quad \xi'\xi = 2$$

$\begin{matrix} \sigma^2 \\ \tau^2 \end{matrix}$	0.1	0.5	1	5	10	25	50	100
0.1	0.1878	0.2190	0.3008	0.6770	0.8260	0.9573	1.0294	1.1081
0.5	0.3478	0.3147	0.3839	0.6879	0.8250	0.9696	1.0253	1.1155
1	0.5150	0.4324	0.4667	0.7231	0.8321	0.9626	1.0385	1.1199
5	0.9570	0.8712	0.8325	0.8408	0.9319	0.9883	1.0397	1.1029
10	0.9899	0.9621	0.9571	0.9330	0.9553	1.0028	1.0367	1.1073
25	0.9991	0.9998	0.9976	0.9988	0.9958	1.0200	1.0370	1.0839
50	0.9998	1.0001	1.0005	1.0050	1.0049	1.0216	1.0472	1.0627
100	1.0001	1.0002	1.0010	1.0044	1.0051	1.0075	1.0240	1.0429
1000	0.9999	0.9999	0.9998	1.0002	1.0001	1.0012	1.0041	1.0048
5000	1.0000	1.0000	1.0000	1.0000	0.9995	1.0005	1.0001	0.9986

表 1.4

 $p = 4 \quad \xi'\xi = 4$

$\begin{matrix} \sigma^2 \\ \tau^2 \end{matrix}$	0.1	0.5	1	5	10	25	50	100
0.1	0.1936	0.1641	0.2040	0.4792	0.6668	0.8582	0.9598	1.0296
0.5	0.2988	0.2236	0.2567	0.5224	0.6776	0.8564	0.9587	1.0397
1	0.4296	0.2956	0.3180	0.5477	0.7074	0.8688	0.9634	1.0353
5	0.8904	0.7275	0.6656	0.7032	0.7756	0.8996	0.9783	1.0342
10	0.9727	0.9014	0.8698	0.8305	0.8684	0.9285	0.9792	1.0383
25	0.9980	0.9891	0.9843	0.9586	0.9423	0.9711	0.9998	1.0436
50	1.0000	0.9988	0.9978	0.9939	0.9951	1.0032	1.0107	1.0466
100	1.0000	1.0000	0.9994	1.0008	1.0006	1.0094	1.0270	1.0306
1000	1.0000	0.9999	1.0005	0.9998	1.0004	1.0032	1.0025	1.0040
5000	1.0000	0.9999	1.0001	1.0002	1.0000	1.0004	0.9999	1.0017

表 1.5

 $p = 8 \quad \xi'\xi = 4$

$\begin{matrix} \sigma^2 \\ \tau^2 \end{matrix}$	0.1	0.5	1	5	10	25	50	100
0.1	0.1148	0.0990	0.1398	0.4528	0.6622	0.8980	1.0344	1.1889
0.5	0.2496	0.1624	0.1991	0.4792	0.6953	0.8993	1.0475	1.2096
1	0.4183	0.2531	0.2680	0.5064	0.6972	0.8987	1.0428	1.2205
5	0.9249	0.7708	0.7054	0.7141	0.8146	0.9599	1.0686	1.2237
10	0.9848	0.9385	0.9060	0.8572	0.8975	0.9811	1.0607	1.2033
25	0.9995	0.9960	0.9937	0.9876	0.9920	1.0231	1.0672	1.1648
50	1.0001	1.0018	1.0025	1.0071	1.0128	1.0303	1.0667	1.1304
100	1.0004	1.0004	1.0011	1.0034	1.0106	1.0213	1.0483	1.0877
1000	1.0000	0.9999	1.0005	0.9998	1.0004	1.0041	1.0071	1.0109
5000	1.0000	0.9999	1.0001	1.0002	1.0000	1.0005	1.0006	1.0016

由模拟结果可以得到如下结论:

1. 由表 1.2 我们可以观察到, 当 $\xi'\xi$ 很小时, 即使 σ^2 很小, 结果也不是很好, 即 $r > 1$ 。但是不论 σ^2 多大, 当 $\xi'\xi$ 增大时, r 都会逐渐变小而后变大, 并且一定会小于 1。同时, 当 $\xi'\xi$ 充分大以后, 不论 σ^2 多大, r 都会趋向于 1。

当 $\xi'\xi \geq 1$ 时, 由表 1.3 至表 1.5 还可以得到,

2. 由表 1.3 我们可以观察到, 随着 σ^2 的增大, r 先减小而后增大, 当 σ^2 很大以后, 结果不是很好, 即 $r > 1$ 。而当 τ^2 增大时, r 先增大而后减少。但当 τ^2 充分大以后, 不论 σ^2 多大, r 都会趋向于 1;

3. 表 1.3 和表 1.5 表明当 p 固定时, 随着 $\xi'\xi$ 增大, 比值 r 会减小, 但 $\xi'\xi$ 的变化不会改变 σ^2 和 τ^2 对 r 的影响;

4. 从表 1.3 至表 1.5 可以看出当 σ^2 和 τ^2 不是很大时, 随着 p 增大, r 会变小; 而当 σ^2 和 τ^2 变得大了一些以后, 随着 p 增大, r 会变大。但同 $\xi'\xi$ 一样, p 值的变化也不会改变 σ^2 和 τ^2 对 r 的影响。

事实上, 在实际问题中 σ^2 和 τ^2 都不是很大, 因此, 当 $\xi'\xi \geq 1$ 时, 我们的方法还是较为实用的。我们将在以后的工作中改进估计量中的不足之处。

第二章 相关情况的矩阵加权组合估计量

§ 2.1 引言

Judge and Mittelhammer(2004)将 Green and Strawderman (1991)的工作推广到了加权估计中变量之间相关的情况。并用了一个回归模型的例子讨论了这种估计的优良性。本章,我们还将采用和第一章类似的做法,将 Judge and Mittelhammer 提出的加权组合估计推广到矩阵加权形式,只是这里的变量是相关的。

本章安排如下: 2.2 节给出使得加权估计量 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 ; 2.3 节讨论了未知参数的估计问题; 2.4 节应用数值模拟来比较两种加权估计的风险之间的关系。2.5 节用 Judge and Mittelhammer (2004) 提出的回归模型的例子进一步比较了两种加权估计的风险。

§ 2.2 矩阵的导出

设 X, Y 分别为 p 维正态随机向量, $E(X) = \theta, E(Y) = \theta + \xi$,

$$\text{cov}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Sigma \\ \Sigma' & \Phi_2 \end{pmatrix}, \text{ 其中 } \Phi_1 \text{ 为 } X \text{ 的协方差矩阵, } \Phi_2 \text{ 为 } Y \text{ 的协方}$$

差矩阵, Σ 为 X 与 Y 协方差矩阵, 且 $\Phi_1, \Phi_2, \Sigma, \theta$ 和 ξ 是未知

的。Judge and Mittelhammer (2004) 给出了使得 $\delta_a^{lc}(X, Y) = aX +$

$(1-a)Y$ 的风险达到最小的 a_0 为 $\frac{tr\Phi_2 + \xi'\xi - tr\Sigma}{tr(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}$, 这时

$$\delta_{a_0}^{lc}(X, Y) = X - \frac{tr\Phi_1 - tr\Sigma}{tr(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}(X - Y),$$

它的风险为

$$MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{tr\Phi_1 tr(\Phi_2 + \xi\xi') - (tr\Sigma)^2}{tr\Phi_1 + tr\Phi_2 + \xi'\xi - 2tr\Sigma}.$$

注释 2.1 在 Judge 和 Mittelhammer 的一文中, 为了要保证

$0 \leq a \leq 1$, 且 $\frac{tr\Phi_1 tr(\Phi_2 + \xi\xi') - (tr\Sigma)^2}{tr\Phi_1 + tr\Phi_2 + \xi'\xi - 2tr\Sigma} \geq 0$, 实际上必须满足:

1. $tr\Phi_1 - tr\Sigma \geq 0$;
2. $tr\Phi_2 + \xi'\xi - tr\Sigma \geq 0$;
3. $tr\Phi_1 + tr\Phi_2 + \xi'\xi - 2tr\Sigma > 0$, $tr\Phi_1 tr(\Phi_2 + \xi\xi') - (tr\Sigma)^2 \geq 0$.

下面的定理 2.1 给出了在一定条件下使得矩阵加权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 及 $\delta_{A_0}^{mc}(X, Y)$ 的风险。

定理 2.1 设 $X \sim N_p(\theta, \Phi_1)$, $Y \sim N_p(\theta + \xi, \Phi_2)$, $cov(X, Y) = \Sigma$,

则当 $B = \Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma'$ 为正定矩阵 ($B > 0$) 时, 使得矩阵加

权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A 为

$A_0 = (\Phi_2 + \xi\xi' - \Sigma)(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}$, 这时, $\delta_{A_0}^{mc}(X, Y)$ 的风险为

$$tr([\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma](\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}).$$

证明: 与定理 1.1 类似可得

$$\begin{aligned} &MES(\delta^{mc}(X, Y)) \\ &= E[(A(X - \theta) + (I - A)(Y - \theta))(A(X - \theta) + (I - A)(Y - \theta))'] \\ &= tr[A'A\Phi_1 + (I - A)'(I - A)(\Phi_2 + \xi\xi') + A'(I - A)\Sigma' + (I - A)'A\Sigma] \\ &= tr[A'A(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma') - A'(\Phi_2 + \xi\xi' - \Sigma') - A(\Phi_2 + \xi\xi' - \Sigma)] \\ &\quad + tr(\Phi_2 + \xi\xi'). \end{aligned}$$

由 $B = \Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma' > 0$, 则 $B = B^{1/2}B^{1/2}$, 且 $B^{1/2}$ 可逆, 从而

$$\begin{aligned} &MES(\delta^{mc}(X, Y)) \\ &= tr[B^{1/2}A'AB^{1/2} - B^{1/2}A'(\Phi_2 + \xi\xi' - \Sigma')B^{-1/2} - B^{-1/2}(\Phi_2 + \xi\xi' - \Sigma)AB^{1/2}] \\ &\quad + tr(\Phi_2 + \xi\xi') \\ &= tr[(AB^{1/2} - (\Phi_2 + \xi\xi' - \Sigma')B^{-1/2})'(AB^{1/2} - (\Phi_2 + \xi\xi' - \Sigma')B^{-1/2})] \\ &\quad + tr((\Phi_2 + \xi\xi') - B^{-1}(\Phi_2 + \xi\xi' - \Sigma)(\Phi_2 + \xi\xi' - \Sigma')). \end{aligned}$$

类似于定理 1.1 的证明, 可得当 $AB^{1/2} = (\Phi_2 + \xi\xi' - \Sigma')B^{-1/2}$ 时, 即 $A = A_0 = (\Phi_2 + \xi\xi' - \Sigma)(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}$ 时, $\delta_A^{mc}(X, Y)$ 的

风险 $MSE(\delta_{A_0}^{mc}(X, Y))$ 达到最小, 其值为

$$\begin{aligned} & MSE(\delta_{A_0}^{mc}(X, Y)) \\ &= tr((\Phi_2 + \xi\xi') - B^{-1}(\Phi_2 + \xi\xi' - \Sigma)(\Phi_2 + \xi\xi' - \Sigma')) \\ &= tr([\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma](\Phi_1 + \Phi_2 + \\ &\xi'\xi - \Sigma - \Sigma')^{-1}). \end{aligned}$$

这时,

$$\delta_{A_0}^{mc}(X, Y) = X - (\Phi_1 - \Sigma)(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}(X - Y).$$

定理 2.1 得证。

下面的定理 2.2 在一定条件下比较了 $MSE(\delta_{A_0}^{mc}(X, Y))$ 和 $MSE(\delta_{a_0}^{lc}(X, Y))$ 之间的关系。

定理 2.2 若 $\Sigma' = \Sigma$, $\Phi_1 - \Sigma > 0$ 及 $\Phi_2 + \xi\xi' - \Sigma > 0$ 成立, 则

$$MSE(\delta_{A_0}^{mc}(X, Y)) \leq MSE(\delta_{a_0}^{lc}(X, Y)).$$

证明: 当满足定理条件时,

$$\begin{aligned} & MSE(\delta_{A_0}^{mc}(X, Y)) \\ &= tr([\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma](\Phi_1 + \Phi_2 + \\ &\xi'\xi - \Sigma - \Sigma')^{-1}) \\ &= tr([\Phi_1(\Phi_2 + \xi\xi') - \Sigma^2](\Phi_1 + \Phi_2 + \xi'\xi - 2\Sigma)^{-1}), \end{aligned}$$

由于 $\Phi_1 - \Sigma > 0$, $\Phi_2 + \xi\xi' - \Sigma > 0$, 根据引理 1.4 可知存在非奇异矩阵 D 使得

$$\Phi_1 - \Sigma = D'D,$$

$$\Phi_2 + \xi\xi' - \Sigma = D'\Lambda D,$$

这里 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_i > 0, i = 1, \dots, p$.

于是,

$$\begin{aligned} \Phi_1(\Phi_2 + \xi\xi') - \Sigma^2 &= (D'D + \Sigma)(D'\Lambda D + \Sigma) - \Sigma^2 \\ &= D'DD'\Lambda D + \Sigma D'\Lambda D + D'D\Sigma, \end{aligned}$$

且

$$\begin{aligned} (\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)^{-1} &= (D'D + D'\Lambda D)^{-1} \\ &= D^{-1}(I + \Lambda)^{-1}(D')^{-1}. \end{aligned}$$

从而,

$$\begin{aligned} &MSE(\delta_{A_0}^{mc}(X, Y)) \\ &= \text{tr}[(\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma)(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}) \\ &= \text{tr}[(\Phi_1(\Phi_2 + \xi\xi') - \Sigma^2)(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)^{-1}] \\ &= \text{tr}[(D'DD'\Lambda D + \Sigma D'\Lambda D + D'D\Sigma)D^{-1}(I + \Lambda)^{-1}(D')^{-1}] \\ &= \text{tr}[(DD'\Lambda(I + \Lambda)^{-1} + (D')^{-1}\Sigma D'\Lambda(I + \Lambda)^{-1} + D\Sigma D^{-1}(I + \Lambda)^{-1})]. \end{aligned}$$

由

$$\text{tr}[(D')^{-1}\Sigma D'\Lambda(I + \Lambda)^{-1}]$$

$$= \text{tr}[(D')^{-1}\Sigma D'\Lambda(I + \Lambda)^{-1}]'$$

$$= \text{tr}[(I + \Lambda)^{-1}\Lambda D\Sigma D^{-1}],$$

可得

$$MSE(\delta_{A_0}^{mc}(X, Y))$$

$$= \text{tr}[(DD'\Lambda(I + \Lambda)^{-1} + D\Sigma D^{-1}(I + \Lambda)^{-1}\Lambda + D\Sigma D^{-1}(I + \Lambda)^{-1}]$$

$$= \text{tr}[(DD'\Lambda(I + \Lambda)^{-1}] + \text{tr}\Sigma.$$

这时

$$\text{tr}\Phi_1 \text{tr}(\Phi_2 + \xi\xi') - (\text{tr}\Sigma)^2$$

$$= [\text{tr}(D'D) + \text{tr}\Sigma][\text{tr}(D'\Lambda D) + \text{tr}\Sigma] - (\text{tr}\Sigma)^2$$

$$= \text{tr}(D'D)\text{tr}(D'\Lambda D) + \text{tr}\Sigma\text{tr}(D'(I + \Lambda)D),$$

$$\text{tr}(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)$$

$$= \text{tr}(D'D + D'\Lambda D) = \text{tr}[D'(I + \Lambda)D],$$

故有

$$\begin{aligned} MES(\delta_{a_0}^{lc}(X, Y)) &= \frac{\text{tr}\Phi_1 \text{tr}(\Phi_2 + \xi\xi') - (\text{tr}\Sigma)^2}{\text{tr}\Phi_1 + \text{tr}\Phi_2 + \xi'\xi - 2\text{tr}\Sigma} \\ &= \frac{\text{tr}(D'D)\text{tr}(D'\Lambda D) + \text{tr}\Sigma\text{tr}(D'(I + \Lambda)D)}{\text{tr}[D'(I + \Lambda)D]} \\ &= \frac{\text{tr}(D'D)\text{tr}(D'\Lambda D)}{\text{tr}[D'(I + \Lambda)D]} + \text{tr}\Sigma. \end{aligned}$$

从而，要证明定理 2.2，只需证明

$$\text{tr}[(DD'\Lambda(I + \Lambda)^{-1}] \leq \frac{\text{tr}(D'D)\text{tr}(D'\Lambda D)}{\text{tr}[D'(I + \Lambda)D]}.$$

由于

$$tr[(DD'\Lambda(I + \Lambda)^{-1})] = tr[D^{-1}\Lambda^{-1}(D')^{-1} + D^{-1}(D')^{-1}]^{-1},$$

和

$$\frac{tr(D'D)tr(D'\Lambda D)}{tr[D'(I + \Lambda)D]} = \frac{tr[D(D')^{-1}]^{-1}tr[D^{-1}\Lambda(D')^{-1}]^{-1}}{tr[D(D')^{-1}]^{-1} + tr[D^{-1}\Lambda(D')^{-1}]^{-1}},$$

并且

$$D^{-1}\Lambda^{-1}(D')^{-1} = (\Phi_2 + \xi\xi' - \Sigma)^{-1} > 0, \quad D^{-1}(D')^{-1} = (\Phi_1 - \Sigma)^{-1} > 0.$$

因此由定理 1.2 得

$$tr[(DD'\Lambda(I + \Lambda)^{-1})] \leq \frac{tr(D'D)tr(D'\Lambda D)}{tr[D'(I + \Lambda)D]}.$$

故定理 2.2 得证。

注释 2.2 在第一章的注释 1.1，我们说明了 ξ 为矩阵 $A_0 = (\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi')$ 的特征向量，同时也可以看出 A_0 是对称正定的。但在这一章变量相关的情况，对于矩阵 $A_0 = (\Phi_2 + \xi\xi' - \Sigma)(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}$ ，只有当 $\Phi_1 - \Sigma > 0$ 及 $\Phi_2 + \xi\xi' - \Sigma > 0$ ，并且 $\Phi_2 + \xi\xi' - \Sigma$ 和 $\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma'$ 是可交换时， A_0 才是对称正定矩阵。而对于 ξ ，只有当 ξ 是 $\Phi_1 - \Sigma$ 和 $\Phi_2 + \xi\xi' - \Sigma$ 的特征向量时， ξ 才能是矩阵 A_0 的特征向量。我们希望在以后的工作中能够在对称正定矩阵中找到使得矩阵加权估计的风险达到最小的矩阵。同时，该矩阵应具有下面的性质： ξ 应为其特征向量。

§ 2.3 未知参数的估计

同第一章类似，由于 ξ ， Φ_1 ， Φ_2 和 Σ 的未知性，我们必须对它们进行估计。

令 $X_i \stackrel{i.i.d.}{\sim} N_p(\theta, \Phi_1)$ ， $Y_i \stackrel{i.i.d.}{\sim} N_p(\theta + \xi, \Phi_2)$ ， $i = 1, 2, \dots, n$ ，并且，

Σ 为 X_i 和 Y_i 的协方差矩阵。这时有，

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)\right] = \text{tr}\Phi_1 + \text{tr}\Phi_2 + \xi'\xi - 2\text{tr}\Sigma,$$

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)'\right] = \Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma'.$$

同时，由于

$$E\left[\frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \bar{X})'(X_i - \bar{X}) - \sum_{i=1}^n (X_i - \bar{X})'(Y_i - \bar{Y})\right)\right] = \text{tr}\Phi_1 - \text{tr}\Sigma,$$

$$E\left[\frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' - \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})'\right)\right] = \Phi_1 - \Sigma.$$

因此，我们可以用下面两个估计量来估计 $\delta_{A_0}^{mc}(X, Y)$ 和 $\delta_{a_0}^{lc}(X, Y)$ 。

$$\begin{aligned} \hat{\delta}^{lc}(X, Y) &= \frac{1}{n} \sum_{i=1}^n \hat{\delta}_i^{lc}(X, Y) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ X_i - \frac{\frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \bar{X})'(X_i - \bar{X}) - \sum_{i=1}^n (X_i - \bar{X})'(Y_i - \bar{Y})\right)}{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (X_i - Y_i) \right\} \end{aligned}$$

$$\begin{aligned}
&= \bar{X} - \frac{n \sum_{i=1}^n (X_i - \bar{X})'(X_i - \bar{X}) - n \sum_{i=1}^n (X_i - \bar{X})'(Y_i - \bar{Y})}{(n-1) \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}), \\
\hat{\delta}^{mc}(X, Y) &= \frac{1}{n} \sum_{i=1}^n \hat{\delta}_i^{mc}(X, Y) \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ X_i - \left[\frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' - \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})' \right) \right] \right. \\
&\quad \left. \times \left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)' \right]^{-1} (X_i - Y_i) \right\} \\
&= \bar{X} - \left[n \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' - n \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})' \right] \left[(n-1) \sum_{i=1}^n (X_i - Y_i) \right. \\
&\quad \left. \times (X_i - Y_i)' \right]^{-1} (\bar{X} - \bar{Y}).
\end{aligned}$$

我们将在 § 2.4 利用数值模拟来比较 $\hat{\delta}^{lc}(X, Y)$ 和 $\hat{\delta}^{mc}(X, Y)$ 的风险之间的关系。

§ 2.4 风险的数值模拟比较

在这一节中，我们将考虑估计 $\hat{\delta}^{lc}(X, Y)$ 和 $\hat{\delta}^{mc}(X, Y)$ 的风险的数值模拟的比较。这里为了简便，只考虑向量维数 $p = 4$ 的情况，事先给定 Φ_1 、 Φ_2 和 Σ 为

$$\Phi_1 = s \begin{pmatrix} c & t & t & t \\ t & c & t & t \\ t & t & c & t \\ t & t & t & c \end{pmatrix}, \Phi_2 = \begin{pmatrix} 1 & t & t & t \\ t & 1 & t & t \\ t & t & 1 & t \\ t & t & t & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{pmatrix}.$$

观察 $\xi'\xi$ 、 t 、 a 、 s 、 c 和 b 的变化对两种估计的风险的比值 $r = MSE(\hat{\delta}^{mc}(X,Y))/MSE(\hat{\delta}^{lc}(X,Y))$ 的影响。在表格中，有一些画斜杠没有值得地方，是由于在这些地方的 b 和 t 值使得矩阵 $\begin{pmatrix} \Phi_1 & \Sigma \\ \Sigma' & \Phi_2 \end{pmatrix}$ 是负定矩阵，因此没法计算值。

表 2.1

$a=0.2 \quad \xi'\xi=4 \quad c=1 \quad s=1$

$b \backslash t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.1	0.2898	0.3919	0.4894	0.6178	0.8254	0.8919	
0.2	0.1964	0.2884	0.3954	0.5794	0.7502	0.8189	0.8870
0.3	0.1217	0.1921	0.2886	0.4925	0.6682	0.7517	
0.5			0.1173	0.2932			
0.70							
0.80							
0.90							
1.00							
1.20							

表 2.2

$a=0.2 \quad \xi'\xi=4 \quad c=1 \quad s=2$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.1	0.3986	0.4986	0.5821	0.7216	0.8247	0.8689	0.9070
0.2	0.3662	0.4704	0.5613	0.7059	0.8222	0.8661	0.9079
0.3	0.3210	0.4333	0.5346	0.6951	0.8151	0.8626	0.9070
0.5	0.2224	0.3545	0.4774	0.6669	0.8039		
0.70			0.3916	0.6243			
0.80			0.3402	0.6024			
0.90				0.5679			
1.00							
1.20							

表 2.3

$a=0.2 \quad \xi'\xi=4 \quad c=2 \quad s=2$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.4727	0.5044	0.5352	0.5799	0.6128	0.6278	0.6404
0.20	0.4717	0.5074	0.5357	0.5819	0.6189	0.6338	0.6480
0.30	0.4688	0.5049	0.5369	0.5866	0.6255	0.6414	0.6546
0.50	0.4622	0.5065	0.5384	0.5956	0.6371	0.6522	0.6675
0.70	0.4524	0.5041	0.5476	0.6107	0.6531	0.6684	
0.80		0.5050	0.5498	0.6166	0.6591	0.6805	
0.90			0.5541	0.6217	0.6700		
1.00			0.5592	0.6328	0.6823		
1.20				0.6611			

表 2.4

$$a = 0.2 \quad \xi'\xi = 4 \quad c = 2 \quad s = 1$$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.3354	0.3735	0.4041	0.4624	0.5031	0.5250	0.5365
0.20	0.2949	0.3342	0.3729	0.4314	0.4775	0.5016	0.5195
0.30	0.2518	0.2971	0.3361	0.4037	0.4574	0.4823	0.5008
0.50	0.1520	0.2009	0.2496	0.3360	0.4023	0.4326	0.4571
0.70			0.1543	0.2513	0.3364		
0.80				0.2026	0.2957		
0.90				0.1516			
1.00							
1.20							

表 2.5

$$a = 0.2 \quad \xi'\xi = 8 \quad c = 2 \quad s = 2$$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.3066	0.3149	0.3157	0.3076	0.2851	0.2698	0.2511
0.20	0.3042	0.3131	0.3196	0.3123	0.2932	0.2803	0.2603
0.30	0.2982	0.3154	0.3208	0.3168	0.2995	0.2863	0.2714
0.50	0.2882	0.3098	0.3242	0.3281	0.3138	0.3041	0.2942
0.70	0.2781	0.3076	0.3205	0.3359	0.3303	0.3245	
0.80		0.3018	0.3243	0.3424	0.3371	0.3305	
0.90			0.3262	0.3472	0.3438		
1.00			0.3326	0.3502	0.3532		
1.20				0.3721			

表 2.6

$a=0.6 \quad \xi'\xi=4 \quad c=1 \quad s=1$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.2988	0.4595	0.6362				
0.20	0.1837	0.2959	0.4609	0.8073			
0.30		0.1860	0.3011	0.6349			
0.50				0.2996	0.6277	0.7993	
0.70					0.3034	0.4588	
0.80					0.1859		
0.90							
1.00							
1.20							

表 2.7

$a=0.6 \quad \xi'\xi=4 \quad c=2 \quad s=1$

$b \setminus t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.2831	0.3248	0.3598	0.4216	0.4689		
0.20	0.2356	0.2810	0.3245	0.3931	0.4485	0.4660	
0.30	0.1876	0.2338	0.2815	0.3592	0.4226	0.4466	0.4680
0.50			0.1856	0.2816	0.3612	0.3948	0.4211
0.70				0.1862	0.2808	0.3264	0.3570
0.80					0.2340	0.2847	0.3229
0.90					0.1869	0.2368	0.2825
1.00						0.1851	
1.20							

表 2.8

$a=0.6 \quad \xi'\xi=4 \quad c=1 \quad s=2$

$b \backslash t$	0.1	0.2	0.3	0.5	0.70	0.80	0.90
0.10	0.3511	0.4579	0.5518	0.6993			
0.20	0.3092	0.4238	0.5277	0.6908	0.8020		
0.30	0.2579	0.3834	0.4971	0.6704	0.8033		
0.50		0.2865	0.4201	0.6385	0.7903	0.8481	0.8994
0.70				0.5890	0.7724	0.8400	0.8981
0.80				0.5554	0.7595	0.8343	
0.90				0.5184	0.7490		
1.00					0.7270		
1.20							

从表 2.1-2.8 的模拟结果可以看出：

1. 随着 t 增大，比值 r 增大，随着 b 增大，比值 r 却是在减小的；
2. 随着 s 增大，比值 r 大多数也是增大的；
3. 随着 c 增大，对于 t 的较小的值，比值 r 是增大的；而对于 t 的较大的值，比值 r 是减小的；
4. 随着 $\xi'\xi$ 的增加，比值 r 是在减小的；
5. 随着 a 的增加，比值 r 呈现出不太规律的变化。

从模拟结果和上面的论述中可知： $\xi'\xi$ 、 t 、 a 、 s 、 c 和 b 的变化会引起 r 很多不同的变化，但 r 始终小于 1，因此我们的方法还是实用的。

§ 2.5 实例比较

2.5.1 回归模型以及估计量的给出

在这一节，我们将用 Judge and Mittelhammer(2004)给出回归模型的实例进一步讨论两种加权估计的风险之间的关系。

对于回归模型

$$\begin{cases} y = X\beta + \varepsilon \\ \varepsilon \sim N_m(0, \sigma^2 I) \end{cases}$$

其中 y 为 m 维观测向量， X 为 $m \times p$ 阶设计阵，秩为 p ， β 为 p 维参数向量。我们这里要估计 p 维参数向量 β 。这时 β 的最小二乘估计为 $\hat{\beta} = (X'X)^{-1} X'y \sim N(\beta, \sigma^2 (X'X)^{-1})$ ，即 $\hat{\beta}$ 为 β 的无偏估计，且 $\hat{\beta}$ 的风险为 $MSE(\hat{\beta}) = \sigma^2 tr(X'X)^{-1}$ 。

同时，Judge and Mittelhammer(2004)给出了 β 的有偏估计量，所谓的 CLS 估计 $\tilde{\beta} = [diag(X'X)]^{-1} X'y = A(X)\hat{\beta}$ ，这里 $A(X) = [diag(X'X)]^{-1} X'X$ ， $diag(X'X)$ 表示 $k \times k$ 对角矩阵。

与上一节类似，可以产生 n 组 β 的估计量 $(\hat{\beta}_i, \tilde{\beta}_i)$ ， $i = 1, 2, \dots, n$ 。这时两种估计量为

$$\hat{\delta}^{lc}(\hat{\beta}, \tilde{\beta}) = \hat{\beta} - \frac{n \sum_{i=1}^n (\hat{\beta}_i - \bar{\beta})' (\hat{\beta}_i - \bar{\beta}) - n \sum_{i=1}^n (\hat{\beta}_i - \bar{\beta})' (\tilde{\beta}_i - \bar{\beta})}{(n-1) \sum_{i=1}^n (\hat{\beta}_i - \tilde{\beta}_i)' (\hat{\beta}_i - \tilde{\beta}_i)} (\hat{\beta} - \bar{\beta}),$$

$$\hat{\delta}^{mc}(\hat{\beta}, \tilde{\beta}) = \bar{\beta} - [n \sum_{i=1}^n (\hat{\beta}_i - \bar{\beta})(\hat{\beta}_i - \bar{\beta})' - n \sum_{i=1}^n (\hat{\beta}_i - \bar{\beta})(\tilde{\beta}_i - \bar{\beta})'] \times$$

$$[(n-1) \sum_{i=1}^n (\hat{\beta}_i - \tilde{\beta}_i)(\hat{\beta}_i - \tilde{\beta}_i)']^{-1} (\hat{\beta} - \bar{\beta}).$$

2.5.2 数值模拟比较

Judge and Mittelhammer (2004) 给出的用于模拟的模型为

$$y_i = X[i, \cdot] \beta + \varepsilon_i = \sum_{j=1}^5 \beta_j X[i, j] + \varepsilon_i, \quad i = 1, 2, \dots, m.$$

模型的矩阵形式为 $y_{m \times 1} = X_{m \times 5} \beta_{5 \times 1} + \varepsilon_{m \times 1}$, 其中 $X_{m \times 5}$ 的第一列全为 1, 剩下的四列的每一行均服从 4 维正态分布 $N_4(\mu, \Omega)$, 这里

$$\mu = (1, 1, 1, 1)', \quad \Omega = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}. \quad \varepsilon_{m \times 1} \text{ 是来自于正态分布}$$

$N_m(\mathbf{1}_{m \times 1}, \sigma^2 I)$, 其中 $\mathbf{1}_{m \times 1} = (1, 1, \dots, 1)'_{1 \times m}$. 与 Judge and Mittelhammer (2004) 类似, 我们将给定 $\beta = (0.1, -0.1, 0.2, -0.2, 0.3)$ 来观察 m, ρ, σ^2 的变化对这两种估计的风险比值 $r = MSE(\hat{\delta}^{mc}(\hat{\beta}, \tilde{\beta})) / MSE(\hat{\delta}^{lc}(\hat{\beta}, \tilde{\beta}))$ 的影响. 然后, 观察 $\beta' \beta$ 的变化对

风险比值 r 的影响。

表 2.9

变量	取 值											
m	15 25 50 100											
$\beta'\beta$	0	0.05	0.45	0.7	1.2	5	10	25	45	140	500	12500
σ^2	0.01 0.0625 1 25 100											
ρ	0.0	0.25	0.5	0.6	0.7	0.8	0.85	0.90	0.95	0.99		

表 2.10

$m = 15 \quad \beta = (0.1, -0.1, 0.2, -0.2, 0.3)$

$\rho \backslash \sigma^2$	0.01	0.0625	1	25	100
0.0	0.3106	0.0971	0.1380	0.4450	0.4090
0.25	0.3983	0.1367	0.1364	0.4553	0.3997
0.5	0.4097	0.1393	0.1364	0.4425	0.3812
0.6	0.3891	0.1218	0.1321	0.4316	0.3705
0.7	0.3333	0.0981	0.1300	0.3994	0.3712
0.8	0.2434	0.0743	0.1231	0.4125	0.3743
0.85	0.1872	0.0589	0.1287	0.4080	0.3864
0.9	0.1280	0.0496	0.1282	0.4323	0.3732
0.95	0.0752	0.0426	0.1633	0.4407	0.4138
0.99	0.0414	0.0751	0.4806	0.7649	0.8036

表 2.11 $m = 25 \quad \beta = (0.1, -0.1, 0.2, -0.2, 0.3)$

$\rho \backslash \sigma^2$	0.01	0.0625	1	25	100
0.0	0.4609	0.1405	0.0992	0.5016	0.4554
0.25	0.5742	0.2094	0.1156	0.4825	0.4304
0.5	0.5898	0.2147	0.1161	0.4404	0.4184
0.6	0.5482	0.1890	0.1101	0.4487	0.4014
0.7	0.4818	0.1513	0.1017	0.4168	0.3892
0.8	0.3836	0.1093	0.0902	0.3915	0.3810
0.85	0.3134	0.0832	0.0871	0.4029	0.3884
0.9	0.2202	0.0654	0.0867	0.4012	0.4044
0.95	0.1303	0.0474	0.1067	0.4123	0.4057
0.99	0.0466	0.0512	0.3007	0.6720	0.7024

表 2.12 $m = 50 \quad \beta = (0.1, -0.1, 0.2, -0.2, 0.3)$

$\rho \backslash \sigma^2$	0.01	0.0625	1	25	100
0.0	0.6318	0.2459	0.0798	0.4814	0.5460
0.25	0.7424	0.3508	0.1026	0.4503	0.5202
0.5	0.7459	0.3523	0.1010	0.4012	0.4917
0.6	0.7383	0.3202	0.0953	0.4022	0.4651
0.7	0.6831	0.2667	0.0828	0.3743	0.4343
0.8	0.5737	0.1884	0.0683	0.3619	0.4022
0.85	0.5021	0.1487	0.0638	0.3483	0.4013
0.9	0.3919	0.1091	0.0564	0.3403	0.3978
0.95	0.2393	0.0654	0.0621	0.3614	0.4375
0.99	0.0714	0.0424	0.1700	0.5962	0.6471

表 2.13 $m = 100 \quad \beta = (0.1, -0.1, 0.2, -0.2, 0.3)$

$\sigma^2 \backslash \rho$	0.01	0.0625	1	25	100
0.0	0.7985	0.3759	0.0819	0.3558	0.6177
0.25	0.8660	0.5323	0.1173	0.3514	0.5489
0.5	0.8602	0.5202	0.1186	0.3168	0.4830
0.6	0.8502	0.4676	0.1092	0.3072	0.4614
0.7	0.8162	0.4043	0.0860	0.2874	0.4214
0.8	0.7471	0.3121	0.0685	0.2769	0.4087
0.85	0.6738	0.2562	0.0579	0.2735	0.4070
0.9	0.5765	0.1789	0.0506	0.2687	0.4009
0.95	0.3907	0.1093	0.0464	0.2873	0.4103
0.99	0.1310	0.0435	0.1012	0.5361	0.6463

表 2.14 $m = 15 \quad \rho = 0$

$\sigma^2 \backslash \beta' \beta$	0.01	0.0625	1	25	100
0.0	0.2938	0.2684	0.2873	0.2742	0.2873
0.05	0.3965	0.1048	0.0506	0.1376	0.1848
0.45	0.8594	0.4966	0.0752	0.0734	0.1103
0.7	0.9099	0.5533	0.0871	0.0656	0.1117
1.2	0.9259	0.5977	0.0918	0.0598	0.0926
5.0	0.9854	0.9190	0.4054	0.0529	0.0500
10.0	0.9917	0.9587	0.5698	0.0734	0.0493
25.0	0.9789	0.9825	0.7990	0.1271	0.0591
45	0.9878	0.9776	0.8665	0.1916	0.0798
140	0.9889	0.9898	0.9533	0.4133	0.1616
500	0.9824	0.9880	0.9793	0.7418	0.4000
12500	0.9893	0.9887	0.9886	0.9781	0.9648

表 2.15

 $m = 15 \quad \rho = 0.75$

σ^2 $\beta' \beta$	0.01	0.0625	1	25	100
0.0	0.1809	0.1914	0.1909	0.1837	0.1758
0.05	0.3479	0.0988	0.0456	0.0926	0.1204
0.45	0.8108	0.4370	0.0645	0.0603	0.0862
0.7	0.8472	0.5066	0.0743	0.0628	0.0895
1.2	0.9094	0.6044	0.1033	0.0513	0.0786
5.0	0.9750	0.8836	0.3572	0.0467	0.0444
10.0	0.9798	0.9338	0.5119	0.0613	0.0443
25.0	0.9862	0.9693	0.7155	0.1117	0.0516
45	0.9873	0.9771	0.8107	0.1754	0.0667
140	0.9892	0.9856	0.9331	0.3565	0.1446
500	0.9911	0.9900	0.9720	0.6750	0.3492
12500	0.9919	0.9926	0.9895	0.9686	0.9171

表 2.16

 $m = 15 \quad \rho = 0.95$

σ^2 $\beta' \beta$	0.01	0.0625	1	25	100
0.0	0.3123	0.2990	0.2945	0.2857	0.2949
0.05	0.1113	0.0439	0.0972	0.1924	0.1896
0.45	0.4925	0.1408	0.0418	0.1539	0.1999
0.7	0.5455	0.1771	0.0410	0.1482	0.2051
1.2	0.6372	0.2379	0.0442	0.1193	0.1959
5.0	0.9126	0.6214	0.1130	0.0491	0.0939
10.0	0.9507	0.7649	0.1820	0.0429	0.0700
25.0	0.9788	0.8936	0.3590	0.0426	0.0459
45	0.9873	0.9371	0.4883	0.0585	0.0411
140	0.9953	0.9783	0.7414	0.1208	0.0503
500	0.9987	0.9926	0.9120	0.3096	0.1119
12500	0.9992	0.9979	0.9951	0.9122	0.7217

表 2.17

 $m = 15 \quad \rho = 0.99$

$\sigma^2 \backslash \beta' \beta$	0.01	0.0625	1	25	100
0.0	0.9688	0.9820	0.9884	0.9683	0.9736
0.05	0.0465	0.0582	0.4245	0.7508	0.6951
0.45	0.1644	0.0512	0.0782	0.7016	0.8364
0.7	0.2064	0.0593	0.0688	0.6601	0.8272
1.2	0.2787	0.0740	0.0526	0.4840	0.7653
5.0	0.6806	0.2517	0.0464	0.1370	0.3986
10.0	0.8049	0.3967	0.0623	0.0858	0.2498
25.0	0.9110	0.6350	0.1099	0.0520	0.1154
45	0.9495	0.7441	0.1729	0.0423	0.0807
140	0.9833	0.8964	0.3614	0.0466	0.0461
500	0.9953	0.9692	0.6813	0.0938	0.0450
12500	1.0001	0.9982	0.9800	0.6680	0.3533

从表 2.10-2.17 的模拟结果可以看出：

1. 当 ρ 和 σ^2 增大时，比值 r 先增大而后减小；当 m 变大时，比值 r 会逐渐变大。
2. 当 $\beta' \beta$ 增大时，比值 r 先减小而后增大，当 $\beta' \beta$ 充分大以后，比值 r 趋向于 1。但 $\rho = 0.99$ 是一个很特殊的点，当 $\beta' \beta = 0$ 时，比值 r 很接近于 1，而后迅速下降，再缓慢趋向于 1。

从表中也可以观察到：虽然， ρ 、 σ^2 、 m 和 $\beta' \beta$ 的变化会引起 r 很多不同的变化，但 r 始终小于 1。因此我们的方法还是实用的。

第三章 矩阵形式的 James-Stein 估计

§ 3.1 引言

设 X 表示 p 维正态随机向量, $E(X) = \theta$, $\text{cov}(X, X) = I$, James and Stein(1961) 给出了风险一致小于 X 的风险的估计量 $\hat{\delta}^{JS} = (1 - \frac{p-2}{X'X})X$ 。同时, 也计算了估计量 $\hat{\delta}(a) = (1 - \frac{a}{X'X})X$ 的风险; 指出当 $a = p-2$ 时, $\hat{\delta}(a) = \hat{\delta}^{JS}$ 的风险达到最小。如果 $\text{cov}(X, X) = \sigma^2 I$, 则 $\hat{\delta}^{JS}$ 和 $\hat{\delta}(a)$ 可分别表示为 $(1 - \frac{(p-2)\sigma^2}{X'X})X$ 和 $(1 - \frac{a\sigma^2}{X'X})X$ 。

在这一章里, 我们根据前两章的思想和形式把估计量 $\hat{\delta}^{JS}$ 推广到矩阵形式, 并讨论了两种估计的风险之间的关系。

本章安排如下: 3.2 节给出矩阵形式的 James-Stein 估计量; 3.3 节应用数值模拟来比较两种估计的风险之间的关系。

§ 3.2 研究的动机和矩阵的导出

如果把估计量 $\hat{\delta}^{JS} = (1 - \frac{(p-2)\sigma^2}{X'X})X$ 中的 $X'X$ 看作是用来估

计 $\theta'\theta + p\sigma^2$ 的, 那么, 类似于前两章的估计量的形式和思想, 可

以把 $\hat{\delta}^{JS}$ 形式的看成是 $\delta^{JS} = (1 - \frac{(p-2)}{p} \times \frac{tr(\sigma^2 I)}{\theta'\theta + p\sigma^2})X$ 的估计量,

结合前两章的内容, 我们可把 δ^{JS} 推广到矩阵的形式, 其表达式为

$$\delta^{mJS} = (I - \frac{(p-2)}{p} \sigma^2 (\theta'\theta + \sigma^2 I)^{-1})X.$$

而对于 $\hat{\delta}(p) = (1 - \frac{p\sigma^2}{X'X})X$ 也有同样的做法。 $\hat{\delta}(p)$ 可以形式的

看成是 $\delta(p) = (1 - \frac{tr(\sigma^2 I)}{\theta'\theta + p\sigma^2})X$ 的估计量。这时, $\delta(p)$ 的矩阵形式

的推广为 $\delta^m(p) = (I - \sigma^2 (\theta\theta' + \sigma^2 I)^{-1})X$ 。

下面定理 3.1 比较了 δ^{mJS} 的风险 $MSE(\delta^{mJS})$ 和 δ^{JS} 的风险 $MSE(\delta^{JS})$ 之间的关系, 证明了 $MSE(\delta^{mJS}) \leq MSE(\delta^{JS})$ 。为了证明定理 3.1, 首先给出引理 3.1。

引理 3.1 若矩阵 A 是对称正定的 p 阶方阵, 则有

$$trA \cdot trA^{-1} \geq p^2.$$

证明: 设 $\lambda_i > 0$, $i=1,2,\dots,p$ 表示 A 的特征根, 则

$$trA = \sum_{i=1}^p \lambda_i, \quad trA^{-1} = \sum_{i=1}^p \lambda_i^{-1}$$

由 Holder 不等式可得,

$$trA \cdot trA^{-1} = \sum_{i=1}^p \lambda_i \cdot \sum_{i=1}^p \frac{1}{\lambda_i} \geq \left(\sum_{i=1}^p (\lambda_i^{1/2} \cdot \lambda_i^{-1/2}) \right)^2 = p^2.$$

引理 3.1 证毕。

定理 3.1 设 X 表示 p 维正态随机向量, $E(X) = \theta$, $cov(X, X) = \sigma^2 I$, 则

$$MSE(\delta^{mJS}) \leq MSE(\delta^{JS}).$$

证明: δ^{JS} 与 δ^{mJS} 的风险分别为

$$MSE(\delta^{JS}) = E(\delta^{JS} - \theta)'(\delta^{JS} - \theta) = p\sigma^2 - \frac{(p^2 - 4)\sigma^4}{\theta'\theta + p\sigma^2},$$

$$\begin{aligned} MSE(\delta^{mJS}) &= E(\delta^{mJS} - \theta)'(\delta^{mJS} - \theta) \\ &= p\sigma^2 - \frac{(p^2 - 4)\sigma^4}{p^2} tr(\theta\theta' + \sigma^2 I)^{-1}. \end{aligned}$$

因此只需证明 $\frac{1}{p^2} tr(\theta\theta' + \sigma^2 I)^{-1} \geq \frac{1}{\theta'\theta + p\sigma^2}$. 而由引理 3.1 知

不等式成立。因此, $MSE(\delta^{mJS}) \leq MSE(\delta^{JS})$.

定理 3.1 证毕。

注释 3.1 对于 $\delta(p)$ 和 $\delta^m(p)$, 可以像前两章一样证得 $\delta(p)$ 是使得 $\delta_b(X) = (1-b)X$ 的风险达到最小的解; 而 $\delta^m(p)$ 则是使得 $\delta_B^m(X) = (I-B)X$ 的风险达到最小的解。下面的定理 3.2 给出了相

应的证明。

定理 3.2 设 X 为 p 维正态随机向量, $E(X) = \theta$, $\text{cov}(X, X) = \sigma^2 I$. 则 $\delta(p) = (1 - \frac{p\sigma^2}{\theta'\theta + p\sigma^2})X$ 是使得 $\delta_b(X) = (1-b)X$ 的风险 $MSE(\delta_b(X))$ 达到最小的解。同时, $\delta^m(p) = (I - \sigma^2(\theta'\theta + p\sigma^2)^{-1})X$ 是使得 $\delta_B^m(X) = (I - B)X$ 的风险 $MSE(\delta_B^m(X))$ 达到最小的解。

证明:

$$\begin{aligned} MSE(\delta_b(X)) &= E(X - \theta - bX)'(X - \theta - bX) \\ &= p\sigma^2 - 2bp\sigma^2 + b^2(p\sigma^2 + \theta'\theta) \\ &= (p\sigma^2 + \theta'\theta)(b - \frac{p\sigma^2}{p\sigma^2 + \theta'\theta})^2 + \frac{p\sigma^2\theta'\theta}{p\sigma^2 + \theta'\theta}, \end{aligned}$$

因此当 $b = \frac{p\sigma^2}{p\sigma^2 + \theta'\theta}$ 时, $\delta_b(X) = \delta(p)$ 的风险达到最小, 这时

$$MSE(\delta(p)) = \frac{p\sigma^2\theta'\theta}{p\sigma^2 + \theta'\theta}.$$

而对于 $\delta_B(X) = (I - B)X$,

$$\begin{aligned} MSE(\delta_B(X)) &= E(X - \theta - BX)'(X - \theta - BX) \\ &= p\sigma^2 + \text{tr}[B'B(\sigma^2 I + \theta'\theta)] - \text{tr}(\sigma^2 B') - \text{tr}(\sigma^2 B) \\ &= p\sigma^2 + \text{tr}\{(\sigma^2 I + \theta'\theta)[B - (\sigma^2 I + \theta'\theta)^{-1}\sigma^2]'\}[B - \\ &\quad (\sigma^2 I + \theta'\theta)^{-1}\sigma^2] + \text{tr}[(\sigma^2 I + \theta'\theta)^{-1}\sigma^2\theta\theta'], \end{aligned}$$

因此, 当 $B = (\sigma^2 I + \theta\theta')^{-1} \sigma^2$ 时, $\delta_B(X) = \delta^m(p)$ 的风险达到最小, 这时 $MSE(\delta^m(p)) = tr[(\sigma^2 I_p + \theta\theta')^{-1} \sigma^2 \theta\theta']$.

定理 3.2 证毕

注释 3.2 类似于定理 3.1 可以证明

$$MSE(\delta^m(p)) \leq MSE(\delta(p)).$$

注释 3.3 对于 p 维正态随机向量的各分量可能相关的情况, 即 $E(X) = \theta$, $cov(X, X) = \Phi$. 也有类似于定理 3.2 结论。这时, 对于 $\delta_b(X) = (1-b)X$, 使其风险 $MSE(\delta_b(X))$ 达到最小的解为 $\delta = (1 - \frac{tr\Phi}{tr\Phi + \theta'\theta})X$, 风险为 $MSE(\delta) = \frac{\theta'\theta \cdot tr\Phi}{tr\Phi + \theta'\theta}$. 而对于 $\delta_B(X) = (I - B)X$, 使其风险 $MSE(\delta_B(X))$ 达到最小的解为 $\delta^m = (I_p - \Phi \times (\Phi + \theta\theta')^{-1})X$, 其风险为 $MSE(\delta^m) = tr[(\Phi + \theta\theta')^{-1} \Phi \theta\theta']$.

同前两章类似, 由 θ 的未知性, 也需要对 δ^{mJS} 和 δ^{JS} 进行估计。

令 $X_i \stackrel{i.i.d.}{\sim} N_p(\theta, \sigma^2 I), i = 1, 2, \dots, n$, 由 $E(\frac{1}{n} \sum_{i=1}^n X'X) = \theta'\theta + p\sigma^2$,

及 $E(\frac{1}{n} \sum_{i=1}^n XX') = \theta\theta' + \sigma^2 I$, 类似于前两章, 可以用下面的两个估

计量来估计 δ^{JS} 和 δ^{mJS} :

$$\tilde{\delta}^{JS} = \left(1 - \frac{(p-2)\sigma^2}{\frac{1}{n} \sum_{i=1}^n X_i' X_i}\right) \bar{X}, \quad \tilde{\delta}^{mJS} = \left(I - \frac{p-2}{p} \sigma^2 \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1}\right) \bar{X}.$$

§ 3.3 风险的数值模拟比较

在这一节中，我们将考虑两个估计的风险的数值模拟比较，观察 σ^2 ， $\theta'\theta$ 和 p 的变化对这两个估计的风险比值 $r = MSE(\tilde{\delta}^{mJS})/MSE(\tilde{\delta}^{JS})$ 的影响。

表 3.1

$p = 4$

$\theta'\theta \backslash \sigma^2$	0.01	0.05	0.1	0.5	1	5	10	25	100
0	1.0472	1.0446	1.0350	1.0470	1.0537	1.0531	1.0416	>1	>1
0.04	0.2772	0.5813	0.7612	1.0328	1.0596	1.0555	1.0529	>1	>1
0.16	0.2730	0.2997	0.4108	0.8129	0.9500	1.0650	1.0601	>1	>1
0.36	0.3603	0.2479	0.2839	0.6042	0.7931	1.0266	1.0510	>1	>1
0.64	0.4689	0.2591	0.2527	0.4604	0.6395	0.9770	1.0477	>1	>1
1	0.5587	0.2928	0.2525	0.3701	0.5234	0.9162	1.0016	>1	>1
4	0.8233	0.5081	0.3844	0.2472	0.2790	0.5827	0.7612	0.9478	1.0549
10	0.9235	0.6993	0.5592	0.2904	0.2495	0.3729	0.5251	0.7717	1.0062
16	0.9470	0.7805	0.6539	0.3386	0.2735	0.3055	0.4065	0.6413	0.9564
50	0.9836	0.9159	0.8536	0.5572	0.4168	0.2504	0.2570	0.3668	0.7102
110	0.9901	0.9562	0.9242	0.7067	0.5859	0.2972	0.2522	0.2701	0.5001
400	0.9974	0.9883	0.9815	0.8967	0.8230	0.5016	0.3802	0.2633	0.2801
10000	0.9998	0.9991	0.9999	0.9947	0.9916	0.9588	0.9131	0.8103	0.5564

表 3.2

$p = 8$

$\theta'\theta \backslash \sigma^2$	0.01	0.05	0.1	0.5	1	5	10	25	100
0	1.8928	1.8528	1.8778	1.8172	1.8480	1.8756	1.8290	>1	>1
0.04	0.1554	0.4920	0.6989	1.0837	1.2104	1.5004	1.6458	>1	>1
0.16	0.1016	0.1823	0.2992	0.7692	0.9531	1.2350	1.3731	>1	>1
0.36	0.1267	0.1152	0.1656	0.5153	0.7414	1.1016	1.2350	>1	>1
0.64	0.1702	0.1027	0.1217	0.3536	0.5598	0.9994	1.1321	>1	>1
1	0.2197	0.1069	0.1060	0.2537	0.4250	0.8869	1.0596	>1	>1
4	0.4957	0.1935	0.1342	0.1099	0.1554	0.4866	0.7014	0.9486	1.1990
10	0.7031	0.3400	0.2217	0.1061	0.1057	0.2563	0.4175	0.7080	1.0511
16	0.7947	0.4466	0.2970	0.1200	0.1028	0.1802	0.2995	0.5556	0.9481
50	0.9229	0.7003	0.5448	0.2227	0.1461	0.1050	0.1377	0.2559	0.6258
110	0.9632	0.8264	0.7214	0.3595	0.2336	0.1077	0.1042	0.1496	0.3960
400	0.9882	0.9461	0.9016	0.6519	0.5041	0.1931	0.1342	0.1034	0.1543
10000	0.9999	0.9999	0.9998	0.9998	0.9978	0.9932	0.9856	0.9584	0.8540

由模拟结果可以看出：

1. 这里出现了和第一章相类似的现象：当 $\theta'\theta$ 很小，即使 σ^2 很小，结果也不是很好，即 $r > 1$ 。但是不论 σ^2 多大，当 $\theta'\theta$ 增大， r 都会逐渐变小而后变大，并且一定会小于 1。同时，当 $\theta'\theta$ 充分大以后，不论 σ^2 多大， r 都会趋向于 1。

2. 当 p 增加时，比值 r 呈现了不太规律的变化。但当 $\theta'\theta$ 充分大之后， r 还是会趋向于 1。

这时我们可以得到同第一章类似结论：实际问题中 σ^2 和 τ^2 都不是很大，因此，当 $\theta'\theta \geq 1$ 以后，我们的方法还是较为实用的。同时，我们将在以后的工作中修正估计量中的不足之处。

第四章 对 Stein 估计的进一步研究

§ 4.1 引言

设 X 为 p 维正态随机向量, $E(X) = \theta$, $\text{cov}(X, X) = I$, Stein(1956) 得出了令统计学者非常惊讶的结果, $p \geq 3$ 时, X 不再是可容许的了。当 a 充分小, b 充分大时, 估计量 $\hat{\delta}^S(a, b) = (1 - \frac{a}{b + X'X})X$ 虽然是有偏的, 但它的风险要小于 X 的风险。随后, James and Stein(1961) 给出了完全控制估计量 X 的 James-Stein 估计 $\hat{\delta}^{JS} = (1 - \frac{p-2}{X'X})X$, 同时, 讨论了 $\hat{\delta}(a) = (1 - \frac{a}{X'X})X$ 的风险。在这一章里我们将讨论 $\hat{\delta}^S(a, b)$ 和 $\hat{\delta}(a)$ 之间的关系。

§ 4.2 $\hat{\delta}^S(a, b)$ 和 $\hat{\delta}(a)$ 关系的讨论

对于上面两个估计, 若 $\text{cov}(X, X) = \sigma^2 I$, 则两个估计量可写为

$\hat{\delta}^S(a, b) = (1 - \frac{a\sigma^2}{b + X'X})X$ 和 $\hat{\delta}(a) = (1 - \frac{a\sigma^2}{X'X})X$ 。Stein(1956) 证明了

当 a 充分小, b 充分大时, $\hat{\delta}^S(a, b)$ 的风险要小于 X 的风险。对于

James-Stein 估计 $\hat{\delta}^{JS} = (1 - \frac{(p-2)\sigma^2}{X'X})X$ 的风险要小于 X 的风险,

James and Stein(1961)给出了精确的证明。随后, Stein(1981), Brandwein and Strawderman(1990)给出了比较简单的证明, 使得计算 James-Stein 型估计的风险时更为简单, 并且证明了 $p \geq 3$,

$0 < a < 2(p-2)$ 时, $\hat{\delta}(a) = (1 - \frac{a\sigma^2}{X'X})X$ 的风险一致的小于 X 的风

险; 当 $a = p-2$ 时, 其风险达到最小。下面的定理 4.1 用与

Brandwein 和 Strawderman(1990)类似的方法的给出 $\hat{\delta}^S(a, b)$ 的风险小于 X 的风险的一个新的证明, 证明了当 $0 < a < 2(p-2)$,

$p \geq 3$, 且 $b \geq 0$ 时, $\hat{\delta}^S(a, b) = (1 - \frac{a\sigma^2}{b + X'X})X$ 的风险要小于 X 的风

险。首先给出 Brandwein 和 Strawderman(1990)得到的引理。

引理 4.1 令 $Y \sim N(\theta, 1)$, 则

$$E[h(Y)(Y - \theta)] = \text{Cov}(Y, h(Y)) = Eh'(Y)$$

这里, $h'(y)$ 存在, 且 $\lim_{y \rightarrow \pm\infty} h(y) \exp[-\frac{1}{2}(y - \theta)^2] = 0$, 同时所有的积分都是有限的。

注释 4.1 事实上若 $Y \sim N(\theta, \sigma^2)$, 可以证明(成平等(1985)),

$$E[h(Y)(Y - \theta)] = \sigma^2 Eh'(Y).$$

定理 4.1 设 X 为 p 维正态随机向量, $E(X) = \theta$, $\text{cov}(X, X) = \sigma^2 I$, 则当 $0 < a < 2(p-2)$, $p \geq 3$, $b \geq 0$ 时, $\hat{\delta}^S(a, b)$ 的风险小

于 X 的风险。这时, $\hat{\delta}^S(a, b)$ 的风险 $MSE(\hat{\delta}^S(a, b))$ 为

$$p\sigma^2 + (a^2 - 2a(p-2))\sigma^4 E \frac{1}{b + X'X} - b(a^2 + 4a)\sigma^4 E \frac{1}{(b + X'X)^2}.$$

证明:

$$\begin{aligned} MSE(\hat{\delta}^S(a, b)) &= E(\hat{\delta}^S(a, b) - \theta)'(\hat{\delta}^S(a, b) - \theta) \\ &= p\sigma^2 + a^2\sigma^4 E \frac{X'X}{(b + X'X)^2} - 2a\sigma^2 E \frac{X'(X - \theta)}{b + X'X}, \end{aligned}$$

用与 Brandwein and Strawderman(1990) 类似的证明方法, 可知

$$\begin{aligned} E \frac{X'(X - \theta)}{b + X'X} &= \sum_{i=1}^p E \left(\frac{X_i(X_i - \theta_i)}{b + \sum_{j=1}^p X_j^2} \right) = \sum_{i=1}^p \sigma^2 E \left(\frac{d}{dx_i} \left(\frac{X_i}{b + \sum_{j=1}^p X_j^2} \right) \right) \\ &= \sigma^2 \sum_{i=1}^p E \left(\frac{b + \sum_{j=1}^p X_j^2 - 2X_i^2}{(b + \sum_{j=1}^p X_j^2)^2} \right) = \sigma^2 E \left(\frac{p(b + X'X) - 2X'X}{(b + X'X)^2} \right) \\ &= \sigma^2 E \left(\frac{p-2}{(b + X'X)} + \frac{2b}{(b + X'X)^2} \right), \end{aligned}$$

则

$$\begin{aligned} MSE(\hat{\delta}^S(a, b)) &= p\sigma^2 + a^2\sigma^4 E \frac{1}{b + X'X} - a^2b\sigma^4 E \frac{1}{(b + X'X)^2} - 2a(p-2)\sigma^4 E \frac{1}{b + X'X} \\ &\quad - 4ab\sigma^4 E \frac{1}{(b + X'X)^2} \\ &= p\sigma^2 + (a^2 - 2a(p-2))\sigma^4 E \frac{1}{b + X'X} - b(a^2 + 4a)\sigma^4 E \frac{1}{(b + X'X)^2}. \end{aligned}$$

因此, 当 $0 < a < 2(p-2)$, $p \geq 3$, $b \geq 0$ 时, $\hat{\delta}^S(a, b)$ 的风险小于 X 的风险。

定理 4.1 证毕。

下面, 我们将简单讨论 $\hat{\delta}^S(a, b)$ 和 $\hat{\delta}(a)$ 的之间的关系:

Casella 和 Hwang(1982)证明了, 当 $X \sim N(\theta, \sigma^2 I)$, 且 $p \geq 3$ 时, 有 $E \frac{1}{X'X} \geq \frac{1}{(p-2)\sigma^2 + \theta'\theta} \geq \frac{1}{p\sigma^2 + \theta'\theta}$ 。因此, 对于 $\hat{\delta}(a) =$

$(1 - \frac{a\sigma^2}{X'X})X$, $X'X$ 作为 $p\sigma^2 + \theta'\theta$ 估计时, 虽然 $X'X$ 是 $p\sigma^2 + \theta'\theta$ 的

无偏估计, 但 $\frac{1}{X'X}$ 却是 $\frac{1}{p\sigma^2 + \theta'\theta}$ 的有偏估计, 且 $E \frac{1}{X'X}$ 与

$\frac{1}{p\sigma^2 + \theta'\theta}$ 有一段距离。而对于 $\hat{\delta}^S(a, b) = (1 - \frac{a\sigma^2}{b + X'X})X$, 虽然

$b + X'X$ 是 $p\sigma^2 + \theta'\theta$ 的有偏估计, 但我们最想估计的是 $\frac{1}{p\sigma^2 + \theta'\theta}$ 。

对于 $b > 0$ 时, $E \frac{1}{X'X}$ 和 $E \frac{1}{X'X + b}$ 的关系为 $E \frac{1}{X'X + b} \leq E \frac{1}{X'X}$, 因

此, $E \frac{1}{X'X + b}$ 比 $E \frac{1}{X'X}$ 更靠近 $\frac{1}{p\sigma^2 + \theta'\theta}$ 。从而有理由认为使用

$\hat{\delta}^S(a, b)$ 要比使用 $\hat{\delta}(a)$ 更好。

我们力图从理论上证明存在这样的 a 和 b , 使得 $\hat{\delta}^S(a, b)$ 的风险小于 $\hat{\delta}(a)$ 。但我们还没有找到太好的办法证明。我们将在 § 4.3 通过模拟讨论两者之间的关系。

§ 4.3 数值模拟比较

在这一节中，首先我们对 $p=4$ 和 8 ， $a=p-2$ ， $\sigma^2=1$ 的情形，给出了的 $MSE(\hat{\delta}^S(a,b))$ 的图形，

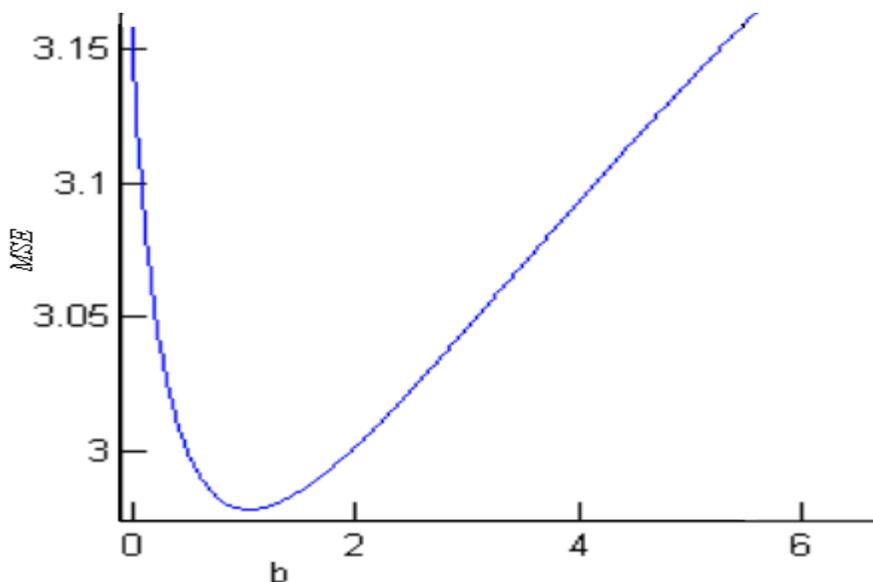


图 4.1 $p=4$

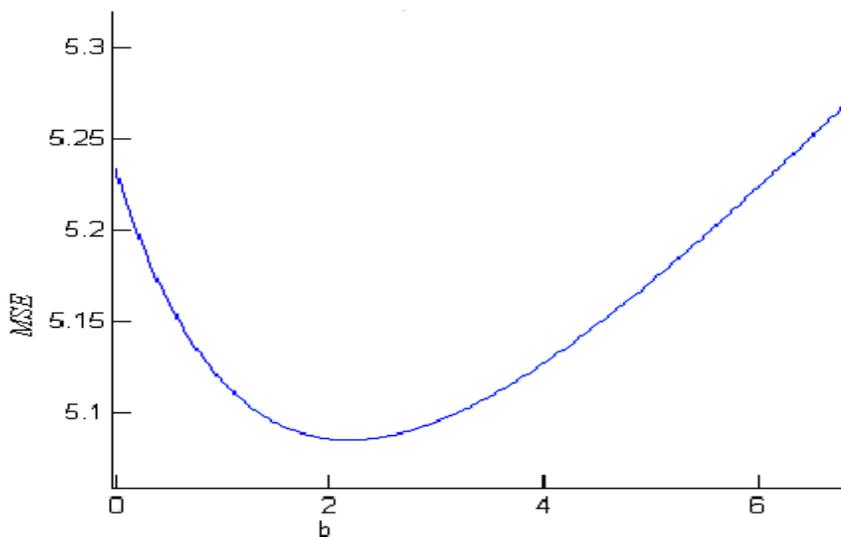


图 4.2 $p=8$

注意到, $b=0$ 时, $MSE(\hat{\delta}^S(a,b))=MSE(\hat{\delta}(a))$, 因此, 由图 4.1 和图 4.2 可以看出存在 a 和 b , 使得

$$MSE(\hat{\delta}^S(a,b)) \leq MSE(\hat{\delta}(a)).$$

下面我们将观察 σ^2 , p , θ , a , b 的变化对两种估计的风险比值 $r = MSE(\hat{\delta}^S(a,b))/MSE(\hat{\delta}(a))$ 的影响。表 4.1 给出了 p 、 $\theta'\theta$ 、 σ^2 、 a 和 b 的不同取值。

表 4.1

变量	取 值
p	4 8
$\theta'\theta$	2 4 8
σ^2	0.1 0.5 1 5 10 1000
a	4 6 8
b	$a-3$ $a-2$ $a-1$ a $a+1$ $a+2$ $a+3$

表 4.2

$a=4$ $\theta'\theta=2$ $p=4$

$\sigma^2 \backslash b$	$a-3$	$a-2$	$a-1$	a	$a+1$	$a+2$	$a+3$
0.1	0.9125	0.9080	0.9138	0.9212	0.9282	0.9345	0.9400
0.5	0.7309	0.7112	0.7250	0.7459	0.7674	0.7876	0.8060
1	0.6164	0.5877	0.5724	0.5924	0.6168	0.6420	0.6662
5	0.5944	0.4648	0.3950	0.3541	0.3299	0.3163	0.3096
10	0.7780	0.6491	0.5641	0.5045	0.4614	0.4297	0.4062
1000	0.9897	0.9799	0.9703	0.9612	0.9523	0.9437	0.9354

表 4.3

$a = 4 \quad \theta'\theta = 4 \quad p = 6$

$\sigma^2 \backslash b$	$a-3$	$a-2$	$a-1$	a	$a+1$	$a+2$	$a+3$
0.1	0.9898	0.9895	0.9922	0.9957	0.9993	1.0027	1.0059
0.5	0.9913	1.0054	1.0247	1.0446	1.0637	1.0815	1.0979
1	0.9403	0.9471	0.9674	0.9915	1.0163	1.0404	1.0634
5	0.9459	0.9212	0.9123	0.9128	0.9193	0.9299	0.9432
10	0.9612	0.9350	0.9175	0.9062	0.8996	0.8965	0.8961
1000	0.9992	0.9985	0.9977	0.9970	0.9962	0.9955	0.9947

表 4.4

$a = 6 \quad \theta'\theta = 8 \quad p = 8$

$\sigma^2 \backslash b$	$a-3$	$a-2$	$a-1$	a	$a+1$	$a+2$	$a+3$
0.1	0.9973	0.9985	1.0001	1.0017	1.0034	1.0052	1.0068
0.5	0.9897	0.9950	1.0015	1.0087	1.0161	1.0235	1.0309
1	0.9858	0.9936	1.0034	1.0145	1.0262	1.0381	1.0501
5	0.9574	0.9545	0.9553	0.9589	0.9647	0.9723	0.9813
10	0.9535	0.9443	0.9375	0.9328	0.9299	0.9286	0.9287
1000	0.9989	0.9986	0.9982	0.9979	0.9976	0.9972	0.9969

由模拟结果可以看出：

1. σ^2 增大时，比值 r 先增大，而后减小，然后再增加；
2. p 、 $\theta'\theta$ 增大时，比值 r 增大；
3. 存在 a 和 b 使得 $r < 1$.

通过上面的模拟结果和论述可知，存在 a 和 b 使得 $\hat{\delta}^S(a, b)$ 优于 $\hat{\delta}(a)$ 的，在以后的工作中我们希望从理论上证明这一结论。

结论与下一步的研究问题

Stein(1956)结论引起了统计界的轰动,掀起了研究可容许估计的热潮。同时,James-Stein 估计的提出也引起了统计学家的浓厚兴趣,对 James-Stein 估计的修正,改进以及应用方面的工作从来没有停止过。因此,对 James-Stein 估计及其相关的问题的研究具有很强的应用背景和实际意义。

本文将 Green and Strawderman(1991) 以及 Judge and Mittelhammer(2004)提出的加权组合估计量推广到了更一般的矩阵加权形式。给出了使得矩阵加权估计量的风险达到最小的矩阵,并讨论了两种估计量的风险之间的关系。论述了矩阵加权估计的优良性及其实用范围。

同时,运用类似的形式和思想给出了矩阵型的 James-Stein 估计量,并讨论了它与 James-Stein 估计的风险之间的关系以及矩阵型的 James-Stein 估计量的优良性。

最后,我们讨论了 Stein(1956)提出的估计量(Stein 估计)与 James-Stein 估计的关系。给出了 Stein 估计优于通常估计量的条件,并初步讨论了 Stein 估计的风险与 James-Stein 估计的风险之间的关系。

下一步的研究问题:

1. 在第四章我们只是从图表中看出存在 a 和 b , 使得 $MSE(\hat{\delta}^S(a,b)) \leq MSE(\hat{\delta}(a))$ 。下一步工作目标之一是从理论上证明

存在区间 H ，当 $a \in H$ 时，存在 b 使得 $MSE(\hat{\delta}^s(a,b)) \leq MSE(\hat{\delta}(a))$ 。
并进一步讨论方差对风险的影响。

2. 第一章和第三章的模拟结果表明矩阵加权估计量在某些情形效果并不是很好，因此，我们将在后续的工作中进一步完善。

3. 下一步工作的另一个目标是能够在对称正定矩阵中找到使得矩阵加权估计的风险达到最小的矩阵。同时，该矩阵应具有下面的性质： ξ 应为其特征向量。

参考文献

1. Alam, K. ; Hawkes, J. S. Minimax property of Stein's estimator. *Communications in Statistics. A. Theory and Methods*, 1979, 8 , 581--590.
2. Baranchik, A. J. A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, 1970, 41, 642--645.
3. Baranchik, A. J. Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *The Annals of Statistics*, 1973, 1 , 312-321.
4. Berry, J. C. Improving the James-Stein estimator using the Stein variance estimator. *Statistics & Probability Letters*, 1994, 20 , 241-245.
5. Blyth, C. R. On minimax statistical decision procedures and their admissibility. *Annals of Mathematical Statistics*, 1951, 22, 22-42.
6. Brandwein, A. C. and Strawderman, W. E. Stein estimation: the spherically symmetric case. *Statistical Science*, 1990, 5, 356-369.
7. Brandwein, A. C. and Strawderman, W. E. Generalizations of James-Stein estimators under spherical symmetry. *The*

- Annals of Statistics*, 1991, 19, 1639–1650.
8. Brown, L.D. On the admissibility of invariant estimators of two location parameters. *Annals of Mathematical Statistics*, 1966, 37, 1087–1136.
 9. Casella, G. and Hwang, J. T. Limit expressions for the risk of James–Stein estimators. *The Canadian Journal of Statistics*, 1982, 10, 305–309.
 10. Efron, B. and Morris, C. Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *Journal of the American Statistical Association*, 1972a, 67, 130–139.
 11. Efron, B. and Morris, C. Empirical Bayes on vector observations: an extension of Stein’s method. *Biometrika*, 1972b, 59, 335–347.
 12. Efron, B. and Morris, C. Stein’s estimation rule and its competitors---an empirical Bayes approach. *Journal of the American Statistical Association*, 1973a, 68, 117–130.
 13. Efron, B. and Morris, C. Combining possibly related estimation problems. *Journal of the Royal Statistical Society. Series B. Methodological*, 1973b, 35, 379–421.
 14. Efron, B. and Morris, C. Data analysis using Stein’s estimator and its generalizations. *Journal of the American Statistical Association*, 1975, 70, 311–319.
 15. Fan, Jianqing and Fang, KaiTai Minimax estimators and

- two-stage Stein estimators of location parameters for elliptically contoured distributions. (Chinese) *Chinese Journal of Applied Probability and Statistics*, 1985, 1, 103-114.
16. Girshick, M. A. and Savage, L. J. Bayes and minimax estimates for quadratic loss functions. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, 1, 53-73..
 17. Green, E. J. and Strawderman, W. E. A James-Stein type estimator for combining unbiased and possibly biased estimators. *Journal of the American Statistical Association*, 1991, 86, 1001-1006.
 18. Groenewald, P. C. N. and van der Merwe, A. J. Asymptotic distributions for the Stein estimator in multiple regression. *South African Statistical Journal*, 1979, 13, 139-152.
 19. Hodges, J. L. and Lehmann, E. L. Some problems in minimax point estimation. *Annals of Mathematical Statistics*, 1950, 21, 182-197.
 20. James, W. and Stein, C. Estimation with quadratic loss. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1961, 1, 361-379.
 21. Judge, G. G. and Mittelhammer, R. C. A semiparametric basis

- for combining estimation problems under quadratic loss. *Journal of the American Statistical Association*, 2004, 99, 479–487.
22. Kumar, S., Tripathi, Y. M. and Misra, N. James-Stein type estimators for ordered normal means. *Journal of Statistical Computation and Simulation*, 2005, 75, 501–511.
23. Kubokawa, T. An approach to improving the James-Stein estimator. *Journal of Multivariate Analysis*, 1991, 36, 121–126.
24. Li, T. F. A note on James-Stein and Bayes empirical Bayes estimators. *Communications in Statistics. A. Theory and Methods*, 1982, 11, 1029–1043.
25. Li, T. F. and Bhoj, D. S. A modified James-Stein estimator with application to multiple regression analysis. *Scandinavian Journal of Statistics. Theory and Applications*, 1988, 15, 33–37.
26. Maruyama, Y. Improving on the James-Stein estimator. *Statistics & Decisions*, 1999, 17, 137–140.
27. Maruyama, Y. A robust generalized Bayes estimator improving on the James-Stein estimator for spherically symmetric distributions. *Statistics & Decisions*, 2003, 21, 69–77.
28. Maruyama, Y. Stein's idea and minimax admissible estimation of a multivariate normal mean. *Journal of*

- Multivariate Analysis* , 2004, 88, 320-334.
29. Moore, T. and Brook, R. J. Risk estimate optimality of James-Stein estimators. *The Annals of Statistics* , 1978, 6 , 917-919.
 30. Namba, A MSE dominance of the PT-2SHI estimator over the positive-part Stein-rule estimator in regression. *Journal of Statistical Planning and Inference*, 2000, 89, 175-185.
 31. Ohtani, K. A comparison of the Stein-rule and positive-part Stein-rule estimators in a misspecified linear regression model. *Econometric Theory*, 1993, 9, 668-679.
 32. Ohtani, K. Inadmissibility of the Stein-rule estimator under the balanced loss function. *Journal of Econometrics*, 1999, 88, 193-201.
 33. Ohtani, K. Further improving the Stein-rule estimator using the Stein variance estimator in a misspecified linear regression model. *Statistics & Probability Letters*, 1996, 29, 191-199.
 34. Ohtani, K. The exact risk of a weighted average estimator of the OLS and Stein-rule estimators in regression under balanced loss. *Statistics & Decisions*, 1998, 16, 35--45.
 35. Oman, S. D. A class of modified Stein estimators with easily computable risk functions. *Journal of Statistical Planning and Inference*, 1983, 7, 359-369.
 36. Phillips, P. C. B. The exact distribution of the

- Stein-rule estimator. *Journal of Econometrics*, 1984, 25, 123-131.
37. Ralescu, S., Brandwein, A. C. and Strawderman, W. E. Stein estimation for nonnormal spherically symmetric location families in three dimensions. *Journal of Multivariate Analysis*, 1992, 42, 35-50.
38. Robert, C. P., Gene Hwang, J. T. and Strawderman, W. E. Is Pitman closeness a reasonable criterion? *Journal of the American Statistical Association*, 1993, 88, 57-76.
39. Shao, P. Yi-Shi; Strawderman, W. E. Improving on the James-Stein positive-part estimator. *The Annals of Statistics*, 1994, 22, 1517-1538.
40. Srivastava, A. K. and Srivastava, V. K. Pitman closeness for Stein-rule estimators of regression coefficients. *Statistics & Probability Letters*, 1993, 18, 85-89.
41. Srivastava, M. S. and Bilodeau, M. Stein estimation under elliptical distributions. *Journal of Multivariate Analysis*, 1989, 28, 247-259.
42. Stein, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956, 1, 187-195.
43. Stein, C. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean.

- Annals of the Institute of Statistical Mathematics*, 1964, 16, 155-160.
44. Stein, C. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 1981, 9, 1135 - 1151.
45. Susarla, V. A property of Stein's estimator for the mean of a multivariate normal distribution. *Statistica Neerlandica*, 1976, 30, 1-5.
46. Takada, Y. Stein's positive part estimator and Bayes estimator. *Annals of the Institute of Statistical Mathematics*, 1979, 31, 177-183.
47. Taniguchi, M. and Hirukawa, J. The Stein-James estimator for short- and long-memory Gaussian processes. *Biometrika*, 2005, 92, 737-746.
48. Ullah, A. and Ullah, S. Double K-class estimators of coefficients in linear regression. *Econometrica*, 1978, 46, 705-722.
49. 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, 第一版, 426-427.
50. 方开泰, 实用多元统计分析, 华东师范大学出版社, 1989, 第一版, 8-9.
51. 张尧庭、方开泰, 多元统计分析引论, 科学出版社, 1997, 第一版第三次印刷, 31-32.

攻博期间发表的学术论文

1. Zhao Shi-shun, Zhu Fu-kang, Wang De-hui, Parameter Estimation for the NEAR(p) Model, *Northeastern Mathematical Journal* 2005, 21, 383-386.
2. 王德辉, 赵世舜, 潘 鸿 熵损失下心理状态数的统计推断, *吉林大学学报 (理学版)*, 2003, 41 卷, 第 4 期, 431-435.
3. 赵世舜, 宋 阳, 宋立新 对称熵损失下两个指数总体均值的序约束估计, *吉林大学学报 (理学版)*, (接收).
4. 赵世舜, 宋 阳, 史宁中 序约束下指数总体均值的混合估计, *应用概率统计*, (接收).

矩阵加权估计及 James-Stein 估计的再研究

姓 名：赵 世 舜

专 业：概率论与数理统计

指导教师：史 宁 中 教授

中文摘要

正态均值用其样本均值去估计有很好的性质，但当把这样的估计推广到 p 元正态分布场合时，Stein(1956)给出了令统计界震惊的结果：设 X 表示 p 维正态随机向量， $E(X) = \theta$ ， $\text{cov}(X, X) = I$ 。则当 $p \geq 3$ 时， X 不再是可容许的了。在平方损失下，当 a 充分小， b 充分大时，估计量 $(1 - \frac{a}{b + XX})X$ 虽然是有偏的，但它的风险要小于 X 的风险。随后，James and Stein(1961)给出了完全控制估计量 X 的 James-Stein 估计， $\hat{\delta}^{JS} = (1 - \frac{p-2}{XX})X$ ，并讨论了估计 $\hat{\delta}(a) = (1 - \frac{a}{XX})X$ ；指出当 $a = p - 2$ 时， $\hat{\delta}(a) = \hat{\delta}^{JS}$ 的风险达到最小。

自 Stein(1956)的结果得出后，各国统计学家讨论了大量的可容许性的问题。而 James-Stein 估计也引起了许多统计学家的兴

趣,对 James-Stein 估计进行了很多的讨论、修正与改进,也将 James-Stein 估计应用于了很多领域。

在众多关于 James-Stein 估计的改进和应用的工作中,令我们非常感兴趣的是 Green and Strawderman (1991)利用 Stein 估计的思想和形式给出了新的加权组合估计量的方法。与以前的方法不同的是这里的辅助变量是可能有偏的。该文章中,作者讨论了正态分布下,这种新的的组合估计量的优良性。Judge and Mittelhammer (2004)将 Green and Strawderman(1991)的工作推广到加权估计中变量之间相关的情形,作者通过数值模拟的方法讨论了加权估计的优良性。本文的很多工作也是在这两篇文章的基础上展开的。

本文主要是将上述两篇文章结果推广到了更一般的矩阵加权形式,讨论了矩阵加权估计和原来的组合估计的风险之间的关系。在此基础上,我们又把 James-Stein 估计推广到矩阵形式的估计,也比较了它与 James-Stein 估计风险之间的关系。最后,我们初步的讨论了 Stein(1956)提出的估计与 James-Stein 估计之间的关系。本文共分四章:第一章讨论独立情况下矩阵加权组合估计量。第二章讨论相关情况下矩阵加权组合估计量,第三章讨论矩阵型的 James-Stein 估计,第四章是对 Stein 给出的估计的进一步讨论。

(一) 独立情况的矩阵加权组合估计量

设 p 维随机向量 $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, 并且 X 与 Y 独立,其中 σ^2, τ^2 已知, θ, ξ 未知。Green and Strawderman (1991)给出了使得 $\delta_a^{lc}(X, Y) = aX + (1-a)Y$ 的风险达到最小的 a_0 为

$$\frac{p\tau^2 + \xi'\xi}{p\tau^2 + p\sigma^2 + \xi'\xi}, \quad \text{这时,} \quad \delta_{a_0}^{lc}(X, Y) = X - \frac{p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi} \times (X - Y),$$

它的风险为 $MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{(p\tau^2 + \xi'\xi)p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi}$.

如果把 $\delta_a^{lc}(X, Y)$ 写成 $aIX + (1-a)IY$, 则 $\delta_a^{lc}(X, Y)$ 成为了一个矩阵加权组合估计量。对于矩阵加权组合估计量, 我们有下面的主要定理和推论:

定理 1.1 设 p 维随机向量 $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, 并且 X 与 Y 独立, 其中 σ^2 , τ^2 已知, θ , ξ 未知。则使得矩阵加权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A_0 为 $(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi')$, 这时, $\delta_{A_0}^{mc}(X, Y)$ 的风险为

$$MSE(\delta_{A_0}^{mc}(X, Y)) = tr(\sigma^2(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1}).$$

定理 1.2 设 p 阶对称方阵 $A > 0$, $B > 0$, 则

$$\frac{trA^{-1} \cdot trB^{-1}}{trA^{-1} + trB^{-1}} \geq tr(A + B)^{-1}.$$

推论 1.1

1. 当 $p = 1$ 时, $MSE(\delta_{A_0}^{mc}(X, Y)) = MSE(\delta_{a_0}^{lc}(X, Y))$;
2. 当 $p \geq 2$ 时, $MSE(\delta_{A_0}^{mc}(X, Y)) \leq MSE(\delta_{a_0}^{lc}(X, Y))$.

同时,我们还讨论了 ξ 与矩阵 A_0 的关系,指出 $\frac{\tau^2 + \xi'\xi}{\tau^2 + \sigma^2 + \xi'\xi}$ 和 ξ 分别为矩阵 A_0 的特征根和特征向量。

又由于 ξ 未知性,我们给出了相应的估计量:

$$\hat{\delta}^{lc}(X, Y) = \bar{X} - \frac{p\sigma^2}{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}),$$

$$\hat{\delta}^{mc}(X, Y) = \bar{X} - \sigma^2 \left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)'\right]^{-1} (\bar{X} - \bar{Y}).$$

并利用数值模拟比较了 $\hat{\delta}^{lc}(X, Y)$ 和 $\hat{\delta}^{mc}(X, Y)$ 的风险之间的关系。讨论了各种参数的变化对两种估计的风险比值的影响,给出了我们方法的实用范围。

(二) 相关情况的矩阵加权组合估计量

设 X, Y 分别为 p 维正态随机向量,且 $E(X) = \theta, E(Y) = \theta + \xi,$

$$\text{cov} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Sigma \\ \Sigma' & \Phi_2 \end{pmatrix},$$

其中 Φ_1 为 X 的协方差矩阵, Φ_2 为 Y 的协方差矩阵, Σ 为 X 与 Y 协方差矩阵, 其中 $\Phi_1, \Phi_2, \Sigma, \theta$ 和 ξ 是未知的。

Judge and Mittelhammer(2004) 给出使得 $\delta_a^{lc}(X, Y) = aX +$

$$(1-a)Y$$

的风险达到最小的 a_0 为 $\frac{\text{tr}\Phi_2 + \xi'\xi - \text{tr}\Sigma}{\text{tr}(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}$, 这时

$$\delta_{a_0}^{lc}(X, Y) = X - \frac{tr\Phi_1 - tr\Sigma}{tr(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}(X - Y),$$

它的风险为

$$MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{tr\Phi_1 tr(\Phi_2 + \xi\xi') - (tr\Sigma)^2}{tr\Phi_1 + tr\Phi_2 + \xi'\xi - 2tr\Sigma}.$$

我们同样将其推广到矩阵形式，有下面的定理：

定理 2.1 设 $X \sim N_p(\theta, \Phi_1)$, $Y \sim N_p(\theta + \xi, \Phi_2)$, $cov(X, Y) = \Sigma$,

则当 $B = \Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma'$ 为正定矩阵 ($B > 0$) 时，使得矩阵加权组合估计 $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ 的风险达到最小的矩阵 A 为 $A_0 = (\Phi_2 + \xi\xi' - \Sigma')(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}$ ，这时 $\delta_{A_0}^{mc}(X, Y)$ 的风险为 $tr([\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma](\Phi_1 + \Phi_2 + \xi'\xi - \Sigma - \Sigma')^{-1})$.

定理 2.2 若 $\Sigma' = \Sigma$ ， $\Phi_1 - \Sigma > 0$ 及 $\Phi_2 + \xi\xi' - \Sigma > 0$ 成立，则

$$MSE(\delta_{A_0}^{mc}(X, Y)) \leq MSE(\delta_{a_0}^{lc}(X, Y)).$$

同第一章类似，我们讨论了 ξ 与 A_0 的关系，并给出了 ξ 是矩阵 A_0 的特征向量的条件。

由于 ξ ， Φ_1 ， Φ_2 和 Σ 未知性，我们给出了相应的估计量：

$$\hat{\delta}^{lc}(X, Y)$$

$$= \bar{X} - \frac{n \sum_{i=1}^n (X_i - \bar{X})'(X_i - \bar{X}) - n \sum_{i=1}^n (X_i - \bar{X})'(Y_i - \bar{Y})}{(n-1) \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}),$$

$$\hat{\delta}^{mc}(X, Y)$$

$$= \bar{X} - [n \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' - n \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})'] [(n-1) \sum_{i=1}^n (X_i - Y_i) \times (X_i - Y_i)']^{-1} (\bar{X} - \bar{Y}).$$

同时，用数值模拟和实例比较了两者风险之间的关系，同样讨论了各种参数的变化对两者风险比值的影响，以及我们方法的实用性。

(三) 矩阵形式的 James-Stein 估计

根据前两部分的思想 and 形式，我们将 James-Stein 估计推广到矩阵形式，得到了如下定理：

定理 3.1 设 X 表示 p 维正态随机向量， $E(X) = \theta$ ， $\text{cov}(X, X) = \sigma^2 I$ ，则

$$MSE(\delta^{mJS}) \leq MSE(\delta^{JS}).$$

定理 3.2 X 为 p 维正态随机向量， $E(X) = \theta$ ， $\text{cov}(X, X) = \sigma^2 I$ 。则 $\delta(p) = (1 - \frac{p\sigma^2}{\theta'\theta + p\sigma^2})X$ 是使得 $\delta_b(X) = (1-b)X$ 的风险

$MSE(\delta_b(X))$ 达到最小的解。同时, $\delta^m(p) = (I - \sigma^2(\theta'\theta + p\sigma^2)^{-1})X$ 是使得 $\delta_B^m(X) = (I - B)X$ 的风险 $MSE(\delta_B^m(X))$ 达到最小的解。

同时, 由于 θ 是未知的, 我们给出了相应的估计量:

$$\tilde{\delta}^{JS} = (1 - \frac{(p-2)\sigma^2}{\frac{1}{n}\sum_{i=1}^n X_i'X_i})\bar{X},$$

$$\tilde{\delta}^{mJS} = (I - \frac{p-2}{p}\sigma^2(\frac{1}{n}\sum_{i=1}^n X_iX_i')^{-1})\bar{X}.$$

并给出了二者风险之间的数值模拟比较结果, 以及我们方法的实用范围。

(四) 对 Stein 估计的进一步研究

这部分我们主要比较了 $\hat{\delta}^s(a,b) = (1 - \frac{a\sigma^2}{b + X'X})X$ 和 $\hat{\delta}(a) = (1 - \frac{a\sigma^2}{X'X})X$ 之间的关系, 导出了 $\hat{\delta}^s(a,b) = (1 - \frac{a\sigma^2}{b + X'X})X$ 的风险, 得到了下面的定理:

定理 4.1 X 为 p 维正态随机向量, $E(X) = \theta$, $cov(X, X) = \sigma^2I$, 则当 $0 < a < 2(p-2)$, $p \geq 3$, $b \geq 0$ 时, $\hat{\delta}^s(a,b)$ 的风险小于 X 的风险。这时, $\hat{\delta}^s(a,b)$ 的风险 $MSE(\hat{\delta}^s(a,b))$ 为

$$p\sigma^2 + (a^2 - 2a(p-2))\sigma^4 E \frac{1}{b+X'X} - b(a^2 + 4a)\sigma^4 E \frac{1}{(b+X'X)^2}.$$

同时，我们用数值模拟比较了 $\hat{\delta}^S(a,b)$ 和 $\hat{\delta}(a)$ 风险之间的关系，说明存在 a 和 b ，使得 $MSE(\hat{\delta}^S(a,b)) \leq MSE(\hat{\delta}(a))$ 。

Matrix-Weighted Combined Estimator and Further Study of James-Stein Estimator

Author: Shi-Shun Zhao

Major: Probability Theory and Mathematical Statistics

Thesis Advisor: Professor **Ning-Zhong Shi**

Abstract

There are many good properties when normal mean is estimated by the corresponding sample mean. But when it is extended to the p -variate normal distribution, what Stein(1956) proved made many statisticians surprised: Let X be a p -variate normal random vector, $E(X) = \theta$, $\text{cov}(X, X) = I$. When $p \geq 3$, X is inadmissible as an estimator of θ . Under square loss function, the risk of estimator of the form $(1 - \frac{a}{(b + \|X\|^2)})X$ is less than that of X for a sufficiently small and b sufficiently large. James and

Stein (1961) gave the James-Stein estimator $\delta^{JS} = (1 - \frac{p-2}{X'X})X$, which dominates X . They also discussed estimator $\hat{\delta}(a) = (1 - \frac{a}{X'X})X$ and indicated that the risk of $\hat{\delta}(a) = \delta^{JS}$ is minimum when $a = p - 2$.

Since the result of Stein(1956), statisticians from all over the world discussed a great number of problems of admissibility. Later, many statisticians put their focuses on James-Stein estimator as well. From then on, James-Stein estimator has been extensively studied, modified, improved, and also applied in many fields.

Among the works of improvement and application, we are attracted by Green and Strawderman (1991). Inspired by the idea and form of Stein(1956), Green and Strawderman (1991) present a new method of combining estimators which differs from the others. The difference is that the auxiliary variate is possibly biased. In their work, Green and Strawderman (1991) showed a better property of the combined estimator. Subsequently, Judge and Mittelhammer (2004) extended the result of Green and Strawderman(1991) to the situation that variate is dependent. Judge and Mittelhammer(2004) also presented a better property of estimator by means of numerical simulation. The main work of our thesis are developed based on the two works.

In this thesis, we mainly extend the above two results

to matrix-weighted combined estimator and discuss the relationship between the risk of matrix-weighted combined estimator and that of the above two estimators. Based on this, we also generalize James-Stein estimator to the estimator of matrix form and compare the relationship between the risk of James-Stein estimator and ours. In the end, we discuss the relationship between the estimator of Stein(1956) and James-Stein estimator. Our thesis is divided into four chapters: the first chapter discusses the matrix-weighted combined estimator in the independent situation; the second chapter discusses the matrix-weighted combined estimator in the dependent situation; the third chapter presents James-Stein estimator of matrix form; and the fourth chapter shows the further study of Stein estimator.

1. Matrix-weighted combined estimator in independent situation

Let X and Y be independent in p -variate random vectors, and $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, where σ^2 and τ^2 are known, but θ and ξ are unknown. Green and Strawderman(1991)

gave $a_0 = \frac{p\tau^2 + \xi'\xi}{p\tau^2 + p\sigma^2 + \xi'\xi}$ which makes the minimum risk of

$\delta_a^{lc}(X, Y) = aX + (1-a)Y$, and then

$$\delta_{a_0}^{lc}(X, Y) = X - \frac{p\sigma^2}{(p\tau^2 + p\sigma^2 + \xi'\xi)}(X - Y),$$

the risk of $\delta_{a_0}^{lc}(X, Y)$ is

$$MSE(\delta_{a_0}^{lc}(X, Y)) = \frac{(p\tau^2 + \xi'\xi)p\sigma^2}{p\tau^2 + p\sigma^2 + \xi'\xi}.$$

If we write $\delta_a^{lc}(X, Y)$ as $aIX + (1-a)IY$, then $\delta_a^{lc}(X, Y)$ becomes an matrix-weighted combined estimator. For the combined estimator of matrix-weighted form, we present the following theorems and corollary.

Theorem 1.1 let p -variate random vectors $X \sim N_p(\theta, \sigma^2 I)$, $Y \sim N_p(\theta + \xi, \tau^2 I)$, X and Y be independent, where σ^2 and τ^2 are known, θ and ξ are unknown. Then, the matrix which make the risk of matrix-weighted combined estimator $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ become minimum, is

$$A_0 = (\tau^2 I + \sigma^2 I + \xi\xi')^{-1}(\tau^2 I + \xi\xi'),$$

and the risk of $\delta_{A_0}^{mc}(X, Y)$ is

$$MSE(\delta_{A_0}^{mc}(X, Y)) = tr(\sigma^2(\tau^2 I + \xi\xi')(\tau^2 I + \sigma^2 I + \xi\xi')^{-1})$$

Theorem 1.2 Let the p th order symmetric $A > 0$ and $B > 0$, then,

$$\frac{trA^{-1} \cdot trB^{-1}}{trA^{-1} + trB^{-1}} \geq tr(A + B)^{-1}.$$

Corollary 1.1

1. If $p = 1$, then $MSE(\delta_{a_0}^{lc}(X, Y)) = MSE(\delta_{A_0}^{mc}(X, Y))$;
2. If $p \geq 2$, then $MSE(\delta_{a_0}^{lc}(X, Y)) \geq MSE(\delta_{A_0}^{mc}(X, Y))$.

We discuss the relationship between ξ and A_0 , and indicate

$\frac{\tau^2 + \xi'\xi}{\tau^2 + \sigma^2 + \xi'\xi}$ and ξ as characteristic root and characteristic

vector of Matrix A_0 , respectively.

Since ξ is unknown, we provide the relevant estimators of $\delta_{a_0}^{lc}(X, Y)$ and $\delta_{A_0}^{mc}(X, Y)$:

$$\hat{\delta}^{lc}(X, Y) = \bar{X} - \frac{p\sigma^2}{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}),$$

$$\hat{\delta}^{mc}(X, Y) = \bar{X} - \sigma^2 \left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)(X_i - Y_i)' \right]^{-1} (\bar{X} - \bar{Y}).$$

We compare the relationships between the risks of $\hat{\delta}^{lc}(X, Y)$ and $\hat{\delta}^{mc}(X, Y)$ by means of numerical simulation and discuss the effects of parameter variations on the ratio of the two risks; finally we propose the practical range of our method.

2. Matrix-weighted combined estimator in dependent situation

Let X, Y be two p -variate normal random vectors, and $E(X) = \theta, E(Y) = \theta + \xi, \text{cov}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Sigma \\ \Sigma' & \Phi_2 \end{pmatrix}$, where Φ_1 is covariance matrix of X, Φ_2 is covariance matrix of Y, Σ is covariance matrix of X and Y , where $\Phi_1, \Phi_2, \Sigma, \theta$ and ξ are unknown. Following Judge and Mittelhammer(2004), noting that when $a = a_0 = \frac{\text{tr}\Phi_2 + \xi'\xi - \text{tr}\Sigma}{\text{tr}(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}$, the risk of $\delta_a^{lc}(X, Y) = aX + (1-a)Y$ becomes minimum. Then

$$\delta_{a_0}^{lc}(X, Y) = X - \frac{\text{tr}\Phi_1 - \text{tr}\Sigma}{\text{tr}(\Phi_1 + \Phi_2 + \xi\xi' - 2\Sigma)}(X - Y),$$

the risk is

$$\text{MSE}(\delta_{a_0}^{lc}(X, Y)) = \frac{\text{tr}\Phi_1 \text{tr}(\Phi_2 + \xi\xi') - (\text{tr}\Sigma)^2}{\text{tr}\Phi_1 + \text{tr}\Phi_2 + \xi'\xi - 2\text{tr}\Sigma}.$$

Similarly, we extend this to matrix form and obtain the following theorems.

Theorems 2.1 Suppose that $X \sim N_p(\theta, \Phi_1), Y \sim N_p(\theta + \xi, \Phi_2),$

$\text{cov}(X, Y) = \Sigma$, when $B = \Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma'$ is positively definite matrix, the matrix, which makes the risk of the matrix-weighted combined estimator $\delta_A^{mc}(X, Y) = AX + (I - A)Y$ reach minimum, is

$$A_0 = (\Phi_2 + \xi\xi' - \Sigma')(\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}.$$

The risk of $\delta_{A_0}^{mc}(X, Y)$ is

$$\text{tr}([\Phi_1(\Phi_2 + \xi\xi') - \Sigma'(\Phi_2 + \xi\xi') + (\Phi_2 + \xi\xi')\Sigma' - \Sigma'\Sigma](\Phi_1 + \Phi_2 + \xi\xi' - \Sigma - \Sigma')^{-1}).$$

Theorem 2.2 let $\Sigma' = \Sigma$, $\Phi_1 - \Sigma > 0$ and $\Phi_2 + \xi\xi' - \Sigma > 0$, then

$$MSE(\delta_{A_0}^{mc}(X, Y)) \leq MSE(\delta_{a_0}^{lc}(X, Y)).$$

Similar to that in Chapter 1, we investigate the relationships between ξ and matrix A_0 and propose a condition for ξ being characteristic vector of A_0 .

Since ξ , Φ_1 , Φ_2 and Σ are unknown, the relevant estimators of $\delta_{a_0}^{lc}(X, Y)$ and $\delta_{A_0}^{mc}(X, Y)$ are presented as :

$$\hat{\delta}^{lc}(X, Y) = \bar{X} - \frac{n \sum_{i=1}^n (X_i - \bar{X})'(X_i - \bar{X}) - n \sum_{i=1}^n (X_i - \bar{X})'(Y_i - \bar{Y})}{(n-1) \sum_{i=1}^n (X_i - Y_i)'(X_i - Y_i)} (\bar{X} - \bar{Y}),$$

$$\begin{aligned} & \hat{\delta}^{mc}(X, Y) \\ &= \bar{X} - [n \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' - n \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})'] [(n-1) \sum_{i=1}^n (X_i - Y_i) \\ & \times (X_i - Y_i)']^{-1} (\bar{X} - \bar{Y}). \end{aligned}$$

We compare the relationships between the above two risks of $\hat{\delta}^{lc}(X, Y)$ and $\hat{\delta}^{mc}(X, Y)$ by means of numerical simulation, and we also discuss the effects of parameter variations on the ratio of the two risks and practicality of our method.

3. James-Stein estimator of Matrix form

In this chapter, we also extend James-Stein estimator to matrix form and obtain the following two theorems,

Theorem 3.1 Let X be p -variate normal random vector $E(X) = \theta$, $\text{cov}(X, X) = \sigma^2 I$, then

$$MSE(\delta^{mJS}) \leq MSE(\delta^{JS}).$$

Theorem 3.2 Let X be p -variate normal random vector

$$E(X) = \theta, \quad \text{cov}(X, X) = \sigma^2 I, \quad \text{then} \quad \delta(p) = \left(1 - \frac{p\sigma^2}{\theta'\theta + p\sigma^2}\right)X$$

minimizes the risk $MSE(\delta_b(X))$ of $\delta_b(X) = (1-b)X$. Meanwhile,

$\delta^m(p) = (I - \sigma^2(\theta'\theta + p\sigma^2)^{-1})X$ minimizes the risk $MSE(\delta_B^m(X))$ of $\delta_B^m(X) = (I - B)X$.

Since θ is unknown, we obtain the relevant estimators of δ^{JS} and δ^{mJS} :

$$\tilde{\delta}^{JS} = (1 - \frac{(p-2)\sigma^2}{\frac{1}{n} \sum_{i=1}^n X_i' X_i}) \bar{X},$$

$$\tilde{\delta}^{mJS} = (I - \frac{p-2}{p} \sigma^2 (\frac{1}{n} \sum_{i=1}^n X_i X_i')^{-1}) \bar{X}.$$

Furthermore, we compare the risks between them by virtue of numerical simulation and obtain the practical range of our method.

4. Further studies on Stein estimator

This part we primarily compare $\hat{\delta}^s(a,b) = (1 - \frac{a\sigma^2}{b + X'X})X$ with $\hat{\delta}(a) = (1 - \frac{a}{X'X})X$, lead to the risk of $\hat{\delta}^s(a,b) = (1 - \frac{a\sigma^2}{b + X'X})X$, and obtain the following theorem:

Theorem 4.1 Let X be p -variate normal random vector $E(X) = \theta$, $cov(X, X) = \sigma^2 I$. When $0 < a < 2(p-2)$, $p \geq 3$, $b \geq 0$, the risk of $\hat{\delta}^s(a,b)$ is less than that of X . The risk

$MSE(\hat{\delta}^S(a,b))$ of $\hat{\delta}^S(a,b)$ is

$$p\sigma^2 + (a^2 - 2a(p-2))\sigma^4 E \frac{1}{b+XX} - b(a^2 + 4a)\sigma^4 E \frac{1}{(b+XX)^2}.$$

Finally, by means of numerical simulation, we compare the relationships of the risks of $\hat{\delta}^S(a,b)$ and $\hat{\delta}(a)$ and indicate the existence of a and b which makes $MSE(\hat{\delta}^S(a,b)) \leq MSE(\hat{\delta}(a))$.