

文章编号: 1000-5862(2015)06-0599-06

高维协方差矩阵估计方法的比较

李小雪, 明瑞星*

(浙江工商大学统计与数学学院, 浙江 杭州 310018)

摘要: 通过模拟比较门限估计方法和收缩估计方法之间的差异, 得出2种方法在实际应用中的使用范围. 由模拟结果可知, 若有确切的证据表明总体协方差矩阵是稀疏矩阵, 则采用门限估计方法, 否则, 采用稳健的收缩估计方法比较恰当.

关键词: 高维协方差矩阵; 稀疏矩阵; 非稀疏矩阵; 门限估计; 收缩估计

中图分类号: F 224 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.06.11

0 引言

设 $X_k = (X_{k1}, X_{k2}, \dots, X_{kp})^T$, $k = 1, 2, \dots, n$ 为来自 p 元正态总体 $N_p(\mu, \Sigma)$ 的独立样本. 以 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ 表示样本均值,

$$S = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})(X_k - \bar{X})^T = (s_{ij})_{p \times p}$$

表示样本协方差矩阵. 众所周知, 当 p 很小时, S 是 Σ 的一个“好”的估计^[1]. 当 p 较大时, S 不是 Σ 的一个“好”的估计^[2-3]. 特别地, 当 p 大于 n 时(此时的协方差矩阵估计称为高维协方差矩阵估计), S 是一个病态矩阵, 它显然不是 Σ 的一个好的估计. 要想得到 Σ 的好的估计, 需要对 S 进行正则化处理, 如 thresholding 方法^[4-8], tapering 方法^[9-15] 和 banding 方法^[16-20]. 事实上, 上述提及的各种方法都需要对协方差矩阵的结构做出相应的假设. 在因子模型的协方差矩阵估计中, 文献[21]假定误差协方差矩阵是一对角矩阵, 文献[22]假定误差协方差矩阵是一稀疏矩阵, 但是在实际应用中, 如何验证误差协方差矩阵是对角矩阵还是稀疏矩阵是一件非常困难的事情. 其次, 对总体协方差矩阵的这些假设一旦与事实相违背, 则得到的最终估计不是一个好的估计, 其均方误差(MSE)比用 S 去估计 Σ 还要大. 自然地, 当对协方差矩阵的认识不够充分时, 会选择一种稳健的估计. 收缩估计正是一种稳健的估计, 它可以不依赖

于对协方差矩阵的任何假设. 收缩估计最初由 C. Stein 提出^[22], 在高维情形下, O. Ledoit 等^[23] 建议用

$$\hat{\Sigma}_s = \lambda I + (1 - \lambda) S, \lambda \in [0, 1] \quad (1)$$

作为协方差矩阵 Σ 的估计, 其中 I 是单位矩阵. 如果对协方差矩阵有着先验的认识(这种先验认识用矩阵 T 表示), J. Schafer 等^[3] 建议用 T 取代(1)式中的 I , 即

$$\hat{\Sigma}_s = \lambda T + (1 - \lambda) S, \lambda \in [0, 1]. \quad (2)$$

本质上, 对协方差矩阵的各种假设是基于对其的一些先验认识, 从这种意义上讲, 收缩估计中 T 的选取也是对协方差矩阵做出某种假设的表现. 相对于对协方差矩阵做出的假设而言, 收缩估计中 T 的选取具有很大的灵活性, 具体参见文献[3].

本文对门限估计和收缩估计进行模拟研究, 为应用统计工作者如何使用门限估计和收缩估计提出一些建议.

1 门限估计

P. J. Bickel 等^[4] 提出用硬门限估计方法来估计总体协方差矩阵 $\Sigma = (\sigma_{ij})_{p \times p}$, 其基本原理是: 设定某个门槛值 θ , 当 $|s_{ij}| < \theta$ 时, σ_{ij} 的估计值为 0; 否则, σ_{ij} 的估计值为 s_{ij} . A. J. Rothman 等^[5] 在文献[4]的基础上把门限估计中的硬门限扩展到了一般门限函数, 即 $\forall \theta \geq 0$, 定义函数 $s_\theta: \mathbf{R} \rightarrow \mathbf{R}$, $\forall z \in \mathbf{R}$, s_θ 满足 (i) $|s_\theta(z)| \leq |z|$; (ii) 对于 $|z| \leq$

收稿日期: 2015-08-10

基金项目: 浙江省高校人文社科重点研究基地(统计学), 浙江省自然科学基金(LY16A01001)和浙江省教育厅课题(1020KZ0413455)资助项目.

通信作者: 明瑞星(1975-), 男, 湖北黄石人, 副教授, 博士, 主要从事金融统计与高维数据分析的研究.

$\theta, s_\theta(z) = 0$; (iii) $|s_\theta(z) - z| \leq \theta$. 文献[5]提出的广义门限估计包含许多门限估计,如硬门限估计、软门限估计、SCAD 门限估计和 adaptive LASSO 门限估计等. 这几种门限估计的原理相似,下面以软门限估计为例来简单阐述其估计原理.

门限估计方法中,最为关键的一个假定是总体协方差矩阵是稀疏矩阵,即总体协方差矩阵的许多非对角元素为 0 或者接近 0. 在软门限估计中,门限函数 $s_\theta(z) = \text{sign}(z) (|z| - \theta)_+$, 其中 $(\cdot)_+$ 表示对于实数 a , $(a)_+ = \max(a, 0)$, 则有

$$\Sigma_t = s_\theta(s_{ij})_{p \times p} = (\text{sign}(s_{ij}) (|s_{ij}| - \theta)_+)_{p \times p}.$$

通常把 θ 称为门限参数,文献[5]采用文献[4]中提出的交叉验证法来确定 θ ,具体步骤:把 n 个样本随机分成 2 部分,第 1 部分的样本个数为 $n_1 = n - n/\log n$ 和第 2 部分的样本个数为 $n_2 = n - n_1$,重复这种分割 N 次,令 S_{1v} 和 S_{2v} 分别表示当第 v 次分割时由第 1 部分的样本和第 2 部分的样本得到的样本协方差矩阵,其中 $v = 1, 2, \dots, N$, 从而有

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{v=1}^N \|s_\theta(S_{1v}) - S_{2v}\|_F^2,$$

则总体协方差矩阵 Σ 的软门限估计为

$$\hat{\Sigma}_t = s_{\hat{\theta}}(s_{ij})_{p \times p} = (\text{sign}(s_{ij}) (|s_{ij}| - \hat{\theta})_+)_{p \times p}.$$

2 收缩估计

O. Ledoit 等^[23]和 J. Schafer 等^[3]提出用(2)式作为总体协方差矩阵的收缩估计矩阵. 在(2)式中最关键的就是 λ 的确定和矩阵 T 的选取. 下面先阐述如何确定 λ . 考虑 $\hat{\Sigma}_s$ 的均方误差(M_{SE}),其均方误差可以分解为

$$M_{SE}(\hat{\Sigma}_s) = B_{ias}(\hat{\Sigma}_s)^2 + \text{Var}(\hat{\Sigma}_s).$$

如果 $\hat{\Sigma}_s$ 是总体协方差矩阵 Σ 的 1 个无偏估计,即 $B_{ias}(\hat{\Sigma}_s) = 0$,则减小均方误差就等价于减小 $\text{Var}(\hat{\Sigma}_s)$. 注意到 M_{SE} 是关于 λ 的一个函数,不妨记为 $R(\lambda)$,此外记 $T = (t_{ij})_{p \times p}$, 这样,

$$R(\lambda) = \sum_{i=1}^p \sum_{j=1}^p \{ \text{Var}(\lambda t_{ij} + (1-\lambda)s_{ij}) + [E(\lambda t_{ij} + (1-\lambda)s_{ij}) - \sigma_{IJ}]^2 \}.$$

将上式展开,

$$R(\lambda) = \sum_{i=1}^p \sum_{j=1}^p \{ \lambda^2 \text{Var}(t_{ij}) + (1-\lambda)^2 \text{Var}(s_{ij}) + 2\lambda(1-\lambda) \text{Cov}(t_{ij}, s_{ij}) + [\lambda E(t_{ij} - s_{ij}) + B_{ias}(s_{ij})]^2 \}.$$

对 $R(\lambda)$ 求导,得

$$\lambda = [\sum_{i=1}^p \sum_{j=1}^p (\text{Var}(s_{ij}) - \text{Cov}(t_{ij} - s_{ij}) - B_{ias}(s_{ij}) \cdot E(t_{ij} - s_{ij}))] / [\sum_{i=1}^p \sum_{j=1}^p E[(t_{ij} - s_{ij})^2]].$$

因为样本协方差阵 S 是总体协方差矩阵的 1 个无偏估计,即 $B_{ias}(s_{ij}) = 0$, 所以有

$$\lambda = [\sum_{i=1}^p \sum_{j=1}^p (\text{Var}(s_{ij}) - \text{Cov}(t_{ij} - s_{ij}))] / [\sum_{i=1}^p \sum_{j=1}^p E[(t_{ij} - s_{ij})^2]].$$

由于 λ 的取值范围为 $[0, 1]$, 所以其极值点或者端点就是取最值的点, 所以

$$\hat{\lambda} = \min(\max(0, \lambda), 1).$$

收缩估计的另外一个关键问题是如何选取目标矩阵 T . 文献[3]给出了 T 的 6 种情形. 文献[24]也给出了 T 的 9 种选法. 其实,目标矩阵如何选取没有统一的定论,需要根据实际情况选取. 一般而言,目标矩阵应该是 1 个低维的矩阵,其包含较少的参数. 本文仅选取 3 种常见的目标矩阵进行研究,其它情形可类似地进行研究.

(i) T 为单位矩阵. 记 $T_a = I$, 总体协方差矩阵的

收缩估计为 $\hat{\Sigma}_{sa} = \lambda T_a + (1-\lambda)S$.

(ii) T 为对角矩阵. 记

$$T_b = (t_{ij})_{p \times p} = \begin{cases} s_{ij}, & i = j, \\ 0, & i \neq j, \end{cases}$$

总体协方差矩阵的收缩估计为

$$\hat{\Sigma}_{sb} = \lambda T_b + (1-\lambda)S.$$

(iii) T 为非对角矩阵. 记

$$T_c = (t_{ij})_{p \times p} = \begin{cases} s_{ij}, & i = j, \\ r \sqrt{s_{ii}s_{jj}}, & i \neq j, \end{cases}$$

其中 r 表示样本相关系数矩阵的平均值,总体协方差矩阵的收缩估计为 $\hat{\Sigma}_{sc} = \lambda T_c + (1-\lambda)S$.

3 模拟

前面分别给出了门限估计和收缩估计的理论基础,本部分通过模拟来具体的比较这 2 种估计方法的优劣. 具体的模拟过程如下: 先从均值为零向量、协方差矩阵为 Σ_0 的多元正态分布中产生 50 个样本(即 $n = 50$),再根据前述的理论确定门限方法的估计矩阵 $\hat{\Sigma}_t$ 和收缩方法的估计矩阵 $\hat{\Sigma}_{sa}$, $\hat{\Sigma}_{sb}$ 和 $\hat{\Sigma}_{sc}$. 将上述过程重复 30 次. 用矩阵的 Frobenius 范数度量估计矩阵和总体协方差矩阵 Σ_0 之间的差

异. 设定 Σ_0 具有下面 3 种结构: Σ_0 为稀疏矩阵、非稀疏矩阵或者随机矩阵(此时 Σ_0 的元素是随机产生的, 无法确定其稀疏性).

3.1 Σ_0 为稀疏矩阵

类似文献 [6] 中模型 1 产生如下的 Σ_0 :

表 1 当 Σ_0 为稀疏矩阵时的估计效果

| p | $\ S - \Sigma_0\ _F$ | $\ \hat{\Sigma}_t - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sa} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sb} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sc} - \Sigma_0\ _F$ |
|-----|----------------------|-----------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| 2 | 0.749 706 | 0.746 830 | 1.016 662 | 0.768 169 | 1.623 049 |
| 5 | 2.465 439 | 2.049 045 | 2.708 103 | 1.733 708 | 2.992 748 |
| 10 | 4.633 239 | 3.009 409 | 4.288 424 | 2.407 053 | 3.878 976 |
| 20 | 9.344 986 | 5.033 152 | 7.100 850 | 3.524 470 | 4.634 618 |
| 50 | 25.076 830 | 12.471 950 | 16.570 070 | 9.956 072 | 10.471 230 |
| 100 | 49.126 630 | 22.899 820 | 30.979 550 | 24.073 720 | 24.239 240 |
| 200 | 98.360 930 | 43.904 870 | 60.218 750 | 53.079 700 | 53.164 350 |
| 300 | 152.928 000 | 67.918 060 | 93.124 660 | 85.590 600 | 85.640 410 |
| 400 | 203.239 700 | 90.482 560 | 122.596 800 | 115.098 800 | 115.135 800 |
| 500 | 247.420 700 | 103.907 100 | 147.899 600 | 140.525 300 | 140.551 500 |

若总体协方差矩阵为稀疏矩阵, 则由表 1 可知, 当维数特别低时(相对于样本量而言, 如 $n = 2$) 样本协方差矩阵是一个好的估计. 当维数比较低时($p < n$), 门限估计方法比收缩估计方法的表现要好, 而收缩估计方法比样本协方差矩阵的表现要好. 当维数 p 大于样本量 n 时, 门限估计方法也比其他方法表现的要好. 另外, 在 3 种收缩估计中, T 为单位矩阵时表现的比其他 2 种收缩估计要差, 而 T 为对角矩阵和非对角矩阵的表现很接近, 几乎相同. 图 1 更加直观地体现了这些结论.

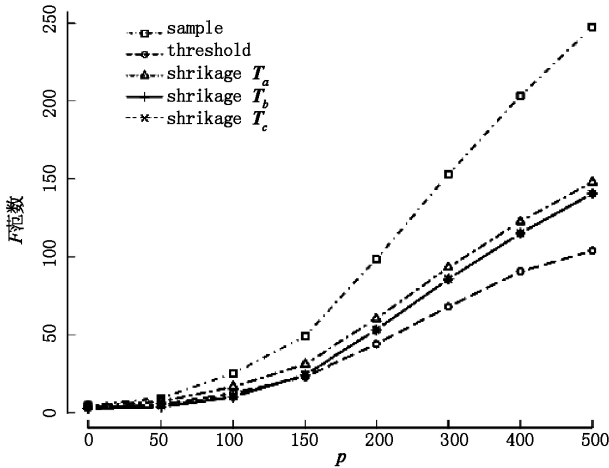


图 1 当 Σ_0 为稀疏矩阵时各种估计方法的表现

由图 1 可以看出, 由于 T 为对角矩阵和非对角矩阵的表现很接近, 所以在图 1 中 2 条线合成 1 条线, 而门限估计比其他方法的表现都要好. 所以, 当

$$\Sigma_0 = (\sigma_{ij}^0)_{p \times p} = \begin{cases} h_i, & i = j, \\ 0.3^{1|i-j|}, & i \neq j, \end{cases}$$

其中 h_1, \dots, h_p 均是取自均匀分布 $U(1, 5)$ 的随机数. 模拟结果见表 1.

总体协方差矩阵是稀疏矩阵时, 选用门限估计方法最为合适.

3.2 Σ_0 为非稀疏矩阵

首先叙述产生非稀疏矩阵的过程. 从正态分布 $N(0, 5)$ 中产生 Σ_0 的对角元素, 从均匀分布 $U(2, 5)$ 中产生 Σ_0 的非对角元素, 得出 Σ_0 后进行谱分解, $\Sigma_0 = \sum_{i=1}^p \lambda_i \mu_i \mu_i'$, 其中 λ_i 为 Σ_0 的特征值, μ_i 为 λ_i 对应的特征向量. 如果 λ_i 为负值或 0, 则用 1 来代替, 即 $\hat{\lambda}_i = \max(\lambda_i, 1)$. 此时, $\Sigma_0 = \sum_{i=1}^p \hat{\lambda}_i \mu_i \mu_i'$, 上面产生 Σ_0 用到的正态分布和均匀分布的参数以及 λ_i 的替代值都可以进行修改, 但是要注意一点, 产生的 Σ_0 的非对角元素的值不能太小也不能为 0, 这是为了保证 Σ_0 为一非稀疏矩阵. 模拟结果见表 2.

若总体协方差矩阵为非稀疏矩阵, 则由表 2 可知, 当维数特别低时(相对于样本量而言, 如 $n = 2$), 样本协方差矩阵比其他估计方法表现的要好, 是一个好的估计. 当维数比较低时($p < n$), 门限估计方法比样本协方差矩阵表现的差, 但是差别不是太大. 当维数 p 大于样本量 n 时, 门限估计方法比样本协方差矩阵的表现差很多, 且样本协方差矩阵比 3 种收缩估计方法的表现差很多. 这表明在总体协方差矩阵是非稀疏矩阵时, 门限估计方法不再适用, 而收缩估计方法比样本协方差矩阵的估计效果要好的多. 具体如图 2 所示.

表 2 当 Σ_0 为非稀疏矩阵时的估计效果

| p | $\ S - \Sigma_0\ _F$ | $\ \hat{\Sigma}_t - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sa} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sb} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sc} - \Sigma_0\ _F$ |
|-----|----------------------|-----------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| 2 | 0.837 686 | 0.907 051 | 0.991 598 | 1.142 594 | 1.180 436 |
| 5 | 3.339 797 | 3.664 142 | 3.384 116 | 3.500 350 | 3.695 759 |
| 10 | 8.137 748 | 8.281 267 | 7.964 004 | 8.013 193 | 9.001 249 |
| 20 | 13.608 750 | 15.352 020 | 13.751 460 | 13.796 020 | 13.540 260 |
| 50 | 42.649 590 | 44.765 930 | 42.112 170 | 42.145 620 | 40.930 300 |
| 100 | 85.958 390 | 91.417 720 | 84.919 140 | 84.946 990 | 76.713 390 |
| 200 | 180.306 900 | 193.087 200 | 176.470 900 | 176.464 500 | 158.032 900 |
| 300 | 334.022 400 | 339.773 100 | 324.450 500 | 324.439 600 | 281.302 100 |
| 400 | 432.133 200 | 468.268 000 | 425.445 800 | 425.440 900 | 332.726 900 |
| 500 | 588.745 800 | 604.403 000 | 569.064 500 | 569.001 300 | 449.255 400 |

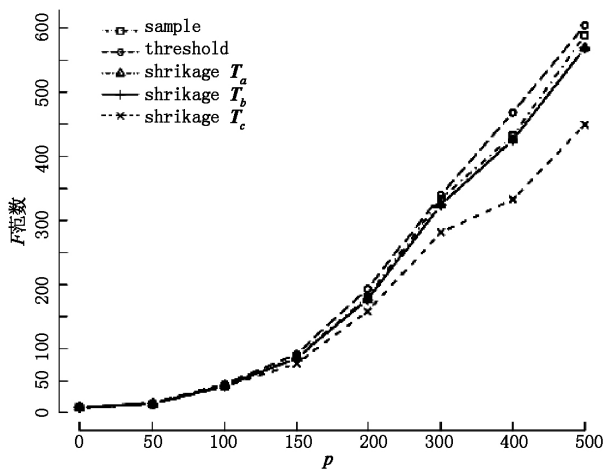


图 2 当 Σ_0 为非稀疏矩阵时各种估计方法的表现

由图 2 可知, 门限估计方法的表现最差, 当 $T = T_c$ 时的收缩估计方法表现最好, 当 $T = T_a$ 和 $T = T_b$ 时的收缩估计方法的表现很接近, 2 条线重合. 所以, 当总体协方差矩阵是非稀疏矩阵时, 选用 $T = T_c$ 时的收缩估计方法最为合适.

3.3 Σ_0 为随机矩阵

随机矩阵的产生类似于非稀疏矩阵的产生. 首先从均匀分布 $U(5, 10)$ 中产生 Σ_0 的对角元素, 其次从指数分布 $\text{Exp}(\eta)$ 中产生 Σ_0 的非对角元素, 其中 η 从均匀分布 $U(0.1, 5)$ 中产生. 由于 η 的随机性, 从指数分布 $\text{Exp}(\eta)$ 中得到的非对角元素也具有随机性. 得到 Σ_0 后为保证其正定性, 类似于非随机矩阵的产生过程, 也进行谱分解. 模拟结果见表 3.

表 3 当 Σ_0 为随机矩阵时的估计效果

| p | $\ S - \Sigma_0\ _F$ | $\ \hat{\Sigma}_t - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sa} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sb} - \Sigma_0\ _F$ | $\ \hat{\Sigma}_{sc} - \Sigma_0\ _F$ |
|-----|----------------------|-----------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| 2 | 2.427 617 | 2.389 344 | 2.781 474 | 2.490 552 | 4.145 089 |
| 5 | 5.927 414 | 5.412 157 | 6.173 041 | 7.715 932 | 7.841 086 |
| 10 | 10.376 620 | 8.845 102 | 9.855 346 | 8.918 269 | 9.542 903 |
| 20 | 26.222 050 | 22.977 250 | 23.098 140 | 22.455 870 | 22.566 600 |
| 50 | 71.535 910 | 61.046 550 | 58.260 000 | 53.671 870 | 51.392 660 |
| 100 | 182.952 800 | 157.302 800 | 141.779 600 | 129.788 600 | 122.720 000 |
| 200 | 423.712 900 | 368.940 900 | 307.030 800 | 285.050 900 | 268.079 000 |
| 300 | 734.674 500 | 650.298 600 | 515.673 600 | 485.016 200 | 459.311 200 |
| 400 | 1 128.743 000 | 1 007.333 000 | 767.009 000 | 725.718 800 | 692.092 600 |
| 500 | 1 546.033 000 | 1 387.433 000 | 1 035.672 000 | 987.033 100 | 947.327 200 |

若总体协方差矩阵为随机矩阵, 则由表 3 可知, 当维数特别低时 (如 $n=2$ 或 5) 样本协方差矩阵是一个好的估计. 当维数比较低时 ($p < n$) 门限估计

方法和收缩估计方法的表现相差不大且比样本协方差矩阵的表现要好. 当维数 p 大于样本量 n 时, 门限估计方法比样本协方差矩阵的表现要好, 但相较于

收缩估计,其表现要差.在3种收缩估计中,当 $T = T_a$ 时的表现最差,当 $T = T_b$ 时次之,当 $T = T_c$ 时的表现最好.具体如图3所示.

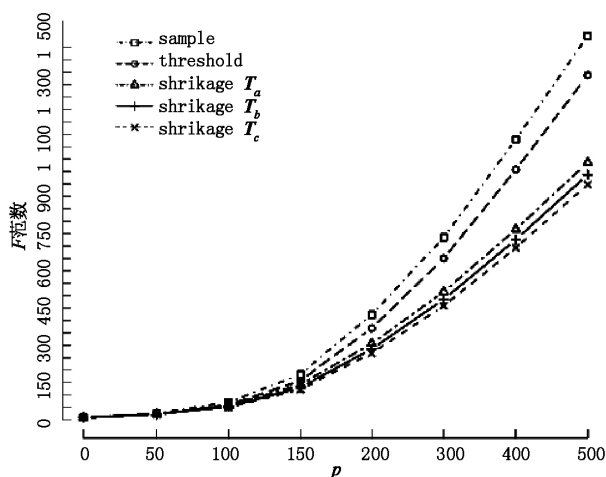


图3 Σ_0 为随机矩阵时各种估计方法的表现

由图3可知,样本协方差矩阵的表现最差,当 $T = T_c$ 时的收缩估计方法表现最好.所以,当无法确定总体协方差矩阵的稀疏性时,选用当 $T = T_c$ 时的收缩估计方法最为合适.

4 结论

通过模拟给出如下的一些建议:当维数特别低时(相对于样本量而言),用样本协方差矩阵估计总体协方差矩阵即可.当有确切的证据表明总体协方差矩阵是稀疏矩阵时,用门限估计方法比较好.如果无法确定总体协方差矩阵的稀疏性,为了保险起见,用收缩估计方法比较好.关于收缩估计中矩阵 T 的选取,由第3部分的模拟可知,当 $T = T_c$ 时的表现或者和其他情形接近,或者比其他情形要好,所以在使用收缩估计方法时,建议选择 $T = T_c$.

最后论述由门限估计和收缩估计这2种方法联想的一些问题.文献[3]中求 λ 是通过令均方误差最小得到 λ 的解析解,在门限估计方法中是通过交叉验证法求出门限参数 θ ,由此受到启发,也可以用交叉验证法来求出 λ ,此时有

$$\tilde{\lambda} = \underset{\lambda}{\operatorname{argmin}} \frac{1}{N} \sum_{p=1}^N \|\lambda T + (1 - \lambda) S_{1_p} - S_{2_p}\|_F^2.$$

则 Σ 的收缩估计为 $\tilde{\Sigma}_s = \tilde{\lambda} T + (1 - \tilde{\lambda}) S$.

一个自然的问题是 $\hat{\Sigma}_s$ 和 $\tilde{\Sigma}_s$ 之间是否存在差异?如果存在,哪种估计的效果比较好?在文献[7-8]中假定 Σ_u 是一稀疏矩阵,据此用门限估计方法来估计 Σ_u ,是否可以考虑把该假定去掉,用收缩估

计方法来估计 Σ_u ?这2种方法之间是否存在差异?如果存在,孰优孰劣?后者的理论方面是否可以得到证明?这都需要进一步研究.

5 参考文献

- [1] Anderson T W. An introduction to multivariate statistical analysis [M]. 3rd ed. New York: John Wiley and Sons, 2003.
- [2] 白志东,郑术蓉,姜丹丹.大维统计分析[M].北京:高等教育出版社,2012.
- [3] Schafer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics [J]. Statistical Applications in Genetics and Molecular Biology 2005 4(1): 1-32.
- [4] Bickel P J, Levina E. Covariance regularization by thresholding [J]. Annals of Statistics, 2008, 36(6): 2577-2604.
- [5] Rothman A J, Levina E, Zhu Ji. Generalized thresholding of large covariance matrices [J]. Journal of the American Statistical Association 2009, 104(485): 177-186.
- [6] Cai Tony T, Zhang Cunhui, Zhou Harrison H. Optimal rates of convergence for covariance matrix estimation [J]. Annals of Statistics 2010, 38(4): 2118-2144.
- [7] Fan Jianqing, Liao Yuan, Mincheva M. High dimensional covariance matrix estimation in approximate factor models [J]. Annals of Statistics 2011, 39(6): 3320-3356.
- [8] Fan Jianqing, Liao Yuan, Mincheva M. Large covariance estimation by thresholding principal orthogonal complements [J]. Journal of the Royal Statistical Society 2011, 75(4): 603-680.
- [9] Furrer R, Bengtsson T. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants [J]. Journal of Multivariate Analysis 2007, 98(2): 227-255.
- [10] Bickel P J, Levina E. Regularized estimation of large covariance matrices [J]. Annals of Statistics 2008, 36(1): 199-227.
- [11] Wu Weibiao, Pourahmadi M. Banding sample autocovariance matrices of stationary processes [J]. Statistica Sinica 2009, 19(4): 1755-1768.
- [12] Cai Tony T, Liu Weidong. Adaptive thresholding for sparse covariance matrix estimation [J]. Journal of the American Statistical Association 2011, 106(494): 672-684.
- [13] Rothman A J, Levina E, Zhu Ji. A new approach to chol- esky-based covariance regularization in high dimensions [J]. Biometrika 2010, 97(3): 539-550.
- [14] Anderes E, Huser R, Nychka D, et al. Nonstationary positive definite tapering on the plane [J]. Journal of Compu-

- tational and Graphical Statistics 2013 22(4): 848-865.
- [15] Stein M L. Statistical properties of covariance tapers [J]. Journal of Computational and Graphical Statistics ,2013 , 22(4): 866-885.
- [16] Bickel P J ,Levina E. Some theory for Fisher's linear discriminant function, "naive Bayes," and some alternatives when there are many more variables than observations [J]. Bernoulli 2004 ,10(6): 989-1010.
- [17] Qiu Yumou ,Chen Songxi. Test for bandedness of high-dimensional covariance matrices and bandwidth estimation [J]. Annals of Statistics 2012 40(3): 1285-1314.
- [18] Qiu Yumou ,Chen Songxi. Band width selection for high dimensional covariance estimation [J]. Journal of the American Statistical Association 2014 ,119(1): 1-35.
- [19] Yi Feng ,Zou Hui. Sure-tuned tapering estimation of large covariance matrices [J]. Computational Statistics and Data Analysis 2013 58(1): 339-351.
- [20] Bien J ,Bunea F ,Xiao Luo. Convex banding of the covariance matrix [J]. Journal of the American Statistical Association 2014 ,doi: 10.1080/01621459.2015.1058265.
- [21] Fan Jianqing ,Fan Yingying ,Lv Jinchi. High dimensional covariance matrix estimation using a factor model [J]. Journal of Econometrics 2008 ,147(1): 186-197.
- [22] Stein C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution [J]. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability ,1956 ,1: 197-206.
- [23] Ledoit O ,Wolf M. A well-conditioned estimator for large-dimensional covariance matrices [J]. Journal of Multivariate Analysis 2004 88(2): 365-411.

The Comparison of Methods for Estimating the High-Dimensional Covariance Matrices

LI Xiaoxue ,MING Ruixing^{*}

(School of Statistics and Mathematics Zhejiang Gongshang University ,Hangzhou Zhejiang 310018 ,China)

Abstract: The differences between the thresholding estimation and the shrinking estimation are reported by a series of simulations and the proper estimation is proposed within these two estimations in practice. The simulations show that if the population covariance matrix is a sparse matrix ,the thresholding estimation is better than that of the shrinking estimation and vice versa.

Key words: high-dimensional covariance matrix; sparse matrix; non-sparse matrix; thresholding estimation; shrinkage estimation

(责任编辑: 曾剑锋)