数理解析研究所講究録
第 1999 巻 2016 年 95-98

ℓ_p-norm based James-Stein estimation with minimaxity and sparsity

丸山　祐造

YUZO MARUYAMA

東京大学・空間情報科学研究センター

CENTER FOR SPATIAL INFORMATION SCIENCE, THE UNIVERSITY OF TOKYO *

Abstract

d 変量正規分布の平均ベクトルの二乗損失関数のもとでの推定問題を考える. $d\geq 3$ のときには, Stein 現象が生じて, 最尤推定量は非許容的になる. このとき James-Stein positive-part 推定量 (JSPP) は 1 つの改良型推定量として知られている. JSPP 推定量をモデル選択の枠組みで考えるとき, null model か full model の二択になっていることが欠点である. Zhou and Hwang (2005) は, 縮小関数を ℓ_2 norm の関数でなく $\ell_p$norm の関数とすることによって 2 つの候補からのモデル選択を可能にし, また同時にミニマクス性を持つ縮小型推定量を提案した. 本稿では, Zhou and Hwang (2005) の結果を拡張して, 彼らが p に課していた制約を除き, 任意の正なる p を用いた ℓ_p norm の関数でミニマクス性とスパース性を併せ持つ推定量を構成できることが示される.

ところで James-Stein positive-part 推定量は, 経験ベイズ推定量として解釈できる. Zhou and Hwang も彼らの推定量がある種のベイズ推定量として解釈できることを示したが, 理論的に不完全である. ℓ_p norm を縮小関数とする縮小型推定量のベイズ的解釈を与えることは今後の課題としたい.

1 イントロダクション

Let $Z\sim N_d(\theta, I_d)$. We are interested in estimation of the mean vector θ with respect to the quadratic loss function $L(\delta,\theta)=\sum_{i=1}^{d}(\delta_i-\theta_i)^2$. Obviously the risk of z is d. We shall say one is as good as the other if the former has a risk no greater than the latter for every θ. Moreover, one dominates the other if it is as good as the other and has smaller risk for some θ. In this case, the latter is called inadmissible. Note that z is a minimax estimator, that is, it minimizes $\sup_\theta E[L(\delta,\theta)]$ among all estimators δ. Consequently any δ is as good as z if and only if it is minimax.

Stein (1956) showed that z is inadmissible when $d\geq 3$. James and Stein (1961) explicitly found a class of minimax estimators $\hat{\theta}_{\mathrm{JS}}=\left(1-c/\|z\|_2^2\right)z$ with $0\leq c\leq 2(d-2)$ and $\|z\|_2^2=\sum_{i=1}^{d}z_i^2$. Baranchik (1964) proposed the James-Stein positive-part estimator

$$\hat{\theta}_{\mathrm{JS}}^{+}=\max\left(0,1-c/\|z\|_2^2\right)z \tag{1.1}$$

*maruyama@csis.u-tokyo.ac.jp

with $0 < c \le 2(d-2)$ which dominates the James-Stein estimator. The problem with the James-Stein positive-part estimator is, however, that it selects only between two models: the origin and the full model. Zhou and Hwang (2005) overcome the difficulty by utilizing the so-called ℓ_p-norm given by

$$\|z\|_p = \left\{\sum_{i=1}^d |z_i|^p\right\}^{1/p} \tag{1.2}$$

and in fact proposed minimax estimators $\hat{\theta}^+_{\mathrm{ZH}}$ with the i-th component given by

$$\hat{\theta}^+_{i\mathrm{ZH}} = \max\left(0, 1 - c/\left\{\|z\|_{2-\alpha}^{2-\alpha}|z_i|^\alpha\right\}\right) z_i \tag{1.3}$$

where $0 \le \alpha < (d-2)/(d-1)$ and $0 < c \le 2\{(d-2) - \alpha(d-1)\}$. When $\alpha > 0$ and

$$|z_i| \le \left\{c/\|z\|_{2-\alpha}^{2-\alpha}\right\}^{1/\alpha}, \tag{1.4}$$

the i-th component of the estimator is zero, which implies that the choice between a full model and reduced models where some coefficients are reduced to zero is possible.

In this paper, we establish minimaxity of a new class of ℓ_p-norm based shrinkage estimators $\hat{\theta}^+_{\mathrm{LP}}$ with the i-th component given by

$$\hat{\theta}^+_{i\mathrm{LP}} = \max\left(0, 1 - c/\left\{\|z\|_p^{2-\alpha}|z_i|^\alpha\right\}\right) z_i \tag{1.5}$$

where $0 \le \alpha < (d-2)/(d-1)$, $p > 0$, $0 < c \le 2(d-2)\gamma(d,p,\alpha)$ and

$$\gamma(d,p,\alpha) = \min(1, d^{(2-p-\alpha)/p})\left\{1 - \alpha(d-1)/(d-2)\right\}.$$

When α is strictly positive in (1.5), sparsity happens as in (1.4). In Zhou and Hwang (2005), $p = 2 - \alpha$ was assumed and the ℓ_p-norm with

$$d/(d-1) < p[= 2-\alpha] < 2$$

seems only applicable for constructing estimators with minimaxity and sparsity simultaneously. We show that it is not so but ℓ_p-norm with any positive p is available for that purpose. As an extreme case ($p = \infty$), we can show that

$$\max\left(0, 1 - 2\frac{(d-2) - \alpha(d-1)}{d\left\{\max|z_i|\right\}^{2-\alpha}|z_i|^\alpha}\right) z_i$$

with $0 \le \alpha < (d-2)/(d-1)$ is minimax. A more general result of minimaxity, corresponding to the result of Efron and Morris (1976), where c is replaced by $\phi(\|z\|_p)$ in (1.5), is given in Section 2.

2 ミニマクス性とスパース性を併せ持つ推定量

In this section, we establish minimaxity result of the shrinkage estimators $\hat{\theta}_\phi$ with the i-th component given by

$$\hat{\theta}_{i\phi} = \left(1 - \phi(\|z\|_p)/\left\{\|z\|_p^{2-\alpha}|z_i|^\alpha\right\}\right) z_i. \tag{2.1}$$

Note the shrinkage factor of (2.1), $1-\phi(\|z\|_p)/\{\|z\|_p^{2-\alpha}|z_i|^\alpha\}$ is symmetric with respect to z_i. As shown in Theorem 4 of Zhou and Hwang (2005), the shrinkage estimator with the symmetry is dominated by the positive-part estimator. Hence the minimaxity of $\hat{\theta}_\phi^+$ follows from the minimaxity of $\hat{\theta}_\phi$.

Under the assumption that $\phi(v)$ is absolutely continuous, so called Stein's (1981) unbiased risk estimator is available.

Lemma 2.1 *Assume $\phi(v)$ is absolutely continuous.*

1. *The risk function of the estimator $\hat{\theta}_\phi$ is*

$$E\left[\|\hat{\theta}_\phi-\theta\|_2^2\right]=d+E\left[\phi(\|z\|_p)\psi_\phi(z)\|z\|_p^{\alpha-p-2}\sum_i|z_i|^{p-\alpha}\right] \tag{2.2}$$

where

$$\begin{aligned}\psi_\phi(z)=\phi(\|z\|_p)\|z\|_p^{p+\alpha-2}\frac{\sum_i|z_i|^{2(1-\alpha)}}{\sum_i|z_i|^{p-\alpha}}-2(1-\alpha)\|z\|_p^p\frac{\sum_i|z_i|^{-\alpha}}{\sum_i|z_i|^{p-\alpha}}\\ -2\left\{\alpha-2+\|z\|_p\phi'(\|z\|_p)/\phi(\|z\|_p)\right\}.\end{aligned} \tag{2.3}$$

2. *Assume $0\le\alpha\le 1$. Then $\psi_\phi(z)\le\Psi_\phi(\|z\|_p)$ where*

$$\Psi_\phi(v)=\max(1,d^{(p+\alpha-2)/p})\phi(v)-2\{d-2-\alpha(d-1)\}-2v\phi'(v)/\phi(v).$$

Assume $d\ge 3$ and $0\le\alpha<(d-2)/(d-1)$. Let

$$\gamma(d,p,\alpha)=\min(1,d^{(2-p-\alpha)/p})\{1-\alpha(d-1)/(d-2)\} \tag{2.4}$$

which is positive from the assumptions. By Lemma 2.1, a sufficient condition for $E[\|\hat{\theta}-\theta\|_2^2]\le d$ with $\phi\ge 0$ is $\Psi_\phi(v)\le 0$ for all $v\ge 0$. Clearly $\phi(v)=c$ where $0<c\le 2(d-2)\gamma(d,p,\alpha)$ with satisfies $\Psi_\phi(v)\le 0$. More generally, by the derivative,

$$\frac{d}{dv}\left\{\frac{v^b\phi(v)}{a-\phi(v)}\right\}=\frac{v^{b-1}\phi(v)}{\{a-\phi(v)\}^2}\left(a\frac{v\phi'(v)}{\phi(v)}+ba-b\phi(v)\right), \tag{2.5}$$

we have a following sufficient condition for minimaxity as in Efron and Morris (1976).

Theorem 2.1 *Assume $d\ge 3$ and $0\le\alpha<(d-2)/(d-1)$. Assume $\phi(v)$ is absolutely continuous and*

$$0\le\phi(v)\le 2(d-2)\gamma(d,p,\alpha)$$

where $\gamma(d,p,\alpha)$ is given by (2.4). Further, for all v with $\phi(v)<2(d-2)\gamma(d,p,\alpha)$

$$g_\phi(v)=\frac{v^{d-2-\alpha(d-1)}\phi(v)}{2(d-2)\gamma(d,p,\alpha)-\phi(v)}$$

is assumed to be non-decreasing. Further if there exists $v_*>0$ such that $\phi(v)=2(d-2)\gamma(d,p,\alpha)$, then $\phi(v)$ is assumed equal to $2(d-2)\gamma(d,p,\alpha)$ for all $v\ge v_*$. Then $\hat{\theta}_\phi$ is minimax.

Recall that ℓ_p norm with any positive p is available in Lemma 2.1 and Theorem 2.1. As an extreme case ($p=\infty$), we have $\lim_{p\to\infty}\gamma(d,p,\alpha)=\{1-\alpha(d-1)/(d-2)\}/d$ and hence

$$\max\left(0,1-2\frac{(d-2)-\alpha(d-1)}{d\{\max|z_i|\}^{2-\alpha}|z_i|^{\alpha}}\right)z_i$$

with $0\le\alpha<(d-2)/(d-1)$ is minimax.

Remark 2.1 *The solution of $\Psi_\phi(v)=0$ or $g_\phi(v)=1/\lambda$ for any $\lambda>0$, is*

$$\phi_{\mathrm{DS}}(v)=\frac{2(d-2)\gamma(d,p,\alpha)}{1+\lambda v^{d-2-\alpha(d-1)}},$$

under which Dasgupta and Strawderman (1997) showed the risk of the estimator with $\phi_{\mathrm{DS}}(v)$ is exactly equal to d when $p=2$ and $\alpha=0$. Actually it is related to the concept of "nearly unbiasedness" or "approximately unbiasedness" in the literature of SCAD (smoothly clipped absolute deviation) including Antoniadis and Fan (2001). Since $\phi_{\mathrm{DS}}(v)$ is monotone decreasing and approaches 0 as $v\to\infty$, unnecessary modeling biases are effectively avoided with $\phi_{\mathrm{DS}}(v)$.

参考文献

Antoniadis, A. and Fan, J. (2001) "Regularization of wavelet approximations," *J. Amer. Statist. Assoc.*, Vol. 96, No. 455, pp. 939–967, With discussion and a rejoinder by the authors.

Baranchik, A. J. (1964) "Multiple regression and estimation of the mean of a multivariate normal distribution,"Technical Report 51, Department of Statistics, Stanford University.

Dasgupta, A. and Strawderman, W. E. (1997) "All estimates with a given risk, Riccati differential equations and a new proof of a theorem of Brown," *Ann. Statist.*, Vol. 25, No. 3, pp. 1208–1221.

Efron, B. and Morris, C. (1976) "Families of minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, Vol. 4, No. 1, pp. 11–21.

James, W. and Stein, C. (1961) "Estimation with quadratic loss," in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, Berkeley, Calif.: Univ. California Press, pp. 361–379.

Stein, C. (1956) "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pp. 197–206, Berkeley and Los Angeles: University of California Press.

——— (1981) "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, Vol. 9, No. 6, pp. 1135–1151.

Zhou, H. H. and Hwang, J. T. G. (2005) "Minimax estimation with thresholding and its application to wavelet analysis," *Ann. Statist.*, Vol. 33, No. 1, pp. 101–125.