Ensemble minimaxity of James-Stein estimators

Yuzo Maruyama | Lawrence D. Brown | Edward I. George

1Graduate School of Business Administration, Kobe University, Kobe, Hyogo, 657-8501, Japan
2Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA

Correspondence
Yuzo Maruyama, Graduate School of Business Administration, Kobe University, 2-1, Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan.
Email: maruyama@port.kobe-u.ac.jp

This article considers the estimation of a multivariate normal mean based on heteroscedastic observations. Under heteroscedasticity, estimators that shrink more on coordinates with larger variances seem desirable. Although they are not necessarily minimax in the ordinary sense, we show that certain James-Stein type estimators can be ensemble minimax, that is, minimax with respect to the ensemble risk considered in the empirical Bayes perspective of Efron and Morris.

KEYWORDS
ensemble minimaxity, minimaxity, stein estimation

1 | INTRODUCTION

Let \( X \sim N_p(\theta, \Sigma) \), where \( p \geq 3, \theta = (\theta_1, \ldots, \theta_p)^T \) and \( \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) \). Let us assume

\[
\sigma_1^2 > \sigma_2^2 > \ldots > \sigma_p^2. \tag{1.1}
\]

We are interested in the estimation of \( \theta \) with respect to the ordinary squared error loss function

\[
L(\delta, \theta) = ||\delta - \theta||^2,
\]

where the risk of an estimator \( \delta(X) \) is \( R(\delta, \theta) = E[L(\delta, \theta)] \). The estimator \( \delta \) is said to be minimax if

\[
\sup_{\theta \in \Theta} R(\delta, \theta) = \inf_{\delta, \theta} \sup_{\theta \in \Theta} R(\delta, \theta).
\]

As in Theorem 1 in Section 2, the MLE \( X \) with constant risk \( \sum \sigma_i^2 \) is minimax for any \( p \) and any \( \Sigma \).

In the homoscedastic case \( \sigma_1^2 = \ldots = \sigma_p^2 \), James and Stein (1961) showed that the shrinkage estimator

\[
\left( 1 - \frac{c}{X^T \Sigma^{-1} X} \right) X \text{ for } c \in (0, 2(p-2)) \tag{1.2}
\]

dominates the MLE \( X \) for \( p \geq 3 \). But in literature discussing the minimax properties of shrinkage estimators under heteroscedasticity, Brown (1975) showed that the James-Stein estimator (1.2) may not be minimax when the variances are not equal. Specifically, it is not minimax for any \( c \in (0, 2(p-2)) \) when \( 2\sigma_1^2 > \sum_{i=1}^p \sigma_i^2 \). Berger (1976) showed that

\[
\left( I - \Sigma^{-1} \frac{c}{X^T \Sigma^{-1} X} \right) X \text{ for } c \in (0, 2(p-2)) \tag{1.3}
\]
is minimax for \( p \geq 3 \) and any \( \Sigma \). However, Casella (1980) argued that the James-Stein estimator (1.3) may not be desirable even if it is minimax. Ordinary minimax estimators, as in (1.3), typically shrink most on the coordinates with smaller variances. From Casella’s (1980) viewpoint, one of the most natural James-Stein variants seems to be
\[
\left(1 - \frac{c}{|X|^2}\right)X \text{ for } c > 0.
\]

Efron and Morris (1971, 1972a, 1972b, 1973) considered the compound risk under the prior \( \theta \sim N_p(0, rI_p) \) with \( r \in (0, \infty) \), ‘ensemble risk’, which is given by
\[
\mathbb{R}(\mathbf{r}, r) = \mathbb{E}_{\mathbf{r}} \left[ \frac{1}{|2\pi r|^p/2} \exp\left( -\frac{||\theta||^2}{2r} \right) \right] \mathbf{r}.
\]

With a set of ensemble risks \( \{\mathbb{R}(\mathbf{r}, r) : r \in (0, \infty)\} \), we can define ensemble minimaxity with respect to a set of priors
\[
\mathcal{P} = \{N_p(0, rI_p) : r \in (0, \infty)\};
\]
that is, an estimator \( \delta \) is said to be ensemble minimax with respect to \( \mathcal{P} \), if
\[
\sup_{r \in (0,\infty)} \mathbb{R}(\mathbf{r}, r) = \inf_{\delta \in \mathcal{P}} \sup_{r \in (0,\infty)} \mathbb{R}(\delta, r).
\]

As a matter of record, the second author in his unpublished manuscript (Brown et al., 2011) previously introduced the concept of ensemble minimaxity. In this article, we follow their spirit but propose a simpler and clearer approach for establishing the ensemble minimaxity of estimators. We note in passing that ensemble minimaxity can also be interpreted as a particular case of Gamma minimaxity studied in the context of robust Bayes analysis by Berger (1979) and Good (1952). However, in such studies, a ‘large’ set consisting of many diffuse priors are usually included in the analysis. Since this is quite different from our formulation of the problem, we use the term ‘ensemble minimaxity’ throughout our paper, following the Efron and Morris papers cited above.

Our article is organized as follows. In Section 2, we review the minimaxity results and explain Casella’s (1980) viewpoint on the incompatibility between minimaxity and well-conditioning. In Section 3, we establish the ensemble minimaxity of various shrinkage estimators including a variant of the James-Stein estimator
\[
\left(1 - \frac{p-2}{(p-2)\sigma_i^2 + |X|^2}\right)X,
\]
as well as the generalized Bayes estimator with respect to the hierarchical prior
\[
\theta|\lambda \sim N_p(0, (\sigma_i^2/\lambda)I - \Sigma), \sigma(\lambda) \sim \lambda^{-2}f_{0,1}(\lambda),
\]
which is a generalization of Stein’s (1974) harmonic prior \( ||\theta||^{-p} \) for the heteroscedastic case.

## 2 \ MINIMAXITY AND CASELLA’S VIEWPOINT

Recall, for \( i = 1, \ldots, p \),
\[
x_i|\theta_i \sim N(\theta_i, \sigma_i^2), \text{ and } \theta_i \sim N(0, r).
\]

Then, the posterior and marginal distributions are given by
\[ \theta_i | x_i \sim N\left( \frac{\tau}{\tau + \sigma_i^2} x_i, \frac{\sigma_i^2}{\tau + \sigma_i^2} \right) \text{ and } x_i \sim N(0, \tau + \sigma_i^2), \]  

(2.1)

respectively, where \( \theta_1, \ldots, \theta_p \mid x \) are mutually independent and \( x_1, \ldots, x_p \) are mutually independent. For the posterior mean of \( \theta \),

\[ \hat{\theta}_i = \left( \frac{\tau}{\tau + \sigma_i^2} \right) x_i, \ldots, \left( \frac{\tau}{\tau + \sigma_p^2} \right) x_p, \]

which is the Bayes estimator under quadratic loss, we have

\[ R(\hat{\theta}, \tau) = \sum_{i=1}^{p} E_{\theta} E_{x} \left[ \left( \frac{\tau x_i}{\tau + \sigma_i^2} - \theta_i \right)^2 \right] = \sum_{i=1}^{p} E_{\theta} \left[ \left( \frac{\tau x_i}{\tau + \sigma_i^2} - \theta_i \right)^2 \right] = \sum_{i=1}^{p} \text{Var}(\theta_i | x_i) = \sum_{i=1}^{p} \frac{\sigma_i^2}{\tau + \sigma_i^2}. \]  

(2.2)

Then, with (2.2), we have the following well-known result.

**Theorem 1.** The estimator \( X \) is minimax.

**Proof.** For any estimator \( \delta \) and any \( \tau > 0 \), we have

\[ \sup_{\theta} R(\delta, \theta) \geq R(\delta, \tau) \geq R(\hat{\theta}, \tau) = \sum_{i=1}^{p} \left( \frac{\tau}{\tau + \sigma_i^2} \right) \sigma_i^2, \]

where the equality follows from (2.2). Hence,

\[ \sup_{\theta} R(\delta, \theta) \geq \sum_{i=1}^{p} \sigma_i^2 = R(X, \theta) = \sup_{\theta} R(X, \theta), \]

which completes the proof. \( \square \)

A class of shrinkage estimators which we consider in this paper, is given by

\[ \delta_g = \left( I - G \frac{\phi(z)}{z} \right) x, \text{ for } z = x^T \Sigma^{-1} x = \sum_{i=1}^{p} \frac{z_i^2}{\sigma_i^2}. \]  

(2.3)

where \( G = \text{diag}(g_1, \ldots, g_p) \). Berger and Srinivasan (1978) showed, in their Corollary 2.7, that, given positive-definite \( C \) and non-singular \( B \), a necessary condition for an estimator of the form

\[ \left( I - B \frac{\phi(x^T C x)}{x^T C x} \right) x \]

to be admissible is \( B = k \Sigma C \) for some constant \( k \), which is satisfied by estimators among the class of (2.3).

Now, consider the assumptions

A1 \( 0 \leq g_i \leq 1 \) for any \( i \).

A2 \( \phi(z) \geq 0 \).

A3 \( \phi \) is absolutely continuous.

A4 \( \phi \) is monotone non-decreasing.
Let

\[ h(\Sigma, G) = 2 \left( \frac{\sum g_i \sigma_i^2}{\max(g_i \sigma_i^2)} - 2 \right). \]

Then, we have the following result, a version of Baranchik's (1964) sufficient condition for ordinary minimaxity.

**Theorem 2.** Suppose \( G \) satisfies A1 and \( h(\Sigma, G) > 0 \). Then, \( \delta_\phi \) is conventional minimax if \( \phi \) satisfies A2, A3, A4 and \( \phi \leq h(\Sigma, G) \).

**Proof.** The risk of \( \delta_\phi \) is expanded as

\[
R(\delta_\phi, \theta) = \sum_{i=1}^{p} \mathbb{E} \left[ \left( \frac{g_i (\phi(\frac{z}{z^2}) \chi_i - \theta_i)^2}{z^2} \right) \right] - 2 \sum_{i=1}^{p} \mathbb{E} \left[ g_i \sigma_i^2 \left( \frac{\phi(z)}{z} + \frac{2 g_i x_i}{\sigma_i^2} \left( \frac{\phi(z)}{z} - \frac{\phi(z)}{z^2} \right) \right) \right]
\]

Noting \( z = \sum g_i \chi_i^2 / \sigma_i^2 \) and using Stein's (1981) identity under Assumption A3, we have

\[
R(\delta_\phi, \theta) - \sum \sigma_i^2 = - \mathbb{E} \left[ \sum g_i \frac{\phi(z)}{z^2} \chi_i^2 \right] - 2 \sum \mathbb{E} \left[ g_i \sigma_i^2 \left( \frac{\phi(z)}{z} + \frac{2 g_i x_i}{\sigma_i^2} \left( \frac{\phi(z)}{z} - \frac{\phi(z)}{z^2} \right) \right) \right]
\]

By A1, we have

\[
\frac{z}{\sum g_i \chi_i^2 / \sigma_i^2} = \frac{\sum g_i \chi_i^2 / \sigma_i^2}{\sum \{g_i \sigma_i^2 \} \{g_i \chi_i^2 / \sigma_i^2 \}} \geq \frac{1}{\max(g_i \sigma_i^2)} \tag{2.4}
\]

Then, by Assumptions A2 and A4, and (2.4), we have

\[
R(\delta_\phi, \theta) - \sum \sigma_i^2 \leq - \mathbb{E} \left[ \sum g_i \frac{\phi(z)}{z^2} \chi_i^2 (h(\Sigma, G) - \phi(z)) \right],
\]

which completes the proof. \( \square \)

For a given \( G \) which satisfies

\[ h(\Sigma, G) = 2 \left( \frac{\sum g_i \sigma_i^2}{\max(g_i \sigma_i^2)} - 2 \right) > 0, \]

Berger (1976) showed that, for any \( \Sigma \),

\[
\max_{\Sigma} h(\Sigma, G) = 2(p - 2), \quad \arg \max_{\Sigma} h(\Sigma, G) = \Sigma^{-1} = \diag \left( \frac{\sigma_1^2}{\sigma_1^2}, \ldots, \frac{\sigma_p^2}{\sigma_p^2}, 1 \right). \tag{2.5}
\]

which seems the right choice of \( G \) since the upper bound for minimaxity, \( 2(p - 2) \) in (2.5) is the largest in the sense that

\[ \max_{\Sigma} h(\Sigma, G) = 2(p - 2). \]

However, from the ‘conditioning’ viewpoint of Casella (1980), which advocates for more shrinkage on higher variance estimates, the descending order
is desirable, whereas $G = \sigma_2^2 \Sigma^{-1}$ corresponding to the ascending order $g_1 < \ldots < g_p$ under $\Sigma$ given by (1.1). As Casella (1980) pointed out, ordinary minimaxity cannot be enjoyed together with well-conditioning given by (2.6) when

$$h(\Sigma, cl) \leq 0 \text{ or equivalently } \sum_{i=1}^{p} \sigma_i^2 \leq 2 \sigma_1^2$$

for some $0 < c \leq 1$. In fact, when $h(\Sigma, cl) \leq 0$ and $c = g_1 > \ldots > g_p$, we have

$$c \sigma_1^2 = g_1 \sigma_1^2, c \sigma_2^2 > g_2 \sigma_2^2, \ldots, c \sigma_p^2 > g_p \sigma_p^2,$$

and hence, $h(\Sigma, G) < 0$ follows. The motivation of Casella (1980, 1985) seems to provide a better treatment for the case. Actually, Brown (1975) pointed out essentially the same phenomenon from a slightly different viewpoint.

Ensemble minimaxity, based on ensemble risk given by (1.4), provides a way of justifying shrinkage estimators with well-conditioning, estimators which are not necessarily ordinary minimax.

3 | ENSEMBLE MINIMAXITY

3.1 | A general theorem for ensemble minimaxity

As in Theorem 1, the ensemble minimaxity of the MLE $X$, with the constant risk $\sum_{i=1}^{p} \sigma_i^2$, is established as follows.

**Theorem 3.** The estimator $X$ is ensemble minimax.

**Proof.** For any estimator $\delta$ and any prior on $\tau$, we have

$$\sup_{\tau} R(\delta, \tau) \geq \int R(\delta, \tau) \pi(\tau) d\tau \geq \sum_{i=1}^{p} \left( \frac{\tau}{\tau + \sigma_i^2} \right) \sigma_i^2.$$

Hence,

$$\sup_{\tau} R(\delta, \tau) \geq \sum_{i=1}^{p} \sigma_i^2 = R(X, \tau) = \sup_{\tau} R(X, \tau),$$

which completes the proof.

Next, we have the following theorem on the ensemble minimaxity of $\delta_\phi$ for general $G$, though we will eventually focus on $\delta_\phi$ for $G$ restricted to the descending order $g_1 > \ldots > g_p$ as in (2.6).

**Theorem 4.** Assume $\phi(z)$ is non-negative, non-decreasing and concave. Assume also that $\phi(z)/z$ is non-increasing. Then,

$$\delta_\phi = \left( I - G \frac{\phi(z)}{z} \right) x, \text{ for } z = x^T G \Sigma^{-1} x = \sum_{i=1}^{p} \frac{g_i x_i^2}{\sigma_i^2}$$

is ensemble minimax if

$$\phi(p \min_{i} (1 + \tau/\sigma_i^2)) \leq 2(p - 1) \frac{\min_{i} (1 + \tau/\sigma_i^2)}{\max_{i} (1 + \tau/\sigma_i^2)} \quad \forall \tau \in (0, \infty).$$

(3.1)
Proof. As in (2.1), the posterior and marginal distributions are

$$
\theta_i | x_i \sim N \left( \frac{x_i}{\tau + \sigma_i^2}, \frac{\tau \sigma_i^2}{\tau + \sigma_i^2} \right) \quad \text{and} \quad x_i \sim N(0, \tau + \sigma_i^2),
$$

respectively, where $\theta_1 | x_1, \ldots, \theta_p | x_p$ are mutually independent and $x_1, \ldots, x_p$ are mutually independent. Then, the Bayes risk is given by

$$
R(\delta, \tau) = \sum_{i=1}^{p} E_{x_i} \left[ \left( 1 - g_i(\theta) x_i - \theta_i \right)^2 \right] \\
= \sum_{i=1}^{p} E_{x_i} \left[ \left( 1 - g_i(\theta) x_i - \theta_i \right)^2 \right] \\
= \sum_{i=1}^{p} E_{x_i} \left[ \left( 1 - g_i(\theta) x_i - E(\theta_i | x_i) + E(\theta_i | x_i) - \theta_i \right)^2 \right] \\
= \sum_{i=1}^{p} E_{x_i} \left[ \left( 1 - g_i(\theta) x_i - E(\theta_i | x_i) \right)^2 \right] + \sum_{i=1}^{p} \text{Var}(\theta_i | x_i) \\
= \sum_{i=1}^{p} E_{x_i} \left[ \left( \frac{\tau \sigma_i^2}{\tau + \sigma_i^2} x_i - g_i(\theta) x_i \right)^2 \right] + \sum_{i=1}^{p} \frac{\tau \sigma_i^2}{\tau + \sigma_i^2}.
$$

Since the first term of the r.h.s. of the above equality is rewritten as

$$
\sum_{i=1}^{p} E_{x_i} \left[ \left( \frac{\tau \sigma_i^2}{\tau + \sigma_i^2} x_i - g_i(\theta) x_i \right)^2 \right] = \sum_{i=1}^{p} \left( \frac{\tau \sigma_i^2}{\tau + \sigma_i^2} \right)^2 E_{x_i} [x_i^2] - 2E_{x_i} \left[ \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i x_i^2 \phi(z)}{\tau + \sigma_i^2} \right] + E_{x_i} \left[ \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i^2 x_i^2 \phi^2(z)}{\tau + \sigma_i^2} \right]
$$

we have

$$
R(\delta, \tau) - \sum_{i=1}^{p} \sigma_i^2 = -2E_{x_i} \left[ \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i x_i^2 \phi(z)}{\tau + \sigma_i^2} \right] + E_{x_i} \left[ \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i^2 x_i^2 \phi^2(z)}{\tau + \sigma_i^2} \right].
$$

Let

$$
w_i = \frac{x_i^2}{\tau + \sigma_i^2}, w = \sum_{i=1}^{p} w_i, \text{ and } t_i = \frac{w_i}{w} \text{ for } i = 1, \ldots, p.
$$

Then,

$$
w \sim \chi^2_p, t = (t_1, \ldots, t_p)^T \sim \text{Dirichlet}(1/2, \ldots, 1/2),
$$

and $w$ and $t$ are mutually independent. With the notation, we have

$$
x_i^2 = wt_i(\sigma_i^2 + \tau) \quad \text{and} \quad z = x^T \Sigma^{-1} x = \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i x_i^2}{\tau + \sigma_i^2} = w \sum_{i=1}^{p} t_i g_i \left( 1 + \frac{\tau}{\sigma_i^2} \right),
$$

and hence,

$$
E_{x_i} \left[ \sum_{i=1}^{p} \frac{\tau \sigma_i^2 g_i x_i^2 \phi^2(z)}{\tau + \sigma_i^2} \right] = E_{w, t} \left[ \frac{\sum t_i g_i (\sigma_i^2 + \tau) \phi(w \sum t_i g_i (1 + \tau/\sigma_i^2))}{\sum t_i g_i (1 + \tau/\sigma_i^2)} \right] \frac{\sum t_i g_i (1 + \tau/\sigma_i^2)}{w \sum t_i g_i (1 + \tau/\sigma_i^2)}
$$

Since $\phi(w)/w$ is non-increasing and $\phi(w)$ is non-decreasing, by the correlation inequality, we have
Given \( \Sigma \), the choice \( G = \Sigma / \sigma_n^2 \) with descending order \( g_1 > \ldots > g_p \), is one of the most natural choices of \( G \) from Casella’s (1980) viewpoint. In this case, we have
\[
\frac{\min g_i(1 + \tau/\sigma_i^2)}{\max g_i(1 + \tau/\sigma_i^2)} = \frac{\min \{\sigma_i^2 + \tau\}}{\max \{\sigma_i^2 + \tau\}} = \frac{\sigma_p^2 + \tau}{\sigma_1^2 + \tau},
\]
\[
p \min g_i(1 + \tau/\sigma_i^2) = \frac{p}{\sigma_1^2} \min \{\sigma_i^2 + \tau\} = \frac{\sigma_p^2 + \tau}{\sigma_1^2}.
\]
and hence the following corollary.

**Corollary 1.** Assume that \(\phi(z)\) is non-negative, non-decreasing and concave. Assume also that \(\phi(z)/z\) is non-increasing. Then,

\[
\delta_\phi = \left(1 - \frac{\mathbf{1}^\top \phi(\|\mathbf{x}\|^2/\sigma_1^2)}{\|\mathbf{x}\|^2}\right) \mathbf{x}
\]
is ensemble minimax if

\[
\phi(p(\sigma_p^2 + \tau)/\sigma_1^2) \leq 2(p - 2)\frac{\sigma_p^2 + \tau}{\sigma_1^2 + \tau}, \quad \forall \tau \in (0, \infty).
\]

### 3.2 An ensemble minimax James-Stein variant

As an example of Corollary 1, we consider

\[
\phi(z) = \frac{c_1 z}{c_2 + z}
\]
for \(c_1 > 0\) and \(c_2 \geq 0\), which is motivated by Stein (1956) and James and Stein (1961). Under \(\Sigma = I_p\), Stein (1956) suggested that there exist estimators dominating the usual estimator \(\mathbf{x}\) among a class of estimators \(\delta_\phi\) with \(\phi\) given by (3.8) for small \(c_1\) and large \(c_2\). Following Stein (1956), James and Stein (1961) showed that \(\delta_\phi\) with \(0 < c_1 < 2(p - 2)\) and \(c_2 = 0\) is ordinary minimax. The choice \(c_2 = 0\) is, however, not good since, by Corollary 1, \(c_1\) cannot be larger than \(2(p - 2)\frac{\sigma_p^2}{\sigma_1^2}\). With positive \(c_2\), we can see as follows, that \(c_1\) can be much larger.

Note that \(\phi(z)\) given by (3.8) is non-negative, increasing and concave and that \(\phi(z)/z\) is decreasing. Then the sufficient condition in (3.7) is

\[
\frac{c_1 p(\sigma_p^2 + \tau)/\sigma_1^2}{c_2 + p(\sigma_p^2 + \tau)/\sigma_1^2} \leq 2(p - 2)\frac{\sigma_p^2 + \tau}{\sigma_1^2 + \tau}, \quad \forall \tau \in (0, \infty),
\]
which is equivalent to

\[
2(p - 2)\left\{\sigma_1^2 c_2 + p(\sigma_p^2 + \tau)\right\} - c_1 p(\sigma_p^2 + \tau) \geq 0 \quad \forall \tau \in (0, \infty)
\]
or

\[
p \tau \left\{2(p - 2) - c_1\right\} + 2(p - 2)c_2^2 \left(\sigma_1^2 + \frac{c_1}{2(p - 2)} - \frac{\sigma_p^2}{\sigma_1^2}\right) \geq 0 \quad \forall \tau \in (0, \infty).
\]

Hence, we have a following result.
Theorem 5. 1. When
\[ 0 < c_1 \leq 2(p-2) \text{ and } c_2 \geq \max \left( 0, p \left( \frac{c_1}{2(p-2)} - \frac{\sigma^2}{\sigma^2_1} \right) \right). \]

the shrinkage estimator
\[ \left( I - \Sigma \frac{c_1}{c_2 \sigma^2_1 + ||x||^2} \right) x \]
is ensemble minimax.

2. It is ordinary minimax if
\[ 2 \left( \sum \sigma^2 / \sigma^2_1 - 2 \right) \geq c_1. \]

Part 2 above follows from Theorem 2.

One of the most interesting estimators with ensemble minimaxity from Part 1 is
\[ \left( I - \Sigma \frac{p-2}{(p-2)\sigma^2_1 + ||x||^2} \right) x \]
with the choice \( c_1 = c_2 = p - 2 \) satisfying (3.9). It is clear that the \( i \)th shrinkage factor
\[ 1 - \frac{(p-2)\sigma^2_i}{(p-2)\sigma^2_1 + ||x||^2} \]
is nonnegative for any \( x \) and any \( \Sigma \), which is an appealing property.

3.3 | A generalized Bayes ensemble minimax estimator

In this subsection, we provide a generalized Bayes ensemble minimax estimator. Following Strawderman (1971), Berger (1976), and Maruyama and Strawderman (2005), we consider the generalized Stein’s (1974) prior
\[ \theta \sim N_p(0, \lambda^{-1} \Sigma G^{-1} - \Sigma), \lambda \sim \lambda^{-2} \Gamma(\alpha,1)(\lambda), \]
where \( G = \text{diag}(g_1, \ldots, g_p) \) satisfies \( 0 < g_i \leq 1 \). Note that for \( \Sigma = G = I_p \), the density of \( \theta \) is exactly \( \pi(\theta) = ||\theta||^{-2-p} \), since \( \lambda^{-1} \Sigma G^{-1} - \Sigma = (1 - \lambda) I_p \) and
\[ \frac{1}{(2\pi)^{p/2}} \int_0^{1/\lambda} \left( \frac{1}{1-\lambda} \right)^{p/2} \exp \left( -\frac{\lambda ||\theta||^2}{2(1-\lambda)} \right) \lambda^{-2} d\lambda = \frac{1}{(2\pi)^{p/2}} \int_0^\infty g^{2p/2} \exp \left( -g ||\theta||^2 / 2 \right) dg \]
\[ = \Gamma(p/2 - 1)2^{-p/2} \frac{||\theta||^{-2-p}}{(2\pi)^{p/2}}. \]

The prior \( \pi(\theta) = ||\theta||^{-2-p} \) is called Stein’s harmonic prior and was originally investigated by Baranchik (1964) and Stein (1974). Berger (1980), and Berger and Strawderman (1996) recommended the use of the prior (3.10) mainly because it is on the boundary of admissibility.

By the way of Strawderman (1971), the generalized Bayes estimator with respect to the prior is given by

\[ \frac{1}{(2\pi)^{p/2}} \int_0^{1/\lambda} \left( \frac{1}{1-\lambda} \right)^{p/2} \exp \left( -\frac{\lambda ||\theta||^2}{2(1-\lambda)} \right) \lambda^{-2} d\lambda = \frac{1}{(2\pi)^{p/2}} \int_0^\infty g^{2p/2} \exp \left( -g ||\theta||^2 / 2 \right) dg \]
\[ = \Gamma(p/2 - 1)2^{-p/2} \frac{||\theta||^{-2-p}}{(2\pi)^{p/2}}. \]
\[
\delta_* = \left(1 - G\frac{\phi_*}{z}\right)x, \text{ for } z = x^T\Sigma^{-1}x
\]

with

\[
\phi_* = \frac{\int_0^1 \lambda^{n/2-1} \exp(-z\lambda/2) d\lambda - \int_0^1 \lambda^{n/2-2} \exp(-z\lambda/2) d\lambda}{\int_0^1 \lambda^{n/2-1} \exp(-z\lambda/2) d\lambda},
\]

where \( \phi_* \) satisfies the following properties

H1 \( \phi_* \) is increasing in \( z \).
H2 \( \phi_* \) is concave.
H3 \( \lim_{z\to\infty} \phi_* (z) = p - 2 \).
H4 \( \phi_* (z)/z \) is decreasing in \( z \).
H5 The derivative of \( \phi_* \) at \( z = 0 \) is \( (p - 2)/p \).

Under the choice \( G = \Sigma/\sigma^2_1 \) with the conditions of Corollary 1, we have the following result.

**Theorem 6.** 1. The estimator \( \delta_* \) is ensemble minimax.
2. The estimator \( \delta_* \) is ordinary minimax when

\[
2\left(\sum \sigma_i^4/\sigma_1^4 - 2\right) \geq p - 2.
\]

3. The estimator \( \delta_* \) is conventional admissible.

**Proof.** [Part 1] Recall that a sufficient condition for ensemble minimaxity is given by Corollary 1. By H1–H5, we have only to check (3.7) in Corollary 1. For \( r \geq \max(0, \sigma^2_1 - 2\sigma^2_0) \), we have

\[
2(p - 2)\frac{\sigma_0^2 + r}{\sigma_1^4 + r} \geq p - 2.
\]

By the properties H1 and H3,

\[
\phi_*\left(p\frac{\sigma_0^2 + r}{\sigma_1^4}ight) \leq p - 2
\]

for \( r \in (0, \infty) \). Hence, for \( r \geq \max(0, \sigma^2_1 - 2\sigma^2_0) \), it follows that

\[
\phi_*\left(p\frac{\sigma_0^2 + r}{\sigma_1^4}ight) \leq 2(p - 2)\frac{\sigma_0^2 + r}{\sigma_1^4 + r}.
\]

So it suffices to show

\[
\phi_*\left(p\frac{\sigma_0^2 + r}{\sigma_1^4}ight) \leq 2(p - 2)\frac{\sigma_0^2 + r}{\sigma_1^4 + r}
\]

when \( \sigma^2_1 - 2\sigma^2_0 > 0 \) and \( 0 < r < \sigma^2_1 - 2\sigma^2_0 \). By the properties H2 and H5, we have \( \phi_* (z) \leq (p - 2)/p \) for all \( z \geq 0 \). Then,
which completes the proof.

[Part 2] follows from Theorem 2.

[Part 3] follows from Theorem 6.4.2 of Brown (1971). □

3.4 | A numerical experiment

Let \( p = 10 \) and

\[
\Sigma = \text{diag}(a^2, a^3, \ldots, a^9, 1)
\]

for \( a = 1.01, 1.05, 1.25, 1.5 \). Approximately \( a^9 \) is 1.09, 1.55, 7.45, 38.4, respectively. We investigate the numerical performance of two ensemble minimax estimators of the form

\[
\delta_\phi = \left( I - \frac{\phi(||x||^2/\sigma^2_1)}{||x||^2/\sigma^2_1} \right) x,
\]

where one is the generalized Bayes estimator (GB) with

\[
\phi_+ (z) = \int_0^1 z^{p/2-1} \exp(-z/2) \frac{dz}{1/2^{p/2-2} \exp(-z/2)}
\]

and the other is the James-Stein variant (JS) with

\[
\phi_{JS}(z) = \left( \frac{p-2}{p-2+z} \right)^4.
\]

### Table 1: Ordinary risk differences

| a  | m  | 0  | 2  | 20 | 40  | 60  | 80  | 100 |
|----|----|----|----|----|-----|-----|-----|-----|
| GB | 1.01 | 0.79 | 0.14 | 1.7 x 10^{-3} | 4.8 x 10^{-4} | 2.5 x 10^{-4} | 1.7 x 10^{-4} | 1.3 x 10^{-4} |
|    | 1.05 | 0.75 | 0.14 | 1.7 x 10^{-3} | 4.3 x 10^{-4} | 2.0 x 10^{-4} | 1.2 x 10^{-4} | 8.0 x 10^{-5} |
|    | 1.25 | 0.63 | 0.19 | 1.9 x 10^{-3} | 2.5 x 10^{-4} | -5.6 x 10^{-5} | -1.7 x 10^{-4} | -2.2 x 10^{-4} |
|    | 1.5  | 0.63 | 0.27 | 2.7 x 10^{-3} | 1.6 x 10^{-4} | -3.0 x 10^{-4} | -4.6 x 10^{-4} | -5.4 x 10^{-4} |
| JS | 1.01 | 0.80 | 0.14 | 1.7 x 10^{-3} | 4.8 x 10^{-4} | 2.5 x 10^{-4} | 1.7 x 10^{-4} | 1.3 x 10^{-4} |
|    | 1.05 | 0.79 | 0.14 | 1.7 x 10^{-3} | 4.3 x 10^{-4} | 2.0 x 10^{-4} | 1.2 x 10^{-4} | 8.0 x 10^{-5} |
|    | 1.25 | 0.72 | 0.19 | 1.9 x 10^{-3} | 2.5 x 10^{-4} | -5.6 x 10^{-5} | -1.7 x 10^{-4} | -2.2 x 10^{-4} |
|    | 1.5  | 0.71 | 0.25 | 2.7 x 10^{-3} | 1.6 x 10^{-4} | -3.0 x 10^{-4} | -4.6 x 10^{-4} | -5.4 x 10^{-4} |
As in Part 2 of Theorem 5 and Part 2 of Theorem 6, a sufficient condition for both estimators to be ordinary minimax is given by

\[
2 \left( \sum_{i=1}^{p} \frac{\sigma_i^4}{\sigma_i^4 - 2} \right) = 2 \left( \sum_{i=1}^{p} \frac{2^{2(i-1)}}{2^{2(i-1)} - 2} \right) \geq \frac{p}{\sigma_i^4 - 1} / \gamma_0.
\]

where the equality is attained by \( a \approx 1.066 \). Hence, the inequality (3.11) is satisfied by \( a = 1.01, 1.05 \) and is not by \( a = 1.25, 1.5 \).

Table 1 provides the relative ordinary risk differences given by

\[
1 - R(\delta_1, \theta) / \text{tr}\Sigma
\]

at

\[
\theta = m \left[ \text{tr}\Sigma \right]^{1/2} \frac{10}{\sqrt{10}} = m \left[ \sum_{i=1}^{10} \frac{10}{\sqrt{10}} \right]^{1/2} \frac{10}{\sqrt{10}}
\]

for \( m = 0, 2, 20, 40, 60, 80, 100 \). For both estimators, we see that, for larger \( m \) and \( a = 1.25, 1.5 \), the differences are negative, strongly suggesting that these two estimators are not ordinary minimax for \( a = 1.25, 1.5 \).

Table 2 provides the relative Bayes risk differences given by

\[
1 - R(\delta_2, r) / \text{tr}\Sigma
\]

for \( r = 1, 2, 5, 20, 40, 60, 80, 100 \). We see that for all \( a \), including \( a = 1.25, 1.5 \), the differences are all positive, which supports the ensemble minimaxity of the two estimators.

In summary, these tables support the theory presented in Theorems 5 and 6.

**ORCID**

Yuzo Maruyama [https://orcid.org/0000-0001-8686-1028](https://orcid.org/0000-0001-8686-1028)

**REFERENCES**

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

Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *The Annals of Statistics*, 4(1), 223–226.

Berger, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *The Annals of Statistics*, 8(4), 716–761.

Berger, J. O., & Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. *The Annals of Statistics*, 6(4), 783–801.

Berger, J. O., & Strawderman, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *The Annals of Statistics*, 24(3), 931–951. [https://doi.org/10.1214/aos/1032526950](https://doi.org/10.1214/aos/1032526950)

Berger, R. L. (1979). Gamma minimax robustness of Bayes rules. *Communications in Statistics A—Theory Methods*, 8(6), 543–560.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Annals of Mathematical Statistics*, 42, 855–903.
Brown, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *Journal of the American Statistical Association, 70*, 417–427.

Brown, L. D., Nie, H., & Xie, X. (2011). Ensemble minimax estimation for multivariate normal means. Retrieved from https://repository.upenn.edu/statistics_papers/44/ unpublished manuscript, (available at: https://repository.upenn.edu/statistics_papers/44/).

Casella, G. (1980). Minimax ridge regression estimation. *The Annals of Statistics, 8*(5), 1036–1056.

Casella, G. (1985). Condition numbers and minimax ridge regression estimators. *Journal of the American Statistical Association, 80*(391), 753–758.

Efron, B., & Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators. I. The Bayes case. *Journal of the American Statistical Association, 66*, 807–815.

Efron, B., & Morris, C. (1972a). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika, 59*, 335–347.

Efron, B., & Morris, C. (1972b). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *Journal of the American Statistical Association, 67*, 130–139.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association, 68*, 117–130.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B, 14*, 107–114.

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

Maruyama, Y., & Strawderman, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *The Annals of Statistics, 33*(4), 1753–1770.

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

Stein, C. (1974). Estimation of the mean of a multivariate normal distribution, *Proceedings of the Prague Symposium on Asymptotic Statistics (Charles Univ., Prague, 1973), Vol. II*. Prague: Charles Univ., pp. 345–381.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics, 9*(6), 1135–1151.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Annals of Mathematical Statistics, 42*(1), 385–388.

---

**How to cite this article:** Maruyama, Y., Brown, L. D., & George, E. I. (2022). Ensemble minimaxity of James-Stein estimators. *Stat, 11*(1), e532. [https://doi.org/10.1002/sta4.532](https://doi.org/10.1002/sta4.532)