Journal of Statistical Planning and Inference 211 (2021) 256–270

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Review

A Gaussian sequence approach for proving minimaxity: A Review

Yuzo Maruyama [a,*,1], William E. Strawderman [b,2]

[a] *Graduate School of Business Administration, Kobe University, Hyogo 657–8501, Japan*
[b] *Department of Statistics, Rutgers University, Piscataway, NJ 08854–8019, USA*

ARTICLE INFO

Article history:
Received 4 October 2018
Received in revised form 22 September 2019
Accepted 16 June 2020
Available online 23 June 2020

MSC:
62C20

Keywords:
Minimaxity
Least favorable prior
Invariance
Equivariance

ABSTRACT

This paper reviews minimax best equivariant estimation in three invariant estimation problems: a location parameter, a scale parameter and a (Wishart) covariance matrix. We briefly review development of the best equivariant estimator as a generalized Bayes estimator relative to right invariant Haar measure in each case. Then we prove minimaxity of the best equivariant procedure by giving a least favorable prior sequence based on non-truncated Gaussian distributions. The results in this paper are all known, but we bring a fresh and somewhat unified approach by using, in contrast to most proofs in the literature, a smooth sequence of non truncated priors. This approach leads to some simplifications in the minimaxity proofs.

Contents

* Corresponding author.
E-mail addresses: maruyama@port.kobe-u.ac.jp (Y. Maruyama), straw@stat.rutgers.edu (W.E. Strawderman).
1 This work was partially supported by JSPS KAKENHI 16K00040, 18H00835, 19K11852.
2 This work was partially supported by grants from the Simons Foundation (#418098 to William Strawderman).

https://doi.org/10.1016/j.jspi.2020.06.007

1. Introduction

We review some results on minimaxity of best equivariant estimators from what we hope is a fresh and somewhat unified perspective. Our basic approach is to start with a general equivariant estimator, and demonstrate that the best equivariant estimator is a generalized Bayes estimator, δ_0, with respect to an invariant prior. We then choose an appropriate sequence of Gaussian priors whose support is the entirety of the parameter space and show that the Bayes risks converge to the constant risk of δ_0. This implies that δ_0 is minimax. All results on best equivariance and minimaxity, which we consider in this paper, are known in the literature. But, using a sequence of Gaussian priors as a least favorable sequence, simplifies the proofs and gives fresh and unified perspective.

In this paper, we consider the following three estimation problems.

Estimation of a location parameter: Let the density function of $\boldsymbol{X}$ be given by

$$f(\boldsymbol{x}-\mu)=f(x_1-\mu,\dots,x_n-\mu). \tag{1}$$

Consider estimation of the location parameter μ under location invariant loss

$$L(\delta-\mu). \tag{2}$$

We study equivariant estimators under the location group, given by

$$\delta(\boldsymbol{x}-\mu)=\delta(\boldsymbol{x})-\mu. \tag{3}$$

Estimation of a scale parameter: Let the density function of $\boldsymbol{X}$ be given by

$$\sigma^{-n}f(\boldsymbol{x}/\sigma), \tag{4}$$

with scale parameter σ, where $\boldsymbol{x}/\sigma=(x_1/\sigma,\dots,x_n/\sigma)$. Consider estimation of the scale σ under scale invariant loss

$$L(\delta/\sigma). \tag{5}$$

We study equivariant estimators under scale group, given by

$$\delta(\boldsymbol{x}/\sigma)=\delta(\boldsymbol{x})/\sigma. \tag{6}$$

Estimation of covariance matrix: We study estimation of $\boldsymbol{\Sigma}$ based on a $p\times p$ random matrix $\boldsymbol{V}$ having a Wishart distribution $\mathcal{W}_p(n,\boldsymbol{\Sigma})$, where the density is given in (14). An estimator $\boldsymbol{\delta}$ is evaluated by the invariant loss including

$$\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}-\log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}|-p. \tag{7}$$

We consider equivariant estimators under the lower triangular group, given by

$$\boldsymbol{\delta}(\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^{\mathrm{T}})=\boldsymbol{A}\boldsymbol{\delta}(\boldsymbol{V})\boldsymbol{A}^{\mathrm{T}}, \tag{8}$$

where $\boldsymbol{A}\in\mathcal{T}^{+}$, the set of $p\times p$ lower triangular matrices with positive diagonal entries.

For the first two cases with squared error loss $(\delta-\mu)^2$ and the entropy loss $\delta/\sigma-\log(\delta/\sigma)-1$, respectively, the so called Pitman (1939) estimators

$$\hat{\mu}_0(\boldsymbol{x})=\frac{\int_{-\infty}^{\infty}\mu f(\boldsymbol{x}-\mu)\mathrm{d}\mu}{\int_{-\infty}^{\infty}f(\boldsymbol{x}-\mu)\mathrm{d}\mu}, \tag{9}$$

$$\hat{\sigma}_0(\boldsymbol{x})=\frac{\int_0^{\infty}\sigma^{-n-1}f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma}{\int_0^{\infty}\sigma^{-n-2}f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma} \tag{10}$$

are well-known to be best equivariant and minimax. Clearly, they are generalized Bayes with respect to $\pi(\mu)=1$ and $\pi(\sigma)=1/\sigma$, respectively. Girshick and Savage (1951) gave the original proof of minimaxity. Kubokawa (2004) also gives a proof and further developments in the restricted parameter setting. Both use a sequence of uniform distributions on expanding intervals as least favorable priors.

For the last case, James and Stein (1961) show that the best equivariant estimator is given by

$$\hat{\boldsymbol{\Sigma}}_0=\boldsymbol{T}\mathrm{diag}(d_1,\dots,d_p)\boldsymbol{T}^{\mathrm{T}} \tag{11}$$

where $\boldsymbol{T}\in\mathcal{T}^{+}$ is from the Cholesky decomposition of $\boldsymbol{V}=\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}$ and $d_i=1/(n+p-2i+1)$ for $i=1,\dots,p$. Note that the group of $p\times p$ lower triangular matrices with positive diagonal entries is solvable, and the result of Kiefer (1957) implies the minimaxity of $\hat{\boldsymbol{\Sigma}}_0$. Tsukuma and Kubokawa (2015) give as a sequence of least favorable priors, the invariant prior truncated on a sequence of expanding sets.

Typical minimaxity proofs in the literature use a sequence of truncated versions of the right invariant prior on nested sets increasing to the whole space. This often makes the proofs somewhat complicated. The motivation for this paper is to find a sequence of smooth priors for which the analysis is more straightforward and less complicated. In particular, in each case, the sequence of priors we employ is based on a Gaussian sequence of possibly transformed parameters.

Section 2 is devoted to developing the best equivariant estimator as a generalized Bayes estimator with respect to a right invariant (Haar measure) prior in each case. The general approach is basically that of Hora and Buehler (1966). Section 3 provides minimaxity proofs of the best equivariant procedure by giving a least favorable prior sequence based on (possibly transformed) Gaussian priors in each cases. We give some concluding remarks in Section 4. Some technical proofs are given in Appendix.

2. Establishing best equivariant procedures

All results in this section are well-known. Our proof of best equivariance for $\hat{\mu}_0$, $\hat{\sigma}_0$ and $\hat{\Sigma}_0$ follows from Hora and Buehler (1966). The reader is referred to Hora and Buehler (1966) for further details on their general development of a best equivariant estimator as the generalized Bayes estimator relative to right invariant Haar measure.

2.1. Estimation of location parameter

Consider a location equivariant estimator which satisfies $\delta(\boldsymbol{x}-\mu)=\delta(\boldsymbol{x})-\mu$. Then we have a following result.

Theorem 2.1. *Let $\boldsymbol{X}$ have distribution (1) and let the loss be given by (2). The generalized Bayes estimator with respect to the invariant prior $\pi(\mu)=1$, $\hat{\mu}_0(\boldsymbol{x})$, is best equivariant under the location group, that is,*

$$\hat{\mu}_0(\boldsymbol{x})=\arg\min_{\delta}\int_{-\infty}^{\infty}L(\delta(\boldsymbol{x})-\mu)f(\boldsymbol{x}-\mu)\mathrm{d}\mu.$$

Proof. The risk of the equivariant estimator (3) is written as

$$\begin{aligned}
R(\delta(\boldsymbol{x}),\mu)&=\int_{\mathbb{R}^n}L(\delta(\boldsymbol{x})-\mu)f(\boldsymbol{x}-\mu)\mathrm{d}\boldsymbol{x} \qquad (12)\\
&=\int_{\mathbb{R}^n}L(\delta(\boldsymbol{x}-\mu))f(\boldsymbol{x}-\mu)\mathrm{d}\boldsymbol{x}\quad (\text{by (3)})\\
&=\int_{\mathbb{R}^n}L(\delta(\boldsymbol{z}))f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\
&=\int_{\mathbb{R}^{n-1}}\int_{-\infty}^{\infty}L(\delta(\boldsymbol{z}_{n-1},z_n))f(\boldsymbol{z}_{n-1},z_n)\mathrm{d}z_n\mathrm{d}\boldsymbol{z}_{n-1}\\
&=\int_{\mathbb{R}^{n-1}}\int_{-\infty}^{\infty}L(\delta(\boldsymbol{z}_{n-1},u_n-\theta))f(\boldsymbol{z}_{n-1},u_n-\theta)\mathrm{d}\theta\mathrm{d}\boldsymbol{z}_{n-1}\\
&\quad\Uparrow\ z_n=u_n-\theta\ (u_n \text{ is an arbitrary constant and } \theta \text{ is variable})\\
&=\int_{\mathbb{R}^{n-1}}\int_{-\infty}^{\infty}L(\delta(\boldsymbol{u}_{n-1}-\theta,u_n-\theta))f(\boldsymbol{u}_{n-1}-\theta,u_n-\theta)\mathrm{d}\theta\mathrm{d}\boldsymbol{u}_{n-1}\\
&\quad\Uparrow\ \boldsymbol{z}_{n-1}=\boldsymbol{u}_{n-1}-\theta\ (\theta \text{ is a constant and } \boldsymbol{u}_{n-1} \text{ is variable})\\
&=\int_{\mathbb{R}^{n-1}}\left(\int_{-\infty}^{\infty}L(\delta(\boldsymbol{u})-\theta)f(\boldsymbol{u}-\theta)\mathrm{d}\theta\right)\mathrm{d}\boldsymbol{u}_{n-1}.\quad (\text{by (3)})
\end{aligned}$$

Then the best equivariant estimator is

$$\hat{\mu}_0(\boldsymbol{x})=\arg\min_{\delta}\int_{-\infty}^{\infty}L(\delta(\boldsymbol{x})-\mu)f(\boldsymbol{x}-\mu)\mathrm{d}\mu.\quad\square$$

Note: The last expression of $R(\delta(\boldsymbol{x}),\mu)$ in (12), is a function of u_n. However, u_n is arbitrary and hence the risk is constant, as demonstrated in the third expression of $R(\delta(\boldsymbol{x}),\mu)$ in (12).

2.2. Estimation of scale

Consider a scale equivariant estimator which satisfies $\delta(\boldsymbol{x}/\sigma)=\delta(\boldsymbol{x})/\sigma$. Then we have a following result.

Theorem 2.2. *Let $\boldsymbol{X}$ have distribution (4) and let the loss be given by (5). Then the generalized Bayes estimator, with respect to the prior $\pi(\sigma)=1/\sigma$, $\hat{\sigma}_0(\boldsymbol{x})$, is best equivariant under the scale group, that is,*

$$\hat{\sigma}_0(\boldsymbol{x})=\arg\min_{\delta}\int_0^{\infty}L(\delta/\sigma)\frac{f(\boldsymbol{x}/\sigma)}{\sigma^n}\frac{\mathrm{d}\sigma}{\sigma}.$$

Proof. The risk of the equivariant estimator is written as

$$
\begin{aligned}
R(\delta(\boldsymbol{x}),\sigma) &= \int_{\mathbb{R}^n} L(\delta(\boldsymbol{x})/\sigma)\sigma^{-n}f(\boldsymbol{x}/\sigma)\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} L(\delta(\boldsymbol{x}/\sigma))\sigma^{-n}f(\boldsymbol{x}/\sigma)\mathrm{d}\boldsymbol{x} \quad (\text{by } (6)) \\
&= \int_{\mathbb{R}^n} L(\delta(\boldsymbol{z}))f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \\
&= \int_{\mathbb{R}^{n-1}} \left(\int_{-\infty}^{0} + \int_{0}^{\infty}\right) L(\delta(\boldsymbol{z}_{n-1},z_n))f(\boldsymbol{z}_{n-1},z_n)\mathrm{d}z_n\mathrm{d}\boldsymbol{z}_{n-1} \\
&= \sum_{j=\{-1,1\}} \int_{\mathbb{R}^{n-1}} \int_{0}^{\infty} L(\delta(\boldsymbol{z}_{n-1},jz_n))f(\boldsymbol{z}_{n-1},jz_n)\mathrm{d}z_n\mathrm{d}\boldsymbol{z}_{n-1} \\
&= \sum_{j=\{-1,1\}} \int_{\mathbb{R}^{n-1}} \int_{0}^{\infty} L(\delta(\boldsymbol{z}_{n-1},ju_n/w))f(\boldsymbol{z}_{n-1},ju_n/w)\frac{u_n}{w^2}\mathrm{d}w\mathrm{d}\boldsymbol{z}_{n-1} \\
&\qquad \Uparrow\ z_n = u_n/w\ (u_n \text{ is an arbitrary positive constant and } w \text{ is variable}) \\
&= \sum_{j=\{-1,1\}} \int_{\mathbb{R}^{n-1}} \int_{0}^{\infty} L(\delta(\boldsymbol{u}_{n-1}/w,ju_n/w))\frac{1}{w^{n-1}}\frac{u_n}{w^2}f(\boldsymbol{u}_{n-1}/w,ju_n/w)\mathrm{d}\boldsymbol{u}_{n-1}\mathrm{d}w \\
&\qquad \Uparrow\ z_i = u_i/w\ (i=1,\dots,n-1) \quad (u_i \text{ is variable and } w \text{ is constant}) \\
&= \int_{\mathbb{R}^{n-1}} u_n \left\{ \sum_{j=\{-1,1\}} \int_{0}^{\infty} L(\delta(\boldsymbol{u}_{n-1},ju_n)/w)\frac{f(\boldsymbol{u}_{n-1}/w,ju_n/w)}{w^{n+1}}\mathrm{d}w \right\} \mathrm{d}\boldsymbol{u}_{n-1}.
\end{aligned} \tag{13}
$$

Then the best equivariant estimator is

$$\hat{\sigma}_0(\boldsymbol{x}) = \arg\min_{\delta} \int_0^\infty L(\delta/\sigma)\sigma^{-n-1}f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma. \quad \square$$

Note: The expression of $R(\delta(\boldsymbol{x}),\sigma)$ in the last line of (13), is a function of u_n. However, u_n is arbitrary and hence the risk is constant, as demonstrated in the third expression of (13).

2.3. Estimation of covariance matrix

Let $\boldsymbol{V}$ have a Wishart distribution $\mathcal{W}_p(n,\boldsymbol{\Sigma})$. Let $\mathcal{T}^+$ be the set of $p\times p$ lower triangular matrices with positive diagonal entries. By the Cholesky decomposition, $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{V}$ can be written as

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta} \text{ and } \boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}$$

for $\boldsymbol{\Theta}=(\theta_{ij})\in\mathcal{T}^+$ and $\boldsymbol{T}=(t_{ij})\in\mathcal{T}^+$. As in Theorem 7.2.1 of Anderson (2003), the probability density function of $\boldsymbol{T}$ is

$$f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) = \frac{1}{C(p,n)}|\boldsymbol{\Theta}\boldsymbol{T}|^n \exp\left[-\frac{1}{2}\mathrm{tr}\left\{(\boldsymbol{\Theta}\boldsymbol{T})(\boldsymbol{\Theta}\boldsymbol{T})^{\mathrm{T}}\right\}\right]\gamma(\mathrm{d}\boldsymbol{T}) \tag{14}$$

where $C(p,n)$ is a normalizing constant given by

$$C(p,n) = 2^{p(n-2)/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\{n+1-i\}/2) \tag{15}$$

and $\gamma(\mathrm{d}\boldsymbol{T})$ is the left-invariant Haar measure on $\mathcal{T}^+$ given by

$$\gamma(\mathrm{d}\boldsymbol{T}) = \prod_{i=1}^{p} t_{ii}^{-i}\mathrm{d}\boldsymbol{T}. \tag{16}$$

An estimator $\boldsymbol{\delta}$ is evaluated by the invariant loss function given by

$$L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}}). \tag{17}$$

Denote the risk function by

$$R(\boldsymbol{\delta},\boldsymbol{\Sigma}) = \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}})f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}).$$

For all $\boldsymbol{A} \in \mathcal{T}^+$, the group transformation with respect to $\mathcal{T}^+$ on a random matrix $\boldsymbol{T}$ and a parameter matrix $\boldsymbol{\Theta}$ is defined by $(\boldsymbol{T}, \boldsymbol{\Theta}) \to (\boldsymbol{AT}, \boldsymbol{\Theta A}^{-1})$. The group $\bar{G}$ operating on $\boldsymbol{\Theta}$ is transitive. Any equivariant estimator of

$$\boldsymbol{\Sigma} = (\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1} = \boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta}^{-1})^{\mathrm{T}}$$

under the lower triangular group is of form given by

$$\delta(\boldsymbol{AT}) = \boldsymbol{A}\delta(\boldsymbol{T})\boldsymbol{A}^{\mathrm{T}}. \tag{18}$$

Then we have a following result.

Theorem 2.3. *Let $\boldsymbol{V} = \boldsymbol{TT}^{\mathrm{T}} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ and let the loss be $L(\boldsymbol{\Theta}\delta\boldsymbol{\Theta}^{\mathrm{T}})$ as in (17). Then the generalized Bayes estimator with respect to the prior*

$$\pi(\boldsymbol{\Theta}) = \gamma(\mathrm{d}\boldsymbol{\Theta}), \tag{19}$$

δ_0, is best equivariant under lower triangular group, that is,

$$\delta_0(\boldsymbol{T}) = \arg\min_{\delta} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\delta\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}). \tag{20}$$

Note that $\gamma(\mathrm{d}\boldsymbol{\Theta})$ is the "left" invariant measure, which seems to contradict the general theory by Hora and Buehler (1966). However this seeming anomaly is due to our parameterization $\boldsymbol{V} = \boldsymbol{TT}^{\mathrm{T}}$, $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}$ and

$$\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta}^{-1})^{\mathrm{T}}. \tag{21}$$

The general theory implies that

$$\nu(\mathrm{d}\boldsymbol{\Theta}^{-1}) = \gamma(\mathrm{d}\boldsymbol{\Theta}) \tag{22}$$

where ν is right invariant Haar measure on $\mathcal{T}^+$ given by

$$\nu(\mathrm{d}\boldsymbol{Z}) = \prod_{i=1}^{p} z_{ii}^{-(p-i+1)}\mathrm{d}\boldsymbol{Z}. \tag{23}$$

In the proof below, in addition to the left invariance of γ, and the right invariance of ν, we use the fact that

$$f_W(\boldsymbol{T}|\boldsymbol{\Theta}) = f_W(\boldsymbol{\Theta T}|\boldsymbol{I}) = f_W(\boldsymbol{I}|\boldsymbol{\Theta T}). \tag{24}$$

Proof of Theorem 2.3. By (14) and (17), the risk of an equivariant estimator can be expressed as

$$\begin{aligned}
R(\delta, \boldsymbol{\Sigma}) &= \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\delta(\boldsymbol{T})\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) \\
&= \int_{\mathcal{T}^+} L(\delta(\boldsymbol{\Theta T})) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{T}) \quad (\text{by } (18)) \\
&= \int_{\mathcal{T}^+} L(\delta(\boldsymbol{Z})) f_W(\boldsymbol{Z}|\boldsymbol{I})\gamma(\mathrm{d}\boldsymbol{Z}) \quad (\boldsymbol{Z} = \boldsymbol{\Theta T},\ \text{and left invariance of } \gamma) \\
&= \int_{\mathcal{T}^+} L(\delta(\boldsymbol{Z})) f_W(\boldsymbol{I}|\boldsymbol{Z}) \prod_{i=1}^{p} z_{ii}^{-i}\mathrm{d}\boldsymbol{Z} \quad (\text{by the form of } f_W) \\
&= \int_{\mathcal{T}^+} L(\delta(\boldsymbol{Z})) f_W(\boldsymbol{I}|\boldsymbol{Z}) \prod_{i=1}^{p} z_{ii}^{p-2i+1}\nu(\mathrm{d}\boldsymbol{Z}) \quad (\text{by } (23)) \\
&= \int_{\mathcal{T}^+} L(\delta(\boldsymbol{WS})) f_W(\boldsymbol{S}|\boldsymbol{W}) \prod_{i=1}^{p} (w_{ii}s_{ii})^{p-2i+1}\nu(\mathrm{d}\boldsymbol{W}) \\
&\qquad \Uparrow \boldsymbol{Z} = \boldsymbol{WS}\ (\boldsymbol{W} \in \mathcal{T}^+ \text{ is variable and } \boldsymbol{S} \in \mathcal{T}^+ \text{ is arbitrary})\ \text{ and right invariance of } \nu \\
&= \prod_{i=1}^{p} s_{ii}^{p-2i+1} \int_{\mathcal{T}^+} L(\boldsymbol{W}\delta(\boldsymbol{S})\boldsymbol{W}^{\mathrm{T}}) f_W(\boldsymbol{S}|\boldsymbol{W})\gamma(\mathrm{d}\boldsymbol{W}), \quad \text{by (18) and the form of } \gamma(\mathrm{d}\boldsymbol{W}).
\end{aligned} \tag{25}$$

Then the best equivariant estimator with respect to the group $\mathcal{T}^+$ can be written as

$$\delta_0(\boldsymbol{T}) = \arg\min_{\delta} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\delta\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}). \quad \square$$

Note: The expression of $R(\delta, \boldsymbol{\Sigma})$ in the last line of (25), is a function of $\boldsymbol{S}$. However, $\boldsymbol{S}$ is arbitrary and hence the risk is constant, as demonstrated in the third expression of (25).

3. Minimaxity

In this section, we choose an appropriate sequence of priors whose support is the entirety of the parameter space and show that the Bayes risks converge to the constant risk of the best equivariant estimator δ_0. By a well-known standard result (see e.g. Lehmann and Casella (1998)), this implies minimaxity of δ_0. In order to deal with explicit expressions for minimax estimators as well as for somewhat technical reasons, in this section, we specify the loss functions to be standard choices in the literature. For the location and scale problem, the squared error loss and the entropy loss

$$L(\delta-\mu)=(\delta-\mu)^2, \quad L(\delta/\sigma)=\delta/\sigma-\log(\delta/\sigma)-1$$

are used respectively. For estimation of covariance matrix, the so called Stein (1956) loss function given by

$$L(\Theta\boldsymbol{\delta}\Theta^{\mathrm{T}})=\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}-\log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}|-p=\mathrm{tr}(\Theta\boldsymbol{\delta}\Theta^{\mathrm{T}})-\log|\Theta\boldsymbol{\delta}\Theta^{\mathrm{T}}|-p \tag{26}$$

is used.

3.1. Estimation of location

In this section, we show the minimaxity of $\hat{\mu}^0$, the best location equivariant estimator under squared error loss. A point of departure from most proofs in the literature is that a smooth sequence of Gaussian densities simplifies the proof. It is also easily applied in the multivariate location family (see Remark 3.1).

Recall that the Bayes estimator corresponding to a (generalized) prior $\pi(\mu)$, under squared error loss, is given by

$$\delta_\pi(\boldsymbol{x})=\arg\min_\delta\int_{-\infty}^{\infty}L(\delta-\mu)f(\boldsymbol{x}-\mu)\pi(\mu)\mathrm{d}\mu \tag{27}$$

$$=\frac{\int\mu f(\boldsymbol{x}-\mu)\pi(\mu)\mathrm{d}\mu}{\int f(\boldsymbol{x}-\mu)\pi(\mu)\mathrm{d}\mu}\quad\text{under } L(t)=t^2. \tag{28}$$

Hence, by Theorem 2.1, the best equivariant estimator is given by

$$\hat{\mu}_0(\boldsymbol{x})=\frac{\int\mu f(\boldsymbol{x}-\mu)\mathrm{d}\mu}{\int f(\boldsymbol{x}-\mu)\mathrm{d}\mu}. \tag{29}$$

Theorem 3.1. *Let $\boldsymbol{X}$ have distribution (1) and let the loss be given by $L(\delta-\mu)=(\delta-\mu)^2$. Assume that there exists an equivariant estimator with finite risk. Then the best equivariant estimator, $\hat{\mu}_0(\boldsymbol{x})$, given by (29), is minimax, and the minimax constant risk is given by*

$$R_0=\int L(\hat{\mu}_0(\boldsymbol{x}))f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\int\left\{\hat{\mu}_0(\boldsymbol{x})\right\}^2f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

Under squared error loss, the Bayes estimator is explicitly written as (28). However, in the following proof, the implicit expression (27) is mainly used to indicate possible extension for more general loss functions. For the same reason, $L(\delta(\boldsymbol{x})-\mu)$ instead of $(\delta(\boldsymbol{x})-\mu)^2$ is used.

Proof of Theorem 3.1. Let

$$\phi(\mu)=\frac{1}{\sqrt{2\pi}}\exp(-\mu^2/2)\text{ and }\phi_k(\mu)=\frac{1}{k}\phi(\mu/k)\text{ for }k>0.$$

The Bayes risk of $\delta(\boldsymbol{x})$ under the prior $\phi_k(\mu)$ is finite and given by

$$r_k(\phi_k,\delta(\boldsymbol{x}))=\iint L(\delta_k^\phi(\boldsymbol{x})-\mu)f(\boldsymbol{x}-\mu)\phi_k(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{x}.$$

Also the corresponding Bayes estimator is given by

$$\delta_k^\phi(\boldsymbol{x})=\arg\min_\delta\int_{-\infty}^{\infty}L(\delta-\mu)f(\boldsymbol{x}-\mu)\phi_k(\mu)\mathrm{d}\mu.$$

Clearly

$$r_k(\phi_k,\delta_k^\phi)\le r_k(\phi_k,\hat{\mu}_0)=R_0<\infty,$$

and therefore, to show $\lim_{k\to\infty}r_k(\phi_k,\delta_k^\phi)=R_0$, it suffices to prove

$$\liminf_{k\to\infty}r_k(\phi_k,\delta_k^\phi)\ge R_0.$$

Making the transformation $\boldsymbol{z}=\boldsymbol{x}-\mu$ and $\nu=\mu/k$ yields

$$\begin{aligned} r_k(\phi_k,\delta_k^{\phi}) &= \iint L(\delta_k^{\phi}(\boldsymbol{x})-\mu)f(\boldsymbol{x}-\mu)\phi_k(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{x} \\ &= \iint L(\delta_k^{\phi}(\boldsymbol{z}+\mu)-\mu)f(\boldsymbol{z})\phi_k(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{z} \quad (\boldsymbol{z}=\boldsymbol{x}-\mu) \\ &= \iint L(\delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu)f(\boldsymbol{z})\phi(\nu)\mathrm{d}\nu\mathrm{d}\boldsymbol{z} \quad (\nu=\mu/k), \end{aligned} \tag{30}$$

where $k\phi_k(k\nu)=\phi(\nu)$. Under squared error loss ($L(t)=t^2$), we have

$$\begin{aligned} \delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu &= \frac{\int_{-\infty}^{\infty}\mu f(\boldsymbol{z}+k\nu-\mu)\phi_k(\mu)\mathrm{d}\mu}{\int_{-\infty}^{\infty} f(\boldsymbol{z}+k\nu-\mu)\phi_k(\mu)\mathrm{d}\mu}-k\nu \\ &= \frac{\int_{-\infty}^{\infty}\theta f(\boldsymbol{z}-\theta)\phi_k(\theta+k\nu)\mathrm{d}\theta}{\int_{-\infty}^{\infty}\theta f(\boldsymbol{z}-\theta)\phi_k(\theta+k\nu)\mathrm{d}\theta} \quad (\theta=\mu-k\nu) \\ &= \frac{\int_{-\infty}^{\infty}\theta f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)\mathrm{d}\theta}{\int_{-\infty}^{\infty}\theta f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)\mathrm{d}\theta} \end{aligned}$$

where the third equality is from $k\phi_k(\theta+k\nu)=\phi(\theta/k+\nu)$. Note, for any fixed $\boldsymbol{z}$, $f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)$ and $|\theta|f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)$ are bounded from above by $\phi(0)f(\boldsymbol{z}-\theta)$ and $\phi(0)|\theta|f(\boldsymbol{z}-\theta)$. Under the assumption of existence of an equivariant estimator, they are both integrable by Lemmas A.2 and A.3 respectively. Since $\lim_{k\to\infty}\phi(\theta/k+\nu)=\phi(\nu)$ for any ν and θ, the dominated convergence theorem implies

$$\delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu=\frac{\lim_{k\to\infty}\int_{-\infty}^{\infty}\theta f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)\mathrm{d}\theta}{\lim_{k\to\infty}\int_{-\infty}^{\infty}f(\boldsymbol{z}-\theta)\phi(\theta/k+\nu)\mathrm{d}\theta}=\hat{\mu}_0(\boldsymbol{z}) \tag{31}$$

and hence

$$\lim_{k\to\infty} L(\delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu)=\{\hat{\mu}_0(\boldsymbol{z})\}^2=L(\hat{\mu}_0(\boldsymbol{z})). \tag{32}$$

Hence by (30) and Fatou's lemma, we obtain that

$$\begin{aligned} \liminf_{k\to\infty} r_k(\phi_k,\delta_k^{\phi}) &= \liminf_{k\to\infty}\iint L(\delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu)f(\boldsymbol{z})\phi(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{z} \\ &\geq \iint \liminf_{k\to\infty} L(\delta_k^{\phi}(\boldsymbol{z}+k\nu)-k\nu)f(\boldsymbol{z})\phi(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{z} \\ &= \iint L(\hat{\mu}_0(\boldsymbol{z}))f(\boldsymbol{z})\phi(\mu)\mathrm{d}\mu\mathrm{d}\boldsymbol{z} \\ &= R_0. \quad \square \end{aligned} \tag{33}$$

Remark 3.1. In the multivariate case, suppose $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_p\in\mathbb{R}^n$ and

$$\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_p\}\sim f(\boldsymbol{x}_1-\mu_1,\ldots,\boldsymbol{x}_p-\mu_p).$$

Let $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^{\mathrm{T}}$. Then the Pitman estimator of $\boldsymbol{\mu}$, the generalized Bayes estimator with respect to $\pi(\boldsymbol{\mu})=1$, is

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_p)=\frac{\int_{\mathbb{R}^p}\boldsymbol{\mu}f(\boldsymbol{x}_1-\mu_1,\ldots,\boldsymbol{x}_p-\mu_p)\mathrm{d}\boldsymbol{\mu}}{\int_{\mathbb{R}^p}f(\boldsymbol{x}_1-\mu_1,\ldots,\boldsymbol{x}_p-\mu_p)\mathrm{d}\boldsymbol{\mu}}. \tag{34}$$

Using

$$\pi_k(\boldsymbol{\mu})=\prod_{i=1}^{p}\phi_k(\mu_i)=\frac{1}{(2\pi k^2)^{p/2}}\exp\left(-\frac{\|\boldsymbol{\mu}\|^2}{2k^2}\right)$$

as the least favorable sequence of priors gives minimaxity under the quadratic loss $\|\boldsymbol{\delta}-\boldsymbol{\mu}\|^2$ of (34).

3.2. Estimation of scale

In this section, we show the minimaxity of the scale Pitman estimator under entropy loss given by

$$L(\delta/\sigma)=\delta/\sigma-\log(\delta/\sigma)-1. \tag{35}$$

Recall that the Bayes estimator corresponding to a (generalized) prior $\pi(\sigma)$, under entropy loss (35), is given by

$$\delta_\pi(\boldsymbol{x}) = \arg\min_\delta \int_{-\infty}^{\infty} L(\delta/\sigma)\sigma^{-n} f(\boldsymbol{x}/\sigma)\pi(\sigma)\mathrm{d}\sigma = \frac{\int \sigma^{-n} f(\boldsymbol{x}/\sigma)\pi(\sigma)\mathrm{d}\sigma}{\int \sigma^{-n-1} f(\boldsymbol{x}/\sigma)\pi(\sigma)\mathrm{d}\sigma}.$$

Hence the generalized Bayes estimator under $\pi(\sigma) = 1/\sigma$, which is best equivariant as shown in Theorem 2.2, is given by

$$\hat{\sigma}_0(\boldsymbol{x}) = \frac{\int \sigma^{-n-1} f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma}{\int \sigma^{-n-2} f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma}. \tag{36}$$

We have a following minimaxity result.

Theorem 3.2. *Let* $\boldsymbol{X}$ *have distribution* (4) *and let the loss be given by* $L(\delta/\sigma) = \delta/\sigma - \log(\delta/\sigma) - 1$. *Assume that there exists an equivariant estimator with finite risk. Then the best equivariant estimator,* $\hat{\sigma}_0(\boldsymbol{x})$*, given by* (36)*, is minimax, and the minimax constant risk is given by*

$$R_0 = \int_{\mathbb{R}^n} L(\hat{\sigma}_0(\boldsymbol{x}))f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} \left\{\hat{\sigma}_0(\boldsymbol{x}) - \log \hat{\sigma}_0(\boldsymbol{x}) - 1\right\} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

Proof. Assume $\log\sigma \sim N(0, k^2)$ or equivalently

$$\pi_k(\sigma) = \frac{1}{k}\phi(\log\sigma/k)\frac{1}{\sigma}, \quad \text{for } k > 0,$$

where $\phi(\cdot)$ is the pdf of $N(0, 1)$. Then the Bayes estimator satisfies

$$\delta_k^\pi = \delta_k^\pi(\boldsymbol{x}) = \arg\min_\delta \int_0^\infty L(\delta/\sigma)\sigma^{-n} f(\boldsymbol{x}/\sigma)\pi_k(\sigma)\mathrm{d}\sigma$$

and the Bayes risk is given by

$$r_k(\pi_k, \delta_k^\pi) = \iint L(\delta_k^\pi(\boldsymbol{x})/\sigma)\sigma^{-n} f(\boldsymbol{x}/\sigma)\pi_k(\sigma)\mathrm{d}\sigma\,\mathrm{d}\boldsymbol{x}.$$

Clearly

$$r_k(\pi_k, \delta_k^\pi) \le r_k(\pi_k, \hat{\sigma}_0(\boldsymbol{x})) = R_0,$$

and therefore, to show $\lim_{k\to\infty} r_k(\phi_k, \delta_k^\phi) = R_0$, it suffices to prove

$$\liminf_{k\to\infty} r_k(\pi_k, \delta_k^\pi) \ge R_0.$$

Making the transformation $\boldsymbol{l} = \boldsymbol{x}/\sigma$ and $\eta^k = \sigma$ yields

$$\begin{aligned} r_k(\pi_k, \delta_k^\pi) &= \iint L(\delta_k^\pi(\boldsymbol{x})/\sigma)\sigma^{-n} f(\boldsymbol{x}/\sigma)\pi_k(\sigma)\mathrm{d}\sigma\,\mathrm{d}\boldsymbol{x} \\ &= \iint L(\delta_k^\pi(\sigma\boldsymbol{l})/\sigma)f(\boldsymbol{l})\pi_k(\sigma)\mathrm{d}\sigma\,\mathrm{d}\boldsymbol{l} \quad (\boldsymbol{l} = \boldsymbol{x}/\sigma) \\ &= \iint L(\delta_k^\pi(\eta^k\boldsymbol{l})/\eta^k)f(\boldsymbol{l})\pi_1(\eta)\mathrm{d}\eta\,\mathrm{d}\boldsymbol{l} \quad (\eta^k = \sigma) \end{aligned} \tag{37}$$

where the relation $k\pi_k(\eta^k) = \pi_1(\eta)$ is used in the third equality. Under the entropy loss, we have

$$\frac{\delta_k^\pi(\eta^k\boldsymbol{l})}{\eta^k} = \frac{1}{\eta^k}\frac{\int \sigma^{-n} f(\eta^k\boldsymbol{l}/\sigma)\pi_k(\sigma)\mathrm{d}\sigma}{\int \sigma^{-n-1} f(\eta^k\boldsymbol{l}/\sigma)\pi_k(\sigma)\mathrm{d}\sigma} = \frac{\int y^{-n} f(\boldsymbol{l}/y)\pi_k(\eta^k y)\mathrm{d}y}{\int y^{-n-1} f(\boldsymbol{l}/y)\pi_k(\eta^k y)\mathrm{d}y} \quad (y = \sigma/\eta^k).$$

Note

$$k\pi_k(\eta^k y) = \frac{1}{\eta^k y}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\log\eta^k + \log y)^2}{2k^2}\right)$$

and $\delta_k^\pi(\eta^k\boldsymbol{l})/\eta^k$ can be expressed as

$$\frac{\delta_k^\pi(\eta^k\boldsymbol{l})}{\eta^k} = \frac{\int y^{-n-1} f(\boldsymbol{l}/y)\{y\eta^k k\pi_k(\eta^k y)\}\mathrm{d}y}{\int y^{-n-2} f(\boldsymbol{l}/y)\{y\eta^k k\pi_k(\eta^k y)\}\mathrm{d}y}. \tag{38}$$

For any fixed $\boldsymbol{l}$, $y^{-n-1}f(\boldsymbol{l}/y)\{y\eta^k k\pi_k(\eta^k y)\}$ and $y^{-n-2}f(\boldsymbol{l}/y)\{y\eta^k k\pi_k(\eta^k y)\}$ are bounded from above by $\phi(0)y^{-n-1}f(\boldsymbol{l}/y)$ and $\phi(0)y^{-n-2}f(\boldsymbol{l}/y)$, respectively. Under the assumption of existence of an equivariant estimator, they are both integrable by Lemmas A.4 and A.5 respectively. Since

$$\lim_{k\to\infty} yk\eta^k\pi_k(\eta^k y) = \phi(\log\eta)$$

for any η and y, the dominated convergence theorem implies

$$\lim_{k\to\infty} \delta_k^{\pi}(\eta^k \boldsymbol{l})/\eta^k = \hat{\sigma}_0(\boldsymbol{l}). \tag{39}$$

Also the continuity of $L(\cdot)$ implies

$$\lim_{k\to\infty} L(\delta_k^{\pi}(\eta^k \boldsymbol{l})/\eta^k) = L(\hat{\sigma}_0(\boldsymbol{l})). \tag{40}$$

Hence by (37) and Fatou's lemma, we obtain that

$$\begin{aligned}
\liminf_{k\to\infty} r_k(\pi_k, \delta_k^{\pi}) &= \liminf_{k\to\infty} \iint L(\delta_k^{\pi}(\eta^k \boldsymbol{l})/\eta^k) f(\boldsymbol{l})\pi_1(\eta)\mathrm{d}\eta\mathrm{d}\boldsymbol{l} \\
&\geq \iint \liminf_{k\to\infty} L(\delta_k^{\pi}(\eta^k \boldsymbol{l})/\eta^k) f(\boldsymbol{l})\pi_1(\eta)\mathrm{d}\eta\mathrm{d}\boldsymbol{l} \\
&= \iint L(\hat{\sigma}_0(\boldsymbol{l})) f(\boldsymbol{l})\pi_1(\eta)\mathrm{d}\eta\mathrm{d}\boldsymbol{l} \\
&= R_0. \quad \square
\end{aligned} \tag{41}$$

For estimating σ^c ($c \in \mathbb{R}$), a similar argument shows that

$$\hat{\sigma}_{0c}(\boldsymbol{x}) = \frac{\int \sigma^{-n-1} f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma}{\int \sigma^{-n-1-c} f(\boldsymbol{x}/\sigma)\mathrm{d}\sigma}$$

is minimax and best equivariant for estimating σ^c under entropy loss

$$L(\delta/\sigma^c) = \delta/\sigma^c - \log(\delta/\sigma^c) - 1,$$

provided an equivariant estimator with finite risk exists.

3.3. Estimation of covariance matrix

As we mentioned in the beginning of this section, we use the so called Stein (1956) loss function given by

$$L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}}) = \mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} - \log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}| - p = \mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}}) - \log|\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}}| - p. \tag{42}$$

James and Stein (1961), in their Section 5, show that the best equivariant estimator is given by

$$\hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{T}\mathrm{diag}(d_1, \ldots, d_p)\boldsymbol{T}^{\mathrm{T}} \tag{43}$$

where $\boldsymbol{T} \in \mathcal{T}^+$ is from the Cholesky decomposition of $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}$ and $d_i = 1/(n + p - 2i + 1)$ for $i = 1, \ldots, p$. As demonstrated in the literature, by e.g. Tsukuma and Kubokawa (2015), the best equivariant estimator under the loss (42) may also be shown to be $\hat{\boldsymbol{\Sigma}}_0$ by using the generalized Bayes representation given in Theorem 2.3 since

$$\begin{aligned}
&\arg\min_{\boldsymbol{\delta}} \int_{\mathcal{T}^+} L(\boldsymbol{\Theta}\boldsymbol{\delta}\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) \\
&= \arg\min_{\boldsymbol{\delta}} \int_{\mathcal{T}^+} \left\{\mathrm{tr}\left(\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{\delta}\right) - \log|\boldsymbol{\delta}|\right\} f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) \\
&= \left(\int_{\mathcal{T}^+} \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta} f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta})\right)^{-1} \int_{\mathcal{T}^+} f_W(\boldsymbol{T}|\boldsymbol{\Theta})\gamma(\mathrm{d}\boldsymbol{\Theta}) \\
&= \left(\int_{\mathcal{T}^+} \boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta} f_W(\boldsymbol{T}|\boldsymbol{\Theta}) \prod \theta_{ii}^{p-2i+1} \nu(\mathrm{d}\boldsymbol{\Theta})\right)^{-1} \int_{\mathcal{T}^+} f_W(\boldsymbol{T}|\boldsymbol{\Theta}) \prod \theta_{ii}^{p-2i+1} \nu(\mathrm{d}\boldsymbol{\Theta}) \\
&= \left(\int_{\mathcal{T}^+} (\boldsymbol{Z}\boldsymbol{T}^{-1})^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{T}^{-1} f_W(\boldsymbol{Z}|\boldsymbol{I}) \prod (t_{ii}^{-1} z_{ii}^{p-2i+1}) \nu(\mathrm{d}\boldsymbol{Z})\right)^{-1} \int_{\mathcal{T}^+} f_W(\boldsymbol{Z}|\boldsymbol{I}) \prod (t_{ii}^{-1} z_{ii}^{p-2i+1}) \nu(\mathrm{d}\boldsymbol{Z}) \\
&= \boldsymbol{T}\left\{\left(\int_{\mathcal{T}^+} \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} f_W(\boldsymbol{Z}|\boldsymbol{I})\gamma(\mathrm{d}\boldsymbol{Z})\right)^{-1} \int_{\mathcal{T}^+} f_W(\boldsymbol{Z}|\boldsymbol{I})\gamma(\mathrm{d}\boldsymbol{Z})\right\}\boldsymbol{T}^{\mathrm{T}} \\
&= \hat{\boldsymbol{\Sigma}}_0
\end{aligned}$$

where the second equality is from Lemma A.1 and

$$\left(\int_{\mathcal{T}^+} \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} f_W(\boldsymbol{Z}|\boldsymbol{I})\gamma(\mathrm{d}\boldsymbol{Z})\right)^{-1} \int_{\mathcal{T}^+} f_W(\boldsymbol{Z}|\boldsymbol{I})\gamma(\mathrm{d}\boldsymbol{Z}) = \mathrm{diag}(d_1, \ldots, d_p) \text{ with } d_i = 1/(n + p - 2i + 1).$$

Note that the group of $p \times p$ lower triangular matrices with positive diagonal entries is solvable, and the result of Kiefer (1957) implies the minimaxity of $\hat{\boldsymbol{\Sigma}}_0$. Tsukuma and Kubokawa (2015) give as a sequence of least favorable priors, a sequence of invariant priors truncated on an expanding set.

In this section, we choose an appropriate sequence of Gaussian priors whose support is the entirety of the parameter space and show that the Bayes risks converge to the constant risk of $\hat{\boldsymbol{\Sigma}}_0$. This implies that $\hat{\boldsymbol{\Sigma}}_0$ is minimax.

As a new parameterization on $\boldsymbol{\Theta}$, let

$$\begin{aligned} \xi_{ii} &= \log \theta_{ii} \text{ for } i = 1, \ldots, p, \\ \xi_{ij} &= \frac{\theta_{ij}}{\theta_{ii}}, \text{ for } 1 \le j < i \le p \end{aligned} \tag{44}$$

and let

$$\boldsymbol{\xi} = (\xi_{11}, \xi_{21}, \xi_{22}, \ldots, \xi_{p1}, \ldots, \xi_{pp})^{\mathrm{T}} \in \mathbb{R}^{p(p+1)/2}. \tag{45}$$

The prior on ξ_{ij} is

$$\xi_{ij} \sim N(0, k_{ij}^2) \text{ for } 1 \le j \le i \le p,$$

and $\xi_{11}, \xi_{21}, \xi_{22}, \ldots, \xi_{p1}, \ldots, \xi_{pp}$ are assumed mutually independent. Equivalently the density is

$$\bar{\pi}_k(\boldsymbol{\xi}) = \prod_{j \le i} k_{ij}^{-1} \phi(\xi_{ij}/k_{ij}). \tag{46}$$

Set

$$k_{ii} = k, \quad k_{ij} = k^{(i-j)k} \text{ for } i > j \tag{47}$$

with $k > 0$, although, in the following, we keep the notation k_{ii} and k_{ij}.

By (44) and (46), we have

$$\pi_k(\boldsymbol{\Theta}) = \prod_{i=1}^{p} \left\{ \frac{1}{k_{ii}} \phi(\log \theta_{ii}/k_{ii}) \frac{1}{\theta_{ii}} \right\} \prod_{j<i} \left\{ \frac{1}{k_{ij}\theta_{ii}} \phi(\theta_{ij}/\{\theta_{ii}k_{ij}\}) \right\}. \tag{48}$$

The prior distributions yield the Bayes estimators

$$\boldsymbol{\delta}_k^{\pi} = \boldsymbol{\delta}_k^{\pi}(\boldsymbol{T}) = \arg\min_{\boldsymbol{\delta}} \int_{\boldsymbol{Z} \in \mathcal{T}^+} L(\boldsymbol{Z}\boldsymbol{\delta}\boldsymbol{Z}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{Z}) \pi_k(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z}$$

with Bayes risks

$$r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) = \iint L(\boldsymbol{\Theta}\boldsymbol{\delta}_k^{\pi}(\boldsymbol{T})\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{T}|\boldsymbol{\Theta}) \gamma(\mathrm{d}\boldsymbol{T}) \pi_k(\boldsymbol{\Theta}) \mathrm{d}\boldsymbol{\Theta}. \tag{49}$$

Theorem 3.3. *Let $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}$ have distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ and let the loss be given by (42). Then the best equivariant estimator, $\hat{\boldsymbol{\Sigma}}_0(\boldsymbol{T})$, given by (43), is minimax, and the minimax constant risk is given by*

$$R_0 = \int_{\boldsymbol{T} \in \mathcal{T}^+} L(\hat{\Sigma}_0(\boldsymbol{T})) f_W(\boldsymbol{T}|\boldsymbol{I}) \gamma(\mathrm{d}\boldsymbol{T}).$$

Proof. We show this theorem along the same lines as in Kubokawa (2004) and Tsukuma and Kubokawa (2015) who modified the method of Girshick and Savage (1951).

Clearly

$$r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) \le r_k(\pi_k, \hat{\boldsymbol{\Sigma}}_0) = R_0,$$

and therefore, to show $\lim_{k\to\infty} r_k(\phi_k, \delta_k^{\phi}) = R_0$, it suffices to prove

$$\liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) \ge R_0.$$

In (49), making the transformation $\boldsymbol{L} = \boldsymbol{\Theta}\boldsymbol{T}$ yields

$$\begin{aligned} r_k(\pi_k, \boldsymbol{\delta}_k^{\pi}) &= \iint L(\boldsymbol{\Theta}\boldsymbol{\delta}_k^{\pi}(\boldsymbol{\Theta}^{-1}\boldsymbol{L})\boldsymbol{\Theta}^{\mathrm{T}}) f_W(\boldsymbol{L}|\boldsymbol{I}_p) \gamma(\mathrm{d}\boldsymbol{L}) \pi_k(\boldsymbol{\Theta}) \mathrm{d}\boldsymbol{\Theta} \\ &= \iint L(\boldsymbol{\Theta}(\boldsymbol{\xi})\boldsymbol{\delta}_k^{\pi}(\boldsymbol{\Theta}(\boldsymbol{\xi})^{-1}\boldsymbol{L})\boldsymbol{\Theta}(\boldsymbol{\xi})^{\mathrm{T}}) f_W(\boldsymbol{L}|\boldsymbol{I}_p) \gamma(\mathrm{d}\boldsymbol{L}) \bar{\pi}_k(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}, \end{aligned} \tag{50}$$

where $\bar{\pi}_k(\boldsymbol{\xi})$ is given by (46) and $\boldsymbol{\Theta}(\boldsymbol{\xi})$ is defined through $\xi_{ii} = \log \theta_{ii}$ for $i = 1, \ldots, p$ and $\xi_{ij} = \theta_{ij}/\theta_{ii}$ for $1 \le j < i \le p$. Further, by change of variables,

$$\xi_{ij} = k_{ij}\omega_{ij} \text{ for } 1 \le j \le i \le p \tag{51}$$

and for notational convenience,

$$\boldsymbol{k}\bullet\boldsymbol{\omega} = (k_{11}\omega_{11}, k_{21}\omega_{21}, k_{22}\omega_{22}, \dots, k_{pp}\omega_{pp}),$$

we have

$$r_k(\pi_k, \boldsymbol{\delta}_k^\pi) = \iint L(\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\bar{\pi}_1(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}. \tag{52}$$

where

$$\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega}) = \boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})^{-1}\boldsymbol{L})\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})^{\mathrm{T}}. \tag{53}$$

Note, by (44) and (51), the (i, i) diagonal component and the non-diagonal (i, j) component, of $\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})$, are

$$\begin{aligned} \boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})_{ii} &= \exp(k_{ii}\omega_{ii}) \text{ for } i = 1, \dots, p, \\ \boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})_{ij} &= k_{ij}\omega_{ij}\exp(k_{ii}\omega_{ii}), \ \text{ for } 1 \le j < i \le p, \end{aligned} \tag{54}$$

respectively.

By Lemma 3.1, we have

$$\lim_{k\to\infty} \boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega}) = \hat{\boldsymbol{\Sigma}}_0(\boldsymbol{L}) \tag{55}$$

and by the continuity of $L(\cdot)$,

$$\lim_{k\to\infty} L(\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega})) = L(\hat{\boldsymbol{\Sigma}}_0(\boldsymbol{L})). \tag{56}$$

Also, by (52) and Fatou's lemma, we have

$$\begin{aligned} \liminf_{k\to\infty} r_k(\pi_k, \boldsymbol{\delta}_k^\pi) &= \liminf_{k\to\infty} \iint L(\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\bar{\pi}_1(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \\ &\ge \iint \liminf_{k\to\infty} L(\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\bar{\pi}_1(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \\ &= \iint L(\hat{\boldsymbol{\Sigma}}_0(\boldsymbol{L})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L})\bar{\pi}_1(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \\ &= \int \bar{\pi}_1(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \int_{\boldsymbol{L}\in\mathcal{T}^+} L(\hat{\boldsymbol{\Sigma}}_0(\boldsymbol{L})) f_W(\boldsymbol{L}|\boldsymbol{I}_p)\gamma(\mathrm{d}\boldsymbol{L}) \\ &= R_0. \quad \square \end{aligned}$$

Lemma 3.1.

$$\lim_{k\to\infty} \boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega}) = \hat{\boldsymbol{\Sigma}}_0(\boldsymbol{L}).$$

Proof. In the definition of $\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega})$ in (53), the Bayes estimator $\boldsymbol{\delta}_k^\pi$ under Stein's loss is

$$\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L}) = \left(\int_{\boldsymbol{Z}\in\mathcal{T}^+} \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} f_W(\boldsymbol{\Theta}^{-1}\boldsymbol{L}|\boldsymbol{Z})\pi_k(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}\right)^{-1} \int_{\boldsymbol{Z}\in\mathcal{T}^+} f_W(\boldsymbol{\Theta}^{-1}\boldsymbol{L}|\boldsymbol{Z})\pi_k(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}, \tag{57}$$

where Lemma A.1 is applied. By change of variables $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Theta}^{-1}$, we have

$$\boldsymbol{\delta}_k^\pi(\boldsymbol{\Theta}^{-1}\boldsymbol{L}) = \boldsymbol{\Theta}^{-1}\left(\int_{\boldsymbol{Y}\in\mathcal{T}^+} \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y}\boldsymbol{\Theta})\mathrm{d}\boldsymbol{Y}\right)^{-1} (\boldsymbol{\Theta}^{\mathrm{T}})^{-1} \int_{\boldsymbol{Y}\in\mathcal{T}^+} f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y}\boldsymbol{\Theta})\mathrm{d}\boldsymbol{Y} \tag{58}$$

and hence

$$\boldsymbol{\delta}_*(\boldsymbol{L} \mid \boldsymbol{k}, \boldsymbol{\omega}) = \left(\int_{\boldsymbol{Y}\in\mathcal{T}^+} \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y}\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega}))\mathrm{d}\boldsymbol{Y}\right)^{-1} \int_{\boldsymbol{Y}\in\mathcal{T}^+} f_W(\boldsymbol{L}|\boldsymbol{Y})\pi_k(\boldsymbol{Y}\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega}))\mathrm{d}\boldsymbol{Y},$$

where $\pi_k(\boldsymbol{\Theta})$ given in (48) is rewritten as

$$\pi_k(\boldsymbol{\Theta}) = \frac{1}{\prod_{j\le i} k_{ij}} \frac{1}{\prod_{i=1}^p \theta_{ii}^i} \prod_{i=1}^p \phi(\log\theta_{ii}/k_{ii}) \prod_{j<i} \phi(\theta_{ij}/\{\theta_{ii}k_{ij}\}). \tag{59}$$

Consider $\pi_k(\boldsymbol{Y}\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega}))$ in the following where θ_{ii} and θ_{ij} in (59) will be replaced with the diagonal components and the non-diagonal components of $\boldsymbol{Y}\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})$. By (54), the (i, i) diagonal component of $\boldsymbol{Y}\boldsymbol{\Theta}(\boldsymbol{k}\bullet\boldsymbol{\omega})$ with $\boldsymbol{Y} \in \mathcal{T}^+$ is

$$y_{ii}\exp(k_{ii}\omega_{ii}) \tag{60}$$

and the non-diagonal (i,j) component for $1 \le j < i \le p$ is

$$y_{ij}\exp(k_{jj}\omega_{jj}) + \sum_{l=j+1}^{i-1} y_{il}\exp(k_{ll}\omega_{ll})k_{lj}\omega_{lj} + y_{ii}k_{ij}\omega_{ij}\exp(k_{ii}\omega_{ii}).$$

Then, for $i > j$, $\theta_{ij}/\{\theta_{ii}k_{ij}\}$ in (59) is replaced with

$$\frac{1}{k_{ij}}\frac{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ij}}{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ii}} = \frac{y_{ij}}{y_{ii}}\frac{\exp(k_{jj}\omega_{jj} - k_{ii}\omega_{ii})}{k_{ij}} + \omega_{ij} + \sum_{l=j+1}^{i-1} w_{lj}\frac{y_{il}}{y_{ii}}\frac{k_{lj}}{k_{ij}}\exp(k_{ll}\omega_{ll} - k_{ii}\omega_{ii}). \tag{61}$$

Recall we set

$$k_{ii} = k, \quad k_{ij} = k^{(i-j)k}.$$

Then (61) is equal to

$$\frac{1}{k_{ij}}\frac{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ij}}{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ii}}\frac{y_{ij}}{y_{ii}}\left(\frac{\exp(\omega_{jj} - \omega_{ii})}{k^{i-j}}\right)^k + \omega_{ij} + \sum_{l=j+1}^{i-1} w_{lj}\frac{y_{il}}{y_{ii}}\left(\frac{\exp(\omega_{ll} - \omega_{ii})}{k^{i-l}}\right)^k$$

and hence it follows that

$$\lim_{k\to\infty}\frac{1}{k_{ij}}\frac{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ij}}{(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ii}} = \omega_{ij}.$$

Similarly, by (60), $\log\theta_{ii}/k_{ii}$ in (59) is replaced with

$$\frac{1}{k_{ii}}\log(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ii} = \frac{\log y_{ii} + k_{ii}\omega_{ii}}{k_{ii}}$$

and we have

$$\lim_{k\to\infty}\frac{1}{k_{ii}}\log(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))_{ii} = \omega_{ii}.$$

Therefore

$$\prod_{i\ge j} k_{ij}\prod_{i=1}^{p}\exp(ik_{ii}\omega_{ii})\pi_k(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega})) \le \frac{\{\phi(0)\}^{p(p+1)/2}}{\prod_{i=1}^{p} y_{ii}^{i}}, \tag{62}$$

and

$$\lim_{k\to\infty}\prod_{i\ge j} k_{ij}\prod_{i=1}^{p}\exp(ik_{ii}\omega_{ii})\pi_k(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega})) = \frac{\prod_{i\ge j}\phi(\omega_{ij})}{\prod_{i=1}^{p} y_{ii}^{i}}. \tag{63}$$

Then, by (62), we have

$$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} f_W(\boldsymbol{L}|\boldsymbol{Y})\left\{\prod_{i\ge j} k_{ij}\prod_{i=1}^{p}\exp(ik_{ii}\omega_{ii})\pi_k(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))\right\} \le \{\phi(0)\}^{p(p+1)/2}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} f_W(\boldsymbol{L}|\boldsymbol{Y})\prod_{i=1}^{p} y_{ii}^{-i}$$

and

$$f_W(\boldsymbol{L}|\boldsymbol{Y})\left\{\prod_{i\ge j} k_{ij}\prod_{i=1}^{p}\exp(ik_{ii}\omega_{ii})\pi_k(\boldsymbol{Y}\Theta(\boldsymbol{k}\bullet\boldsymbol{\omega}))\right\} \le \{\phi(0)\}^{p(p+1)/2} f_W(\boldsymbol{L}|\boldsymbol{Y})\prod_{i=1}^{p} y_{ii}^{-i}$$

which are both integrable under the region $\boldsymbol{Y} \in \mathcal{T}^+$. By the dominated convergence theorem with (63),

$$\lim_{k\to\infty}\delta_*(\boldsymbol{L} \mid \boldsymbol{k},\boldsymbol{\omega}) = \left(\int_{\boldsymbol{Y}\in\mathcal{T}^+}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} f_W(\boldsymbol{L}|\boldsymbol{Y})\gamma(\mathrm{d}\boldsymbol{Y})\right)^{-1}\int_{\boldsymbol{Y}\in\mathcal{T}^+} f_W(\boldsymbol{L}|\boldsymbol{Y})\gamma(\mathrm{d}\boldsymbol{Y}) = \hat{\Sigma}_0(\boldsymbol{L}). \quad \square$$

4. Concluding remarks

We have reviewed some known results on establishing minimaxity of best equivariant procedures. While none of the results established are new, the proofs of minimaxity are somewhat divergent from the typical minimaxity proofs in the literature in that the least favorable sequence is smooth and strictly positive on the support of the approximated right invariant measure: it is not a sequence of truncated versions of the invariant prior on expanding sets. In this sense, our proofs are in the same spirit as the common textbook proof of minimaxity of the mean of a normal distribution. In fact the same sequence of priors that works in the normal case is shown to work in the general location case. Hence the present method provides a degree of unification and simultaneously simplifies the proofs.

We note that the Gaussian kernel is not necessary and could be replaced by a bounded, continuous, positive density.

Our choices of particular loss featured in each of the problems also simplified the analyses, in the sense that, in each case, the form of the Bayes estimate could be explicitly given. This facilitated the use of Fatou's Lemma in establishing the limiting Bayes risk as being equal to the constant risk of the best equivariant estimator.

An approach for more general loss function could be constructed by requiring that all Bayes estimators (for priors with full support) be unique, and that the loss is sufficiently smooth that statements such as (31) hold in each problem.

It is also worth noting that, as in Tsukuma and Kubokawa (2015), in the problem of estimating a covariance matrix, a specific least favorable sequence of priors is established.

Acknowledgments

The authors thank the referees for their constructive and insightful comments that helped improve this manuscript.

Appendix. Lemmas and proofs

Lemma A.1. *For given positive definite matrix $\boldsymbol{H}$, consider the following function of $\boldsymbol{\Psi}$,*

$$Q(\boldsymbol{\Psi}) = \mathrm{tr}\boldsymbol{\Psi}\boldsymbol{H} - \log|\boldsymbol{\Psi}|.$$

Then

$$\arg\min_{\boldsymbol{\Psi}} Q(\boldsymbol{\Psi}) = \boldsymbol{H}^{-1}.$$

Proof. See proof of Theorem 3.1.5 of Muirhead (1982). □

Lemma A.2. *Let*

$$A = \left\{\boldsymbol{x} : 0 < \int_{-\infty}^{\infty} f(\boldsymbol{x}-\mu)\mathrm{d}\mu < \infty\right\}.$$

where $f(\boldsymbol{x}-\mu)$ is given by (1). Then $\Pr(A) = 1$.

Proof. Note, for any location parameter $\theta \in \mathbb{R}$,

$$\int_{\mathbb{R}^n} f(\boldsymbol{x}-\theta)\mathrm{d}\boldsymbol{x} = 1.$$

Let $y_i = x_i - x_n$ for $i = 1, \ldots, n-1$ and $t = x_n - \theta$. Then

$$\int_{\mathbb{R}^{n-1}} \left\{\int_{\mathbb{R}} f(y_1 + t, \ldots, y_{n-1} + t, t)\mathrm{d}t\right\} \mathrm{d}\boldsymbol{y} = 1,$$

where the integral inside is the marginal density of $\boldsymbol{y} \in \mathbb{R}^{n-1}$. Making the transformation $t = v - \mu$ in the integral inside, where μ is a variable and v is an arbitrary constant, we have

$$\int_{\mathbb{R}} f(y_1 + t, \ldots, y_{n-1} + t, t)\mathrm{d}t = \int_{-\infty}^{\infty} f(y_1 + v - \mu, \ldots, y_{n-1} + v - \mu, v - \mu)\mathrm{d}\mu.$$

Note any probability density should be strictly positive and finite with probability 1. Then, for $v = x_n$, we have

$$\begin{aligned} 1 &= \Pr\left(\int_{\mathbb{R}} f(y_1 + t, \ldots, y_{n-1} + t, t)\mathrm{d}t \in (0, \infty)\right) \\ &= \Pr\left(\int_{-\infty}^{\infty} f(y_1 + x_n - \mu, \ldots, y_{n-1} + x_n - \mu, x_n - \mu)\mathrm{d}\mu \in (0, \infty)\right), \end{aligned}$$

which completes the proof. □

Lemma A.3. *Let*

$$B = \left\{\boldsymbol{x} : \int_{-\infty}^{\infty} |\mu| f(\boldsymbol{x}-\mu)\mathrm{d}\mu < \infty\right\}.$$

Assume that there exists an equivariant estimator δ_0 with finite expected squared loss. Then $\Pr(B) = 1$.

Proof. The constant risk of δ_0 is $E[\delta_0^2] < \infty$. Hence $E[|\delta_0|] < \infty$ follows. Also $E[|\delta_0|| \boldsymbol{Y}] < \infty$ with probability 1 follows where $\boldsymbol{Y} = (X_1 - X_n, \dots, X_{n-1} - X_n)$. By Theorem 1.6 of Chapter 3 of Lehmann and Casella (1998), δ_0 can be expressed as $X_n - u(\boldsymbol{Y})$ with some $u(\boldsymbol{y})$. Since $|x_n| \le |x_n - u(\boldsymbol{y})| + |u(\boldsymbol{y})|$, $E[|X_n|| \boldsymbol{Y} = \boldsymbol{y}] < \infty$ with probability 1 follows.

Note

$$\int_{\mathbb{R}^n} |x_n| f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} \left\{ \int_{-\infty}^{\infty} |v - \mu| f(y_1 + v - \mu, \dots, y_{n-1} + v - \mu, v - \mu) \mathrm{d}\mu \right\} \mathrm{d}\boldsymbol{y}$$
$$= \int_{\mathbb{R}^{n-1}} E[|X_n|| \boldsymbol{Y} = \boldsymbol{y}] g(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

where

$$g(\boldsymbol{y}) = \int_{-\infty}^{\infty} f(y_1 - \mu, \dots, y_{n-1} - \mu, -\mu) \mathrm{d}\mu,$$

which is the density of $\boldsymbol{y} \in \mathbb{R}^{n-1}$ and

$$E[|X_n|| \boldsymbol{Y} = \boldsymbol{y}] = \{g(\boldsymbol{y})\}^{-1} \int_{-\infty}^{\infty} |v - \mu| f(y_1 + v - \mu, \dots, y_{n-1} + v - \mu, v - \mu) \mathrm{d}\mu.$$

By the triangle inequality $|\mu| < |\mu - v| + |v|$ and Lemma A.2, the lemma follows. □

Lemma A.4. *Let*

$$A = \left\{ \boldsymbol{x} : 0 < \int_0^{\infty} \sigma^{-n-1} f(\boldsymbol{x}/\sigma) \mathrm{d}\sigma < \infty \right\}.$$

Then $\Pr(A) = 1$.

Proof. Note, for any scale parameter $\nu > 0$,

$$\int_{\mathbb{R}^n} \nu^{-n} f(\boldsymbol{x}/\nu) \mathrm{d}\boldsymbol{x} = 1.$$

Then

$$1 = \int_{\mathbb{R}^n} \nu^{-n} f(\boldsymbol{x}/\nu) \mathrm{d}\boldsymbol{x}$$
$$= \int_{\mathbb{R}^{n-1}} \left\{ \int_0^{\infty} \sum_{j=\{-1,1\}} \nu^{-n} f(x_1/\nu, \dots, x_{n-1}/\nu, jt/\nu) \mathrm{d}t \right\} \mathrm{d}x_1, \dots \mathrm{d}x_{n-1}$$
$$= \int_{\mathbb{R}^{n-1}} \left\{ \int_0^{\infty} \sum_{j=\{-1,1\}} \nu^{-1} t^{n-1} f(z_1 t, \dots, z_{n-1} t, jt/\nu) \mathrm{d}t \right\} \mathrm{d}\boldsymbol{z} \quad \frac{x_i}{\nu} = z_i t \quad (i = 1, \dots, n-1) \quad \frac{\mathrm{d}x_i}{\mathrm{d}z_i} = \nu t$$
$$= \int_{\mathbb{R}^{n-1}} \left\{ \sum_{j=\{-1,1\}} \int_0^{\infty} \frac{v^n}{\nu \sigma^{n+1}} f(z_1 v/\sigma, \dots, z_{n-1} v/\sigma, \{jv/\nu\}/\sigma) \mathrm{d}\sigma \right\} \mathrm{d}\boldsymbol{z} \quad t = \frac{v}{\sigma} \quad \frac{\mathrm{d}t}{\mathrm{d}\sigma} = \frac{v}{\sigma^2},$$

where v is an arbitrary positive constant. Hence the integral inside is the marginal density of $\boldsymbol{z} \in \mathbb{R}^{n-1}$. A similar argument to that of Lemma A.2 completes the proof. □

Lemma A.5. *Let*

$$B = \left\{ \boldsymbol{x} : \int_0^{\infty} \sigma^{-n-2} f(\boldsymbol{x}/\sigma) \mathrm{d}\sigma < \infty \right\}.$$

Assume that there exists an equivariant estimator δ_0 with finite expected entropy loss. Then $\Pr(B) = 1$.

Proof. Note, for $e = \exp(1)$,

$$x - \log x - 1 \ge (e - 1)x/e - 1 \text{ for } x \ge e. \tag{A.1}$$

The constant risk of δ_0 is $E_1[\delta_0 - \log \delta_0 - 1] < \infty$. Together with the inequality (A.1), we have $E[\delta_0 \mid \delta_0 > e] < \infty$ and

$$E[\delta_0] \le e + E[\delta_0 \mid \delta_0 > e] < \infty.$$

Also $E[\delta_0 \mid \boldsymbol{Z} = \boldsymbol{z}] < \infty$ with probability 1 follows where

$$\boldsymbol{Z} = (X_1/|X_n|, \dots, X_{n-1}/|X_n|).$$

By Theorem 3.1 of Chapter 3 of Lehmann and Casella (1998), δ_0 can be expressed as $|X_n|/u(\boldsymbol{Z})$ with some $u(\boldsymbol{z})$. Hence $E[|X_n|| \boldsymbol{Z}=\boldsymbol{z}]<\infty$ with probability 1 follows. Further $E[|X_n|| \boldsymbol{Z}=\boldsymbol{z}]$ is expressed as

$$E[|X_n|| \boldsymbol{Z}=\boldsymbol{z}]=\frac{\int_0^\infty \sum_{j=\{-1,1\}} tt^{n-1}f(tz_1,\ldots,tz_{n-1},jt)\mathrm{d}t}{\int_0^\infty \sum_{j=\{-1,1\}} t^{n-1}f(tz_1,\ldots,tz_{n-1},jt)\mathrm{d}t}=v\frac{\int_0^\infty \sum_{j=\{-1,1\}} \sigma^{-n-2}f(z_1v/\sigma,\ldots,z_{n-1}v/\sigma,jv/\sigma)\mathrm{d}\sigma}{\int_0^\infty \sum_{j=\{-1,1\}} \sigma^{-n-1}f(z_1v/\sigma,\ldots,z_{n-1}v/\sigma,jv/\sigma)\mathrm{d}\sigma}.$$

Since the denominator is strictly positive with probability 0 by Lemma A.4, the lemma follows. □

References

Anderson, T.W., 2003. An introduction to multivariate statistical analysis, third ed. In: Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
Girshick, M.A., Savage, L.J., 1951. Bayes and minimax estimates for quadratic loss functions. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950. University of California Press, Berkeley and Los Angeles, pp. 53–73.
Hora, R.B., Buehler, R.J., 1966. Fiducial theory and invariant estimation. Ann. Math. Stat. 37, 643–656.
James, W., Stein, C., 1961. Estimation with quadratic loss. In: Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I. Univ. California Press, Berkeley, Calif., pp. 361–379.
Kiefer, J., 1957. Invariance, minimax sequential estimation, and continuous time processes. Ann. Math. Stat. 28, 573–601.
Kubokawa, T., 2004. Minimaxity in estimation of restricted parameters. J. Japan Statist. Soc. 34, 229–253.
Lehmann, E.L., Casella, G., 1998. Theory of point estimation, second ed. In: Springer Texts in Statistics, Springer-Verlag, New York.
Muirhead, R.J., 1982. Aspects of multivariate statistical theory. In: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York.
Pitman, E.J.G., 1939. The estimation of the location and scale parameters of a continuous population of any given form. Biometrika 30, 391–421.
Stein, C., 1956. Some Problems in Multivariate Analysis Part I. Technical Report 6, Department of Statistics, Stanford University.
Tsukuma, H., Kubokawa, T., 2015. Minimaxity in estimation of restricted and non-restricted scale parameter matrices. Ann. Inst. Statist. Math. 67, 261–285.