Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 92 (2005) 324–342
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Estimation of the entropy of a multivariate normal distribution

Neeraj Misra,[a,b] Harshinder Singh,[a,c,*] and Eugene Demchuk[c,d]

[a] *Department of Statistics, West Virginia University, Morgantown, WV 26506-6330, USA*
[b] *Department of Mathematics, Indian Institute of Technology Kanpur, Kanpur 208 016, India*
[c] *Health Effects Laboratory Division, National Institute for Occupational Safety and Health, Morgantown, WV 26505-2888, USA*
[d] *School of Pharmacy, West Virginia University, Morgantown, WV 26506-9500, USA*

Received 27 June 2002

Abstract

Motivated by problems in molecular biosciences wherein the evaluation of entropy of a molecular system is important for understanding its thermodynamic properties, we consider the efficient estimation of entropy of a multivariate normal distribution having unknown mean vector and covariance matrix. Based on a random sample, we discuss the problem of estimating the entropy under the quadratic loss function. The best affine equivariant estimator is obtained and, interestingly, it also turns out to be an unbiased estimator and a generalized Bayes estimator. It is established that the best affine equivariant estimator is admissible in the class of estimators that depend on the determinant of the sample covariance matrix alone. The risk improvements of the best affine equivariant estimator over the maximum likelihood estimator (an estimator commonly used in molecular sciences) are obtained numerically and are found to be substantial in higher dimensions, which is commonly the case for atomic coordinates in macromolecules such as proteins. We further establish that even the best affine equivariant estimator is inadmissible and obtain Stein-type and Brewster–Zidek-type estimators dominating it. The Brewster–Zidek-type estimator is shown to be generalized Bayes.

AMS 1991 subject classifications: 62H12

Keywords: Affine equivariant estimators; Brewster–Zidek-type estimator; Entropy; Generalized Bayes estimator; Inadmissible estimator; Quadratic loss function; Risk function; Stein-type estimator; Wishart distribution

*Corresponding author. Department of Statistics, West Virginia University, Morgantown, WV 26506, USA.
E-mail address: hsingh@stat.wvu.edu (H. Singh).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2003.10.003

1. Introduction

The atomic coordinates of molecules fluctuate randomly. The extent of these fluctuations determines the thermodynamic properties and shapes of molecules. The evaluation of thermodynamic properties, including entropy, is an important problem in molecular biology, chemistry and molecular physics [7,10]. The internal entropy of a molecule depends on the random fluctuations in its atomic coordinates and, for the evaluation of entropy, researchers have developed probabilistic models for modelling these fluctuations. The simplest model, known as the *normal mode* or *harmonic* analysis, is based on the expansion of energy function. At stationary points and when the atomic fluctuations are small, the potential energy function is reasonably approximated by the second term of the expansion, thus resulting in a quadratic-form dependence of the potential energy on atomic displacements and, consequently, in a multivariate normal probability density function [4]. Karplus and Kushik [8] and Levy et al. [11] added a fitting approach to this idea, which in their interpretation is known as the *quasi-harmonic* analysis. They proposed modelling the p coordinates of a macromolecule, which may not be strictly harmonic, by a p-variate normal distribution. The entropy of a p-variate normal random variable X_1, having probability density function $f_{\mu,\Sigma}(x)$, where $\mu \in \mathscr{R}^p$ is the mean vector and Σ is the $p \times p$ positive definite covariance matrix, is given by

$$\begin{aligned} H_p(\Sigma) &= E_{\mu,\Sigma}(-\ln f_{\mu,\Sigma}(X_1)) \\ &= \frac{p}{2}[1 + \ln(2\pi)] + \frac{\ln|\Sigma|}{2}. \end{aligned} \tag{1.1}$$

Normal mode and quasi-harmonic analysis are widely used for the analysis of conformational dynamics in biological macromolecules, including ligand–receptor interactions [6], protein folding [16] and gene regulation [5]. These models are especially appropriate for estimation of entropy at the very core of tight macromolecular assemblies and in crystallographic studies. A common practice in molecular sciences is to estimate $H_p(\Sigma)$ by its maximum likelihood estimator (mle) $H_p(S/n)$, where S/n is the sample covariance matrix based on a random sample of size n. In the statistical literature on estimation of various functions of the covariance matrix Σ, it has been observed that usual estimates based on S/n are often not optimal and better alternatives can be found. This motivates us to search for better alternatives to $H_p(S/n)$ as an estimator of $H_p(\Sigma)$, which will be useful to researchers working in molecular sciences. We, therefore, deal with the decision-theoretic estimation of $H_p(\Sigma)$.

Let $X_1, \ldots, X_n$ be a random sample drawn from a p-variate normal distribution $N_p(\mu, \Sigma)$ $(n \geqslant p+1)$, where the mean vector $\mu \in \mathscr{R}^p$ and the $p \times p$ positive definite covariance matrix Σ are assumed to be unknown. Based on $X_1, \ldots, X_n$, we desire to estimate the entropy $H_p(\Sigma)$, given by (1.1), under the quadratic loss function. Note that, under the quadratic loss function, the problem of estimating $H_p(\Sigma)$ is equivalent to that of estimating $\ln|\Sigma|$. For an estimator δ of $\ln|\Sigma|$, the corresponding estimator of the entropy $H_p(\Sigma)$ is given by $\delta^E = [p\{1 + \ln(2\pi)\} + \delta]/2$. For notational simplicity, we prefer to deal with the estimation of $\ln|\Sigma|$ under the quadratic loss function

$$L(\delta, \ln|\Sigma|) = (\delta - \ln|\Sigma|)^2. \tag{1.2}$$

The risk function and the bias function of an estimator δ of $\ln|\Sigma|$ will be denoted by $R(\delta,\theta) = E_\theta(L(\delta, \ln|\Sigma|))$ and $B(\delta,\theta) = E_\theta(\delta - \ln|\Sigma|)$, respectively, where $\theta = (\mu, \Sigma)$. Note, that the risk function and the bias function of the corresponding estimator $\delta^E = [p\{1+\ln(2\pi)\} + \delta]/2$ of the entropy $H_p(\Sigma)$ are given by $R_E(\delta^E, \theta) = R(\delta,\theta)/4$ and $B_E(\delta^E,\theta) = B(\delta,\theta)/2$, respectively.

Although problems of estimating the generalized variance $|\Sigma|$ and the generalized precision $|\Sigma^{-1}|$ have received considerable attention in the past [9,13,17–19,21,22], to the best of our knowledge, estimation of $\ln|\Sigma|$ (which is equivalent to estimation of the entropy $H_p(\Sigma)$) has not been addressed before.

Define

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad X = \sqrt{n}\bar{X} \quad \text{and} \quad S = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^t,$$

so that the statistic (X, S) is minimal-sufficient, $X \sim N_p(\sqrt{n}\mu, \Sigma)$, $S \sim W_p(n-1, \Sigma)$ and X and S are independently distributed; here $W_p(n-1,\Sigma)$ denotes the p-variate Wishart distribution with $n-1$ degrees of freedom and covariance matrix Σ. Since the statistic (X,S) is minimal-sufficient, it is reasonable to consider only those estimators which depend on observations only through (X,S).

In Section 2, we consider affine equivariant estimators and derive the best affine equivariant estimator. Interestingly, the best affine equivariant turns out to be unbiased, which is a rare statistical phenomenon, and it also turns out to be a generalized Bayes estimator. We also establish the admissibility of the best affine equivariant estimator among estimators that depend on $|S|$ alone. In Section 3, we establish the inadmissibility of the best affine equivariant estimator in the class of all estimators by deriving a Stein-type estimator dominating it. Section 4 provides a Brewster–Zidek-type estimator dominating the best affine equivariant estimator. The Brewster–Zidek-type estimator is shown to be generalized Bayes.

2. Affine equivariant estimators

Note that the estimation problem under study is invariant under the affine transformations:

$$(X, S) \to (CX + D, CSC^t), \quad (\mu, \Sigma) \to (C\mu + D, C\Sigma C^t),$$

where C is any non-singular $p \times p$ matrix and D is any $p \times 1$ vector. Under this affine transformation, $\ln|\Sigma| \to \ln|\Sigma| + \ln|C|^2$ and therefore it is reasonable to require that an estimator $\delta(X,S)$ satisfies

$$\delta(CX + D, CSC^t) = \delta(X,S) + \ln|C|^2, \tag{2.1}$$

for all $p \times p$ non-singular real matrices C and all $p \times 1$ vectors D. An estimator of the form (2.1) is called an affine equivariant estimator. A standard argument shows that any affine equivariant estimator is of the form

$$\delta_c(X, S) = \ln |S| - c, \tag{2.2}$$

for some real constant c.

Let $c_1 = p \ln n$, so that δ_{c_1} is the mle of $\ln |\Sigma|$. We desire to find better alternatives to the estimator δ_{c_1}, which is the commonly used estimator in molecular sciences. It is well known [12, p. 100] that

$$\frac{|S|}{|\Sigma|} \sim \prod_{i=1}^{p} \chi^2_{n-i}, \tag{2.3}$$

where the χ^2_{n-i}, for $i = 1, \ldots, p$, denote independent central chi-square random variables with $n - i$ degrees of freedom. Thus, the risk and bias of any affine equivariant estimator do not depend on θ and therefore, for any real c, we denote $R(\delta_c, \theta)$ and $B(\delta_c, \theta)$ by $R(\delta_c)$ and $B(\delta_c)$, respectively. Let $\Gamma(x),\ x > 0$, denote the gamma function and let $\psi(x) = \frac{d}{dx}(\ln \Gamma(x)),\ x > 0$, denote the digamma function [1]. The following theorem gives the best affine equivariant estimator of $\ln |\Sigma|$.

Theorem 2.1. *Under the quadratic loss function* (1.2), *the best affine equivariant estimator of* $\ln |\Sigma|$ *is given by* $\delta_{c_0}(X, S)$, *where* $c_0 = p \ln 2 + \sum_{i=1}^{p} \psi(\frac{n-i}{2})$. *Moreover,* $\delta_{c_0}(X, S)$ *is also an unbiased estimator.*

Proof. For a real constant c, the risk of the estimator $\delta_c(X, S)$, as defined by (2.2), is given by

$$R(\delta_c) = E_\theta \left(\ln \frac{|S|}{|\Sigma|} - c \right)^2,$$

which is minimized (for any θ) at

$$c \equiv C_{OPT} = E_\theta \left(\ln \frac{|S|}{|\Sigma|} \right).$$

Using (2.3), we get

$$\begin{aligned} C_{OPT} &= \sum_{i=1}^{p} E(\ln \chi^2_{n-i}) \\ &= p \ln 2 + \sum_{i=1}^{p} \psi\left(\frac{n-i}{2} \right) \\ &= c_0. \end{aligned}$$

Clearly δ_{c_0} is also unbiased for $\ln |\Sigma|$. Hence the result follows. □

Remark 2.1. (i) Interestingly, the best affine equivariant estimator δ_{c_0} is also an unbiased estimator of $\ln |\Sigma|$, which is a rare statistical phenomenon. It follows that the best affine equivariant estimator δ_{c_0} is a better alternative to δ_{c_1} both in terms of bias and risk.

(ii) Using Jensen's inequality to the function $\ln \chi^2_{n-i}$, it follows that

$$E(\ln \chi^2_{n-i}) < \ln(n-i),$$

implying that

$$c_0 < \sum_{i=1}^{p} \ln(n-i) < p \ln n = c_1.$$

Thus, the maximum likelihood estimator δ_{c_1} is negatively biased and, therefore, it follows that the mle δ_{c_1} under-estimates $\ln|\Sigma|$.

The following theorem provides the expression for the bias of the estimator δ_{c_1} and expressions for the risks of estimators δ_{c_0} and δ_{c_1}. This theorem also describes behaviors of $|B(\delta_{c_1})|$ and $\Delta = R(\delta_{c_1}) - R(\delta_{c_0})$, as functions of p.

Theorem 2.2. *Let* $c_0 = p\ln 2 + \sum_{i=1}^{p} \psi(\frac{n-i}{2}), c_1 = p \ln n$, *and for* $z>0$ *let* $\psi^{(1)}(z) = \frac{d}{dz}\psi(z)$ *denote the trigamma function. Then*

(i) $B(\delta_{c_1}) = -[p \ln \frac{n}{2} - \sum_{i=1}^{p} \psi(\frac{n-i}{2})]$.
(ii) $R(\delta_{c_0}) = \sum_{i=1}^{p} \psi^{(1)}(\frac{n-i}{2})$.
(iii) $R(\delta_{c_1}) = R(\delta_{c_0}) + [p \ln \frac{n}{2} - \sum_{i=1}^{p} \psi(\frac{n-i}{2})]^2$.
(iv) *The absolute bias of the mle* δ_{c_1} *is an increasing function of* p $(1 \leqslant p \leqslant n-1)$.
(v) *The risk difference* $\Delta = R(\delta_{c_1}) - R(\delta_{c_0})$ *is an increasing function of* p $(1 \leqslant p \leqslant n-1)$.

Proof. The proof of assertion (i) is obvious. Assertion (ii) follows from using (2.3) along with the fact that, for a χ^2_v random variable $\mathrm{Var}(\ln \chi^2_v) = \psi^{(1)}(\frac{v}{2})$. The proof of assertion (iii) is again obvious. For proving assertions (iv) and (v), we are required to establish that $D_1(p) = p \ln \frac{n}{2} - \sum_{i=1}^{p} \psi(\frac{n-i}{2})$ is an increasing function of $p \in \{1, 2, \ldots, n-1\}$. We have

$$\begin{aligned} D_1(p+1) - D_1(p) &= \ln \frac{n}{2} - \psi\left(\frac{n-p-1}{2}\right) \\ &> \ln \frac{n}{2} - \ln \frac{n-p-1}{2} \\ &> 0, \ \forall \ p = 1, 2, \ldots, n-2, \end{aligned}$$

since, for $z>0$, $\psi(z) < \ln z$ (using Jensen's inequality). Hence the result follows.

For various combinations of n and p $(n \geqslant p+1)$, values of the absolute bias $|B(\delta_{c_1})|$ and the risk improvement $RI(\delta_{c_0}, \delta_{c_1}) = (R(\delta_{c_1}) - R(\delta_{c_0}))/R(\delta_{c_1}) \times 100\%$ are tabulated in Table 1. It is evident from Table 1 that the best affine equivariant estimator δ_{c_0} gives significant (up to 100%) improvements over the mle δ_{c_1} for higher values of p. Moreover the mle δ_{c_1} is heavily biased for large values of p. This, along

Table 1
Values of $|B(\delta_{c_1})|$ and $RI(\delta_{c_0}, \delta_{c_1})$

n	p	$\|B(\delta_{c_1})\|$	$RI(\delta_{c_0}, \delta_{c_1})$
10	1	0.2206	16.3598
	2	0.5738	38.2119
	5	2.6732	80.3434
25	1	0.0830	7.3561
	2	0.2106	19.9629
	5	0.8812	61.8996
	10	3.1290	89.8411
50	1	0.0408	3.8331
	2	0.1026	11.1022
	5	0.4188	44.6436
	10	1.4136	81.2315
	25	8.7296	98.1347
100	1	0.0202	1.9575
	2	0.0506	5.8802
	5	0.2046	28.6553
	10	0.6766	68.1360
	25	3.8584	96.216
	50	16.3916	99.4754
500	1	0.0040	0.3983
	2	0.0100	1.2345
	5	0.0402	7.4131
	10	0.1310	29.7521
	25	0.7128	83.1548
	50	2.7460	97.2731
	100	11.0778	99.6364

with results (iv) and (v) of Theorem 2.2, suggests that the use of the mle δ_{c_1} may give undesirable results in higher dimensions, which are common in molecular biology where macromolecules, such as proteins, have a very large number of atomic coordinates. □

In the following theorem, we establish that the best affine equivariant estimator δ_{c_0} is also generalized Bayes.

Theorem 2.3. *The best affine equivariant estimator δ_{c_0} is generalized Bayes with respect to the prior on (μ, Σ) with density*

$$\Pi(\mu, \Sigma) = \frac{1}{|\Sigma|^{\frac{p+1}{2}}}, \quad \Sigma > 0, \ \mu \in \mathscr{R}^p,$$

where $\Sigma > 0$ means that Σ is positive definite.

Proof. Since, given (μ, Σ), X and S are independently distributed as $N_p(\sqrt{n}\mu, \Sigma)$ and $W_p(n-1, \Sigma)$ respectively, the (formal) posterior density of (μ, Σ), given (X, S), is

$$K_1(S)\frac{1}{|\Sigma|^{\frac{n+p+1}{2}}}e^{-\frac{1}{2}(X-\sqrt{n}\mu)^t\Sigma^{-1}(X-\sqrt{n}\mu)}e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}S)}, \quad \mu\in\mathscr{R}^p,\ \Sigma>0,$$

where $\operatorname{tr}(A)$ denotes the trace of matrix A,

$$K_1(S) = \frac{n^{\frac{p}{2}}|S|^{\frac{n-1}{2}}}{(2\pi)^{\frac{p}{2}}2^{\frac{(n-1)p}{2}}\Gamma_p\left(\frac{n-1}{2}\right)}$$

and $\Gamma_p(\cdot)$ denotes the p-variate gamma function.

Therefore, the posterior density of Σ, given (X, S), is

$$\frac{|S|^{\frac{n-1}{2}}}{2^{\frac{(n-1)p}{2}}\Gamma_p\left(\frac{n-1}{2}\right)|\Sigma|^{\frac{n+p}{2}}}e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}S)}, \quad \Sigma>0,$$

which is the density of inverted Wishart distribution with $n+p$ degrees of freedom and parameter matrix S (see [12], Problem 3.6). Thus, given (X, S),

$$\Sigma^{-1} \sim W_p(n-1, S^{-1})$$
$$\Rightarrow \frac{|\Sigma^{-1}|}{|S^{-1}|} \sim \prod_{i=1}^{p} \chi^2_{n-i}, \quad ([12], \text{Theorem } 3.2.15),$$

where the χ^2_{n-i}, for $i = 1, \dots, p$, are independent chi-square random variables.

Since the loss function is squared error, the generalized Bayes estimator of $\ln|\Sigma|$ is given by

$$\begin{aligned}\delta_{GB1}(X,S) &= E_\Sigma(\ln|\Sigma||(X,S)) \\ &= \ln|S| - \sum_{i=1}^{p} E(\ln\chi^2_{n-i}) \\ &= \delta_{c_0}(X,S).\end{aligned}$$

Hence the result follows.

Motivated by the work of Pal [14] who, under the entropy loss, established the admissibility of the best affine equivariant estimator of the generalized variance $|\Sigma|$ among estimators depending on $|S|$ alone, we address the admissibility of the best affine equivariant estimator, δ_{c_0}, of $\ln|\Sigma|$ among estimators depending on $|S|$ (or $\ln|S|$) alone. As pointed out by Pal [14], this problem is quite meaningful, especially when the population mean vector μ is known, as then one is forced to base his/her estimate entirely on a Wishart matrix. The following lemma due to Brown [3] (also see Pal [14]) will be useful in this direction. □

Lemma 2.1. *Let η be a real location parameter for the real random variable Q. Let R_0 denote the risk of the best invariant estimator Q of η under a loss function*

$L^*(\delta, \eta) = W^*(\delta - \eta)$, where $W^*(\cdot)$ is a non-negative function defined on the real line, and let $R^*(\delta, \eta)$ denote the risk (expected loss) function of an estimator δ. Assume that $R_0 < \infty$ and that

(i) $\lim_{i\to\infty} R^*(Q + d_i, \eta) = R^*(Q, \eta) \Rightarrow \lim_{i\to\infty} d_i = 0$,
(ii) $\int_0^\infty \{\sup_\gamma E_{\eta=0}((W^*(Q) - W^*(Q+\gamma))I_{(0,\lambda)}(|Q|))\}d\lambda < \infty$, *and*
(iii) $E_{\eta=0}(|Q|W^*(Q)) < \infty$,

where $I_A(.)$ denotes the indicator function of a set A. Then the best invariant estimator Q is admissible for estimating η under the loss function $L^*(\cdot, \cdot)$.

Theorem 2.4. *Under the loss function $L(\cdot, \cdot)$ (given by (1.2)), the best affine equivariant estimator δ_{c_0} is an admissible estimator of $\ln|\Sigma|$ in the class of estimators depending on $|S|$ alone.*

Proof. Let $Q = \ln|S|$ and $\eta = \ln|\Sigma| + c_0$. Then, under the loss function (1.2), δ_{c_0} is an admissible estimator of $\ln|\Sigma|$ in the class of estimators depending on $|S|$ alone if and only if Q is an admissible estimator of η in the class of estimators depending on $|S|$ alone. Thus, it suffices to verify conditions (i)–(iii) of Lemma 2.1.

(i) We have

$$
\begin{aligned}
R(Q + d_i, \eta) &= d_i^2 + \mathrm{Var}\left(\ln\frac{|S|}{|\Sigma|}\right)\\
&= d_i^2 + \mathrm{Var}\left(\ln\left(\prod_{j=1}^p \chi^2_{n-j}\right)\right)\\
&= d_i^2 + \sum_{j=1}^p \mathrm{Var}(\ln \chi^2_{n-j}),
\end{aligned}
$$

where the χ^2_{n-j}, for $j = 1, \ldots, p$, are independent chi-square random variables.

Thus,

$$
\begin{aligned}
&\lim_{i\to\infty} R(Q + d_i, \eta) = R(Q, \eta)\\
&\quad\Rightarrow \lim_{i\to\infty} d_i^2 + \sum_{j=1}^p \mathrm{Var}(\ln \chi^2_{n-j}) = \sum_{j=1}^p \mathrm{Var}(\ln \chi^2_{n-j})\\
&\quad\Rightarrow \lim_{i\to\infty} d_i = 0.
\end{aligned}
$$

(ii) Define,

$$M_\lambda(\gamma) = E_{\eta=0}((Q^2 - (Q+\gamma)^2)I_{(0,\lambda)}(|Q|)), \quad \gamma \in \mathscr{R}^1, \ \lambda > 0.$$

Clearly, for fixed $\lambda > 0$,

$$\sup_{\gamma\in\mathscr{R}^1} M_\lambda(\gamma) = \frac{(E_{\eta=0}(QI_{(0,\lambda)}(|Q|)))^2}{E_{\eta=0}(I_{(0,\lambda)}(|Q|))} = \frac{(h(\lambda))^2}{g(\lambda)}, \text{ say,}$$

where

$$h(\lambda) = E_{\eta=0}(QI_{(0,\lambda)}(|Q|)),\ \ \lambda > 0,$$

and

$$g(\lambda) = E_{\eta=0}(I_{(0,\lambda)}(|Q|)),\ \ \lambda > 0.$$

We are required to verify that

$$I = \int_0^\infty \frac{(h(\lambda))^2}{g(\lambda)}\, d\lambda < \infty.$$

Note that, under $\eta = 0$,

$$Q \sim \sum_{i=1}^{p} \ln \chi^2_{n-i} - c_0,$$

where the χ^2_{n-i}, for $i = 1, \ldots, p$, are independent chi-square random variables. Thus the moments of Q (under $\eta = 0$) can be expressed in terms of gamma and polygamma functions [1, p. 260]. Thus it follows that, under $\eta = 0$, Q has finite moments of all orders. Also, for any $\lambda > 0$,

$$\begin{aligned}(h(\lambda))^2 &= (E_{\eta=0}(QI_{(0,\lambda)}(|Q|)))^2\\ &\leqslant E_{\eta=0}(Q^2)E_{\eta=0}(I_{(0,\lambda)}(|Q|))\end{aligned}$$

$$\Rightarrow \frac{(h(\lambda))^2}{g(\lambda)} \leqslant E_{\eta=0}(Q^2) < \infty.$$

Moreover,

$$\lim_{\lambda\to\infty} \frac{h(\lambda)}{g(\lambda)} = E_{\eta=0}(Q) = 0.$$

Thus, it suffices to verify that

$$I_N = \int_N^\infty |h(\lambda)|\, d\lambda < \infty,$$

for large N.

We have

$$\begin{aligned}h(\lambda) &= E_{\eta=0}(QI_{(0,\lambda)}(|Q|))\\ &= -E_{\eta=0}(QI_{(\lambda,\infty)}(|Q|))\ \text{(since } E_{\eta=0}(Q) = 0).\end{aligned}$$

Thus,

$$\begin{aligned}|h(\lambda)| &= |E_{\eta=0}(QI_{(\lambda,\infty)}(|Q|))|\\ &\leqslant \sqrt{E_{\eta=0}(Q^2)}\sqrt{P_{\eta=0}(|Q| > \lambda)}\\ &\leqslant \sqrt{E_{\eta=0}(Q^2)}\sqrt{\frac{E_{\eta=0}(Q^4)}{\lambda^4}}.\end{aligned}$$

Therefore, for large N,

$$I_N \leqslant \sqrt{E_{\eta=0}(Q^2)}\sqrt{E_{\eta=0}(Q^4)} \int_N^\infty \frac{1}{\lambda^2}\, d\lambda < \infty .$$

(iii) We have

$$E_{\eta=0}(|Q|W(Q)) = E_{\eta=0}(|Q|Q^2) \leqslant \sqrt{E_{\eta=0}(Q^2)}\sqrt{E_{\eta=0}(Q^4)} < \infty ,$$

since Q has finite moments of all orders.

Now we address the problem of admissibility/inadmissibility of the best affine equivariant estimator δ_{c_0} in the class of all estimators. We observe that the estimator δ_{c_0} is inadmissible in the class of all estimators. In the following section, we obtain a Stein-type estimator dominating the best affine equivariant estimator δ_{c_0}.

3. Inadmissibility of the best affine equivariant estimator and an improved stein-type estimator

Using the idea of Stein [20], we will derive an estimator that dominates the best affine equivariant estimator δ_{c_0}. To do so, we explore a larger class than the affine equivariant estimators. One such class of estimators contains all estimators of the form

$$\delta_\phi(X, S) = \ln|S + XX^t| - \phi(T), \tag{3.1}$$

where $T = |S||S + XX^t|^{-1}$ and $\phi(.)$ is some real-valued function defined on the unit interval $[0, 1]$. Note that the choice $\phi(t) \equiv \phi_c(t) = c - \ln t$, for $c \in (-\infty, \infty)$, yields the class of affine equivariant estimators. Xiaoqian and Wankai [22] considered a similar class of estimators for the problem of estimating the generalized precision.

Let F be a $p \times p$ non-singular matrix such that $F\Sigma F^t = I_p$, the $p \times p$ identity matrix, and let $\tau = \sqrt{n}F\mu$. Let P be a $p \times p$ orthogonal matrix whose first row is $(\frac{\tau_1}{\sqrt{\lambda}}, \dots, \frac{\tau_p}{\sqrt{\lambda}})$, where $\lambda = ||\tau||^2$ and $||.||$ denotes the usual Euclidean norm. Define $U = PFX$, $W = PFSF^tP^t$ and $\mu_0 = (\sqrt{\lambda}, 0, \dots, 0)^t$, so that $U \sim N_p(\mu_0, I_p)$ and $W \sim W_p(n-1, I_p)$ are independently distributed. Then, the risk function of any estimator of the form (3.1) is given by

$$R(\delta_\phi, \theta) = E_\lambda(\ln|W + UU^t| - \phi(T))^2,$$

where $T = |W||W + UU^t|^{-1}$. It follows that the risk function of any estimator of the form (3.1) depends on θ only through λ and therefore, for notational simplicity, we denote $R(\delta_\phi, \theta)$ by $R_\lambda(\delta_\phi)$.

Following Xiaoqian and Wankai [22], let $V = (V_{ij})$ be the random lower-triangular matrix, having positive diagonal elements, such that $W + UU^t = VV^t$ and let $Y = V^{-1}U$. Then $|W + UU^t| = \prod_{i=1}^p V_{ii}^2$ and $T = |W||W + UU^t|^{-1} = 1 - Y^tY$.

Let K be a Poisson random variable with mean $\lambda/2$. Then, from Xiaoqian and Wankai [22], it follows that $(V_{11}^2, T), V_{22}^2, \ldots, V_{pp}^2$ are mutually independent; $V_{ii}^2 \sim \chi^2_{n-i+1},\ i = 2, \ldots, p$ and that, given $K = k$, $V_{11}^2 \sim \chi^2_{n+2k}$ and $T \sim \mathrm{Beta}(\frac{n-p}{2}, \frac{p}{2} + k)$ (beta distribution) are independently distributed. Under the above notation, the risk function of any estimator of the form (3.1) is given by

$$R_\lambda(\delta_\phi) = E_\lambda\left(\sum_{i=1}^{p} \ln V_{ii}^2 - \phi(T)\right)^2,\ \lambda > 0.$$

The above representation of the risk function can also be obtained using the ideas of Shorrock and Zidek [17]. In the following theorem, we establish the inadmissibility of the best affine equivariant estimator δ_{c_0} by providing a Stein-type estimator dominating it.

Theorem 3.1. *Define* $\phi_{ST}(t) = \max\{p \ln 2 + \sum_{i=1}^{p} \psi(\frac{n-i+1}{2}), c_0 - \ln t\},\ t \in [0, 1]$. *Then, under the quadratic loss function* (1.2), *the best affine equivariant estimator* δ_{c_0} *is inadmissible and is dominated by* $\delta_{\phi_{ST}}(X, S) = \ln|S + XX^t| - \phi_{ST}(T)$.

Proof. Consider the risk difference

$$\begin{aligned} R_\lambda(\delta_{c_0}) - R_\lambda(\delta_{\phi_{ST}}) &= E_\lambda\left(\sum_{i=1}^{p} \ln V_{ii}^2 - c_0 + \ln T\right)^2 - E_\lambda\left(\sum_{i=1}^{p} \ln V_{ii}^2 - \phi_{ST}(T)\right)^2 \\ &= E_\lambda(\Delta_1(T, \lambda)),\ \lambda > 0, \end{aligned}$$

where, for $t \in [0, 1]$ and $\lambda > 0$,

$$\begin{aligned} &\Delta_1(t, \lambda) \\ &\quad = [\phi_{ST}(t) - c_0 + \ln t]\left[2\sum_{i=2}^{p} E(\ln V_{ii}^2) + 2E_\lambda(\ln V_{11}^2 | T = t) - c_0 + \ln t - \phi_{ST}(t)\right]. \\ &\quad = [\phi_{ST}(t) - c_0 + \ln t] \\ &\qquad \times \left[2p \ln 2 + 2\sum_{i=2}^{p} \psi\left(\frac{n-i+1}{2}\right) + 2E_\lambda\left(\psi\left(\frac{n}{2} + K\right)\right) - c_0 + \ln t - \phi_{ST}(t)\right]. \end{aligned}$$

It is enough to establish that $\Delta_1(t, \lambda) \geqslant 0$, for each $t \in [0, 1]$ and each $\lambda > 0$, with strict inequality on a set of positive probability for some $\lambda > 0$. Clearly, for $\ln t \leqslant c_0 - p \ln 2 - \sum_{i=1}^{p} \psi(\frac{n-i+1}{2})$ and any $\lambda > 0$, $\Delta_1(t, \lambda) = 0$. Now suppose that $\ln t > c_0 - p \ln 2 - \sum_{i=1}^{p} \psi(\frac{n-i+1}{2}) = -\sum_{i=1}^{p} \{\psi(\frac{n-i+1}{2}) - \psi(\frac{n-i}{2})\}$. Then, since $\psi(x)$ is an increasing function of $x\in(0,\infty)$ [1], it follows that

$$\begin{aligned}\Delta_1(t,\lambda) = & \left[\ln t+\sum_{i=1}^{p}\left\{\psi\left(\frac{n-i+1}{2}\right)-\psi\left(\frac{n-i}{2}\right)\right\}\right] \\ & \times\left[2E_\lambda\left(\psi\left(\frac{n}{2}+K\right)\right)+\sum_{i=2}^{p}\psi\left(\frac{n-i+1}{2}\right)-\psi\left(\frac{n}{2}\right)\right. \\ & \left. -\sum_{i=1}^{p}\psi\left(\frac{n-i}{2}\right)+\ln t\right] \\ \geqslant & \left[\ln t+\sum_{i=1}^{p}\left\{\psi\left(\frac{n-i+1}{2}\right)-\psi\left(\frac{n-i}{2}\right)\right\}\right]^2 \\ > & 0,\ \forall\ t\in[0,1] \text{ and } \forall\ \lambda>0.\end{aligned}$$

Since $P_\lambda(\ln T>-\sum_{i=1}^{p}\{\psi(\frac{n-i+1}{2})-\psi(\frac{n-i}{2})\})>0$, for each $\lambda>0$, the result follows.

The Stein-type estimator obtained in Theorem 3.1 is non-smooth and therefore seems to be inadmissible. In the search of a smooth estimator dominating the best affine equivariant estimator δ_{c_0}, in the following section, we will derive a Brewster-Zidek-type improvement over the estimator δ_{c_0}. $\square$

4. Improved brewster-zidek-type estimator

In this section, following the innovative idea of Brewster and Zidek [2], we will derive a smooth estimator that dominates the best scale equivariant estimator δ_{c_0}. To do so, we first consider estimators of the form

$$\delta^*_{d,r}(X,S)=\begin{cases}\ln|S|-c_0, & \text{for } 0\leqslant T\leqslant r,\\ \ln|S|-d, & \text{for } r<T\leqslant 1,\end{cases} \tag{4.1}$$

where $d\in(-\infty,\infty)$ and $r\in(0,1)$ are real constants.

Note that, for each $r\in(0,1)$, $\delta^*_{c_0,r}=\delta_{c_0}$. Let $I(\cdot)$ denote the indicator function. Then, for given $r\in(0,1)$ and $\lambda>0$, the risk function $R_\lambda(\delta^*_{d,r})$ is minimized at $d\equiv d_r(\lambda)$, where

$$d_r(\lambda)=\frac{E_\lambda\left(\ln\left(\frac{|S|}{|\Sigma|}\right)I(r<T\leqslant 1)\right)}{E_\lambda(I(r<T\leqslant 1))}. \tag{4.2}$$

In the following lemma, we will derive some properties of the function $d_r(\lambda),\ \lambda>0,\ r\in(0,1)$.

Lemma 4.1. (i) *Let* $\delta^*_{d,r}(.,.)$ *be as defined in* (4.1). *Then, for fixed* $r\in(0,1)$ *and* $\lambda>0$, $R_\lambda(\delta^*_{d,r})$ *is minimized at* $d=d_r(\lambda)$, *where* $d_r(\lambda)$ *is given by* (4.2).

(ii) *For fixed* $r \in (0,1)$, $\inf_{\lambda>0} d_r(\lambda) = d(r)$, *where*

$$d(r) = \frac{\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}-1}\{\ln t + p\ln 2 + \sum_{i=1}^{p}\psi(\frac{n-i+1}{2})\}\,dt}{\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}-1}\,dt}. \tag{4.3}$$

(iii) $d(r)$ *is an increasing function of* $r \in (0,1)$, *and*

$$c_0 < d(r) < p\ln 2 + \sum_{i=1}^{p}\psi\left(\frac{n-i+1}{2}\right), \quad \forall\ r \in (0,1).$$

Proof. We have already observed that the assertion (i) is true. For proving the assertion (ii), consider a fixed $r \in (0,1)$. Then, from (4.2) and under the notation of Section 3, we have

$$\begin{aligned} d_r(\lambda) &= \frac{E_\lambda\left(\ln\left(\frac{|S|}{|\Sigma|}\right)I(r<T\leqslant 1)\right)}{E_\lambda(I(r<T\leqslant 1))} \\ &= \frac{E_\lambda(\{\ln T + \sum_{i=1}^{p}\ln V_{ii}^2\}I(r<T\leqslant 1))}{E_\lambda(I(r<T\leqslant 1))} \\ &= \frac{E_\lambda(L_1(K))}{E_\lambda(L_2(K))}, \ \ \lambda>0, \end{aligned}$$

where K is a Poisson random variable with mean $\lambda/2>0$ and, for $k=1,2,\ldots,$

$$\begin{aligned} L_1(k) = {}& \left[\beta\left(\frac{n-p}{2},\frac{p}{2}+k\right)\right]^{-1}\left[\int_r^1 \ln(t)t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-1}\,dt\right. \\ &\left. + \left\{\psi\left(\frac{n}{2}+k\right) + \sum_{i=2}^{p}\psi\left(\frac{n-i+1}{2}\right) + p\ln 2\right\}\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-1}\,dt\right], \\ L_2(k) = {}& \left[\beta\left(\frac{n-p}{2},\frac{p}{2}+k\right)\right]^{-1}\left[\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-1}\,dt\right]; \end{aligned}$$

here $\beta(\cdot,\cdot)$ denotes the beta function.

We may write

$$d_r(\lambda) = E_\lambda[L_3(K_1)], \quad \lambda>0,$$

where, for $k=0,1,\ldots, L_3(k) = L_1(k)/L_2(k)$ and K_1 is a random variable whose probability mass function is proportional to $\frac{\lambda^k L_2(k)}{2^k k!}, k=0,1,\ldots.$ We have

$$L_3(k) = L_4(k) + \psi\left(\frac{n}{2}+k\right) + \sum_{i=2}^{p}\psi\left(\frac{n-i+1}{2}\right) + p\ln 2, \ \ k=0,1,\ldots,$$

where

$$L_4(k) = \frac{\int_r^1 \ln(t)t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-1}\,dt}{\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-1}\,dt}, \ \ k=0,1,\ldots.$$

Integrating the integrals in the numerator and denominator above by parts with $\ln(t)(1-t)^{\frac{p}{2}+k-1}$ and $(1-t)^{\frac{p}{2}+k-1}$ as the differentiating factor, and after some standard adjustments, we get, for $k = 1, 2, \ldots,$

$$\begin{aligned}
&L_4(k)\\
&\quad = L_4(k-1)\left[\frac{p+2k-2-2\ln(r)r^{\frac{n-p}{2}}(1-r)^{\frac{p}{2}+k-1}\{\int_r^1 \ln(t)t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-2}\,dt\}^{-1}}{p+2k-2-2r^{\frac{n-p}{2}}(1-r)^{\frac{p}{2}+k-1}\{\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-2}\,dt\}^{-1}}\right]\\
&\qquad - \frac{2}{n+2k-2} > L_4(k-1) - \frac{2}{n+2k-2},
\end{aligned}$$

since

$$\int_r^1 \ln(t)t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-2}\,dt > \ln(r)\int_r^1 t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}+k-2}\,dt, \quad \forall\ r\in(0,1). \tag{4.4}$$

Now, using the fact that for $z>0$, $\psi(z+1) = \psi(z) + 1/z$ [1] we get, for $k = 1, 2, \ldots,$

$$\begin{aligned}
L_3(k) &\geqslant L_4(k-1) + \psi\left(\frac{n}{2}+k-1\right) + \sum_{i=2}^{p}\psi\left(\frac{n-i+1}{2}\right) + p\ln 2\\
&= L_3(k-1).
\end{aligned}$$

Thus, $L_3(k)$ is an increasing function of $k\in\{0, 1, \ldots\}$. Now, let K_2 be a random variable that is degenerate at 0. Then, for each $\lambda>0$, the random variable K_1 is stochastically larger than the random variable K_2, and therefore, for each $\lambda>0$,

$$\begin{aligned}
d_r(\lambda) &= E_\lambda[L_3(K_1)]\\
&\geqslant E_\lambda[L_3(K_2)]\\
&= L_3(0)\\
&= d(r),
\end{aligned}$$

where $d(r)$, $r>0$, is given by (4.3). Now the assertion follows by noting that, for each $r\in(0,1)$, $\lim_{\lambda\downarrow 0} d_r(\lambda) = d(r)$.

It remains to prove the assertion (iii). The increasing behavior of the function $d(r)$, $r\in(0,1)$, follows by direct differentiation and using (4.4) for $k = 1$. Therefore, for each $r\in(0,1)$,

$$\lim_{r\downarrow 0} d(r) < d(r) < \lim_{r\uparrow 1} d(r).$$

Using l'Hôspital's rule, it follows that $\lim_{r\uparrow 1} d(r) = p\ln 2 + \sum_{i=1}^{p}\psi(\frac{n-i+1}{2})$. Also, note that

$$\begin{aligned}
\int_0^1 \ln(t)t^{\frac{n-p}{2}-1}(1-t)^{\frac{p}{2}-1}\,dt &= \left[\frac{d}{dx}\beta(x,y)\right]_{x=\frac{n-p}{2},y=\frac{p}{2}}\\
&= [\beta(x,y)\{\psi(x)-\psi(x+y)\}]_{x=\frac{n-p}{2},y=\frac{p}{2}}\\
&= \beta\left(\frac{n-p}{2},\frac{p}{2}\right)\left[\psi\left(\frac{n-p}{2}\right)-\psi\left(\frac{n}{2}\right)\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{r\downarrow 0} d(r) &= p\ln 2 + \sum_{i=1}^{p} \psi\left(\frac{n-i+1}{2}\right) + \psi\left(\frac{n-p}{2}\right) - \psi\left(\frac{n}{2}\right)\\ &= c_0,\end{aligned}$$

and the result follows.

Since, for each $\lambda > 0$ and $r \in (0,1)$, $R_\lambda(\delta^*_{d,r})$ is minimized at $d = d_r(\lambda)$ and, for each fixed $r \in (0,1)$, $\inf_{\lambda>0} d_r(\lambda) = d(r)$, using the convexity of the function $R_\lambda(\delta^*_{d,r})$ (as a function of d, for fixed $\lambda > 0$ and $r \in (0,1)$), we conclude that, for each $\lambda > 0$ and $r \in (0,1)$, $R_\lambda(\delta^*_{d,r})$ is a strictly decreasing function of d, for $d \in (-\infty, d(r))$. Since, for each $r \in (0,1)$, we have $c_0 < d(r)$, the following result follows.

Theorem 4.1. *Let $d(r)$, $r>0$, be defined by* (4.3). *Then, for any $r \in (0,1)$, the estimator $\delta_{d(r),r}(\cdot,\cdot)$, as defined by* (4.1), *dominates the best affine equivariant estimator $\delta_{c_0}(\cdot,\cdot)$.*

Further select r', such that $0 < r < r' < 1$. Since $d(r)$ is an increasing function of $r \in (0,1)$, we have $c_0 < d(r) < d(r')$. Now, by considering estimators of the form

$$\delta^*_{d,r',r}(X,S) = \begin{cases} \ln|S| - c_0, & \text{for } 0 \leqslant T \leqslant r,\\ \ln|S| - d(r), & \text{for } r < T \leqslant r',\\ \ln|S| - d, & \text{for } r' < T \leqslant 1\end{cases}$$

and repeating the above arguments, it can be seen that the estimator $\delta_{d(r),r}(\cdot,\cdot)$ (and therefore the best affine equivariant estimator δ_{c_0}) is further dominated by the estimator

$$\delta^*_{d(r'),r',r}(X,S) = \begin{cases} \ln|S| - c_0, & \text{for } 0 \leqslant T \leqslant r,\\ \ln|S| - d(r), & \text{for } r < T \leqslant r',\\ \ln|S| - d(r'), & \text{for } r' < T \leqslant 1.\end{cases}$$

Now, using the idea of [2], we select a finite partition of the unit interval $[0,1]$, represented by $0 = r_{i,0} < r_{i,1} < \ldots < r_{i,n_i-1} < r_{i,n_i} = 1$, for each $i = 1,2,\ldots$, and a corresponding estimator defined by

$$\delta^*_i(X,S) = \begin{cases} \ln|S| - c_0, & \text{for } r_{i,0} \leqslant T \leqslant r_{i,1}\\ \ln|S| - d(r_{i,1}), & \text{for } r_{i,1} < T \leqslant r_{i,2}\\ \cdot & \cdot\\ \cdot & \cdot\\ \cdot & \cdot\\ \ln|S| - d(r_{i,n_i-1}), & \text{for } r_{i,n_i-1} < T \leqslant r_{i,n_i}.\end{cases}$$

Then, providing $\max_{1\leqslant j\leqslant n_i} |r_{i,j} - r_{i,j-1}| \to 0$, as $i \to \infty$, the sequence of estimators $\delta_i^*(X, S)$ converge point wise to

$$\begin{aligned}\delta_{\phi_{BZ}}(X,S) &= \ln|S| - d(T)\\ &= \ln|S+XX^t| - \phi_{BZ}(T),\end{aligned} \tag{4.5}$$

where the function $d(x)$, $x \in [0,1]$ is given by (4.3) and $\phi_{BZ}(t) = d(t) - \ln t$, $t \in [0,1]$.

Now we have the following theorem.

Theorem 4.2. *Let* $d(x)$, $x \in [0,1]$, *be defined by* (4.3). *Then, under the quadratic loss function* (1.2), *the best affine equivariant estimator* δ_{c_0} *is dominated by the estimator* $\delta_{\phi_{BZ}}$, *given by* (4.5).

Proof. Since, for each $i \in \{1, 2, \ldots\}$, δ_i^* has smaller risk than that of δ_{c_0}, the result follows by an application of Fatou's lemma.

In the following theorem, we establish that the Brewster-Zidek-type estimator $\delta_{\phi_{BZ}}$ is also generalized Bayes with respect to a prior similar to the one considered by [15] for estimating the variance in multivariate normal distribution. $\square$

Theorem 4.3. *The Brewster-Zidek-type estimator* $\delta_{\phi_{BZ}}$ *is generalized Bayes with respect to the prior on* (μ, Σ) *with density*

$$\Pi_1(\mu,\Sigma) = \frac{1}{|\Sigma|^{\frac{p+2}{2}}} \int_0^\infty \frac{z^{\frac{p}{2}-1}}{1+z} e^{-\frac{n\mu^t\Sigma^{-1}\mu}{2}z}\, dz, \quad \mu \in \mathscr{R}^p, \quad \Sigma > 0.$$

Proof. The (formal) posterior density of (μ, Σ), given (X, S), is proportional to

$$\begin{aligned}&\frac{1}{|\Sigma|^{\frac{n+p+2}{2}}} e^{-\frac{1}{2}(X-\sqrt{n}\mu)^t\Sigma^{-1}(X-\sqrt{n}\mu)} e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}S)} \int_0^\infty \frac{z^{\frac{p}{2}-1}}{1+z} e^{-\frac{n\mu^t\Sigma^{-1}\mu}{2}z}\, dz\\ &\quad = \frac{1}{|\Sigma|^{\frac{n+p+2}{2}}} \int_0^\infty \frac{z^{\frac{p}{2}-1}}{1+z} e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}(S+\frac{z}{1+z}XX^t))} e^{-\frac{n(1+z)}{2}(\mu-\frac{X}{\sqrt{n}(1+z)})^t\Sigma^{-1}(\mu-\frac{X}{\sqrt{n}(1+z)})}\, dz.\end{aligned}$$

Thus, the posterior distribution of Σ, given (X, S), is proportional to

$$\frac{1}{|\Sigma|^{\frac{n+p+1}{2}}} \int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}} e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}(S+\frac{z}{1+z}XX^t))}\, dz. \tag{4.6}$$

We know that the density of a $m \times m$ positive definite inverted Wishart matrix B, having N degrees of freedom and positive definite parameter matrix V, is given by

$$f_{N,m}(B|V) = \frac{|V|^{\frac{N-m-1}{2}}}{2^{\frac{m(N-m-1)}{2}}\Gamma_m(\frac{N-m-1}{2})|B|^{\frac{N}{2}}} e^{-\frac{1}{2}\mathrm{tr}(B^{-1}V)}, \quad B>0. \tag{4.7}$$

We write that B is $W_m^{-1}(N, V)$.

Thus, the normalizing factor in (4.6) is

$$K_1(X,S) = \left[\Gamma_p\left(\frac{n}{2}\right)2^{\frac{np}{2}}\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}\left|S + \frac{z}{1+z}XX^t\right|^{-\frac{n}{2}} dz\right]^{-1}$$

and therefore the posterior density of Σ, given (X,S), is

$$q(\Sigma|X,S) = \frac{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}\left|S + \frac{z}{1+z}XX^t\right|^{-\frac{n}{2}} f_{n+p+1,p}(\Sigma|S + \frac{z}{1+z}XX^t)\, dz}{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}|S + \frac{z}{1+z}XX^t|^{-\frac{n}{2}}\, dz},$$

where $f_{n+p+1,p}(\cdot|\cdot)$ is as defined in (4.7).

Thus, the generalized Bayes estimator of $\ln|\Sigma|$ is given by

$$\begin{aligned}\delta_{GB2}(X,S) &= E_\Sigma(\ln|\Sigma||X,S)\\ &= \frac{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}(1+\frac{z}{1+z}X^tS^{-1}X)^{-\frac{n}{2}}E(\ln|B_z|)\, dz}{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}(1+\frac{z}{1+z}X^tS^{-1}X)^{-\frac{n}{2}}\, dz},\end{aligned}$$

where, for given (X,S), B_z is $W_p^{-1}(n+p+1, S+\frac{z}{1+z}XX^t)$. It follows that

$$\begin{aligned}&B_z^{-1} \sim W_p\left(n, \left(S + \frac{z}{1+z}XX^t\right)^{-1}\right)\\ &\Rightarrow \left|\frac{B_z^{-1}}{\left(S + \frac{z}{1+z}XX^t\right)^{-1}}\right| \sim \prod_{i=1}^p \chi^2_{n-i+1},\end{aligned}$$

where the χ^2_{n-i+1}, for $i = 1, \ldots, p$, are independent chi-square random variables. Thus,

$$\begin{aligned}E(\ln|B_z|) &= \ln\left|S + \frac{z}{1+z}XX^t\right| - \sum_{i=1}^p E(\ln\chi^2_{n-i+1})\\ &= \ln|S| + \ln\left(1 + \frac{z}{1+z}X^tS^{-1}X\right) - p\ln 2 - \sum_{i=1}^p \psi\left(\frac{n-i+1}{2}\right)\end{aligned}$$

and therefore

$$\delta_{GB2}(X,S) = \ln|S| - \frac{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}(1+\frac{z}{1+z}X^tS^{-1}X)^{-\frac{n}{2}}\{p\ln 2 + \sum_{i=1}^{p}\psi(\frac{n-i+1}{2}) - \ln(1+\frac{z}{1+z}X^tS^{-1}X)\}\,dz}{\int_0^\infty \frac{z^{\frac{p}{2}-1}}{(1+z)^{\frac{p+2}{2}}}(1+\frac{z}{1+z}X^tS^{-1}X)^{-\frac{n}{2}}\,dz}.$$

On making the transformation $(1+\frac{z}{1+z}X^tS^{-1}X)^{-1} = t$ in the above integrals and on using the fact that $(1+X^tS^{-1}X)^{-1} = T$, the above expression turns out to be same as $\delta_{\phi_{BZ}}(X,S)$.

Remark 4.1. We could not resolve the question of admissibility/inadmissibility of the Brewster-Zidek-type estimator. For various combinations of n,p and λ, we compared the performances of estimators δ_{c_0}, $\delta_{\phi_{ST}}$ and $\delta_{\phi_{BZ}}$ using Monte Carlo simulations. We observed that there is not much difference in the values of their risks. In most cases, the differences in the values of their risks occur in second or third decimal places. The estimators $\delta_{\phi_{ST}}$ and $\delta_{\phi_{BZ}}$ provide less than 6% improvements over the best affine equivariant estimator δ_{c_0}. These simulation results are consistent and trivial, and therefore are not reported. Since the best affine equivariant estimator δ_{c_0} is simple to evaluate and is also unbiased, we recommend its use in applications.

Acknowledgments

The authors are grateful to the editor and the referee for their valuable comments which improved the quality of the paper.

References

[1] M. Abramowitz, A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.
[2] J.F. Brewster, J.V. Zidek, Improving on equivariant estimators, Ann. Statist. 2 (1974) 21–38.
[3] L.D. Brown, On the admissibility of invariant estimators of one or more location parameters, Ann. Math. Statist. 37 (1966) 1087–1136.
[4] N. Go, H.A. Scheraga, Analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules, J. Chem. Phys. 51 (1969) 4751–4767.
[5] S.A. Harris, E. Gavathiotis, M.S. Searle, M. Orozco, C.A. Laughton, Cooperativity in drug-DNA recognition: a molecular dynamics study, J. Amer. Chem. Soc. 123 (50) (2001) 12658–12663.
[6] B. Heymann, H. Grubmuller, Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces, Biophys. J. 81 (3) (2001) 1295–1313.
[7] S. Jusuf, P.J. Loll, P.H. Axelsen, The role of configurational entropy in biochemical cooperativity, J. Amer. Chem. Soc. 124 (14) (2002) 3490–3491.

[8] M. Karplus, J.N. Kushik, Method for estimating the configurational entropy of macromolecules, Macromolecules 14 (1981) 325–332.
[9] T. Kubokawa, Y. Konno, Estimating the covariance matrix and the generalized variance under a symmetric loss, Ann. Inst. Statist. Math. 42 (1990) 331–343.
[10] L.D. Landau, E.M. Lifshitz, Statistical Physics, Part 1, 3rd Edition, Butterworth-Heinemann, Oxford, 1980.
[11] R.M. Levy, M. Karplus, J. Kushik, D. Perahia, Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix, Macromolecules 17 (1984) 1370–1374.
[12] R.J. Muirhead, Aspects Of Multivariate Statistical Theory, Wiley, New York, 1982.
[13] N. Pal, Decision-theoretic estimation of generalized variance and generalized precision, Comm. Statist. Theor. Methods 17 (1988) 4221–4230.
[14] N. Pal, Estimation of generalized variance under entropy losses: admissibility results, Calcutta Statist. Assoc. Bull. 38 (1989) 147–156.
[15] A.L. Rukhin, M.M.A. Ananda, Risk behavior of variance estimators in multivariate normal distribution, Statist. Probab. Lett. 13 (1992) 159–166.
[16] H. Schafer, L.J. Smith, A.E. Mark, W.F. van Gunsteren, Entropy calculations on the molten goluble state of a protein: Side-chain entropies of alpha-lactalbumin, Proteins-Structure Funct. Genet. 46 (2) (2002) 215–224.
[17] R.W. Shorrock, J.V. Zidek, An improved estimator of the generalized variance, Ann. Statist. 4 (1976) 629–638.
[18] B.K. Sinha, On improved estimators of the generalized variance, J. Mult. Anal. 6 (1976) 617–625.
[19] B.K. Sinha, M. Ghosh, Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix and the generalized variance under entropy loss, Statist. Decisions 5 (1987) 201–227.
[20] C. Stein, Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Ann. Inst. Statist. Math. 16 (1964) 155–160.
[21] N. Sugiura, Y. Konno, Risk of improved estimators of generalized variance and precision, in: A.K. Gupta (Ed.), Advances in Multivariate Statistical Analysis, Riedel, Holland, 1987, pp. 353–371.
[22] S. Xiaoqian, P. Wankai, Improved estimation of the generalizedprecision under the entropy loss, Acta Math. Appl. Sinica 16 (2000) 162–170.