MONOTONIC MINIMAX ESTIMATORS OF A 2×2 COVARIANCE MATRIX

by

Francois Perron
University of Montreal

Technical Report #90-15

Department of Statistics
Purdue University

March 1990

*

Monotonic Minimax Estimators of a 2×2 Covariance Matrix

Francois Perron
University of Montreal

Abstract

Let $S : 2 \times 2$ have a nonsingular Wishart distribution with unknown matrix Σ and n degrees of freedom. For estimating Σ two families of minimax estimators, with respect to the entropy loss, are presented. These estimators are of the form $\hat{\Sigma}(S) = R\Phi(L)R^t$ where R is orthogonal, L and Φ are diagonal and $RLR^t = S$. Conditions under which the components of Φ and L follow the same order relation are established (i.e. writing $\Phi =$ diag (φ_1, φ_2) and $L =$ diag (ℓ_1, ℓ_2) with $\ell_1 \geq \ell_2$ we have $\varphi_1 \geq \varphi_2$).

Key words: minimax, admissible, group, orthogonally equivariant, risk, dominate.

1. Introduction.

In this paper we consider minimax estimators of the covariance matrix Σ of a bivariate normal population using the entropy loss $\mathcal{L}(\mathcal{L}(\Sigma,\hat{\Sigma}) = \mathrm{tr}(\Sigma^{-1}\hat{\Sigma}) - \log(\det(\Sigma^{-1}\hat{\Sigma})) - 2)$. The estimators we introduce are orthogonally equivariant and are based on a statistic S which has a Wishart distribution with parameter Σ and n degrees of freedom ($S \sim W_2(\Sigma, n)$). As orthogonally equivariant our estimators are of the form

$$\hat{\Sigma}(S) = R\Phi(\ell)R^t \tag{1.1}$$

where R is orthogonal, $\ell = (\ell_1, \ell_2)^t$, $\ell_1 \geq \ell_2 > 0$, $\Phi = \mathrm{diag}(\varphi)$ and $S = R\ \mathrm{diag}(\ell)R^t$.

Finding orthogonally equivariant estimators for Σ and estimating the eigenvalues $(\lambda_1, \lambda_2;\ \lambda_1 \geq \lambda_2 > 0)$ of Σ are two closely related problems. Muirhead (1987) proposed estimating λ_i by φ_i, $i = 1, 2$. A reasonable condition on φ_1, φ_2 is to impose $\varphi_1 \geq \varphi_2$. Such estimators will be called monotonic. When an estimator is not monotonic Stein suggests modifying the estimate using isotonic regression. This modification is described in detail in Lin and Perlman (1985).

The best equivariant estimator with respect to the group of lower triangular matrices with positive diagonal elements (G_T^+) is minimax and has constant risk. This estimator has the form

$$\hat{\Sigma}_T(S) = TDT^t$$

where $T \in G_T^+$, $TT^t = S$, $D = \mathrm{diag}(d)$ and $d^t = (d_1, d_2)$ with $d_1 = (n+1)^{-1}$ and $d_2 = (n-1)^{-1}$. The minimax risk being

$$\mathcal{R}(I, \hat{\Sigma}_T) = -\log(d_1 d_2) - E(\log(\chi^2_{n-1}) + \log(\chi^2_n)) \tag{1.2}$$

(James and Stein 1961). Averaging $\hat{\Sigma}_T$ over the orthogonal group (cf Sharma and Krishnamoorthy 1983 or Takemura 1984) we get the monotonic minimax estimator $\hat{\Sigma}_0$. This estimator dominates $\hat{\Sigma}_T$ and, referring to expression (1.1), is given by

$$\varphi_i(\ell) = \ell_i\{w(\ell)d_i + (1 - w(\ell))(2d_0 - d_i) \tag{1.3}$$

$$\text{with} \qquad w(\ell) = \sqrt{\ell_1}/(\sqrt{\ell_1} + \sqrt{\ell_2}) \tag{1.4}$$

and $d_0 = (d_1 + d_2)/2$. The risk being substantially reduced in a neighborhood of the identity matrix.

Intuitively, keeping φ_1 and φ_2 close together lead to a small risk when Σ is a multiple of the identity matrix. For this reason, one might be interested in having flexibility on the choice of φ_1 and φ_2 within the class of monotonic minimax estimators. The purpose of this article is to provide two classes of monotonic minimax estimators ($\mathcal{C}$ and $\mathcal{D}$). The estimators are obtained by modifying the definition of w in expression (1.4). The elements of $\mathcal{C}$ are characterized by a function $h: \mathbb{R}_+ \to \mathbb{R}_+$ and are given by expressions (1.1) and (1.3) with

$$w(\ell) = h(\ell_1)/(h(\ell_1) + h(\ell_2)) \tag{1.5}$$

and denoted $\hat{\Sigma}^h$. Similarly, the elements of $\mathcal{D}$ are characterized by a function $h: \mathbb{R}_+ \to (-1, 1)$ with

$$w(\ell) = (1 + h(\ell_2/\ell_1))/2 \tag{1.6}$$

and are denoted $\tilde{\Sigma}^h$. Monotonicity and minimaxity properties of the elements of $\mathcal{C}$ and $\mathcal{D}$ are studied in section 2 and 3 respectively. These properties are proven by solving differential inequalities (cf Efron and Morris 1976 for an example). In order to prove minimaxity, Stein's technique is applied. In this technique, an unbiased estimator of the risk of an orthogonal equivariant estimator is used. This estimator involves the functions φ_i and their derivatives.

2. Properties of $\mathcal{C}$.

In the first section, we introduced the class $\mathcal{C}$ along with the notation φ_i. This class is an extension of $\hat{\Sigma}_0$. Let φ_i^h be determined by expressions (1.3) and (1.5) and set $\varphi_i^h = \ell_i \Psi_i^h$ for $i = 1, 2$. For any function $h: \mathbb{R}_+ \to \mathbb{R}_+$ the point d_0 is a middle point between Ψ_1^h and Ψ_2^h and the range of Ψ_1^h is included in (d_1, d_2). The function h is a parameter indicating how near Ψ_1^h is to d_1. When h is nondecreasing, $\Psi_1^h \le d_0$ and when h is nonincreasing, $\Psi_1^h \ge d_0$. Roughly speaking, the greater h is increasing, the more Ψ_1^h is near to d_1. An intermediate case being $\Psi_1^h = d_0$ which corresponds to a function h which is flat. The limiting cases are $\Psi_1^h = d_1$ and $\Psi_1^h = d_2$. The case $\Psi_1^h = d_1$ has been proposed by Stein in a series of lectures given at the University of Washington, Seattle 1982. This estimator is minimax (cf Dey and Srinivasan 1985).

Proposition 2.1. If h_1 and g are two functions from R_+ to R_+, $h_2(x) = h_1(x)g(x)$ and g is nondecreasing then $\Psi_1^{h_2} \le \Psi_1^{h_1}$.

Proof. From expressions (1.4) and (1.5) we have $\Psi_1^{h_j} = w_j d_1 + (1-w_j)d_2$ with $w_j = (1 + h_j(\ell_2)/h_j(\ell_1))^{-1}$. Since $d_1 < d_2$ and $\ell_1 \ge \ell_2 > 0$ we get

$$w_2(\ell) = \left(1 + \frac{h_1(\ell_2)}{h_1(\ell_1)}\frac{g(\ell_2)}{g(\ell_1)}\right)^{-1} \ge \left(1 + \frac{h_1(\ell_2)}{h_1(\ell_1)}\right)^{-1} = w_1(\ell)$$

and $\Psi_{1^2}^{h} \le \Psi_{1^1}^{h}$. QED.

Theorem 2.1. (monotonicity property). The relation $(\varphi_1^h(\ell) - \varphi_2^h(\ell))/(\ell_1 - \ell_2) > 0$ holds for all $\ell_1 > \ell_2 > 0$ if and only if $h(x) = x^n r(x)$ for some differentiable, nonincreasing function $r\colon \mathsf{R}_+ \to \mathsf{R}_+$.

Proof. Define the function $r\colon \mathsf{R}_+ \to \mathsf{R}_+$ by $r(x) = x^{-n}h(x)$.
After computations we get

$$(\varphi_1^h(\ell) - \varphi_2^h(\ell))/(\ell_1 - \ell_2) = (n - g(\ell))/(n^2 - 1) \tag{2.1}$$

with $g(\ell) = (\ell_1 + \ell_2)(h(\ell_1) - h(\ell_2))/\{(h(\ell_1) + h(\ell_2))(\ell_1 - \ell_2)\}$. In order to complete the proof we shall prove the necessity and the sufficiency parts separately.

(**Necessity**). If $(\varphi_1^h(\ell) - \varphi_2^h(\ell))/(\ell_1 - \ell_2) > 0$ holds for all $\ell_1 > \ell_2 > 0$, then $\lim_{\ell_2 \to \ell_1} n - g(\ell) = n - \frac{h'(\ell_1)}{h(\ell_1)}\ell_1 = \ell_1 \frac{d}{d\ell_1}\log(\ell_1^n / h(\ell_1)) \ge 0$ for all $\ell_1 > 0$. Therefore $h(x) = x^n r(x)$ where $r\colon \mathsf{R}_+ \to \mathsf{R}_+$ is nondecreasing.

(**Sufficiency**). If $h(x) = x^n r(x)$ where $r\colon \mathsf{R}_+ \to \mathsf{R}_+$ is nonincreasing then

$$\begin{aligned} n - g(\ell) &= n - \frac{(\ell_1 + \ell_2)}{(\ell_1 - \ell_2)}\left\{\frac{2\ell_1^n \ell_2^n (r(\ell_1) - r(\ell_2))}{(\ell_1^n + \ell_2^n)(\ell_1^n r(\ell_1) + \ell_2^n r(\ell_2))} + \frac{(\ell_1^n - \ell_2^n)}{(\ell_1^n + \ell_2^n)}\right\} \\ &\ge n - \frac{(\ell_1 + \ell_2)}{(\ell_1 - \ell_2)}\frac{(\ell_1^n - \ell_2^n)}{(\ell_1^n + \ell_2^n)} \end{aligned}$$

for all $\ell_1 > \ell_2 > 0$. This inequality combined with expression (2.1) indicates that it is sufficient to consider $h(x) = x^n$ in order to complete the proof of the sufficiency part. If $h(x) = x^n$ then

$$(h(\ell_1) + h(\ell_2))(\ell_1 - \ell_2)(n - g(\ell)) = (n-1)(\ell_1^{n+1} - \ell_2^{n+1}) + (n+1)(\ell_1 \ell_2^n - \ell_1^n \ell_2)$$
$$= u(\ell) (\text{say}).$$

It remains to show that $u(\ell) > 0$ if $\ell_1 > \ell_2 > 0$. First, notice that $u(\ell) = 0$ if $\ell_1 = \ell_2 > 0$. Secondly, notice that

$$\frac{\partial}{\partial \ell_1} u(\ell) = (n+1)\ell_2^n \left[1 + (n-1)(\ell_1/\ell_2)^n - n(\ell_1/\ell_2)^{n-1}\right] = (n+1)\ell_2^n t(\ell_1/\ell_2) (\text{say})$$

with $t(1) = 0$ and $t'(x) = n(n-1)x^{n-2}(x-1) > 0$ for $x > 1$. Therefore t is positive on $\{x : x > 1\}$, u is positive for $\ell_1 > \ell_2 > 0$ and $\varphi_1(\ell) > \varphi_2(\ell)$ for $\ell_1 > \ell_2 > 0$. QED.

In order to prove the minimax property we shall use an estimate $\hat{\mathcal{R}}(\Sigma, \hat{\Sigma})$ of $\mathcal{R}(\Sigma, \hat{\Sigma})$ for an orthogonally equivariant estimator $\hat{\Sigma}$ given in the form of expression (1.1). We also define the function α as

$$\alpha(\Sigma, \hat{\Sigma}) = \mathcal{R}(I, \hat{\Sigma}_T) - \hat{\mathcal{R}}(\Sigma, \hat{\Sigma}) \tag{2.2}$$

where $\mathcal{R}(I, \hat{\Sigma}_T)$ is given by expression (1.3). For $\hat{\Sigma}$ fixed, $\hat{\Sigma}$ an orthogonally equivariant estimator of $\Sigma, \alpha(\Sigma, \hat{\Sigma})$ is a function depending on ℓ only. Therefore, an orthogonally equivariant estimator $\hat{\Sigma}$ is minimax if $\alpha(\Sigma, \hat{\Sigma}) \geq 0$.

Lemma 2.1. (Stein 1977). If $\hat{\Sigma}$ is an orthogonally equivariant estimator of Σ then an unbiased estimator $\hat{\mathcal{R}}(\Sigma, \hat{\Sigma})$ of $\mathcal{R}(\Sigma, \hat{\Sigma})$ is given by

$$\hat{\mathcal{R}}(\Sigma, \hat{\Sigma}) = (n-1)(\Psi_1 + \Psi_2) + 2(\ell_1 \Psi_1 - \ell_2 \Psi_2)/(\ell_1 - \ell_2) - 2$$

$$+2(\ell_1 \frac{\partial}{\partial \ell_1} \Psi_1(\ell) + \ell_2 \frac{\partial}{\partial \ell_2} \Psi_2(\ell)) - \log(\Psi_1 \Psi_2) - E\left(\log(\chi^2_{n-1}) + \log(\chi^2_n)\right).$$

Let Δ be the function defined by

$$\Delta(h, \ell) = w(1-w)\left(\ell_1 \frac{h'(\ell_1)}{h(\ell_1)} + \ell_2 \frac{h'(\ell_2)}{h(\ell_2)}\right) - \left(\frac{(1-w)\ell_1 - w\ell_2}{\ell_1 - \ell_2}\right).$$

Theorem 2.2. If $\Delta(h,\ell)+\left(\frac{n^2-1}{4}\right)\log\left(1+\frac{4w(1-w)}{(n^2-1)}\right)\geq 0$ for all $\ell_1>\ell_2>0$ then $\hat{\Sigma}^h$ is minimax.

Proof. Computations give $\alpha(\Sigma,\hat{\Sigma}^h)(\ell)=\frac{4}{n^2-1}\Delta(h,\ell)+\log\left(1+\frac{4w(1-w)}{(n^2-1)}\right)\geq 0$ for all $\ell_1>\ell_2>0$ by assumptions. QED.

Corollary 2.1. If $\Delta(h,\ell)\geq 0$ for all $\ell_1>\ell_2>0$ then $\hat{\Sigma}^h$ is minimax.

Theorem 2.3. We have $\Delta(h,\ell)\geq 0$ for all $\ell_1>\ell_2>0$ if and only if $h(x)=\sqrt{x}v(x)$ where $v\colon \mathbb{R}_+\to\mathbb{R}_+$ is differentiable and nondecreasing.

Proof. Let $h(x)=\sqrt{x}v(x)$. Computations give

$$\Delta(h,\ell)=\frac{\sqrt{\ell_1\ell_2}}{(h(\ell_1)+h(\ell_2))^2}\left[\ell_1v'(\ell_1)v(\ell_2)+\ell_2v'(\ell_2)v(\ell_1)+\sqrt{\ell_1\ell_2}\,\frac{(v^2(\ell_1)-v^2(\ell_2))}{\ell_1-\ell_2}\right]$$

It is clear that if $v\colon\mathbb{R}_+\to\mathbb{R}_+$ is nondecreasing then $\Delta(h,\ell)\geq 0$ for all $\ell_1>\ell_2>0$. On the other hand if $\Delta(h,\ell)\geq 0$ for all $\ell_1>\ell_2>0$ then $\lim\limits_{\ell_2\to\ell_1}\Delta(h,\ell)=\ell_1\frac{v'(\ell_1)}{v(\ell_1)}\geq 0$ for all $\ell_1\geq 0$ which implies that v is nondecreasing. QED.

3. Properties of $\mathcal{D}$.

The class $\mathcal{D}$ is another extension of $\hat{\Sigma}_0$. In particular $\hat{\Sigma}_0=\tilde{\Sigma}^h$ for $h(x)=(1+\sqrt{x})/(1-\sqrt{x})$. More generally $\hat{\Sigma}^h=\tilde{\Sigma}^g$ for $h(x)=x^\alpha$ and $g(x)=(1-x^\alpha)/(1+x^\alpha)$. However $\mathcal{C}\not\subset\mathcal{D}$ and $\mathcal{D}\not\subset\mathcal{C}$. As before let $\varphi_i^h=\ell_i\Psi_i^h$ where φ_i^h are now given by expressions (1.3) and (1.6), $i=1,2$ and let $x=\ell_2/\ell_1$. For any function $h\colon(0,1)\to(-1.1)$, d_0 is the middle point between Ψ_1^h and Ψ_2^h and the range of Ψ_1^h is included in (d_1,d_2). When $h(x)$ is positive $\Psi_1^h(\ell)>d_0$ and when $h(x)$ is negative $\Psi_1^h(\ell)<d_0$. Having $h(x)=0$ implies $\Psi_1^h(\ell)=\Psi_2^h(\ell)$. The nearer h is to 1 the nearer Ψ_1^h is to d_1. The limiting cases are $\Psi_1^h=d_1,d_2$ corresponding to $h=1,\ -1$ respectively. Finally a useful expression for Ψ_1^h is

$$\Psi_1^h(\ell)=d_0-h(x)/(n^2-1).$$

Theorem 3.1. (monotonicity property). The relation $(\varphi_1^h(\ell)-\varphi_2^h(\ell))/(\ell_1-\ell_2)>0$ holds for all $\ell_1>\ell_2>0$ if and only if $h(x)<n(1-x)/(1+x)$ for all $x\in(0,1)$.

Proof. $(\varphi_1^h(\ell) - \varphi_2^h(\ell))/(\ell_1 - \ell_2) = (n^2-1)^{-1}\{n - h(x)(1+x)/(1-x)\} \geq 0$ for all $\ell_1 > \ell_2 > 0$ by assumptions. QED.

Let Δ be the function defined by

$$\Delta(h,x) = h(x)(1+x)/(1-x) - 1 - 2xh'(x).$$

Theorem 3.2. If $2\Delta(h,x) + \log(1 + (1-h^2(x))/(n^2-1)) \geq 0$ for all $x \in (0,1)$ then $\tilde{\Sigma}^h$ is minimax.

Proof. Referring to expression (2.2) computations give

$$\alpha(\Sigma, \tilde{\Sigma}^h)(\ell) = (n^2-1)^{-1}(2\Delta(h,x) + \log(1 + (1-h^2(x))/(n^2-1))) \geq 0 \text{ for all } x \in (0,1)$$

by assumptions which implies that $\alpha(\Sigma, \tilde{\Sigma}^h)(\ell) \geq 0$ for all $\ell_1 > \ell_2 > 0$. QED.

Corollary 3.1. If $\Delta(h,x) \geq 0$ for all $x \in (0,1)$, then $\tilde{\Sigma}^h$ is minimax.

Theorem 3.3. The inequality $\Delta(h,x) \geq 0$ holds for all $x \in (0,1)$ if and only if $h(x) = (1-\sqrt{x})/(1+\sqrt{x}) + v(x)\sqrt{x}/(1-x)$ where v is differentiable and nonincreasing on $(0,1)$.

Proof. Let $h(x) = (1-\sqrt{x})/(1+\sqrt{x}) + v(x)\sqrt{x}/(1-x)$. Computations give $\Delta(h,x) = -2x\sqrt{x}v'(x)/(1-x) \geq 0$ for all $x \in (0,1)$ by assumptions. QED.

References

Dey, D.K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, *Ann. Statist.* **13**, 1581–1591.

Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution, *Ann. Statist.* **4**, 11–21.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 361–380, Univ. of California Press.

Lin, S.P. and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. *Multivariate Analysis*, **6**, 411–429, ed. P.R. Krishnaiah North Holland, Amsterdam.

Muirhead, R.J. (1987). Developments in eigenvalue estimation. *Advances in Multivariate Statistical Analysis*, 277–288, ed. A.K. Gupta, D. Reidel, Boston.

Sharma, D. and Krishnamoorthy, K. (1983). Orthogonal equivariant estimators of bivariate normal covariance matrix and precision matrix. *Calcutta Statistical Association Bulletin*, **32**, 23–45.

Stein, C. (1977). Lectures on the theory of estimation of many parameters (In Russian). Studies in the Statistical Theory of Estimation, Part I (Ibragimov, I.A. and Nikulin, M.S., eds). *Proceeding of Scientifics Seminars of the Steklov Institute*, Leningrad Division **74**, 4–65.

Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsubuka J. Math.* **8**, 367–376.

Department of Mathematics and Statistics
University of Montreal
C.P. 6128, succ. A
Montreal, (Quebec) Canada
H3C 3J7