TESTING OPTIMALITY OF EXPERIMENTAL DESIGNS FOR A REGRESSION MODEL WITH RANDOM VARIABLES

by

Rameshwar Gupta and Donald Richards

University of New Brunswick, St. John
and
University of North Carolina, Chapel Hill

Key Words and Phrases: Optimality in the mean, multivariate normal, left-spherical, optimality criteria, zonal polynomial, hypergeometric function of matrix argument.

AMS 1980 Subject Classification: Primary 62K05, Secondary 62H10

ABSTRACT

Tsukanov (*Theor. Prob. Appl.* 26 (1981) 173-177) considers the regression model: $E(y|Z) = Fp + Zq$, $D(y|Z) = \sigma^2 I_n$, where $y(n\times 1)$ is a vector of measured values, $F(n\times k)$ contains the control variables, $Z(n\times \ell)$ contains the observed values, and $p(k\times 1)$ and $q(\ell\times 1)$ are vectors being estimated. Assuming that $Z = FL + R$, where $L(k\times \ell)$ is non-random, and the rows of $R(n\times \ell)$ are i.i.d. $N(0,\Sigma)$, we extend Tsukanov's results in three ways by (i) computing $E(\det H_p)$, where H_p is the covariance matrix of $\hat{p}$, the ℓ.s.e of p, (ii) considering "optimality in the mean" for the largest root criterion, (iii) considering these questions when the matrix R has a left-spherical distribution.

1. INTRODUCTION

In an experiment, let x be a vector of controlled variables which identify certain conditions associated with the experiment. Also, let $f_1(x),\ldots,f_k(x)$ be a prescribed set of linearly independent functions. In n independent repetitions of the experiment, let F be the $n \times k$ array of values of $f(x) = [f_1(x), f_2(x),\ldots,f_k(x)]$, and Z be the $n \times \ell$ matrix of outcomes of the $\ell \times 1$ random vector z. Initially, we will assume that $z \sim N_\ell(\mu,\Sigma)$ where the covariance matrix Σ is positive definite.

Recently, Tsukanov (1981) has discussed certain optimality questions pertaining to the regression model

$$E(y|Z) = Fp + Zq, \qquad Cov(y|Z) = \sigma^2 I_n . \tag{1}$$

Here, $y(n\times 1)$ contains the values observed in the n trials, while $p(k\times 1)$ and $q(\ell\times 1)$ contain the coefficients to be estimated. Specifically, Tsukanov considers optimality tests for the model (1) using A- and D-optimality criteria applied to the conditional covariance matrices H, H_p and H_q of the least squares estimates of the vectors $[p^T q^T]$, p^T and q^T respectively:

$$H = \left[\begin{array}{c|c} H_p & \\ \hline & H_q \end{array}\right] = \left[\begin{array}{c|c} F^TF & F^TZ \\ \hline Z^TF & Z^TZ \end{array}\right]^{-1}$$

When z is related to f(x), we assume that

$$Z = FL + R , \tag{2}$$

where L $(k\times\ell)$ is non-random, and the matrix of residuals R has i.i.d. $N_\ell(0,\Sigma)$ rows.

A brief summary runs as follows; Section 2 closes a problem of Tsukanov concerning E det H_p, when the relation (2) holds, and Section 3 discusses E-optimality in the mean for the matrices H, H_p and H_q. Finally, Section 4 treats these problems under the weaker assumption that the matrix R is left-spherical (Dawid(1977)).

2. D-OPTIMALITY IN THE MEAN

Following Tsukanov (1981), we say that the designs which minimise the mean of the generalised variance of the coefficient estimators are *D-optimal in the mean.* Theorem 1 below completes Section 3 in Tsukanov (1981).

Theorem 1. Suppose that H_p *exists and* $n > k + \ell + 1$. *Then*

$$E \det H_p = \det[F^TF]^{-1} \prod_{i=1}^{\ell} \left\{\frac{n-i-1}{n-k-i-1}\right\} \sum_{i=0}^{\ell} \frac{tr_i \Sigma^{-1} L^T F^T F L}{(n-2)^{(i)}} . \quad (3)$$

Here, $tr_i\Omega$ denotes the i-th elementary symmetric function of Ω, and $a^{(i)} = a(a-1)\cdots(a-i+1)$, $i = 0,1,\ldots,$ is the *falling factorial.*

In the particular case $\ell=1$, (3) shows that

$$E \det H_p = \det[F^TF]^{-1}\left(\frac{n-2}{n-k-2} + \frac{L^TF^TFL}{(n-2)\Sigma}\right)$$

which is given as Theorem 3 in Tsukanov (1981). It is also somewhat surprising that (3) appears as a *finite* series.

Proof of Theorem 1. We may write

$$\det H_p = \det[F^TF]^{-1} \det[U+V]/\det U, \quad (4)$$

where $U = Z^TAZ$, $V = Z^TBZ$ and $A = I_n - B = I_n - F[F^TF]^{-1}F^T$ is idempotent of rank $r = n-k$. Since $AB = 0$, an extension of Craig's theorem (Mardia et al (1979), problem 3.4.20) shows that U and V are mutually independent.

As Tsukanov shows, $U \sim W_\ell(r,\Sigma)$. To find the distribution of V, choose an $n\times n$ orthogonal matrix K satisfying $K^TBK = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$, and set $Z = HY$. Then $V = Y^T\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}Y$. Further, the rows of Y are independent and normally distributed with covariance matrix Σ. Hence $V \sim W_\ell(k,\Sigma,\Omega)$, a non-central Wishart distribution. The matrix of non-centrality parameters is $\Omega = \Sigma^{-1}E(Y^T)E(Y) = \Sigma^{-1}L^TF^TFL$.

Thus, $\Lambda = (\det U)\det[U+V]^{-1}$ follows the non-central Wilks' Λ distribution, so that (Constantine (1963)),

$$E(\Lambda^{-1}) = \frac{\Gamma_\ell(\frac{1}{2}r-1)\Gamma_\ell(\frac{1}{2}n)}{\Gamma_\ell(\frac{1}{2}r)\Gamma_\ell(\frac{1}{2}n-1)}\,{}_1F_1(-1;\tfrac{1}{2}n-1;-\tfrac{1}{2}\Omega) \tag{5}$$

where $\Gamma_\ell(\cdot)$ is the multivariate gamma function and ${}_1F_1(\alpha;\beta;\Omega)$ is the confluent hypergeometric function of matrix argument. Since

$${}_1F_1(-1;\tfrac{1}{2}n;-\tfrac{1}{2}\Omega) = \sum_{i=0}^{\ell}(\mathrm{tr}_i\,\Omega)/n^{(i)} \quad ,$$

(Shah and Khatri (1974)), then in conjunction with (4) we obtain (3) from (5).

3. E-OPTIMALITY IN THE MEAN

Let us now suppose that the efficiency of the experiment is being estimated using the largest characteristic root, λ_{max}, of the covariance matrices H, H_p and H_q. This widely used test is referred to as E-optimality. In some instances, E-optimality is more easily shown than A- or D-optimality. However, here the expectations are more recondite for the E-optimal criterion. It is not surprising that the simplest results appear in terms of zonal polynomials.

Theorem 2. If $m = \frac{1}{2}(n-k-\ell-1)$ *is a positive integer, then*

$$E\lambda_{max}(H_q) = \mathrm{tr}\Sigma^{-1}/2m + \tfrac{1}{2}\sum_{j=m+1}^{m\ell} (j^{-1}-(j-1)^{-1})(\mathrm{tr}\Sigma^{-1})^{-j+1} \sum_{\kappa}^{*} C_{\kappa}(\Sigma^{-1}) . \quad (6)$$

Here, $C_{\kappa}(\cdot)$ is the zonal polynomial corresponding to the partition $\kappa = (\kappa_1,\kappa_2,\ldots,\kappa_{\ell})$ of the integer j; $\sum_{\kappa}^{*}$ denotes summation over all κ for which $\kappa_1 \leq m$.

Proof of Theorem 2. As shown in Tsukanov (1981), $H_q^{-1} = R^T AR \sim W_{\ell}(r,\Sigma)$, $r = n-k$. Since $\lambda_{max}(H_q) = 1/\lambda_{min}(H_q^{-1})$, (6) can be established in a straightforward manner using Khatri's (1972) representation for the distribution of $\lambda_{min}(H_q^{-1})$.

Before considering $\lambda_{max}(H_p)$, we first identify the distribution of H_p.

Theorem 3. If $n \geq k + \ell$ and $k \geq \ell$, then $[F^TF]^{\frac{1}{2}} H_p [F^TF]^{\frac{1}{2}} - I_k$ has a matric-variate non-central F-distribution.

Proof. From Tsukanov (1981), we may write

$$[F^TF]^{\frac{1}{2}} H_p [F^TF]^{\frac{1}{2}} - I_k = F_1 Z H_q Z^T F_1^T \quad , \quad (7)$$

where $F_1 = [F^TF]^{-\frac{1}{2}}F^T$. Since the k×n matrix F_1 is pseudo-orthogonal $(F_1F_1^T = I_k)$, then the rows of F_1Z are independent, and normally distributed, with common covariance matrix Σ. Further, Z is independent of H_q^{-1} since $F_1A = 0$. Therefore, $F_1ZH_qZ^TF_1^T$ has a non-central F-distribution (James (1964), p. 485). The non-centrality parameter is $\Omega = [F^TF]^{\frac{1}{2}}L\Sigma^{-1}L^T[F^TF]^{\frac{1}{2}}$.

In the simplest cases, $E\lambda_{max}(H_p)$ appears in terms of the zonal polynomials of Davis (1979). Simple bounds are provided by the following result.

Theorem 4. Suppose that $n \geq k + \ell$, $k \geq \ell$ and $E(z) = 0$. Then, there are constants c_1, c_2 such that

$$c_1\lambda_{max}([F^TF]^{-1}) \leq E\lambda_{max}(H_p) \leq c_2\lambda_{max}([F^TF]^{-1}). \quad (8)$$

Proof. Using (7), we have

$$\lambda_{max}(H_p) = \lambda_{max}([F^TF]^{-1}(I_k + F_1ZH_qZ^TF_1^T))$$

$$\leq \lambda_{max}([F^TF]^{-1})\ \lambda_{max}(I_k + F_1ZH_qZ^TF_1^T) ,$$

where the inequality is obtained from Anderson and Dasgupta (1963). Taking expectations obtains the right side of (8) with $c_2 = 1+E\lambda_{max}(F_1ZH_qZ^TF_1^T)$. The second inequality follows similarly. The exact value of c_1 and c_2 can be deduced from the results of Khatri (1967) but the expressions seem to be complicated.

No satisfactory results are available for $\lambda_{max}(H)$.

4. EXTENSION TO LEFT-SPHERICAL RESIDUALS

For the rest of the paper, $\Delta_1 \overset{d}{=} \Delta_2$ will mean that the random entities Δ_1 and Δ_2 have the same distribution.

We now consider testing for optimality when the matrix of residuals, R, is *left-spherical* (Dawid (1977)). That is, for any n×n orthogonal matrix K, $KR \overset{d}{=} R$. In order to obtain positive results, we shall assume - unless explicitly stated otherwise - that R has a denisty function which is proportional to $(\det \Sigma)^{-n/2}\, g(\Sigma^{-1}R^TR)$, for some symmetric function g. Non-normal variables in this class include certain matric-variate t distributions (Dickey (1967)). We also assume that all expectations discussed do exist; in general, the conditions needed depend on the function g.

4.1. MD-optimality

Theorem 5. Assume that E(z) = 0. Then,

(i) $E \det H_q \propto \det \Sigma^{-1}$,

(ii) $E \det H \propto \det[F^TF]^{-1} \det \Sigma^{-1}$,

(iii) $E \det H_p \propto \det[F^TF]^{-1}$.

Proof. (i) Recall that $H_q^{-1} = R^T AR$, where A is idempotent of rank r = n-k. Since R is left-spherical then for a suitable orthogonal matrix K,

$$H_q^{-1} \stackrel{d}{=} R^T K^T A K R = R^T \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} R = R_1^T R_1 ,$$

where $R_1 (r \times \ell)$ consists of the first r rows of R. It is not difficult to check that R_1 has a density proportional to $(\det \Sigma)^{-r/2} g(\Sigma^{-1} R_1^T R_1)$. Therefore,

$$E \det H_q \propto (\det \Sigma)^{-r/2} \int_{R_1} (\det R_1^T R_1)^{-1} g(\Sigma^{-1} R_1^T R_1) dR_1 \propto \det \Sigma^{-1}, \qquad (9)$$

with (9) following by transforming $R_1 \longrightarrow R_1 \Sigma^{\frac{1}{2}}$ in the integral.

(ii) This follows directly from (i) since $\det H = \det [F^T F]^{-1} \det H_q$.

(iii) We use the notation of the proof of Theorem 1. As shown in the proof of (i), $U = Z^T AZ \stackrel{d}{=} R_1^T R_1$. In a similar fashion, we may show that $V = Z^T BZ \stackrel{d}{=} R_2^T R_2$ where R_2 consists of the last k rows of R. Therefore,

$$\det H_p \stackrel{d}{=} \det [F^T F]^{-1} \det R^T R / \det R_1^T R_1$$

$$= \det [F^T F]^{-1} / \det [I + R_2 (R_1^T R_1)^{-1} R_2^T] .$$

Since the matrix $R_2 (R_1^T R_1)^{-1} R_2^T$ has a matric-variate F distribution (Dawid(1977)), then $\det H_p$ has the same distribution as under normality, so (iii) follows from Theorem 1.

Remarks. (i) As in the normal case, (i) and (ii) are valid if $E(z) \neq 0$. However, the constants of proportionality are usually quite complicated. (One exception is the case when R has density (Dickey (1967)) const. $(\det \Sigma)^{-n/2} \det(I+\Sigma^{-1} R^T R)^{-\alpha}$, $\alpha = (\nu + n + \ell - 1)/2$. Then,

$$E \det H_q = \det \Sigma^{-1} \prod_{i=1}^{\ell} \frac{\nu+\ell-i}{n-k-i-1} ,$$

which is quite similar to the normal case.)

(iii) Although Theorem 5 (iii) only holds for $E(z) = 0$, we should note that it is valid for arbitrary left-spherical R.

4.2. MA-*optimality*

Theorem 6. Assume that $E(z) = 0$. Then,

(i) $E \operatorname{tr} H_q \propto \operatorname{tr} \Sigma^{-1}$,

(ii) $E \operatorname{tr} H_p$ is the same as under normality.

Proof. We shall show that $E H_q \propto \Sigma^{-1}$. Let $S = \Sigma^{1/2} H_q \Sigma^{1/2} \overset{d}{=} \Sigma^{1/2}(R_1^T R_1)^{-1}\Sigma^{1/2}$. Then, for symmetric $\ell \times \ell$ $W = (w_{\alpha\beta})$,

$$E \exp(i \operatorname{tr} WH_q) \propto \int_S \exp(i \operatorname{tr} \Delta S^{-1})\ g(S)\ (\det S)^{(r-\ell-1)/2}\, dS,$$

$\Delta = \Sigma^{-1/2} W \Sigma^{-1/2}$. Since g is symmetric, we can assume that Δ is diagonal, $\Delta = \operatorname{diag}(\Delta_1, \Delta_2, \ldots, \Delta_\ell)$.

Write $\exp(i \operatorname{tr} \Delta S^{-1}) = i \operatorname{tr} \Delta S^{-1} + \phi(\Delta S^{-1})$, where $\partial\phi(\Delta S^{-1})/\partial w_{\alpha\beta} = 0$ at $W = 0$, $1 \le \alpha, \beta \le \ell$. Then,

$$\int_S (\operatorname{tr} \Delta S^{-1})\ g(S)\ (\det S)^{(r-\ell-1)/2} dS = \sum_{j=1}^{\ell} \Delta_j \int_S s^{jj} g(S)\ (\det S)^{(r-\ell-1)/2} dS, \tag{10}$$

$S^{-1} = (s^{\alpha\beta})$. Since g is symmetric, an orthogonal (in fact, *permutation*) transformation will show that the integrals on the right hand side of (10) all have the same value. Hence, $E \operatorname{tr} \Delta S^{-1} \propto \sum_{j=1}^{k} \Delta_j = \operatorname{tr} \Sigma^{-1} W$, and replacing W by $\Sigma^{1/2} W \Sigma^{1/2}$ shows that $E \operatorname{tr} WS^{-1} \propto \operatorname{tr} W$. Therefore,

$$E\, s^{\alpha\beta} = \frac{1}{2i}(1 + \delta_{\alpha\beta}) \frac{\partial}{\partial w_{\alpha\beta}} E \exp(i \operatorname{tr} WS^{-1}) \Big|_{W=0}$$

$$= \tfrac{1}{2}(1 + \delta_{\alpha\beta}) \frac{\partial}{\partial w_{\alpha\beta}} E \operatorname{tr} WS^{-1} \Big|_{W=0} \propto \delta_{\alpha\beta}\ .$$

That is, $E\, S^{-1} \propto I_\ell$, so that $E H_q \propto \Sigma^{-1}$. Finally, (ii) is implied by a consequence of the proof of Theorem 5, viz., that H_p has the same distribution as under normality.

Acknowledgment. We are grateful to Professor I. M. Chakravarti for calling our attention to some relevant references.

REFERENCES

Anderson, T.W. and Dasgupta, S. (1963). Some inequalities on characteristic roots of matrices. *Biometrika 50*, 522-524.

Constantine, A.G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist. 34*, 1270-1285.

Davis, A.W. (1979). Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory. *Ann. Inst. Statist. Math. 31*, Part A, 465-485.

Dawid, A.P. (1977). Spherical matrix distributions and a multivariate model. *J. Roy. Statist. Soc.* B *39*, 254-261.

Dickey, J.M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist. 38*, 511-518.

James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist. 35*, 475-501.

Khatri, C.G. (1967). Some distribution problems connected with the characteristic roots of $S_1S_2^{-1}$. *Ann. Math. Statist. 38*, 944-948.

Khatri, C.G. (1972). On the exact finite series distribution of the smallest or the largest root of matrices in three situations. *J. Multivariate Anal. 2*, 201-207.

Mardia, K.V., Kent, J.B. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, New York

Shah, B.K. and Khatri, C.G. (1974). Proof of conjectures about the expected values of the elementary symmetric functions of a noncentral Wishart matrix. *Ann. Statist. 2*, 833-836.

Tsukanov, A.V. (1981). On tests of optimality of experimental designs for a regression model with random variables. *Theory Prob. Appl. 26*, 173-177.