Ann. Inst. Statist. Math.
33 (1981), Part A, 279-285

THE POWER OF THE LIKELIHOOD RATIO TEST FOR ADDITIONAL INFORMATION IN A MULTIVARIATE LINEAR MODEL

YASUNORI FUJIKOSHI

(Received Dec. 15, 1980; revised Mar. 20, 1981)

Summary

This paper deals with the likelihood ratio test for additional information in a multivariate linear model. It is shown that the power of the likelihood ratio test procedure has a monotonicity property. Asymptotic approximations for the power are also obtained.

1. Introduction

Let Y be an observed $N\times p$ matrix of p variables $y_1,\cdots,y_p$ on each of N individuals. We assume that the N rows of Y are independently distributed according to p variate normal distributions with the common covariance matrix Σ and expectations given by

$$\mathrm{E}(Y)=A\Xi \tag{1.1}$$

where A is a known $N\times k$ matrix of rank k and Ξ is a $k\times p$ matrix of unknown parameters. We partition Ξ and Σ as

$$\Xi=(\underset{p_1}{\Xi_1},\ \underset{p_2}{\Xi_2}) \qquad \text{and} \qquad \Sigma=\begin{array}{cc} \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} & \begin{array}{c}p_1\\ p_2\end{array}\\ \begin{array}{cc}p_1 & \quad p_2\end{array} & \end{array} \tag{1.2}$$

respectively. Consider the problem of testing the hypothesis (Rao [9], [10])

$$H_0: C\Gamma=0 \qquad \text{against} \qquad H_1: C\Gamma\neq 0 \tag{1.3}$$

where C is a known $q\times k$ matrix of rank q, $\Gamma=\Xi_2-\Xi_1\boldsymbol{\beta}$ and $\boldsymbol{\beta}=\Sigma_{11}^{-1}\Sigma_{12}$. From McKay [7] the hypothesis H_0 can be interpreted as the hypothesis that $\boldsymbol{y}_2'=(y_{p_1+1},\cdots,y_p)$ supplies no additional information about departures from nullity of the hypothesis $\tilde{H}_0: C\Xi=0$, independently of

AMS 1970 subject classifications: Primary 62H15; Secondary 62H10.
Key words and phrases: Power of likelihood ratio test, additional information, asymptotic expansions, nonnull distribution, multivariate linear model.

$\boldsymbol{y}_1'=(y_1, \cdots, y_{p_1})$.

Let W and B be the matrices of sums and products due to error and departure from the hypothesis in the problem of testing $\tilde{H}_0: C\Xi=0$ against $\tilde{H}_1: C\Xi\neq 0$, i.e., $W=Y'(I-A(A'A)^{-1}A')Y$ and $B=Y'A(A'A)^{-1}\cdot C'\{C(A'A)^{-1}C'\}^{-1}C(A'A)^{-1}A'Y$. Then the likelihood ratio criterion for testing H_0 against H_1 is an increasing function of

$$\Lambda=|S_e|/|S_e+S_h| \tag{1.4}$$

where $S_e=W_{22.1}=W_{22}-W_{21}W_{11}^{-1}W_{12}$, $S_h=W_{22}+B_{22}-(W_{21}+B_{21})(W_{11}+B_{11})^{-1}\cdot(W_{12}+B_{12})-W_{22.1}$ and W_{ij} and B_{ij} are submatrices of W and B, respectively, partitioned in the manner of Σ. In this paper we shall discuss about the distributions of S_e and S_h. Using the distributional results we shall show that the power of the likelihood ratio test procedure has a monotonicity property. Further asymptotic nonnull distributions of Λ are obtained.

2. The distributions of S_e and S_h

Consider the partitioning of Y into the sub-observation matrices of the first p_1 variables and the last p_2 variables as (Y_1, Y_2). Given Y_1, the N rows of Y_2 are independently distributed according to p_2 variate normal distributions with the common covariance matrix $\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\cdot\Sigma_{11}^{-1}\Sigma_{12}$ and expectations given by

$$\mathrm{E}\,(Y_2|Y_1)=A\Gamma+Y_1\boldsymbol{\beta}\ . \tag{2.1}$$

Under the conditional setup the testing problem (1.3) can be regarded as one of a linear hypothesis in a multivariate linear model. Further, since Λ is also the likelihood ratio statistic under the conditional model, it is implicitly known that the conditional distribution of Λ given Y_1 is a Wilks' lambda distribution. Here we shall give a direct proof of the result and further distributional results on S_h. Since W and B are independently distributed as a central Wishart distribution $W_p(N-k, \Sigma)$ and a noncentral Wishart distribution $W_p(q, \Sigma; \Xi'C'\{C(A'A)^{-1}\cdot C'\}^{-1}C\Xi)$, respectively, we may write W and B as $U'U$ and $X'X$, respectively, where the rows of $U: (N-k)\times p$ and $X: q\times p$ are independently distributed according to $N_p(\cdot, \Sigma)$, $\mathrm{E}\,(U)=0$ and $\mathrm{E}\,(X)=\{C(A'A)^{-1}C'\}^{-1/2}\cdot C\Xi=\eta$. Let U, X and η decompose as $U=(U_1, U_2)$, $X=(X_1, X_2)$ and $\eta=(\eta_1, \eta_2)$, respectively, where $U_1: (N-k)\times p_1$, $X_1: q\times p_1$ and $\eta_1: q\times p_1$. Noting that $(I+X_1W_{11}^{-1}X_1')^{-1}=I-X_1(X_1'X_1+W_{11})^{-1}X_1'$,

$$S_h=(X_2-X_1W_{11}^{-1}W_{12})'(I+X_1W_{11}^{-1}X_1')^{-1}(X_2-X_1W_{11}^{-1}W_{12})\ . \tag{2.2}$$

Hence we can express S_e and S_h as $Z_2'FZ_2$ and $Z_2'GZ_2$, respectively, with

$Z_2'=(X_2', U_2')$,

$$F=\begin{pmatrix} 0 & 0 \\ 0 & I-U_1(U_1'U_1)^{-1}U_1' \end{pmatrix},$$

$$G=(I, -X_1W_{11}^{-1}U_1')'(I+X_1W_{11}^{-1}X_1')^{-1}(I, -X_1W_{11}^{-1}U_1') .$$

Given $Z_1'=(X_1', U_1')$ the rows of Z_2 are independently distributed according to p_2 variate normal distributions with the common covariance matrix $\Sigma_{22.1}$ and expectations given by

$$\mathrm{E}(Z_2|Z_1)=\begin{pmatrix} \tilde{\eta}_2+X_1\boldsymbol{\beta} \\ U_1\boldsymbol{\beta} \end{pmatrix}$$

where

$$\tilde{\eta}_2=\{C(A'A)^{-1}C'\}^{-1/2}C\Gamma . \tag{2.3}$$

Noting, that $F^2=F$, $G^2=G$, $FG=0$ and $\mathrm{E}(Z_2|Z_1)G\,\mathrm{E}(Z_2'|Z_1)=\tilde{\eta}_2'(I+X_1 \cdot W_{11}^{-1}X_1')^{-1}\tilde{\eta}_2$, we have

THEOREM 1.

(1) *S_e has a central Wishart distribution $W_{p_2}(N-k-p_1, \Sigma_{22.1})$. Given $R=(I+X_1W_{11}^{-1}X_1')^{-1}$,*

(2) *S_h has a noncentral Wishart distribution $W_{p_2}(q, \Sigma_{22.1}; \tilde{\eta}_2'R\tilde{\eta}_2)$,*

(3) *S_e and S_h are independent.*

Further, the rows of X_1 are independently distributed according to p_1 variate normal distributions with the common covariance matrix Σ_{11} and expectations given by $\mathrm{E}(X_1)=\eta_1$, W_{11} has a central Wishart distribution $W_{p_1}(N-k, \Sigma_{11})$ and X_1 and W_{11} are independent.

Remark 1. Under the assumption of $\eta_1=0$, i.e., $C\Xi_1=0$, the distributions of S_e and S_h are essentially the same as ones of the matrices due to error and departure from hypothesis in a general MANOVA model (cf. Fujikoshi [5]).

3. Monotonicity of the power of Λ

When we consider the distribution of Λ, we may assume without loss of generality that

$$\begin{aligned} &S_e\sim W_{p_2}(N-k-p_1, I),\ S_h\sim W_{p_2}(q, I; \Delta) \text{ given } R, \\ &\text{where } \Delta=\zeta_2'R\zeta_2 \text{ and } \zeta_2=\tilde{\eta}_2\Sigma_{22.1}^{-1/2}, \\ &W_{11}\sim W_{p_1}(N-k, I) \text{ and } \mathrm{E}(X_1)=\zeta_1=\eta_1\Sigma_{11}^{-1/2} . \end{aligned} \tag{3.1}$$

Since the conditional power of the likelihood ratio test procedure depends only on the characteristic roots $\delta_1 \geqq \cdots \geqq \delta_{p_2}$ of Δ, the unconditional power depends on both $\Theta=\zeta_2\zeta_2'$ and ζ_1. We can write the power and the conditional power of the likelihood ratio test procedure as $\beta_\Lambda(\Theta, \zeta_1)$ and $\beta_\Lambda(D_\delta|R)$, respectively, where $D_\delta=\mathrm{diag}(\delta_1, \cdots, \delta_{p_2})$. Then

$$\beta_\Lambda(\Theta, \zeta_1)=\mathrm{E}(\beta_\Lambda(D_\delta|R)) . \tag{3.2}$$

Using the result due to Das Gupta, Anderson and Mudholkar [4] that $\beta_\Lambda(D|R)$ increases monotonically in each δ_i, we have

THEOREM 2. *If* $\Theta^*=\zeta_2^*\zeta_2^{*\prime} \geqq \Theta=\zeta_2\zeta_2'$, then

$$\beta_\Lambda(\Theta^*, \zeta_1) \geqq \beta_\Lambda(\Theta, \zeta_1) . \tag{3.3}$$

PROOF. It is sufficient to show that $\delta_i^* \geqq \delta_i$, where $\delta_1^* \geqq \cdots \geqq \delta_{p_2}^*$ are the characteristic roots of $\Delta^*=\zeta_2^{*\prime}R\zeta_2^*$. Then the inequality can be proved as follows:

$$\begin{aligned}\delta_i^* &= \mathrm{ch}_i(R^{1/2}\Theta^* R^{1/2}) \\ &= \mathrm{ch}_i\{R^{1/2}\Theta R^{1/2}+R^{1/2}(\Theta^*-\Theta)R^{1/2}\} \\ &\geqq \mathrm{ch}_i(R^{1/2}\Theta R^{1/2})=\delta_i \qquad \text{(ch. [3], pp. 33–34)}\end{aligned}$$

where $\mathrm{ch}_i(\cdot)$ means the ith largest characteristic root of a matrix.

Remark 2. Let ϕ be any test on the characteristic roots of $S_h S_e^{-1}$. Then we can write the corresponding power and the conditional power as $\beta_\phi(\Theta, \zeta_1)$ and $\beta_\phi(D_\delta|R)$, respectively. Then from the proof of Theorem 2 it follows that if $\beta_\phi(D_\delta|R)$ increases monotonically in each δ_i, $\beta_\phi(\Theta, \zeta_1)$ has the same monotonicity property as in Theorem 2. For the monotonicity results for $\beta_\phi(D_\delta|R)$, see Das Gupta, Anderson and Mudholkar [4] and Perlman and Olkin [8].

Remark 3. Under the assumption of $C\Xi_1=0$ it follows from Remark 1 and Fujikoshi [5] that the power of the likelihood ratio test procedure depends only on the characteristic roots $\omega_1 \geqq \cdots \geqq \omega_{p_2}$ of $\Omega=\zeta_2'\zeta_2=\Sigma_{22.1}^{-1/2}\Xi_2'C'\{C(A'A)^{-1}C'\}^{-1}C\Xi_2\Sigma_{22.1}^{-1/2}$ and increases monotonically in each ω_i.

4. Asymptotic nonnull distribution of Λ

In this section we consider asymptotic approximations for the power of the likelihood ratio test procedure when p and q are fixed and the sample size N is large. We may write the power with a level of significance α as

$$\beta_\Lambda(\Theta, \zeta_1; \alpha)=\mathrm{P}(-m \log \Lambda \geqq u) \tag{4.1}$$

where $m=N-k-p_1+(q-p_2-1)/2$ and the u is determined such that $\mathrm{P}(-m\log\Lambda\geqq u|H_0)=\alpha$. Since under H_0 Λ has a Wilks' lambda distribution $\Lambda(p, N-k-p_1, q)$, the u can be approximated by the formula (cf. Anderson [1], p. 208)

$$\begin{aligned}(4.2)\qquad &\mathrm{P}(-m\log\Lambda\leqq x|H_0)\\ &=\mathrm{P}(\chi_f^2\leqq x)+\frac{p_2q}{48m^2}(p_2^2+q^2-5)\{\mathrm{P}(\chi_{f+4}^2\leqq x)\\ &\quad-\mathrm{P}(\chi_f^2\leqq x)\}+O(m^{-4})\end{aligned}$$

where $f=p_2q$. In the following we consider asymptotic nonnull distributions of $-m\log\Lambda$.

First we assume that $\zeta=(\zeta_1, \zeta_2)=O(1)$. Then we have $\Delta=O_p(1)$. From Sugiura and Fujikoshi [11] we can write the conditional characteristic function of $-m\log\Lambda$ as

$$(4.3)\qquad \phi(t|X_1, W_{11})=\psi(\Delta)+O_p(m^{-2})$$

where

$$\begin{aligned}(4.4)\qquad \psi(\Delta)&=(1-2it)^{-f/2}\exp\left[\frac{it}{1-2it}\operatorname{tr}\Delta\right]\\ &\cdot\times\left[1+\frac{1}{4m}\sum_{j=1}^{3}a_j(\Delta)(1-2it)^{-j}\right]\end{aligned}$$

and

$$(4.5)\qquad \begin{aligned}&a_1(\Delta)=(p_2+q+1)\operatorname{tr}\Delta\ ,\\ &a_2(\Delta)=-(p_2+q+1)\operatorname{tr}\Delta+\operatorname{tr}\Delta^2\ ,\qquad a_3(\Delta)=-\operatorname{tr}\Delta^2\ .\end{aligned}$$

Noting that $\Delta=\Omega-m^{-1}\zeta_2'X_1(m^{-1}W_{11})^{-1}X_1'\zeta_2+m^{-2}\zeta_2'\{X_1(m^{-1}W_{11})^{-1}X_1'\}^2R\zeta_2$, it can be shown that

$$\begin{aligned}(4.6)\qquad \mathrm{E}(\psi(\Delta))&=\psi(\Omega)+\frac{1}{2m}(1-2it)^{-f/2}\exp\left[\frac{it}{1-2it}\operatorname{tr}\Omega\right]\\ &\times[1-(1-2it)^{-1}](p_1\operatorname{tr}\Omega+\operatorname{tr}\zeta_1'\Theta\zeta_1)+O(m^{-2})\end{aligned}$$

where $\Omega=\zeta_2'\zeta_2$ and $\Theta=\zeta_2\zeta_2'$. By formally inverting (4.6) we obtain an asymptotic expansion given by Theorem 3. The derivation of the asymptotic expansion will be justified by proving $\mathrm{E}(O_p(m^{-2}))=O(m^{-2})$ and the existence of a valid asymptotic expansion for $\mathrm{P}(-m\log\Lambda\leqq x)$.

THEOREM 3. *Under $\zeta=O(1)$, the following asymptotic formula for the nonnull distribution of $-m\log\Lambda$ holds for large N.*

(4.7) $\mathrm{P}(-m\log\Lambda\leqq x)$

$$=G_f(x:\omega^2)+\frac{1}{4m}\sum_{j=1}^{3}a_j(\Omega)G_{f+2j}(x:\omega^2)$$
$$+\frac{1}{2m}(p_1\operatorname{tr}\Omega+\operatorname{tr}\zeta_1'\Theta\zeta_1)[G_f(x:\omega^2)-G_{f+2}(x:\omega^2)]+O(m^{-2})$$

where $m=N-k-p_1+(q-p_2-1)/2$, $f=p_2q$, $\omega^2=(\operatorname{tr}\Omega)/2$, $\Omega=\zeta_2'\zeta_2$, $\Theta=\zeta_2\zeta_2'$, *the coefficients* $a_j(\cdot)$ *are given by* (4.5) *and* $G_f(x:\omega^2)$ *denotes the distribution function of a noncentral chi-square variate with* f *degrees of freedom and noncentrality parameter* ω^2.

Next we assume that $\zeta=\sqrt{m}L$, where L is a fixed matrix. Consider the asymptotic distribution of

$$\tilde{\Lambda}=\sqrt{m}\{-\log\Lambda-\log|I+Q|\} \tag{4.8}$$

where $Q=L_2'(I+L_1L_1')^{-1}L_2$, $L=(L_1,L_2)$ and $L_1:q\times p_1$. Under $\zeta=\sqrt{m}L$ we have

$$\frac{1}{m}\Lambda=Q+\frac{1}{\sqrt{m}}M'(L_1VL_1'-ZL_1'-L_1Z')M+O_p(m^{-1}) \tag{4.9}$$

where $M=(I+L_1L_1')^{-1}L_2$, $Z=X_1-\zeta_1$ and $V=\sqrt{m}(m^{-1}W_{11}-I)$. Let $T_e=\sqrt{m}((1/m)S_e-I)$ and $T_h=\sqrt{m}((1/m)S_h-Q)$. Then it is seen (cf. Sugiura [12]) from (4.9) that

$$\begin{aligned}&\mathrm{E}\,[\exp\{it(\operatorname{tr}AT_e+\operatorname{tr}BT_h)\}]\\&\quad=\exp[-t^2\{\operatorname{tr}A^2+2\operatorname{tr}QB^2+\operatorname{tr}(L_1'MBM'L_1)^2\\&\qquad+2\operatorname{tr}L_1'MBM'MBM'L_1\}](1+O(m^{-1/2}))\end{aligned} \tag{4.10}$$

where A and B are any symmetric matrices. This show that T_e and T_h converge in law to $p(p+1)/2$ variate normal distributions as m tends to infinity. We can express $\tilde{\Lambda}$ as

$$\tilde{\Lambda}=\operatorname{tr}\{(I+Q)^{-1}-I\}T_e+\operatorname{tr}(I+Q)^{-1}T_h+O_p(m^{-1/2})\,. \tag{4.11}$$

Applying a theorem on limiting distributions (cf. Anderson [2]), we have

THEOREM 4. *Under* $\zeta=\sqrt{m}L$ *the limiting distribution of* $\tilde{\Lambda}$ *is* $N(0,\sigma^2)$, *where* $\sigma^2=2\{p_2-\operatorname{tr}(I+Q)^{-2}\}+2\operatorname{tr}\{L_1'M(I+Q)^{-1}M'L_1\}^2+4\operatorname{tr}L_1'M(I+Q)^{-1}\cdot M'M(I+Q)^{-1}M'L_1$, *where* $Q=L_2'(I+L_1L_1')^{-1}L_2$ *and* $M=(I+L_1L_1')^{-1}L_2$.

The statistic Λ defined by (1.4) is also the likelihood ratio test statistic for testing $H_0:C\Gamma=0$ given $C\Xi_1=0$. The approximations for the nonnull distribution of Λ given $C\Xi_1=0$ are given by (4.7) and Theorem 4 with $\zeta_1=0$ and $\zeta_2=\{C(A'A)^{-1}C'\}^{-1/2}C\Xi_2\Sigma_{22.1}^{-1/2}$. Further asymptotic expansions in this case have been obtained by Fujikoshi [6].

Acknowledgements

The author would like to thank the referee for his helpful comments.

HIROSHIMA UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*, Wiley, New York.

[2] Anderson, T. W. (1963). Asymptotic theory for principal component analysis, *Ann. Math. Statist.*, **34**, 122-148.

[3] Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics, 1*, Interscience, New York.

[4] Das Gupta, S., Anderson, T. W. and Mudholkar, G. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis, *Ann. Math. Statist.*, **35**, 200-205.

[5] Fujikoshi, Y. (1973). Monotonicity of the power functions of some tests in general MANOVA models, *Ann. Statist.*, **1**, 388-391.

[6] Fujikoshi, Y. (1974). Asymptotic expansions of the non-null distributions of three statistics in GMANOVA, *Ann. Inst. Statist. Math.*, **26**, 289-297.

[7] McKay, R. J. (1977). Simultaneous procedures for variable selection in multiple discriminant analysis, *Biometrika*, **64**, 283-290.

[8] Perlman, M. D. and Olkin, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems, *Ann. Statist.*, 8, 1326-1341.

[9] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, Wiley, New York.

[10] Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis, *Multivariate Analysis—I* (ed. P. R. Krishnaiah), Academic Press, New York, 87-103.

[11] Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence, *Ann. Math. Statist.*, **40**, 942-952.

[12] Sugiura, N. (1973). Further asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis, *Ann. Inst. Statist. Math.*, **25**, 153-163.