Ann. Inst. Statist. Math.
Vol. 45, No. 4, 615–620 (1993)

# ON RATES OF CONVERGENCE OF INFORMATION THEORETIC CRITERION IN RANK DETERMINATION OF ONE-WAY RANDOM EFFECTS MODELS*

Y. WU

*Department of Mathematics and Statistics, York University, 4700 Keele Street, North York, Ontario, Canada M3J 1P3*

(Received February 28, 1992; revised September 2, 1992)

**Abstract.** In this paper, upper bounds on the probabilities of wrong determination of the rank of covariance matrix of random effects in one-way random effects models are given, based on the information theoretic criterion. Under weak conditions, the bounds are shown of exponential-type.

*Key words and phrases*: Rate of convergence, information theoretic criterion, one-way random effects model, rank determination.

## 1. Introduction

Consider a one-way multivariate random effects model:

$$\boldsymbol{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij} \tag{1.1}$$

for $i = 1, \dots, k$; $j = 1, \dots, m$, where $\boldsymbol{\mu} \in R^p$ is the general mean vector, $\boldsymbol{\alpha}_i$ is the vector of random effects of $i$-th column and $\boldsymbol{x}_{ij}$ denotes the $j$-th observation on $i$-th column and $\boldsymbol{e}_{ij}$ is distributed as multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_1$. Also, $\boldsymbol{\alpha}_i$ is distributed independent of $\boldsymbol{e}_{ij}$ as multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\psi$. Let $\Sigma_2$ denote the covariance matrix of $\boldsymbol{x}_{ij}$, then $\Sigma_2 = \psi + \Sigma_1$. It is of interest to test whether the rank of $\psi$ is $r$. If $\psi$ is not of full rank, then we can take advantage of this knowledge in estimating $\psi$. Specially, that $r = 0$ is equivalent to that there are no column effects. Anderson (1984, 1985) and Schott and Saw (1984) have independently derived the likelihood ratio test statistic for testing the hypothesis on the rank of $\psi$. Zhao *et al.* (1986) suggests to use information theoretic criterion (ITC) to determine the rank of $\psi$. They have showed that the criterion is strongly consistent.

The main contribution of our paper is to give exponential-type bounds on the probabilities of error determination under certain conditions, based on the ITC

* The work was partially supported by Natural Sciences and Engineering Research Council of Canada and Faculty Research Grant, Faculty of Arts, York University.

proposed in Zhao *et al.* (1986). The bounds tend to zero rapidly as the sample size increases. The paper is organized as follows.

In Section 2, the preceding problem is generalized and some preliminaries are given. In Section 3, several theorems about the rates of convergence of the ITC are given and proved.

## 2. Preliminaries

In Model (1.1), let $S_b$ and $S_w$ denote the between groups and within group sums of squares and cross products matrices respectively. Then $S_b$ and $S_w$ are distributed independently as central Wishart matrices with $k-1$ and $k(m-1)$ degrees of freedom respectively, $E(S_b) = (k-1)(\Sigma_1 + m\psi)$ and $E(S_w) = k(m-1)\Sigma_1$. Hence, we can generalize the problem as follows.

Suppose that $\boldsymbol{y}$ and $\boldsymbol{u}$ are $p \times 1$ complex random vectors which are distributed with mean vector $\mathbf{0}$, $E\boldsymbol{uu}^* = \Sigma_1 > 0$, $E\boldsymbol{yy}^* = \Sigma_2 > 0$, where $(\cdot)^*$ is the conjugate transpose of $(\cdot)$. Also, $\Sigma_2 = \Gamma + \sigma^2\Sigma_1$, where $\sigma^2$ is known and $\Gamma$ is a $p \times p$ nonnegative definite matrix of rank $q$ with $q < p$. Assume that two independent sets of observations $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n_1}\}$ and $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2}\}$ of $\boldsymbol{u}$ and $\boldsymbol{y}$ are available. Let $n_1 S_1 = \sum_{i=1}^{n_1} \boldsymbol{u}_i \boldsymbol{u}_i^*$ and $n_2 S_2 = \sum_{i=1}^{n_2} \boldsymbol{y}_i \boldsymbol{y}_i^*$. Then $E(S_1) = \Sigma_1$ and $E(S_2) = \Sigma_2$, respectively. We need to estimate $q$.

Denote the eigenvalues of $S_2 S_1^{-1}$ by $\delta_1 \geq \cdots \geq \delta_p$ and denote the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ by $\lambda_1 \geq \cdots \geq \lambda_p$. Let $H_k$ denote the hypothesis that

$$H_k : \lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} = \cdots = \lambda_p = \sigma^2,$$

and let $M_k$ denote the model for which $H_k$ is true. Without loss of generality, we may assume that $\sigma^2 = 1$. Then, let

$$I(k, C(n)) = \frac{n}{2} \sum_{j=1+\min(k,\tau)}^{p} [\log(\alpha_n + \beta_n\delta_j) - \beta_n \log \delta_j] + C(n)\nu(k,p),$$

where $\alpha_n = n_1/n$, $\beta_n = n_2/n$, $n = n_1 + n_2$, $\tau = \#\{i \leq p : \delta_i > 1\}$, $\nu(k,p) = k(2p-k+1)$, the number of free parameters in $H_k$, and $C(n)$ is chosen to satisfy the following conditions:

$$\lim_{n\to\infty} \frac{C(n)}{n} = 0; \qquad \lim_{n\to\infty} \frac{C(n)}{\log\log n} = \infty. \tag{2.1}$$

In Zhao *et al.* (1986), an estimate $\hat{q}$ of $q$ is proposed to satisfy

$$I(\hat{q}, C(n)) = \min\{I(0, C(n)), \ldots, I(p-1, C(n))\}. \tag{2.2}$$

This criterion was proven to be strongly consistent in Zhao *et al.* (1986).

Denote matrices $\Sigma_1$, $\Sigma_2$, $S_1$ and $S_2$ as

$$\Sigma_l = (\sigma_{ij}^{(l)}), \quad S_l = (s_{ij}^{(l)}), \qquad l = 1, 2$$

and let their respective eigenvalues be $\lambda_1^{(l)} \geq \cdots \geq \lambda_p^{(l)}$ and $\delta_1^{(l)} \geq \cdots \geq \delta_p^{(l)}$ for $l = 1, 2$. The following lemma is needed in the proof of the main theorem.

LEMMA 2.1. *Assume that $p$ eigenvalues of $\Sigma_1$ take $t$ different values, say $\mu_1 > \cdots > \mu_t > 0$. For the matrices $\Sigma_l$, $S_l$, $l = 1, 2$, if $\max|\sigma_{ij}^{(l)} - s_{ij}^{(l)}| \leq \alpha \leq \lambda_p^{(1)}/p$, then $|\lambda_i - \delta_i| \leq M\alpha$ where $M = p^2[K\lambda_1^{(2)}(1/\sqrt{\lambda_p^{(1)}} + 1/\sqrt{\lambda_p^{(1)} - p\alpha}) + p/(\lambda_p^{(1)} - p\alpha)]$, and $K = (2tp^3/\sqrt{\mu_t})[(p^3/\min_{i\neq j}|\mu_i - \mu_j|)(1 + \mu_1/\mu_t) + 1/\mu_t] + p^2/(2\sqrt{\mu_t}(\mu_t - p\alpha)^2)$.*

For the proof, see Tam and Wu (1991).

The following inequality is also needed in the proof of the main theorem:

$$x(1-x) \leq \log(1+x) < x \tag{2.3}$$

for $|x| < 1/2$.

## 3. Rates of convergence of the ITC

In this section, we study the rates of convergence of the criterion (2.2).

Let

$$G(k) = -\sum_{j=1+\min(k,\tau)}^{p} [\log(\xi + \zeta\delta_j) - \zeta \log \delta_j]$$

where $\xi + \zeta = 1$, $0 < \xi < 1$. Define $\bar{I}(k, \bar{C}) = -G(k) + \nu(k, p)\bar{C}$ where $\bar{C} = 2C(n)/n$. It is easy to see that an estimate $\hat{q}$ of $q$ such that

$$\bar{I}(\hat{q}, \bar{C}) = \min\{\bar{I}(0, \bar{C}), \ldots, \bar{I}(p-1, \bar{C})\}$$

is same as the one given by (2.2). We have the main theorem as follows.

THEOREM 3.1. *Let $S_l : p \times p$ be an estimate of $\Sigma_l$, $l = 1, 2$ as before. We assume that the following conditions are true:*

(i) $\max|\sigma_{ij}^{(l)} - s_{ij}^{(l)}| \leq \alpha$, $l = 1, 2$.

(ii) *There is a fixed interval $[a, b] \subset (0, 1)$ such that $\xi \in [a, b]$.*

(iii) $0 < \alpha < \lambda_p^{(1)}/p$.

(iv)

$$\varepsilon < \min\left\{\frac{\lambda_q - 1}{2}, \frac{1}{2}\right\} \tag{3.1}$$

*where $\varepsilon = M\alpha$, $M$ is given in Lemma 2.1.*

(v) $(p-q)\varepsilon^2 < \bar{C}$.

(vi) $2p\bar{C} < g(\gamma_0)$, *where $g(\gamma_0) = \min\{\log((1-\gamma) + \gamma(\lambda_q + 1)/2) - \gamma \log((\lambda_q + 1)/2) : \gamma$ in an interval $[a, b] \subset (0, 1)\}$.*

*Then*

$$P(\hat{q} \neq q \mid H_q) \leq \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{l=1}^{2} P(|s_{ij}^{(l)} - \sigma_{ij}^{(l)}| \geq \alpha). \tag{3.2}$$

PROOF. By (i) and Lemma 2.1, we get that $|\lambda_i - \delta_i| \leq \varepsilon$, $i = 1, 2, \ldots, p$. Obviously, $\delta_q \geq \lambda_q - |\lambda_q - \delta_q| \geq \lambda_q - \varepsilon \geq (\lambda_q + 1)/2 > 1$. Thus $\min(q, \tau) = q$ as $\tau \geq q$.

Let $k > q$, then $\min(k, \tau) \geq q$. Thus we obtain:

$$\bar{I}(q, \bar{C}) - \bar{I}(k, \bar{C}) = \left\{ \sum_{i=1+q}^{\min(k,\tau)} [\log(\xi + \zeta\delta_i) - \zeta \log \delta_i] \right\} + (\nu(q,p) - \nu(k,p))\, \bar{C}.$$

For $i > q$, $\lambda_i = 1$. Since $|\lambda_i - \delta_i| = |1 - \delta_i| \leq \varepsilon < 1/2$, by the inequality (2.3), we have

$$\log(\xi + \zeta\delta_i) - \zeta \log \delta_i \leq \zeta(\delta_i - 1) - \zeta(\delta_i - 1)[1 - (\delta_i - 1)] = \zeta(\delta_i - 1)^2.$$

Thus,

$$[\bar{I}(q, \bar{C}) - \bar{I}(k, \bar{C})]/\bar{C} \leq \frac{1}{\bar{C}} \left\{ \sum_{i=1+q}^{\min(k,\tau)} \zeta(\delta_i - 1)^2 - (k - q)(2p - k - q + 1) \right\}.$$

By Condition (v), it follows that $(1/\bar{C}) \sum_{i=1+q}^{p} \zeta(\delta_i - 1)^2 < 1$. As $(k-q)(2p-k-q+1) > 1$, $[\bar{I}(q, \bar{C}) - \bar{I}(k, \bar{C})]/\bar{C} < 0$ or $\bar{I}(q, \bar{C}) < \bar{I}(k, \bar{C})$.

Let $k \leq q$. Then $\delta_i \geq \cdots \geq \delta_q > 1$ for $i \leq k$. By condition (iv), $\delta_i \geq (\lambda_q + 1)/2 > 1$ for all $i = k+1, \ldots, q$. The function $f(\delta) = \log(\xi + \zeta\delta) - \zeta \log \delta$ is increasing for $\delta > 1$. As $\min(q, \tau) = q$ and $\min(k, \tau) = k$,

$$\begin{aligned} G(q) - G(k) &= \sum_{i=k+1}^{q} [\log(\xi + \zeta\delta_i) - \zeta \log \delta_i] \\ &\geq (q - k) \left[ \log\left(\xi + \zeta \frac{\lambda_q + 1}{2}\right) - \zeta \log \frac{\lambda_q + 1}{2} \right]. \end{aligned}$$

The function $g(\gamma) = \log((1-\gamma) + \gamma(\lambda_q + 1)/2) - \gamma \log{(\lambda_q + 1)/2}$ is continuous. For some interval $[a, b] \subset (0, 1)$, there is a $\gamma_0$ so that $0 < g(\gamma_0) \leq g(\gamma)$ since $g(\gamma) > 0$ for $\gamma \in [a, b] \subset (0, 1)$. Thus

$$G(q) - G(k) > (q - k) g(\gamma_0)$$

and

$$\begin{aligned} \bar{I}(q, \bar{C}) - \bar{I}(k, \bar{C}) &= -G(q) + G(k) + 2[\nu(q,p) - \nu(k,p)]\bar{C} \\ &< -(q - k) g(\gamma_0) + (q - k)(2p - k - q + 1)\bar{C} \\ &= (q - k)[(2p - k - q + 1)\bar{C} - g(\gamma_0)]. \end{aligned}$$

By (vi), $\bar{I}(q,\bar{C}) < \bar{I}(k,\bar{C})$. Hence $\hat{q} = q$ and

$$P(\hat{q} \neq q \mid H_q) \leq \sum_{j=1}^{p}\sum_{i=1}^{p}\sum_{l=1}^{2} P(|s_{ij}^{(l)} - \sigma_{ij}^{(l)}| \geq \alpha). \qquad \square$$

*Remark* 3.1. From the proof, it is easy to see that $\nu(k,p)$ can be chosen as a strictly increasing function of $k$.

*Remark* 3.2. The result here is different from the result in Bai *et al.* (1989). In Bai *et al.* (1989), $\Sigma_1$ is an identity matrix and an independent estimation of $\Sigma_1$ is not required.

The proof is existential. When we are interested in the limiting properties, we can choose $\alpha = \alpha(n)$ and $C(n) = n\bar{C}/2$ satisfying the following properties:

$$\begin{aligned} &\alpha(n) \downarrow 0, \\ &\frac{C(n)}{n} \to 0, \\ &\frac{C(n)}{n\alpha^2(n)} \to \infty. \end{aligned} \tag{3.3}$$

Then for large $n$, conditions in (3.1) hold. If $C(n)$ satisfies (2.1), we can choose $\alpha = \alpha(n)$ so that (3.3) is true. We thus have the following.

COROLLARY 3.1. *If $\alpha(n)$ and $C(n)$ are chosen so as to satisfy condition* (3.3) *and $\nu(k,p)$ is strictly decreasing function, then upper bound* (3.2) *on the probability of wrong determination holds.*

From (3.2) we see that the probability of correct determination increases as the value of $\alpha$ increases. Therefore, from (3.3) we observe that the probability of correct determination increases as the value of $C(n)$ increases, since a larger value of $C(n)$ allows us to take a larger value of $\alpha(n)$.

Up to now, no special assumption is made about $S_l$, $l = 1, 2$. For the remaining, we impose some moment conditions and obtain results as in Bai *et al.* (1989). The proofs are similar to those in Bai *et al.* (1989) and will not be given.

THEOREM 3.2. *Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ are identically and independently distributed (i.i.d.) vectors of order $p \times 1$ such that $E(\boldsymbol{y}_1) = \mathbf{0}$, $E(\boldsymbol{y}_1\boldsymbol{y}_1^*) = \Sigma_2$ and $E|\boldsymbol{y}_1|^{2\eta} < \infty$ for $\eta > 1$ and that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ are i.i.d. vectors of order $p \times 1$ such that $E(\boldsymbol{u}_1) = \mathbf{0}, E(\boldsymbol{u}_1\boldsymbol{u}_1^*) = \Sigma_1$ and $E|\boldsymbol{u}_1|^{2\eta} < \infty$. Also, let $C(n)$ and $\alpha(n)$ be chosen so that they satisfy* (3.3). *Then for any $s > \eta$, we have*

$$P(\hat{q} \neq q \mid H_q) = O(n/(n\alpha)^{\eta}) + O((n\alpha^2)^{-s}) \qquad as \quad n \to \infty.$$

COROLLARY 3.2. *In Theorem* 3.2, *if we take $\alpha = \alpha(n) \downarrow 0$ as a slowly varying function and $C(n) = n\alpha$ then*

$$P(\hat{q} \neq q \mid H_q) = O(n^{1-\eta}(\alpha)^{-\eta}) \qquad as \quad n \to \infty.$$

The exponential-type bound on the probability of wrong determination is given as:

THEOREM 3.3. *Suppose that* $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ *are i.i.d. with* $E(\boldsymbol{y}_1) = \mathbf{0}$, $E(\boldsymbol{y}_1 \boldsymbol{y}_1^*) = \Sigma_2$ *and* $E\{\exp(\eta|\boldsymbol{y}_1|^2)\} < \infty$ *for some* $\eta > 0$. *Also let* $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ *be i.i.d. with* $E(\boldsymbol{u}_1) = \mathbf{0}$, $E(\boldsymbol{u}_1 \boldsymbol{u}_1^*) = \Sigma_1$ *and* $E\{\exp(\eta|\boldsymbol{u}_1|^2)\} < \infty$. *Then*

$$P(\hat{q} \neq q \mid H_q) \leq C \exp\{-bn\alpha^2\}$$

*as* $n \to \infty$ *for some constants* $b > 0$ *and* $C > 0$.

COROLLARY 3.3. *If* $\alpha(n) \downarrow 0$ *is a slowly varying function,* $C(n) = n\alpha$ *and the conditions of Theorem* 3.3 *are true, then for any* $\delta > 0$

$$P(\hat{q} \neq q \mid H_q) \leq C \exp(-bn^{1-\delta}).$$

*Remark* 3.3. If $n_i S_i \sim W_p(n_i, \Sigma_i)$, $i = 1, 2$, and $S_1$ and $S_2$ are independent, then the results in Theorem 3.2 and Theorem 3.3 hold.

Back to the original problem, the solutions follow from the above remark.

## Acknowledgements

The author acknowledges gratefully many helpful suggestions by the referees.

## REFERENCES

Anderson, T. W. (1984). Estimation linear statistical relationships, *Ann. Statist.*, **12**, 1–45.

Anderson, T. W. (1985). Components of variance in MANOVA, *Multivariate Analysis* (ed. P. R. Krishnaiah), **6**, 1–8, North-Holland, Amsterdam.

Bai, Z. D., Krishnaiah, P. R. and Zhao, L. C. (1989). On rate of convergence of efficient detection criteria in signal processing with white noise, *IEEE Trans. Inform. Theory*, **35**, 380–388.

Schott, J. R. and Saw, J. G. (1984). A multivariate one-way classification model with random effects, *J. Multivariate Anal.*, **15**, 1–12.

Tam, K. W. and Wu, Y. (1991). On the convergence of general information theoretic criteria in signal processing when the covariance matrix is arbitrary, *IEEE Trans. Inform. Theory*, **37**, 1667–1671.

Zhao, L. C., Krishnaiah, P. R. and Bai, Z. D. (1986). On detection of the number of signals when the noise covariance matrix is arbitrary, *J. Multivariate Anal.*, **20**, 26–48.