Selection of Variables in Two-group Discriminant Analysis by Error Rate and Akaike's Information Criteria

YASUNORI FUJIKOSHI

Hiroshima University, Japan

Communicated by P. R. Krishnaiah

This paper deals with two criteria for selection of variables for the discriminant analysis in the case of two multivariate normal populations with different means and a common covariance matrix. One is based on the estimated error rate of misclassification. The other uses Akaike's information criterion. The asymptotic distributions and error rate risks of the criteria are obtained. The result will prove that the two criteria are asymptotically equivalent in the sense of their asymptotic distributions and error rate risks being identical.

0 1985 Academic Press, Inc.

1. INTRODUCTION

Let \( x = (x_1, ..., x_p)' \) be an observation vector consisting of the available variables associated with an object which is to be allocated to one of two multivariate normal populations \( \Pi_g: N_p[\mu_g, \Sigma], \ g = 1, 2 \). The mean vectors \( \mu_1, \mu_2 \) and the covariance matrix \( \Sigma \) of full rank are unknown. Suppose that random samples of sizes \( N_g \) from each population \( \Pi_g \ (g = 1, 2) \), are available. In many applications, it is desired to find the "best" subset of variables for classifying an observation \( x \) as coming from \( \Pi_1 \) or \( \Pi_2 \). A number of methods have been suggested for selection of variables. For a summary of the methods, see, e.g., Eisenbeis and Gilbert (1973), Lachenbruch (1975), Habbema and Hermans (1977), McLachlan (1980), Krishnaiah (1982). However, it seems that the theoretical study of the methods has been little done.

This paper is concerned with the methods based on two criteria. One is based on the estimated error rate proposed by McLachlan (1976, 1980).
The other is based on an application of Akaike's information criterion (Akaike, 1974). The purpose of this paper is to study the statistical property of the two criteria. We give the asymptotic distribution of the selected subset of variables based on each criterion. Using the asymptotic distribution we give the asymptotic error rate risk when the subset of variables is selected by each criterion. The result will prove that the two criteria are asymptotically equivalent in the sense of their asymptotic distributions and error rate risks being identical.

2. TWO CRITERIA FOR SELECTION OF VARIABLES

We will identify the subvector \( x(j) = (x_{j_1}, ..., x_{j_k})' \) of \( x \) by the corresponding subset \( j = \{ j_1, ..., j_k \} \) of the set of subscripts \( 1, 2, ..., p \). Let \( J \) be the family of all possible subsets of \( \{ 1, ..., p \} \). Then the problem of selection of variables may be regarded as how to select the best subset of variables \( j \) from \( J \). If we use only a subset of variables \( x(j) \), then we may classify a new observation \( x \) by means of the classification statistic

\[
w(j) = (\bar{x}_1(j) - \bar{x}_2(j))' S(j)^{-1}\{x(j) - \frac{1}{2}(\bar{x}_1(j) + \bar{x}_2(j))\},
\]

where \( \bar{x}_g \) and \( S \) are the sample means and pooled sample covariance matrix, and \( \bar{x}_g(j) \) and \( S(j) \) denote the \( \bar{x}_1 \) and \( S \) corresponding to \( x(j) \). The rule is to classify \( x \) as coming from \( \Pi_1 \) if \( w(j) > 0 \) and from \( \Pi_2 \) if \( w(j) \leq 0 \). The expected error rate with equal a priori probabilities is given by

\[
R(j) = E\{L(j)\},
\]

where \( L(j) = \frac{1}{2}\{L_1(j) + L_2(j)\} \) and \( L_g(j) \) is the conditional error of mis-allocation for \( x \) coming from \( \Pi_g \), i.e.,

\[
L_g(j) = \Phi \left( \frac{(-1)^g \frac{1}{2}(\bar{x}_1(j) - \bar{x}_2(j))' S(j)^{-1}\{\mu_g(j) - \frac{1}{2}(\bar{x}_1(j) + \bar{x}_2(j))\}}{(\bar{x}_1(j) - \bar{x}_2(j))' S(j)^{-1}\Sigma(j) S(j)^{-1}(\bar{x}_1(j) - \bar{x}_2(j))} \right)^{1/2},
\]

where \( \Phi \) denotes the standard normal distribution function, and \( \bar{x}_g(j) \) and \( \Sigma(j) \) denote the \( \mu \) and \( \Sigma \) corresponding to \( x(j) \).

One of the natural methods for selection of variables is to select the subset \( j \) which minimizes an estimate of \( R(j) \). As an asymptotic unbiased estimate McLachlan (1976, 1980) proposed

\[
M(j) = \Phi(G(j)),
\]
where $D$ and $D(j)$ are the sample Mahalanobis distance between $\Pi_1$ and $\Pi_2$, $n = N - 2$, $N = N_1 + N_2$, and

$$G(j) = -\frac{1}{2}D(j) + \frac{1}{2}(k(j) - 1)[N_1^{-1} + N_2^{-1}]/D(j) + (32n)^{-1}D(j)[4(4k(j) - 1) - D(j)^2].$$ (2.5)

We denote the selection method based on $M(j)$ by $j_M$, i.e., $M(j_M) = \text{Min}_{j \subseteq j_M} M(j)$. Since $\Phi$ is a monotonic increasing function, $j_M$ minimizes also $G(j)$, i.e., $G(j_M) = \text{Min}_{j \subseteq j_M} G(j)$.

The other method considered here is based on a model selection criterion. We shall define a parametric model $\Omega(j)$ which leads $x(j)$ to be the “best” subsets of variables. As one of such parametric models we adopt the no additional information model, defined by

$$\Omega(j); a_k = 0 \text{ for any } k \in j^c \quad \text{and} \quad a_k = 0 \text{ for any } k \in j.$$ (2.6)

where $a = (a_1, ..., a_p)' = \Sigma^{-1}(\mu_1 - \mu_2)$ is the vector of coefficients of linear discriminant function for populations $\Pi_1, \Pi_2$, and $j^c$ is the complement of $j$ with respect to the entire set $\{1, 2, ..., p\}$. The first condition in (2.6) is the same as Rao’s (1973, p. 551) hypothesis that $x(j)$ provides no additional information. As a model selection criterion we use Akaike’s (1974) information criterion

$$AIC(j) = -2 \log f(\hat{\Theta}(j)) + 2p(j),$$ (2.7)

where $f(\Theta)$ is the likelihood function of the initial $N$ observations on $x$, $\hat{\Theta}(j)$ is the maximum likelihood estimate of $\Theta = [\mu_1, \mu_2, \Sigma]$ under $\Omega(j)$, and $p(j) = p + k(j) + \frac{1}{2}p(p + 1)$ is the dimensionality of $\Theta$ under $\Omega(j)$. By the argument similar to the derivation of the likelihood ratio criterion for testing “$a_k = 0$ for any $k \in j^c$” (see, e.g., Rao (1973, Sect. 8c.4)) we have

$$A(j) = AIC(j) - AIC(\{1, 2, ..., p\}) = N \log \{1 + (p-k(j)) F(j)/(N-p-1)\} + 2(k(j) - p),$$ (2.8)

where

$$F(j) = \{(N-p-1)/(p-k(j))\}D^2 - D(j)^2)/\{n(N_1^{-1} + N_2^{-1}) + D(j)^2\}.$$ (2.9)

We select the subset of variables $x(j)$ to minimize $A(j)$ and denote this selection by $j_A$. 

---

**SELECTION OF VARIABLES**

29
3. A RELATIONSHIP BETWEEN $R(j)$ AND $\Omega(j)$

We will show that the expected error rate $R(j)$ is closely related to the no additional information model $\Omega(j)$. Let $j_0$ be a fixed subset in $J$. We may assume $j_0 = \{1, \ldots, k_0\}$ without loss of generality. We call $\Omega(j_0)$ true model if $\{\mu_1, \mu_2, \Sigma\}$ satisfies the condition $\Omega(j_0)$ in (2.6). Letting

$$J_1 = \{j \in J; j \supseteq j_0\} \quad \text{and} \quad J_2 = J_1 \cap J,$$

(3.1)

it is known (Fujikoshi, 1983) that $\Omega(j_0)$ is true if and only if

$$\Delta(j) = \Delta \quad \text{for any } j \in J_1 \quad \text{and} \quad \Delta(j) < \Delta \quad \text{for any } j \in J_2,$$

(3.2)

where $\Delta$ and $\Delta(j)$ are the population Mahalanobis distance between $\Pi_1$ and $\Pi_2$ based on $x$ and $x(j)$, respectively.

**Theorem 1.** The model $\Omega(j_0)$ is true if and only if

1. $\infty > \lim N(R(j) - R(j_0)) > 0$ for $j \in J_1 - \{j_0\}$,
2. $\lim N(R(j) - R(j_0)) > 0$ for $j \in J_2$,

where "lim" denotes the limit when $N_1 \to \infty$, $N_2 \to \infty$, and $N_1/N_2 \to \rho$, a fixed constant.

**Proof.** From Corollary 2 in Okamoto (1963) we have

$$R(j) = \Phi(-\frac{1}{2}\Delta(j)) + \phi(-\frac{1}{2}\Delta(j))\left[\frac{1}{16}\{4(k(j) - 1)/\Delta(j)\right.$$

$$+ \Delta(j)\}({N_1}^{-1} + {N_2}^{-1}) + \frac{1}{4}(k(j) - 1)\Delta(j)/n] + O_2,$$

(3.3)

where $\phi$ denotes the standard normal density function and $O_2$ denotes the term of the second order with respect to $N_1^{-1}$, $N_2^{-1}$, and $n^{-1}$. Using this formula we can see the equivalence of (3.2) and (i), (ii). This completes the proof.

From Theorem 1 we can regard $\Omega(j)$ as a minimal realization of the parametric model such that $R(j)$ is minimum, in the sense of (i), (ii) in Theorem 1.

4. GOODNESS OF CRITERIA AND PRELIMINARY LEMMAS

Let $j_*$ be a selection method, i.e., a mapping from $X$ to $J$, where $X$ denotes the observation matrix based on the samples of sizes $N_g$ from each population $\Pi_g$ ($g = 1, 2$). We assesses the goodness of the criterion in terms of

$$\{p_{N_*}(j) = \Pr(j_*=j); j \in J\},$$

(4.1)
and

\[ R_{n,*} = E_X \{ L(\hat{j}_*) \}. \quad (4.2) \]

It may be noted that in our problem (4.2) is more appropriate than (4.1). Shibata (1976) considered these typed measures of goodness in selection of the order of an autoregressive model by Akaike's information criterion.

For the derivation of (4.1) and (4.2) we may assume, without loss of generality,

**Assumption 1.** The model \( \Omega(j_0) \) is true, where

\[ j_0 = \{1, ..., k_0\}. \]

In the following we will give some Lemmas useful in deriving asymptotic expressions of (4.1) and (4.2).

**Lemma 1.** For any positive constant \( b \),

\[ \Pr(D(j)^2 - \Delta(j)^2 > b) \leq O(e^{-\sqrt{n}b}). \quad (4.3) \]
\[ \Pr(D(j)^2 - \Delta(j)^2 < -b) \leq O(e^{-\sqrt{n}b}). \quad (4.4) \]

**Proof.** Using a relationship between \( D^2 \) and Hotelling's \( T^2 \) we can express \( D(j)^2 \) as the ratio of independent \( \chi^2 \)-variates,

\[ D(j)^2 = \left\{ \frac{Nn/(N_1N_2)}{\chi^2_{k(j)}(\lambda^2)/\chi^2_{n-k(j)+1}} \right\} \]

where \( \lambda^2 = (N_1N_2/N) \Delta(j)^2 \). Using Chebyshev's inequality we have, for \( \theta > 0 \),

\[ \Pr(D(j)^2 - \Delta(j)^2 > b) \]
\[ = \Pr \left( \frac{1}{n} \left( \Delta(j)^2 + b \right) \chi^2_{n-k(j)+1} < \left\{ \frac{N/(N_1N_2)}{\chi^2_{k(j)}(\lambda^2)} \right\} \right) \]
\[ \leq E \left[ \exp \left\{ -\frac{\theta}{n} \left( \Delta(j)^2 + b \right) \chi^2_{n-k(j)+1} \right\} E \left\{ \exp \left( \frac{\theta N}{N_1N_2} \chi^2_{k(j)}(\lambda^2) \right) \right\} \right] \]
\[ = \left\{ 1 + \frac{\theta}{2n} \left( \Delta(j)^2 + b \right) \right\}^{-\left( n - k(j) + 1 \right)/2} \left( 1 - 2 \frac{\theta N}{N_1N_2} \right)^{-k(j)/2} \]
\[ \cdot \exp \left[ \theta \Delta(j)^2 \left\{ 1 - 2 \frac{\theta N}{N_1N_2} \right\} \right]. \]

Letting \( \theta = \sqrt{n} \), we obtain (4.3). Similarly we can prove (4.4).
LEMMA 2. Under Assumption 1 there exists a \( p \times p \) matrix \( B \) such that

1. \( B = (b_{ij}) \), \( B_{11} ; k_0 \times k_0 \) and \( B_{22} \) is a lower triangular matrix,
2. \( B(\mu_1 - \mu_2) = (A, 0, \ldots, 0)' = \delta, \)
3. \( BSB' = I_p = \) the identity matrix of order \( p \).


Using the matrix \( B \) in Lemma 2 we define \( y, u, \) and \( V \) by

\[
Bd = \delta + (1/\sqrt{n}) y, \quad B(\bar{x}_1 - \mu_1) = (1/\sqrt{n}) u,
\]

\[
BSB' = I_p + (1/\sqrt{n}) V,
\]

where \( d = \bar{x}_1 - \bar{x}_2 \). Then the limiting distribution of \( y = (y_1, \ldots, y_p)' \) is \( N_p[0, (2 + p + \rho^{-1})I_p] \), the limiting distribution of \( u = (u_1, \ldots, u_p)' \) is \( N_p[0, (1 + \rho)I_p] \), and the limiting distribution of \( V = [v_{kl}] \) is normal with mean 0 and \( E[v_{kk}^2] = 2, E[v_{kl}^2] = 1, k \neq l \). Further the \( \frac{1}{2} p(p+3) \) elements \( y_k \) and \( v_{kl} \) \((k < l)\) are independent in the limiting distribution. For \( j \in J_1 \), we define a \((k(j) - k_0) \times (p - k_0)\) matrix \( T(j) \) of zeros and ones and a \((p - k_0) \times (p - k_0)\) matrix \( K(j) \) by

\[
x(j) = (x_1', T(j)')',
\]

\[
K(j) = B_{22}^{-1} T(j)' \{ T(j) \Sigma_{22}^{-1} T(j)' \}^{-1} T(j) B_{22}^{-1},
\]

where \( \Sigma_{22}^{-1} = \Sigma_{22}^{-1} \Sigma_{11}^{-1} \Sigma_{12} \),

\[
x = \begin{pmatrix} x_1 & k_0 \\ x_2 & p - k_0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma_{11} ; k_0 \times k_0.
\]

LEMMA 3. Under Assumption 1 it holds that for \( j \in J_1 \),

\[
D(j)^2 = D(j_0)^2 + \frac{1}{n} \sigma^2 z_j^2 K(j) z_j + O_p(n^{-3/2}),
\]

where \( \sigma = \{A^2 + 2 + \rho + \rho^{-1}\}^{1/2}, \quad z = (z_1, \ldots, z_p)' \), \( z_k = (y_k - \Delta v_{1k})/\sigma \), and \( z_2 = (z_{k_0+1}, \ldots, z_p)' \). The limiting distribution of \( z_2 \) is \( N_{p-k_0} [0, I_{p-k_0}] \).

Proof. Using (4.7) it is seen that

\[
D(j)^2 = D(j_0)^2 + (d_2 - S_{21} S_{11}^{-1} d_1)'
\]

\[
\times \{ T(j)' \{ T(j) S_{22}^{-1} T(j)' \}^{-1} T(j) \} (d_2 - S_{21} S_{11}^{-1} d_1),
\]

where \( d = (d_1, d_2)' \), \( d_1 ; k_0 \times 1 \) and \( S_{kl} \) are the submatrices of \( S \) partitioned as in (4.9). Expressing the second term of (4.11) in terms of \( y \) and \( V \) in (4.5) and (4.6) we obtain the desired result.
5. ASYMPTOTIC DISTRIBUTIONS OF $j_M$ AND $j_A$

We consider the asymptotic distributions of $j_M$ and $j_A$ under Assumption 1 when $N_1 \to \infty$, $N_2 \to \infty$, and $N_1/N_2 \to \rho$, a positive constant. The asymptotic distribution of $j_A$ is obtained as a special case of Fujikoshi (1983) as follows: Let $\lim P_{N,A}(j) = p_A(j)$. Then for $j \in J_1$,

$$P_{N,A}(j) = \Pr(S'WW - W) = \xi^t \epsilon < 2(k(m) - k(j)) \quad \text{for} \quad m \in J_1$$

(5.1)

and for $j \in J_2$,

$$p_A(j) = 0,$$

(5.2)

where $\xi = (\xi_1, \ldots, \xi_{p-k_0})'$ and $\xi_k$'s are independent random variables with $N[0, 1]$ distributions. In the following we give the asymptotic distribution of $j_M$ and an improvement for the convergences of $p_{N,M}(j)$ and $p_{N,A}(j)$ for $j \in J_2$.

**Lemma 5.** Under Assumption 1 it holds that for $j \in J_1$ and $i \in J_2$,

$$\Pr(G(i) < G(j)) = O(e^{-\sqrt{n}a})$$

(5.3)

$$\Pr(A(i) < A(j)) = O(e^{-\sqrt{n}a}),$$

(5.4)

where $\alpha = \min_{j \in J_1, i \in J_2} \{\Delta(j) - \Delta(i)\} > 0$ and $b = \frac{1}{16}\alpha^2$.

**Proof.** We can write

$$\Pr(G(i) < G(j)) = \Pr(D(i) - D(j) > \frac{1}{n} h(D(i), D(j))),$$

where $h(x, y) = \{nN/(N_1N_2)\} \{(k(i) - 1)/x - (k(j) - 1)/y\} + \frac{1}{16} \{(k(i) - 1)x - x^2 - 4(k(j) - 1)y + y^2\}$. Let $Q$ be the set of $(D(i), D(j))$ such that $|D(i) - D(j)| < \frac{1}{2} \alpha$ and $|D(j) - A(j)| < \frac{1}{2} \alpha$. Then from Lemma 1 we have $\Pr(Q^c) \leq O(e^{-\sqrt{n}b})$. Further, there exists a positive number $n_0$ such that if $n > n_0$, $|(1/n)h(D(i), D(j))| < \frac{1}{2} \alpha$ for $(D(i), D(j)) \in Q$. Therefore we have, for $n > n_0$,

$$\Pr(G(i) < G(j)) \leq \Pr\left\{\left\{(D(i) - D(j)) > \frac{1}{n} h(D(i), D(j))\right\} \cap Q\right\} + \Pr(Q^c)$$

$$\leq \Pr(X(i) - X(j) > \frac{1}{2} \alpha) + \Pr(Q^c)$$

$$\leq \Pr(X(i) > \frac{1}{2} \alpha) + \Pr(X(j) < -\frac{1}{2} \alpha) + \Pr(Q^c)$$

$$\leq O(e^{-\sqrt{n}b}),$$

where $X(j) = D(j) - A(j)$. This completes the proof.
THEOREM 2. The asymptotic distribution of $j_M$ is the same as one of $j_A$. Under Assumption 1 it holds that

(i) for $j \in J_1$,
\[
\lim p_{N,M}(j) = p_M(j) = p_A(j) \quad \text{in (5.1), and} \quad (5.5)
\]

(ii) for $j \in J_2$ and any positive constant $h$,
\[
\lim N^h p_{N,M}(j) = \lim N^h p_{N,A}(j) = 0. \quad (5.6)
\]

Proof. Equation (5.6) follows from Lemma 5. From Lemma 5 we obtain that for $j \in J_1$,
\[
0 \leq \tilde{p}_{N,M}(j) - p_{N,M}(j) \leq O(e^{-\sqrt{n}}), \quad (5.7)
\]
\[
0 \leq \tilde{p}_{N,A}(j) - p_{N,A}(j) \leq O(e^{-\sqrt{n}}), \quad (5.8)
\]

where $\tilde{p}_{N,M}(j) = \Pr(G(j) \leq G(m) \text{ for } m \in J_1)$ and $\tilde{p}_{N,A}(j) = \Pr(A(j) \leq A(m) \text{ for } m \in J_1)$. From Fujikoshi (1983) or Lemma 4 we obtain that for $j, m \in J_1$,
\[
A(j) - A(m) = z_2(K(m) - K(j)) z_2 + 2(k(m) - k(j)) + O_p(n^{-1/2}). \quad (5.9)
\]

Similarly we obtain that for $j, m \in J_1$,
\[
(4An/\sigma^2)(G(j) - G(m)) = \text{the right-hand side of (5.9).} \quad (5.10)
\]

The formulas (5.7)–(5.10) imply (5.5).

We are sometimes interested in selecting the "best" subset from a subfamily $\bar{J}$ of $J$. Let the selection methods obtained by minimizing $M(j)$ and $A(j)$ for $j \in \bar{J}$ denote by $j_M$ and $j_A$, repectively. By the same way as in the proof of Theorem 2 we obtain.

THEOREM 3. Let $\bar{J}$ be a subfamily of $J$ such that $\bar{J}_1 = J_1 \cap \bar{J} \neq \emptyset$. Then $\bar{j}_M$ and $\bar{j}_A$ have the same asymptotic distribution, i.e., $\lim \tilde{p}_{N,M}(j) = \lim \tilde{p}_{N,A}(j) = \tilde{p}(j)$, where $\tilde{p}_{N,*} = \Pr(\bar{j}_* = j)$. Further, under Assumption 1 it holds that

(i) for $j \in J_1$,
\[
\tilde{p}(j) = \Pr(\xi'(K(m) - K(j)) \xi \leq 2(k(m) - k(j)) \text{ for } m \in \bar{J}_1),
\]

(ii) for $j \in \bar{J}_2 = J_2 \cap \bar{J}$ and any positive constant $h$,
\[
\lim N^h \tilde{p}_{N,M}(j) = \lim N^h \tilde{p}_{N,A}(j) = 0.
\]
SELECTION OF VARIABLES

As one of the interested subfamilies we consider the family of \( p \) subsets

\[
J - \{1, 2, \ldots, \hat{p}\},
\]

(5.11)

where \( k = \{1, \ldots, k\} \). This is the case when an assessment of the relative importance of individual variables is given a priori and the initial order of the variables \( x_1, \ldots, x_p \) makes sense. Then, since \( B_{22} \) is a lower triangular matrix, we have

\[
\begin{pmatrix}
I_{k - k_0} & 0 \\
0 & 0
\end{pmatrix}, \quad k_0 < k < p.
\]

(5.12)

This implies that for \( k_0 < k < p \).

\[
\bar{p}(k) = s(k - k_0) t(p - k),
\]

(5.13)

where \( s(k) = \Pr(\bigcap_{i=1}^{k} (U_i > 0)) \), \( t(k) = \Pr(\bigcap_{i=1}^{k} (U_i \leq 0)) \), \( s(0) = t(0) = 1 \), \( U_i = (W_i - 2) + \cdots + (W_i - 2) \) and \( W_i \)'s are independent random variables with \( \chi_i^2 \) distributions. We note that \( \bar{p}(k) \) in (5.13) is the same typed one as in Shibata (1976). For explicit formulas of \( s(k) \) and \( t(k) \), see Spitzer (1956) and Shibata (1976).

6. ASYMPTOTIC ERROR RATE RISKS

For simplicity, instead of (4.1) we evaluate

\[
r_{N,*} = E_x\{L(j_*) - L(j_0)\}
\]

(6.1)

which shows how much the risk increases by applying \( j_* \), from the risk \( E_x\{L(j_0)\} = \text{R}(j_0) \) when \( j_0 \) is known.

**Lemma 6.** It holds that for \( j \in J_1 \),

\[
L(j) = L(j_0) + \frac{c}{n} z_2^* K(j) z_2 + O_p(n^{-3/2}),
\]

(6.1)

where \( c = \sigma^2 \phi(-\frac{1}{2}A)/(4A) \), \( \phi \) is the pdf of \( N[0, 1] \), and \( z_2 \) and \( K(j) \) are defined by Lemma 3 and (4.8), respectively.
Proof. By the same way as in the proof of Lemma 3 we have
\[ d(j)'S(j)^{-1} \Sigma(j) S(j)^{-1} d(j) = d(j_0)'S(j_0)^{-1} \Sigma(j_0) S(j_0)^{-1} d(j_0) \]
\[ + \frac{3}{n} \sigma^2 z_j^2 K(j) z_j \]
\[ - \frac{2}{n} \sigma y_j^2 K(j) z_j + O_p(n^{-3/2}), \]  
(6.2)
\[ d(j)'S(j)^{-1}(\tilde{x}_1(j) - \mu_1(j)) = d(j_0)'S(j_0)^{-1}(\tilde{x}_1(j_0) - \mu_1(j_0)) \]
\[ + \frac{1}{n} \sigma u_j^2 K(j) z_j + O_p(n^{-3/2}), \]  
(6.3)
\[ d(j)'S(j)^{-1}(\tilde{x}_2(j) - \mu_2(j)) = d(j_0)'S(j_0)^{-1}(\tilde{x}_2(j_0) - \mu_2(j_0)) \]
\[ + \frac{1}{n} \sigma (u_j - y_j)^2 K(j) z_j + O_p(n^{-3/2}), \]  
(6.4)
where \( y = (y_1', y_2')', y_1: k_0 \times 1 \) and \( u = (u_1', u_2')', u_1: k_0 \times 1 \). Substituting (4.10) and (6.2)-(6.4) to (2.3) we obtain (6.1).

**Theorem 4.** The selection methods \( J_M \) and \( J_A \) have the same asymptotic increase in risk and under Assumption 1 it is given by
\[ \lim N r_{N,M} = \lim N r_{N,A} = c \sum_{j \in J - \{j_0\}} b(j), \]  
(6.5)
where \( c = \sigma^2 \phi(-\frac{1}{2}a)/(4a) \),
\[ b(j) = E\{ (\xi_j' K(j) \xi_j) I_{(\xi_j' (K(m) - K(k)) \xi_j \leq 2(k(m) - k(j)) \text{ for } j \in J_1)} \}, \]  
(6.6)
and \( I_{(\cdot)} \) denotes an indicator function of \( (\cdot) \).

Proof. Let \( J^* \) be \( J_M \) or \( J_A \). Using (5.6) and Lemma 6 we can write
\[ r_{N,*} = \sum_{j \in J} E X_j \{ [L(j) - L(j_0)] I_{(j_*) = j} \} \]
\[ = \frac{c}{n} \sum_{j \in J - \{j_0\}} E X_j \{ (z_j^2 K(j) z_j) I_{(j_*) = j} \} + O(n^{-3/2}). \]  
(6.7)
This implies the desired result.

By the same way as in the proof of Theorem 4 it is easily seen that the selection methods \( J_M \) and \( J_A \) in Theorem 3 have the same asymptotic increase in risk. Let \( \tilde{r}_{N,M} \) and \( \tilde{r}_{N,A} \) be the increases in risk for \( J_M \) and \( J_A \). Then \( \lim N \tilde{r}_{N,M} = \lim N \tilde{r}_{N,A} = \) the right-hand side of (6.5) replaced \( J_1 \) by
\( J_1 = J_1 \cap \bar{J}. \) Especially, for the case when \( J \) is given by (5.11), the result is equal to \( e \sum_{k=k_0+1} b(k) \), where
\[
b(k) = E\left[ \{ U_k - k_0 + 2(k - k_0) \} I(U_m - k_0 - U_k - k_0 \leq 0 \text{for } m = k_0 \ldots p) \right]
\] (6.8)
and \( U_k \)'s are defined in (5.13). Shibata (1976) has given an exact expression of \( b(k) \) only by using tail probabilities of \( \chi^2 \) random variables.

**ACKNOWLEDGMENTS**

The author wishes to thank the referee for his valuable comments and suggestions.

**REFERENCES**