The Annals of Statistics
1984, Vol. 12, No. 2, 758–765

ASYMPTOTIC PROPERTIES OF CRITERIA FOR SELECTION OF VARIABLES IN MULTIPLE REGRESSION

BY RYUEI NISHII

Hiroshima University, Japan

In normal linear regression analysis, many model selection rules proposed from various viewpoints are available. For the information criteria AIC, FPE, C_p, PSS and BIC, the asymptotic distribution of the selected model and the asymptotic quadratic risk based on each criterion are explicitly obtained.

1. Introduction. In normal linear regression analysis, many model selection rules are available. Most of these selection rules are obtained by estimates of risk functions, for example $(-1) \times$ (Kullback-Leibler information) or predictive mean squared error. (See the review articles by Hocking, 1976, Thompson, 1978a,b.) It is known that some criteria are asymptotically equivalent to each other. But the equivalence of two criteria may only assert that the value of one criterion function converges asymptotically to the other criterion function under the null hypothesis. Such an equivalence does not always imply that the risk based on one criterion will converge to the risk based on the other criterion. The equivalence of criteria should be defined by the risk function based on one criterion converging to the risk function based on the other.

Consider the information criteria AIC, FPE, C_p, PSS and BIC, which are respectively proposed by Akaike (1973, 1970), Mallows (1973), Allen (1971) and Schwarz (1978). This paper is concerned with applications of these criteria for selection of variables and prediction in multiple regression. We give explicitly the asymptotic distributions of the selected model and the quadratic risk when the model is selected by these criteria. This result will prove that AIC, FPE, C_p and PSS are asymptotically equivalent in the above two senses under general conditions. However, BIC or its generalization GIC has different properties from those of AIC. See Sugiura (1978) and Hashimoto et. al (1981) for numerical studies.

2. Regression model and information criteria. Consider the multiple regression model

$$\mathbf{y} = X\beta + \mathbf{e}, \tag{2.1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of observations, X is a known $N \times K$ design matrix, $\beta = (\beta_1, \cdots, \beta_K)'$ is an unknown parameter vector and $\mathbf{e}$ is an $N \times 1$ error vector whose elements are assumed to be independently normally distributed with mean 0 and unknown variance σ^2, i.e., $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 I_N)$. Under this setup, we select

Received November 1982; revised October 1983.
AMS 1980 *subject classifications.* Primary 62J05, secondary 62F12.
Key words and phrases. AIC, BIC, C_p, FPE, PSS, regression analysis, selection of variables.

variables which enable us to give good prediction for future observations by some information criterion. We say the model is $j = \{j_1, \cdots, j_k\}$ $(1 \le j_1 < \cdots < j_k \le K)$ if and only if $\beta_{j_1} \ne 0, \cdots, \beta_{j_k} \ne 0$ and other elements of β are all zeros. Then the number of unknown parameters is defined by $k(j) = k + 1$ because σ^2 is treated as unknown. Let D_j be a $K \times k$ matrix of zeros and ones such that XD_j contains columns $j_1, \cdots, j_k$ of X. When the model j is given, we regard (2.1) as

$$\mathbf{y} = X\beta(j) + \mathbf{e}$$

where $\beta(j) = D_jD_j'\beta = D_j(\beta_{j_1}, \cdots, \beta_{j_k})'$. Let J be a set of models j under consideration. For example, J is a hierarchic model $J_H = \{\bar{j}_1, \cdots, \bar{j}_K\}$ where $\bar{j}_t = \{1, \cdots, t\}$ $(t = 1, \cdots, K)$. Throughout this paper we make the following assumption:

ASSUMPTION 1. The true model is $j_0 = \{1, \cdots, k_0\}$ and J includes j_0. The matrix $X'X$ is positive definite, and $M = \lim_{N\to\infty} N^{-1}X'X$ exists and is positive definite.

This assumption implies rank $(XD_j) = k$, that is, $D_j'X'XD_j$ is positive definite. For the model $j \in J$ we define the following quantities:

$$\hat{\beta}(j) = D_j(D_j'X'XD_j)^{-1}D_j'X'\mathbf{y} = \text{MLE of } \beta(j),$$

$$Q(j) = XD_j(D_j'X'XD_j)^{-1}D_j'X'$$

$$= \text{projection operator w.r.t. column space of } XD_j,$$

$$\hat{\sigma}^2(j) = N^{-1}\mathbf{y}'\{I_N - Q(j)\}\mathbf{y} = \text{MLE of } \sigma^2.$$

We discuss the following information criteria:

$$\text{AIC}(j) = N \log \hat{\sigma}^2(j) + ak(j),$$

$$C_p(j) = N\hat{\sigma}^2(j)/\hat{\sigma}^2(\bar{j}_K) + a\{k(j) - 1\},$$

$$\text{FPE}(j) = [N + a\{k(j) - 1\}]\hat{\sigma}^2(j),$$

$$\text{PSS}(j) = \mathbf{y}'\{I_N - Q(j)\}\{I_N - \Lambda(j)\}^{-2}\{I_N - Q(j)\}\mathbf{y},$$

$$\text{GIC}(j) = N \log \hat{\sigma}^2(j) + a_Nk(j),$$

where a is a positive constant (usually a is defined to be 2), $\Lambda(j)$ is a diagonal matrix whose diagonal elements are given by those of $Q(j)$, and $a_N > 0$ is a sequence such that $\lim_{N\to\infty} a_N = \infty$ and $\lim_{N\to\infty} N^{-1}a_N = 0$. When $a_N = \log N$, GIC is known as BIC. We select model j in the set J by minimizing the value of a criterion. In this paper, our asymptotic study of criteria has K and β fixed as $N \to \infty$. For alternative asymptotics, see Stone (1979) and Shibata (1981).

3. Goodness of criterion and general results. Let $\hat{j}$ be the model selected from J by some information criterion. We assess the goodness of the criterion in terms of

(i) $$\{p_N(j) = \Pr\{\hat{j} = j\}; j \in J\}, \quad \text{and}$$

(ii) $$R_N = E_y[\| X\beta - X\hat{\beta}(\hat{j})\|^2],$$

where $\|\cdot\|$ denotes the Euclidean norm. The expected mean squared error of future observations is given by $R_N + N\sigma^2$. The risk function R_N is expressed by $R_N = \sum_{j\in J} R_N(j)$ where $R_N(j) = E_y[\| X\beta - X\hat{\beta}(j)\|^2 I_{(\hat{j}=j)}]$ and $I_{(\cdot)}$ denotes an indicator function of $(\cdot)$. Shibata (1976) considered these measures of goodness in selection of the order of an autoregressive model by AIC, and Fujikoshi (1982) considered them in selection of variables in two-group discriminant analysis.

Let $J_1 = \{j \in J \mid j \not\supseteq j_0\}$ and $J_2 = \{j \in J \mid j \supseteq j_0\}$. We define the following conditions on a criterion:

CONDITION 1. $\lim_{N\to\infty} N p_N(j) = 0$ for $j \in J_1$.

CONDITION 2. $\lim_{N\to\infty} p_N(j) = 0$ for $j \in J_2 - \{j_0\}$.

Some implications of these conditions for $\{R_N(j)\}$ are given by:

THEOREM 1. (i) *If a criterion satisfies Condition* 1, *then* $\lim_{N\to\infty} R_N(j) = 0$ *for* $j \in J_1$. (ii) *If a criterion satisfies Condition* 2, *then* $\lim_{N\to\infty} R_N(j) = 0$ *for* $j \in J_2 - \{j_0\}$.

PROOF. For $j \in J$, we have

$$R_N(j) = \beta' X'\{I_N - Q(j)\}X\beta \cdot p_N(j) + E[\mathbf{e}'Q(j)\mathbf{e} \cdot I_{(\hat{j}=j)}] = I_1 + I_2.$$

By Schwarz's inequality,

$$I_2 \le [E\{\mathbf{e}'Q(j)\mathbf{e}\}^2 p_N(j)]^{1/2} = \sigma^2[\{k(j)^2 - 1\}p_N(j)]^{1/2}.$$

(i) Let $j \in J_1$. Then

$$\lim_{N\to\infty} N^{-1}\beta' X'\{I_N - Q(j)\}X\beta = \beta'\{M - MD_j(D_j'MD_j)^{-1}D_j'M\}\beta > 0$$

or

$$\beta' X'\{I_N - Q(j)\}X\beta = O(N)$$

by Assumption 1. Therefore we have I_1, $I_2 \to 0$ as $N \to \infty$ because $p_N(j) = o(N^{-1})$.

(ii) Let $j \in J_2 - \{j_0\}$. Then $\beta' X'\{I_N - Q(j)\}X\beta = \beta' X'\{Q(j_0) - Q(j)\}X\beta = 0$. By Condition 2, we have $I_2 \to 0$ as $N \to \infty$.

REMARK. For a criterion satisfying Conditions 1 and 2, we have

$$\lim_{N\to\infty} R_N = \lim_{N\to\infty} R_N(j_0) = k_0\sigma^2.$$

4. Asymptotic properties of criteria. In this section, we derive the asymptotic distribution of $\hat{j}$ and the limit of R_N for the information criteria AIC, C_p, FPE, PSS and GIC.

Let $M^{1/2}$ be an upper triangular matrix of order K satisfying $(M^{1/2})'M^{1/2} = M$ and, for $\ell \in J_2$, let L_ℓ be a $(K - k_0) \times k_\ell^*$ matrix defined by

$$M^{1/2}D_\ell = \begin{array}{c} \\ \boldsymbol{k}_0 \\ \boldsymbol{K}-\boldsymbol{k}_0 \end{array} \!\!\begin{array}{c} \begin{array}{cc} \boldsymbol{k}_0 & \boldsymbol{k}_\ell^* \end{array} \\ \begin{pmatrix} * & * \\ 0 & L_\ell \end{pmatrix} \end{array},$$

where $k_\ell^* = k(\ell) - k(j_0)$. For $\ell \in J_2$, we define

$$\xi_\ell = \mathbf{z}'A_\ell\mathbf{z} \quad \text{and} \quad \xi_\ell^{(a)} = \xi_\ell - ak_\ell^*, \tag{4.1}$$

where $\mathbf{z} \sim N(\mathbf{0}, I_{K-k_0})$ and $A_\ell = L_\ell(L_\ell'L_\ell)^{-1}L_\ell'$. In the case of $k_0 = K$, L_ℓ, $\mathbf{z}$ and A_ℓ are defined to be zeros.

LEMMA 1. *Let $\ell \in J_2$. Then* $\text{AIC}(j_0) - \text{AIC}(\ell)$ *converges in law to the random variable $\xi_\ell^{(a)}$ as $N \to \infty$.*

THEOREM 2. (Asymptotic properties of $\{p_N(j)\}$ and R_N for AIC).
(i) (a) *For $j \in J_1$ and any positive constant h,*

$$\lim_{N\to\infty} N^h p_N(j) = 0.$$

(b) *For $j \in J_2$, $p_N(j)$ converges to*

$$p(j) = \Pr\{\xi_j^{(a)} \geq \xi_\ell^{(a)} \text{ for } \ell \in J_2\}. \tag{4.2}$$

(ii) *The risk function R_N converges to*

$$R = \sigma^2[k_0 + \textstyle\sum_{j\in J_2} E\{\xi_j I_{(\xi_j^{(a)} \geq \xi_\ell^{(a)} \text{ for } \ell \in J_2)}\}]. \tag{4.3}$$

PROOF. We shall prove only (i) (a). Let $j \in J_1$. From Assumption 1, we have

$$\begin{aligned} p_N(j) &= \Pr\{\text{AIC}(j) \leq \text{AIC}(\ell) \text{ for } \ell \in J\} \leq \Pr\{\text{AIC}(j) \leq \text{AIC}(j_0)\} \\ &= \Pr\{X + Y_N + N^{1/2}c_N \leq Z_N\}, \end{aligned}$$

where

$$\begin{aligned} X &= 2(\lambda_N N)^{-1/2}\mathbf{e}'QX\boldsymbol{\beta} \sim N(0, 1), \\ Y_N &= (\lambda_N N)^{-1/2}\mathbf{e}'Q\mathbf{e} \to_P 0, \\ Z_N &= b_N\lambda_N^{-1/2}N^{-3/2}\mathbf{e}'\{I_N - Q(j_0)\}\mathbf{e} \to_P 0, \end{aligned}$$

$b_N = N\{\exp(apN^{-1}) - 1\}$, $p = k(j_0) - k(j)$, $Q = Q(j_0) - Q(j)$, $\lambda_N = 4\sigma^2N^{-1} \cdot \boldsymbol{\beta}'X'Q^2X\boldsymbol{\beta} > 0$, and $c_N = \lambda_N^{-1/2}N^{-1}\boldsymbol{\beta}'X'QX\boldsymbol{\beta} > 0$. By the assumption, λ_N, b_N and c_N converge, respectively, to $\lambda = 4\sigma^2(M^{1/2}\boldsymbol{\beta})'\{S(j_0) - S(j)\}^2M^{1/2}\boldsymbol{\beta} > 0$, ap and $c = \lambda^{-1/2}(M^{1/2}\boldsymbol{\beta})'\{S(j_0) - S(j)\}M^{1/2}\boldsymbol{\beta} > 0$ as $N \to \infty$ where $S(j) = M^{1/2}D_j(D_j'MD_j)^{-1}D_j'(M^{1/2})'$ for $j \in J$. The formula (4.5) is dominated by

$$\Pr\{X \leq -N^{1/2}c_N + 2N^{1/4}\} + \Pr\{-Y_N > N^{1/4}\} + \Pr\{Z_N > N^{1/4}\}$$

since $\text{P}(F) \leq \text{P}(F \cap G \cap H) + \text{P}(G^c) + \text{P}(H^c)$ for events F, G and H. Let $d_N = c_N - 2N^{-1/4} = O(1)$. For large N, an inequality for the standard normal distribution function implies

$$\Pr\{X \le -N^{1/2}d_N\} \le N^{-1/2}d_N^{-1}\phi(N^{1/2}d_N) = o(e^{-c_N^2 N/2})$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. By a relationship between a distribution function and its moment generating function, we have

$$\begin{aligned}\Pr\{-Y_N > N^{1/4}\} &\le e^{-N^{1/4}}E[e^{-Y_N}]\\ &= e^{-N^{1/4}}|I_N + 2\sigma^2(N\lambda_N)^{-1/2}Q|^{-1/2} = o(e^{-N^{1/4}}),\\ \Pr\{Z_N > N^{1/4}\} &\le e^{-N^{1/4}}E[e^{Z_N}]\\ &= e^{-N^{1/4}}(1 - 2\sigma^2 b_N\lambda_N^{-1/2}N^{-3/2})^{-(N-k_0)/2} = o(e^{-N^{1/4}}).\end{aligned}$$

This completes the proof.

In case $J = J_H$, $p(j)$ and R (defined in (4.2) and (4.3) respectively) can be reduced to computable forms and this result is essentially due to Shibata (1976). In case J is the family of all subsets of $\bar{j}_K$ and M is diagonal, $p(j)$ and R can be reduced to

$$\{\Pr(\chi_1^2 \ge a)\}^{k-k_0}\{\Pr(\chi_1^2 < a)\}^{K-k} \quad \text{and} \quad \sigma^2\{k_0 + (K-k_0)\Pr(\chi_3^2 \ge a)\},$$

respectively, where $j \in J_2$, $k(j) = k + 1$ and χ_q^2 denotes a random variable which has a chi-squared distribution with q degrees of freedom.

REMARK. We note that Lemma 1 and Theorem 2 remain valid if AIC is replaced by FPE or C_p.

Theorem 2 shows that $p(j_0)$ is a monotone increasing function of a. This may suggest that, for large N, the constant a should be chosen large. But when a is large, $\{p_N(j); j \in J_1\}$ converge to zero more slowly than for small a, while both $p_N(j)$ for large a and for small a are $o(N^{-h})$ for $j \in J_1$ and $h > 0$.

Next, we consider PSS introduced by Allen (1971) as an estimate of the mean squared error of prediction in multiple regression. In the general case, cross-validation was derived from an idea similar to PSS, and Stone (1977) proved the asymptotic equivalence of choice of model by AIC and cross-validation *under the null hypothesis* and some restrictions. Our next purpose is to give the asymptotic equivalence of AIC and PSS under weaker conditions than those of Stone (1977) in the context of multiple regression.

ASSUMPTION 2. Let $c_\alpha^{(N)} \ge 0$ be diagonal elements of $Q(\bar{j}_K)$. Then

$$\lim_{N\to\infty}\max\{c_1^{(N)}, \cdots, c_N^{(N)}\} = 0.$$

This assumption implies that, for any $j \in J$,

$$\lim_{N\to\infty}\max\{c_1^{(N)}(j), \cdots, c_N^{(N)}(j)\} = 0,$$

where $c_\alpha^{(N)}(j) \ge 0$ are diagonal elements of $Q(j)$.

Under Assumption 2, we have the following two lemmas:

LEMMA 2. (a). *For* $j \in J$, *let* $\Lambda(j)$ *be a diagonal matrix of order* N *whose diagonal elements are given by* $c_\alpha^{(N)}(j)$ *and let* $A(j) = \{I_N - Q(j)\}\{I_N - \Lambda(j)\}^{-2} \cdot \{I_N - Q(j)\}$. *Then*

$$\lim_{N\to\infty} N^{-1}X'A(j)X = M - MD_j(D_j' MD_j)^{-1}D_j' M.$$

(b) *For* $j \in J_1$, *let*

$$\mu_N = 4N^{-1}\sigma^2\beta'X'A(j)^2X\beta > 0$$

and

$$\mu = 4\sigma^2\beta'\{M - MD_j(D_j'MD_j)^{-1}D_j'M\}\beta > 0.$$

Then $\lim_{N\to\infty}\mu_N = \mu$.

LEMMA 3. *Let* $j \in J_2$. *Then* $\mathrm{PSS}(j_0) - \mathrm{PSS}(j)$ *converges in law to* $\xi_j^{(2)}$ *as* $N \to \infty$, *where* $\xi_j^{(2)}$ *is given in* (4.1).

PROOF. It is sufficient to prove that $\mathrm{PSS}(j) - \mathrm{FPE}(j) \to_P 0$ as $N \to \infty$ for $j \in J_2$ and $a = 2$ in $\mathrm{FPE}(j)$, because $\mathrm{FPE}(j_0) - \mathrm{FPE}(j) \to_L \xi_j^{(2)}$. Set $\Delta(j) = \{I_N - \Lambda(j)\}^{-2} - I_N - 2\Lambda(j)$. For simplicity, we treat j as fixed and omit it from the notations. Then

$$\mathrm{PSS}(j) - \mathrm{FPE}(j) = 2[\mathbf{e}'(I - Q)\Lambda(I - Q)\mathbf{e} - \{k(j) - 1\}N^{-1}\mathbf{e}'(I - Q)\mathbf{e}] + \mathbf{e}'(I - Q)\Delta(I - Q)\mathbf{e}.$$

By large-number theory, we have $\text{p-lim}_{N\to\infty} N^{-1}\mathbf{e}'(I_N - Q)\mathbf{e} = \sigma^2$. We shall prove

$$U_N = \sigma^{-2}\mathbf{e}'(I_N - Q)\Lambda(I_N - Q)\mathbf{e} \to_P k(j) - 1$$

and

$$V_N = \sigma^{-2}\mathbf{e}'(I_N - Q)\Delta(I_N - Q)\mathbf{e} \to_P 0.$$

Assumption 2 yields $E[U_N] = \mathrm{tr}\Lambda - \mathrm{tr}\, Q\Lambda = k(j) - 1 + o(1)$, $V[U_N] = \mathrm{tr}(I - Q)\Lambda(I - Q)\Lambda(I - Q) = o(1)$, $E[V_N] = o(1)$ and $V[V_N] = o(1)$. Let W be a random variable such that $\Pr\{W > 0\} = 1$. Then $\Pr\{W > w\} < E[W]/w$ for $w > 0$. Putting $W = \{U_N - k(j) + 1\}^2$ or V_N^2, we have the required statements.

THEOREM 3. *Under Assumption* 2, *the asymptotic properties of* $\{p_N(j)\}$ *and* R_N *for* PSS *are the following*:

(i) (a) *For* $j \in J_1$ *and any constant* $h > 0$, $p_N(j) = o(N^{-h})$.
  (b) *For* $j \in J_2$, $p_N(j)$ *converges to* $p(j)$ *defined in* (4.2) *with* $a = 2$.
(ii) *The risk function* R_N *converges to* R *defined in* (4.3) *with* $a = 2$.

PROOF. (1) (a). For $j \in J_1$, we have

$$p_N(j) \le \Pr\{\mathrm{PSS}(j) \le \mathrm{PSS}(j_0)\} = \Pr\{X + Y_N + N^{1/2}e_N \le 0\}$$

where

$$X = 2(\mu_N N)^{-1/2}\mathbf{e}'A(j)X\beta \sim N(0, 1),$$

$$Y_N = (\mu_N N)^{-1/2}\mathbf{e}'\{A(j) - A(j_0)\}\mathbf{e} \to_P 0,$$

$$e_N = \mu_N^{-1/2}N^{-1}\beta'X'A(j)X\beta > 0,$$

μ_N is defined in Lemma 2, and $\lim_{N\to\infty} e_N = (4\sigma^2)^{-1}\mu^{1/2} > 0$ by Lemma 2. Applying the same technique as in the proof of Theorem 2, we have

$$\Pr\{X \le -Y_N - N^{1/2}e_N\} \le \Pr\{X \le -N^{1/2}c_N + N^{1/4}\} + \Pr\{-Y_N > N^{1/4}\}.$$

Hence $p_N(j)$ is $o(N^{-h})$ for $j \in J_1$ and for any positive constant h. Lemma 3 yields (i) (b) and (ii).

As we have seen in previous theorems, criteria AIC, FPE, C_p and PSS are asymptotically equivalent under some conditions. Asymptotically, these criteria have positive probability of selecting models that properly include the true model. However, GIC obtained by a generalization of BIC, has slightly different asymptotic properties: GIC is a "consistent" estimator of the true model as follows:

THEOREM 4. (Asymptotic properties of $\{p_N(j)\}$ and R_N for GIC).

(i) (a) *Let $j \in J_1$. Then $p_N(j) = o(N^{-h})$ for any positive constant h.*
(b) *Let $j \in J_2 - \{j_0\}$. Then $p_N(j) = o(1)$.*

(ii) *The risk function R_N converges to $k_0\sigma^2$ as $N \to \infty$.*

PROOF. (i) (a). For $j \in J_1$,

$$p_N(j) \le \Pr\{X + Y_N + N^{1/2}c_N - \lambda_N^{-1/2}b_N^* N^{-1/2} \le Z_N^*\}$$

where X, Y_N, c_N, p and λ_N are defined in the proof of Theorem 2,

$$Z_N^* = \lambda_N^{-1/2}b_N^* N^{-3/2}[\mathbf{e}'\{I_N - Q(j_0)\}\mathbf{e} - \sigma^2 N],$$

and $b_N^* = N\{\exp(pN^{-1}a_N) - 1\} = O(a_N)$. Hence $p_N(j) = o(N^{-h})$ for any $h > 0$.

(i) (b). For $j \in J_2 - \{j_0\}$, we have

$$p_N(j) \le \Pr\{\text{GIC}(j) \le \text{GIC}(j_0)\} = \Pr\{\chi \ge N^{-1}b_N^*\chi_N\}$$

$$\le \Pr\{\chi \ge b_N^*(1 - a_N^{-1/2})\} + \Pr\{\chi_N \le N(1 - a_N^{-1/2})\}$$

where $p = k(j) - k(j_0) > 0$, $\chi = \sigma^{-2}N\{\hat\sigma^2(j_0) - \hat\sigma^2(j)\} = \sigma^2\mathbf{e}'\{Q(j) - Q(j_0)\}\mathbf{e} \sim \chi_p^2$ and $\chi_N = \sigma^{-2}N\hat\sigma^2(j) = \sigma^{-2}\mathbf{e}'\{I_N - Q(j)\}\mathbf{e} \sim \chi^2_{N-k(j)+1}$. The assumption that $a_N \to \infty$ yields $\lim_{N\to\infty} b_N^*(1 - a_N^{-1/2}) = \infty$. Therefore $\Pr\{\chi \ge pb_N^*(1 - a_N^{-1/2})\} = o(1)$. By an inequality on a chi-squared distribution, $\Pr\{\chi_k^2 \le k - \delta\} \le \exp\{-(4k)^{-1}\delta^2\}$ for $\delta > 0$ (see Shibata, 1981), so we have

$$\Pr\{\chi_N < N - Na_N^{-1/2}\} \le \exp(-\tfrac14 Na_N^{-1}) = o(1)$$

since $\lim_{N\to\infty} N^{-1}a_N = 0$. Hence $p_N(j)$ is $o(1)$.

(ii) From (i), GIC satisfies Conditions 1 and 2. By the Remark following Theorem 1, we have the required result, completing the proof.

As we have seen in Theorems 2–4 and a Remark, all criteria discussed in this paper have $\Pr\{\hat{j} = j\} = o(N^{-h})$ for $j \in J_1$ and $h > 0$. This result implies that

$$\Pr\{\hat{j} = j\} = \Pr\{\text{CRITERION}(j) \le \text{CRITERION}(\ell) \text{ for } \ell \in J_2\} + o(N^{-h}),$$

$$R_N = \textstyle\sum_{j \in J_2} R_N(j) + o(N^{-h}).$$

Hence asymptotic expansions of $\Pr\{\hat{j} = j\}$ and R_N are given by those of

$$\Pr\{\text{CRITERION}(j) \le \text{CRITERION}(\ell) \text{ for } \ell \in J_2\} \quad \text{or} \quad \textstyle\sum_{j \in J_2} R_N(j).$$

The asymptotic properties of the criteria may be studied by these expansions, if possible.

Acknowledgement. I would like to express my deepest gratitude to Professor Y. Fujikoshi for suggesting this study and for helpful discussions, and also to the referee for his valuable comments. This research was supported in part by Grant-in-Aid for Encouragement of Young Scientists of the Ministry of Education of Japan.

REFERENCES

AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22** 203–217.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* 267–281. (B. N. Petrov and F. Czáki, eds.). Akademiai Kiadó, Budapest.

ALLEN, D. (1971). The prediction sum of squares as a criterion for selecting prediction variables. Univ. of Kentucky, Dept. of Statistics, Technical Report, No. 23.

FUJIKOSHI, Y. (1982). Selection of variables in two-group discriminant analysis by error rate and Akaike's information criteria. Unpublished manuscript.

HASHIMOTO, A., HONDA, M., INOUE, T. and TAGURI, M. (1981). Selection of regression models by several information criterions. *Rep. Statist. Appl. Res., JUSE* **28** 57–72.

HOCKING, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32** 1–49.

MALLOWS, C. L. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126.

SHIBATA, R. (1981). An optimal selection of regression variables. *Biometrika* **68** 45–54.

STONE, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. Roy. Statist. Soc.* B. **39** 44–47.

STONE, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *J. Roy. Statist. Soc.* B **41** 276–278.

SUGIURA, N. (1978). Numerical experiment on information criteria (in Japanese). Reported in the 11th Symposium on Statistical Data Analysis at Osaka University, organized by M. Okamoto, Osaka Univ.

THOMPSON, M. L. (1978a). Selection of variables in multiple regression: Part I. A review and evaluation. *Internat. Statist. Rev.* **46** 1–19.

THOMPSON, M. L. (1978b). Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples. *Internat. Statist. Rev.* **46** 126–146.

DEPARTMENT OF MATHEMATICS
HIROSHIMA UNIVERSITY
HIGASHI-SENDA-MACHI
HIROSHIMA
730, JAPAN