Ann. Inst. Statist. Math.
34 (1982), Part A, 123-133

LOCAL ASYMPTOTIC ADMISSIBILITY OF A GENERALIZATION OF AKAIKE'S MODEL SELECTION RULE*

CHARLES J. STONE

(Received Dec. 3, 1980; revised Sept. 29, 1981)

Summary

A model selection rule of the form minimize [$-2\log$(maximized likelihood)+complexity] is considered, which is equivalent to Akaike's minimum AIC rule if the complexity of a model is defined to be twice the number of independently adjusted parameters of the model. Under reasonable assumptions, when applied to a locally asymptotically normal sequence of experiments, the model selection rule is shown to be locally asymptotically admissible with respect to a loss function of the form [inaccuracy+complexity], where the inaccuracy is defined as twice the Kullback-Leibler measure of the discrepancy between the true model and the fitted version of the selected model.

1. Introduction

Let Θ_0 be an open subset of R^d, where $1 \leqq d < \infty$. Consider a sequence $\{(\Omega_n; P_{n,\theta}, \theta \in \Theta_0)\}$ of experiments in which each $P_{n,\theta}$ is absolutely continuous with respect to a reference measure μ_n on Ω_n, and let $p_n(\theta)=p_n(\cdot;\theta)$ denote the density of $P_{n,\theta}$ with respect to μ_n. Let $\mathcal{K}$ be a finite set of integers such that $0 \in \mathcal{K}$ and let Θ_k, $k \in \mathcal{K}\setminus\{0\}$, be distinct proper submanifolds of Θ_0. Let $\bar{\theta}_{nk}$ be, say, the maximum likelihood estimator of θ restricted to Θ_k and based on the nth experiment.

A *model selection rule* $\hat{k}_n$, based on the nth experiment, is a $\mathcal{K}$-valued random variable on Ω_n. Let $\bar{k}_n$ be that rule which chooses $k \in \mathcal{K}$ to minimize

$$-2\log p_n(\bar{\theta}_{nk})+C_k\,,$$

* This research was supported by NSF Grant No. MCS 80-02732.
AMS 1980 subject classifications: Primary 62F12; Secondary 62C15.
Key words and phrases: Minimum AIC, complexity, local asymptotic normality, local asymptotic admissibility.

any tie breaking rule being allowed. Here the numbers C_k, $k \in \mathcal{K}$, are real-valued constants; e.g. $C_k = c \dim \Theta_k$ for some $c > 0$. In particular if $C_k = 2 \dim \Theta_k$, $k \in \mathcal{K}$, then $\bar{k}_n$ is the minimum AIC rule introduced by Akaike [1], [2], which has been the subject of much recent literature by Akaike and others.

For $\theta, \tau \in \Theta_0$, let

$$l_n(\theta, \tau) = 2 \, \mathrm{E}_{n,\theta} \log \left(p_n(\theta)/p_n(\tau)\right)$$

denote twice the Kullback-Leibler measure of the discrepancy between the true probability measure $P_{n,\theta}$ and the alternative measure $P_{n,\tau}$. Then $l_n(\theta, \tau) \geqq 0$, with equality holding if and only if $P_{n,\theta} = P_{n,\tau}$. The purpose of this paper is to prove (in particular) that under reasonable assumptions, $\{\bar{k}_n\}$ is locally asymptotically admissible with respect to the loss

$$L_n(\theta, k) = l_n(\theta, \bar{\theta}_{nk}) + C_k \, .$$

The first term, $l_n(\theta, \bar{\theta}_{nk})$, of this loss is a measure of the inaccuracy of the fitted version of the selected model. It is asymptotically equivalent to a quadratic loss function (see Section 3). The second term, C_k, can take into account a variety of attributes of the selected model. One such attribute is the cost of measuring the variables required to implement the model—say, a multiple linear regression model. Measurement cost, which was emphasized by Lindley [15], is especially relevant in certain applications; e.g., in medical diagnostics, for a given accuracy, tomography is preferable to exploratory surgery. A second attribute is the complexity of the selected model. The general principle that for a given level of accuracy a simpler or more parsimonious model is preferable to a more complex one is known as Occam's Razor. The extent to which model complexity should be incorporated into the loss function depends on the enlightened preferences of the decision maker (see von Neumann and Morgenstern [23]). In this connection Blalock states on page 8 of [5] that

> The dilemma of the scientist is to select models that are at the same time simple enough to permit him to think with the aid of the model but also sufficiently realistic that the simplifications required do not lead to predictions which are highly inaccurate. The more complex the model, the more difficult it becomes to decide exactly which modification to make and which new variables to introduce. Put simply, the basic dilemma faced in all sciences is that of how much to over-simplify reality.

Tukey [22], Anderson [3], Kiefer [12], Demster [8], Box [6] and Faden

and Rausser [9] express similar preferences. On the other hand Bunge [7] argues against giving much weight to the desire for simplicity in the pursuit of scientific truth. No particular form or interpretation of the numbers C_k, $k \in \mathcal{K}$, is required for the theoretical results which follow. The additive form, $l_n(\theta, \bar{\theta}_{nk})+C_k$, for combining the two types of losses is simple and reasonable; for an extensive discussion of this issue see Chapter 3 of Keeney and Raiffa [11].

The local asymptotic admissibility result is described in Section 3 and proven in Section 4. The description and proof both depend on a corresponding admissibility result for a normal limit experiment, which will be summarized in convenient form in Section 2.

Although inclusion of the C_k's in the loss function has been shown to be practically justifiable, it is by no means traditional. For the usual quadratic loss function (i.e., with all the C_k's set equal to zero), Theorems 1 and 2 below reduce to the rather uninteresting result that the rule which selects the largest model is admissible or locally asymptotically admissible among all model selection rules. Recently Shibata [17] and Taniguchi [21] derived a much more interesting asymptotic optimality property of the minimum AIC rule corresponding to the usual quadratic loss function, which this rule was designed to handle. But the optimality was obtained in specialized settings involving infinite *nested* sequences of models. It is unlikely that a similar optimality result could be obtained for general *nonnested* collections of models.

2. Admissible model selection rules for a normal experiment

Consider an experiment $(\Omega; P_v, v \in R^d)$ on which there is defined a d-dimensional sufficient statistic T having the following properties: Under P_0, T has a multivariate normal distribution with mean 0 and nonsingular covariance matrix denoted by $\mathcal{J}^{-1}$ (in short $\mathcal{L}_0(T)=N(0, \mathcal{J}^{-1})$. For $v \in R^d$, P_v is absolutely continuous with respect to P_0 and has density $p(v)=p(\cdot; v)=g(T; v)$, where

$$g(t; v)=\frac{\exp\left[-(t-v)\cdot\mathcal{J}(t-v)/2\right]}{\exp\left(-t\cdot\mathcal{J}t/2\right)}=\exp\left(t\cdot\mathcal{J}v-\frac{v\cdot\mathcal{J}v}{2}\right);$$

here "$\cdot$" denotes the usual Euclidean inner product on R^d. As a consequence, $\mathcal{L}_v(T)=N(v, \mathcal{J}^{-1})$ for $v \in R^d$. It is also supposed that there is a random variable U on Ω which under P_0 is independent of T and uniformly distributed on $[0, 1]$.

Let H denote a positive definite symmetric $d \times d$ matrix. Consider the inner product norm $\| \ \|$ on R^d defined by $\|v\|^2=v\cdot Hv$. If $H=\mathcal{J}$ and $v, w \in R^d$, then

$$\|v-w\|^2=2\,\mathrm{E}_v \log\,(p(v)/p(w))\,,$$

which is twice the Kullback-Leibler measure of the discrepancy between P_v and P_w.

Let $\mathcal{K}_0$ be a finite set of integers such that $0\in\mathcal{K}_0$. Let V_k, $k\in\mathcal{K}_0\backslash\{0\}$, be distinct proper subspaces of $V_0=R^d$. For $k\in\mathcal{K}_0$ let $v_k(\cdot)$ denote the orthogonal projection of V_0 onto V_k relative to the norm $\|\ \|$. Then $v_0(\cdot)$ is the identity transformation on V_0 and $v_0(T)=T$ is the maximum likelihood estimator of v. For $k\in\mathcal{K}_0$ the function $v_k(\cdot)$ is linear and if $H=\mathcal{J}$, then $v_k(T)$ is the maximum likelihood estimator of v restricted to V_k.

A model selection rule $\hat{k}$ for this experiment is a $\mathcal{K}_0$-valued random variable on Ω. Two such model selection rules $\hat{k}$ and k^* are said to be *equivalent* ($\hat{k}\equiv k^*$) if $P_0(\hat{k}=k^*)=1$, in which case $P_v(\hat{k}=k^*)=1$ for $v\in V_0$. Let C_k, $k\in\mathcal{K}_0$, denote real-valued constants. Consider the loss function

$$L(v,k)=\|v_k(T)-v\|^2+C_k\,,\qquad v\in V_0 \text{ and } k\in\mathcal{K}_0\,.$$

Define the risk function for a model selection rule $\hat{k}$ by

$$R(v,\hat{k})=\mathrm{E}_v\,L(v,\hat{k})\,,\qquad v\in V_0\,.$$

Observe that this risk function is everywhere finite and that if $\hat{k}\equiv k^*$, then $R(\cdot,\hat{k})=R(\cdot,k^*)$. A model selection rule k^* is called admissible if there is no model selection rule $\hat{k}$ such that $R(\cdot,\hat{k})\leqq R(\cdot,k^*)$ and $R(v,\hat{k})<R(v,k^*)$ for some $v\in V_0$.

Let $\bar{k}=\bar{k}(T)$ be the model selection rule which chooses $k\in\mathcal{K}_0$ to minimize

$$L(T,k)=\|v_k(T)-T\|^2+C_k\,;$$

ties, which occur with probability zero, can be broken, say, by minimizing k. Then $\bar{k}(\cdot)$ is continuous almost everywhere with respect to $L_v(T)$ for each $v\in V_0$. If $H=\mathcal{J}$, the model selection rule $\bar{k}$ chooses $k\in\mathcal{K}_0$ to minimize $-2\log p(v_k(T))+C_k$.

The following result, which implies in particular that $\bar{k}$ *is admissible*, is contained in Theorem 1′ of Stone [19]. (For a stronger result when $d\leqq 2$, see Theorem 1 of Stone [20].)

THEOREM 1. *Let $\hat{k}$ be any model selection rule. Then $R(\cdot,\hat{k})\leqq R(\cdot,\bar{k})$ if and only if $\hat{k}=\bar{k}$, in which case $R(\cdot,\hat{k})=R(\cdot,\bar{k})$.*

3. Locally asymptotically admissible model selection rules

Consider again the original sequence $\{(\Omega_n; P_{n,\theta}, \theta \in \Theta_0)\}$ of experiments. Let $\theta_0 \in \Theta_0$ be fixed and let the reference measure μ_n be defined to be P_{n,θ_0} (so that $p_n(\theta_0)=1$). Random variables Y_n and Z_n (possibly multidimensional) on Ω_n for $n \geqq 1$ are said to be *locally asymptotically equivalent* ($Y_n \overset{a}{=} Z_n$) if $\mathcal{L}_{n,\theta_0}(|Y_n - Z_n|) \to \mathcal{L}(0)$ (in the sense of weak convergence); Y_n is said to be *locally asymptotically infinite* ($Y_n \overset{a}{=} \infty$) if

$$\lim_n P_{n,\theta_0}(Y_n \geqq M) = 1 \qquad \text{for all } M < \infty .$$

It is supposed that there are R^d-valued random variables T_n on Ω_n for $n \geqq 1$ such that

$$\mathcal{L}_{n,\theta_0}(T_n) \to \mathcal{L}_0(T) \tag{1}$$

and

$$p_n(\theta_0 + b_n v) \overset{a}{=} g(T_n; v), \qquad v \in V_0, \tag{2}$$

where $\{b_n\}$ is a fixed sequence of positive constants. It follows from (1) and (2), which together correspond to the *local asymptotic normality* condition of Hájek [10], that if $Y_n \overset{a}{=} Z_n$, then $\mathcal{L}_{n,\theta_0+b_n v}(|Y_n - Z_n|) \to \mathcal{L}(0)$ for all $v \in V_0$ (see Lemma 2 below).

Let $\bar{\theta}_{nk}$, $k \in \mathcal{K}$, be Θ_k-valued estimators of θ based on the nth experiment (i.e., Θ_k-valued random variables on Ω_n). Let $\mathcal{K}_0$ now be defined in terms of θ_0 by $\mathcal{K}_0 = \{k \in \mathcal{K} : \theta_0 \in \Theta_k\}$. It is supposed that

$$b_n^{-1}(\bar{\theta}_{nk} - \theta_0) \overset{a}{=} v_k(T_n), \qquad k \in \mathcal{K}_0 . \tag{3}$$

Let $\bar{l}_n(\theta, \tau) = \bar{l}_n(\omega_n; \theta, \tau)$, $\omega_n \in \Omega_n$ and $\theta, \tau \in \Theta_0$, be a jointly measurable function of ω_n, θ and τ. It is supposed that

$$\bar{l}_n(\bar{\theta}_{n0}, \bar{\theta}_{nk}) \overset{a}{=} b_n^{-2} \|\bar{\theta}_{nk} - \bar{\theta}_{n0}\|^2, \qquad k \in \mathcal{K}_0 , \tag{4}$$

and

$$\bar{l}_n(\bar{\theta}_{n0}, \bar{\theta}_{nk}) \overset{a}{=} \infty, \qquad k \in \mathcal{K} \setminus \mathcal{K}_0 . \tag{5}$$

Let C_k, $k \in \mathcal{K}$, denote real-valued constants and set

$$\bar{L}_n(\theta, k) = \bar{l}_n(\theta, \bar{\theta}_{nk}) + C_k, \qquad \theta \in \Theta_0 \text{ and } k \in \mathcal{K} .$$

Let $\bar{k}_n$ be the model selection rule based on the nth experiment which chooses $k \in \mathcal{K}$ to minimize

$$\bar{L}_n(\bar{\theta}_{n0}, k) = \bar{l}_n(\bar{\theta}_{n0}, \bar{\theta}_{nk}) + C_k ,$$

any tie-breaking rule being allowed. In particular if

$$\bar{l}_n(\theta, \tau)=2 \log (p_n(\theta)/p_n(\tau)), \qquad \theta, \tau \in \Theta_0,$$

then $\bar{k}_n$ chooses $k \in \mathcal{K}$ to minimize

$$2 \log (p_n(\bar{\theta}_{n0})/p_n(\bar{\theta}_{nk}))+C_k$$

or equivalently to minimize

$$-2 \log p_n(\bar{\theta}_{nk})+C_k;$$

so that the definition of $\bar{k}_n$ is equivalent to that given in Section 1. Let $l_n(\theta, \tau)$, $\theta, \tau \in \Theta_0$, be a function which is measurable in τ. It is supposed that for each $v \in V_0$,

$$(6) \qquad l_n(\theta_0+b_n b, \bar{\theta}_{nk}) \stackrel{a}{=} b_n^{-2}\|\bar{\theta}_{nk}-\theta_0-b_n v\|^2, \qquad k \in \mathcal{K}_0,$$

and

$$(7) \qquad l_n(\theta_0+b_n v, \bar{\theta}_{nk}) \stackrel{a}{=} \infty, \qquad k \in \mathcal{K} \backslash \mathcal{K}_0.$$

Set

$$L_n(\theta, k)=l_n(\theta, \bar{\theta}_{nk})+C_k, \qquad \theta \in \Theta_0 \text{ and } k \in \mathcal{K}.$$

To obtain examples where (1)–(7) hold, let the nth experiment consist of the first n trials in an i.i.d. sequence, let $\mathcal{J}=\mathcal{J}(\theta_0)$ be the Fisher information matrix for a single trial evaluated at θ_0 and assumed to be nonsingular, and set $b_n=n^{-1/2}$ for $n \geqq 1$. Suppose that, for $k \in \mathcal{K}_0$, V_k is the tangent space to Θ_k at θ_0 and that, for $k \in \mathcal{K} \backslash \mathcal{K}_0$, θ_0 is not in the closure of Θ_k. Then under mild regularity conditions (see Rao [16]) (1)–(7) hold where $\bar{\theta}_{nk}$, $\bar{l}_n$ and l_n are defined as follows: $\bar{\theta}_{n0} \in \Theta_0$ is a consistent maximum likelihood estimator of θ_0; for $k \in \mathcal{K} \backslash \{0\}$, $\bar{\theta}_{nk}$ is chosen in Θ_k to minimize $\|\bar{\theta}_{nk}-\bar{\theta}_{n0}\|^2$; and

$$\bar{l}_n(\theta, \tau)=l_n(\theta, \tau)=b_n^{-2}\|\tau-\theta\|^2, \qquad \theta, \tau \in \Theta_0.$$

Here $\bar{k}_n$ chooses $k \in \mathcal{K}$ to minimize

$$\bar{L}_n(\bar{\theta}_{n0}, k)=b_n^{-2}\|\bar{\theta}_{nk}-\bar{\theta}_{n0}\|^2+C_k.$$

Suppose additionally that $H=\mathcal{J}$. Then under mild regularity conditions as in [16], (1)–(7) hold where $\bar{\theta}_{nk}$, $\bar{l}_n$ and l_n are defined as follows: for $k \in \mathcal{K}$, $\bar{\theta}_{nk}$ is a maximum likelihood estimator of θ_0 restricted to Θ_k —which is consistent if $k \in \mathcal{K}_0$;

$$\bar{l}_n(\theta, \tau)=2 \log (p_n(\theta)/p_n(\tau)), \qquad \theta, \tau \in \Theta_0$$

(so $\bar{k}_n$ is determined as in the previous paragraph); and

$$l_n(\theta, \tau)=2\,\mathrm{E}_{n,\theta}\log\left(p_n(\theta)/p_n(\tau)\right).$$

Two sequences $\{\hat{k}_n\}$ and $\{k_n^*\}$ of model selection rules (the nth rule being based on the nth experiment) are asymptotically equivalent if and only if $\lim_n P_{n,\theta_0}(\hat{k}_n=k_n^*)=1$. It follows from (1) and (2) (see Lemma 1 below) that if $\hat{k}_n\overset{a}{=}k_n^*$, then

$$\lim_n P_{n,\theta_0+b_nv}(\hat{k}_n=k_n^*)=1\,,\qquad v\in V_0\,.$$

The *local asymptotic risk function* $R_\infty(v, \{\hat{k}_n\})$, $v\in V_0$, is defined by

$$R_\infty(v, \{\hat{k}_n\})=\lim_\alpha \overline{\lim_n}\,\mathrm{E}_{n,\theta_0+b_nv}\min\left[L_n(\theta_0+b_nv, \hat{k}_n), \alpha\right],$$

where $\lim_\alpha$ means $\lim_{\alpha\to\infty}$. It follows from Lemma 3 below that if (1)–(6) hold and $\hat{k}_n\overset{a}{=}\bar{k}_n$, then $\overline{\lim_n}$ can be replaced by $\lim_n$ in the formula defining $R_\infty(v, \{\hat{k}_n\})$.

PROPOSITION 1. Suppose (1)–(6) hold. Then $\bar{k}_n\overset{a}{=}\bar{k}(T_n)$ and $R_\infty(\cdot, \{\bar{k}_n\})=R(\cdot, \bar{k})$.

The sequence $\{k_n^*\}$ of model selection rules is said to be *locally asymptotically admissible* if there is no other sequence $\{\hat{k}_n\}$ such that $R_\infty(\cdot, \{\hat{k}_n\})\leqq R_\infty(\cdot, \{k_n^*\})$ and $R_\infty(v, \{\hat{k}_n\})<R_\infty(v, \{k_n^*\})$ for some $v\in V_0$. The main result of this paper, which follows, implies that *if* (1)–(7) *hold, then* $\{\bar{k}_n\}$ *is locally asymptotically admissible.* (For a stronger result when $d\leqq 2$ see Theorem 2 of Stone [20].)

THEOREM 2. *Suppose* (1)–(7) *hold and let* $\{\hat{k}_n\}$ *be any sequence of model selection rule. Then* $R_\infty(\cdot, \{\hat{k}_n\})\leqq R_\infty(\cdot, \{\bar{k}_n\})$ *if and only if* $\hat{k}_n\overset{a}{=}\bar{k}_n$, *in which case* $R_\infty(\cdot, \{\hat{k}_n\})=R_\infty(\cdot, \{\bar{k}_n\})$.

The proofs of Proposition 1 and Theorem 2 are given in Section 4, the proof of Theorem 2 depending crucially on Theorem 1. The statement and method of proof of Theorem 2 is clearly influenced by the work of Hájek [10] and LeCam [13], [14]. But they do not fit directly into LeCam's general treatment since the loss $L_n(\theta, k)=l_n(\theta, \bar{\theta}_{nk})+C_k$ depends through $\bar{\theta}_{nk}$ on $\omega_n\in\Omega_n$.

4. Proofs

Proposition 1 and Theorem 2 will now be verified.

LEMMA 1. *Suppose* (2) *holds. If* $\mathcal{L}_{n,\theta_0}(T_n, Y_n)\to\mathcal{L}_0(T, Y)$, *then*

$$\mathcal{L}_{n,\theta_0+b_nv}(T_n, Y_n) \to \mathcal{L}_v(T, Y), \qquad v \in V_0.$$

PROOF. Since $g(t; v)$ is continuous in t,

$$\mathcal{L}_{n,\theta_0}(T_n, Y_n, g(T_n; v)) \to \mathcal{L}_0(T, Y, g(T; v)).$$

Thus by (2)

$$\mathcal{L}_{n,\theta_0}(T_n, Y_n, p_n(\theta_0+b_nv)) \to \mathcal{L}_0(T, Y, g(T; v)).$$

Let $\psi(t, y)$ be a bounded continuous function of t and y. Since the random variables $p_n(\theta+b_nv)$, $n \geqq 1$, are uniformly integrable for each $v \in V_0$ (see Theorem 5.4 of Billingsley [4])

$$\begin{aligned}\mathrm{E}_{n,\theta_0+b_nv}\, \psi(T_n, Y_n) &= \mathrm{E}_{n,\theta_0}\, \psi(T_n, Y_n) p_n(\theta_0+b_nv) \\ &\to \mathrm{E}_0\, \psi(T, Y) g(T; v) = \mathrm{E}_v\, (T, Y),\end{aligned}$$

so the desired conclusion holds.

LEMMA 2. *Suppose* (1)-(5) *hold. Then* $\bar{k}_n \overset{a}{=} \bar{k}(T_n)$ *and*

$$\mathcal{L}_{n,\theta_0+b_nv}(T_n, \bar{k}_n) \to \mathcal{L}_v(T, \bar{k}(T)); \qquad v \in V_0.$$

PROOF. Let k_n^* be a model selection rule based on the nth experiment which chooses $k \in \mathcal{K}_0$ minimize $\bar{L}_n(\bar{\theta}_{n0}, k)$ and is such that $k_n^* = \bar{k}_n$ whenever $\bar{k}_n \in \mathcal{K}_0$. Then $k_n^* \overset{a}{=} \bar{k}_n$ by (1) and (3)-(5). Thus to verify the first conclusion it suffices to show that $k_n^* \overset{a}{=} \bar{k}(T_n)$. But this follows from (1), (3), (4), the continuity of $v_k(\cdot)$ for $k \in \mathcal{K}_0$, and the fact that almost surely with respect to $\mathcal{L}_0(T)$ there is a unique $k \in \mathcal{K}_0$ which minimizes $L(T, k)$. Therefore $\bar{k}_n \overset{a}{=} \bar{k}(T_n)$. Since $\bar{k}(\cdot)$ is continuous almost surely with respect to $\mathcal{L}_0(T)$, it follows from (1) that

$$\mathcal{L}_{n,\theta_0}(T_n, \bar{k}(T_n)) \to \mathcal{L}(T, \bar{k}(T)).$$

Thus by Lemma 1

$$\mathcal{L}_{n,\theta_0+b_nv}(T_n, \bar{k}(T_n)) \to \mathcal{L}_v(T, \bar{k}(T)).$$

The second conclusion of the lemma now follows from the first conclusion.

LEMMA 3. *Suppose* (1)-(6) *hold. If* $\hat{k}_n \overset{a}{=} \bar{k}_n$, *then for* $v \in V_0$ *and* $\alpha \geqq 0$

$$\lim_n \mathrm{E}_{n,\theta_0+b_nv} \min [L_n(\theta_0+b_nv, \hat{k}_n), \alpha] = \mathrm{E}_v \min [L(v, \bar{k}(T)), \alpha].$$

PROOF. Let $v \in V_0$ be fixed. By (3) and (6)

$$L_n(\theta_0+b_nv, k) \overset{a}{=} \|v_k(T_n) - v\|^2 + C_k, \qquad k \in \mathcal{K}_0.$$

Thus by Lemma 2,

$$L_n(\theta_0+b_nv, \hat{k}_n)\overset{a}{=}\|v_{\bar{k}(T_n)}(T_n)-v\|^2+C_{\bar{k}(T_n)}.$$

Consequently by (1), the continuity properties of $v_k(\cdot)$ and $\bar{k}(\cdot)$, and Lemma 1,

$$\lim_n \mathrm{E}_{n,\theta_0+b_nv} \min [L_n(\theta_0+b_nv, \hat{k}_n), \alpha]$$
$$=\mathrm{E}_v \min [\|v_{\bar{k}(T)}(T)-v\|^2+C_{\bar{k}(T)}, \alpha]=\mathrm{E}_v \min [L(v, \bar{k}(T)), \alpha]$$

as desired.

Proposition 1 is an immediate consequence of Lemmas 2 and 3.

LEMMA 4. *Suppose* (1)–(7) *hold. If* $R_\infty(\cdot, \{\hat{k}_n\})\leqq R(\cdot, \bar{k}(T))$, *then* $\mathcal{L}_{n,\theta_0}(T_n, \hat{k}_n)\to \mathcal{L}_0(T, \bar{k}(T))$.

PROOF. Since $R_\infty(0, \{\hat{k}_n\})\leqq R(0, \bar{k}(T))<\infty$, it follows from (7) that $\lim_n P_{n,\theta_0}(\hat{k}_n \in \mathcal{K}_0)=1$. Let $\{n_j\}$ be a strictly increasing sequence of positive integers such that $\mathcal{L}_{n_j,\theta_0}(T_{n_j}, \hat{k}_{n_j})$ converges weakly to some probability distribution G on $R^d\times\mathcal{K}_0$. Then there is a $\mathcal{K}_0$-valued random variable $\hat{k}$ on Ω such that $\mathcal{L}_0(T, \hat{k})=G$. (Here the uniformly distributed random variable U described in Section 2 is used.) By Lemma 1

$$\mathcal{L}_{n_j,\theta_0+b_{n_j}v}(T_{n_j}, \hat{k}_{n_j})\to \mathcal{L}_v(T, \hat{k}), \qquad v\in V_0.$$

It follows as in the proof of Lemma 3 that

$$\lim_j \mathrm{E}_{n_j,\theta_0+b_{n_j}v} \min [L_{n_j}(\theta_0+b_{n_j}v, \hat{k}_{n_j}), \alpha]=\mathrm{E} \min [L(v, \hat{k}), \alpha]$$

for $v\in V_0$ and $\alpha\geqq 0$ and hence that

$$R(\cdot, \bar{k}(T))\geqq R_\infty(\cdot, \{\hat{k}_n\})\geqq R(\cdot, \hat{k}).$$

Thus $\hat{k}\equiv\bar{k}(T)$ by Theorem 1 and hence

$$\mathcal{L}_{n_j,\theta_0}(T_{n_j}, \hat{k}_{n_j})\to \mathcal{L}_0(T, \bar{k}(T)).$$

Consequently $\mathcal{L}_{n,\theta_0}(T_n, \hat{k}_n)\to \mathcal{L}_0(T, \bar{k}(T))$ as desired.

LEMMA 5. *Let* ψ *be a function on* R^d *which is continuous almost surely with respect to* $\mathcal{L}_0(T)$. *If* $\mathcal{L}_{n,\theta_0}(T_n, Y_n)\to \mathcal{L}_0(T, \psi(T))$, *then* $Y_n\overset{a}{=}\psi(T_n)$.

PROOF. Choose $\varepsilon>0$. Set

$$A=\{(t, y): |y-\psi(t)|\geqq\varepsilon\}.$$

Then $P_0((T, \psi(T))\in\partial A)=0$, so

$$\begin{aligned} 0 &= P_0((T, \psi(T)) \in A) \\ &= \lim_n P_{n,\theta_0}((T_n, Y_n) \in A) \\ &= \lim_n P_{n,\theta_0}(|Y_n - \psi(T_n)| \geqq \varepsilon) \end{aligned}$$

as desired. (The continuity assumption on ψ can be dropped; but a less elementary argument based, e.g., on Section 3.1.1 of Skorohod [18] is then required.)

Theorem 2 is an immediate consequence of Proposition 1 and Lemmas 3–5.

UNIVERSITY OF CALIFORNIA, BERKELEY

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), Akademiai Kaido, Budapest, 267-281.

[2] Akaike, H. (1974). A new look at statistical model identification, *IEEE Trans. Automat. Contr.*, AC-**19**, 716-723.

[3] Anderson, T. W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem, *Ann. Math. Statist.*, **33**, 255-265.

[4] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[5] Blalock, H. M., Jr. (1961). *Causal Inferences in Nonexperimental Research*, University of North Carolina Press, Chapel Hill.

[6] Box, G. E. P. (1976). Science and statistics, *J. Amer. Statist. Ass.*, **71**, 791-799.

[7] Bunge, M. (1963). *The Myth of Simplicity: Problems of Scientific Philosophy*, Prentice-Hall, Englewood Cliffs.

[8] Demster, A. P. (1971). Model searching and estimation in the logic of inference (with discussion), *Foundations of Statistical Inference* (eds. V. P. Godambe and D. A. Sprott), Holt, Rinehart and Winston, Toronto, 56-81.

[9] Faden, A. M. and Rausser, G. C. (1976). Econometric policy model construction: The post-Bayesian approach, *Ann. Economic. and Social Measurement*, **5**, 349-362.

[10] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation, *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, **1**, 175-194.

[11] Keeney, R. L. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.

[12] Kiefer, J. (1968). *The Future of Statistics: Proceedings of a Conference on the Future of Statistics Held at the University of Wisconsin, Madison, Wisconsin, June 1967* (ed. P. G. Watts), Academic Press, New York, 139-142.

[13] LeCam, L. (1972). Limits of experiments, *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, **1**, 245-261.

[14] LeCam, L. (1979). On a theorem of J. Hájek, *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (ed. J. Jurečková), Academia, Prague, 119-135.

[15] Lindley, D. V. (1968). The choice of variables in multiple regression (with discussion), *J. R. Statist. Soc.*, B, **30**, 31-66.

[16] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd Ed., Wiley, New York.

[17] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Ann. Statist.*, **8**, 147-164.

[18] Skorohod, A. V. (1956). Limit theorems for stochastic processes, *Theory Prob. Appl.*, **1**, 261-299.

[19] Stone, C. J. (1981). Admissible selection of an accurate and parsimonious normal linear regression model, *Ann. Statist.*, **9**, 475-485.

[20] Stone, C. J. (1981). Admissibility and local asymptotic admissibility of procedures which combine estimation and model selection, *Proc. Third Purdue Symp. on Statist. Decision Theory and Related Topics*, to appear.

[21] Taniguchi, M. (1980). On selection of the order of the spectral density model for a stationary process, *Ann. Inst. Statist. Math.*, **32**, A, 401-419.

[22] Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis, *Technometrics*, **3**, 191-219.

[23] von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economics Behavior*, 3rd Ed., Princeton University Press, Princeton.