The Annals of Statistics
1975, Vol. 3, No. 2, 539-546

A NOTE ON THE CONSISTENCY OF MAXIMUM LIKELIHOOD ESTIMATES FOR FINITE FAMILIES OF STOCHASTIC PROCESSES

BY P. E. CAINES[1]

Stanford University

We consider families of stochastic processes indexed by a finite number of alternative parameter values. For general classes of stochastic processes it is shown that maximum likelihood estimates converge almost surely to the correct parameter value. This is established by use of a submartingale property of the sequence of maximized likelihood ratios together with a technique first employed by Wald [24] in the case of independent identically distributed random variables.

1. Introduction, notation and assumptions. The maximum likelihood technique has been one of the principal large sample methods of statistics since its invention by Fisher [14], [15]. Much of the extensive literature on this technique is concerned with the convergence properties of the maximum likelihood estimate (MLE). In a classical paper in 1949 Wald [24] proved the strong consistency of the MLE in the case of independent identically distributed random variables. He used no differentiability assumptions on the density functions involved. This was a considerable contribution since such assumptions were common in the literature, especially in that part of it concerned with the closely related likelihood equation technique (see for example [10], [16]).

Most of the identification problems occurring in system theory involve processes with dependent observations. The particular problem of parameter estimation for finite dimensional linear systems has naturally been of great interest to engineers and statisticians. (See for instance the survey article [4].) Partly following Wald, and Stuart and Kendall [22], many authors have attempted to use the laws of large numbers to establish weak and strong consistency results for linear systems. (See for example [1]—[3], [5]—[7], [12], [23].) The latter result has, in fact, recently been established [8], [19] by using the ergodic theorem, as suggested by Åström and Bohlin [3].

The extension of consistency results to various classes of general, in particular nonlinear, stochastic processes has been studied by several authors (see for instance [17], [18]). Unfortunately the conditions required for the application of these results are frequently difficult to verify in practice. In an elegant paper

Received April 1973; revised April 1974.

[1] Now with Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California, 94720. The work of this author was partially supported by the National Science Foundation Grants GK-10656x3/GK-31627 and the National Aeronautics and Space Administration Grant NGL-05-020-007.

AMS 1970 *subject classifications.* 6215 Asymptotic Theory; 6220 Estimation: Parametric Case.
Key words and phrases. Parameter estimation, maximum likelihood estimation.

in 1965, however, Roussas [20] proved the consistency of the maximum likelihood estimate for stationary ergodic Markov processes. He also established in another paper [21] the asymptotic normality of the MLE. The first result employed Birkhoff's ergodic theorem and the second a martingale property of the differential of the log-likelihood function.

In this note we treat the highly restricted situation where there exist only a finite, albeit arbitrarily large, number of alternative parameter values (including the true value $\mathring{\theta}$) in the set Θ. This finiteness restriction enables us to prove the strong consistency of the MLE for a general class of families of processes. (Namely those satisfying (A1) and (A2) below.) From a theoretical point of view substantial generality in one direction has been purchased by corresponding restrictions in the other. We believe our result has practical significance because implementation of identification algorithms takes place upon computers, and mathematically speaking these are finite objects.

It has long been well known that a sequence of likelihood functions forms a martingale and that this allows one to employ the martingale convergence theorem to study the MLE (see for instance Doob [11] pages 348–350 and Feller [13] pages 211–212). Despite this, the observation that the maximized likelihood ratio (MLR) sequence forms a positive submartingale appears to be new. (Indeed, Doob's assumptions make the sequence of likelihood ratios a positive supermartingale.) Once this property is invoked an extension of Wald's technique to general dependent sample processes is almost immediate. Unfortunately this proof technique cannot be applied directly to families of processes indexed by an infinite number of parameters. This is due to the fact that one of the bounds employed in Section 2 below may fail to hold in such cases.

Consider a measurable space $(\Omega, \mathscr{A})$ and a finite abstract set of parameters of cardinality $N + 1$. Without loss of generality we shall take this set to be the collection of integers $\{0, 1, \cdots, N\}$. Let $\{P_k; 0 \leqq k \leqq N\}$ be a family of probability measures on $\mathscr{A}$. It is assumed that for every $k \in \{0, \cdots, N\}$ $\{Y_n, n \geqq 0\}$ is an $\mathbb{R}^m$-valued stochastic process defined on the probability space $(\Omega, \mathscr{A}, P_k)$. Without loss of generality we may assume $(\Omega, \mathscr{A})$ is the infinite Cartesian product $\prod_{i=0}^{\infty}(\mathbb{R}^m, \mathscr{B})$ where $\mathscr{B}$ is the m-dimensional Borel field in $\mathbb{R}^m$ and P_k is the probability measure induced in $\mathscr{A}$ by a set of (consistent) probability distributions $\{p_{k,n}(\cdot), n \geqq 0\}$ on $\{\prod_{i=0}^{n}(\mathbb{R}^m, \mathscr{B}), n \geqq 0\}$ according to Kolmogorov's consistency theorem.

Let $P_{k,n}$ denote the restriction of P_k to the σ-field $\mathscr{A}_n = \mathscr{B}(Y_0, Y_1, \cdots, Y_n)$, $n \geqq 0$. Throughout the remainder of this note we hold the following

ASSUMPTION (A1). For each $n \geqq 0$ the members of the family $\{P_{k,n}(\cdot), k \in \{0, \cdots, N\}\}$ are mutually absolutely continuous.

As a consequence, outside $P_{j,n}$ null sets for all $j \in \{0, 1, \cdots, N\}$, we have $[dP_{k,n}/dP_{j,n}] = f_{k,j}(Y_0, Y_1, \cdots, Y_n)$ for any $k, j \in \{0, \cdots, N\}$ where $f_{k,j}(\cdot)$ is a specified version of a Radon–Nikodym derivative of the measure $P_{k,n}$ with respect to the measure $P_{j,n}$. We shall arbitrarily assume that 0 is the true parameter in the set $\{0, \cdots, N\}$; in other words the process $\{Y_n, n \geqq 0\}$ is generated according to the measures $\{P_{0,n}; n \geqq 0\}$ or equivalently the distributions $\{p_{0,n}(\cdot); n \geqq 0\}$. By forming each of the likelihoods with respect to, say, $\{dP_{N,n}, n \geqq 0\}$ we may simplify the notation $f_{k,N}(\cdot)$ to read $f_k(\cdot)$.

We also need the following definitions. Any mapping $\hat{k}_n = \hat{k}(Y_0, Y_1, \cdots, Y_n)$ on Ω into a subset $\mathscr{M} \subset \{0, \cdots, N\}$ which is $\mathscr{A}_n$ measurable will be called an *estimate*. An estimate such that $f_{\hat{k}_n}(Y^n) \geqq \max\{f_j(Y^n); j \in \mathscr{M}\}$ is called a *maximum likelihood estimate* (MLE), where Y^n denotes $(Y_0, Y_1, \cdots, Y_n)$. (This integer-valued MLE is obviously the finite parameter specialization of the usual Θ-valued estimate, where Θ is some infinite subset of $\mathbb{R}^\nu$ for some ν.)

We also need to define the notion of *likelihood ratio*. This is the ratio $f_k(Y^n)/f_0(Y^n)$; it has the property that it takes its maximum value at the same subset of $\mathscr{M}$ as does the likelihood function. Let $\mathscr{M} = \{0, \cdots, N\}$. Then we shall call the estimates $\{\hat{k}_n, n \geqq 0\}$ (*strongly*) *consistent* if $\hat{k}_n \neq 0$ infinitely often with P_0-probability 0, i.e. $P_0(\omega \mid \hat{k}_n(\omega) = 0, n \in \{n^0\}) = 1$ and $P_0(\omega \mid \hat{k}_n(\omega) = j, n \in \{n^j\}) = 0$ for $j \in \{1, \cdots, N\}$, where $\{n^j\}$, $j \in \{0, \cdots, N\}$ denote subsequences of $\{n\} = \{0, 1, \cdots\}$.

The family $\{P_{0,n}; n \geqq 0\}$ will be the only family of measures that governs the observed process $\{Y_n; n \geqq 0\}$. The extent to which we require the measures $\{P_{k,n}; n \geqq 0\}$, $1 \leqq k \leqq N$, to differ from the true family $\{P_{0,n}; n \geqq 0\}$ is given by

ASSUMPTION (A2). Given $\varepsilon > 0$ there exists $\alpha(\varepsilon) > 1$ such that for any integer N the event $[0 \leqq h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) < \alpha$, for all $n > N]$ when $\mathscr{M} = \{1, \cdots, N\}$ is of P_0 probability less than ε where

$$h_k(Y_n \mid Y^{n-1}) \triangleq f_k(Y_n \mid Y^{n-1})/f_0(Y_n \mid Y^{n-1}),$$
$$f_k(Y_n \mid Y^{n-1}) = f_k(Y^n)/f_k(Y^{n-1}), \qquad \text{and}$$
$$f_k(Y_0 \mid Y^{-1}) = f_k(Y_0) = f_k(Y^0).$$

By definition

$$f_k(Y^n) = [dP_{k,n}/dP_{N,n}] = \textstyle\prod_{j=0}^n f_k(Y_j \mid Y^{j-1}).$$

Consequently we may write
$f_k(Y^n)/f_0(Y^n) = \prod_{j=0}^n (f_k(Y_j \mid Y^{j-1})/f_0(Y_j \mid Y^{j-1})) = h_k(Y_n \mid Y^{n-1}) \cdot f_k(Y^{n-1})/f_0(Y^{n-1})$ for the kth likelihood ratio at the nth sample. We see that $\{h_k(Y_n \mid Y^{n-1}); 1 \leqq k \leqq N\}$ compares the true conditional probability measure of the next observation with that given by each of the measures $\{P_{k,n}(\cdot \mid \mathscr{A}_{n-1}), 1 \leqq k \leqq N\}$, for $n \geqq 0$. Analogous "prediction conditions" are to be found elsewhere in the literature on maximum likelihood estimation. See for instance [23] for the case of linear systems. Clearly it would be of interest to have a general characterization of the classes of processes satisfying such prediction conditions.

2. Main result. Using the notions defined in the previous section we shall prove the following result.

THEOREM. *Under assumptions* (A1) *and* (A2) *maximum likelihood estimates are strongly consistent.*

PROOF. Let us denote the maximized likelihood ratio of the kth probability density over the 0th probability density with respect to any subset $\mathscr{M} \subset \{0, \cdots, N\}$ by $x^{\mathscr{M}}(Y^n)$. When no confusion is likely to occur we shall abbreviate this to $x(Y^n)$ or x_n.

Then it is clear that

$$0 \leqq h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) \cdot x(Y^{n-1}) = f_{\hat{k}_{n-1}}(Y^n)/f_0(Y^n) \leqq f_{\hat{k}_n}(Y^n)/f_0(Y^n)$$
$$= x(Y^n) .$$

Further

$$x(Y^n) = f_{\hat{k}_n}(Y^n)/f_0(Y^n) = h_{\hat{k}_n}(Y_n \mid Y^{n-1}) \cdot f_{\hat{k}_n}(Y^{n-1})/f_0(Y^{n-1})$$
$$\leqq h_{\hat{k}_n}(Y_n \mid Y^{n-1}) \cdot x(Y^{n-1}) .$$

The latter inequality is included for completeness since we shall only be concerned with the first inequality in the sequel. However, the two inequalities taken together yield the following basic inequality for maximized likelihood ratios:

$$0 \leqq h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) \cdot x_{n-1} \leqq x_n \leqq h_{\hat{k}_n}(Y_n \mid Y^{n-1}) \cdot x_{n-1} . \tag{1}$$

It should be noticed that an analogous inequality holds in the case of a continuum of parameter values Θ under mild conditions e.g. $f_\theta(Y^n)$ continuous in θ with Θ a compact metric space (using an obvious modification of the notion established before).

We shall demonstrate that the MLR inequality above shows the process $\{x^{\mathscr{M}}(Y^n); n \geqq 0\}$ to be a positive submartingale. Let E_0 denote expectation with respect to P_0 measure. Then clearly $x_n \geqq 0$ for $n \geqq 0$ and the submartingale property follows from

$$\begin{aligned} E_0(x_n \mid \mathscr{A}_{n-1}) &\geqq E_0(h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1})x_{n-1} \mid \mathscr{A}_{n-1}) \\ &=_{\text{a.s.}} x_{n-1}E_0(h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) \mid \mathscr{A}_{n-1}) \\ &= x_{n-1}E_0\left(\frac{f_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1})}{f_0(Y_n \mid Y^{n-1})}\middle| \mathscr{A}_{n-1}\right) \\ &=_{\text{a.s.}} x_{n-1}\frac{f_0(Y^{n-1})}{f_{\hat{k}_{n-1}}(Y^{n-1})}E_0\left(\frac{f_{\hat{k}_{n-1}}(Y^n)}{f_0(Y^n)}\middle| \mathscr{A}_{n-1}\right) . \end{aligned} \tag{2}$$

Next, by setting,

$$I(Y^{n-1}) = E_0\left(\frac{f_{\hat{k}_{n-1}}(Y^n)}{f_0(Y^n)}\middle| \mathscr{A}_{n-1}\right), \qquad A_j = (\hat{k}_{n-1} = j) ,$$

one has

$$I(Y^{n-1}) =_{\text{a.s.}} \textstyle\sum_j I_{A_j}E_0\left(\frac{f_j(Y^n)}{f_0(Y^n)}\middle| \mathscr{A}_{n-1}\right) . \tag{3}$$

But

$$E_0\left(\frac{f_j(Y^n)}{f_0(Y^n)}\,\Big|\,\mathscr{A}_{n-1}\right) = E_0\left(\frac{f_j(Y_n \mid Y^{n-1})f_j(Y^{n-1})}{f_0(Y_n \mid Y^{n-1})f_0(Y^{n-1})}\,\Big|\,\mathscr{A}_{n-1}\right)$$

$$=_{\text{a.s.}} \frac{f_j(Y^{n-1})}{f_0(Y^{n-1})} E_0\left(\frac{f_j(Y_n \mid Y^{n-1})}{f_0(Y_n \mid Y^{n-1})}\,\Big|\,\mathscr{A}_{n-1}\right) = \frac{f_j(Y^{n-1})}{f_0(Y^{n-1})} .$$

Then (3) gives

$$I(Y^{n-1}) =_{\text{a.s.}} \textstyle\sum_j I_{A_j} \dfrac{f_j(Y^{n-1})}{f_0(Y^{n-1})} = \dfrac{f_{\hat{k}_{n-1}}(Y^{n-1})}{f_0(Y^{n-1})}$$

and the desired result then follows from (2).

Now the inequality

$$E_0(x^{\mathscr{M}}(Y^n)) \leqq \textstyle\sum_{k=0}^{N} E_0(f_k(Y^n)/f_0(Y^n)) = N + 1$$

shows that for any subset $\mathscr{M} \subset \{0, \cdots, N\}$ the sequence of expectations of the positive submartingale $\{x_n;\, n \geqq 0\}$ is uniformly bounded. (It is this bound which may fail when the parameter set is infinite.) It immediately follows from the submartingale convergence theorem ([11], pages 324–328, [9], pages 307–308) that $\{x_n;\, n \geqq 0\}$ converges almost surely to a limiting random variable which shall be denoted by x^*. Although this is true for any subset $\mathscr{M}$ of $\{0, \cdots, N\}$, it is convenient to assume $\mathscr{M} = \{1, \cdots, N\}$ from this point on in the proof. The notation $\{x_n;\, n \geqq 0\}$ should be interpreted accordingly.

We now employ a technique originally developed by Wald [24] to obtain the main result. Suppose it was not the case that $\hat{k}_n$ with range $\{0, \cdots, N\}$ took only the value 0 infinitely often with P_0 probability 1. Then the event [$\max\{f_j(Y^n);\, j \in \{1, \cdots, N\}\} \geqq \max\{f_j(Y^n);\, j \in \{0, \cdots, N\}\}$, infinitely often] would have some nonzero probability with respect to the measure P_0. Notice that one possible value of the MLR with respect to $\{0, \cdots, N\}$ is 1 for each n. Consequently if the MLE $\hat{k}_n$ is not consistent the event [$\max\{f_j(Y^n)/f_0(Y^n);\, j \in \{1, \cdots, N\}\} \geqq 1$, infinitely often] would have nonzero P_0 measure. It follows that the consistency of $\hat{k}_n$ would be established by contradiction if the event [$x_n \geqq 1$, infinitely often] had P_0 probability 0. In fact we now prove the stronger statement that under assumptions (A1) and (A2) $x_n \to x^* = 0$, as $n \to \infty$, almost surely $[P_0]$.

Assume $x^* = 0$ with P_0 probability $1 - 3\varepsilon$, $\varepsilon > 0$. Then referring to assumption (A2) choose $\alpha(\varepsilon) > 1$ sufficiently small that for some M_ε the event [$0 \leqq h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) < \alpha$, for all $n > M_\varepsilon$] is of P_0 probability less than ε.

Now by the continuity of probability measures there exists $\delta(\varepsilon) > 0$ such that

$$P_0\left[x^* > \frac{\alpha + 1}{\alpha - 1}\,\delta\right] > 2\varepsilon .$$

The convergence of $\{x_n\}$ to x^* implies the existence of $N(\delta(\varepsilon), \varepsilon) \geqq M_\varepsilon$, denoted N_ε, such that

$$P_0[\sup_{n \geq N_\varepsilon} |x_n - x^*| < \delta] > 1 - \varepsilon .$$

Hence

$$P_0\left[\sup_{n \geq N_\varepsilon} |x_n - x^*| < \delta;\; x^* > \frac{\alpha + 1}{\alpha - 1}\,\delta\right] > \varepsilon .$$

Together with the left-hand side of (1) this statement implies the event

$$\left[0 \leqq h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) \leqq \frac{x_n}{x_{n-1}} \leqq \frac{x^* + \delta}{x^* - \delta} < \alpha, \quad \text{for all} \quad n > N_\varepsilon\right]$$

is of P_0 probability greater than ε. This contradicts assumption (A2), given our choice of α, and we conclude $x^* = 0$ almost surely $[P_0]$. Consequently we have proved the main theorem.

3. Example. Consider the real-valued scalar process $\{Y_n; n \geqq 1\}$ generated by the recurrence equation

$$Y_{n+1} = F_n(Y^n) + W_{n+1} \qquad n \geqq 0 \tag{4}$$

where Y^n denotes $(Y_0, Y_1, \cdots, Y_n)$, Y_0 is given, $\{W_n; n \geqq 1\}$ is a real-valued i.i.d. process and $\{F_n(\cdot); n \geqq 0\}$ is a family of measurable functions from $\mathbb{R}^{n+1}$ to $\mathbb{R}$ for each n. Let W_n take the set of values $\{w_i; i \geqq 1\}$ with probabilities $\{p_i^0; i \geqq 1\}$ with respect to the true measure P_0 and with probabilities $\{p_i^1; i \geqq 1\}$ with respect to an alternative measure P_1. We assume that $0 < p_i^0 < 1$ and $0 < p_i^1 < 1$ for all i and that $p_i^0 \neq p_i^1$ for at least one i.

We remark that the process generated by the recurrence relation (4) is clearly not Gaussian nor necessarily Markovian.

Assumption (A1) is immediately satisfied in this example. If the prediction condition (A2) is also satisfied we are assured by the theorem in Section 2 that the maximum likelihood estimate $\hat{k}_n$ is strongly consistent at the true parameter 0. This is straightforward to verify. The set of parameters excluding 0 consists only of the parameter 1. Then

$$\begin{aligned} h_1(Y_n \mid Y^{n-1}) = \frac{f_1(Y_n \mid Y^{n-1})}{f_0(Y_n \mid Y^{n-1})} &= \frac{p_1^1}{p_1^0} \qquad \text{when} \quad Y_n - F(Y^{n-1}) = w_1 \\ &= \frac{p_2^1}{p_2^0} \qquad \text{when} \quad Y_n - F(Y^{n-1}) = w_2 \\ &\ \ \vdots \qquad\qquad\qquad\qquad \vdots \end{aligned}$$

Since at least one p_i^0 differ from p_i^1 there exists a partition (P, P') of the integers $\{1, 2, \cdots\}$ such that $p_i^1/p_i^0 > 1$ for $i \in P$ and $p_i^1/p_i^0 \leqq 1$ for $i \in P'$ with neither P nor P' empty. Now choose α so that

$$1 < \alpha < \sup_{i \in P} \frac{p_i^1}{p_i^0};$$

then there exists at least one i, denoted $\hat{i}$, such that $1 < \alpha < p_{\hat{i}}^1/p_{\hat{i}}^0$. Writing i.o. as an abbreviation for infinitely often, we have

$$\begin{aligned} P[h_{\hat{k}_{n-1}}(Y_n \mid Y^{n-1}) > \alpha, \ \text{i.o.}] &= P[h_1(Y_n \mid Y^{n-1}) > \alpha, \ \text{i.o.}] \\ &\geqq P[W_n = w_{\hat{i}}, \ \text{i.o.}] = 1 . \end{aligned}$$

It follows that condition (A2) holds for this example and consequently the maximum likelihood estimate is strongly consistent.

4. Conclusion. An analysis of maximum likelihood estimates may be viewed as an examination of infinite products of conditional densities. The example above showed that the prediction condition can reduce this problem to simpler and more amenable tasks.

As far as extensions are concerned, it obviously remains to find simple sufficient conditions for (A2) which may be employed in various large classes of problems.

Acknowledgments. The author wishes to acknowledge many helpful conversations with Terrence Fine and with members of the Information Systems Laboratory, Stanford University. In addition, the author wishes to express his appreciation of the referee's detailed comments, especially in connection with the proof of the submartingale property in Section 2.

REFERENCES

[1] Aoki, M. and Yue, P. C. (1970). On certain convergence questions in system identification. *SIAM J. Control* **8** No. 2 239–256.

[2] Åström, K. J. and Bohlin, T. (1965). Numerical identification of linear dynamic systems from operating records. *Proc. IFAC Conference on Self-Adaptive Control Systems, Teddington.* 96–110.

[3] Åström, K. J., Bohlin, T. and Wensmark, S. (1965). Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records. Technical Paper TP.18.150. IBM Nordic Laboratory, Sweden.

[4] Åström, K. J. and Eykhoff, P. (1971). System identification—a survey. *Automatica* **7** 123–162.

[5] Bar-Shalom, Y. (1971). On the asymptotic properties of the maximum likelihood estimate obtained from dependent observations. *J. Roy. Statist. Soc. Ser. B* **33**, No. 1 72–77.

[6] Bar-Shalom, Y. (1972). Optimal simultaneous state estimation and parameter identification in linear discrete-time systems. *IEEE Trans. Automatic Control* **AC-17** No. 3 308–319.

[7] Caines, P. E. (1971). The parameter estimation of state variable models of multivariable linear systems. *Proc. Fourth UKAC Convention, Manchester*, in *IEE Conference Publication No.* 78 98–106.

[8] Caines, P. E. and Rissanen, J. (1974). Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes. *IEEE Trans. Information Theory* **IT-20** No. 1, 102–104.

[9] Chung, K. L. (1968). *A Course in Probability Theory.* Harcourt, Brace and World, New York.

[10] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

[11] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[12] Durbin, J. (1960). Estimation of parameters in time series regression models. *J. Roy. Statist. Soc. Ser. B* **22** 139–153.

[13] Feller, W. (1966). *An Introduction to Probqbility Theory and its Applications*, **2**. Wiley, New York.

[14] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

[15] Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725.

[16] Hurzurbazar, V. S. (1948). The likelihood equation, consistency and the maximum of the likelihood function. *Ann. Eugenics* **14** 185–200.

[17] Kraft, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. Calif. Publ. Statist.* **2** 125–242.

[18] LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.* **1** 277–330.

[19] RISSANEN, J. and CAINES, P. E. Consistency of maximum likelihood estimators for multivariate Gaussian processes with rational spectrum. Submitted to *Annals of Statistics.*

[20] ROUSSAS, G. C. (1965). Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimate. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** 69–73.

[21] ROUSSAS, G. C. (1967). Asymptotic normality of the maximum likelihood estimate in Markov processes. *Metrika* **14** 62–70.

[22] STUART, A. and KENDALL, M. G. (1948). *Advanced Theory of Statistics.* Griffin, London.

[23] TSE, E. and ANTON, J. J. (1972). On the identifiability of parameters. *IEEE Trans. Automatic Control* **AC-17** No. 5 637–646.

[24] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist* **20** 595–601.

ELECTRONICS RESEARCH LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA