Estimating Functions in Chaotic Systems

Subhash Lele

Journal of the American Statistical Association, Volume 89, Issue 426 (Jun., 1994), 512-516.

Estimating Functions in Chaotic Systems

Subhash Lele*

Berliner considered Bayesian and likelihood-based approaches for estimation and prediction in a chaotic system with measurement error. This article proposes the use of estimating functions for this problem. Logistic and exponential maps are analyzed. Estimators are shown to be consistent and asymptotically normal. Small-sample behavior is studied with simulations.

KEY WORDS: Measurement errors; Nuisance parameters.

1. INTRODUCTION

The last decade or two has seen an explosion of development in the theoretical as well as applied aspects of simple nonlinear systems that lead to chaotic behavior. I refer the reader to the *Journal of the Royal Statistical Society*, Ser. B, Vol. 54, No. 2 (1992), hereinafter referred to as JRSS-B, which contains papers presented at a special meeting of the Royal Statistical Society in 1991, and also Berliner (1991). These papers discuss various statistical aspects of chaotic systems. They also contain a large bibliography related to the subject of chaos. The main approach used in the JRSS-B papers is nonparametric. On the other hand, Berliner (1991) considered parametric models that lead to chaos on a subset of the parameter space. He considered a likelihood-based approach for estimation of the parameters of the underlying deterministic system in the presence of measurement error. He also considered Bayesian prediction for these so-called "unpredictable processes."

Estimating functions (Godambe 1960, 1985, 1991) have proved to be a promising alternative to maximum likelihood estimation. They usually lead to simple numerical calculations and also possess a robustness property (Godambe and Thompson 1984) in that they need only the specification of the first few, usually two, moments instead of the complete specification of the underlying distribution.

The purpose of this article is to show the applicability of the estimating functions approach to the setup considered by Berliner (1991). I will consider only the estimation and not the prediction problem. The models I consider are the logistic map and exponential map; these are described in detail in Section 2. Introduction of measurement error in these models is described in Section 3. Estimating functions for these maps are derived, simulation results for the estimators are presented, and asymptotic properties of consistency and normality are proven in Section 4. A discussion of some further problems is presented in Section 5.

2. CHAOTIC SYSTEMS: TWO EXAMPLES

Following Smith's (1992) description, let $\{X_t\}$ denote a scalar time series satisfying $X_t = F(X_{t-d}, X_{t-d+1}, \ldots, X_{t-1})$ for some integer d known as the embedding dimension and nonlinear function $F(\cdot)$. For many $F(\cdot)$, this deterministic system does not converge to a fixed point or limit cycle but exhibits the apparently random behavior known as chaos. In general, as $t \to \infty$, the d vectors $Y_t = (X_{t-d+1}, \ldots, X_t)$ become arbitrarily close to a limiting set known as the attractor. If the system converges to a stable fixed point or a limit cycle, then the attractor consists of one or a finite number of points. In a chaotic system, the attractor is typically a fractal (i.e., a set whose dimension is nonintegral), and in that case the attractor is known as a strange attractor. Following is a description of two such dynamical systems. May (1976) attested to the importance of these simple maps for various practical situations.

2.1 Logistic Map (Berliner 1991)

Consider a dynamical system where for all t in $T = \{0, 1, 2, \ldots\}$, $x_{t+1} = ax_t(1 - x_t)$ and where $x_0 \in [0, 1]$ and $a \in [0, 4]$. If $a \in [0, 1]$, then $x_t \to 0$ as $t \to \infty$. If $a \in (1, 3]$, then $x_t \to (1 - (1/a))$ as $t \to \infty$. For $a > 3$, the behavior is quite interesting. For $3 < a \leq 1 + 6^{1/2}$, x_t asymptotically oscillates between two points forever. For slightly larger a, the asymptotic oscillation is on a 4-point attractor, then 16, and so on. This period doubling continues to a "period" of 2^∞ at $a = 3.569\ldots$. For larger a the behavior is even stranger and "unpredictable."

2.2 Exponential Map (May 1976)

Consider another dynamical system $x_{t+1} = x_t\exp\{r(1 - x_t)\}$, where $x_0 \in [0, \infty)$ and $r \in [0, 4]$. This model describes a population with a propensity to simple exponential growth at low densities and a tendency to decrease at high densities. The steepness of this nonlinear behavior is tuned by the parameter r. The model is plausible for a single-species population regulated by an epidemic disease at high density. For $r \in [0, 2]$, $x_t \to 1$ as $t \to \infty$. For $r > 2$, x_t asymptotically oscillates between two or more points. The "chaotic" region begins at $r > 2.6924\ldots$. In theory, the chaotic region ends at $r = \infty$; however, in practice, as r becomes large, x_t eventually gets carried so low as to be effectively 0, thus producing extinction in models of biological populations.

3. MEASUREMENT ERROR IN CHAOTIC SYSTEMS

In the formulation in the previous section, given the series $\{x_t, t = 0, 1, 2, \ldots\}$, calculation of the parameter values of a or r is trivial. But in practice what one usually observes is

* Subhash Lele is Associate Professor, Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21205. This work was initiated during the author's visit to Patuxent Wildlife Research Center, Laurel, MD. This visit was supported by the U.S. Fish and Wildlife Service's Global Climate Change Program. The author thanks Robert Serfling for kind encouragement and many useful comments, Brian Dennis and Joan Aron for teaching him the basics of chaotic models, and John Sauer, Richard Barker, Jim Nichols and Bill Link for initiating him in the area of measurement error analysis.

Journal of the American Statistical Association
June 1994, Vol. 89, No. 426, Theory and Methods

a "noise-introduced series" $\{y_t, t = 0, 1, 2, \ldots\}$. For example, in ecology one may estimate x_t (the population size at time t) using capture–recapture methods or quadrat sampling data, or there might be pure measurement error added to the observations due to the measuring device.

I consider the following kinds of measurement errors:

Additive error. Here $Y_t = x_t + \sigma e_t$, $\sigma > 0$, where e_t's are independent identically distributed random variables with mean 0 and variance 1. Berliner (1991) assumed these to be Gaussian.

Multiplicative error. Here $Y_t = x_t Z_t$; where Z_t's are independent identically distributed random variables with mean 1, variance σ^2, and range on $[0, \infty)$.

Note that for the logistic map, although x_t's are in [0, 1], Y_t's are not necessarily so. But for the exponential map with multiplicative error, x_t's and Y_t's both have the same range $[0, \infty)$.

To keep close correspondence with Berliner (1991), I consider additive error for the logistic map and multiplicative error for the exponential map and assume that σ is known. x_0, r, and a are unknown. I am interested in estimating a or r with x_0 considered to be a nuisance parameter.

4. ESTIMATING FUNCTIONS

In the following I briefly describe the estimating functions approach and then apply it to the two models in Section 2.

Let θ be a one-dimensional parameter taking values in Θ. Let S denote the sample space. An estimating function for θ is defined (Godambe 1960) as any function $g : S \times \Theta \to R$ such that $E_\theta[g(\mathbf{Y}, \theta)] = 0$. This is also called a "zero unbiased estimating function." An estimator $\hat{\theta}$ is obtained by solving the empirical version $g(\mathbf{y}, \theta) = 0$. Under suitable regularity conditions (see the Appendix), these estimators can be shown to be consistent and asymptotically normal.

4.1 Logistic Map with Additive Error

In this case,

$$x_{t+1} = a x_t (1 - x_t)$$

and

$$y_{t+1} = x_{t+1} + \sigma e_{t+1}. \qquad (*)$$

Analogous to the equality $\sum_{t=0}^{n-1} (x_{t+1} - a x_t(1 - x_t)) = 0$, one may write

$$\sum_{t=0}^{n-1} (y_{t+1} - a y_t (1 - y_t)) = 0. \qquad (1)$$

But note that the estimating function

$$g(\mathbf{Y}, a) = \sum_{t=0}^{n-1} [Y_{t+1} - a Y_t (1 - Y_t)] \qquad (2)$$

is not zero unbiased. Hence $\hat{a}$ obtained from (1) may not be (in fact, *is* not) consistent for a. But a simple adjustment to (2) yields a zero-unbiased estimating function, $g(\mathbf{Y}, a) = \sum_{t=0}^{n-1} [Y_{t+1} - a(Y_t - Y_t^2 + \sigma^2)]$. Solving the empirical version of this equation, namely $\sum_{t=0}^{n-1} [y_{t+1} - a(y_t - y_t^2 + \sigma^2)] = 0$ one gets the estimator

$$\hat{a} = \frac{\sum_{t=0}^{n-1} y_{t+1}}{\sum_{t=0}^{n-1} (y_t - y_t^2 + \sigma^2)}.$$

Using first principles one can demonstrate the strong consistency and asymptotic normality of $\hat{a}$. Alternatively, one can apply the theorem stated in the Appendix to get the required result. In the following theorem, the result is stated formally.

Theorem 1. For model $(*)$, under the conditions stated in Section 3,

(a) $\hat{a} \to a$ w.p. 1, and
(b) $\sqrt{n}(\hat{a} - a) \to N(0, V_a)$ in law.

Table 1 gives simulation results for this model. For these simulations, I have taken $x_0 = .1$, $\sigma = .1$, and e_t's to be $N(0, 1)$ variates. Four entries in each cell correspond to the mean, standard deviation, skewness, and kurtosis of the distribution of $\hat{a}$ numerically estimated from 500 independent realizations of length n with given a. Both consistency and asymptotic normality are fairly apparent from the results. Changing x_0 did not have any noticeable effect on the results. The value of σ is the same as considered by Berliner (1991). Also note that even with samples of sizes 20 or 50, the estimator is well behaved.

4.2 Exponential Map

In this case,

$$x_{t+1} = x_t \exp\{r(1 - x_t)\}$$

and

$$y_{t+1} = x_{t+1} Z_{t+1}. \qquad (**)$$

I also assume that Z_t's are independent identically distributed log-normal random variables with parameters $-(\sigma^2/2)$ and σ^2. Thus $E(Z_t) = 1$, $\text{var}(Z_t) = e^{\sigma^2}$, and $E(\log Z_t) = -(\sigma^2/2)$, $\text{var}(\log Z_t) = \sigma^2$.

Consider the following equation:

$$g(\mathbf{Y}, r) = \sum_{t=0}^{n-1} [(\log Y_{t+1} - \log Y_t) - r(1 - Y_t)]. \qquad (3)$$

It is easy to check that this is a zero-unbiased estimating function. But note that

$$\frac{1}{n} E\left[\frac{\partial}{\partial r} g(\mathbf{Y}, r)\right] = E\left[\frac{1}{n}\left(\sum_{t=0} Y_t - 1\right)\right]$$
$$= \frac{1}{n} \sum_{t=0}^{n-1} (x_t - 1) \to EX - 1.$$

My numerical experience suggests that for all values of r, this is 0. Thus this violates the condition $|E[(\partial/\partial\theta) g(\mathbf{Y}, \theta)]| > 0$ for the result in the Appendix to hold. Numerical experience also suggests that $\hat{r}$, obtained by solving (3), is very unstable and inconsistent.

Table 1. Simulation Results for the Model $x_{t+1} = ax_t(1 - x_t)$ and $Y_{t+1} = x_{t+1} + \sigma e_{t+1}$, $x_0 = .1$, $\sigma = .1$, $e_t \sim$ iid $N(0, 1)$

Sample size	1.5	2.5	3.25	3.50	3.75	3.95	
20	1.4988	2.5056	3.2584	3.5017	3.7933	3.9908	mean
	.0712	.1517	.2631	.3362	.4109	.4544	standard deviation
	.0001	.0008	.0103	.0268	.0275	.0661	skewness
	3.1936	3.0552	3.5844	3.8651	3.0679	3.7731	kurtosis
50	1.4966	2.5097	3.2483	3.5155	3.7709	3.9653	mean
	.0397	.0977	.1525	.2064	.2331	.2860	standard deviation
	.0000	.0002	.0010	.0033	.0069	.0121	skewness
	2.9991	2.9806	2.8399	3.1379	3.2888	3.7084	kurtosis
100	1.4993	2.4938	3.2555	3.5045	3.7484	3.9566	mean
	.0261	.0633	.1193	.1495	.1606	.1924	standard deviation
	.0000	.0000	.0005	.0008	.0018	.0029	skewness
	3.0265	2.8015	3.3289	2.9800	3.2197	3.2472	kurtosis
500	1.5004	2.5003	3.2504	3.5006	3.7513	3.9516	mean
	.0104	.0292	.0481	.0622	.0703	.0930	standard deviation
	.0000	.0000	.0000	.0000	.0000	.0001	skewness
	2.8054	2.9822	2.8089	2.8635	3.3975	3.4093	kurtosis
1,000	1.4996	2.4999	3.2467	3.5015	3.7513	3.9506	mean
	.0077	.0206	.0375	.0444	.0515	.0656	standard deviation
	.0000	.0000	.0000	.0000	.0000	.0000	skewness
	3.1843	3.0055	2.8782	2.9864	3.1055	2.7990	kurtosis

NOTE: Mean, standard deviation, skewness, and kurtosis are estimated from 500 independent realizations.

Now consider the estimating function

$$g(\mathbf{Y}, r) = \sum_{t=0}^{n-1} \left\{ [(\log Y_{t+1} - \log Y_t)^2 - 2\sigma^2] - r^2\left(1 - 2Y_t + \frac{Y_t^2}{1 + e^{\sigma^2}}\right)\right\}.$$

Note that

$$E[(\log Y_{t+1} - \log Y_t)^2] = (\log x_{t+1} - \log x_t)^2 + 2\sigma^2$$

and

$$E\left[1 - 2Y_t + \frac{Y_t^2}{1 + e^{\sigma^2}}\right] = (1 - x_t)^2.$$

Hence this estimating function is zero unbiased. Moreover,

$$\frac{1}{n}\left|E\left[\frac{\partial}{\partial r^2} g(\mathbf{Y}, r)\right]\right| \to E(1 - X)^2 > 0,$$

unless X is degenerate at 1.

The following theorem is a consequence of the result in the Appendix.

Theorem 2. For model (**), under the conditions stated in Section 3 and for $r > 2$,

$$\hat{r} = \text{abs}\left[\frac{\sum_{t=0}^{n-1} (\log y_{t+1} - \log y_t)^2 - 2\sigma^2}{\sum_{t=0}^{n-1}\left(1 - 2y_t + \frac{y_t^2}{1 + e^{\sigma^2}}\right)}\right]^{1/2}$$

and

(a) $\hat{r} \to r$ w.p. 1
(b) $\sqrt{n}(\hat{r} - r) \to N(0, V_r)$ in law.

Table 2 reports the simulation results for this model. For simulations, I have taken $x_0 = 1.3$, $\sigma = .3$, and Z_t's to be log-normal random variates with parameters $-(\sigma^2/2)$ and σ^2. Four entries in each cell correspond to the mean, standard deviation, skewness, and kurtosis of the distribution of $\hat{r}$ numerically estimated from 500 independent realizations of length n and specified r. Both consistency and asymptotic normality are fairly apparent from the results. Changing x_0 did not have any noticeable effect on the results. Note that even with small sizes of 20 or 50, the estimator is well behaved.

It should be mentioned that to derive the zero-unbiased estimating function for this model, I have assumed a particular model for Z_t's viz. log-normal errors. Strictly speaking, all that one needs to know is var(Z_t) and var(log Z_t) to get zero unbiasedness.

I would also like to point out a curious result that if $r \in [1, 2]$ (i.e., when $\{x_t\}$ process is stable and $x_t \to 1$ as $t \to \infty$), r cannot be estimated consistently using the aforementioned procedure. This happens perhaps because $(1/n)E[(\partial/\partial(r^2))g(\mathbf{Y}, r)] \to 0$ in this case. Unstable and chaotic processes lead to consistent estimation, whereas stable processes lead to inconsistency.

5. DISCUSSION

This article demonstrates that estimating functions can be successfully applied for the estimation of parameters of chaotic systems in the presence of measurement error. A comparison with the likelihood approach yields the following points:

a. The nuisance parameter x_0 is eliminated very easily. Of course, whether or not x_0 is a nuisance parameter could be debated. But note that in the discussion of his paper, Berliner (1991) noted that if n is large, even for prediction purposes, then it is computationally better to ignore x_0 and consider only a previous few observations.

Table 2. Simulation Results for the Model $x_{t+1} = x_t \exp\{r(1 - x_t)\}$ and $Y_{t+1} = x_{t+1}Z_{t+1}$ with $x_0 = 1.3$, $z_t \sim \text{lognormal}(-\sigma^2/2, \sigma^2)$, with $\sigma = .3$

Sample size	2.5	3.0	3.25	3.50	3.75	3.95	
20	2.5471	3.1621	3.3592	3.6351	3.8417	4.2634	mean
	0.2490	0.4373	.4320	.4440	.5808	.7223	standard deviation
	−.0050	.0205	−.0004	−.0089	.0329	.0503	skewness
	3.0901	2.8671	2.8739	2.7659	2.7154	2.5502	kurtosis
50	2.5194	2.9463	3.3026	3.4746	3.8498	4.0687	mean
	.1713	.2778	.2654	.3098	.3697	.4074	standard deviation
	−.0005	−.0041	.0002	−.0012	−.0004	.0031	skewness
	2.6883	3.5359	2.7830	2.9494	2.6608	2.8774	kurtosis
100	2.5078	3.0169	3.2321	3.5595	3.7962	3.9957	mean
	.1106	.1996	.1940	.2461	.2643	.3019	standard deviation
	−.0003	−.0000	.0008	.0034	−.0021	−.0029	skewness
	2.9286	3.5018	2.6623	3.0184	2.8794	2.9628	kurtosis
500	2.5012	2.9947	3.2595	3.5074	3.7567	3.9511	mean
	.0573	.0811	.0864	.1050	.1205	.1326	standard deviation
	.0000	.0000	−.0001	−.0001	.0000	.0003	skewness
	3.0732	2.9921	2.9812	3.4521	2.8987	3.0630	kurtosis
1,000	2.5047	3.0046	3.2489	3.5016	3.7571	3.9451	mean
	.0736	.03792	.0609	.0738	.0817	.0925	standard deviation
	.0000	.0000	.0000	.0000	.0000	.0000	skewness
	2.8002	2.8411	3.2843	3.3123	2.6021	3.2468	kurtosis

NOTE: Mean, standard deviation, skewness, and kurtosis are estimated from 500 independent realizations.

b. As noted by Berliner (1991), the likelihood surface for chaotic systems looks very choppy; thus finding the maximum is a difficult task. Estimating functions seem to yield computationally simple (and for the cases considered in this article, unique) estimators.

c. Asymptotic properties for the estimators obtained by using estimating functions are easily derived.

d. Following Lele (1991), one can estimate the asymptotic variance of these estimates by jackknifing the linear estimating functions.

e. Assumption of the knowledge of σ^2 is, of course, somewhat unrealistic. But there are instances where repeated observations with the same x_t are available (Sauer, personal communication). In such cases one may estimate σ^2 from these repeated observations and behave as if it is the true value in the spirit of pseudolikeli[illegible] suggested by Gong and Samaniego (1986).

f. If one does not have repeated observations, then one can possibly estimate σ^2 by applying the semiparametric approach described by Kiefer and Wolfowitz (1956) and Lindsay (1983). For example, consider the logistic map with Gaussian errors. In this case,

$$Y_t \mid x_t \sim N(x_t, \sigma^2) \qquad t = 1, 2, \ldots .$$

Moreover, let $Q(\cdot)$ denote the invariant distribution of the $\{x_t\}$ process. Then, for sufficiently large t,

$$f_{Y_t}(y_t) = \int_0^1 f(y_t \mid x_t, \sigma^2)\, dQ(x_t).$$

Then, under certain regularity and identifiability conditions (Kiefer and Wolfowitz 1956), it is possible to estimate σ^2 consistently by the method of maximum likelihood. Existence but no particular form for $Q(\cdot)$ is assumed. Unfortunately, this needs the assumption of a known form of the error distribution.

The models considered in this article are simple, but they demonstrate the point that using estimating functions is a reasonable approach for statistical analysis of chaotic systems with measurement error. Multiparameter models would necessarily require more ingenuity in deriving zero-unbiased estimating functions than was needed here. It is also not obvious whether and how estimating functions can be used for the prediction problem.

APPENDIX: ASYMPTOTIC PROPERTIES

Consistency and asymptotic normality for the estimators based on zero-unbiased estimating functions can be proved as follows.

Let $\{Y_1, Y_2, \ldots, Y_n, \ldots\}$ be a sequence of independent and identically distributed random variables. Let S denote the sample space and Θ denote the parameter space with Θ being an open interval in $\Re$.

Let $S_n = \underbrace{S \times S \times \cdots \times S}_{n \text{ times}}$ and $Y_{(n)} = (Y_1, Y_2, \ldots, Y_n)$.

Let $g : S_n \times \Theta \to R$ be a function such that $E_\theta(g(Y_{(n)})) = 0$ for all $\theta \in \Theta$ and for all n.

The function $g(\cdot)$ is called a zero-unbiased estimating function. Further assume that $g(\cdot)$ is a linear estimating function (Godambe 1985, Lele 1991); that is, it has the form

$$g(Y_{(n)}, \theta) = \sum_{t=1}^{n} g(Y_t, \theta),$$

where $Y_t = (Y_t, Y_j, j \in N(t))\cdot N(t)$ denotes the finite collection of "neighbors" of t. For the cases considered in this article, $\mathbf{Y}_t = (Y_t, Y_{t-1})$. An estimator $\hat{\theta}$ of θ is the solution of

$$\sum_{t=1}^{n} g(y_t, \tau) = 0,$$

where $(y_1, y_2, \ldots, y_n)$ is the realization of $(Y_1, Y_2, \ldots, Y_n)$.

Additional Assumptions for Consistency of $\hat{\theta}$

C1. $g(Y_{(n)}, \tau)$ is strictly monotone in τ for all n.

C2. $E_\theta[g(Y_{(n)}, \tau)]$ is finite and strictly monotone in τ for all n.

C3. $\{g(Y_t, \tau)\}$ is a weakly stationary, m-dependent sequence with finite variances and covariances for all $\tau \in \Theta$.

Under these assumptions, the following result is a consequence of the strong law of large numbers for m-dependent sequences (Stout 1974, p. 207):

$$\frac{1}{n}[g(Y_{(n)}, \tau) - E_\theta(g(Y_{(n)}, \tau))] \to 0 \quad \text{with probability } 1 \quad \forall \tau \in \Theta.$$

Given this result and the uniqueness of $\hat{\theta}$, by lemma 7.2.1A of Serfling (1980, p. 249), it follows that

$$\hat{\theta} \to \theta \quad \text{with probability } 1.$$

Asymptotic Normality of $\hat{\theta}$

This proof follows by the linearization technique described by Serfling (1980, pp. 144–148).

Let $S_n(\tau) = (1/\sqrt{n}) \sum_{t=1}^{n} g(Y_t, \tau)$.

Assume that:

N1. $E|g(Y_t, \theta)|^3 \leq R < \infty$ for all t.

N2. $\lim_{n\to\infty} 1/n \sum_{t=1}^{n} E_\theta[(\partial/\partial\theta)g(Y_t, \theta)] = c_0$ where $c_0 \neq 0$ and $< \infty$.

N3. $|(1/n) \sum_{t=1}^{n} E_\theta[(\partial^2/\partial\tau^2)g(Y_t, \tau)]| \leq M < \infty$ for some M, for all and $\tau \in \Theta$.

N4. $\hat{\theta} \to \theta$ in probability.

N5. $\lim_{n\to\infty}(1/n)E(\sum_{t=1}^{n} g(Y_t, \theta))^2 = \eta < \infty$.

Under suitable smoothness conditions on $g(Y_{(n)}, \tau)$, expand $S_n(\hat{\theta})$ around θ to get

$$S_n(\hat{\theta}) = S_n(\theta) + \sqrt{n}(\hat{\theta} - \theta)\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\tau}g(Y_t, \tau)|_{\tau=\theta}\right] + \frac{(\hat{\theta}-\theta)^2}{2!}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\tau^2}g(Y_t, \tau)|_{\tau=\theta^*}\right],$$

where θ^* is such that $|\theta^* - \theta| \leq |\hat{\theta} - \theta|$.

Application of the central limit theorem for stationary m-dependent sequences (Fraser 1957, p. 219, thm. 4.2), it follows that

$$S_n(\theta) \to N(0, \eta) \quad \text{in law.}$$

Under conditions similar to C3 for the sequences $\{(\partial/\partial\tau)g(Y_t, \tau)|_{\tau=\theta}\}$ and $\{(\partial^2/\partial\tau^2)g(Y_t, \tau)|_{\tau=\theta^*}\}$, namely of finite variances and covariances, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\tau}g(Y_t, \tau)|_{\tau=\theta} \to c_0 \quad \text{with probability } 1$$

and

$$R_n = \frac{(\hat{\theta}-\theta)^2}{2!}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\tau^2}g(Y_t, \tau)|_{\tau=\theta^*}\right] \to 0 \quad \text{with probability } 1.$$

These together yield the result

$$\sqrt{n}(\hat{\theta} - \theta) \to N\left(0, \frac{\eta}{c_0^2}\right) \quad \text{in law.}$$

[Received June 1992. Revised August 1993.]

REFERENCES

Berliner, L. M. (1991), "Likelihood and Bayesian Prediction for Chaotic Systems," *Journal of the American Statistical Association*, 86, 938–952.
Fraser, D. A. S. (1957), *Nonparametric Methods in Statistics*, New York: John Wiley.
Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208–1211.
——— (1985), "The Foundations of Finite Sample Estimation in Stochastic Processes," *Biometrika*, 72, 419–428.
——— (ed.) (1991), *Estimating Functions*, Cambridge, U.K.: Cambridge University Press.
Godambe, V. P., and Thompson, M. (1984), "Robust Estimation Through Estimating Equations," *Biometrika*, 71, 115–125.
Gong, G., and Samaniego, F. J. (1986), "Pseudo-Maximum Likelihood Estimation: Theory and Applications," *The Annals of Statistics*, 9, 861–869.
Kiefer, J., and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters," *Annals of Mathematical Statistics*, 27, 887–906.
Lele, S. (1991), "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes," *Journal of the Royal Statistical Society*, Ser. B, 53, 253–268.
Lindsay, B. G. (1983), "The Geometry of Mixture Likelihoods, Part II: The Exponential Family," *The Annals of Statistics*, 11, 783–792.
May, R. M. (1976), "Simple Mathematical Models With Very Complicated Dynamics," *Nature*, 261, 459–467.
Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.
Smith, R. L. (1992), "Estimating Dimension in Noisy Chaotic Time Series," *Journal of the Royal Statistical Society*, Ser. B, 54, 3329–3352.
Stout, W. F. (1974), *Almost Sure Convergence*, New York: Academic Press.