Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

GRID–BASED SIMULATION AND THE METHOD OF CONDITIONAL LEAST SQUARES

Katherine B. Ensor

Department of Statistics
Rice University
Houston, TX 77251-1892, U.S.A.

Peter W. Glynn

Department of Engineering–Economic
Systems and Operations Research
Stanford University
Stanford, CA 94305-4022, U.S.A.

ABSTRACT

This paper is concerned with the use of simulation to compute the conditional expectations that arise in the method of conditional least squares. Our approach involves performing simulations at each point on a discrete grid imbedded within a statistical parameter space. Our main result concerns the number of grid points and amount of simulation necessary in order to obtain a degree of accuracy comparable to that in the case in the which the conditional expectations are available in closed form.

1 INTRODUCTION

In this paper, we discuss a method known as "conditional least squares" that is widely used for purposes of statistical parameter estimation in the stochastic process setting; see Hall & Heyde (1980) for an introduction to the method. This method requires minimizing a function over the parameter space that involves conditional expectations defined in terms of the stochastic process under consideration. In certain applications, it is natural to compute the conditional expectations via Monte Carlo simulation. In doing so, it is clearly practical only to perform simulations at a finite number of different parameter values. This leads naturally to the concept of "grid–based simulation", in which simulations are performed at various points comprising a grid.

In Section 2, we introduce the method of conditional least squares in the context of parameter estimation for continuous time Markov chains (CTMCs). Section 3 concerns the asymptotic analysis of conditional least squares under the assumption that the relevant conditional expectations can be computed in closed form. Finally, Section 4, we study the use of grid–based simulation in the CTMC context and prove our main result (see Theorems 1 and 2). We show that if the CTMC is observed at n equally spaced time points, then one needs to simulate on the order of n^2 time units of the CTMC so as to ensure that the simulation based minimization gives roughly the same solution quality as that associated with the case in which the conditional expectations are available in closed form. This is accomplished by starting with a coarse grid, and refining it successively as more information on the likely location of the minimizer becomes available.

2 PROBLEM FORMULATION

Suppose that we observe a stationary finite–state continuous–time Markov chain $X = (X(t) : t \geq 0)$, with the intention of using the observed data to make inferences about the generator underlying X. In this paper, we shall adopt a parametric statistical formulation for this inference problem. Specifically, we shall require that the generator underlying X be a member of a parametric family $(A(\theta) : \theta \in \Lambda)$ of generators defined on the state space S associated with X. Our goal, then, is to develop a means of estimating the "true" value of the d-dimensional parameter θ, call it θ^*, underlying X. For example, in the context of the M/M/1/∞ single-server queueing model, this would correspond to attempting to estimate the vector $\theta^* = (\lambda^*, \mu^*)$, where λ^* and μ^* are the arrival and service rates for the queue, respectively.

Without any significant loss of generality, we shall assume that the parameter set Λ is the d-dimensional unit hypercube. We shall further assume that:

A1. **i)** $|S| < \infty$;

ii) $A(\theta^*)$ is irreducible;

iii) $A(\cdot)$ is three times continuously differentiable on Λ;

iv) $\{ (x, y) : A(\theta, x, y) \neq 0 \}$ is independent of $\theta \in \Lambda$;

v) θ^* lies in the interior of Λ.

If X is observed continuously over some finite interval $[0,t]$, the method of maximum likelihood applies directly here (see, for example, Billingsley 1961). In particular, let $A(\theta,x,y)$ be the (x,y)'th element of the generator $A(\theta)$, let $Y=(Y_n : n \geq 0)$ be the embedded discrete–time Markov chain associated with X, and let $J(t)$ be the number of jumps of X over $[0,t]$. Then, the likelihood function $L_c(\theta,t)$ can be easily written down explicitly:

$$L_c(\theta,t) = \prod_{i=0}^{J(t)-1} A(\theta,Y_i,Y_{i+1}) \cdot \exp(\int_0^t A(\theta,X(s),X(s)ds). \quad (1)$$

The maximum likelihood estimator for θ^* is then taken to be the maximizer of $L_c(\cdot,t)$ over Λ.

However, in many applications, X is observed only discretely, see for example Bridges, Ensor and Thompson (1992). In particular, we shall be concerned with the situation in which X is observed only at the integer times $0,1,2,\ldots,n$. Let $P_\theta(\cdot)$ be the probability measure on the path space of X under which X evolves according to a stationary process with generator $A(\theta)$, and set $P(\theta,t,x,y) = P_\theta(X(t)=y \mid X(0)=x)$. If we put $X_i \stackrel{\Delta}{=} X(i)$, then the likelihood function $L_n(\theta)$ associated with the discrete sample $(X_0,\ldots,X_n)$ is given by

$$L_n(\theta) = \prod_{i=0}^{n-1} P(\theta,1,X_i,X_{i+1}).$$

However, in contrast to (1), the likelihood function here is not a simple function of the matrices $(A(\theta) : \theta \in \Lambda)$ that are typically directly specified by the modeler. Rather, in order to compute $L_n(\theta)$, it is necessary to compute the $P(\theta,1,x,y)$'s from $A(\theta)$. Setting $P(\theta,t) = (P(\theta,t,x,y) : x,y \in S)$, this may be accomplished either by taking advantage of the fact that

$$P(\theta,t) = \exp(A(\theta)t)$$

or by noting that the transition semigroup $(P(\theta,t) : t \geq 0)$ is the unique solution of both the backward Kolmogorov differential equations

$$\begin{aligned} P'(\theta,t) &= A(\theta)P(\theta,t) \\ \text{such that } P(\theta,0) &= I \end{aligned}$$

and the forward Kolmogorov differential equations

$$\begin{aligned} P'(\theta,t) &= P(\theta,t)A(\theta) \\ \text{such that } P(\theta,0) &= I. \end{aligned}$$

Clearly, significant numerical effort will generally be required to compute $L_n(\theta)$ for a fixed value of θ. Given that $L_n(\cdot)$ needs to be maximized over Λ in order to compute the maximum likelihood estimator, the numerical challenge is even more daunting.

In an effort to develop a more tractable numerical approach to such inference problems, Klimko and Nelson (1978) proposed the method of conditional least squares. In particular, for a given $f : S \to \mathbb{R}$, the idea is to define the estimator θ_n^* as the minimizer of the sum of squares

$$\frac{1}{n}\sum_{i=0}^{n-1}(f(X_{i+1} - E_\theta[f(X_{i+1})|X_i])^2.$$

Setting $g(\theta,x) = E_\theta[f(X_1) \mid X_0 = x]$, we note that the method of conditional least squares requires the computation of $g(\theta) = (g(\theta,x) : x \in S)$. This can be accomplished by, for example, solving the linear system of differential equations

$$\begin{aligned} u'(\theta,t) &= A(\theta)u(\theta,t) \\ \text{subject to } u(\theta,0) &= f, \end{aligned} \quad (2)$$

in which case $g(\theta) = u(\theta,1)$. However, our interest in this paper stems from the fact that $g(\theta)$ can also be computed via Monte Carlo simulation. The simulation alternative is particularly attractive, relative to (2), when $|S|$ is large. The idea that simulation has a useful role to play in the statistical estimation context has received significant attention from the statistics and econometric communities; see, for example, Cook and Stefanski (1994), Diggle & Gratton (1984), Duffie and Singleton (1993), Ensor (1994), Lee (1992), Keane (1994), Maa et al. (1993), McFadden (1989), Pakes and Pollard (1989), Thompson, Brown and Atkinson (1988).

Before concluding this section, it should be noted that conditional least squares typically exacts a cost from a statistical standpoint. While more numerically tractable then maximum likelihood, the asymptotic variance of θ_n^* tends to be larger than that of the maximum likelihood estimator. Thus, θ_n^* does not extract as much of the statistical information present in the sample as does the method of maximum likelihood. This is typical of the trade off between statistical efficiency and computation tractability that is common to the area of statistical inference for stochastic processes.

3 LIMIT THEORY FOR CONDITIONAL LEAST SQUARES IN THE CTMC SETTING

In this section, we fully work out the asymptotic limit theory for conditional least squares in the CTMC set-

ting. (The existing literature tends to focus on discussion of the method in a general framework, under hypotheses that need to be verified on a case–by–case basis).

Our first goal is to verify consistency of θ_n^* as an estimator of θ^*. Our argument requires that we start by establishing smoothness of $g(\cdot)$. Letting e_i be the i'th unit vector in $\mathbb{R}^d$. Assume, without any loss of generality, that e_i is an admissable direction from the point θ_0, in the sense that $\theta_0 + he_i$ belongs to Λ for h sufficiently small. Then we can use A1 iv), (1), and Taylor's theorem to write

$$h^{-1}(P(\theta_0 + he_i, t, x, y) - P(\theta_0, t, x, y)) \quad (3)$$
$$= E_{\theta_0}\left[I(X(t) = y)\frac{\partial_i L_c(\xi, t)}{L_c(\theta_0, t)} \mid X(0) = x\right]$$

where ξ lies on the line segment connecting θ_0 and $\theta_0 + he_i$ and $\partial_i L_c(\xi, t)$ is the i'th component of the gradient of $L_c(\cdot, t)$ with respect to θ, evaluated at ξ. Now,

$$\frac{\partial_i L_c(\theta, t)}{L_c(\theta, t)} \partial_i \log L_c(\theta, t)$$
$$= \sum_{i=0}^{J(t)-1} \frac{\partial_i A(\theta, Y_i, Y_{i+1})}{A(\theta, Y_i, Y_{i+1})}$$
$$+ \int_0^t \partial_i A(\theta, X(s), X(s)) ds$$

By A1 i) and iii) it follows that $|\partial_i L_c(\xi, t)| \leq (a + bJ(t))L_c(\xi, t)$ for deterministic constants a and b. Furthermore, for c arbitrarily small, positive, and deterministic, we can find h_0 so that for $|h| < h_0$, $L_c(\xi, t)/L_c(\theta_0, t) \leq (1 + c)^{J(t)}d$ with $d > 0$ and deterministic. Since $J(t)$ is stochastically dominated by a Poisson random variable having mean equal to $\sup\{ -A(\theta_0 + he_i, x, x) : |h| < h_0, x \in S \}$ we may conclude that $E_{\theta_0}(a + bJ(t))(1 + c)^{J(t)}d < \infty$ for c sufficiently samll, thereby permitting the application of the dominated convergence theorem in (3). Hence, $P(\cdot, t, x, y)$ is differentiable on Λ. One may easily proceed to show that $P(\cdot, t, x, y)$ is, in fact, three times differentiable under A1 iii).

Let $P(\cdot)$ and $E(\cdot)$ denote the probability and expectation operators on the path–space of X associated with $A(\theta^*)$. Also, let $\pi = (\pi(x) : x \in s)$ be the stationary distribution of X under P, and let $P(x, y) = P(\theta^*, 1, x, y)$. The strong law of large numbers for irreducible CTMC's guarantees that for $x, y \in S$,

$$\pi_n(x, y) \triangleq \frac{1}{n}\sum_{i=0}^{n-1} I(X_i = x, X_{i+1} = y) \quad (4)$$
$$\to \pi(x)P(x, y) \quad P \text{ a.s.}$$

as $n \to \infty$. Setting $\alpha(\theta) = E[(f(X_1) - g(\theta, X_0))^2]$, we note that

$$\alpha_n(\theta) = \sum_{x,y}(f(y) - g(\theta, x))^2 \pi_n(x, y)$$

and hence

$$\sup_{\theta \in \Lambda} |\alpha_n(\theta) - \alpha(\theta)|$$
$$\leq \sup\{ (f(y) - g(\theta, x))^2 : x, y \in S, \theta \in \Lambda \}$$
$$\cdot \max_{x,y} |\pi_n(x, y) - \pi(x)P(x, y)|.$$

The supremum of $|g(\theta, x)|$ over $\theta \in \Lambda$ and $x \in S$ is finite because of the continuity of $P(\cdot, 1, x, y)$ over Λ (a compact set), and the finiteness of S. In view of (4), we have therefore proved the following result.

Proposition 1 *Under A1, $\alpha_n(\cdot)$ converges uniformly P a.s. to $\alpha(\cdot)$ over Λ.*

Let θ_n^* be any (measurable) selection from the set of global minimizers of $\alpha_n(\cdot)$. In view of Proposition 1, strong consistency of θ_n^* to θ^* follows if we show that θ^* is the unique global minimizer of $\alpha(\cdot)$. For this, we need an identifiability assumption.

A2. If $\theta_1, \theta_2 \in \Lambda$ and $g(\theta_1) = g(\theta_2)$, then $\theta_1 = \theta_2$.

Observe that

$$\alpha(\theta) = \alpha(\theta^*) + 2E\Big[(f(X_1) - g(\theta^*, X_0))$$
$$\cdot(g(\theta^*, X_0) - g(\theta, X_0))\Big]$$
$$+E\Big[(g(\theta^*, X_0) - g(\theta, X_0))^2\Big].$$

Since $f(X_1) - g(\theta^*, X_0)$ is a martingale difference under P, the second term on the right-hand side vanishes. So, under A2, θ^* is indeed the unique global minimizer of $\alpha(\cdot)$, proving our next result.

Proposition 2 *Under A1-A2, $\theta_n^* \to \theta^*$ P a.s. as $n \to \infty$.*

To deal with the central limit theory for the estimator θ_n^*, note that since both $\alpha_n(\cdot)$ and $\alpha(\cdot)$ are smooth, it is evident that $\nabla\alpha_n(\theta_n^*) = \nabla\alpha(\theta^*) = 0$ (note also conditon A1 v)), so that

$$\nabla\alpha_n(\theta_n^*) - \nabla\alpha_n(\theta^*) = \nabla\alpha(\theta^*) - \nabla\alpha_n(\theta^*).$$

Let $\partial_{ij}^2\alpha_n(\xi)$ be the second partial derivative of α_n with respect to θ_i and θ_j, evaluated at ξ. Then, by

Taylor's theorem, there exists ξ_{ni} lying on the line segment connecting θ_n^* and θ^* such that

$$(\partial_{ij}^2 \alpha_n(\xi_{ni}) : 1 \le j \le d)(\theta_n^* - \theta^*) = \partial_i \alpha(\theta^*) - \partial_i \alpha_n(\theta^*)$$

holds for $1 \le i \le d$. Set $H_n = (\partial_{ij}^2 \alpha_n(\xi_{ni}) : 1 \le i, j \le d)$. Then,

$$H_n(\theta_n^* - \theta^*) = \nabla\alpha(\theta^*) - \nabla\alpha_n(\theta^*).$$

The proof of Proposition 1 also carries over to showing that $\partial_{ij}^2 \alpha_n(\cdot)$ converges uniformly P a.s. to $\partial_{ij}^2 \alpha(\cdot)$ on Λ. Since $\theta_n^* \to \theta^*$ P a.s. as $n \to \infty$, it follows that $H_n \to H$ P a.s. as $n \to \infty$, where $H = (\partial_{ij}^2 \alpha(\theta^*) : 1 \le i, j \le d)$. Because θ^* is the unique global minimum of $\alpha(\cdot)$, H is positive definite, and consequently H_n^{-1} exists for n sufficiently large, and $H_n^{-1} \to H^{-1}$ P a.s. as $n \to \infty$. Hence,

$$\theta_n^* - \theta^* = H_n^{-1}(\nabla\alpha(\theta^*) - \nabla\alpha_n(\theta^*)).$$

But

$$\nabla\alpha(\theta^*) - \nabla\alpha_n(\theta^*) = -\frac{1}{n}\sum_{i=1}^{n} D_i,$$

where $D_i = 2(f(X_i) - g(\theta^*, X_{i-1}))\nabla g(\theta^*, X_{i-1})$. Now, $(D_i : i \ge 1)$ is a stationary sequence of square-integrable martingale differences, and consequently the martingale central limit theorem (CLT) (see Ethier and Kurtz (1986)) yields

$$\sqrt{n}\,(\nabla\alpha(\theta^*) - \nabla\alpha_n(\theta^*)) \xrightarrow{d} N(0, C)$$

as $n \to \infty$ where $N(0, C)$ is a d–dimensional multivariate normal random variable having covariance matrix $C = ED_1 D_1^T$. We have therefore established the following CLT for θ_n^*.

Theorem 1 *Under A1-A2,*

$$\sqrt{n}\,(\theta_n^* - \theta^*) \xrightarrow{d} H^{-1}N(0, C)$$

as $n \to \infty$.

The above analysis presupposes that $g(\cdot)$ can be easily evaluated, so that θ_n^* can be computed without difficulty. As indicated earlier, we are especially concerned with problems in which $g(\cdot)$ is computed via simulation, thereby introducing additional error into our estimator of θ^*; this is the subject of Section 4.

4 GRID–BASED SIMULATION

Clearly, the conditional expectations associated with $g(\theta)$ can easily be computed via simulation of X. Specifically, suppose that $(W_j(i, \theta, x) : i, j \ge 1, \theta \in \Lambda, x \in S)$ is a collection of independent random variables in which $(W_j(i, \theta, x) : i, j \ge 1)$ is identically distributed with common distribution $P_\theta(f(X_1) \in \cdot \mid X_0 = x)$. We let $\tilde{P}(\cdot)$ and $\tilde{E}(\cdot)$ denote the probability and expectation operator associated with the probability space that supports the $W_j(i, \theta, x)$'s and the process X. Then,

$$g(\theta, x, i, m) = \frac{1}{m}\sum_{j=1}^{m} W_j(i, \theta, x)$$

is an estimator of $g(\theta, x)$. Furthermore, $\alpha_n(\theta)$ can be calculated numerically via the Monte Carlo estimator

$$\alpha_n(\theta, i, m) = \sum_{x,y}(f(y) - g(\theta, x, i, m))^2 \pi_n(x, y). \quad (5)$$

We note that, computationally speaking, $\alpha_n(\theta, i, m)$ requires only that $g(\theta, x, i, m)$ be calculated for states $x \in \{ X_0, \ldots, X_{n-1} \}$; this observation can result in significant computational savings when $|S|$ is (very) large.

However, it is clearly impossible to compute the function $\alpha_n(\cdot, i, m)$ over the entire parameter space Λ. Instead, one needs to restrict attention to a finite subset of Λ. Our approach will be to generate $\alpha_n(\cdot, i, n)$ on a uniform grid (hence, the term "grid–based simulation"). The grid will then be successively refined as more information becomes available on the likely location of the minimizer of $\alpha_n(\cdot)$.

More specifically, the iteration proceeds as follows. Suppose that at iteration i, we have a "guess" $\theta_n^*(i-1)$ available as to the likely location of some point in the set $\operatorname{argmin}\{ \alpha_n(\theta) : \theta \in \Lambda \}$. For $z > 0$, let

$$I(z) = \{ (i_1, \ldots, i_d) : i_j \in \mathcal{Z},\ |i_j| \le z,\ 1 \le j \le d \}.$$

For δ positive, we then proceed to generate $\alpha_n(\theta, i, m_n(i))$ over the grid points $\theta \in \Lambda_n(i)$, where

$$\Lambda_n(i) = (\theta_n^*(i-1) + n^{-(i+1)\delta} I(n^{2\delta})) \cap \Lambda.$$

We next select $\theta_n^*(i)$ to be any point in the set of global minimizers of $\{ \alpha_n(\theta, i, m_n(i)) : \theta \in \Lambda_n(i) \}$, and move on to the next iteration. The algorithm is initiated by setting $\theta_n^*(0) = 0$, and is terminated at iteration k with the final computed parameter estimator $\widehat{\theta}_n = \theta_n^*(k)$. Our choice for the sequence $\{ m_n(i) : i \ge 1 \}$ is $m_n(i) = [n^{4(i+2)\delta}]$ (where $[\cdot]$ denotes the greatest integer).

Let $\|\cdot\|$ be the norm on $\mathbb{R}^d$ defined by $\|x\| = \max_{1 \le i \le d} |x_i|$. Our main mathematical result of this section is the following.

Proposition 3 *Assume A1-A2. For each fixed* $k \ge 1$ *and* $\delta > 0$,

$$\tilde{P}\left(\|\theta_n^* - \theta_n^*(i)\| \le n^{-i\delta},\ 1 \le i \le k\right) \to 1$$

as $n \to \infty$.

Proof. Let $\mathcal{G}_{ni} = \sigma(X_0, \ldots, X_1,\ W_j(l,\theta,x) : \theta \in \Lambda,\ x \in S, j \geq 1,\ l \leq i)$. Then,

$$\tilde{P}\left(\| \theta_n^* - \theta_n^*(i) \| \leq n^{-i\delta},\ 1 \leq i \leq k\right) \tag{6}$$
$$= \tilde{E}\Bigg[\prod_{i=1}^{k-1} I(\| \theta_n^* - \theta_n^*(i) \| \leq n^{-i\delta}) \cdot \tilde{P}(\| \theta_n^* - \theta_n^*(k) \| \leq n^{k\delta} \mid \mathcal{G}_{n,k-1})\Bigg].$$

On the event $\{ \| \theta_n^* - \theta^*(k-1) \| \leq n^{(k-1)\delta} \}$, the convex hull of $\Lambda_n(k)$ contains θ_n^*. Hence, there exists a point $\theta_{nc}^*(k) \in \Lambda_n(k)$ for which $\| \theta_n^* - \theta_{nc}^*(k) \| \leq n^{-\delta(k+1)}$. We next observe that a sufficient condition for $\theta_n^*(k)$ to be within $n^{-\delta k}$ of θ_n^* (in the norm $\| \cdot \|$) is that $\theta_{nc}^*(k)$ have a strictly smaller objective value (with regard to the objective function $\alpha_n(\cdot, k, m_n(k))$ than all those points $\theta \in \Lambda_n(k)$ such that $\| \theta_n^* - \theta \| > n^{-k\delta}$. We now proceed to establish that this event occurs with high probability under our hypotheses.

Observe that for any positive deterministic $\tilde{a}$, there exists deterministic n_0 such that for $n > n_0$,

$$\begin{aligned}
&\{\ \alpha_n(\theta, k, m_n(k)) > \alpha_n(\theta_{nc}^*(k), k, m_n(k)), \\
&\qquad \theta \in \Lambda_n(k),\ \| \theta - \theta_n^* \| > n^{-k\delta}\ \} \\
&\supseteq \{\ \alpha_n(\theta) > \alpha_n(\theta_{nc}^*(k)) + \tilde{a} n^{-2k\delta},\ \theta \in \Lambda_n(k), \\
&\qquad \| \theta - \theta_n^* \| > n^{-k\delta}\ \} \\
&\cap \{\ |\alpha_n(\theta) - \alpha_n(\theta, k, m_n(k))| \leq n^{-2(k+1)\delta}, \\
&\qquad \theta \in \Lambda_n(k)\ \}.
\end{aligned}$$

Hence, for n sufficiently large,

$$\begin{aligned}
&\tilde{P}\Big(\alpha_n(\theta, k, m_n(k)) > \alpha_n(\theta_{nc}^*(k), k, m_n(k)), \qquad (7)\\
&\qquad \theta \in \Lambda_n(k), \| \theta - \theta_n^* \| > n^{-k\delta} \mid \mathcal{G}_{n,k-1}\Big) \\
&\geq I\Big(\alpha_n(\theta) > \alpha_n(\theta_{nc}^*(k)) + \tilde{a} n^{-2k\delta}, \theta \in \Lambda_n(k), \\
&\quad \| \theta - \theta_n^* \| > n^{-k\delta}\Big) \\
&\quad \cdot \prod_{\theta \in \Lambda_n(k)} \tilde{P}\Big(\ |\alpha_n(\theta) - \alpha_n(\theta, k, m_n(k))\ | \\
&\qquad \leq n^{-2(k+1)\delta} \mid \mathcal{G}_{n,k-1}\Big).
\end{aligned}$$

(We used above the fact that $\theta_{nc}^*(k)$ is $\mathcal{G}_{n,k-1}$ measurable and the independence of the $W_j(i,\theta,x)$'s.) Since θ_n^* is a minimizer of $\alpha_n(\cdot)$, $\nabla \alpha_n(\theta_n^*) = 0$ so Taylor's theorem (see, for example Sen and Singer (1993)) yields

$$\alpha_n(\theta) = \alpha_n(\theta_n^*) + \frac{1}{2}(\theta - \theta_n^*)^T H_n(\xi_n(\theta))(\theta - \theta_n^*),$$

where $\xi_n(\theta)$ lies on the line segment connecting θ and θ_n^*, and $H_n(x)$ is the Hessian of $\alpha_n(\cdot)$ evaluated at x. Now, as asserted earlier, $H_n(\cdot)$ is a continuous matrix function which converges uniformly on Λ, as $n \to \infty$, to the limiting continuous matrix function $H(\cdot)$, where $H(\cdot)$ is the Hessian of $\alpha(\cdot)$. Furthermore, θ^* is the unique global minimizer of $\alpha(\cdot)$, so $H(\cdot)$ is positive definite in a neighborhood of θ^*. Now, the minimal and maximal eigenvalues of a positive definite matrix are continuous functions of their matrix argument. As a consequence, the minimal and maximal eigenvalues of $H_n(\cdot)$ converge uniformly to the corresponding eigenvalues of $H(\cdot)$. Thus, it is evident that there exists positive constants ϵ, $\tilde{c}$, $\tilde{d}$ such that for any $x \in \mathbb{R}^d$, $\| \theta - \theta^* \| < \epsilon$,

$$\tilde{c} \| x \|^2 \leq x^T H_n(\theta) x \leq \tilde{d} \| x \|^2$$

for n large enough $\tilde{P}$ a.s. (We also use here the fact that our norm can be bounded above and below by constant multiples of the Euclidian norm.) Now, for $k \geq 2$, the diameter of $\Lambda_n(k)$ shrinks as $n \to \infty$. Consequently, on $\{ \| \theta_n^* - \theta_n^*(k-1) \| \leq n^{-\delta(k-1)} \}$, and for n large enough,

$$\tilde{c} \| \theta - \theta_n^* \|^2 \leq \alpha_n(\theta) - \alpha_n(\theta_n^*) \leq \tilde{d} \| \theta - \theta_n^* \|^2 \tag{8}$$

for $\theta \in \Lambda_n(k)$, $k \geq 2$. In view of the fact that $\| \theta_{nc}^*(k) - \theta_n^* \| \leq n^{-\delta(k+1)}$ it follows from (8) that for $k \geq 2$, there exists a $\mathcal{G}_{n,1}$ measurable random variable N_k that is finite $\tilde{P}$ a.s. such that

$$\begin{aligned}
&I\Big(\alpha_n(\theta) > \alpha_n(\theta_{nc}^*(k)) + \tilde{a} n^{-2k\delta}, \qquad (9)\\
&\qquad \theta \in \Lambda_n(k),\ \| \theta - \theta_n^* \| > n^{-k\delta}\Big) \\
&\qquad = I\,(n \geq N_k)
\end{aligned}$$

on $\{ \| \theta_n^* - \theta_n^*(k-1) \| \leq n^{-\delta(k-1)} \}$ (if we choose a small enough δ. For $k = 1$, we use the fact that $\alpha_n(\cdot)$ converges uniformly to $\alpha(\cdot)$ outside any ϵ–neighborhood of θ_n^* and use the estimates (8) inside the ϵ–neighborhood, to arrive at (9).

Turning now to the second factor on the right–hand side of (7), observe that the $W_j(i,\theta,x)$'s are a family of uniformly bounded random variables (bounded by $\max(\ |f(x)|\ ) : x \in S$). It is easy to see that

$$\tilde{P}\Big(\mid \alpha_n(\theta) - \alpha_n(\theta,k,m_n(k)) \mid \leq n^{-2(k+1)\delta} \mid \mathcal{G}_{n,k-1} \Big) \tag{10}$$
$$\geq \prod_{x\in S} \tilde{P}\Big(\mid g(\theta,x) - g(\theta,x,k,m_n(k)) \mid \leq \tilde{b}n^{-2(k+1)\delta} \Big)$$

for $\tilde{b}$ chosen small enough and deterministic (see (5)). But for $p > 0$,

$$\tilde{P}\Big(\mid g(\theta,x) - g(\theta,x,k,m_n(k)) \mid \leq \tilde{b}n^{-2(k+1)\delta} \Big) \tag{11}$$
$$= 1 - \tilde{P}\Big(\mid g(\theta,x) - g(\theta,x,k,m_n(k)) \mid > \tilde{b}n^{-2(k+1)\delta} \Big)$$
$$\geq 1 - \tilde{b}^{-p} n^{2(k+1)\delta p} \cdot \tilde{E} \mid g(\theta,x) - g(\theta,x,k,m_n(k)) \mid^p$$
$$\geq 1 - \tilde{b}^{-p} n^{2(k+1)\delta p} c(p) \cdot \max\big(\mid f(x) \mid^p : x \in S \big) m_n(k)^{-p/2}$$
$$= 1 - \tilde{b}^{-p} c(p) n^{-2\delta p} \max\big(\mid f(x) \mid^p : x \in S \big)$$
$$\triangleq 1 - r n^{-2\delta p};$$

the Burkholder inequality was applied in the second inequality; see Hall & Heyde (1980), p. 23. Application of (7), (9), (10), and (11), together with repeated conditioning in (6), yields the inequality

$$\tilde{P}\left(\| \theta_n^* - \theta_n^*(i) \| \leq n^{-i\delta},\ 1 \leq i \leq k \right)$$
$$\geq \tilde{P}(N_1 \leq n, \ldots, N_k \leq n) \cdot \left(1 - r n^{-2\delta p}\right)^{|S| \cdot |\Lambda_n(k)| \cdot k}$$
$$\geq \tilde{P}(N_1 \leq n, \ldots, N_k \leq n) \cdot \left(1 - r n^{-2\delta p}\right)^{|S| \cdot (2n^{2\delta}+1)^d k}.$$

By choosing p sufficiently large and letting $n \to \infty$, we obtain the desired result. □

We note that by choosing $k\delta > \frac{1}{2}$, Proposition 3 and Theorem 1 combine to yield a CLT for $\widehat{\theta}_n$.

Theorem 2 *Assume A1-A2 and suppose* $k\delta > \frac{1}{2}$. *Then*

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) \xrightarrow{d} H^{-1} N(0, C)$$

as $n \to \infty$, *where* H *and* C *are as in Theorem 1.*

In terms of the computational effort required to calculate $\widehat{\theta}_n$, note that the i'th iteration requires simulation at $(2n^{2\delta}+1)^d$ points. Each simulation at the i'th iteration, for a given point, requires $n^{4(i+2)\delta}$ replications. Thus, the total work required at iteration i is of the order of $n^{2\delta d + 4(i+2)\delta}$. Summing over the k iterations, we conclude that the total work is of order $n^{2\delta d + 4(k+2)\delta}$. But k and δ can be chosen arbitrarily, subject to the constraint $k\delta > \frac{1}{2}$. Hence, by (for example), choosing the number of iterations k large, and $\delta = (\frac{1}{2} + \eta)/k$ for η positive, we note that we can make the exponent $2\delta d + 4(k+2)\delta$ as close as we wish to 2. Thus, roughly speaking, the computational effort required to compute $\widehat{\theta}_n$ is of order n^2.

This should come as no surprise. In the limit, an accuracy of order $n^{-1/2}$ in the location of the minimizer requires that we perform "function evaluations" that have accuracy n^{-1} (because of the locally quadratic structure of the objective function). To obtain simulations of accuracy n^{-1} requires a run-length of order n^2.

While the analysis of this paper is asymptotic, it does suggest that in implementing grid-based simulation, it is important to slowly refine the grid (i.e. k large), and that using a course grid (δ small) reduces the impact of dimensionality considerations (i.e. the impact of d being large).

ACKNOWLEDGMENTS

This collaborative work began during the first author's National Science Foundation sponsored visits to Stanford University. The work of the second author was supported by the Army Research Office under Contract No. DAAL03-91-G-0319.

REFERENCES

Billingsley, P. 1961. *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

Bridges, E., K. B. Ensor and J. R. Thompson. 1992. Marketplace competition in the personal computer industry. *Decision Sciences* 23:467-477.

Cooke, J. R. and L. A. Stefanski. 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89:1314–1328.

Diggle, P. J. and R. J. Gratton. 1984. Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society Series B* 46:193–227.

Duffie, D. and K. J. Singleton. 1993. Simulated moments estimation of Markov models of asset prices.

Econometrica 61:929–952.

Ensor, K. B.. 1994. Properties of simulation based estimators of stochastic processes. *Proceedings of the Thirty-Ninth Conference on the Design of Experiments. ARO Report No. 94-2*, 15-22.

Ethier, S. N. and T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence.* New York: John Wiley & Sons.

Hall, P. and C. C. Heyde. 1980. *Martingale Limit Theory and its Application.* New York: Academic Press.

Keane, M. P. 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62:95–116.

Klimko, L. A. and P. I. Nelson. 1978. On conditional least squares estimation for stochastic processes. *The Annals of Statistics* 6:629–642.

Maa, J.-F., Pearl, D. K., Batoszyński, R., and Horn, D. 1993. Simulation–based optimization and estimation. *ASA Proceedings of the Statistical Computing Section*, 102–106.

Lee, L.–F. 1992. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory* 8:518–552.

McFadden, D. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57:995–1026.

Pakes, A. and D. Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57:1027–1057.

Sen, P. K. and J. M. Singer. 1993. *Large Sample Methods in Statistics. An Introduction with Applications.* Chapman & Hall.

Thompson, J. R., B.W. Brown, and E. N. Atkinson. 1987. SIMEST An algorithm for simulation–based estimation of parameters characterizing a stochastic process. In *Cancer Modeling*, eds. J. R. Thompson and B. W. Brown,387–415. Marcel Dekker.

AUTHOR BIOGRAPHIES

KATHERINE B. ENSOR is Associate Professor in the Department of Statistics at Rice University. She joined the faculty at Rice University in 1987 after receiving her Ph.D. in Statistics from Texas A&M University. Her research interests include time series analysis, spatial processes, environmental statistics, and estimation for stochastic processes based on simulation methods. She serves on the editorial board of *The Journal of Statistical Computation and Simulation* and *Communications in Statistics.*

PETER W. GLYNN received his Ph.D. from Stanford. University, after which he joined the faculty of the Department of Industrial Engineering at the University of Wisconsin-Madison. In 1987, he returned to Stanford, where he currently holds the Thomas Ford Faculty Scholar Chair in the EES/OR Department. He was a co–winner of the 1993 Outstanding Simulation Publication Award sponsored by the TIMS College on Simulation. His research interests include discrete-event simulation, computational probability, queueing, and general theory for stochastic systems.