Proceedings of the 1993 Winter Simulation Conference
G.W. Evans, M. Mollaghasemi, E.C. Russell, W.E. Biles (eds.)

PARAMETRIC INFERENCE FOR GENERALIZED SEMI-MARKOV PROCESSES

Halim Damerdji

Department of Industrial Engineering
North Carolina State University
Raleigh, NC 27695–7906, U.S.A.

ABSTRACT

Part of the inputs to the simulation of a real-world discrete-event stochastic system will be obtained via a statistical analysis of the system. Also, a number of discrete-event stochastic systems can be modeled as a generalized semi-Markov process.

The basic building blocks of a generalized semi-Markov process are the probability distributions of the event lifetimes and the routing probabilities. We assume the form of these distributions and routing probabilities is known but depends on an unknown vector parameter, which must be estimated from the data. In this parametric inference setting we show that, under certain conditions, the maximum likelihood estimator exists, is consistent, and obeys a central limit theorem. A related estimator, easier to compute, is shown to also be consistent. Use of this other estimator results in no loss in statistical efficiency.

1 INTRODUCTION

Simulation is often used to study design changes and alternatives to an already existing system. There will, then, most likely be a statistical analysis of this system so as to estimate some of the needed inputs to the simulation.

The dynamics of a number of real-world systems is well captured by a *generalized semi-Markov process* (GSMP), i.e., the system under study may adequately be modeled by a GSMP. Central to a GSMP are its set of *states* and set of *events*. When an event triggers a transition, the GSMP moves from its current state to some other state with a certain routing probability. It then stays in this new state for a certain length of time, until another transition is triggered by some event, and so on. The routing probabilities and event lifetime distributions are the basic building blocks of the GSMP.

In this paper, we only tackle the parametric statistical inference case, that is, we assume the event-lifetime distributions and routing probabilities have a known form but depend on a finite number of unknown parameters. These must be estimated from one long observation of the process. We consider maximum likehood estimation and show that, under certain conditions, the maximum likelihood estimators (MLE's) do exist and do obey a central limit theorem (CLT). (Of course, the CLT gives then a rate of convergence.) We also consider estimators closely related to the MLE's but easier to compute, and show that there is no loss in statistical efficiency in considering these estimators as surrogates to the MLE's. This paper is based on Damerdji (1992).

GSMP's are defined in Section 2, while Section 3 is a review of some of the properties of maximum likelihood estimation. The likelihood function of the GSMP is introduced in Section 4. The main results of the study are given in Section 5 and the conclusion in Section 6.

2 THE GSMP

The state space S of the GSMP is countable in the general theory, but here, we assume it is finite. The event set, denoted E, is finite, i.e., there is a finite number $|E|$ of events. Denote the GSMP process by $\{Q(u) : u \geq 0\}$. We assume that we not only observe the GSMP up to some large time t, but that we also observe which events trigger the transitions. (Observing the GSMP alone may not provide that information.) Another assumption is that we know what the system "looks like" when we start observing the GSMP at time 0. (If the GSMP models a job-shop, say, we would probably know how long the current jobs have been in service, the elapsed time since the last order, when the machines failed last, etc.)

The routing probabilities and event-lifetime distributions depend on an unknown r-dimensional parameter $\theta \in \Theta$, where Θ is an open set in $\mathcal{R}^r$. If the GSMP is in some state s and event i triggers the transition, the GSMP moves to a state s' with probability $p(s'; s, i; \theta)$. At the time of the transition (in fact, right after), new events j may be generated with a lifetime following the distribution function $F(.; s', j, s, i; \theta)$, and certain "old events" may get cancelled. The other "old events" simply continue their lifetime, until they either die out, in which case they will then be triggering a transition themselves, or get cancelled before then. We bar event cancelling in this study, and so all events will be observed over their full lifetimes. It is discussed in Section 4 that different arguments must be made when event cancelling is allowed.

A $GI/G/1$ queue can be modeled as a GSMP with state space $S = \{0, 1, 2, \ldots\}$, where the current state represents the current number of customers. (Note that S would not be finite here.) When there is no customer in the system (i.e., when the GSMP is in state $s = 0$), the only possible event which may occur next is the arrival of a customer. When there is one or more customers in the system (i.e., when $s \geq 1$), the possible events are an arrival or an end-of-service. If the next event is an arrival (resp. an end-of-service), the GSMP moves from state s to state $s + 1$ (resp. $s - 1$) with probability one. The interarrival times and service times are distributed according to their respective cumulative distribution functions.

Embedded in the GSMP is a general state-space discrete-time Markov chain $\{X_n = (s_n, c_n) : n \geq 0\}$, where s_n is the state occupied right after the nth transition. If $N(t)$ is the number of transitions by time t, the GSMP will be at that time in state $Q(t) = s_{N(t)}$. The component c_n of the embedded Markov chain is actually a vector of $(|E| + 1)$ "clocks." Clock 0 indicates the elapsed time between transitions $n - 1$ and n. Each other clock is associated with an event: if event i is active at time T_n, where T_n is the epoch of the nth transition, its corresponding clock reading $c_{i,n}$ indicates the elapsed time since it was generated last. Following Glynn's (1988) notation, an inactive event is set, by convention, to have a clock reading of -1. Newly generated events have a clock reading of 0, i.e., $c_{i,n} = 0$ if event i is newly generated. If event i is an old event, that is that i was active before the transition occured and event i did not trigger the transition, then $c_{i,n} = c_{i,n-1} + c_{0,n}$. Here, we are looking at clocks with time running up, as in König, Matthes, and Nawrotzki (1967) and Glynn (1988). Another representation of the vector clocks is with time running down, i.e., the clocks show the residual lifetime of these events, as in Whitt (1980); this is a natural representation in a discrete-event simulation. In our study, the clocks represent the elapsed times, since we are in a statistical experiment setting, and cannot therefore anticipate the future as in the other representation.

For simplicity of notation, let us assume that the distribution of an event j depends only on the event itself and the unknown parameter, and so is denoted $F_j(.; \theta)$. (Recall that the general formulation of GSMP's allows the distribution of an event to also depend on the current and previous states and the event which triggered the last transition.) Let $\bar{F}_j(.; \theta) \equiv 1 - F_j(.; \theta)$ be the residual lifetime distribution of event j. We assume that the support of the distributions is $(0, \infty)$. This also implies that the supports of the event distributions do not depend on the unknown parameter. We also assume that the distribution function $F_j(.; \theta)$ of an event j admits a density, denoted $f_j(.; \theta)$. Following Whitt's (1980) notation, let $E(s)$ be the set of active events in state s, and $N(s', s, i)$ (resp. $O(s', s, i)$) the set of new events (resp. old events) in state s' when it is event i which just triggered the transition from state s.

From Glynn (1988) and Damerdji (1992), the transition density function $h(x, x'; \theta)$ of the embedded Markov chain is given by

$$
\begin{aligned}
h\Big(x = (s, c), x' = (s', c'); \theta\Big) = & \\
& \sum_{i \in E(s)} \Big\{ p(s'; s, i; \theta) \frac{f_i(c_i + c_0'; \theta)}{\bar{F}_i(c_i; \theta)} \\
& \bullet \prod_{j \in O(s', s, i)} \frac{\bar{F}_j(c_j'; \theta)}{\bar{F}_j(c_j; \theta)} I[c_j' = c_j + c_0'] \\
& \bullet \prod_{j \in N(s', s, i)} I[c_j' = 0] \prod_{j \notin E(s)} I[c_j' = -1] \Big\}.
\end{aligned}
$$

Here, $I[.]$ denotes the indicator function. These indicators appear for consistency. Note that the transition density function is a complicated expression of some basic building blocks of the GSMP. Because of the indicators, the transition density function is also not a smooth function.

3 REVIEW OF MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood (ML) estimation is undoubtedly one of the most powerful parametric statistical inference techniques. The maximum likelihood estimator is the estimator which is explained best by the data in the following sense. Suppose $X_0, \ldots, X_n$ is the observed sample, with known distribution P^θ. This known distribution depends, however, on an unknown parameter θ. We assume there exists a parameter $\theta^0 \in \Theta$, which is the true parameter of the distribution the population is drawn from. The objective is of course to estimate the true parameter given the observed sample. Technically, the likelihood function (LF) $\mathcal{L}_n(\theta)$ is defined as the Radon-Nikodym derivative of the absolutely continuous part of P^θ with respect to P^{θ^0}, both restricted to the history of the process up to time n. See Billinglsey (1986) for the various definitions.

The MLE $\hat{\theta}_n$, if it exists, is defined as the maximizer of the LF, i.e., $\mathcal{L}_n(\hat{\theta}_n) \geq \mathcal{L}_n(\theta)$ for all $\theta \in \Theta$. The LF is simple to compute for i.i.d. random variables and also for certain stochastic processes (e.g., birth and death processes and countable Markov chains). However, the LF may be intractable for a number of other stochastic processes.

Maximum likelihood estimation has a number of desirable properties, which we now discuss. For simplicity, let us assume, in this section, that the parameter θ is univariate. First, it is often possible to show that a CLT of the form $\sqrt{n}(\hat{\theta}_n - \theta^0) \Longrightarrow N(0,\zeta^2)$, as $n \to \infty$, is in force. (The symbol "$\Longrightarrow$" stands for convergence in distribution, while $N(0,\zeta^2)$ denotes the centered Normal distribution with variance ζ^2.) If such a CLT holds, the MLE $\hat{\theta}_n$ converges then to the true parameter θ^0 at rate $1/\sqrt{n}$. Under certain regularity conditions, the variance ζ^2 is equal to the Rao-Cramér lower bound, in which case the MLE would be asymptotically efficient. Under certain conditions, the MLE may also be asymptotically sufficient, and there would then be no loss of statistical information (asymptotically) when grouping the data into the statistic. For all these reasons, ML estimation, when feasible, leads to estimators with a number of desirable properties, at least asymptotically.

A difficulty in trying to apply the method is that one must compute the LF, which may be infeasible as previously mentioned. Also, one must impose stringent conditions in order to guarantee the existence and the consistency of the MLE. It is also known that the MLE is often a biased estimator. From a numerical point of view, it is not guaranteed that the optimization routine used in order to maximize the LF will converge to the MLE. Nonetheless, despite these drawbacks, ML estimation is a powerful statistical inference technique. See Cox and Hinkley (1974) for a comprehensive treatment.

Basawa and Prakasa Rao (1980) discuss ML estimation for a number of discrete-time stochastic processes (e.g., autoregressive processes, moving average processes, branching processes, and Markov chains), and continuous-time processes (continuous-time Markov chains, diffusion processes, and point processes). The major references on ML estimation for Markov chains are Billinsgley (1961a and 1961b). Billingsley (1961a) considers discrete-time countable state-space Markov chains, while ML estimation for discrete-time general state-space Markov chains is undertaken in Billinsgley (1961b).

Maximum likelihood estimation for continuous-time stochastic processes is in general harder. The more modern approach is via martingale theory. The other approach is via a transformation of the problem into a discrete-time problem. For example, Billingsley (1961b) tackles continuous-time countable state-space Markov chains by using the fact that the continuous-time Markov chain can be embedded into a discrete-time general state-space Markov chain. The arguments follow then from the theory of ML estimation applied to the discrete-time scale. Moore and Pyke (1968) also use an embedding technique to study ML estimation for semi-Markov processes. This will also be the approach used in our study.

4 THE LIKELIHOOD FUNCTIONS

We first discuss the differences between Billingsley (1961b) and our work here. Recall that to be of practical interest, the various assumptions made in order to guarantee the existence of the MLE must be on the basic building blocks of the GSMP, i.e., the routing probabilities and the event distributions. Billingsley (1961b) gives conditions on the transition density function. However, in our setting, the density function involves the basic building blocks in a nontrivial way, and it is therefore not evident how one could translate these general conditions into conditions on the basic building blocks. Consequently, the approach we use will lead to conditions on the basic building blocks. Another difference is that we use structural properties of the GSMP, and appeal to simple CLT's for i.i.d. random variables in order to obtain our various CLT's. Billingsley (1961b), on the other hand, uses a CLT for martingales. Our conditions here too will depend on the basic building blocks only. Another (measure-theoretic) difference between these two works is the following: the measure associated with the transition density function $h(x, x'; \theta)$ of the embedded Markov chain of the GSMP depends on the point x (see Damerdji, 1992). This does not fit into Billingsley's (1961b) framework, and hence, we could not apply some of Billingsley's arguments. Billingsley (1961b) also assumes conditions which directly guarantee ergodicity of the embedded Markov chain. In our study, we use strong laws of large numbers results from Glynn (1988).

We now provide heuristics in order to compute the likelihood function $L_t(\theta)$ of the GSMP observed up to time t. (See Section 10 of Damerdji (1992) for a proof.) Some more notation is needed at this point. Let Y_j be the random variable which denotes the lifetime of event j, that is, Y_j has distribution function $F_j(.;\theta)$. Let i_k be the index of the event which triggers the $(k+1)$st transition and m_k the index of occurrence of i_k. Let Y_{i_k,m_k} be the lifetime of event i_k when it triggered the transition. (For example, if the 24th transition gets triggered by Event 5, and this particular Event 5 corresponds to the one that got generated for the second time, the lifetime $Y_{5,2}$ of Event 5 is denoted $Y_{i_{23},m_{23}}$.) Therefore, if x_k and x_{k+1} are the states of the embedded Markov chain at transitions k and $k+1$, respectively, we have that the value of $f_i(c_i+c_0';\theta)$ which appears in the expression of $h(x_k,x_{k+1};\theta)$ is $f_{i_k}(Y_{i_k,m_k};\theta)$.

Observing the GSMP up to time t is equivalent to observing the embedded Markov chain $X_0,\ldots,X_{N(t)}$, plus observing the GSMP from $T_{N(t)}$ until t. The likelihood function $\mathcal{L}_n(\theta)$ of the embedded Markov chain $X_0,\ldots,X_n$, observed up to its nth transition, is given by

$$\mathcal{L}_n(\theta) \quad = \quad \eta(x_0;\theta)\prod_{k=0}^{n-1}\frac{h(x_k,x_{k+1};\theta)}{h(x_k,x_{k+1};\theta^0)}.$$

(The term $\eta(x_0;\theta)$ is the Radon-Nikodym density stemming from the initial distribution. See Damerdji (1992).) In observing $X_0,\ldots,X_n$, we would know exactly which events triggered the transitions, i.e., we would know $i_0,\ldots,i_{n-1}$. The contribution of $h\Big(x_k=(s_k,c_k),x_{k+1}=(s_{k+1},c_{k+1});\theta\Big)$ to the likelihood function $\mathcal{L}_n(\theta)$ is then

$$p(s_{k+1};s_k,i_k;\theta)\frac{f_{i_k}(Y_{i_k,m_k};\theta)}{\bar F_{i_k}(c_{i_k,k};\theta)} \bullet \prod_{j\in O(s_{k+1},s_k,i_k)}\frac{\bar F_j(c_{j,k+1};\theta)}{\bar F_j(c_{j,k};\theta)}.$$

Although this last term has a complicated expression, it turns out that a number of simplifications occur in $\prod_{k=0}^{n-1}h(x_k,x_{k+1};\theta)$. The above term can be rewritten as

$$p(s_{k+1};s_k,i_k;\theta)f_{i_k}(Y_{i_k,m_k};\theta) \bullet\Big(\frac{1}{\bar F_{i_k}(c_{i_k,k};\theta)}\prod_{j\in O(s_{k+1},s_k,i_k)}\frac{\bar F_j(c_{j,k+1};\theta)}{\bar F_j(c_{j,k};\theta)}\Big).$$

When multiplying out the $h(x_k,x_{k+1};\theta)$'s, note that all the residual lifetimes in the numerator will have a counterpart in the denominator of $h(x_{k+1},x_{k+2};\theta)$, and hence would cancel out. We are then left with

$$\prod_{k=0}^{n-1}h(x_k,x_{k+1};\theta) = \prod_{k=0}^{n-1}\Big(p(s_{k+1};s_k,i_k;\theta)f_{i_k}(Y_{i_k,m_k};\theta)\Big) \bullet\frac{1}{\prod_{j\in E(s_0)}\bar F_j(c_{j,0};\theta)}\prod_{j\in O(s_n,s_{n-1},i_{n-1})}\bar F_j(c_{j,n};\theta)$$

The terms which were in the denominator (resp. numerator) of $h(x_0,x_1;\theta)$ (resp. $h(x_{n-1},x_n;\theta)$) did not cancel out since they did not have a counterpart. It is important to note that if event cancelling were permitted, certain terms would not cancel out, and should appear in the expression of the likelihood function. So far, we have been discussing the likelihood function of the embedded discrete-time Markov chain $X_0,\ldots,X_n$. However, we are interested in the likelihood function of the GSMP up to time t. We observe $X_0,\ldots,X_{N(t)}$, but there will also be extra terms in the LF stemming from the fact that there was no transition between $T_{N(t)}$ and t. Incorporating these extra terms will cancel out the numerator-terms (due to the end effects) appearing in the previous expression. Heuristically, we then obtain

$$L_t(\theta) = \prod_{k=0}^{N(t)-1}\frac{p(s_{k+1};s_k,i_k;\theta)f_{i_k}(Y_{i_k,m_k};\theta)}{p(s_{k+1};s_k,i_k;\theta^0)f_{i_k}(Y_{i_k,m_k};\theta^0)} \bullet\prod_{j\in E(s_0)}\frac{\bar F_j(c_{j,0};\theta^0)}{\bar F_j(c_{j,0};\theta)} \bullet\prod_{j\in E(s_{N(t)})}\frac{\bar F_j(c_{j,N(t)}+t-T_{N(t)};\theta)}{\bar F_j(c_{j,N(t)}+t-T_{N(t)};\theta^0)}.$$

See Glynn (1988) for the likelihood function of a GSMP up to a stopping time (see Billingsley (1986) for the definition of a stopping time), and Glynn and Iglehart (1989) for the expression of the likelihood function of the GSMP with the other clock representation.

It is typically simpler to work with the log-likelihood function instead of the likelihood function. Let

$$g(x_k,x_{k+1};\theta)\equiv\log\frac{p(s_{k+1};s_k,i_k;\theta)f_{i_k}(Y_{i_k,m_k};\theta)}{p(s_{k+1};s_k,i_k;\theta^0)f_{i_k}(Y_{i_k,m_k};\theta^0)},$$

and call the approximate log-likelihood function (ALLF)

$$\tilde l_t(\theta) \quad\equiv\quad \sum_{k=0}^{N(t)-1} g(x_k,x_{k+1};\theta).$$

The quantity $\eta(x_0;\theta)$ is based on a single observation and so, has negligible effect as the sample size gets large. We then drop this term from the expression of the likelihood function. If $z_t(\theta)$ is defined as

$$z_t(\theta) \equiv - \sum_{j\in E(s_0)} \log \frac{\bar{F}_j(c_{j,0};\theta)}{\bar{F}_j(c_{j,0};\theta^0)} + \sum_{j\in E(s_{N(t)})} \log \frac{\bar{F}_j(c_{j,N(t)}+t-T_{N(t)};\theta)}{\bar{F}_j(c_{j,N(t)}+t-T_{N(t)};\theta^0)},$$

the log-likelihood function (LLF) $l_t(\theta) \equiv \log L_t(\theta)$ of the GSMP can then be rewritten as

$$l_t(\theta) \;=\; \tilde{l}_t(\theta) \;+\; z_t(\theta).$$

The term $z_t(\theta)$ is due to the end-effects. Note that the ALLF is simpler to maximize than the LLF, since the latter involves the end-effects. If no loss of statistical efficiency entails, it is then preferable to maximize the ALLF instead. We will show that a maximizer of the ALLF exists. We call this maximizer the approximate maximum likelihood estimator (AMLE). We also show that the true MLE exists, and that, importantly, there is no loss in statistical efficiency in computing the AMLE instead. This is discussed in the next section.

As an example, let us consider an $M/M/1$ queue, with arrival rate λ and service rate μ. So, θ is here the vector with coordinates λ and μ. If i_k corresponds to an arrival event, we have that $h(x_k, x_{k+1};\theta) = (\lambda\exp\{-\lambda Y_{i_k,m_k}\})/(\lambda_0\exp\{-\lambda_0 Y_{i_k,m_k}\})$, and $h(x_k, x_{k+1};\theta) = (\mu\exp\{-\mu Y_{i_k,m_k}\})/(\mu_0\exp\{-\mu_0 Y_{i_k,m_k}\})$ if i_k corresponds to a departure event. If $A(t)$ and $D(t)$ are, respectively, the number of arrivals and number of departures by time t, then, if we omit in the expression of the likelihood function the end-effect terms, we have that

$$L_t(\theta) \;\approx\; \frac{\lambda^{A(t)}\exp\{-\lambda\sum_{i=1}^{A(t)} A_i\}}{\lambda_0^{A(t)}\exp\{-\lambda_0\sum_{i=1}^{A(t)} A_i\}} \bullet \frac{\mu^{D(t)}\exp\{-\mu\sum_{i=1}^{D(t)} S_i\}}{\mu_0^{D(t)}\exp\{-\mu_0\sum_{i=1}^{D(t)} S_i\}},$$

where A_i and S_i are, respectively, the ith interarrival time and ith service time.

5 THE MAIN RESULTS

In order to carry through the arguments, a number of assumptions must be made. We only state a few here and refer the reader to Damerdji (1992) for the complete conditions.

A vexing requirement is that the state space be finite. Another vexing condition is that events admit a positive density over the whole positive line. This excludes discrete variables and also continuous variables with finite support. Another restrictive condition is that events have exponentially bounded distributions (see Barlow and Proschan (1975) for a definition).

Some more notation is needed. Let $\mathbf{s}_0$, $\mathbf{i}_0$, $X_{\mathbf{i}_0}$, and $\mathbf{s}_1$ be the random variables corresponding to, respectively, the state of the GSMP at time 0, the event which triggers the first transition, the lifetime of that event, and the next state visited. Consider $\sigma(\theta)$ the $r\times r$-matrix with entries

$$\sigma_{uv}(\theta) = E\Big[\frac{\partial}{\partial\theta_u}\log\big(p(\mathbf{s}_1;\mathbf{s}_0,\mathbf{i}_0;\theta)f_{\mathbf{i}_0}(X_{\mathbf{i}_0};\theta)\big) \bullet \frac{\partial}{\partial\theta_v}\log\big(p(\mathbf{s}_1;\mathbf{s}_0,\mathbf{i}_0;\theta)f_{\mathbf{i}_0}(X_{\mathbf{i}_0};\theta)\big)\Big].$$

The expectation is with respect to the stationary distribution, which exists and is unique (Glynn, 1988). To ensure that no two parameters of the vector θ are redundant, the matrix $\sigma(\theta)$ is assumed to be nonsingular, and thus invertible.

Let λ be the limit of $N(t)/t$ as $t\to\infty$. (The constant λ exists and is positive with probability one (see Damerdji, 1992).) The exact conditions under which the following results hold are given in Damerdji (1992).

RESULT 1:

(i) For t sufficiently large, the AMLE $\tilde{\theta}_t$ exists and is consistent with probability one.
(ii) We have that

$$\sqrt{t}(\tilde{\theta}_t \;-\; \theta^0) \;\Longrightarrow\; N\left(0,\lambda^{-1}\sigma(\theta^0)^{-1}\right) \quad \text{as } t\to\infty.$$

This result says that the approximate maximum likelihood estimator exists and converges to the true parameter at rate $1/\sqrt{t}$.

RESULT 2:

(i) For t sufficiently large, the MLE θ_t exists and is consistent with probability one.
(ii) We have that

$$\sqrt{t}(\theta_t \;-\; \theta^0) \;\Longrightarrow\; N\left(0,\lambda^{-1}\sigma(\theta^0)^{-1}\right) \qquad \text{as } t\to\infty.$$

The MLE also exists and converges to θ^0 at rate $1/\sqrt{t}$. The above two results do not imply that the

MLE and AMLE converge to one another at rate $1/\sqrt{t}$, but, under an additional condition, we do obtain the following (see Damerdji, 1992).

RESULT 3:

We have that

$$\sqrt{t}(\tilde{\theta}_t - \theta_t) \implies 0 \qquad \text{as } t \to \infty.$$

We get, then, that the approximate maximum likelihood estimator and the maximum likelihood estimator get close to one another at rate faster than $1/\sqrt{t}$. This implies that there is no loss in statistical efficiency in computing the AMLE instead of the MLE.

6 CONCLUSION

Maximum likelihood estimation for generalized semi-Markov processes was undertaken in this study. Central to our approach is the fact that the GSMP admits an embedded (general state-space) Markov chain. Although the transition density function of this Markov chain has a complicated expression, the likelihood function of the GSMP has a relatively simple expression. An even simpler function was considered. It is in general much easier to compute the AMLE, the maximizer of this other function. We discussed that the MLE also exists, and that both the MLE and AMLE obey a central limit theorem. There is no loss in statistical efficiency in computing the AMLE instead of the MLE.

To allow priority rules, such as preemptive-resume queueing disciplines, the general theory of GSMP's allows the incorporation of rates on the clocks (see Whitt, 1980). We did not consider rates in this study, but we believe our approach could generalize to include such a scheme.

Event cancelling was barred in our study, as the likelihood function would be more complicated. This is a restriction, and work to include event cancelling in the theory is warranted since a number of real-world systems do have such a property.

ACKNOWLEDGMENTS

This work was part of the author's dissertation, at the University of Wisconsin-Madison, under the guidance of Professor Peter W. Glynn. The author is most grateful to him.

REFERENCES

Barlow, R. E. and Proschan, F. 1975. *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston, New York.

Basawa, I. V. and Prakasa Rao, B. L. S. 1980. *Statistical Inference for Stochastic Processes*. Academic Press, London.

Billingsley, P. 1961a. Statistical methods in Markov chains. *Annals of Mathematical Statistics*, **32**, 12–40.

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

Billingsley, P. 1986. *Probability and Measure*, 2nd edn. Wiley, New York.

Cox, D. R. and Hinkley, D. V. 1974. *Theoretical Statistics*. Chapman and Hall, London.

Damerdji, H. 1992. Maximum likelihood estimation for generalized semi-Markov processes. Technical Report #92–11, Department of Industrial Engineering, North Carolina State University, Raleigh, NC.

Glynn, P. W. 1988. A GSMP formalism for discrete-event systems. *Proceedings of the IEEE*, **77**, 14–23.

Glynn, P. W. and Iglehart, D. L. 1989. Importance sampling for stochastic simulations. *Management Science*, **35**, 1367–1392.

König, D., Matthes, K., and Nawrotzki, W. K. 1967. *Verallgemeinerungen der Erlangschen und Engsetschen Formeln*. Akademie-Verlag. Berlin.

Moore, E. H. and Pyke, R. 1968. Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Stat. Math.*, **20**, 411–424.

Whitt, W. 1980. Continuity of generalized semi-Markov processes. *Mathematics of Operations Research*, **5**, 494–501.

AUTHOR BIOGRAPHY

HALIM DAMERDJI is an Assistant Professor in the Department of Industrial Engineering at North Carolina State University. His research interests are in simulation and applied stochastic processes. He is a member of ORSA.