Likelihood Ratio Derivative Estimation for Finite-Time Performance Measures in Generalized Semi-Markov Processes

Marvin K. Nakayama

Department of Computer and Information Science

New Jersey Institute of Technology

Newark, NJ 07102

Perwez Shahabuddin

Department of Industrial Engineering and Operations Research

Columbia University

New York, NY 10027

August 31, 1998

Abstract

This paper investigates the likelihood ratio method for estimating derivatives of finite-time performance measures in generalized semi-Markov processes (GSMPs). We develop readily verifiable conditions for the applicability of this method. Our conditions mainly place restrictions on the *basic building blocks* (i.e., the transition probabilities, the distribution and density functions of the event lifetimes, and the initial distribution) of the GSMP, which is in contrast to the structural conditions needed for infinitesimal perturbation analysis. We explicitly show

that our conditions hold in many practical settings, and in particular, for large classes of queueing and reliability models. One intermediate result which we obtain in this study, which is of independent value, is to formally show that the random variable representing the number of occurring events in a GSMP in a finite time horizon, has finite exponential moments in a neighborhood of zero.

1   Introduction

When running a simulation, one is often interested in estimating derivatives of a performance measure with respect to parameters of the input distributions (e.g., routing probabilities, interarrival time distributions and service time distributions in queueing networks; lifetime distributions and repair time distributions of components in reliability models). One reason for computing derivative estimates is that they are useful for sensitivity analysis. For example, in the reliability context, a designer of a system may want to determine where to focus design efforts to improve overall system performance. In addition, efficient derivative estimation plays an important role in simulation-based optimization algorithms (e.g., see Glynn 1986, 1989a).

One approach for estimating derivatives via simulation is the likelihood ratio method (also called the score function method); e.g., see Aleksandrov, Sysoyev, and Shemeneva (1968), Arsham et al. (1989), Glynn (1986,1989a,1990), Reiman and Weiss (1986,1989), Rubinstein (1986,1989), L'Ecuyer (1990,1995), Nakayama, Goyal, and Glynn (1994), and Nakayama (1995). Overviews of this technique may be found in Bratley, Fox and Schrage (1987), Glynn (1987), L'Ecuyer (1991) and Rubinstein and Shapiro (1993). The validity of this technique has been established explicitly for certain performance measures of classes of discrete-space, discrete-time Markov chains (Glynn 1986), discrete-space, continuous-time Markov chains (Reiman and Weiss 1989, and Nakayama, Goyal, and Glynn 1994), and general-state-space Markov chains that have a regenerative structure, also called Harris chains (Glasserman and Glynn 1992). The performance measures that have been studied have mainly been tran-

2

sient ones: performance measures that are based on the behavior of the stochastic process in some time interval from the beginning, the length of which is either fixed or random (a stopping-time). Steady-state performance measures have been studied insofar as they can be expressed as functions of stopping-time based transient performance measures using some regenerative structure. The main impediment in verifying the validity lies in justifying the interchange of derivative and expectation that is required in this method. L'Ecuyer (1990,1995), Reiman and Weiss (1989) and Rubinstein and Shapiro (1993) have general conditions that ensure when this interchange is justified. However, these conditions are difficult to verify in many specific contexts. In this paper, we establish conditions that are readily verifiable in many contexts for the validity of the likelihood ratio method in the setting of generalized semi-Markov processes (GSMPs).

A GSMP is a general mathematical framework for modeling many discrete-event systems, and there has been a considerable amount of work studying different aspects of this class of stochastic processes; e.g., see Shassberger (1976), Whitt (1980), Glynn (1989b), Haas and Shedler (1987), and Glasserman and Yao (1992a, 1992b). Also, because their dynamics closely follow those of event-driven simulations, GSMPs have been useful in modeling systems analyzed through simulation and for studying various simulation methodologies; e.g., see Glynn and Iglehart (1988), Glasserman (1991a), and Glasserman and Yao (1992c). In this paper we use the GSMP framework to study the likelihood ratio derivative estimation method applied to finite-time performance measures, i.e., performance measures that are expectations of random variables that can be determined by a fixed time t. The generality of the GSMP framework enables us to verify the applicability of the likelihood ratio derivative estimation method for large classes of reliability and queueing models. In fact, this whole project was initially motivated by the need to do simulation-based derivative estimation in a large class of reliability models of the type in Heildelberger, Shahabuddin, Nicola (1994), Nicola, Nakayama, Heildelbeger and Goyal (1993) and Nicola, Shahabuddin, Heidelberger, Glynn (1993). These are non-Markovian versions of models constructed and analyzed by the SAVE software package

(Blum et al. 1993).

An alternative method for estimating derivatives using simulation is infinitesimal perturbation analysis (IPA); e.g., see Ho and Cao (1983), Heidelberger et al. (1988), and Glasserman (1991a). Glasserman (1991a,1991b,1991c) established conditions under which this method will give rise to unbiased estimates of derivatives of finite-time performance measures of GSMPs. One main difference between the conditions for IPA and our conditions for the likelihood ratio method is that ours impose restrictions on the *basic building blocks* (i.e., the transition probabilities, the distribution and density functions of the event lifetimes, and the initial distribution) of the GSMP, whereas those for IPA relate to the underlying structure of the GSMP. We give examples of large classes of stochastic models for which one can justify the validity of the likelihood ratio method but not IPA, due to the limitations posed by the structural conditions. On the other hand, the IPA method admits a larger class of event-time distributions than the likelihood ratio method. Also, the variance of likelihood ratio method derivative estimators grows quite fast as the length of the observation increases; e.g., see Glynn (1987) for results in the setting of discrete-time Markov chains. Thus, for large time horizons t, the likelihood ratio method is statistically inefficient for estimating derivatives. IPA derivative estimators do not suffer from this drawback.

The rest of the paper is organized as follows. Section 2 contains a description of the GSMP model with which we will work. Section 3 reviews the basic idea of the likelihood ratio method for estimating derivatives. The section also provides a proof of the validity of the method in the GSMP setting under a set of technical conditions. One intermediate result that is derived here, shows that the random variable representing the number of occurring events in the GSMP in $[0, t]$, has finite exponential moments around the neighborhood of zero. This is done by bounding the GSMP by an age-dependent branching process, and using results on finite exponential moments of such branching processes. Section 4 gives some simple sufficient conditions for our conditions in Section 3 to hold. In particular, we show that the conditions hold in many settings arising in practice. Finally, in Section 5 we explicitly demonstrate that

the conditions in Section 3 hold for certain classes of reliability models and queueing network models.

2 Mathematical Model

We model the evolution of our stochastic system as a generalized semi-Markov process (GSMP), which we now describe. Our development of the model closely follows Glasserman's (1991a) approach. We denote the (physical) state space of the process by $\mathbf{S}$, where we assume that $\mathbf{S}$ is either finite or countably infinite. For example, in the context of multi-class queueing networks, a state in $\mathbf{S}$ contains information about the number of customers of each type at each station, along with any description of the queueing at the various stations. We let $\mathbf{A}$ denote the set of events. In the example of multi-class queueing networks, the events are the completion of service and external customer arrival at each station. We make the following assumption:

A1 *The set $\mathbf{A}$ is finite.*

A case where this assumption is not satisfied is the infinite-server queue. For a state $s \in \mathbf{S}$, we define $\mathcal{E}(s)$ as the set of active events in state s; i.e., these are the events that may occur when the system is in state s. In our queueing-system example, a service-completion event at a particular queue is only possible in states in which the server at that queue is busy.

We introduce a scalar parameter θ with respect to which derivatives are computed. For example, θ might represent the mean arrival time of customers or a routing probability in a queueing network. After computing the derivative, we then want to evaluate the derivative at the parameter value $\theta = \theta_0$, where $\theta_0 \in \Theta$ and Θ is some open set. We can easily extend our setting of calculating a derivative with respect to a scalar-valued parameter θ to work with computing the gradient with respect to a vector-valued parameter $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)})$ by considering each $\theta^{(i)}$ separately.

For each event $e \in \mathbf{A}$, we let $F(\,\cdot\,; e, \theta)$ denote the distribution used to generate event times for e under parameter value θ. We assume that each distribution function $F(\,\cdot\,; e, \theta)$ has a density $f(\,\cdot\,; e, \theta)$ (with respect to Lebesgue measure) and that $F(0; e, \theta) = 0$. An initial state is chosen using the probability mass function $\mu(\,\cdot\,; \theta)$, and given that a state s is selected as the initial state, we generate event times for each event $e \in \mathcal{E}(s)$ from $F(\,\cdot\,; e, \theta)$. For each $\theta \in \Theta$, $s \in \mathbf{S}$, and $e \in \mathbf{A}$, let $p(\,\cdot\,; s, e, \theta)$ be a probability mass function on $\mathbf{S}$. Under parameter value θ, if the system is currently in a state $s \in \mathbf{S}$ and some event $e \in \mathcal{E}(s)$ occurs, then the system immediately moves to a new state $s' \in \mathbf{S}$ with probability $p(s'; s, e, \theta)$. Thus, the $p(s'; s, e, \theta)$, $(s, s') \in \mathbf{S} \times \mathbf{S}$, $e \in \mathbf{A}$, are called the *transition probabilities*. (The general theory of GSMPs allows for more generality; e.g., see Glynn 1989b.)

For any state $s \in \mathbf{S}$, we define the set of possible clock readings as

$$\mathbf{C}(s) = \{(c(e), e \in \mathbf{A}) : c(e) > 0 \text{ if } e \in \mathcal{E}(s), \ c(e) = +\infty \text{ if } e \notin \mathcal{E}(s)\}.$$

We identify $(c(e), e \in \mathbf{A})$ with an $|\mathbf{A}|$-dimensional vector that we write as c. When the system is in state s, the clock $c(e)$ for any active event $e \in \mathcal{E}(s)$ is run down at rate $0 \leq \lambda(s; e) < \infty$. We assume that $\lambda(s; e) > 0$ for some $e \in \mathcal{E}(s)$. In most applications $\lambda(s; e) = 1$. However, letting $\lambda(s; e)$ assume values other than 1 allows greater modeling capabilities. For example, we can model pre-emptive service by setting $\lambda(s; e) = 0$ if e is a service-completion event that is pre-empted in state s. We assume that the clock rates are bounded; i.e.,

A2 *There exists some constant $\lambda_* < \infty$ such that the clock rates $\lambda(s; e) \leq \lambda_*$ for all $s \in \mathbf{S}$ and all $e \in \mathbf{A}$.*

In most applications, A2 holds, but one example where it does not is an infinite-capacity open queueing system in which the rate at which the server processes customers is proportional to the queue length.

We assume that in each state $s \in \mathbf{S}$, there is at least one active event $e \in \mathcal{E}(s)$. Given that the system is currently in state s with clock vector c, we define

$$t^*(s, c) = \min\{c(e)/\lambda(s; e) : e \in \mathcal{E}(s), \ \lambda(s; e) > 0\}$$

as the holding time in the current state. Also, we define

$$e^*(s, c) = \operatorname{argmin}\{c(e)/\lambda(s; e) : e \in \mathcal{E}(s), \ \lambda(s; e) > 0\},$$

which is the event triggering the transition out of the current state. (Since all of the clock times have continuous distributions, two clocks are never equal with probability one for all parameter values θ.)

For states $s, s' \in \mathbf{S}$ and an event $e \in \mathbf{A}$, let $\mathcal{C}(s'; s, e)$ denote the set of events that are active in state s but are cancelled upon the transition to state s' triggered by the event e. Then we define the set of new events $\mathcal{N}(s'; s, e)$ and the set of old events $\mathcal{O}(s'; s, e)$ as $\mathcal{N}(s'; s, e) = \mathcal{E}(s') - (\mathcal{E}(s) - \mathcal{C}(s'; s, e) - e)$ and $\mathcal{O}(s'; s, e) = \mathcal{E}(s) - e - \mathcal{C}(s'; s, e)$, respectively. After a transition from state s to state s' triggered by event e, for each (new) event $e' \in \mathcal{N}(s'; s, e)$, a new clock value $c(e')$ is generated from the distribution $F(\ \cdot\ ; e', \theta)$. Also, for each (old) event $e' \in \mathcal{O}(s'; s, e)$, we update its new clock value to be $c(e') - \lambda(s; e')t^*(s, c)$ immediately after entering state s'. Each (cancelled) event $e' \in \mathcal{C}(s'; s, e)$ was originally active in state s but may no longer be active in state s'. Note that we allow for the possibility that a cancelled event e' is immediately reactivated (i.e., $e' \in \mathcal{C}(s'; s, e)$ and $e' \in \mathcal{N}(s'; s, e)$), in which case a new clock value is immediately generated for the event e'.

We define the discrete-time Markov process $\{(Y_n, c_n) : n \geq 0\}$, where Y_n is the n-th state visited and c_n is the vector of clock readings just after the n-th transition. The state space of the process is $\Sigma = \bigcup_{s \in \mathbf{S}}(\{s\} \times \mathbf{C}(s))$, and its transition probability is as follows: for $A \subset \Sigma$ of the form $A = \{s'\} \times \{c' \in \mathbf{C}(s') : c'(e) \leq x(e), \ e \in \mathcal{E}(s')\}$ for some fixed (possibly infinite) real numbers $(x(e), \ e \in \mathbf{A})$, the transition probability from (s, c) to A under parameter value θ is

$$\pi((s, c), A, \theta) = p(s'; s, e^*, \theta) \prod_{e \in \mathcal{N}(s'; s, e^*)} F(x(e); e, \theta) \prod_{e' \in \mathcal{O}(s'; s, e^*)} 1\left\{\frac{c(e')}{\lambda(s; e')} - t^* \leq x(e')\right\},$$

where $t^* = t^*(s, c)$, $e^* = e^*(s, c)$, and $1\{B\}$ is the indicator function of the probability event $\{B\}$.

We define our continuous-time GSMP as a piecewise-constant process that stays in state Y_n for an interval of length $t^*(Y_n, c_n)$. More precisely, let $\tau_0 = 0$ and define $\tau_n = \sum_{k=0}^{n-1} t^*(Y_k, c_k)$,

7

which is the epoch of the n-th transition. For notational simplicity, define $e_n = e^*(Y_{n-1}, c_{n-1})$, $n \geq 1$. Also, define $N(t) = \sup\{n \geq 0 : \tau_n \leq t\}$, which is the number of transitions that occur by time $t \geq 0$. Then, we define our GSMP $Z = \{Z_t : t \geq 0\}$, where $Z_t = Y_{N(t)}$. Let P_θ denote the probability measure governing our GSMP under parameter value θ, and let E_θ be the expectation operator induced by P_θ. Let $(\Omega, \mathcal{F}, P_\theta)$ denote the probability space of the GSMP, and let $\mathcal{F}_t$ denote the filtration of the GSMP up to a (deterministic) time $t \geq 0$ (or loosely speaking, the history of the process up to time t); e.g., see Section 36 of Billingsley (1986).

Let $\alpha(t, \theta) = E_\theta[V(t, \theta)]$ be our performance measure, where $V(t, \theta)$ is some $\mathcal{F}_t$-measurable random variable. We will be interested in estimating the derivative $\frac{d}{d\theta}\alpha(t, \theta)$ at the parameter value $\theta = \theta_0$. Let P_{θ_0} be the probability measure of the GSMP when we fix $\theta = \theta_0$ and E_{θ_0} is the expectation operator induced by P_{θ_0}.

We now impose some restrictions on the random variable $V(t, \theta)$:

A3 *There exists some constant $h > 0$ such that $V'(t, \theta, \omega) \equiv \frac{d}{d\theta}V(t, \theta, \omega)$ exists for P_{θ_0}-almost all $\omega \in \Omega$ and all $\theta \in (\theta_0 - h, \theta_0 + h)$, and there exist random variables W and W' such that $E_{\theta_0}[W^2] < \infty$ and $E_{\theta_0}[W'^2] < \infty$, and $|V(t, \theta)| \leq W$ and $|V'(t, \theta)| \leq W'$ with P_{θ_0}-probability 1 for all $\theta \in (\theta_0 - h, \theta_0 + h)$.*

We show in Section 4 that many performance measures of interest satisfy the above condition.

3 Likelihood Ratio Derivative Method

In this section we review the likelihood ratio method for estimating $\frac{d}{d\theta}\alpha(t, \theta) = \frac{d}{d\theta}E_\theta[V(t, \theta)]$, which is to be evaluated at the parameter value $\theta = \theta_0$. To compute $\frac{d}{d\theta}E_\theta[V(t, \theta)]$, we would like to bring the derivative operator inside the expectation. However, the expectation operator depends on θ, and so some care must be taken. Assuming that P_θ is absolutely continuous with respect to P_{θ_0} restricted to the filtration $\mathcal{F}_t$, we have that

$$\alpha(t, \theta) = \int V(t, \theta) \, dP_\theta = \int V(t, \theta) \, \frac{dP_\theta}{dP_{\theta_0}} \, dP_{\theta_0} = E_{\theta_0}[V(t, \theta)L(t, \theta)], \tag{1}$$

where $L(t, \theta) = dP_\theta/dP_{\theta_0}$ is the Radon-Nykodym derivative of P_θ with respect to P_{θ_0} restricted to the filtration $\mathcal{F}_t$, or simply the likelihood ratio. (A probability measure Q_1 is absolutely continuous with respect to another probability measure Q_2 if the sets of Q_2-measure 0 are also of Q_1-measure 0; e.g., see Section 32 of Billingsley 1986.) This is known as a "change of measure," and Glynn (1989) and Damerdji (1994) established its validity for GSMPs in certain contexts. Thus we have expressed the performance measure $\alpha(t, \theta)$ as an expectation of some random quantity taken with respect to a probability measure that is independent of the parameter θ. The key question that remains now is whether we can bring the derivative operator inside the expectation; i.e., whether

$$\frac{d}{d\theta} E_{\theta_0}[V(t, \theta)L(t, \theta)] = E_{\theta_0}\left[\frac{d}{d\theta}(V(t, \theta)L(t, \theta))\right]. \tag{2}$$

Assumption A1 of L'Ecuyer (1990) and Assumption A1$'(k)$ of L'Ecuyer (1995) (see also Rubinstein and Shapiro 1993) provide conditions when (2) holds on a general probability space. However, in the specific context of the GSMP, these are difficult to verify directly. Before answering the interchange question in the GSMP setting, let us first determine the exact form of the likelihood ratio $L(t, \theta)$. Heuristically, the likelihood ratio $L(t, \theta)$ is given by

$$L(t, \theta) = \frac{\mu(Y_0; \theta)}{\mu(Y_0; \theta_0)} \prod_{e \in \mathcal{E}(Y_0)} \frac{F(dc_0(e); e, \theta)}{F(dc_0(e); e, \theta_0)} \prod_{k=1}^{N(t)} \left(\frac{p(Y_k; Y_{k-1}, e_k, \theta)}{p(Y_k; Y_{k-1}, e_k, \theta_0)} \prod_{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k)} \frac{F(dc_k(e); e, \theta)}{F(dc_k(e); e, \theta_0)} \right); \tag{3}$$

see Glynn and Iglehart (1989) for further details. However, if the above version of the likelihood ratio is used, we would need very restrictive conditions on the distribution functions $F(\,\cdot\,; e, \theta)$ to justify the interchange of derivative and expectation. (For example, we would have to change A5 below to alternatively require $f(\,\cdot\,; e, \theta)/f(\,\cdot\,; e, \theta_0) \to 1$ uniformly on $[0, \infty)$ as $\theta \to \theta_0$, which is not even satisfied when $F(\,\cdot\,; e, \theta)$ is the exponential distribution with mean $1/\theta$.) To get around this, we instead work with a slightly modified GSMP, which leads to a different likelihood ratio but does not alter the distribution of $V(\theta, t)$. We do this by changing the distributions from which clock times are generated: for all $e \in \mathbf{A}$, define the new

distribution function

$$
F_*(x; e, \theta) = \begin{cases} F(x; e, \theta) & \text{if } x \leq \lambda_* t \\ F(\lambda_* t; e, \theta) & \text{if } \lambda_* t < x < \lambda_*(t+1) \\ 1 & \text{if } x \geq \lambda_*(t+1) \end{cases}
$$

where we recall that t is the time horizon. Note that $F_*(x; e, \theta)$ is a mixed (continuous and discrete) distribution that is the same as $F(x; e, \theta)$ for $x \leq \lambda_* t$, and has a point mass of probability $\bar{F}(\lambda_* t; e, \theta)$ at the point $\lambda_*(t+1)$, where $\bar{F}(x; e, \theta) \equiv 1 - F(x; e, \theta)$. Unlike before, now more than one event can occur at the same time. We can use any rule we want to deal with this. For example, we can order the set of events in some manner, and if more than one event occurs at the same time, we take the "lowest" of these events as the "occurring" event (for determining the next physical state visited, etc.). In any case, both the original and the modified GSMP (i.e., the process Z defined in the previous section) have the same distribution of sample paths in $[0, t]$. Since $V(\theta, t)$ is only dependent on the sample path of the GSMP from time 0 to t (i.e., it is $\mathcal{F}_t$-measurable), using the new distributions for generating clock times does not change the distribution of $V(\theta, t)$. By applying this approach, we will only require mild conditions on the distribution functions $F(\,\cdot\,; e, \theta)$ to prove the interchange.

The new likelihood ratio is given by

$$
\begin{aligned}
L(t, \theta) \;=\; & \frac{\mu(Y_0; \theta)}{\mu(Y_0; \theta_0)} \prod_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* \leq t}} \frac{f(c_0(e); e, \theta)}{f(c_0(e); e, \theta_0)} \prod_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* > t}} \frac{\bar{F}(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta_0)} \\[2mm]
& \cdot \prod_{k=1}^{N(t)} \left(\frac{p(Y_k; Y_{k-1}, e_k, \theta)}{p(Y_k; Y_{k-1}, e_k, \theta_0)} \prod_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* \leq t}} \frac{f(c_k(e); e, \theta)}{f(c_k(e); e, \theta_0)} \prod_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* > t}} \frac{\bar{F}(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta_0)} \right).
\end{aligned}
$$

Now assuming that (2) holds, we see that applying the product rule of differentiation yields

$$
\frac{d}{d\theta} \alpha(t, \theta) = E_{\theta_0} \left[V'(t, \theta) L(t, \theta) + V(t, \theta) L'(t, \theta) \right], \tag{4}
$$

where

$$
L'(t, \theta) = L(t, \theta) \left[\frac{\mu'(Y_0; \theta)}{\mu(Y_0; \theta)} + \sum_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* \leq t}} \frac{f'(c_0(e); e, \theta)}{f(c_0(e); e, \theta)} + \sum_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* > t}} \frac{\bar{F}'(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta)} \right. \tag{5}
$$

$$
\left. + \sum_{k=1}^{N(t)} \left(\frac{p'(Y_k; Y_{k-1}, e_k, \theta)}{p(Y_k; Y_{k-1}, e_k, \theta)} + \sum_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* \leq t}} \frac{f'(c_k(e); e, \theta)}{f(c_k(e); e, \theta)} + \sum_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* > t}} \frac{\bar{F}'(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta)} \right) \right]
$$

with $\mu'(\,\cdot\,; \theta) = \frac{d}{d\theta}\mu(\,\cdot\,; \theta)$, $f'(\,\cdot\,; \cdot, \theta) = \frac{d}{d\theta}f(\,\cdot\,; \cdot, \theta)$, $\bar{F}'(\,\cdot\,; \cdot, \theta) = \frac{d}{d\theta}\bar{F}(\,\cdot\,; \cdot, \theta)$, and $p'(\,\cdot\,; \cdot, \cdot, \theta) = \frac{d}{d\theta}p(\,\cdot\,; \cdot, \cdot, \theta)$. Glynn (1990) gives the derivative of a likelihood ratio expression similar to (3), but for the case of a fixed number of transitions (in contrast to a fixed time horizon). Also, Glynn did not establish the validity of the interchange of the derivative and expectation in the likelihood ratio derivative method applied to GSMPs.

Now we develop conditions on the basic building blocks of the GSMP under which (2) is valid. (We later give in Section 4 some easily verifiable sufficient conditions for the following conditions to hold.) First, we impose some regularity conditions on the densities generating the event lifetimes.

A4 *For each event $e \in \mathbf{A}$, there exists $h > 0$ such that the set $\Lambda_f(e) \equiv \{x : 0 \leq x \leq \lambda_* t,\ f(x; e, \theta) > 0\}$ is independent of θ for $\theta \in (\theta_0 - h, \theta_0 + h)$.*

A5 *For each event $e \in \mathbf{A}$ such that $\Lambda_f(e) \neq \emptyset$, for each $\epsilon > 0$ there exists $h > 0$ such that $1 - \epsilon \leq f(x; e, \theta)/f(x; e, \theta_0) \leq 1 + \epsilon$ for all $x \in \Lambda_f(e)$ and all $\theta \in (\theta_0 - h, \theta_0 + h)$, i.e., $f(\,\cdot\,; e, \theta)/f(\,\cdot\,; e, \theta_0) \to 1$ uniformly on $\Lambda_f(e)$ as $\theta \to \theta_0$.*

A6 *For each event $e \in \mathbf{A}$ such that $\Lambda_f(e) \neq \emptyset$, there exists $h > 0$ such that $f(\,\cdot\,; e, \theta)$ is differentiable in θ on $\Lambda_f(e)$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$ and $f'(\,\cdot\,; e, \cdot)/f(\,\cdot\,; e, \cdot)$ is uniformly bounded on $\Lambda_f(e) \times (\theta_0 - h, \theta_0 + h)$; i.e., $\sup_{x \in \Lambda_f(e),\ \theta \in (\theta_0 - h, \theta_0 + h)} |f'(x; e, \theta)/f(x; e, \theta)| < \infty$.*

Note that Condition A4 does not permit the support of the distribution of the lifetime of an event e to depend on θ. Thus, for example, we disallow an event e to have a uniform distribution with support $(0, \theta)$.

We will now present some properties of $\bar{F}(\,\cdot\,; e, \,\cdot\,)$ that are implied by the previous conditions on $f(\,\cdot\,; e, \,\cdot\,)$. The proof of this lemma may be found in the Appendix. Let $\Lambda_{\bar{F}}(e) = \{x : 0 \leq x \leq \lambda_* t, \ \bar{F}(x; e, \theta) > 0\}$.

Lemma 1 *Assume that Condition A4–A6 hold. Then*

(i) there exists $h > 0$, such that $\Lambda_{\bar{F}}(e)$ is independent of θ for $\theta \in (\theta_0 - h, \theta_0 + h)$;

(ii) for each event $e \in \mathbf{A}$,

$$\bar{F}(\,\cdot\,; e, \theta) \to \bar{F}(\,\cdot\,; e, \theta_0) \tag{6}$$

uniformly on $[0, \lambda_* t]$ as $\theta \to \theta_0$; i.e., $\sup_{x \in [0, \lambda_* t]} |\bar{F}(x; e, \theta) - \bar{F}(x; e, \theta_0)| \to 0$ as $\theta \to \theta_0$;

(iii) for each event $e \in \mathbf{A}$ such that $\bar{F}(\lambda_* t; e, \theta_0) > 0$, for each $\epsilon > 0$ there exists $h > 0$ such that $1 - \epsilon \leq \bar{F}(\lambda_* t; e, \theta)/\bar{F}(\lambda_* t; e, \theta_0) \leq 1 + \epsilon$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$, i.e., $\bar{F}(\lambda_* t; e, \theta)/\bar{F}(\lambda_* t; e, \theta_0) \to 1$ as $\theta \to \theta_0$;

(iv) for each event $e \in \mathbf{A}$ such that $\bar{F}(\lambda_* t; e, \theta_0) > 0$, there exists $h > 0$ such that $\bar{F}(\lambda_* t; e, \theta)$ is differentiable in θ for all $\theta \in (\theta_0 - h, \theta_0 + h)$ and $\bar{F}'(\lambda_* t; e, \,\cdot\,,)/\bar{F}(\lambda_* t; e, \,\cdot\,,)$ is bounded on $(\theta_0 - h, \theta_0 + h)$; i.e., $\sup_{\theta \in (\theta_0 - h, \theta_0 + h)} |\bar{F}'(\lambda_* t; e, \theta)/\bar{F}(\lambda_* t; e, \theta)| < \infty$.

We now impose some regularity conditions on the initial distribution.

A7 *There exists $h > 0$ such that the set $\Lambda_\mu \equiv \{s : s \in \mathbf{S}, \ \mu(s; \theta) > 0\}$ is independent of θ for $\theta \in (\theta_0 - h, \theta_0 + h)$.*

A8 *For each $\epsilon > 0$ there exists $h > 0$ such that $1 - \epsilon \leq \mu(s; \theta)/\mu(s; \theta_0) \leq 1 + \epsilon$ for all $s \in \Lambda_\mu$ and $\theta \in (\theta_0 - h, \theta_0 + h)$, i.e., $\mu(\,\cdot\,; \theta)/\mu(\,\cdot\,; \theta_0) \to 1$ uniformly on Λ_μ as $\theta \to \theta_0$.*

A9 *There exists $h > 0$ such that $\mu(\,\cdot\,; \theta)$ is differentiable in θ on Λ_μ for all $\theta \in (\theta_0 - h, \theta_0 + h)$ and $\mu'(\,\cdot\,; \,\cdot\,)/\mu(\,\cdot\,; \,\cdot\,)$ is uniformly bounded on $\Lambda_\mu \times (\theta_0 - h, \theta_0 + h)$; i.e., $\sup_{s \in \Lambda_\mu, \ \theta \in (\theta_0 - h, \theta_0 + h)} |\mu'(s; \theta)/\mu(s; \theta)| < \infty$.*

Finally, we impose some regularity conditions on the transition probabilities.

A10 *There exists $h > 0$ such that the set $\Lambda_p \equiv \{(s', s, e) : s' \in \mathbf{S}, \ s \in \mathbf{S}, \ e \in \mathbf{A}, \ p(s'; s, e, \theta) > 0\}$ is independent of θ for $\theta \in (\theta_0 - h, \theta_0 + h)$.*

A11 *For each $\epsilon > 0$ there exists $h > 0$ such that $1 - \epsilon < p(s'; s, e, \theta)/p(s'; s, e, \theta_0) < 1 + \epsilon$ for all $(s', s, e) \in \Lambda_p$ and $\theta \in (\theta_0 - h, \theta_0 + h)$, i.e., $p(\ \cdot\ ; \cdot, \cdot, \theta)/p(\ \cdot\ ; \cdot, \cdot, \theta_0) \to 1$ uniformly on Λ_p as $\theta \to \theta_0$.*

A12 *There exists $h > 0$ such that $p(\ \cdot\ ; \cdot, \cdot, \theta)$ is differentiable in θ on Λ_p for all $\theta \in (\theta_0 - h, \theta_0 + h)$ and $p'(\ \cdot\ ; \cdot, \cdot, \cdot)/p(\ \cdot\ ; \cdot, \cdot, \cdot)$ is uniformly bounded on $\Lambda_p \times (\theta_0 - h, \theta_0 + h)$; i.e., $\sup_{(s', s, e) \in \Lambda_p, \ \theta \in (\theta_0 - h, \theta_0 + h)} |p'(s'; s, e, \theta)/p(s'; s, e, \theta)| < \infty$.*

In some sense, Conditions A4–A12 ensure that the basic building blocks of our GSMP are well-behaved functions of θ in a neighborhood of θ_0. Now we establish the validity of (2).

Theorem 1 *If Conditions A1–A12 hold, then $E_{\theta_0}[V(t, \theta)L(t, \theta)]$ is differentiable and (2) is valid.*

To show that Theorem 1 holds, we will first need to prove the following result which is of independent value.

Lemma 2 *Suppose Conditions A1, A2, A4 and A5 hold. Then there exists constants $z_0 > 0$ and $h > 0$ such that $E_\theta[e^{zN(t)}] < \infty$ for all $z \in [-z_0, z_0]$ and $\theta \in (\theta_0 - h, \theta_0 + h)$.*

Proof. Observe that $N(t)$ has the same distribution whether we use $F_*(\ \cdot\ ; e, \theta)$ or $F(\ \cdot\ ; e, \theta)$ to generate lifetimes for the event $e \in \mathbf{A}$. Hence, we work with $F(\ \cdot\ ; e, \theta)$.

For the case of systems with no event cancellation, we can show the result using the methods employed in the proofs of Corollary 3.3 of Glasserman (1991a) and Theorem 2.1 on page 155 of Prabhu (1965), that use results from renewal theory. For systems with event cancellation, we need a new approach, which is given below and uses results from the theory of branching processes.

13

We will establish the result by first constructing a new process that bounds $N(t)$. Then we will show that this new process has the same distribution as a multi-type, age-dependent branching process (see, e.g., Athreya and Ney (1972) for a definition). We then bound this multi-type age-dependent branching process by a single-type, age-dependent branching process, and finally use a result of Nakayama, Shahabuddin, and Sigman (1996). For simplicity, we will assume that all of the clock rates $\lambda(s; e) = 1$. The proof can be easily generalized when the clock rates are unequal and bounded, as in Condition A2. For each $e \in \mathbf{A}$, recall that lifetimes of the event e follow a distribution $F(\,\cdot\,; e, \theta)$. Some of the lifetimes may not expire in the GSMP because of event cancellation.

In what follows, we frequently will be using the following multi-type, age-dependent branching process. To simplify notation, let $m = |\mathbf{A}|$. By Condition A1, $m < \infty$. The branching process has m different types of objects, indexed by e. Objects of type e have lifetime distributions $F(\,\cdot\,; e, \theta)$. The branching process begins at time 0, at which time one object of each type is born, i.e., a total of m objects are born. Each object lives for a random time (specified by its corresponding lifetime distribution) and then dies. When any object dies, it instantaneously gives birth to exactly one object of each type. The process thus continues.

We now construct a sample path of our first bounding process as follows. We begin by generating a sample path of the original GSMP in $[0, t]$. Then we do the following:

- In the sample-path generation of the original GSMP, at each transition point (equivalent to an event expiration point of a non-cancelled event, but also includes $t = 0$) a subset of events (which may also be a null-set) in $\mathbf{A}$ are scheduled. In the new process, at each transition point of the GSMP we schedule each of the other events in $\mathbf{A}$ that have not already been scheduled at that transition point. These new events are never cancelled and whenever any of them expire, we start a multi-type branching process of the sort mentioned before.

- In the original GSMP, events are cancelled. In the new process, we let each of the

cancelled events run to expiration. And whenever any of them expires, we again start a multi-type branching process of the kind mentioned before.

Let $B(t)$ be the total number of events scheduled in $[0, t]$ in this constructed process. Clearly $B(t) \geq N(t)$ and so $E_\theta[e^{zB(t)}] \geq E_\theta[e^{zN(t)}]$ for all $z \geq 0$. We now claim that this constructed process is a sample path of a multi-type, age-dependent branching process.

On first glance this claim may seem incorrect. This is because the event expiration times of cancellable events (if not cancelled) depend on the event expiration times of events that cause the cancellation (by their expiration). Then the independent lifetimes assumption of a branching process seems to be violated. However, a careful argument reveals otherwise. Note that the only difference between the constructed process and an ordinarily generated, multi-type, age-dependent branching process is the order of generation of the lifetimes. In the ordinary case, the lifetimes may have been generated sequentially, i.e., we first generate the lifetimes of each of the m objects in the first generation, then of each of the m^2 objects in the second generation, and so on. In the constructed process, the order in which we generate lifetimes is random. Not only is it random, but the next lifetime one chooses to generate depends on the position and length of the lifetimes already generated. However, the lifetimes themselves are independent of anything in the system and hence the order of generation will not change the distribution of the branching process.

Now consider a new branching process defined as follows. The branching process begins at time 0, at which time m objects are born. When any object dies, it gives birth to exactly m objects. The lifetime of each object follows a distribution H, which depends only on θ_0. We define it as follows. First, define a distribution function G with $G(x; \theta) = 1 - \prod_{e \in \mathbf{A}} \bar{F}(x; e, \theta)$ for $x \geq 0$, and 0 otherwise. Note that if X_e has distribution $F(\cdot; e, \theta)$, $e \in \mathbf{A}$, then $X = \min_{e \in \mathbf{A}} X_e$ has distribution $G(\cdot; \theta)$. Now using Condition A4, A5, and part(ii) of Lemma 1 we can easily show that $G(x; \theta) \to G(x; \theta_0)$ uniformly on the set $[0, t]$, as $\theta \to \theta_0$. Hence, we can choose a sufficiently small $h > 0$ such that $|G(x; \theta) - G(x; \theta_0)| < 1/(2m)$ for all $x \leq t$ and $\theta \in (\theta_0 - h, \theta_0 + h)$. Let $\epsilon = \sup\{|G(x; \theta) - G(x; \theta_0)| : x \leq t, \theta \in (\theta_0 - h, \theta_0 + h)\}$, and

note that $\epsilon \leq 1/(2m)$. Now define the new distribution function H as follows: $H(x; \theta_0) = 0$ if $x < 0$; $H(x; \theta_0) = G(x; \theta_0) + \epsilon$ if $G(x; \theta_0) + \epsilon < 1$ and $x \geq 0$; and $H(x; \theta_0) = 1$ otherwise. Observe that the distribution $H(\cdot; \theta_0)$ is stochastically smaller (see Ross 1983) than $G(\cdot; \theta)$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$, which implies that $H(\cdot; \theta_0)$ is stochastically smaller than $F(\cdot; e, \theta)$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$ and $e \in \mathbf{A}$. Also, $H(0; \theta_0)m < 1$ since $G(0; \theta) = 0$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$ because we assumed that $F(0; e, \theta) = 0$ for all $e \in \mathbf{A}$.

Note that the above defines a single-type, age-dependent branching process. Let $B'(t)$ be the number of objects born up to and including time t, Also, since $H(\cdot; \theta_0)$ is stochastically smaller than $G(\cdot; \theta)$ for all $\theta \in (\theta_0 - h, \theta_0 + h)$, we have that $B'(t)$ is stochastically larger than $B(t)$ for all $t \geq 0$ and $\theta \in (\theta_0 - h, \theta_0 + h)$. Hence, $E[e^{zB'(t)}] \geq E_\theta[e^{zB(t)}]$ for all $z \geq 0$ and $\theta \in (\theta_0 - h, \theta_0 + h)$. Finally, since $H(0; \theta_0)m < 1$, we can use Theorem 3.1 of Nakayama, Shahabuddin, and Sigman (1996) to show that there exists some $z_0 > 0$ such that $E[e^{zB'(t)}] < \infty$ for all $0 \leq z \leq z_0$, and so the proof is complete. ∎

Proof of Theorem 1.

We need to show that for all sufficiently small δ, the absolute values of the difference quotients $D(\delta) \equiv \delta^{-1}[V(t, \theta_0 + \delta)L(t, \theta_0 + \delta) - V(t, \theta_0)]$ are dominated uniformly by a P_{θ_0}-integrable random variable. By the mean value theorem, for each sufficiently small δ, there exists $\eta \in (\theta - |\delta|, \theta + |\delta|)$ such that

$$D(\delta) = V'(t, \eta)L(t, \eta) + V(t, \eta)L'(t, \eta). \tag{7}$$

Let $\kappa = \max\{|\mathcal{E}(s)| : s \in \mathbf{S}\}$, which is finite since $|\mathbf{A}| < \infty$. Choose $\epsilon > 0$ such that

$$E_{\theta_0}((1 + \epsilon)^{4(2\kappa+1)N(t)}) < \infty. \tag{8}$$

Clearly, by Lemma 2, this is possible. Now choose a sufficiently small $h > 0$ so that Conditions A3, A4, A6, A7, A9, A10, A12 and part (iv) of Lemma 1 hold; Conditions A5, A8, A11 and part (iii) of Lemma 1 hold for the above ϵ; and (7) holds for all $|\delta| \leq h$. Hence in (7), the η's corresponding to these δ's always lie in $(\theta_0 - h, \theta_0 + h)$.

16

First consider the term $V(t, \eta)L'(t, \eta)$. From (5),

$$L'(t, \eta) = \frac{\mu(Y_0; \eta)}{\mu(Y_0; \theta_0)} \prod_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* \leq t}} \frac{f(c_0(e); e, \eta)}{f(c_0(e); e, \theta_0)} \prod_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* > t}} \frac{\bar{F}(\lambda_* t; e, \eta)}{\bar{F}(\lambda_* t; e, \theta_0)}$$

$$\cdot \prod_{k=1}^{N(t)} \frac{p(Y_k; Y_{k-1}, e_k, \eta)}{p(Y_k; Y_{k-1}, e_k, \theta_0)} \prod_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* \leq t}} \frac{f(c_k(e); e, \eta)}{f(c_k(e); e, \theta_0)} \prod_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* > t}} \frac{\bar{F}(\lambda_* t; e, \eta)}{\bar{F}(\lambda_* t; e, \theta_0)}$$

$$\cdot \left[\frac{\mu'(Y_0; \eta)}{\mu(Y_0; \eta)} + \sum_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* \leq t}} \frac{f'(c_0(e); e, \eta)}{f(c_0(e); e, \eta)} + \sum_{\substack{e \in \mathcal{E}(Y_0) \\ c_0(e)/\lambda_* > t}} \frac{\bar{F}'(\lambda_* t; e, \eta)}{\bar{F}(\lambda_* t; e, \eta)} \right.$$

$$\left. + \sum_{k=1}^{N(t)} \left(\frac{p'(Y_k; Y_{k-1}, e_k, \eta)}{p(Y_k; Y_{k-1}, e_k, \eta)} + \sum_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* \leq t}} \frac{f'(c_k(e); e, \eta)}{f(c_k(e); e, \eta)} + \sum_{\substack{e \in \mathcal{N}(Y_k; Y_{k-1}, e_k) \\ c_k(e)/\lambda_* > t}} \frac{\bar{F}'(\lambda_* t; e, \eta)}{\bar{F}(\lambda_* t; e, \eta)} \right) \right].$$

All of the density functions $f(\cdot; \cdot, \cdot)$ in the above expression are evaluated at points no greater than $\lambda_* t$. Now we define

$$M_1(h) = \sup \left\{ \sup \left\{ \frac{\mu(s; \theta)}{\mu(s; \theta_0)} : s \in \Lambda_\mu, \ |\theta - \theta_0| < h \right\}, 1 \right\},$$

$$M_2(h) = \sup \left\{ \left| \frac{\mu'(s; \theta)}{\mu(s; \theta)} \right| : s \in \Lambda_\mu, \ |\theta - \theta_0| < h \right\},$$

$$M_3(h) = \sup \left\{ \sup \left\{ \frac{f(x; e, \theta)}{f(x; e, \theta_0)} : e \in \mathbf{A}, \ x \in \Lambda_f(e), \ |\theta - \theta_0| < h \right\}, 1 \right\},$$

$$M_4(h) = \sup \left\{ \left| \frac{f'(x; e, \theta)}{f(x; e, \theta)} \right| : e \in \mathbf{A}, \ x \in \Lambda_f(e), \ |\theta - \theta_0| < h \right\},$$

$$M_5(h) = \sup \left\{ \sup \left\{ \frac{\bar{F}(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta_0)} : e \in \mathbf{A}, \ |\theta - \theta_0| < h \right\}, 1 \right\},$$

$$M_6(h) = \sup \left\{ \left| \frac{\bar{F}'(\lambda_* t; e, \theta)}{\bar{F}(\lambda_* t; e, \theta)} \right| : e \in \mathbf{A}, \ |\theta - \theta_0| < h \right\},$$

$$M_7(h) = \sup \left\{ \left\{ \frac{p(s'; s, e, \theta)}{p(s'; s, e, \theta_0)} : (s', s, e) \in \Lambda_p, \ |\theta - \theta_0| < h \right\}, 1 \right\},$$

$$M_8(h) = \sup \left\{ \left| \frac{p'(s'; s, e, \theta)}{p(s'; s, e, \theta)} \right| : (s', s, e) \in \Lambda_p, \ |\theta - \theta_0| < h \right\}.$$

By Conditions A1, A4–A12, and part(iii) and (iv) of Lemma 1, $M_2(h)$, $M_4(h)$, $M_6(h)$, and $M_8(h)$ are finite, and $M_1(h)$, $M_3(h)$, $M_5(h)$, and $M_7(h)$ lie in the interval $[1, 1 + \epsilon]$.

Then, letting $\phi_1(h) = M_1(h) \, (\, (M_3(h) \, M_5(h))^\kappa \, (\, M_7(h) \,)^{N(t)+1}$ and $\phi_2(h) = M_2(h) + (N(t) + 1) \, (\, \kappa(M_4(h) + M_6(h)) + M_8(h) \,)$, we obtain $|L'(t, \eta)| \leq \phi_1(h)\phi_2(h)$, and so

$$|V(t, \eta)L'(t, \eta)| \;\leq\; |V(t, \eta)| \, \phi_1(h) \, \phi_2(h) \;\leq\; W \, \phi_1(h) \, \phi_2(h) \;\equiv\; \phi(h)$$

with P_{θ_0}-probability 1 by Condition A3. Repeated applications of the Schwarz inequality yields $E_{\theta_0}[\phi(h)] \leq E_{\theta_0}^{1/2}[W^2] \, E_{\theta_0}^{1/4}[\phi_1(h)^4] \, E_{\theta_0}^{1/4}[\phi_2(h)^4]$. By Condition A3, $E_{\theta_0}^{1/2}[W^2] < \infty$. In addition, using the fact that $\phi_1(h) \leq (1 + \epsilon)^{2\kappa+2}(1 + \epsilon)^{(2\kappa+1)N(t)}$ and (8) we have that $E_{\theta_0}^{1/4}[\phi_1(h)^4] < \infty$. Also, Lemma 2 implies that all moments of $N(t)$ exist under P_{θ_0} and hence

$$E_{\theta_0}^{1/4}\left[\phi_2(h)^4\right] \leq (\kappa(M_4(h) + M_6(h)) + M_8(h) + M_2(h)) \, E_{\theta_0}^{1/4}\left[(N(t) + 1)^4\right] < \infty.$$

Hence, $\phi(h)$ is P_{θ_0}-integrable. Similarly, we can show that $|V'(t, \eta)L(t, \eta)|$ is uniformly dominated over $(\theta_0 - h, \theta_0 + h)$, by a P_{θ_0}-integrable random variable. Finally, applying the dominated convergence theorem completes the proof. ∎

It will be worthwhile comparing our conditions with similar ones in the literature. First, consider the "amiability" condition used in Theorem 3 of Reiman and Weiss (1989). One may use Theorem 3 of Reiman and Weiss to validate the exchange of derivative and expectation only in special cases of our setup, i.e., where θ is present only in one distribution and it never appears as a variable in the performance measure. However, even if we impose these restrictions in our setting, the amiability condition is difficult to verify directly. One can show, only with the help of Lemma 2 and a mean value theorem, that our assumptions (with Condition A3 replaced by the weaker and more tractable assumption used in Proposition 1 below) imply the amiability condition.

Secondly, consider assumption A3(k) for $k = 1$ of L'Ecuyer (1995). Though the setting used there is different from ours, we can still make some comparisons. For example, our Condition A3 is similar to his conditions on the performance measure. Also, in some sense,

our type of distributional conditions imply the distributional conditions in A3(k) of L'Ecuyer; i.e, if our type of distributional conditions hold, then one can use a constant close to 1 as his bounding random variable $\Gamma_{1i}(\zeta)$ and some constant as his bounding random variable $\Gamma_{2i}(\zeta)$. If we use $N(t) + 1$ as an analogue to his discrete stopping time τ, then an analogous result to Lemma 2 given below will prove that the conditions specified in A3(k) on $\Gamma_{1i}(\zeta)$ and $\Gamma_{2i}(\zeta)$ hold.

We now compare our conditions to those employed by Glasserman (1991a,1991b) to show the unbiasedness of IPA derivative estimators in the GSMP context. First, note that our conditions mainly impose restrictions on the basic building blocks of the GSMP. On the other hand, the conditions for IPA are primarily on the structure of the GSMP. In particular, Glasserman requires a "commuting condition" which places limitations on the relationship between active events in certain states and the transition probabilities from those states and from possible immediate successor states. The condition essentially ensures that if slightly altering the value of the parameter θ results in a change in the order of two events on a sample path, then the sample path can (in some sense) correct itself on the next event. Moreover, Glasserman does not permit the initial distribution nor the transition probabilities to depend on the parameter θ.

Another difference is the type of performance measure $\alpha(t, \theta) = E_\theta[V(t, \theta)]$ allowed. In our results, we restrict $V(t, \theta)$ to be any $\mathcal{F}_t$-measurable random variable that satisfies Condition A3, where t is deterministic. Glasserman allows $V(t, \theta)$ to be an additive functional up to a time T, but T is allowed to be either deterministic, the time of the n-th transition, or the time of the n-th occurrence of a particular event.

4   Sufficient Conditions for A1-A12 to Hold

We now show that our assumptions hold in many settings arising in practice. (In Section 5 we explicitly demonstrate that the following sufficient conditions are satisfied for large classes of reliability and queueing models.) First consider Condition A3.

Proposition 1 *Suppose Conditions A1, A2, A4 and A5 hold. Also, suppose $\alpha(t,\theta) = E_\theta[V(t,\theta)]$, where $V(t,\theta)$ is an $\mathcal{F}_t$-measurable random variable such that $V'(t,\theta,\omega)$ exists for P_{θ_0}-almost all $\omega \in \Omega$ and all $\theta \in (\theta_0 - h, \theta_0 + h)$ for some $h > 0$ and that satisfies the following:*

   (i) *there exist finite constants $K_1 > 0$ and $K_2 > 0$ such that $|V(t,\theta)| \le K_1(N(t)+t)^{K_2}$ with P_{θ_0}-probability 1 for all $\theta \in (\theta_0 - h, \theta_0 + h)$ for some $h > 0$;*

   (ii) *there exist finite constants $K_3 > 0$ and $K_4 > 0$ such that $|V'(t,\theta)| \le K_3(N(t)+t)^{K_4}$ with P_{θ_0}-probability 1 for all $\theta \in (\theta_0 - h, \theta_0 + h)$ for some $h > 0$.*

Then Condition A3 holds.

Proof. We need to show that $W \equiv K_1(N(t)+t)^{K_2}$ and $W' \equiv K_2(N(t)+t)^{K_3}$ have finite second moments when the parameter value $\theta = \theta_0$. But this follows from Lemma 2. ∎

We now show directly that many performance measures arising in practice satisfy Condition A3.

Proposition 2 *Suppose the performance measure $\alpha(t,\theta) = E_\theta[V(t,\theta)]$ satisfies one of the following:*

   (i) $V(t,\theta) = 1\{A\}$, *where $\{A\}$ is some $\mathcal{F}_t$-measurable event;*

   (ii) $V(t,\theta) = \int_0^t g(Z_u, \theta) du$, *where $g(\cdot,\theta)$ is some real-valued function on $\mathbf{S}$ with $g(s,\cdot)$ continuously differentiable for each $s \in \mathbf{S}$, and either **(a)** $|\mathbf{S}| < \infty$, or **(b)** $g(\cdot,\cdot)$ and $g'(\cdot,\cdot)$ are uniformly bounded on $\mathbf{S} \times (\theta_0 - h, \theta_0 + h)$ for some $h > 0$, where $g'(\cdot,\theta) = \frac{d}{d\theta}g(\cdot,\theta)$.*

Then Condition A3 holds.

Note that $V'(t,\theta) = 0$ if $V(t,\theta)$ has form (i) above, and $V'(t,\theta) = \int_0^t g'(Z_u,\theta) du$ if $V(t,\theta)$ has form (ii).

Proof. First note that $V(t,\theta)$ in **(i)** and **(ii)** are $\mathcal{F}_t$-measurable. When condition **(i)** holds, let $W = W' = 1$, which trivially satisfies Condition A3. Now consider form **(ii)**. Assuming

that either condition **(ii)(a)** or **(ii)(b)** holds and since $g(s, \cdot)$ is continuously differentiable, we have that for sufficiently small $h > 0$, $\bar{g} \equiv \sup\{|g(s, \theta)| : s \in \mathbf{S}, \ \theta \in (\theta_0 - h, \theta_0 + h)\} < \infty$ and $\bar{g}' \equiv \sup\{|g'(s, \theta)| : s \in \mathbf{S}, \ \theta \in (\theta_0 - h, \theta_0 + h)\} < \infty$. Hence, $|V(t, \theta)| \leq W \equiv \bar{g}t$ and $|V'(t, \theta)| \leq W' \equiv \bar{g}'t$, which satisfies Condition A3. ∎

Now we show that two rich classes of distributions, the Gamma distribution and the Weibull distribution, with the θ in each case being the scale parameter, satisfy the distributional conditions given in A4–A6. The Gamma distribution is widely used in all types of stochastic modeling; the Weibull distribution is one of the most frequently employed distributions in reliability modeling.

Proposition 3 *Suppose the lifetime of an event e has a gamma density with shape parameter $a > 0$ and scale parameter $\theta > 0$; i.e., $f(x; e, \theta) = \theta e^{-\theta x}(\theta x)^{a-1}/\Gamma(a)$. Fix $\theta_0 > 0$. Then for the given e,*

   (i) *the set $\{x : 0 \leq x \leq t, \ f(x; e, \theta) > 0\}$ is independent of θ for $\theta > 0$;*

   (ii) *$f(x; e, \theta)/f(x; e, \theta_0) \to 1$ uniformly on $[0, t]$ as $\theta \to \theta_0$;*

   (iii) *$f'(x; e, \theta)/f(x; e, \theta)$ is uniformly bounded on $[0, t] \times (\theta_0/2, 3\theta_0/2)$.*

Proof. Part (i) is trivial. For part (ii), note that $f(x; e, \theta)/f(x; e, \theta_0) = (\theta/\theta_0)^a \, e^{(\theta_0 - \theta)x}$, which converges to 1 uniformly on $[0, t]$ as $\theta \to \theta_0$.

For part (iii), note that $f'(x; e, \theta) = (a\theta^{a-1}e^{-\theta x}x^{a-1} - \theta^a e^{-\theta x}x^a)/\Gamma(a)$, and so

$$\left|\frac{f'(x; e, \theta)}{f(x; e, \theta)}\right| = \left|\frac{a}{\theta} - x\right| \leq \frac{2a}{\theta_0} + t < \infty$$

for all $x \in [0, t]$ and all $\theta \in (\theta_0/2, 3\theta_0/2)$. ∎

Proposition 4 *Suppose the lifetime of an event e has a Weibull density with shape parameter $a > 0$ and scale parameter $\theta > 0$; i.e., $f(x; e, \theta) = \theta(\theta x)^{a-1} \exp\{-(\theta x)^a\}$. Fix $\theta_0 > 0$. Then for the given e,*

21

(i) *the set* $\{x : 0 \leq x \leq t, \ f(x; e, \theta) > 0\}$ *is independent of* θ *for* $\theta > 0$;

(ii) $f(x; e, \theta)/f(x; e, \theta_0) \to 1$ *uniformly on* $[0, t]$ *as* $\theta \to \theta_0$. ;

(iii) $f'(x; e, \theta)/f(x; e, \theta)$ *is uniformly bounded on* $[0, t] \times (\theta_0/2, 3\theta_0/2)$.

We omit the proof since it is similar to that of Proposition 3.

Note that we do not need to consider any distributions that do not depend on the parameter θ when checking if Conditions A4–A6 are satisfied. The same is true for the initial distribution $\mu(\cdot; \theta)$ and Conditions A7–A9, and the transition probabilities $p(\cdot; \cdot, \cdot, \theta)$ and Conditions A10–A12.

The following result shows that if the state space is finite and the convergence and bounds in Conditions A8, A9, A11, and A12 hold, then they are automatically uniform. The proof is trivial and is therefore omitted.

Proposition 5 *Suppose that* $|\mathbf{S}| < \infty$. *Then*

(i) *Condition A8 holds if* $\mu(s; \theta)/\mu(s; \theta_0) \to 1$ *for each* $s \in \mathbf{S}$;

(ii) *Condition A9 holds if for each* $s \in \mathbf{S}$, *there exists some* $h > 0$ *(which may depend on s) such that* $\mu'(s; \cdot)/\mu(s; \cdot)$ *is bounded on* $(\theta_0 - h, \theta_0 + h)$;

(iii) *Condition A11 holds if for each* $(s', s, e) \in \Lambda_p$, $p(s'; s, e, \theta)/p(s'; s, e, \theta_0) \to 1$ *as* $\theta \to \theta_0$;

(iv) *Condition A12 holds if for each* $(s', s, e) \in \Lambda_p$, *there exists some* $h > 0$ *(which may depend on s) such that* $p'(s'; s, e, \cdot)/p(s'; s, e, \cdot)$ *is bounded on* $(\theta_0 - h, \theta_0 + h)$.

5 Examples

We now give some examples of systems for which Conditions A1–A12 hold. The first example deals with a large class of reliability models, and the second example considers a large class of multiclass queueing networks.

5.1 Reliability Models

Consider the class of reliability models studied in Heidelberger, Shahabuddin, and Nicola (1994), Nicola, Shahabuddin, Heidelberger, and Glynn (1993), and Nicola, Nakayama, Heidelberger, and Goyal (1993). These are systems composed of K components, where each component can fail and get repaired. Component lifetimes and repair times are generally distributed. There are R classes of repairmen in the system, each consisiting of a finite number of repairmen. The repair of each component is assigned to one repairman class. The repair strategy used by a repairman class may be pre-emptive or non-pre-emptive priority with FCFS or random order service used between members of the same priority class. We allow for the following interpendencies among the components. First, the operation of an operational component may depend on certain other components being operational. In addition, the repair of a failed component may depend on certain other components being operational. Also, the failure of a component may cause other components to fail instantaneously with certain probabilities. This is called failure propagation, which is an example of how cancelled events arise. The system is considered operational if certain sets of components are operational. The physical state of the system after the nth transition is given by $Y_n = \{Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(K)}; Q_n^{(1)}, Q_n^{(2)}, \ldots Q_n^{(R)}\}$ where $Y_n^{(i)}$ is 1 or 0 depending on whether the ith component is operational or failed after the n-th transition, and $Q_n^{(j)}$ is the repair queue at the j-th repairman class. Note that $|\mathbf{S}| < \infty$. It is easy to see that the process $\{Z_t : t \geq 0\}$ (recall that $Z_t = Y_{N(t)}$) is a GSMP. As we previously mentioned, the system is considered operational if certain sets of components are operational. More formally, $\mathbf{S} = \mathbf{U} \cup \mathbf{F}$, with $\mathbf{U} \cap \mathbf{F} = \emptyset$, where $\mathbf{U}$ is the set of operational states and $\mathbf{F}$ is the set of failed states. The system is considered operational at time t if $Z_t \in \mathbf{U}$; otherwise, it is considered down. We assume that at time $t = 0$, all components in the system are operational and new (i.e., all components have age $= 0$). Thus, the initial distribution is independent of the parameter θ, and so Conditions A7–A9 are automatically satisfied.

The events in the system are either component failures or component repairs. Clearly,

$|\mathbf{A}| < \infty$ as there are a maximum of $2K$ events in the entire system since each of the K components can have a failure and repair event associated with it. Hence, Condition A2 is satisfied since $|\mathbf{S}| < \infty$. If there is no failure propagation in the system then the outcome of any event e, given that the system is currently in some state s, is fixed; i.e., $p(s'; s, e, \theta) = 1$ for exactly one state s' and 0 otherwise. However, if we have failure propagation in the system then there may be some randomness in the final state reached from any given state after an event, and Proposition 5 provides sufficient conditions for Conditions A11 and A12 to hold.

The transient performance measures in which one may be interested are the unreliability and the expected interval unavailability. The unreliability is the probability of a system failure before some fixed time t. Hence, in this case $\alpha(t, \theta) = E_\theta[V(t, \theta)]$ where $V(t, \theta)$ is of form (i) in Proposition 2 and A is the event $\{Z_s \in \mathbf{F} \text{ for some } s \leq t\}$. Thus, Condition A3 holds.

The interval unavailability is the fraction of time in $[0, t]$ that the system is down. In this case, $\alpha(t, \theta) = E_\theta[V(t, \theta)]$, where $V(t, \theta)$ is of form (ii) in Proposition 2 with $g(Z_s, \theta) = 1\{Z_s \in \mathbf{F}\}/t$. Note that $g(Z_s, \theta)$ is independent of θ and therefore $V'(t, \theta, \omega) = 0$ for all $\omega \in \Omega$. Also, $V(t, \theta) \leq 1$ with P_{θ_0}-probability 1, and so Condition A3 is satisfied.

In the reliability setting one is usually interested in the derivatives of the performance measures with respect to (i) parameters of the failure time distribution of individual components; (ii) parameters of the repair time distribution of individual components; (iii) the failure propagation probabilities. For the first two cases, the likelihood ratio derivative estimation method is applicable if the distributions that depend on the parameter with respect to which the derivative is taken satisfy Conditions A4–A6. Some special cases where these conditions are satisfied are when the distribution and the parameter are as in Propositions 3 and 4.

In the third case, the density functions $f(\,\cdot\,; e, \theta)$ are independent of the parameter θ, and so Conditions A4–A6 are automatically satisfied. Also, $p(s'; s, e, \theta)$ typically will be of the form $c_1 + c_2\theta$, where c_1 and c_2 are constants. Clearly, Condition A10 is satisfied for $\theta \in (0, \infty)$. Since $|\mathbf{S}|$ and $|\mathbf{A}|$ are finite, we only have to check the conditions given in Proposition 5. For the particular form of $p(s'; s, e, \theta)$ described above, these conditions are satisfied for all

$\theta_0 \in (0, \infty)$.

For this class of reliability systems, it turns out that Glasserman's (1991a,1991b) conditions for the unbiasedness of IPA derivatives typically are not satisfied. In particular, Glasserman assumed that events are never cancelled, but this could occur in this setting since we allowed for failure propagation. Also, his "commuting condition" will not hold for many repair strategies. For example, if there is only one repairman who fixes all failed components in a first-come first-served fashion, then this condition will be violated. Moreover, Glasserman's work does not cover performance measures such as the unreliability. However, IPA can admit many more event-time distributions. For example, one can use a repair time that is uniformly distributed over $(0, \theta)$ (the θ also being the parameter) in IPA, but it is not allowed by the results in this paper.

5.2   Queueing Networks

Consider a multiclass queueing network where service times for each customer class at each station are generally distributed random variables. We assume that there is a finite number of servers at each station, and that each customer class uses Markovian routing; i.e., when a customer of type i completes service at station j, he immediately goes to station k with probability $R_i(j, k)$. We consider both open and closed networks. For the case of open networks, we assume that the arrival process of a class at each station is a renewal process with generally distributed inter-renewal times. The events in the system are customer arrivals (one for each customer class at each station) and customer service completions (one for each customer class for each server at each station). Hence, again the event set is finite.

In most applications Condition A2 is satisfied. As mentioned earlier, one situation in which it is not is in an open network with the service rate of customers at a station is proportional to the number of customers in the system.

If we assume that the network is closed and always starts in the same customer configuration, then the initial distribution is independent of the parameter θ, and so Conditions A7–A9

are automatically satisfied. Similarly, if we assume that the network is open and always starts with no customers present, then the initial distribution is independent of the parameter θ, and Conditions A7–A9 automatically hold.

Some of the performance measures of interest in such systems are (i) the expected time-average queue length over $[0, t]$ at a given station; (ii) for the case of open networks, the expected time-average number of customers over $[0, t]$ in the entire system; (iii) for the case of open networks, the probability that total number of customers in the system exceeds some threshold in time interval $[0, t]$ (if the network has a finite combined buffer, then this is the probability that there is at least one customer loss due to buffer overflow in $[0, t]$).

We now show that Condition A3 holds for each of the above performance measures. Since in all cases $V(t, \theta)$ is independent of θ and $V'(t, \theta, \omega) = 0$ for all $\omega \in \Omega$, we only have to show that $E_{\theta_0}[(V(t, \theta))^2] < \infty$. For case (iii), this is obvious. Case (i) for closed networks is also obvious. For case (i) for open networks and case (ii) and no customers are initially in the system, this follows from Lemma 2 when Conditions A2, A4 and A5 hold. This is because, say for case (ii), if we let $Q(s)$ be the total number of customers in the system at time s, then

$$V(t, \theta) = \frac{1}{t} \int_{s=0}^{t} Q(s)ds \leq \frac{1}{t} \int_{s=0}^{t} N(s)ds \leq \frac{1}{t} \int_{s=0}^{t} N(t)ds = N(t),$$

and so Proposition 1 implies that Condition A3 holds under Conditions A2, A4 and A5.

One may be interested in the derivatives of the performance measures with respect to (i) a parameter of the service time distribution of a class at a particular station; (ii) in the case of open networks, a parameter of the interarrival-time distribution of a certain customer class at a particular station; (iii) the routing probability of a certain customer class at a particular station. As in the case with the reliability models, for the first two cases, the likelihood ratio derivative estimation method is applicable if the distributions that depend on the parameter with respect to which the derivative is taken satisfy Conditions A4–A6. Some widely used cases when these conditions are satisfied are when the distribution and the parameter are of the type given in Proposition 3 and Proposition 4.

In the third case, for the case of closed networks, due to the finiteness of the state space, we can again apply Proposition 5. For the case of open networks, Proposition 5 is no longer useful as now the state space is no longer finite. However, since the number of stations is finite and since there is Markovian routing, there are only finitely many different values that the $p(s'; s, e, \theta)$, $s, s' \in \mathbf{S}$, $e \in \mathbf{A}$, can assume. Hence, to show that Conditions A11 and A12 hold, we only have to make sure that each of the transition probabilities is continuous at θ_0, and that each of the $p'(s'; s, e, \theta)/p(s'; s, e, \theta)$ is also bounded in some neighborhood of θ_0. These conditions hold if $p(s'; s, e, \theta)$ is of the form $c_1 + c_2\theta$, where c_1 and c_2 are constants.

For the multi-class queueing systems considered here, Glasserman's (1991a,1991b) conditions for unbiasedness of IPA derivatives typically are not satisfied. Specifically, his "commuting condition" will not hold for multi-class networks of our generality. For example, if a particular station that is visited by more than one customer class is fed by more than one source, then this condition is violated. Also, Glasserman's work does not cover performance measures such as the probability of a buffer overflow occurring before time t. Finally, Glasserman disallows the routing probabilities (more generally, the state-dependent transition probabilities) to depend on the parameter θ. However, as we mentioned before, IPA admits many more event-time distributions.

We should mention that Conditions A1–A12 will hold for more general queueing networks. For example, it can be shown that they typically are valid when there is state-dependent routing. [1]

A   Appendix

Proof of Lemma 1:

(i) Obvious.

(ii) We wish to show that for any $\tilde{\epsilon} > 0$ there exists $\delta > 0$, such that for all θ satisfying $|\theta - \theta_0| < \delta$ and for all $y \in [0, \lambda_* t]$,

$$|\bar{F}(y; e, \theta) - \bar{F}(y; e, \theta_0)| \leq \tilde{\epsilon}. \tag{9}$$

From Conditions A4 and A5, for any $\tilde{\epsilon} > 0$, there exists $h_1 > 0$ (smaller than the h is Condition A4), such that for all θ satisfying $|\theta - \theta_0| < h_1$, $1 - \tilde{\epsilon} \leq f(x; e, \theta)/f(x; e, \theta_0) \leq 1 + \tilde{\epsilon}$ or

$$f(x; e, \theta_0) - \tilde{\epsilon}f(x; e, \theta_0) \leq f(x; e, \theta) \leq f(x; e, \theta_0) + \tilde{\epsilon}f(x; e, \theta_0) \tag{10}$$

for all $x \in \Lambda_f(e)$. From Condition A4, for $x \in [0, \lambda_* t] - \Lambda_f(e)$, $f(x; e, \theta) = f(x; e, \theta_0) = 0$ for all θ satisfying $|\theta - \theta_0| \leq h_1$. Therefore (10) still holds. Choose $\delta = h_1$. Hence from (10), for any $\tilde{\epsilon} > 0$ there exists $\delta > 0$, such that for all θ satisfying $|\theta - \theta_0| < \delta$, we have that

$$F(y; e, \theta_0) - \tilde{\epsilon}F(y; e, \theta_0) \leq F(y; e, \theta) \leq F(y; e, \theta_0) + \tilde{\epsilon}F(y; e, \theta_0)$$

or $\quad \bar{F}(y; e, \theta_0) + \tilde{\epsilon}F(y; e, \theta_0) \geq \bar{F}(y; e, \theta) \geq \bar{F}(y; e, \theta_0) - \tilde{\epsilon}F(y; e, \theta_0),$

for all $y \in [0, \lambda_* t]$. Subtracting $\bar{F}(y; e, \theta_0)$ from each term and using the fact $F(y; e, \theta_0) \leq 1$, we get (9).

(iii) This follows immediately from part (ii) of the Lemma.

(iv) By Conditions A4 and A5, for any $\tilde{\epsilon} > 0$, there exists $h_1 > 0$ (smaller than the h in Condition A4) such that $f(x; e, \theta) \leq f(x; e, \theta_0)(1 + \tilde{\epsilon})$ for all $\theta \in (\theta_0 - h_1, \theta_0 + h_1)$. For $\delta > 0$ and $x \in \Lambda_f(e)$, define $D(\delta; x, \theta) = (f(x; e, \theta + \delta) - f(x; e, \theta))/\delta$. Then by Conditions A4 and A6, there exists a h_2 smaller than h_1, such that for all $\theta \in (\theta_0 - h_2, \theta_0 + h_2)$ and $x \in \Lambda_f(e)$, $f'(x; e, \theta) \equiv \lim_{\delta \to 0} D(\delta; x, \theta)$ exists and $|f'(x; e, \theta)| \leq Kf(x; e, \theta) \leq (1 + \tilde{\epsilon})Kf(x; e, \theta_0)$, for some constant $K > 0$. From part(iii) of Lemma 1, there exists h_3 such that for all $\theta \in (\theta_0 - h_3, \theta_0 + h_3)$ $\bar{F}(\lambda_* t; e, \theta) \geq 0.5\bar{F}(\lambda_* t; e, \theta_0)$.

Choose the h in part (iv) of Lemma 1 to be $\min(h_2, h_3)/2$. By the mean value theorem, for all $\theta \in (\theta_0 - h, \theta_0 + h)$ (since, as shown in the previous paragraph, $f'(x; e, \theta)$ exists), for all

sufficiently small δ there exists $\eta \in (\theta - \delta, \theta + \delta)$ such that $D(\delta; x, \theta) = f'(x; e, \eta)$. Then for all $\delta \leq h_2/2$ we have that $\eta \in (\theta_0 - h_2, \theta_0 + h_2)$, and therefore $|f'(x; e, \eta)| \leq Kf(x, \theta_0)(1 + \tilde{\epsilon})$. Also

$$\int_{x \in \Lambda_f(e)} Kf(x, \theta_0)(1 + \tilde{\epsilon})dx \leq K(1 + \tilde{\epsilon}) < \infty.$$

Then by the dominated convergence theorem, for all $\theta \in (\theta_0 - h, \theta_0 + h)$, $\int_{x \in \Lambda_f(e)} |f'(x; e, \theta)|dx < \infty$ and $F'(\lambda_* t; e, \theta) = \int_{x \in \Lambda_f(e)} f'(x; e, \theta)dx$, and hence

$$
\begin{aligned}
\frac{|\bar{F}'(\lambda_* t; e, \theta)|}{\bar{F}(\lambda_* t; e, \theta)} &= \frac{|F'(\lambda_* t; e, \theta)|}{\bar{F}(\lambda_* t; e, \theta)} = \frac{|\int_{x \in \Lambda_f(e)} f'(x; e, \theta)dx|}{\bar{F}(\lambda_* t; e, \theta)} \\
&\leq \frac{|\int_{x \in \Lambda_f(e)} Kf(x; e, \theta)dx|}{\bar{F}(\lambda_* t; e, \theta)} \leq K\left(\frac{1}{0.5\bar{F}(\lambda_* t; e, \theta_0)} - 1\right) < \infty. \quad \blacksquare
\end{aligned}
$$

References

Athreya, K. B. and P. E. Ney, *Branching Processes*, Springer-Verlag, Berlin, (1972).

Aleksandrov, V. M., V.I. Sysoyev and V. V. Shemeneva, "Stochastic Optimization," *Engineering Cybernetics* 5 (1968), 11–16.

Arsham, H., A. Feuerberger, D.L. McLeish, J.Kreimer and R.Y. Rubinstein, "Sensitivity Analysis and the "What If" Problem in Simulation Analysis," *Math. Comput. Modelling* 12 (1989), 193-219.

Billingsley, P., *Probability and Measure, Second Edition*, John Wiley and Sons, New York, (1986).

Blum, A., P. Heidelberger, S.S. Lavenberg, M.K. Nakayama and P. Shahabuddin, *System Availability Estimator (SAVE) Language Reference and User's Manual, Version 4.0*, Research Report RA 219 S, IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, (1993).

Bratley, P., B. L. Fox and S.E. Schrage, *A Guide to Simulation* (2nd ed.), Springer-Verlag, New-York, (1987).

Damerdji, H., "Maximum Likelihood Ratio Estimation for Generalized Semi-Markov Processes," *Discrete Event Dynamic Systems: Theory and Applications* 6 (1996), 73–104.

Glasserman, P., *Gradient Estimation Via Perturbation Analysis*, Kluwer Academic Press, Norwell, MA, (1991a).

Glasserman, P., "Structural Conditions for Perturbation Analysis Derivative Estimates: Finite Time Performance Indices," *Oper. Res.* 39 (1991b), 724–738.

Glasserman, P., "Structural Conditions for Perturbation Analysis of Queuing Systems," *Journal of the ACM* 38 (1991c), 1005–1025.

Glasserman, P. and P. W. Glynn, "Gradient Estimation for Regenerative Processes," *Proceedings of the 1992 Winter Simulation Conference*, Swain, J. J., Goldsman, D., Crain, R. C. and Wilson, J. R. (Eds.), IEEE Press, (1992), 280–288.

Glasserman, P. and D. D. Yao, "Monotonicity in Generalized Semi-Markov Processes," *Math. Oper. Res.* 17 (1992a), 1–21.

Glasserman, P. and D. D. Yao, "Generalized Semi-Markov Processes: Antimatroid Structure and Second-Order Properties," *Math. Oper. Res.* 17 (1992b), 444–469.

Glasserman, P. and D. D. Yao, "Some Guidelines and Guarantees for Common Random Numbers," *Management Sci.* 38 (1992c), 884–908.

Glynn, P. W., "Stochastic Approximation for Monte Carlo Optimization," *Proceedings of the 1986 Winter Simulation Conference*, Wilson, J., Henriksen, J. and Roberts, S. (Eds.), IEEE Press, (1986), 356–364.

Glynn, P. W., Likelihood ratio gradient estimation: an overview. *Proceedings of the 1987 Winter Simulation Conference*, Thesen, A., Grant, H. and Kelton, W. D. (Eds.), IEEE Press, (1987), 366–375.

Glynn, P. W., "Optimization of stochastic systems via simulation," *Proceedings of the 1989 Winter Simulation Conference*, MacNair, E.A., Musselman, K.J. and Heidelberger, P. (Eds.), IEEE Press, (1989a), 90-105.

Glynn, P. W., "A GSMP Formalism for Discrete-Event Systems," *Proceedings of the IEEE*

77 (1989b), 14–23.

Glynn, P. W., "Likelihood Ratio Derivative Estimators for Stochastic Systems," *Comm. ACM* 33, 10 (1990), 75–84.

Glynn, P. W. and D. L. Iglehart, "Simulation Methods for Queues: An Overview," *Queueing Systems: Theory and Applications* 3 (1988), 221–256.

Glynn, P. W. and D. L. Iglehart, "Importance Sampling for Stochastic Simulations," *Management Science* 35 (1989), 1367–1393.

Haas, P. J. and G. S. Shedler, "Regenerative Generalized Semi-Markov Processes," *Stochastic Models* 3 (1987), 409–438.

Heidelberger, P., X. R. Cao, M. A. Zazanis, R. Suri, "Convergence Properties of Infinitesimal Perturbation Analysis Estimates," *Management Sci.* 34 (1988), 1281–1302.

Heidelberger, P., P. Shahabuddin, and V. F. Nicola, "Bounded Relative Error in Estimating Transient Measures of Highly Dependable Non-Markovian Systems," *ACM Transactions on Modeling and Computer Simulation*, 4 (1994), 137–164.

Ho, Y. C. and X. R. Cao, "Optimization and Perturbation Analysis of Queueing Networks," *Journal of Optimization Theory and Applications* 40 (1983), 559–582.

L'Ecuyer, P., "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques," *Management Sci.* 36 (1990), 1364–1383.

L'Ecuyer, P., "An Overview of Derivative Estimation," *Proceedings of the 1991 Winter Simulation Conference*, Nelson, B.L., Kelton, W.D. and Clark, G.M. (Eds.), IEEE Press, (1991), 207-217.

L'Ecuyer, P., "Note: On the Interchange of Derivative and Expectation for Likelihood Ratio Derivative Estimators," *Management Sci.* 41 (1995), 738–748.

Nakayama, M. K., "Asymptotics of Likelihood Ratio Derivative Estimators in Simulations of Highly Reliable Markovian Systems," *Management Sci.* 41 (1995), 524–554.

Nakayama, M. K., A. Goyal, and P. W. Glynn, "Likelihood Ratio Sensitivity Analysis for Markovian Models of Highly Dependable Systems," *Oper. Res.* 42 (1994), 137–157.

Nakayama, M.K., P. Shahabuddin and K. Sigman, "A Note on Exponential Moments for Branching Processes and Busy Periods for Queues," preprint (1996).

Nicola, V.F., M. K. Nakayama, P. Heidelberger, and A. Goyal, "Fast Simulation of Highly Dependable Systems with General Failure and Repair Processes.," *IEEE Transactions on Computers* 14 (1993), 1440–1452.

Nicola, V. F., P. Shahabuddin, P. Heidelberger, and P. W. Glynn, "Fast Simulation of Steady State Availability in Non-Markovian Highly Dependable Systems," *Proceedings of the Twenty-Third Annual International Symposium on Fault Tolerant Computing* IEEE Computer Society Press, (1993), 38–47.

Prabhu, N. U., *Stochastic Processes*, Macmillan, New York, (1965).

Reiman, M. I. and A. Weiss, "Sensitivity Analysis via Likelihood Ratios," *Proceedings of the 1986 Winter Simulation Conference*, Wilson, J.R., Henriksen, J.O. and Roberts, S.D. (Eds.), IEEE Press, (1986), 285–289.

Reiman, M. I. and A. Weiss, "Sensitivity Analysis for Simulations via Likelihood Ratios," *Oper. Res.* 37 (1989), 830–844.

Ross, S., *Stochastic Processes*, J. Wiley, New York, (1983).

Rubinstein, R. Y., "The Score Function Approach for Sensitivity Analysis of Computer Simulation Models," *Math. Comput. Simulation* 28 (1986), 351–379.

Rubinstein, R. Y., Sensitivity analysis and performance extrapolation for computer simulation models. *Oper. Res.* **37** (1989), 72–81.

Rubinstein, R. Y. and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, J. Wiley, (1993).

Shassberger, R., "On the Equilibrium Distribution of a Class of Finite-State Generalized Semi-Markov Processes," *Math. Oper. Res.* 1 (1976), 395–406.

Whitt, W., "Continuity of Generalized Semi-Markov Processes," *Math. Oper. Res.* 5 (1980), 494–501.