SENSITIVITY ANALYSIS FROM SAMPLE PATHS USING LIKELIHOODS

Philip Heidelberger and Don Towsley

COINS Technical Report 87-131
December 1, 1987

Sensitivity Analysis from Sample Paths Using Likelihoods

Philip Heidelberger

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, New York 10598

Don Towsley [1]

Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01003

Abstract

We modify the likelihood-based method for obtaining derivatives with respect to the rate of a Poisson process so that it is not necessary to know the exact value of that rate. This type of modification is necessary if the method is to be used on a sample path from a real system. The modification to the likelihood estimator is simply to use the value of the Poisson rate estimated during the sample interval. For regenerative systems, this produces a strongly consistent, asymptotically normal and asymptotically unbiased estimate of the derivative. The strong law and central limit theorem are generalized to the case of estimating a derivative with respect to an unknown parameter from the exponential class of probability density functions. Numerical results for the M/M/1 queue illustrate little difference between the estimates for the derivative of the expected delay with respect to arrival rate obtained when the arrival rate is known and unknown. However, both estimates are highly biased for small sample sizes. This bias can be reduced by jackknifing.

Version Dated December 1st, 1987

[1] The work of this author was partially supported by the Office of Naval Research under grant number N00014-87-K-0796.

1. Introduction

In recent papers, Glynn (1986), Reiman and Weiss (1986) and Rubinstein (1986) describe a simple method for estimating derivatives with respect to an input parameter in a discrete event simulation. This method has its roots in classical statistics (see, e.g., Section 32 of Cramér (1946)). As a specific example, the method can be used to estimate the derivative of expectations with respect to the parameter of a Poisson process in many queueing systems. The estimator, called the likelihood ratio estimator in Reiman and Weiss (1986), is based on the derivative of logarithm of the likelihood (also called the efficient score, see Rubinstein (1986)) of the sample path and is known to produce valid derivative estimates for a large class of systems.

There is another class of estimators for computing derivatives based on the technique of perturbation analysis; see, e.g., Ho and Cassandras (1983) for an introduction to this subject and Heidelberger, Cao, Zazanis and Suri (1987) (and the references therein) for a discussion of the convergence properties of Infinitesimal Perturbation Analysis derivative estimates. Perturbation Analysis can also be applied to sample paths generated from a real system. In this paper, we will only consider likelihood-based derivative estimates.

As described in Reiman and Weiss (1986), the likelihood-based estimator requires that the parameter of the Poisson process be known exactly. In this paper, we modify the likelihood-based estimator for derivatives so that the underlying parameter need not be known a priori. The primary application for this problem is in control and optimization of real time systems. In many situations, gradient information is required for control policies for real systems. For example, the routing policies in Gallager (1977) and Bertsekas, Gafni and Gallager (1984) and the load balancing policies in Kurose and Singh (1986) and Lee (1987) require gradient information in order to obtain optimal performance. In this context, the sequence of arrival times to the system is observable, although, unlike in a discrete event simulation, the mechanism generating arrivals may be unknown or not directly controllable. However, it is often reasonable to assume that the arrival process possesses a particular stochastic structure even though its underlying parameters are unknown. In particular, because of well known limit theorems concerning the superposition of point processes (see, e.g., page 223 of Karlin and Taylor (1975)) a Poisson process may often be assumed to model the arrival process. Within our context, the arrival rate λ of the Poisson process is then unknown.

To make things concrete, let Y_k denote the sum of the waiting times over the k'th busy period in an M/G/1 queue with arrival rate λ, let α_k denote the number of customers served during the busy period and let τ_k denote the length of the k'th busy period. By regenerative process theory, the expected stationary waiting time $E[W]$ is given by the ratio

$$E[W] = \frac{E[Y_k]}{E[\alpha_k]} \tag{1.1}$$

(see, e.g., Çinlar (1975) or Crane and Iglehart (1975)). Consider the problem of estimating $\frac{d}{d\lambda}E[W]$. This requires estimating both $\frac{d}{d\lambda}E[Y_k]$ and $\frac{d}{d\lambda}E[\alpha_k]$. Letting N_k denote the number of arrivals in the k'th busy period, then as shown in Reiman and Weiss (1986), $H_k(\lambda) = (\frac{N_k}{\lambda} - \tau_k)$ is the derivative (with respect to λ) of the logarithm of the likelihood over cycle k and that $\frac{d}{d\lambda}E[Y_k] = E[Y_k(\frac{N_k}{\lambda} - \tau_k)]$. Note that in this example, $\alpha_k = N_k$, but that this need not be the case in more general queueing systems. If n busy periods are observed (or simulated), then $\frac{d}{d\lambda}E[Y_k]$ can be estimated by

$$\hat{y}_n(\lambda) = \frac{\sum_{k=1}^{n} Y_k(\frac{N_k}{\lambda} - \tau_k)}{n}. \tag{1.2}$$

In this paper, we show that if a strongly consistent estimate $\hat{\lambda}_n$ is used instead of λ, then resulting estimator $\hat{y}_n(\hat{\lambda}_n)$ is strongly consistent, asymptotically normal and asymptotically unbiased for $\frac{d}{d\lambda}E[Y_k]$. We further state and prove a central limit theorem for the derivative of a steady state quantity that can be represented by a ratio as in Equation 1.1. This provides a means for placing confidence intervals on the derivative.

While the above discussion was specific to the M/G/1 queue, the results generalize to essentially arbitrary regenerative systems with Poisson arrivals. In the general case, Y_k and α_k are random variables defined on the k'th cycle, τ_k is the length of the k'th cycle and N_k is the number of Poisson arrivals during the k'th cycle. In the general case, for the expression $H_k(\lambda) = (\frac{N_k}{\lambda} - \tau_k)$ to be valid, we assume that the end of a regenerative cycle always coincides with an arrival from the Poisson process (see discussion below).

As discussed in Reiman and Weiss (1986), these results also generalize to estimating the derivative with respect to the rate of an exponential service time in a queueing network; in this case N_k is the number of departures in the k'th cycle and τ_k is the sum of the service times at the particular queue in the k'th cycle (as in the case of the Poisson arrival rate, we assume that the cycles either end with a departure or the server is idle at the end of a cycle). These results are also applicable to estimating the derivative of the expected value of a so-called transient performance measure using the method of independent replications; in this context, k indexes replications rather than regenerative cycles.

The strong law and central limit theorem further generalize to estimating derivatives with respect to an unknown parameter of an input sequence drawn from the exponential class of probability density functions (pdf; see, e.g., Chapter 7 of Hogg and Craig (1970)). The exponential class, which includes many standard distributions, arises in the study of sufficient statistics and maximum likelihood estimation. This generalization is useful for systems with non-exponential arrival or service time processes.

We conducted experiments on an M/M/1 queue to study the effect that the lack of knowledge of λ has on the estimate of $\frac{d}{d\lambda}$ E[W]. We observed that, for small sample sizes, the estimates of $\frac{d}{d\lambda}$ E[W] can be highly biased due to the nonlinearity of the derivative of a ratio. However, there is only a slight increase in the bias when $\hat{\lambda}_n$ is used instead of λ. Furthermore, the bias of both estimators can be substantially reduced by the technique of jackknifing (see, e.g., Miller (1974)). There is also only a slight difference in the variances of the estimators when λ is known as opposed to estimated. Consequently, lack of knowledge of the arrival rate does not significantly affect estimation of the derivative. The primary consequence of this work is that we have shown that the likelihood-based technique can be used to in certain situations to estimate derivatives on a sample path generated by systems in real time.

The remainder of this paper is organized as follows. Section 2 contains the mathematical development including the strong law and asymptotic unbiasedness for the Poisson case and the strong law and central limit theorem for the case of an unknown parameter from the exponential class. The numerical study of the M/M/1 queue is reported in Section 3. Finally, Section 4 summarizes the main results of the paper and indicates where further work is required to generalize the results of this paper.

2. Mathematical Development

Because of its importance in applications, we begin by specializing the results to derivative estimation with respect to the unknown rate of a Poisson process. Treating the Poisson process separately has the advantage of providing the practitioner with explicit formulae and exposing the fundamental concepts. Reiman and Weiss (1986) show that $H_k(\lambda) = (\frac{N_k}{\lambda} - \tau_k)$ is the derivative (with respect to λ) of the logarithm of the likelihood over the k'th regenerative cycle. We assume throughout that the likelihood-based method of derivative estimation is valid, i.e., $E[Y_k H_k(\lambda)] = \frac{d}{d\lambda} E[Y_k]$ (see, e.g., Reiman and Weiss (1986)). Define the estimator of λ to be:

$$\hat{\lambda}_n = \frac{\sum_{k=1}^{n} N_k}{\sum_{k=1}^{n} \tau_k}. \tag{2.1}$$

Notice that $\lim_{n\to\infty} \hat{\lambda}_n = \lambda$ a.s. provided $E[N_k] < \infty$ and $0 < E[\tau_k] < \infty$. Let $\hat{y}_n(\lambda)$ be defined as in Equation 1.2 and define $\hat{y}_n(\hat{\lambda}_n)$ to be the right hand side of Equation 1.2 but with $\hat{\lambda}_n$ replacing λ: $\hat{y}_n(\hat{\lambda}_n)$ is the estimate of $\frac{d}{d\lambda} E[Y_k]$ when the parameter λ is unknown.

Proposition 2.1

Let $\lambda > 0$ and assume that $E[|Y_k|] < \infty$, $0 < E[\tau_k] < \infty$, $E[N_k] < \infty$, $E[|Y_k N_k|] < \infty$, $E[|Y_k \tau_k|] < \infty$. If $E[Y_k H_k(\lambda)] = \frac{d}{d\lambda} E[Y_k]$, then

$$\lim_{n\to\infty} \hat{y}_n(\lambda) = \lim_{n\to\infty} \hat{y}_n(\hat{\lambda}_n) = \frac{d}{d\lambda} E[Y_k] \quad \text{a.s.} \tag{2.2}$$

Proof: The proof for $\hat{y}_n(\lambda)$ follows directly from the strong law of large numbers (see Chapter 4 of Chung (1974)). Consider $\hat{y}_n(\hat{\lambda}_n)$:

$$\begin{aligned} \hat{y}_n(\hat{\lambda}_n) &= \frac{\sum_{k=1}^{n} Y_k N_k}{\hat{\lambda}_n n} - \frac{\sum_{k=1}^{n} Y_k \tau_k}{n} \\ &\to \frac{1}{\lambda} E[Y_k N_k] - E[Y_k \tau_k] = \frac{d}{d\lambda} E[Y_k]. \end{aligned} \tag{2.3}$$

The convergence in Equation 2.3 (which occurs with probability one) is guaranteed by the strong law of large numbers under the stated moment conditions. □

Notice that the convergence in 2.3 is easily obtained because $\hat{\lambda}_n$ enters as only a simple multiplicative factor in $\hat{y}_n(\hat{\lambda}_n)$. As will be seen later, this is characteristic of the exponential class of pdfs.

Proposition 2.2

Assume the conditions of Proposition 2.1 hold. If, in addition, $N_k \geq 1$ a.s., $E[\,|Y_k N_k|^{2+\delta}] < \infty$ and $E[\,|\tau_k|^{2+\delta}] < \infty$ for some $\delta > 0$, then

$$\lim_{n \to \infty} E[\hat{y}_n(\hat{\lambda}_n)] = \frac{d}{d\lambda} E[Y_k]. \tag{2.4}$$

Proof: By Equation 2.3 and the fact that the regenerative cycles are independent and identically distributed (iid),

$$E[\hat{y}_n(\hat{\lambda}_n)] = E[\frac{Y_1 N_1}{\hat{\lambda}_n}] - E[Y_1 \tau_1]. \tag{2.5}$$

Thus we need only show that $\lim_{n \to \infty} E[Y_1 N_1 / \hat{\lambda}_n] = E[Y_1 N_1 / \lambda]$. Since $\lim_{n \to \infty} Y_1 N_1 / \hat{\lambda}_n = Y_1 N_1 / \lambda$ a.s., it suffices to show that the family of random variables $\{Y_1 N_1 / \hat{\lambda}_n\}$ is uniformly integrable (see page 95 of Chung (1974)). Since $N_k \geq 1$, $1/\hat{\lambda}_n \leq \bar{\tau}_n$ where $\bar{\tau}_n = \sum_{k=1}^{n} \tau_k / n$. Therefore $|Y_1 N_1 / \hat{\lambda}_n| \leq D_n \equiv |Y_1 N_1 \bar{\tau}_n|$. Therefore, it suffices to show that $E[D_n^{1+\epsilon}]$ is bounded in n for some $\epsilon > 0$. By Cauchy-Schwartz,

$$E[D_n^{1+\epsilon}] \leq E[\,|Y_1 N_1|^{2+2\epsilon}]^{1/2}\, E[\bar{\tau}_n^{2+2\epsilon}]^{1/2}. \tag{2.6}$$

By the convexity of $h(x) = x^{\beta}$ for $x \geq 0$ and $\beta \geq 1$, $\bar{\tau}_n^{2+2\epsilon} \leq (1/n) \sum_{k=1}^{n} \tau_k^{2+2\epsilon}$ and therefore the right hand side of Inequality 2.6 is finite and bounded under the stated moment assumptions provided $\epsilon \leq \delta/2$.□

The assumption that $N_k \geq 1$ merely states that there is a least one arrival in every regenerative cycle.

We next generalize the estimation context. Glynn and Iglehart (1987) present an expression for the likelihood in an essentially arbitrary discrete event simulation (formally, in a Generalized Semi-Markov Process (GSMP)). Let us fix a parameter θ and let $f(x, \theta)$ denote the pdf associated with one of the "clocks" in a GSMP (a clock may represent an arrival or service process). Assuming the clocks are sampled independently, then the likelihood $L_k(\theta)$ over the k'th cycle can be written as a product of terms

$$L_k(\theta) = \prod_{j=1}^{N_k} f(X_{kj}, \theta)\, L_{k1} \qquad (2.7)$$

where X_{kj} is the j'th value sampled from the clock in cycle k, N_k is the number of times the clock is sampled and L_{k1} does not depend on θ (L_{k1} is itself a product of terms). The representation of Equation 2.7 again assumes that either the clock associated with θ is inactive at the end of a cycle or that the start of a new cycle corresponds to a new setting of the clock. To apply the likelihood-based derivative estimation technique, we need to compute

$$H_k(\theta) = \frac{d}{d\theta} \ln(L_k(\theta)) = \sum_{j=1}^{N_k} \frac{d}{d\theta} \ln(f(X_{kj}, \theta)). \qquad (2.8)$$

We assume that f belongs to the exponential class of pdfs by which we mean (see Chapter 7 of Hogg and Craig (1970))

$$f(x, \theta) = \begin{cases} \exp(\, p(\theta)A(x) + B(x) + q(\theta)\,) & a < x < b \\ 0 & \text{elsewhere} \end{cases} \qquad (2.9)$$

where a and b do not depend on θ, $p(\theta)$ is continuous and both $A'(x) \not\equiv 0$ and $B(x)$ are continuous functions of x for $a < x < b$. An analogous definition exists for a discrete valued random variable. Examples of continuous distributions in this class include the exponential, gamma, Weibull, normal and truncated normal. However, the uniform distribution on (a,b) is not in this family.

When $f(x, \theta)$ has this form, then

$$H_k(\theta) = \frac{d}{d\theta} \ln(L_k(\theta)) = p'(\theta)\, A_k + q'(\theta)\, N_k \qquad (2.10)$$

where $A_k = \sum_{j=1}^{N_k} A(X_{kj})$. For example, in the case of an exponential density $f(x, \theta) = \theta e^{-\theta x}$ so that $p(\theta) = -\theta$, $q(\theta) = \ln(\theta)$ and $A(x) = x$. Therefore, $H_k(\theta) = (\frac{N_k}{\theta} - A_k)$ as described above (in the case of Poisson arrivals $A_k = \tau_k$).

We are now in a position to generalize Proposition 2.1. Analogous to Equation 1.2, define

$$\hat{y}_n(\theta) = \frac{\sum_{k=1}^{n} Y_k H_k(\theta)}{n} \qquad (2.11)$$

and let $\hat{y}_n(\hat{\theta}_n)$ be the right hand side of Equation 2.11 but with the estimate $\hat{\theta}_n$ replacing θ. We assume that the likelihood-based method of derivative estimation is valid, i.e., $E[Y_k H_k(\theta)] = \frac{d}{d\theta} E[Y_k]$. This requires some additional regularity conditions (see Cramér (1946), Reiman and Weiss (1986) or Glynn (1986)), however, these conditions are typically not restrictive in practice.

Proposition 2.3

Assume $H_k(\theta) = p'(\theta)A_k + q'(\theta)N_k$ where $p'(\ )$ and $q'(\ )$ are finite and continuous in a neighborhood of θ. If $\lim_{n\to\infty} \hat{\theta}_n = \theta$ a.s., $E[\,|Y_k A_k|\,] < \infty$, $E[\,|Y_k N_k|\,] < \infty$ and $E[Y_k H_k(\theta)] = \frac{d}{d\theta} E[Y_k]$, then

$$\lim_{n\to\infty} \hat{y}_n(\theta) = \lim_{n\to\infty} \hat{y}_n(\hat{\theta}_n) = \frac{d}{d\theta} E[Y_k] \quad \text{a.s.} \tag{2.12}$$

Proof: Because of the factorization of $H_k(\theta)$ in Equation 2.10,

$$\hat{y}_n(\hat{\theta}_n) = p'(\hat{\theta}_n)\frac{\sum_{k=1}^{n} A_k Y_k}{n} + q'(\hat{\theta}_n)\frac{\sum_{k=1}^{n} N_k Y_k}{n}, \tag{2.13}$$

and the result follows immediately from the strong law of large numbers and the continuity of $p'(\ )$ and $q'(\ )$. □

The analog of Proposition 2.2 concerning the asymptotic unbiasedness of $\hat{y}_n(\hat{\theta}_n)$ could be established under a variety of possible assumptions, although these will not be pursued here.

To estimate the derivative of a ratio as defined in Equation 1.1, let $\bar{Y}_n = \sum_{k=1}^{n} Y_k/n$, $\bar{\alpha}_n = \sum_{k=1}^{n} \alpha_k/n$, $\hat{\alpha}_n(\theta) = \sum_{k=1}^{n} \alpha_k H_k(\theta)/n$

$$\hat{d}_n(\theta) = \frac{\hat{y}_n(\theta)\,\bar{\alpha}_n - \hat{\alpha}_n(\theta)\,\bar{Y}_n}{\bar{\alpha}_n^2} \tag{2.14}$$

and let $\hat{d}_n(\hat{\theta}_n)$ be the right hand side of Equation 2.13 but with $\hat{\theta}_n$ replacing θ.

Proposition 2.4

If $E[\,|Y_k|\,] < \infty$, $E[\,|\alpha_k|\,] < \infty$, $E[\alpha_k] \neq 0$ and if the conditions of Proposition 2.3 hold for both $\hat{y}_n(\theta)$ and $\hat{\alpha}_n(\theta)$, then

$$\lim_{n\to\infty} \hat{d}_n(\theta) = \lim_{n\to\infty} \hat{d}_n(\hat{\theta}_n) = \frac{d}{d\theta} E[W] \quad \text{a.s.} \tag{2.15}$$

We next turn to central limit theorems for $\hat{y}_n(\hat{\theta}_n)$ and $\hat{d}_n(\hat{\theta}_n)$. These limit theorems are direct applications of results in Section 28 of Cramér (1946) concerning the asymptotic normality of a nonlinear function of means. For comparison purpose, we also state central limit theorems for $\hat{y}_n(\theta)$ and $\hat{d}_n(\theta)$, but see Reiman and Weiss (1986) who use somewhat different techniques to establish these results.

For simplicity, assume that $\theta = E[B_k]/E[C_k]$; for the Poisson process $B_k = N_k$ and $C_k = \tau_k$. Let X_k be the 8-dimensional vector

$$X_k = (Y_k, \alpha_k, A_k Y_k, N_k Y_k, A_k \alpha_k, N_k \alpha_k, B_k, C_k), \tag{2.16}$$

$\underset{\sim}{\mu} = E[X_k]$ and let C be the variance-covariance matrix of X_k. Let μ_i and $X_k(i)$ denote the i'th components of $\underset{\sim}{\mu}$ and X_k, respectively, and let $\overline{X}_n = \sum_{k=1}^{n} X_k/n$. In this notation, $\hat{\theta}_n = \sum_{k=1}^{n} B_k / \sum_{k=1}^{n} C_k = \overline{X}_n(7)/\overline{X}_n(8)$. If $E[\,|X_k|\,] < \infty$ and $|C| < \infty$, then

$$\sqrt{n}\left(\overline{X}_n - \underset{\sim}{\mu}\right) \Rightarrow N(0, C) \tag{2.17}$$

where $\Rightarrow$ denotes convergence in distribution (see Billingsley (1968)) and N(0 , C) denotes a multidimensional normally distributed random vector with means zero and variance-covariance matrix C. Using this central limit theorem and Taylor series expansions, Cramér (1946) shows that if, in addition, h is a real valued function such that h and $\frac{\partial h}{\partial \mu_i}$ are finite and continuous in a neighborhood of $\underset{\sim}{\mu}$, then

$$\sqrt{n}\ (h(\overline{X}_n) - h(\underset{\sim}{\mu})) \Rightarrow N(0, \sigma^2) \tag{2.18}$$

where $\sigma^2 = \sum_{ij} h_i C_{ij} h_j$ and $h_i = \frac{\partial h}{\partial \mu_i}\Big|_{\underset{\sim}{\mu}}$ (provided $0 < \sigma^2 < \infty$). The central limit theorems for $\hat{y}_n(\theta)$, $\hat{y}_n(\hat{\theta}_n)$, $\hat{d}_n(\theta)$ and $\hat{d}_n(\hat{\theta}_n)$ all follow as special cases of Equation 2.18.

Proposition 2.5

Assume the conditions of Proposition 2.4 hold. If, in addition, $E[\,|X_k|\,] < \infty$, $|C| < \infty$, $\mu_2 \neq 0$, $\mu_8 \neq 0$ and $p'(\ ), q'(\ ), p''(\ )$ and $q''(\ )$ are finite and continuous in a neighborhood of θ, then

$$\sqrt{n}\left(\hat{y}_n(\theta) - \frac{d}{d\theta}E[Y_k]\right) \Rightarrow N(0, \sigma_1^2), \tag{2.19}$$

$$\sqrt{n}\left(\hat{y}_n(\hat{\theta}_n) - \frac{d}{d\theta}E[Y_k]\right) \Rightarrow N(0, \sigma_2^2), \tag{2.20}$$

$$\sqrt{n}\left(\hat{d}_n(\theta) - \frac{d}{d\theta}\,\mathrm{E}[W]\right) \Rightarrow N(0, \sigma_3^2), \tag{2.21}$$

$$\sqrt{n}\left(\hat{d}_n(\theta) - \frac{d}{d\theta}\,\mathrm{E}[W]\right) \Rightarrow N(0, \sigma_4^2), \tag{2.22}$$

where $\sigma_1^2 = \sum_{i=1}^{6}\sum_{j=1}^{6} a_i C_{ij} a_j$, $\sigma_2^2 = \sum_{i=1}^{8}\sum_{j=1}^{8} a_i C_{ij} a_j$, $\sigma_3^2 = \sum_{i=1}^{6}\sum_{j=1}^{6} b_i C_{ij} b_j$, $\sigma_4^2 = \sum_{i=1}^{8}\sum_{j=1}^{8} b_i C_{ij} b_j$, and the a_i's and b_i's are defined in Table 1 (provided these variance terms are finite).

Proof: For example, since $\hat{y}_n(\hat{\theta}_n) = p'(\overline{X}_n(7)/\overline{X}_n(8))\overline{X}_n(3) + q'(\overline{X}_n(7)/\overline{X}_n(8))\overline{X}_n(4)$, Equation 2.18 can be applied to the function $h(\underset{\sim}{\mu}) = p'(\mu_7/\mu_8)\mu_3 + q'(\mu_7/\mu_8)\mu_4$. The other results follow similarly by appropriate definition and differentiation of the function h. □

Notice the similarity between the variance terms σ_1^2 and σ_2^2. In particular, $\sigma_2^2 = \sigma_1^2$ + the variance due to estimating θ + the covariance between $\hat{\theta}_n$ and $\{\overline{X}_n(i)\ ,\ i = 1, \ldots, 6\}$. The variance terms σ_3^2 and σ_4^2 are similarly related.

For a Poisson process, the central limit theorem for $\hat{y}_n(\hat{\lambda}_n)$ simplifies and can be arrived at more directly. Let $\hat{Z}(\lambda) = Y_k H_k(\lambda) = Y_k(\frac{N_k}{\lambda} - \tau_k)$, $G_k(\lambda) = \lambda H_k(\lambda) = N_k - \lambda\tau_k$, $\sigma_Z^2 = \mathrm{Var}[Z_k(\lambda)]$, $\sigma_G^2 = \mathrm{Var}[G_k(\lambda)]$ and $\sigma_{ZG} = \mathrm{Cov}[Z_k(\lambda), G_k(\lambda)]$. Since $\hat{y}_n(\hat{\lambda}_n) = \hat{y}_n(\lambda) - (\hat{\lambda}_n - \lambda)\sum_{k=1}^{n} Y_k N_k/(n\hat{\lambda}_n\lambda)$, standard weak convergence arguments (see Billingsley (1968)) establish that

$$\begin{aligned} \sqrt{n}\left(\hat{y}_n(\lambda) - \frac{d}{d\lambda}\,\mathrm{E}[Y_k]\right) &\Rightarrow N(0, \sigma_Z^2) \\ \sqrt{n}\left(\hat{y}_n(\hat{\lambda}_n) - \frac{d}{d\lambda}\,\mathrm{E}[Y_k]\right) &\Rightarrow N(0, \sigma_5^2) \end{aligned} \tag{2.23}$$

where $\sigma_5^2 = \sigma_Z^2 - 2a\sigma_{ZG} + a^2\sigma_G^2$ and $a = \mathrm{E}[N_k Y_k]/(\mathrm{E}[\tau_k]\lambda^2)$. It is straightforward to verify that, in this case, 2.23 agrees with 2.19 and 2.20 .

We conclude this section by noting that the central limit theorems remain valid if the variance terms are replaced by strongly consistent estimates. For example, if $\lim_{n\to\infty} \hat{\sigma}_4^2(n) = \sigma_4^2$ a.s., then $\sqrt{n}\,(\hat{d}_n(\hat{\theta}_n) - \frac{d}{d\theta}\,\mathrm{E}[W])/\,\hat{\sigma}_4(n) \Rightarrow N(0,1)$. As discussed in Crane and Iglehart (1975), such variance estimation is straightforward and, furthermore, confidence intervals can be formed from such central limit theorems.

Table 1

Constants for the Central Limit Theorems of Proposition 2.5

a_1	0
a_2	0
a_3	$p'(\theta)$
a_4	$q'(\theta)$
a_5	0
a_6	0
a_7	c_3/μ_8
a_8	$-\mu_7 c_3/\mu_8^2$
b_1	$-c_2/\mu_2^2$
b_2	$2\mu_1 c_2/\mu_2^3 - c_1/\mu_2^2$
b_3	$p'(\theta)/\mu_2$
b_4	$q'(\theta)/\mu_2$
b_5	$-\mu_1 p'(\theta)/\mu_2^2$
b_6	$-\mu_1 q'(\theta)/\mu_2^2$
b_7	$c_3/(\mu_2\mu_8) - \mu_1 c_4/(\mu_2^2\mu_8)$
b_8	$-\mu_7 c_3/(\mu_2\mu_8^2) + \mu_1\mu_7 c_4/(\mu_2^2\mu_8^2)$
c_1	$p'(\theta)\mu_3 + q'(\theta)\mu_4$
c_2	$p'(\theta)\mu_5 + q'(\theta)\mu_6$
c_3	$p''(\theta)\mu_3 + q''(\theta)\mu_4$
c_4	$p''(\theta)\mu_5 + q''(\theta)\mu_6$

3. Experimental Results

In this section we report on simulation experiments when applying the methodology described in Section 2 to a simple queueing system. In particular, we consider estimation of the derivative of the mean stationary waiting time in the M/M/1 queue with Poisson arrival rate λ, service rate μ and traffic intensity $\rho = \lambda/\mu$. The primary performance measures of interest are the bias and variance of estimates of $\frac{d}{d\lambda}$ E[W] for both known and unknown λ.

To study these quantities, we performed experiments as follows. We simulated M iid replications of the queueing system where each replication consisted of a simulation of n regenerative cycles (throughout the paper $M = 10{,}000$). Let $\hat{W}_n(i)$, denote the estimate for E[W] on replication i and let $\hat{d}_n(\lambda, i)$ and $\hat{d}_n(\hat{\lambda}_n, i)$ denote the estimates on replication i for $\frac{d}{d\lambda}$ E[W] using the known and unknown values of λ, respectively. Let $\overline{W}_n$, $\bar{d}_n(\lambda)$ and $\bar{d}_n(\hat{\lambda}_n)$ be the sample averages of $\hat{W}_n(i)$, $\hat{d}_n(\lambda, i)$ and $\hat{d}_n(\hat{\lambda}_n, i)$, respectively, over the M replications, e.g., $\bar{d}_n(\hat{\lambda}_n) = (1/M)\sum_{i=1}^{M} \hat{d}_n(\hat{\lambda}_n, i)$. We estimate the relative bias of an estimator by the sample average minus the corresponding steady state value divided by the steady state value, e.g., $(\bar{d}_n(\hat{\lambda}_n) - \frac{d}{d\lambda}\mathrm{E}[W])/\frac{d}{d\lambda}\mathrm{E}[W]$. If an estimator is unbiased, then this estimate has expectation zero and converges to zero with probability one as the sample size M increases. We also computed the variance of the relative bias in the usual fashion (which is used in Figures 1 and 3).

Figure 1 plots the relative bias in estimates of $\frac{d}{d\lambda}$ E[W] using the known value of λ for different traffic intensities ρ and run lengths n (along with 95% confidence intervals for $\rho = 0.9$). As expected, the absolute amount of bias increases as either the traffic intensity increases or the run length decreases. Figure 2 shows that, for a particular value of ρ, the increase in bias due to estimating λ is very slight. Furthermore, estimates of $\frac{d}{d\lambda}$ E[W] are more highly biased than are estimates of E[W]. As observed elsewhere, e.g., in Reiman and Weiss (1986), Figure 3 shows that the variances of the estimates of $\frac{d}{d\lambda}$ E[W] are significantly higher than the variance of the estimate of E[W]. Again, however, there is little difference between the variances of the derivative estimates when λ is known and when it is estimated, i.e., from Proposition 2.4, most of the asymptotic variance of $\hat{d}_n(\hat{\lambda}_n)$, σ_4^2/n, is due to the contribution of σ_3^2/n, the asymptotic variance of $\hat{d}_n(\lambda)$. In this example, the variance is actually lowered slightly when λ is estimated. The slight increase in variance for small n could be due either to the nonlinear form of the derivative estimate or to an insufficiently large value of M.

Because of the high variance of these derivative estimates, a number of authors (see, e.g., Rubinstein (1986)) have suggested using a variance reduction technique such as control variables. The studies presented here suggest that one should also be concerned about the bias of the derivative estimates, especially in small samples. This bias is due to the nonlinear form of the derivative of a ratio. Typically, the leading term in the bias expansion of such a nonlinear form is given by a constant divided by n (see Sections 27 and 28 of Cramér (1946), Quenouille (1956) or Tin (1965)). Thus, jackknifing (see Miller (1974) for a survey on the jackknife) may be applied to remove this term from the bias expansion. Figure 4 shows that jackknifing significantly reduces the bias of the estimates of $E[W]$ and $\frac{d}{d\lambda} E[W]$. As discussed in Miller (1974), there are also straightforward variance estimation procedures and corresponding central limit theorems based on jackknifing.

Figure 1

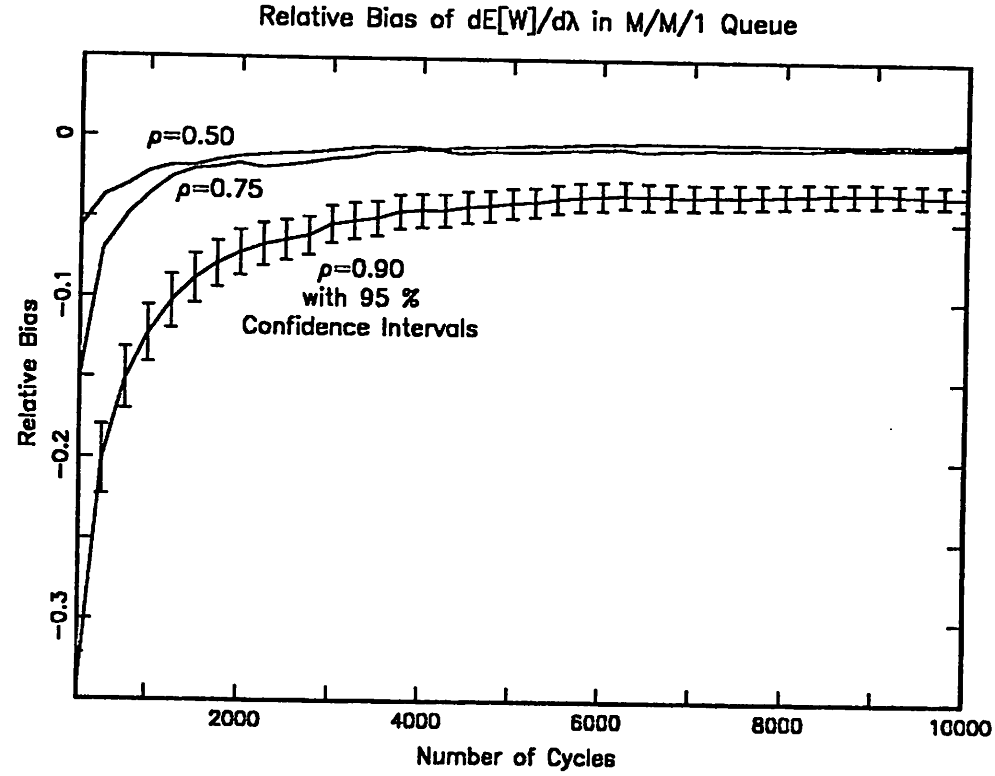

Figure 2

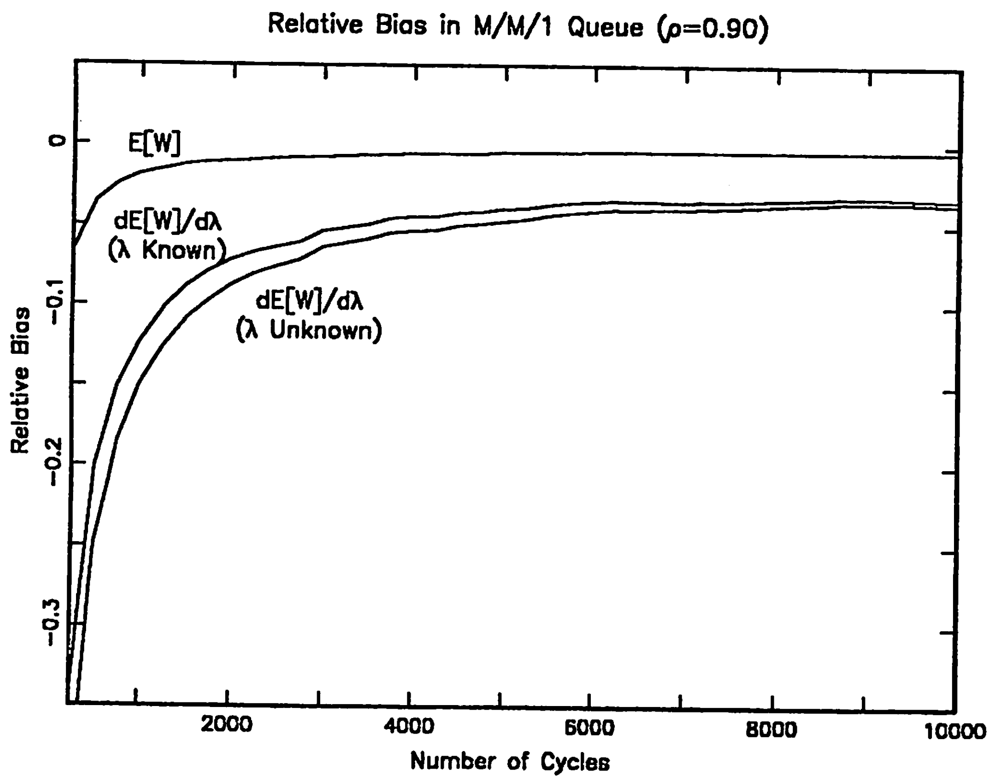

Figure 3

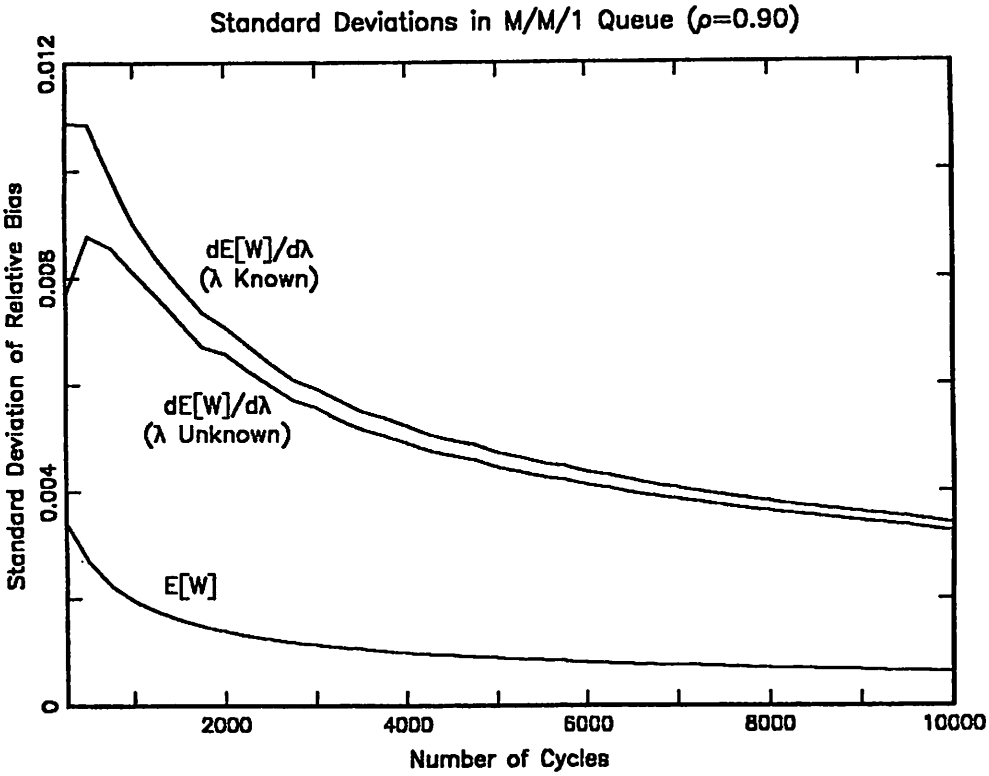

Figure 4