Operations Research
Vol. 58, No. 2, March–April 2010, pp. 357–370
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5802 | 0357

DOI 10.1287/opre.1090.0739

Pathwise Estimation of Probability Sensitivities Through Terminating or Steady-State Simulations

L. Jeff Hong
Department of Industrial Engineering and Logistics Management,
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China, hongl@ust.hk

Guangwu Liu
Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong, China,
msgw.liu@cityu.edu.hk

A probability is the expectation of an indicator function. However, the standard pathwise sensitivity estimation approach, which interchanges the differentiation and expectation, cannot be directly applied because the indicator function is discontinuous. In this paper, we design a pathwise sensitivity estimator for probability functions based on a result of Hong [Hong, L. J. 2009. Estimating quantile sensitivities. *Oper. Res.* **57**(1) 118–130]. We show that the estimator is consistent and follows a central limit theorem for simulation outputs from both terminating and steady-state simulations, and the optimal rate of convergence of the estimator is $n^{-2/5}$ where n is the sample size. We further demonstrate how to use importance sampling to accelerate the rate of convergence of the estimator to $n^{-1/2}$, which is the typical rate of convergence for statistical estimation. We illustrate the performances of our estimators and compare them to other well-known estimators through several examples.

Subject classifications: stochastic simulation; gradient estimation; probability function; perturbation analysis.
Area of review: Simulation.
History: Received October 2007; revisions received June 2008, November 2008, April 2009; accepted April 2009. Published online in *Articles in Advance* November 19, 2009.

1. Introduction

Probabilities are important measures of random performances. They are widely used in practice. In the financial industry, for example, default probabilities are important measures of credit risk. The Black-Scholes-Merton framework models the default probability of a firm with a bond maturing at time T as the probability that the firm's market value is below the face value of the bond at time T (Duffie and Singleton 2003). In the service industries, for example, service quality is often measured by probability that the waiting times of customers are more than a certain service standard.

Suppose that the random performance is a function of some parameters. Then, the probability of the random performance exceeding a certain value is also a function of these parameters. The partial derivatives of the function are called *probability sensitivities*, which provide information on how changes of these parameters affect the output probability. They are useful in practice. If the parameters are decision variables, their sensitivities may be used to control and optimize the probability. For example, the service rates of a queueing system can be adjusted to control the exceedance probability of waiting time. If the parameters are uncontrollable, their sensitivities may be used to assess and hedge the risk. For example, the sensitivities of a default probability with respect to market parameters may help a financial institute to limit the risk exposure of a loan portfolio to the changes of the market conditions.

Simulations are often used to evaluate probabilities when the models of the random performance are complicated. In the example of the default probability, a firm's market value may be modeled as a complicated diffusion process. Then, simulation is often the only way to estimate the default probability. In this example, the trajectories of the diffusion process are simulated and the firm values at the bond maturity are observed. They are then used to estimate the default probability. The simulation observations in this example are independent and identically distributed (i.i.d.). We call this type of simulation *a terminating simulation*. In some other examples, however, simulation observations may be dependent. For instance, to simulate customers' waiting times in a steady-state queueing system, we often observe the waiting times of a sequence of customers entering the system through a single simulation run to avoid a lengthy warm-up period (see, for instance, Law and Kelton 2000). Then, the observations are dependent but (approximately) stationary. We call this type of simulation *a steady-state simulation*. In this paper, we are interested in estimating probability sensitivities using the same simulation observations that are used to estimate probabilities for both terminating and steady-state simulations. Then, we may obtain both the estimate of probability and the estimates of probability sensitivities in a single simulation experiment.

Estimating probability sensitivities has been studied in the literature. Because a probability function is the expectation of an indicator function, estimating its sensitivities may be viewed as a special case of estimating sensitivities of an expectation function, which has been studied extensively in the simulation literature. Readers can refer to L'Ecuyer (1991) and Fu (2006) for comprehensive reviews. The typical methods include the likelihood ratio (LR) method, the weak derivative (WD) method, and perturbation analysis (PA, also known as the pathwise method). The LR method differentiates the input distribution. It is widely applicable if the density function is available. However, the variance of the LR estimator is often large, which significantly degrades its performance. The WD method often seeks to represent the derivative of the density function into the difference of two densities. It is often difficult to choose which WD representation to use, and it often requires many simulations to estimate a derivative (especially when the derivative is taken with respect to a multidimensional vector). Both the LR and WD methods may encounter difficulties if the parameter is not in the input distribution. In the PA family, because the indicator function is discontinuous, infinitesimal perturbation analysis (IPA) cannot be applied directly and smoothed perturbation analysis (SPA) is required. SPA writes the probability function as the expectation of a conditional probability function. If the conditional probability function is smooth and its pathwise derivative can be evaluated easily, the probability sensitivity can be written as the expectation of the pathwise derivative, which can be estimated through simulation. To apply SPA, one needs to decide what to condition on such that the conditional probability function can be evaluated. This is often problem dependent, and may not be easy to determine.

In this paper, we develop an estimator of probability sensitivity based on a result of Hong (2009), who shows that a probability sensitivity (with respect to a parameter of the simulation) can be written as another sensitivity with respect to a value that is not in the simulation. We briefly summarize the assumptions and the result, and discuss the verification of the assumptions in §2. We then develop an estimator based on the result. The estimator does not require the knowledge of the densities. Therefore, it can be applied even when densities are unknown. We prove that the estimator is consistent and follows an asymptotic normal distribution for both terminating and steady-state simulations in §§3 and 4, respectively. However, the estimator has a slower rate of convergence compared to the typical $n^{-1/2}$ of many simulation estimators. When density information is available, we may use importance sampling (IS) to accelerate the rate of convergence of the estimator to $n^{-1/2}$ in some situations. We introduce the IS estimator and discuss its rate of convergence in §5. In §6, we compare our estimators to the SPA and LR estimators through three numerical examples. The paper is concluded in §7, and some lengthy proofs and discussions are in the electronic companion, which is available as part of the online version that can be found at http://or.journal.informs.org/.

2. Background

2.1. Analytical Results of Probability Sensitivity

Let $L(\theta)$ denote the random performance that we are interested in, where θ is the parameter with respect to which we differentiate. In this paper, we assume that θ is one dimensional and $\theta \in \Theta$, where $\Theta \subset \Re$ is an open set. If θ is multidimensional, we may treat each dimension as a one-dimensional parameter while fixing other dimensions constants. Let $p_y(\theta) = \Pr\{L(\theta) \leqslant y\}$. We are interested in estimating $p'_y(\theta) = dp_y(\theta)/d\theta$.

Let $L'(\theta) = dL(\theta)/d\theta$ be the pathwise derivative of $L(\theta)$. If $L(\theta) = l(\theta, X)$ with some function l and random variable (or vector) X, then $L'(\theta) = \partial_\theta l(\theta, X)$ (where ∂ denotes the partial derivative with respect to the subscripted argument). For example, $L(\theta)$ may be the random return of a financial portfolio that has θ share of a stock with an annual return X, i.e., $L(\theta) = \theta X$. Then $L'(\theta) = X$. When $L(\theta)$ cannot be represented by a closed-form function, $L'(\theta)$ may still be evaluated numerically through perturbation analysis (PA) in many situations (Glasserman 1991). For example, $L(\theta)$ may be the sojourn time of a customer in a G/G/1 queue and θ may be the mean service time. Although the closed-form expression of $L(\theta)$ is not available, $L'(\theta)$ can be evaluated through PA. In this paper, we assume that $L'(\theta)$ can be evaluated for all $\theta \in \Theta$.

Assumption 1. *For any $\theta \in \Theta$, $L'(\theta)$ exists with probability* 1 (*w.p.*1) *and there exists a random variable $\mathcal{K}$, which may depend on θ, such that $\mathrm{E}(\mathcal{K}) < \infty$ and $|L(\theta + \Delta\theta) - L(\theta)| \leqslant \mathcal{K}|\Delta\theta|$ for any $\Delta\theta$ that is close enough to* 0.

Assumption 1 is a typical assumption used in pathwise derivative estimation. It guarantees the validity of interchanging differentiation and expectation when evaluating $d\mathrm{E}\{r[L(\theta)]\}/d\theta$ for any Lipschitz-continuous function $r(\cdot)$. Glasserman (1991) develops the commuting conditions for generalized semi-Markov processes under which this assumption holds. Broadie and Glasserman (1996) demonstrate the use of this assumption in estimating price sensitivities of financial derivatives.

Let $F(t, \theta) = \Pr\{L(\theta) \leqslant t\}$ denote the cumulative distribution function of $L(\theta)$. We make the following assumption on the smoothness of $F(t, \theta)$.

Assumption 2. *For any $\theta \in \Theta$, $F(t, \theta)$ is $\mathcal{C}^1$ continuous at (y, θ).*

Assumption 2 basically requires that $F(t, \theta)$ is continuously differentiable at (y, θ). Note that $p_y(\theta) = F(y, \theta)$. It is natural to assume that $F(y, \theta)$ is differentiable in θ because we estimate $p'_y(\theta)$. Furthermore, note that $\partial_t F(t, \theta)$ is the density function of $L(\theta)$. If $L(\theta)$ has a density at $t = y$, then $F(t, \theta)$ is differentiable in t at $t = y$. When a function is differentiable, it is generally continuously differentiable except for some obscure examples (e.g., Example 6.3.4 of Marsden and Hoffman 1993). We believe that Assumption 2 is typically satisfied for practical problems where $L(\theta)$ is a continuous random variable in the neighborhood of y, although its verification can be difficult. In §2.2, we provide more discussions on the verification of Assumption 2.

Given the assumptions, we can prove the following theorem of Hong (2009). Because the assumptions we make are simpler than the ones in Hong (2009), we include our proof of the theorem in the electronic companion.

Theorem 1 (Hong 2009). *Suppose that Assumptions* 1 *and* 2 *are satisfied. Then*

$$p'_y(\theta) = -\partial_y \mathrm{E}[L'(\theta)\cdot 1_{\{L(\theta)\leqslant y\}}]. \tag{1}$$

Hong (2009) further assumes that the conditional expectation $\mathrm{E}[L'(\theta) \mid L(\theta)=t]$ is continuous at $t=y$. Then, by the fundamental theorem of calculus,

$$\begin{aligned} p'_y(\theta) &= -\partial_y \mathrm{E}[L'(\theta)\cdot 1_{\{L(\theta)\leqslant y\}}] \\ &= -\partial_y \int_{-\infty}^{y} \mathrm{E}[L'(\theta)\mid L(\theta)=t] f(t;\theta)\,dt \\ &= -f(y;\theta)\cdot \mathrm{E}[L'(\theta)\mid L(\theta)=y], \end{aligned} \tag{2}$$

where $f(t;\theta)$ is the density function of $L(\theta)$.

In this paper, we directly apply Equation (1) instead of Equation (2) to derive an estimator of $p'_y(\theta)$ for both terminating and steady-state simulations. Therefore, we do not have to assume the continuity of $\mathrm{E}[L'(\theta)\mid L(\theta)=t]$ at $t=y$, which is often difficult to verify for practical problems.

2.2. Verification of Assumption 2

Note that $F(t,\theta)=\Pr\{L(\theta)\leqslant t\} = \mathrm{E}[1_{\{L(\theta)\leqslant t\}}]$. To verify Assumption 2, we need to analyze the differentiability of $\mathrm{E}[1_{\{L(\theta)\leqslant t\}}]$. However, the indicator function $1_{\{\cdot\}}$ is not Lipschitz continuous. Therefore, we cannot interchange the differentiation and expectation if we want to study $\partial_t \mathrm{E}[1_{\{L(\theta)\leqslant t\}}]$ and $\partial_\theta \mathrm{E}[1_{\{L(\theta)\leqslant t\}}]$ (Broadie and Glasserman 1996). To overcome this difficulty, we have to either smooth or remove the indicator function to make the term inside of the expectation Lipschitz continuous. In this subsection, we propose two methods to achieve that.

2.2.1. Conditional Monte Carlo Method. Suppose that there exists a random vector X_θ such that $\Pr\{L(\theta)\leqslant t \mid X_\theta\} = g_\theta(t, X_\theta)$. To simplify the notation, we let $G(t,\theta)=g_\theta(t,X_\theta)$. Note that $G(t,\theta)$ is a random variable given (t,θ). Then,

$$F(t,\theta) = \mathrm{E}[\Pr\{L(\theta)\leqslant t \mid X_\theta\}] = \mathrm{E}[G(t,\theta)]. \tag{3}$$

Therefore, verifying Assumption 2 is equivalent to verifying that $\mathrm{E}[G(t,\theta)]$ is $\mathscr{C}^1$ continuous at (y,θ). In the following lemma, we provide a set of conditions under which it is satisfied. The proof of the lemma is just a straightforward application of Broadie and Glasserman (1996), and it is included in the electronic companion.

Lemma 1. *Suppose that $G(t,\theta)$ is $\mathscr{C}^1$ continuous at (y,θ) w.p.1 and there exists a random variable B, which may depend on (y,θ), such that $\mathrm{E}(B)<\infty$ and*

$$|G(y+\Delta y, \theta+\Delta\theta) - G(y,\theta)| \leqslant B(|\Delta y| + |\Delta\theta|)$$

for any $(\Delta y, \Delta\theta)$ that is close enough to $(0,0)$. Then, $\mathrm{E}[G(t,\theta)]$ is $\mathscr{C}^1$ continuous at (y,θ).

Lemma 1 is closely related to SPA, which uses the relation $p'_y(\theta) = \partial_\theta F(y,\theta) = \mathrm{E}[\partial_\theta G(y,\theta)]$ and directly estimates $\mathrm{E}[\partial_\theta G(y,\theta)]$. To apply SPA, only the differentiability and Lipschitz continuity of $G(y,\theta)$ with respect to θ is required. To verify Assumption 2, however, we require the differentiability and Lipschitz continuity of $G(t,\theta)$ with respect to (t,θ). Therefore, the conditions in Lemma 1 are stronger than the conditions for applying SPA. Nevertheless, for most practical problems, the conditions of Lemma 1 are satisfied if SPA is applicable.

There are also examples where the conditions of Lemma 1 are satisfied, but SPA may not be applied practically. In Lemma 1, we only require that $G(t,\theta)$ satisfy certain properties, we do not need to know how to compute $\partial_\theta G(y,\theta)$. To implement SPA, however, it is critical to know how to compute $\partial_\theta G(y,\theta)$. In §6.3, we provide a portfolio risk example where the conditions of Lemma 1 are satisfied, but $\partial_\theta G(y,\theta)$ cannot be computed practically; thus, SPA cannot be applied practically.

2.2.2. Importance-Sampling Method. Suppose that we may use an importance-sampling distribution to generate $L(\theta)$ such that $L(\theta)\leqslant t$ w.p.1. Let $H(t,\theta)$ denote the likelihood ratio, then

$$\begin{aligned} F(t,\theta) &= \mathrm{E}[1_{\{L(\theta)\leqslant t\}}] = \tilde{\mathrm{E}}[1_{\{L(\theta)\leqslant t\}} H(t,\theta)] \\ &= \tilde{\mathrm{E}}[H(t,\theta)], \end{aligned} \tag{4}$$

where $\tilde{\mathrm{E}}$ denotes the expectation taken with respect to the importance-sampling measure. Then, verifying Assumption 2 is equivalent to verifying that $\tilde{\mathrm{E}}[H(t,\theta)]$ is $\mathscr{C}^1$ continuous at (y,θ). We have the following lemma, which is completely parallel to Lemma 1.

Lemma 2. *Suppose that $H(t,\theta)$ is $\mathscr{C}^1$ continuous at (y,θ) w.p.1, and there exists a random variable B that may depend on (y,θ) such that $\tilde{\mathrm{E}}(B)<\infty$ and*

$$|H(y+\Delta y, \theta+\Delta\theta) - H(y,\theta)| \leqslant B(|\Delta y| + |\Delta\theta|)$$

for any $(\Delta y, \Delta\theta)$ that is close enough to $(0,0)$. Then, $\tilde{\mathrm{E}}[H(t,\theta)]$ is $\mathscr{C}^1$ continuous at (y,θ).

By Lemma 2, we also have $p'_y(\theta) = \partial_\theta F(y,\theta) = \tilde{\mathrm{E}}[\partial_\theta H(y,\theta)]$. If $\partial_\theta H(y,\theta)$ can be computed practically under the importance-sampling measure, we may estimate $p'_y(\theta)$ by directly estimating $\tilde{\mathrm{E}}[\partial_\theta H(y,\theta)]$. There are also examples where the conditions of Lemma 2 are satisfied, but computing $\partial_\theta H(y,\theta)$ requires substantial efforts and is thus not practical to implement. We provide such an example in §6.1.5.

3. Estimating Probability Sensitivity via Terminating Simulations

Because θ is fixed when estimating $p'_y(\theta)$, to simplify the notation, we let L and D denote $L(\theta)$ and $L'(\theta)$, respectively, and let $f(t)$ denote $f(t;\theta)$. Let $h(t) = \mathrm{E}[D\cdot 1_{\{L\leqslant t\}}]$. By Theorem 1, $h(t)$ is differentiable at $t=y$ and $h'(y) = -p'_y(\theta)$. Then,

$$p'_y(\theta) = -h'(y) = -\lim_{\delta\to 0^+}\frac{1}{2\delta}[h(y+\delta)-h(y-\delta)]$$
$$= -\lim_{\delta\to 0^+}\frac{1}{2\delta}\mathrm{E}[D\cdot 1_{\{y-\delta\leqslant L\leqslant y+\delta\}}]. \tag{5}$$

Let $(L_1,D_1),(L_2,D_2),\ldots,(L_n,D_n)$ be the simulation observations of (L,D), and let δ_n, $n=1,2,\ldots,$ be a sequence of positive constants such that $\delta_n\to 0$ and $n\delta_n\to\infty$ as $n\to\infty$. Then we can estimate $p'_y(\theta)$ by

$$\bar{M}_n = -\frac{1}{2n\delta_n}\sum_{i=1}^{n} D_i\cdot 1_{\{y-\delta_n\leqslant L_i\leqslant y+\delta_n\}}. \tag{6}$$

By Equation (5), we can see that $\bar{M}_n$ is essentially a finite difference (FD) estimator. However, it is different from typical FD estimators of $p'_y(\theta)$ that estimate $(1_{\{L(\theta+\Delta\theta)\leqslant y\}} - 1_{\{L(\theta)\leqslant y\}})/\Delta\theta$ or $(1_{\{L(\theta+\Delta\theta)\leqslant y\}} - 1_{\{L(\theta-\Delta\theta)\leqslant y\}})/2\Delta\theta$. First, typical FD estimators do not use pathwise derivatives $L'(\theta)$, but $\bar{M}_n$ does. Intuitively, to evaluate $p'_y(\theta) = \partial_\theta \mathrm{E}[1_{\{L(\theta)-y\leqslant 0\}}]$, one needs to differentiate the indicator function $1_{\{L(\theta)\leqslant y\}}$ with respect to $L(\theta)$, and $L(\theta)$ with respect to θ by the chain rule. Typical FD estimators combine these two differentiations together, whereas $\bar{M}_n$ separates them. It uses an FD method to estimate the first differentiation and uses $L'(\theta)$ directly for the second. Therefore, $\bar{M}_n$ can be viewed as a combination of the FD method and the pathwise method (or IPA). Second, typical FD estimators require simulating at different parameters, e.g., $\theta+\Delta\theta$ and θ. When θ is d-dimensional, for instance, at least $d+1$ simulation runs may be needed to obtain an observation of the estimator. To compute $\bar{M}_n$, however, only one simulation run is needed because y is not a parameter of the simulation model.

The biggest advantage of $\bar{M}_n$ is its simplicity. To compute $\bar{M}_n$, we only need a sample of (L,D) that is often observable from simulation. However, to compute some other estimators, e.g., the SPA or the LR estimators, one has to analyze the structure of the problem and uses distribution information that is not observable from simulation. If the distributions are not available or are difficult to evaluate, these estimators may become difficult to implement.

In the rest of this section, we assume that (L_1,D_1), $(L_2,D_2),\ldots,(L_n,D_n)$ are observations from terminating simulations. Then they are i.i.d. We show that for these observations, $\bar{M}_n$ is a consistent estimator of $p'_y(\theta)$ and follows a central limit theorem under certain conditions.

3.1. Consistency of $\bar{M}_n$

Let $h_\gamma(t) = \mathrm{E}[|D|^\gamma\cdot 1_{\{L\leqslant t\}}]$ for any $\gamma>0$. If $h_\gamma(t)$ is differentiable at $t=y$,[1] then similar to Equation (5),

$$h'_\gamma(y) = \lim_{\delta\to 0^+}\frac{1}{2\delta}\mathrm{E}[|D|^\gamma\cdot 1_{\{y-\delta\leqslant L\leqslant y+\delta\}}]. \tag{7}$$

Because $\mathrm{E}(\bar{M}_n) = -(1/2\delta_n)\mathrm{E}[D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]$, by Equation (5), $\mathrm{E}(\bar{M}_n)\to p'_y(\theta)$ as $n\to\infty$ if $\delta_n\to 0$ as $n\to\infty$. Note that

$$\mathrm{Var}(\bar{M}_n) = \frac{1}{4n\delta_n^2}\mathrm{Var}[D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]$$
$$\leqslant \frac{1}{4n\delta_n^2}\mathrm{E}[D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]. \tag{8}$$

Suppose that $h_2(t)$ is differentiable at $t=y$, then by Equations (7) and (8), $\mathrm{Var}(\bar{M}_n)\to 0$ as $n\to\infty$ if $\delta_n\to 0$ and $n\delta_n\to\infty$ as $n\to\infty$.

By Chebyshev's inequality, for any $\xi>0$, $\Pr\{|\bar{M}_n - \mathrm{E}(\bar{M}_n)|\geqslant\xi\}\leqslant \mathrm{Var}(\bar{M}_n)/\xi^2$. Then $\Pr\{|\bar{M}_n - \mathrm{E}(\bar{M}_n)|\geqslant\xi\}\to 0$, which implies that $\bar{M}_n - \mathrm{E}(\bar{M}_n)\to 0$ in probability as $n\to\infty$. Because $\lim_{n\to\infty}\mathrm{E}(\bar{M}_n) = p'_y(\theta)$, we have $\bar{M}_n\to p'_y(\theta)$ in probability as $n\to\infty$. Therefore, $\bar{M}_n$ is a consistent estimator of $p'_y(\theta)$ as $n\to\infty$. We summarize this result in the following theorem.

THEOREM 2. *Suppose that Assumptions 1 and 2 are satisfied, and $h_2(t)$ is differentiable at $t=y$. If $\delta_n\to 0$ and $n\delta_n\to\infty$ as $n\to\infty$, then $\bar{M}_n\to p'_y(\theta)$ in probability as $n\to\infty$.*

3.2. Asymptotic Normality of $\bar{M}_n$

Let $o(a)$ denote a term such that $\lim_{a\to 0}|o(a)/a| = 0$, and let $O(a)$ denote a term such that $\limsup_{a\to 0}|O(a)/a|<\infty$. To study the asymptotic normality of $\bar{M}_n$, we need a deeper analysis of the asymptotic behaviors of the bias and variance of $\bar{M}_n$. We assume that $h'''(t)$ exists at $t=y$. Note that $h'(y) = -p'_y(\theta)$. Then,

$$\mathrm{E}(\bar{M}_n) = -\frac{1}{2\delta_n}[h(y+\delta_n)-h(y-\delta_n)]$$
$$= p'_y(\theta) - \frac{1}{6}h'''(y)\delta_n^2 + o(\delta_n^2), \tag{9}$$

$$\mathrm{Var}(\bar{M}_n) = \frac{1}{4n\delta_n^2}\{\mathrm{E}[D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}] - \mathrm{E}^2[D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]\}$$
$$= \frac{1}{4n\delta_n^2}\{[h_2(y+\delta_n)-h_2(y-\delta_n)] - [h(y+\delta_n)-h(y-\delta_n)]^2\}$$
$$= \frac{1}{2n\delta_n}\{h'_2(y)+o(1)\}. \tag{10}$$

Therefore, $\delta_n^{-2}[\mathrm{E}(\bar{M}_n) - p'_y(\theta)]\to -h'''(y)/6$ and $2n\delta_n\,\mathrm{Var}(\bar{M}_n)\to h'_2(y)$ as $n\to\infty$.

Let

$$R_{n,i} = -\frac{1}{\sqrt{2n\delta_n}}[D_i \cdot 1_{\{y-\delta_n \leqslant L_i \leqslant y+\delta_n\}} - \mathrm{E}(D \cdot 1_{\{y-\delta_n \leqslant L \leqslant y+\delta_n\}})]. \quad (11)$$

Then, $\sum_{i=1}^n R_{n,i} = \sqrt{2n\delta_n}[\bar{M}_n - \mathrm{E}(\bar{M}_n)]$ and $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n R_{n,i}) = 2n\delta_n \mathrm{Var}(\bar{M}_n) = h_2'(y) + o(1)$. In the next lemma, we show that $R_{n,i}$ satisfies the Lindeberg condition (Billingsley 1995). The proof of the lemma is provided in the electronic companion.

Lemma 3. *Suppose that Assumptions* 1 *and* 2 *are satisfied,* $h_2(t)$ *and* $h_{2+\gamma}(t)$ *are differentiable at* $t = y$ *for some* $\gamma > 0$, *and* $h_2'(y) > 0$. *If* $\delta_n \to 0$ *and* $n\delta_n \to \infty$ *as* $n \to \infty$, *then for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^n \mathrm{E}[R_{n,i}^2 \cdot 1_{\{|R_{n,i}| \geqslant \epsilon\sigma_n\}}] = 0.$$

Because $R_{n,i}$ satisfies the Lindeberg condition, by the Lindeberg central limit theorem (Billingsley 1995) we have $\sigma_n^{-1} \cdot \sum_{i=1}^n R_{n,i} \Rightarrow N(0,1)$, where $\Rightarrow$ denotes "converge in distribution" and $N(0,1)$ denotes the standard normal random variable. Becasue $\sum_{i=1}^n R_{n,i} = \sqrt{2n\delta_n}[\bar{M}_n - \mathrm{E}(\bar{M}_n)]$ and $\lim_{n\to\infty} \sigma_n^2 = h_2'(y)$, then $\sqrt{2n\delta_n}[\bar{M}_n - \mathrm{E}(\bar{M}_n)] \Rightarrow \sqrt{h_2'(y)} \cdot N(0,1)$. Because

$$\sqrt{2n\delta_n}[\bar{M}_n - p_y'(\theta)] = \sqrt{2n\delta_n}[\bar{M}_n - \mathrm{E}(\bar{M}_n)] + \sqrt{2n\delta_n}[\mathrm{E}(\bar{M}_n) - p_y'(\theta)],$$

by Equation (9) we can easily prove the following theorem on the asymptotic normality of $\bar{M}_n$.

Theorem 3. *Suppose that Assumptions* 1 *and* 2 *are satisfied,* $h_2(t)$ *and* $h_{2+\gamma}(t)$ *are differentiable at* $t = y$ *for some* $\gamma > 0$, $h'''(t)$ *exists at* $t = y$, *and* $h_2'(y) > 0$. *If* $n\delta_n^5 \to a$ *as* $n \to \infty$, *then*

$$\sqrt{2n\delta_n}[\bar{M}_n - p_y'(\theta)] \Rightarrow -\frac{\sqrt{2a}}{6} h'''(y) + \sqrt{h_2'(y)} \cdot N(0,1) \quad \text{as } n \to \infty.$$

Remark 1. Theorem 3 assumes that $h_2'(y) > 0$, which is equivalent to $\lim_{n\to\infty} 2n\delta_n \mathrm{Var}(\bar{M}_n) > 0$. If $h_2'(y) = 0$, then $\lim_{n\to\infty} 2n\delta_n \mathrm{Var}(\bar{M}_n) = 0$. By Chebyshev's inequality, one can easily prove that $\sqrt{2n\delta_n}[\bar{M}_n - p_y'(\theta)] \to -(\sqrt{2a}/6)h'''(y)$ in probability. Hence, the conclusion of Theorem 3 also holds.

Theorem 3 shows that the rate of convergence of $\bar{M}_n$ is $(n\delta_n)^{-1/2}$. It is $n^{-2/5}$ when $a > 0$, and it is slower than $n^{-2/5}$ when $a = 0$. When $a > 0$, however, the asymptotic normal distribution has a nonzero mean. Because the mean is typically difficult to estimate, confidence intervals of $p_y'(\theta)$ may be difficult to construct. When $a = 0$, the asymptotic normal distribution has a zero mean, and asymptotically valid confidence intervals can be constructed.

Let $\sigma_\infty^2 = h_2'(y)$. Because $p_y'(\theta) = -h'(y)$, then we may use the same approach that estimates $p_y'(\theta)$ to estimate σ_∞^2. Let $\bar{V}_n^2 = (1/2n\delta_n)\sum_{i=1}^n D_i^2 \cdot 1_{\{y-\delta_n \leqslant L_i \leqslant y+\delta_n\}}$. If we assume that $h_4(t)$ is differentiable at $t = y$, we can easily show that $\bar{V}_n^2$ is a consistent estimator of σ_∞^2 by the same techniques used in the proof of Theorem 2. Suppose that $a = 0$, i.e., $n\delta_n^5 \to 0$ as $n \to \infty$. Then an asymptotically valid $100(1-\beta)\%$ confidence interval of $p_y'(\theta)$ is

$$\left(\bar{M}_n - z_{1-\beta/2}\bar{V}_n/\sqrt{2n\delta_n},\ \bar{M}_n + z_{1-\beta/2}\bar{V}_n/\sqrt{2n\delta_n}\right), \quad (12)$$

where $z_{1-\beta/2}$ is the $1-\beta/2$ quantile of the standard normal distribution.

4. Estimating Probability Sensitivity via Steady-State Simulations

In this section we assume that $(L_1, D_1), (L_2, D_2), \ldots, (L_n, D_n)$ are observations from a steady-state simulation.[2] Therefore, they are typically stationary and dependent. We show that under certain conditions on (L_i, D_i), the estimator $\bar{M}_n$ of Equation (6) is still consistent and follows a central limit theorem.

Suppose that $\{(L_i, D_i),\ i = 1, 2, \ldots\}$ is a stationary sequence. Let $\mathcal{F}_k$ be the σ-algebra generated by $\{(L_i, D_i),\ i = 1, 2, \ldots, k\}$ and $\mathcal{G}_k$ be the σ-algebra generated by $\{(L_i, D_i),\ i = k, k+1, \ldots\}$. Following Billingsley (1968), let $\phi(k) = \sup\{|\Pr(B \mid A) - \Pr(B)|\colon A \in \mathcal{F}_s, \Pr(A) > 0, B \in \mathcal{G}_{s+k}\}$. Then the sequence is ϕ-mixing if $\phi(k) \to 0$ as $k \to \infty$. Intuitively, the condition means that the dependence between the future and the present of a ϕ-mixing process goes to zero as the time between them goes to infinity. Many stochastic processes are ϕ-mixing. For instance, m-dependent processes and stationary Markov processes with finite state space (Billingsley 1968) are ϕ-mixing, and positive recurrent regenerative processes are also ϕ-mixing (Glynn and Iglehart 1985). In this section, we make the following assumption on $\{(L_i, D_i),\ i = 1, 2, \ldots\}$.

Assumption 3. *The sequence* $\{(L_i, D_i),\ i = 1, 2, \ldots\}$ *satisfies that* $\sum_{k=1}^\infty \sqrt{\phi(k)} < \infty$.

Note that Assumption 3 implies that the sequence is ϕ-mixing. Schruben (1983) argues that stationary finite-state[3] discrete-event simulations can be described as a finite-state, aperiodic, and irreducible Markov process with $\phi(k) = a\rho^k$ for some $\rho < 1$. Then the assumption holds. Assumption 3 has been widely used to study steady-state behaviors of discrete-event simulations. For example, Schruben (1983) makes this assumption to study the estimators of steady-state means, Heidelberger and Lewis (1984) use this assumption to analyze the estimators of steady-state quantiles, and Chien et al. (1997) make a stronger assumption to study asymptotic properties of the batch means method.

Define $Z_{n,i} = D_i \cdot 1_{\{y-\delta_n \leqslant L_i \leqslant y+\delta_n\}}$. Let $\mathcal{F}_{n,k}^Z$ be the σ-algebra generated by $\{Z_{n,i},\ i = 1, 2, \ldots, k\}$ and $\mathcal{G}_{n,k}^Z$ be the σ-algebra generated by $\{Z_{n,i},\ i=k,k+1,\ldots,n\}$ for any $1\leqslant k\leqslant n$, and let

$$\phi_n^Z(k)=\sup\{|\Pr(B\mid A)-\Pr(B)|\colon A\in\mathcal{F}_{n,s}^Z,\ \Pr(A)>0,\ B\in\mathcal{G}_{n,s+k}^Z\}.$$

Because $Z_{n,i}$ is a Borel-measurable function of (L_i,D_i), then $\mathcal{F}_{n,k}^Z$ and $\mathcal{G}_{n,k}^Z$ are subsets of $\mathcal{F}_k$ and $\mathcal{G}_k$, respectively. Then, by the definitions of $\phi_n^Z(k)$ and $\phi(k)$, we have

$$\phi_n^Z(k)\leqslant\phi(k) \tag{13}$$

for all $k=1,2,\ldots,n$ and all $n=1,2,\ldots$.

Note that

$$\begin{aligned}\operatorname{Var}(\bar M_n)&=\operatorname{Var}\left(\frac{1}{2n\delta_n}\sum_{i=1}^n Z_{n,i}\right)\\&=\frac{1}{4n\delta_n^2}\left[\operatorname{Var}(Z_{n,1})+2\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\operatorname{Cov}(Z_{n,1},Z_{n,i+1})\right].\end{aligned} \tag{14}$$

By the covariance inequality (Billingsley 1968, p. 170) and Equation (13),

$$|\operatorname{Cov}(Z_{n,1},Z_{n,i+1})|\leqslant 2\sqrt{\phi_n^Z(i)}\operatorname{E}(Z_{n,1}^2)\leqslant 2\sqrt{\phi(i)}\operatorname{E}(Z_{n,1}^2).$$

Because $\operatorname{E}(Z_{n,1}^2)=\operatorname{E}[D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]$, by Equation (7), $\lim_{n\to\infty}(1/2\delta_n)\operatorname{E}(Z_{n,1}^2)=h_2'(y)$. Then

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{2\delta_n}\sum_{i=1}^n\left|\left(1-\frac{i}{n}\right)\operatorname{Cov}(Z_{n,1},Z_{n,i+1})\right|\\&\quad\leqslant\lim_{n\to\infty}\frac{1}{2\delta_n}\sum_{i=1}^n\left|\operatorname{Cov}(Z_{n,1},Z_{n,i+1})\right|=2h_2'(y)\sum_{i=1}^\infty\sqrt{\phi(i)}<\infty\end{aligned}$$

by Assumption 3. Therefore, $1/2\delta_n\sum_{i=1}^n(1-i/n)\cdot\operatorname{Cov}(Z_{n,1},Z_{n,i+1})$ converges as $n\to\infty$ by the ratio comparison test (Marsden and Hoffman 1993). Furthermore, by Equation (10),

$$\lim_{n\to\infty}\frac{1}{2\delta_n}\operatorname{Var}(Z_{n,1})=\lim_{n\to\infty}\frac{1}{2\delta_n}\operatorname{Var}[D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]=h_2'(y). \tag{15}$$

Therefore, by Equation (14), there exists a limit that we denote as σ_∞^2, such that

$$\lim_{n\to\infty}2n\delta_n\operatorname{Var}(\bar M_n)=\sigma_\infty^2. \tag{16}$$

To further understand σ_∞^2, by Equation (14), we write

$$2n\delta_n\operatorname{Var}(\bar M_n)=\frac{1}{2\delta_n}\operatorname{Var}(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})\left[1+2\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\rho_{n,i}\right],$$

where $\rho_{n,i}=\operatorname{Cov}(Z_{n,1},Z_{n,i+1})/\operatorname{Var}(Z_{n,1})$. Let $\alpha_n=1+2\sum_{i=1}^{n-1}(1-i/n)\rho_{n,i}$. By Equations (15) and (16), there exists a nonnegative limit, which we denote as α_∞, such that $\lim_{n\to\infty}\alpha_n=\alpha_\infty$ and $\lim_{n\to\infty}2n\operatorname{Var}(\bar M_n)=\alpha_\infty h_2'(y)$ if $h_2'(y)>0$. Therefore, the dependence in the ϕ-mixing sequence inflates the asymptotic variance of $\bar M_n$ by a factor of α_∞ compared to the one of the i.i.d. sequence.

4.1. Consistency of $\bar M_n$

Note that $\operatorname{E}(\bar M_n)=-(1/2\delta_n)\operatorname{E}[D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}]$. Then, by Equation (5), $\operatorname{E}(\bar M_n)\to p_y'(\theta)$ as $n\to\infty$. By Equation (16), $\operatorname{Var}(\bar M_n)\to 0$ as $n\to\infty$. By Chebyshev's inequality, we can easily prove that $\bar M_n$ is a consistent estimator of $p_y'(\theta)$. We summarize the result in the following theorem.

THEOREM 4. *Suppose that Assumptions* 1 *to* 3 *are satisfied, and $h_2(t)$ is differentiable at $t=y$. If $\delta_n\to 0$ and $n\delta_n\to\infty$ as $n\to\infty$, then $\bar M_n\to p_y'(\theta)$ in probability as $n\to\infty$.*

4.2. Asymptotic Normality of $\bar M_n$

To establish the asymptotic normality of $\bar M_n$, we need the following lemma of Utev (1990).

LEMMA 4 (UTEV 1990). *Let $\xi_{n,1},\ldots,\xi_{n,k_n}$, $n=1,2,\ldots$, be a triangular array of random variables with zero mean and finite variances, and let $\mathcal{F}_{n,k}^\xi$ be the σ-algebra generated by $\{\xi_{n,i},\ i=1,2,\ldots,k\}$ and $\mathcal{G}_{n,k}^\xi$ be the σ-algebra generated by $\{\xi_{n,i},\ i=k,k+1,\ldots,k_n\}$ for any $1\leqslant k\leqslant k_n$. Moreover, let $\phi_n^\xi(k)=\sup\{|\Pr(B\mid A)-\Pr(B)|\colon A\in\mathcal{F}_{n,s}^\xi,\Pr(A)>0,B\in\mathcal{G}_{n,s+k}^\xi\}$ for $k=1,2,\ldots,$ and $\sigma_n^2=\operatorname{Var}[\sum_{i=1}^{k_n}\xi_{n,i}]$. Suppose that there exists a sequence of natural numbers $j_1,j_2,\ldots$ such that $\sup_n\phi_n^\xi(kj_n)\to 0$ as $k\to\infty$ and $j_n\sigma_n^{-2}\cdot\sum_{i=1}^{k_n}\operatorname{E}[\xi_{n,i}^2\cdot 1_{\{|\xi_{n,i}|\geqslant\epsilon\sigma_n/j_n\}}]\to 0$ as $n\to\infty$ for any $\epsilon>0$. Then $\sigma_n^{-1}\cdot\sum_{i=1}^{k_n}\xi_{n,i}\Rightarrow N(0,1)$ as $n\to\infty$.*

We define $R_{n,i}$ as in Equation (11). Then, $\sum_{i=1}^n R_{n,i}=\sqrt{2n\delta_n}[\bar M_n-\operatorname{E}(\bar M_n)]$. Let $\mathcal{F}_{n,k}^R$ be the σ-algebra generated by $\{R_{n,i},\ i=1,2,\ldots,k\}$ and $\mathcal{G}_{n,k}^R$ be the σ-algebra generated by $\{R_{n,i},\ i=k,k+1,\ldots,n\}$ for any $1\leqslant k\leqslant n$. Moreover, let $\sigma_n^2=\operatorname{Var}[\sum_{i=1}^n R_{n,i}]$ and

$$\phi_n^R(k)=\sup\{|\Pr(B\mid A)-\Pr(B)|\colon A\in\mathcal{F}_{n,s}^R,\ \Pr(A)>0,B\in\mathcal{G}_{n,s+k}^R\},\quad k=1,2,\ldots.$$

Because $R_{n,i}$ is a Borel-measurable function of (L_i,D_i), then similar to Equation (13), we can show that $\phi_n^R(k)\leqslant\phi(k)$ for any n and k. By Assumption 3, $\phi(k)\to 0$ as $k\to\infty$. Then,

$$\sup_n\phi_n^R(k)\to 0\quad\text{as }k\to\infty. \tag{17}$$

Because $\sigma_n^2=2n\delta_n\operatorname{Var}(\bar M_n)$, by Equation (16), $\lim_{n\to\infty}\sigma_n^2=\sigma_\infty^2$. Then, with the same proof of Lemma 3, we can show that when $\sigma_\infty^2>0$,

$$\lim_{n\to\infty}\frac{1}{\sigma_n^2}\sum_{i=1}^n\operatorname{E}[R_{n,i}^2\cdot 1_{\{|R_{n,i}|\geqslant\epsilon\sigma_n\}}]=0 \tag{18}$$

for any $\epsilon>0$, because only σ_n^2 in Equation (18) is affected by the dependence in the ϕ-mixing sequence, and it converges to a positive constant as in the i.i.d. case.

Combining Equations (17) and (18), by Lemma 4, we have $\sigma_n^{-1}\cdot\sum_{i=1}^n R_{n,i}\Rightarrow N(0,1)$ when we let $k_n=n$ and $j_n=1$ for all n in Lemma 4. Then, with the same analysis in §3.2, we have the following theorem on the asymptotic normality of $\bar M_n$ for dependent sequences.

Theorem 5. *Suppose that Assumptions 1 to 3 are satisfied, $h_2(t)$ and $h_{2+\gamma}(t)$ are differentiable at $t=y$ for some $\gamma>0$, $h'''(t)$ exists at $t=y$, and $\sigma_\infty^2>0$. If $n\delta_n^5\to a$ as $n\to\infty$,*

$$\sqrt{2n\delta_n}[\bar{M}_n - p_y'(\theta)] \Rightarrow -\frac{\sqrt{2a}}{6}h'''(y)+\sigma_\infty\cdot N(0,1) \quad \text{as } n\to\infty.$$

Remark 2. Similar to the remark of Theorem 3, the conclusion of Theorem 5 also holds if $\sigma_\infty^2=0$.

Theorem 5 shows that the rate of convergence of $\bar{M}_n$ for dependent sequences is the same as that for i.i.d. sequences. The only difference is that the asymptotic variance for dependent sequences is inflated by a factor of α_∞.

To construct an asymptotically valid confidence interval of $p_y'(\theta)$ using a dependent sequence $\{(L_i,D_i),\ i=1,2,\ldots\}$, we need to set $a=0$ in Theorem 5 to avoid the estimation of $(\sqrt{2a}/6)h'''(y)$. Furthermore, we need an approach to consistently estimating $\sigma_\infty^2=\alpha_\infty h_2'(y)$. Because α_∞ is unknown and difficult to estimate, σ_∞^2 for dependent sequences is more difficult to estimate than that for i.i.d. sequences.

We suggest using the batch means method to estimate σ_∞^2. We divide the n observations of (L_i,D_i) into k_n adjacent batches, and each batch has m_n observations. We require that both $m_n\to\infty$ and $k_n\to\infty$ as $n\to\infty$. For example, a reasonable choice may be $m_n=k_n=\sqrt{n}$. Let

$$\bar{M}_{m_n}^{(j)} = -\frac{1}{2m_n\delta_{m_n}}\sum_{i=1}^{m_n} D_{(j-1)m_n+i}\cdot 1_{\{y-\delta_{m_n}\leqslant L_{(j-1)m_n+i}\leqslant y+\delta_{m_n}\}},$$

for $j=1,\ldots,k_n$. Then the variance estimator $\bar{V}_n^2$ can be expressed as

$$\bar{V}_n^2 = \frac{2m_n\delta_{m_n}}{k_n-1}\sum_{j=1}^{k_n}\left[\bar{M}_{m_n}^{(j)} - \frac{1}{k_n}\sum_{j=1}^{k_n}\bar{M}_{m_n}^{(j)}\right]^2 = \frac{2m_n\delta_{m_n}}{k_n-1}\sum_{j=1}^{k_n}[\bar{M}_{m_n}^{(j)}-\bar{A}_n]^2, \tag{19}$$

where $\bar{A}_n = k_n^{-1}\cdot\sum_{j=1}^{k_n}\bar{M}_{m_n}^{(j)}$. Note that $\bar{V}_n^2/2m_n\delta_{m_n}$ is the sample variance of $\bar{M}_{m_n}^{(1)},\bar{M}_{m_n}^{(2)},\ldots,\bar{M}_{m_n}^{(k_n)}$, and $\bar{M}_{m_n}^{(i)}$ and $\bar{M}_{m_n}^{(j)}$ are approximately independent when m_n is large. Then $\bar{V}_n^2/2m_n\delta_{m_n}$ is a reasonable estimator of $\mathrm{Var}(\bar{M}_{m_n}^{(1)})$. From Equation (16), we know that $2m_n\delta_{m_n}\mathrm{Var}(\bar{M}_{m_n}^{(1)})\to\sigma_\infty^2$ as $n\to\infty$. Therefore, $\bar{V}_n^2$ is a reasonable estimator of σ_∞^2. In the electronic companion, we prove that $\bar{V}_n^2$ is a consistent estimator of σ_∞^2 under some additional technical conditions.

Suppose that $a=0$, i.e., $n\delta_n^5\to 0$ as $n\to\infty$. Then, by Theorem 5, an asymptotically valid $100(1-\beta)\%$ confidence interval of $p_y'(\theta)$, when the observations $(L_1,D_1),(L_2,D_2),\ldots,(L_n,D_n)$ are dependent, is

$$\left(\bar{M}_n - z_{1-\beta/2}\bar{V}_n/\sqrt{2n\delta_n},\ \bar{M}_n + z_{1-\beta/2}\bar{V}_n/\sqrt{2n\delta_n}\right). \tag{20}$$

5. Accelerating the Rate of Convergence

From §§3 and 4, we see that the rate of convergence of $\bar{M}_n$ is $(n\delta_n)^{-1/2}$, which is slower than the typical $n^{-1/2}$. Recall that $\bar{M}_n=-(1/2n\delta_n)\cdot\sum_{i=1}^n D_i\cdot 1_{\{y-\delta_n\leqslant L_i\leqslant y+\delta_n\}}$. Intuitively, the reason that the rate of convergence is $(n\delta_n)^{-1/2}$ is because only the samples in the important region $\{y-\delta_n\leqslant L\leqslant y+\delta_n\}$ play roles in the estimator, and the total number of such samples is of order $n\delta_n$. Therefore, a natural idea to accelerate the rate of convergence of the estimator is to use importance sampling (IS) to have all the samples falling into the important region. In this section, we show how to use IS to accelerate the rate of convergence of the estimator. For simplicity, the analysis in this section is based on i.i.d. sequences, although the method can also be applied to dependent sequences.

5.1. A Simple Situation

We first consider the situation where $p_n=\Pr\{y-\delta_n\leqslant L\leqslant y+\delta_n\}$ is known. We show that a simple IS scheme can always reduce the variance of the estimator and accelerate the rate of convergence to $n^{-1/2}$. Because p_n is rarely known in practice, this only provides a preliminary analysis. However, this helps to understand the basics of the IS scheme. The situation where p_n is not known is considered in §5.2.

Let $f(t)$ denote the density of L. Let $\tilde{f}(t)=f(t)/p_n$ for $t\in[y-\delta_n,y+\delta_n]$, and $\tilde{f}(t)=0$ otherwise. We let $\tilde{f}$ be the IS distribution of L. Samples generated from the IS distribution are always in $[y-\delta_n,y+\delta_n]$. Because f and $\tilde{f}$ are mutually absolutely continuous in the region $[y-\delta_n, y+\delta_n]$, the likelihood ratio $f(t)/\tilde{f}(t)=p_n$ in the region. Then, the IS estimator of the probability sensitivity is

$$\bar{M}_n^{IS} = -\frac{1}{2n\delta_n}\sum_{i=1}^n p_n\cdot D_i\cdot 1_{\{y-\delta_n\leqslant L_i\leqslant y+\delta_n\}} = -\frac{1}{2n\delta_n}\sum_{i=1}^n p_n\cdot D_i,$$

where the observations (L_i,D_i) are generated under the IS distribution.

Let $\tilde{\mathrm{E}}$ and $\widetilde{\mathrm{Var}}$ denote the expectation and variance under the IS distribution. Then,

$$\tilde{\mathrm{E}}(\bar{M}_n^{IS}) = -\frac{1}{2\delta_n}\tilde{\mathrm{E}}(p_n\cdot D) = -\frac{1}{2\delta_n}\tilde{\mathrm{E}}[p_n\cdot D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}] = \mathrm{E}(\bar{M}_n),$$

$$\begin{aligned}\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) &= \frac{1}{4n\delta_n^2}[p_n\tilde{\mathrm{E}}(p_n\cdot D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) - \tilde{\mathrm{E}}^2(p_n\cdot D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})] \\ &= \frac{1}{4n\delta_n^2}[p_n\mathrm{E}(D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) - \mathrm{E}^2(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})].\end{aligned} \tag{21}$$

Because $p_n\leqslant 1$,

$$\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) \leqslant \frac{1}{4n\delta_n^2}[\mathrm{E}(D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) - \mathrm{E}^2(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})] = \mathrm{Var}(\bar{M}_n).$$

Therefore, the IS estimator $\bar{M}_n^{IS}$ has a smaller variance than the original estimator $\bar{M}_n$ when δ_n is the same for the both estimators.

Furthermore, $p_n = \Pr\{y-\delta_n \leqslant L \leqslant y+\delta_n\} = 2f(y)\delta_n + o(\delta_n)$. Then, by Equation (21),

$$\begin{aligned}\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) &= \frac{1}{4n\delta_n^2}\{[2f(y)\delta_n + o(\delta_n)]\cdot[2h_2'(y)\delta_n + o(\delta_n)] \\ &\qquad - [2h'(y)\delta_n + o(\delta_n)]^2\} \\ &= \frac{1}{n}[f(y)h_2'(y) - [h'(y)]^2 + o(1)].\end{aligned}$$

Hence, $n\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) \to f(y)h_2'(y) - [h'(y)]^2$ as $n\to\infty$. Therefore, the variance of $\bar{M}_n^{IS}$ is of $O(n^{-1})$. Note that the bias of $\bar{M}_n^{IS}$ is of $O(\delta_n^2)$. Then the rate of convergence of $\bar{M}_n^{IS}$ is $n^{-1/2}$ if $n\delta_n^4 \to a$ with $a \geqslant 0$ as $n\to\infty$.

From this simple situation, we see that IS can indeed accelerate the rate of convergence of the sensitivity estimator. In the next subsection, we show how to extend this approach to a more practical situation.

5.2. A More Practical Situation

Because the density of L is typically unknown, then p_n is also unknown. In this subsection, we consider a more practical situation. Suppose that $L = L(X_1, X_2, \ldots, X_k)$ and $D = D(X_1, X_2, \ldots, X_k)$, where $X_1, X_2, \ldots, X_k$ are independent random variables with known densities $f_1, f_2, \ldots, f_k$, respectively. Then the joint density of $(X_1, X_2, \ldots, X_k)$ is $f(x_1, x_2, \ldots, x_k) = f_1(x_1)f_2(x_2)\cdots f_k(x_k)$. In many examples, the simulation output is generated by a sequence of independent random variables. Then our situation applies.

Suppose that $X_1, X_2, \ldots, X_k$ are generated sequentially in the simulation to obtain observations of (L, D). Let

$$\begin{aligned}A = \{&(x_1,x_2,\ldots,x_k)\in\Re^k\colon L(x_1,x_2,\ldots,x_k)\in[y-\delta_n, y+\delta_n] \\ &\text{and } f_1(x_1)\cdots f_k(x_k) > 0\}.\end{aligned}$$

Then A corresponds to the set $\{y-\delta_n \leqslant L \leqslant y+\delta_n\}$ in the simple situation of §5.1. It is the important region. We assume that $\Pr\{A\} > 0$. Let A_1 be the projection of the set A to the first dimension, $A_2(X_1)$ be the projection of the set A to the second dimension given X_1, $A_3(X_1, X_2)$ be the projection of the set A to the third dimension given $X_1, X_2, \ldots,$ and $A_k(X_1, X_2, \ldots, X_{k-1})$ be the projection of the set A to the kth dimension given $X_1, X_2, \ldots, X_{k-1}$. Then, we may define the IS distribution as

$$\begin{aligned}&\tilde{f}(x_1,x_2,\ldots,x_k) \\ &= \frac{f_1(x_1)}{\Pr\{X_1\in A_1\}}\cdot\frac{f_2(x_2)}{\Pr\{X_2\in A_2(x_1)\}}\cdots\frac{f_k(x_k)}{\Pr\{X_k\in A_k(x_1,\ldots,x_{k-1})\}}\end{aligned}$$

if $(x_1, x_2, \ldots, x_k) \in A$; and $\tilde{f}(x_1, x_2, \ldots, x_k) = 0$ otherwise. Note that $\Pr\{X_1\in A_1\}\cdot\Pr\{X_2\in A_2(x_1)\}\cdots\Pr\{X_k \in A_k(x_1, \ldots, x_{k-1})\} > 0$ almost surely in the set A, if $\Pr\{A\} > 0$.

Under the IS distribution, we first simulate X_1 given that $X_1 \in A_1$, then simulate X_2 given that $X_2 \in A_2(X_1)$, and so on. Then we can compute L and D, and $L \in [y-\delta_n, y+\delta_n]$ w.p.1. Because the IS distribution $\tilde{f}$ is absolutely continuous with respect to f in the set A, the likelihood ratio

$$\begin{aligned}P_n &= \frac{f(X_1, X_2, \ldots, X_k)}{\tilde{f}(X_1, X_2, \ldots, X_k)} \\ &= \Pr\{X_1\in A_1\} \\ &\quad\cdot\Pr\{X_2\in A_2(X_1)\}\cdots\Pr\{X_k\in A_k(X_1,\ldots,X_{k-1})\}.\end{aligned}$$

Because $(X_1, X_2, \ldots, X_k)$ is a random vector, P_n is also a random variable and $P_n \leqslant 1$. Then the IS estimator is

$$\begin{aligned}\bar{M}_n^{IS} &= -\frac{1}{2n\delta_n}\sum_{i=1}^n P_{n,i}\cdot D_i\cdot 1_{\{y-\delta_n\leqslant L_i\leqslant y+\delta_n\}} \\ &= -\frac{1}{2n\delta_n}\sum_{i=1}^n P_{n,i}\cdot D_i,\end{aligned} \tag{22}$$

where the observations $(L_i, D_i, P_{n,i})$ are generated under the IS distribution.

Similar to the analysis in §5.1, we can show that $\tilde{\mathrm{E}}(\bar{M}_n^{IS}) = \mathrm{E}(\bar{M}_n)$ and

$$\begin{aligned}\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) &= \frac{1}{4n\delta_n^2}[\mathrm{E}(D^2\cdot P_n\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) \\ &\qquad - \mathrm{E}^2(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})] \\ &\leqslant \frac{1}{4n\delta_n^2}[\mathrm{E}(D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) \\ &\qquad - \mathrm{E}^2(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})] = \mathrm{Var}(\bar{M}_n).\end{aligned}$$

Therefore, $\bar{M}_n^{IS}$ has a smaller variance than $\bar{M}_n$ when both estimators use the same δ_n.

Note that $\tilde{\mathrm{E}}(P_n) = \mathrm{E}(1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) = 2f(y)\delta_n + o(\delta_n)$. In many situations, we can prove that $P_n = K_n\delta_n$, where $\mathrm{E}(K_n\cdot D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})$ is often of $O(\delta_n)$. Then

$$\begin{aligned}\widetilde{\mathrm{Var}}(\bar{M}_n^{IS}) &= \frac{1}{4\delta_n^2}[\tilde{\mathrm{E}}(P_n^2\cdot D^2) - \tilde{\mathrm{E}}^2(P_n\cdot D)] \\ &= \frac{1}{4n\delta_n^2}[\delta_n\tilde{\mathrm{E}}(P_nK_n\cdot D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) \\ &\qquad - \tilde{\mathrm{E}}^2(P_n\cdot D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})] \\ &= \frac{1}{4n\delta_n^2}[\delta_n\mathrm{E}(K_n\cdot D^2\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}}) \\ &\qquad - \mathrm{E}^2(D\cdot 1_{\{y-\delta_n\leqslant L\leqslant y+\delta_n\}})],\end{aligned}$$

which is of $O(n^{-1})$. Then the rate of convergence of $\bar{M}_n^{IS}$ is $n^{-1/2}$, if $n\delta_n^4 \to a$ with $a \geqslant 0$ as $n\to\infty$.

5.3. Impact to the Mean Square Error

By Theorem 3, to maximize the rate of convergence of $\bar{M}_n$, we face the trade-off between the bias and variance. The optimal choice is to set $\delta_n = O(n^{-1/5})$. Then the bias and variance of $\bar{M}_n$ are of $O(n^{-2/5})$ and $O(n^{-4/5})$, and the mean square error (MSE) is of $O(n^{-4/5})$.

The IS scheme does not change the mean of the estimator, i.e., $\mathrm{E}(\bar{M}_n) = \tilde{\mathrm{E}}(\bar{M}_n^{IS})$, then the bias of $\bar{M}_n^{IS}$ is still of $O(\delta_n^2)$ by Equation (9). However, it allows us to choose a smaller δ_n to reduce the bias of $\bar{M}_n^{IS}$. Because the variance of $\bar{M}_n^{IS}$ is of $O(n^{-1})$, which is independent of δ_n, we no longer face the trade-off between the bias and variance when selecting δ_n. Therefore, the theoretically optimal δ_n for $\bar{M}_n^{IS}$ is $\delta_n = 0$, and the bias becomes 0. Then, the MSE of $\bar{M}_n^{IS}$ is the same as the variance of $\bar{M}_n^{IS}$, which is of $O(n^{-1})$.

To use the IS scheme, however, we require $\delta_n > 0$. Otherwise, the likelihood ratio becomes 0. Therefore, we may choose δ_n arbitrarily close to 0 (depending on the precision of the computer) to minimize the bias of $\bar{M}_n^{IS}$. In the numerical examples reported in §6, we set $\delta_n = 10^{-10}$ for all the IS examples. Then, $\bar{M}_n^{IS}$ is practically unbiased.

6. Examples

In this section, we present three examples to illustrate the performances of $\bar{M}_n$ and $\bar{M}_n^{IS}$. We also compare them to the SPA and LR estimators whenever the estimators are available. In the electronic companion, we show that Assumptions 1 and 2 are satisfied by all the examples. In all the examples, we need to select a δ_n for $\bar{M}_n$. We follow a selection procedure that selects c such that $\delta_n = c \cdot n^{-1/5}$, and the procedure is fully described in the electronic companion. For the IS estimator, we set $\delta_n = 10^{-10}$ for all examples. We also construct confidence intervals for the examples. We set $\delta_n = c \cdot n^{-1/3}$ to ensure the validity of the confidence intervals, where c is determined by the same selection procedure as in the electronic companion. All the numerical results reported in this section are based on 1,000 independent replications.

6.1. A Financial Example

Suppose that the value of an asset follows the following diffusion process:

$$dS(t) = \mu(t, S(t))dt + \sigma(t, S(t))dB_t, \tag{23}$$

where B_t is a standard Brownian motion and $S(0) = S_0$. We are interested in estimating the sensitivity of $\Pr\{S(T) \leqslant y\}$ with respect to S_0 for some $y > 0$. Note that $p_y(S_0) = \Pr\{S(T) \leqslant y\}$ can be viewed as the default probability (Duffie and Singleton 2003) or the payoff of a digital option at maturity (Glasserman 2004).

To simulate $S(T)$, we use Euler scheme to discretize $S(t)$ (Glasserman 2004). Under the scheme

$$S_{i+1} = S_i + \mu_i \Delta t + \sigma_i \sqrt{\Delta t}\, Z_{i+1}, \quad i = 0, 1, \ldots, k-1, \tag{24}$$

where k is the number of time steps in the discretization, $\Delta t = T/k$, $t_i = (i/k)T$, $S_i = S(t_i)$, $\mu_i = \mu(t_i, S_i)$ and $\sigma_i = \sigma(t_i, S_i)$, and $\{Z_1, Z_2, \ldots, Z_k\}$ are independent standard normal random variables. Then, we can use simulation to generate $Z_1, Z_2, \ldots, Z_k$ to obtain $S(T) = S_k$. Furthermore, under the approximation scheme,

$$\frac{\partial S(T)}{\partial S_0} = \frac{\partial S_k}{\partial S_{k-1}} \frac{\partial S_{k-1}}{\partial S_{k-2}} \cdots \frac{\partial S_2}{\partial S_1} \frac{\partial S_1}{\partial S_0}$$

and

$$\frac{\partial S_{i+1}}{\partial S_i} = 1 + \frac{\partial \mu_i}{\partial S_i}\Delta t + \frac{\partial \sigma_i}{\partial S_i}\sqrt{\Delta t} Z_{i+1}, \quad i = 1, 2, \ldots, k-1.$$

Then the pathwise derivative $\partial S(T)/\partial S_0$ can also be computed in the simulation. Therefore, $\bar{M}_n$ can be computed easily by Equation (6).

6.1.1. IS Estimator. By Equation (24), $S(T) = S_k = S_{k-1} + \mu_{k-1}\Delta t + \sigma_{k-1}\sqrt{\Delta t} Z_k$, where Z_k is independent of S_{k-1}. Then, by the IS scheme of §5.2, we let $A_1 = A_2 = \cdots = A_{k-1} = \Re$ and

$$\begin{aligned} A_k(S_{k-1}) &= \{z_k \in \Re\colon y - \delta_n \leqslant S(T) \leqslant y + \delta_n \mid S_{k-1}\} \\ &= \left\{ \frac{y - \delta_n - S_{k-1} - \mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}} \leqslant Z_k \leqslant \frac{y + \delta_n - S_{k-1} - \mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}} \right\}. \end{aligned}$$

Under the IS scheme, we first generate S_{k-1} using the Euler scheme, then generate Z_k from a standard normal distribution truncated in the set $A_k(S_{k-1})$. Hence, the likelihood ratio is

$$P_n = \Phi\left(\frac{y + \delta_n - S_{k-1} - \mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right) - \Phi\left(\frac{y - \delta_n - S_{k-1} - \mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right). \tag{25}$$

Then, the IS estimator $\bar{M}_n^{IS}$ can be computed by Equation (22).

Now we analyze $\bar{M}_n^{IS}$ to obtain a better understanding of the estimator. By the stochastic version of Taylor's expansion (Lehmann 1999), when δ_n is close to 0,

$$P_n \approx \phi\left(\frac{y - S_{k-1} - \mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right) \frac{2\delta_n}{\sigma_{k-1}\sqrt{\Delta t}}.$$

Then, by Equation (22),

$$\bar{M}_n^{IS} \approx -\frac{1}{n\sqrt{\Delta t}} \sum_{j=1}^{n} \phi\left(\frac{y - S_{k-1}^j - \mu_{k-1}^j \Delta t}{\sigma_{k-1}^j \sqrt{\Delta t}}\right) \cdot \frac{\partial S_k^j}{\partial S_0} \cdot \frac{1}{\sigma_{k-1}^j}, \tag{26}$$

where the superscript j represents the jth observation of the simulation. By Equation (24),

$$\begin{aligned} \frac{\partial S_k^j}{\partial S_0} &= \frac{\partial S_{k-1}^j}{S_0} + \frac{\partial \mu_{k-1}^j}{\partial S_0}\Delta t + \frac{\partial \sigma_{k-1}^j}{\partial S_0}\sqrt{\Delta t} Z_k^j \\ &\approx \frac{\partial S_{k-1}^j}{S_0} + \frac{\partial \mu_{k-1}^j}{\partial S_0}\Delta t + \frac{\partial \sigma_{k-1}^j}{\partial S_0} \cdot \frac{y - S_{k-1}^j - \mu_{k-1}^j \Delta t}{\sigma_{k-1}^j}, \end{aligned} \tag{27}$$

where the last equation holds because $\sqrt{\Delta t} Z_k \approx (y - S_{k-1} - \mu_{k-1}\Delta t)/\sigma_{k-1}$ when δ_n is close to 0.

By Equation (26), $\bar{M}_n^{IS}$ has a rate of convergence of $n^{-1/2}$ when Δt is fixed. When Δt goes to zero, however, the rate of convergence of the estimator may become slower. Note that the rate of convergence of $\bar{M}_n$ is not affected by the size of Δt. Therefore, $\bar{M}_n$ may be even more efficient than $\bar{M}_n^{IS}$ when Δt is small.[4]

6.1.2. SPA Estimator. Based on Equation (24) and by conditioning on S_{k-1}, we have

$$p_y(S_0) = \mathrm{E}[\mathrm{E}(1_{\{S_{k-1}+\mu_{k-1}\Delta t+\sigma_{k-1}\sqrt{\Delta t}Z_k\leqslant y\}} \mid S_{k-1})] = \mathrm{E}\left[\Phi\left(\frac{y-S_{k-1}-\mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right)\right].$$

Then,

$$p_y'(S_0) = \mathrm{E}\left[\frac{\partial}{\partial S_0}\left\{\Phi\left(\frac{y-S_{k-1}-\mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right)\right\}\right] = -\mathrm{E}\left[\frac{1}{\sigma_{k-1}\sqrt{\Delta t}}\phi\left(\frac{y-S_{k-1}-\mu_{k-1}\Delta t}{\sigma_{k-1}\sqrt{\Delta t}}\right)\cdot Y\right],$$

where

$$Y = \frac{\partial S_{k-1}}{\partial S_0} + \frac{\partial \mu_{k-1}}{\partial S_0}\Delta t + \frac{y-S_{k-1}-\mu_{k-1}\Delta t}{\sigma_{k-1}}\frac{\partial \sigma_{k-1}}{\partial S_0}.$$

Therefore, the SPA estimator can be expressed as

$$\bar{M}_n^{SPA} = -\frac{1}{n\sqrt{\Delta t}}\sum_{j=1}^{n}\phi\left(\frac{y-S_{k-1}^j-\mu_{k-1}^j\Delta t}{\sigma_{k-1}^j\sqrt{\Delta t}}\right)\cdot Y^j\cdot\frac{1}{\sigma_{k-1}^j},$$

where

$$Y^j = \frac{\partial S_{k-1}^j}{S_0} + \frac{\partial \mu_{k-1}^j}{\partial S_0}\Delta t + \frac{\partial \sigma_{k-1}^j}{\partial S_0}\cdot\frac{y-S_{k-1}^j-\mu_{k-1}^j\Delta t}{\sigma_{k-1}^j}.$$

Compared to Equations (26) and (27), we find that $\bar{M}_n^{SPA}$ and $\bar{M}_n^{IS}$ are essentially the same estimator for this example.

6.1.3. LR Estimator. Let $f_i(s_i,\cdot)$ denote the conditional density of S_{i+1} given that $S_i = s_i$ for all $i = 0, 1, \ldots, k-1$. Then, we have

$$\begin{aligned}p_y'(S_0) &= \frac{\partial}{\partial S_0}\int_{-\infty}^{y}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f_0(S_0,s_1)\cdot f_1(s_1,s_2)\cdots f_{k-1}(s_{k-1},s_k)\,ds_1\cdots ds_{k-1}ds_k\\ &= \int_{-\infty}^{y}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial \log f_0(S_0,s_1)}{\partial S_0}f_0(S_0,s_1)\cdot f_1(s_1,s_2)\cdots f_{k-1}(s_{k-1},s_k)\,ds_1\cdots ds_{k-1}ds_k\\ &= \mathrm{E}\left[1_{\{S(T)\leqslant y\}}\cdot\frac{\partial \log f_0(S_0,S_1)}{\partial S_0}\right].\end{aligned}$$

By Equation (24), $f_0(S_0,s_1) = (1/\sigma_0\sqrt{\Delta t})\phi((s_1 - S_0 - \mu_0\Delta t)/\sigma_0\sqrt{\Delta t})$, then

$$\frac{\partial \log f_0(S_0,S_1)}{\partial S_0} = \frac{s_1-S_0-\mu_0\Delta t}{\sigma_0\sqrt{\Delta t}}\cdot\left[\frac{1}{\sigma_0\sqrt{\Delta t}}\left(1+\frac{\partial\mu_0}{\partial S_0}\Delta t\right) + \frac{s_1-S_0-\mu_0\Delta t}{\sigma_0^2\Delta t}\cdot\frac{\partial\sigma_0}{\partial S_0}\right] - \frac{1}{\sigma_0}\frac{\partial\sigma_0}{\partial S_0}.$$

Therefore, an LR estimator can be expressed as:

$$\bar{M}_n^{LR} = \frac{1}{n}\sum_{j=1}^{n}1_{\{S_k^j\leqslant y\}}\cdot\frac{\partial \log f_0(S_0,S_1^j)}{\partial S_0}.$$

The LR estimator has a rate of convergence of $n^{-1/2}$ when Δt is fixed. Similarly to the IS and SPA estimators, the rate of convergence of the estimator may become slower when Δt goes to 0.

6.1.4. Numerical Experiments. Suppose that $\mu(t,S(t)) = b[\mu - S(t)]$ and $\sigma(t,S(t)) = \sigma$. Then the diffusion process of Equation (23) is known as an Ornstein-Uhlenbeck (OU) process. For the OU process,

$$S(T) = S_0e^{-bT} + \mu(1-e^{-bT}) + \sigma\sqrt{\frac{1-e^{-2bT}}{2b}}Z,$$

where Z is a standard normal random variable. Therefore, the close-form expressions of $p_y(S_0)$ and $p_y'(S_0)$ can be derived, and the true value of $p_y'(S_0)$ can be calculated. We let $b = 10\%$, $\mu = 100$, $\sigma = 20$, $T = 0.25$, $y = 80$, and $S_0 = 100$. Then the true value of $p_y'(S_0)$ is -0.0051. We use this example to test the performances of all the estimators.

We report the estimated relative root mean square errors (RRMSE, defined as the percentage of root mean square error relative to the absolute value of the true probability sensitivity) of all the estimators for different sample sizes (n) and different numbers of time steps (k) in Table 1. From Table 1, we see that $\bar{M}_n$ always has a better performance than the LR estimator $\bar{M}_n^{LR}$ for all sample sizes and all numbers of time steps. The IS estimator $\bar{M}_n^{IS}$ and the SPA estimator $\bar{M}_n^{SPA}$ have almost same performances as demonstrated theoretically. The performance of $\bar{M}_n$ is not affected by the number of time steps.[5] However, the performances of all other estimators deteriorate as the number of time steps increases. When the number of time steps is small, $\bar{M}_n^{IS}$ and $\bar{M}_n^{SPA}$ outperform $\bar{M}_n$; when the number of time steps becomes large, it reverses. It can also be seen that $\bar{M}_n^{IS}$ and $\bar{M}_n^{SPA}$ have faster convergence rates than $\bar{M}_n$, which coincides with the analytical results in §5.

We also construct 90% confidence intervals based on Equation (12). We find that the observed coverage probabilities are approximately 90%, which is consistent with valid confidence intervals.

6.1.5. Extension of the Financial Example. Suppose that we are interested in $\Pr\{\min_{1\leqslant i\leqslant k} S_i \leqslant y\}$. It is the probability that a firm's asset is ever below a threshold value y

Table 1. Performance comparison of the four estimators for the financial example.

		RRMSE(%)			
	k	$\bar{M}_n$	$\bar{M}_n^{IS}$	$\bar{M}_n^{SPA}$	$\bar{M}_n^{LR}$
$n = 20{,}000$	10	5.5	2.7	2.7	8.1
	20	5.3	3.2	3.2	10.5
	50	5.4	4.3	4.3	16.0
	200	5.2	6.0	6.0	29.2
$n = 40{,}000$	10	4.1	1.9	1.9	6.2
	20	4.2	2.4	2.4	7.4
	50	4.0	3.0	3.1	10.9
	200	4.0	4.3	4.3	21.0
$n = 80{,}000$	10	3.3	1.4	1.4	4.8
	20	3.2	1.6	1.6	5.6
	50	3.0	2.2	2.2	8.1
	200	3.1	3.2	3.2	15.2

in the period of $(0, T]$ if $S(t)$ models the firm's asset. It can be used to measure the firm's default risk between 0 and T. In the rest of this section, we estimate $p'_y(\theta) = d\Pr\{\min_{1\leqslant i\leqslant k} S_i(\theta) \leqslant y\}/d\theta$ with $\theta = S_0$ under the OU process specified in §6.1.4.

Because $d[\min_{1\leqslant i\leqslant k} S_i(\theta)]/d\theta = S'_{i^*}(\theta)$ w.p.1, where $S_{i^*}(\theta) = \min_{1\leqslant i\leqslant k} S_i(\theta)$, the pathwise derivative of $\min_{1\leqslant i\leqslant k} S_i(\theta)$ is typically easy to obtain. Then, our estimator $\bar{M}_n$ can be straightforwardly applied, although it is difficult to use the importance-sampling estimator. The LR method can also be applied in this example. However, it is not clear how SPA can be applied.

We use Lemma 2 to verify Assumption 2 for this example. Because $\Pr\{\min_{1\leqslant i\leqslant k} S_i \leqslant y\} = 1 - \Pr\{\min_{1\leqslant i\leqslant k} S_i > y\}$, estimating the sensitivity of $\Pr\{\min_{1\leqslant i\leqslant k} S_i \leqslant y\}$ is equivalent to that of $\Pr\{\min_{1\leqslant i\leqslant k} S_i > y\}$ in terms of methodology. Note that $\{\min_{1\leqslant i\leqslant k} S_i > y\}$ is equivalent to $\{S_1 > y, \ldots, S_k > y\}$. Then, we may generate S_i given S_{i-1} conditioning on $S_i > y$. This scheme defines an IS distribution. Under this IS distribution, $\min_{1\leqslant i\leqslant k} S_i > y$ w.p.1 and

$$\Pr\left\{\min_{1\leqslant i\leqslant k} S_i > y\right\} = \tilde{\mathrm{E}}\left[\prod_{i=0}^{k-1} p_i(\tilde{S}_i)\right] \tag{28}$$

where $p_i(x) = \Pr\{S_{i+1} > y \mid S_i = x\}$ and $\tilde{S}_i$ is generated under the IS distribution. Then Assumption 2 may be verified for this example. The details of the verification are in the electronic companion.

Conceptually, we may directly estimate $d\Pr\{\min_{1\leqslant i\leqslant k} S_i > y\}/dS_0$ by using the pathwise method on the right-hand side of Equation (28). However, under the IS distribution, $d\tilde{S}_i/dS_0$ becomes very complicated. Then, $\partial_{S_0} \prod_{i=0}^{k-1} p_i(\tilde{S}_i)$ becomes impractical to calculate, especially when k is large. Therefore, this approach cannot be applied in practice. A similar example was given in Glasserman (2004) when he discussed the sensitivity estimation for barrier options.

To numerically compare the performance of $\bar{M}_n$ with the LR estimator $\bar{M}_n^{LR}$, we follow the same settings of §6.1.4. The true probability sensitivities for $k = 5, 10, 20$ are -0.0065, -0.0073, and -0.0080, respectively. The estimated RRMSEs of $\bar{M}_n$ and $\bar{M}_n^{LR}$ are summarized in Table 2. From the table we can see that $\bar{M}_n$ outperforms $\bar{M}_n^{LR}$ when k is relatively large. Estimation error of $\bar{M}_n^{LR}$ increases as k, whereas that of $\bar{M}_n$ is stable in k.

6.2. A Queueing Example

Let $L(\theta)$ be the customer's sojourn time in the steady state of a G/G/1 queue, where $\theta = (\theta_1, \theta_2)'$ and θ_1 and θ_2 denote the mean interarrival and mean service times, respectively. Let $p_y(\theta) = \Pr\{L(\theta) \leqslant y\}$. We are interested in estimating $\partial_{\theta_2} p_y(\theta)$. We use a steady-state simulation to generate (dependent) observations of $L(\theta)$ and $\partial_{\theta_2} L(\theta)$ by perturbation analysis (Ho and Cao 1983, Cao 1985). Therefore, $\bar{M}_n$ can be computed easily by Equation (6).

Table 2. Performance comparison of the estimators for the extension of the financial example.

	RRMSE(%)		
	k	$\bar{M}_n$	$\bar{M}_n^{LR}$
$n = 20{,}000$	5	7.1	7.7
	10	6.7	9.1
	20	6.4	11.0
$n = 20{,}000$	5	5.3	5.7
	10	4.8	6.3
	20	4.8	7.8
$n = 40{,}000$	5	4.0	4.1
	10	3.7	4.6
	20	3.6	5.6

6.2.1. IS Estimator. In a G/G/1 queue, the sojourn time of customer k only depends on the sojourn time of customer $k-1$, and the interarrival time and service time of customer k. Let L_k, I_k, and S_k denote the sojourn time, interarrival time, and service time of customer k, respectively. Then, $L_k = (L_{k-1} - I_k)^+ + S_k$, where $(L_{k-1} - I_k)^+$ is the waiting time of customer k and $a^+ = \max\{a, 0\}$. Furthermore, L_{k-1}, I_k, and S_k are independent, and the distribution functions of I_k and S_k are known.

To apply the IS scheme, we first generate L_{k-1} through simulation without IS. If $L_{k-1} \leqslant y + \delta_n$, then for any I_k, there exists $S_k > 0$ such that $L_k \in (y - \delta_n, y + \delta_n)$. If $L_{k-1} > y + \delta_n$, however, I_k must be larger than $L_{k-1} - (y + \delta_n)$; otherwise, $L_k \notin (y - \delta_n, y + \delta_n)$ for any $S_k > 0$. Therefore, given L_{k-1}, I_k has to satisfy $I_k \geqslant [L_{k-1} - (y + \delta_n)]^+$. Furthermore, given L_{k-1} and I_k, S_k has to satisfy $[y - \delta_n - (L_{k-1} - I_k)^+]^+ \leqslant S_k \leqslant y + \delta_n - (L_{k-1} - I_k)^+$. Therefore, the likelihood ratio

$$\begin{aligned} P_n = {}& \Pr\{I_k \geqslant [L_{k-1} - (y + \delta_n)]^+ \mid L_{k-1}\} \\ & \cdot \Pr\{[y - \delta_n - (L_{k-1} - I_k)^+]^+ \\ & \qquad \leqslant S_k \leqslant y + \delta_n - (L_{k-1} - I_k)^+ \mid L_{k-1}, I_k\}. \end{aligned} \tag{29}$$

Note that L_k is computed based on L_{k-1}. If L_{k-1} is generated through the IS scheme, then the likelihood of L_k is more complicated than P_n. To resolve this problem in the steady-state simulation, we suggest using the following approach. We first run the simulation without the IS in the warm-up period. Starting from the first customer after the warm-up period, we generate L_k and $\tilde{L}_k$ with and without IS, respectively, based on L_{k-1}, which is generated without IS. Then the likelihood ratio of each observation can be calculated using Equation (29), and we can use $\bar{M}_n^{IS}$ to estimate the sensitivity.

Now we analyze $\bar{M}_n^{IS}$ to obtain a better understanding of the estimator. If we let $W_k = (L_{k-1} - I_k)^+$ represent the waiting time of the kth customer, then $L_k = W_k + S_k$. To ensure that $L_k \in [y - \delta_n, y + \delta_n]$, the IS scheme requires that $W_k \leqslant y + \delta_n$ and $y - W_k - \delta_n \leqslant S_k \leqslant y - W_k + \delta_n$. Let

F_W and f_S denote the distribution function of W_k and the density of S_k, respectively. Then, when δ_n is close to 0,

$$P_n = \Pr\{W_k \leqslant y+\delta_n\}\cdot\Pr\{y-W_k-\delta_n \leqslant S_k \leqslant y-W_k+\delta_n\}$$
$$\approx F_W(y)\cdot f_S(y-W_k)\cdot 2\delta_n.$$

Hence, $\bar{M}_n^{IS} \approx -(1/n)\sum_{k=1}^n F_W(y)\cdot f_S(y-W_k)\cdot \partial_{\theta_2} L_k$, where both W_k and L_k are generated by the IS scheme. Note that, under the IS scheme, $L_k = W_k + S_k$ and $L_k \approx y$ when δ_n is close to 0. Then $S_k \approx y - W_k$. We have

$$\bar{M}_n^{IS} \approx -\frac{1}{n}\sum_{k=1}^n F_W(y)\cdot f_S(y-W_k) \cdot(\partial_{\theta_2} W_k + \partial_{\theta_2} S_k\,|_{S_k=y-W_k}), \quad (30)$$

where both W_k and L_k are generated under the IS measure.

6.2.2. SPA and LR Estimators. Fu and Hu (1997) have derived an SPA estimator for this example. It can be expressed as

$$\bar{M}_n^{SPA} = -\frac{1}{n}\sum_{k=1}^n f_S(y-W_k)\cdot(\partial_{\theta_2} W_k + \partial_{\theta_2} S_k\,|_{S_k=y-W_k})\cdot 1_{\{W_k\leqslant y\}}.$$

Now we compare the SPA estimator and the IS estimator of Equation (30). We find that the IS estimator can be viewed as an IS-enhanced SPA estimator with W_k being generated from the IS distribution $\tilde{f}_W(x) = f_W(x)/F_W(y)$ when $x \leqslant y$ and $\tilde{f}_W(x) = 0$ when $x > y$, where $f_W(x)$ is the original density of W_k. Therefore, we expect that the IS estimator is more efficient than the SPA estimator, especially when $\{W_k \leqslant y\}$ is a rare event. This example also shows that the SPA estimator and the IS estimator are very similar, although they are not identical, because they use essentially the same distribution information.

Fu and Hu (1997) have also derived an LR estimator for this example. It can be expressed as

$$\bar{M}_n^{LR} = \frac{1}{n}\sum_{k=1}^n LR_k \cdot 1_{\{L_k\leqslant y\}},$$

where $LR_k = \sum_{i=1}^k \partial_{\theta_2} \log f_i(L_i - 1, L_i)$, $f_i(x_{i-1},\cdot)$ is the conditional density of L_i given $L_{i-1} = x_{i-1}$, and L_0 is the sojourn time of the last customer of the warm-up period. Fu and Hu (1997) point out that the variance of the LR estimator may increase without bound as the sample size n increases in the estimation. Therefore, the LR estimator may not be useful for this example.

6.2.3. Numerical Experiments. Suppose that both the interarrival and service times follow exponential distributions. Then, the queueing system is an M/M/1 queue. When the queue is stable, i.e., $\theta_1 > \theta_2$, $L(\theta)$ is exponentially distributed with rate $1/\theta_2 - 1/\theta_1$ (Ross 1996). Therefore, $p_y(\theta) = \Pr\{L(\theta)\leqslant y\} = 1-\exp[-(1/\theta_2 - 1/\theta_1)y]$. Hence, $\partial_{\theta_2} p_y(\theta) = -y\exp[-(1/\theta_2 - 1/\theta_1)y]/\theta_2^2$. In this example, we let $\theta_1 = 10$, $\theta_2 = 8$, and $y = 2$. Then $p_y(\theta) = 4.88\%$, and $\partial_{\theta_2} p_y(\theta) = -2.9726\times 10^{-2}$. We use this example to test the performances of all the estimators.

We report the MSEs of all the estimators for different sample sizes (n) in Table 3. All MSEs are estimated through 1,000 independent replications. From Table 3, we find that the LR estimator is the worst among all estimators, and its variance blows up when the sample size increases. This confirms the observation of Fu and Hu (1997). Besides the LR estimator, all other estimators converge as the sample size increases. We also find that $\bar{M}_n$ performs less well than $\bar{M}_n^{IS}$ and $\bar{M}_n^{SPA}$, and $\bar{M}_n^{IS}$ outperforms $\bar{M}_n^{SPA}$, as we expect. Even considering the extra effort of computing the IS estimator, we find that it is still more efficient than the SPA estimator for this example.

Table 3. Performance comparison of the four estimators for the queueing example.

	RRMSE (%)			
n	$\bar{M}_n$	$\bar{M}_n^{IS}$	$\bar{M}_n^{SPA}$	$\bar{M}_n^{LR}$
5,000	23.2	11.8	19.1	>700
10,000	17.4	8.3	13.8	>1,000
50,000	9.2	3.6	5.9	>2,300
100,000	6.8	2.6	4.3	>3,300

We also construct 90% confidence intervals based on Equation (20). We find that the observed coverage probabilities are approximately 90%, which is consistent with valid confidence intervals.

6.3. A Portfolio Risk Example

In the financial industry, investors may hold large portfolios that may consist of many stocks, options, and other securities. To quantify the risk exposure of the portfolio and manage the risk, the investors may be interested in the probability that the loss of the portfolio at a future date is greater than some given threshold value, as well as the sensitivities of the probability with respect to various market parameters such as interest rate, stock prices, and volatilities.

To illustrate the basic ideas, we consider a simple portfolio that has two call options underlying two stocks, respectively. Let $S_{t,1}$ and $S_{t,2}$ denote the prices dynamics of the two stocks, and let $V_{t,1}$ and $V_{t,2}$ denote the prices of the two call options at time t. Suppose that $S_{t,i}$ follows a geometric Brownian motion process with drift μ_i (under the real-world probability measure) and volatility σ_i for both $i = 1, 2$. Then we may use the Black-Sholes formula to obtain the closed form of $V_{t,i}$ for any t that is earlier than the maturity date of the option. Note that $V_{t,i}$ is a random variable, because it depends on $S_{t,i}$, which is a random variable. To make this dependence explicit, we let $V_{t,i} = g_i(S_{t,i})$, $i = 1, 2$. Suppose that we form the portfolio at time 0 and both of the options mature at time T. Then, at any time $0 < \tau < T$, the loss of the portfolio is $L = V_0 - V_{\tau,1} - V_{\tau,2}$, where V_0 is the initial investment to obtain the two options. Furthermore, let $p = \Pr\{L \geqslant y\}$ for some $y > 0$. Suppose that we are interested in finding $\partial p/\partial\sigma_i$ for $i = 1, 2$.

Assume that $S_{t,1}$ and $S_{t,2}$ are independent. Then, $S_{\tau,i} = S_{0,i}e^{(\mu_i-\sigma_i^2/2)\tau+\sigma_i\sqrt{\tau}Z_i}$, $i = 1, 2$, where Z_1 and Z_2 are independent standard normal random variables. Therefore, L is an explicit function of (Z_1, Z_2) and we can compute $\partial_{\sigma_i} L$

easily. To implement $\bar{M}_n$, we only need to verify Assumption 2. Let $F_{S,2}(\cdot)$ and $F_2(\cdot)$ denote the cumulative distribution functions of $S_{\tau,2}$ and $V_{\tau,2}$, respectively. Note that the explicit form of $F_{S,2}(\cdot)$ can be derived easily. Then,

$$F_2(t) = \Pr\{g_2(S_{\tau,2}) \leqslant t\} = \Pr\{S_{\tau,2} \leqslant g_2^{-1}(t)\} \\ = F_{S,2}(g_2^{-1}(t)), \tag{31}$$

where $g_2^{-1}(\cdot)$ denotes the inverse function of $g_2(\cdot)$ and $g_2(\cdot)$ is invertible in this example. Then,

$$\Pr\{L \geqslant y\} = \mathrm{E}[\Pr\{V_0 - g_1(S_{\tau,1}) - g_2(S_{\tau,2}) \geqslant y \mid S_{\tau,1}\}] \\ = \mathrm{E}[F_2(V_0 - y - g_1(S_{\tau,1}))]. \tag{32}$$

Therefore, Assumption 2 can be verified using Lemma 1 when $\tau < T/2$. The details of the verification are provided in the electronic companion.

Conceptually, SPA can also be applied to this example. By Equations (31) and (32), we have

$$\frac{\partial \Pr\{L \geqslant y\}}{\partial \sigma_i} = \mathrm{E}\left[\frac{\partial}{\partial \sigma_i} F_2(V_0 - y - g_1(S_{\tau,1}))\right] \\ = \mathrm{E}\left[\frac{\partial}{\partial \sigma_i} F_{S,2}(g_2^{-1}((V_0 - y - g_1(S_{\tau,1})))\right].$$

To implement SPA, however, $g_2(\cdot)$ needs to be inverted for every observation of $S_{\tau,1}$. Because $g_2(\cdot)$ cannot be inverted explicitly, one has to use root-finding algorithms. This may become computationally intensive when the sample size is large. Therefore, it may not be practical to implement the SPA estimator for this example.

By the Black-Sholes formula, $V_{\tau,i} = S_{\tau,i}\Phi(d_{1,i}) - e^{-r(T-\tau)}K\Phi(d_{2,i})$, where r is the risk-free interest rate, K is the strike price, $d_{1,i} = [\log(S_{\tau,i}/K) + (r + \sigma_i^2/2)(T-\tau)]/(\sigma_i\sqrt{T-\tau})$, and $d_{2,i} = d_{1,i} - \sigma_i\sqrt{T-t}$. Then, σ_i does not appear only in $S_{\tau,i}$. Therefore, the LR method cannot be applied to this example. Furthermore, our IS estimator cannot be applied to this example either, because it is hard to find an IS distribution that ensures $\{y - \delta_n \leqslant L \leqslant y + \delta_n\}$ w.p.1.

To numerically test the performances of the estimators, we let $S_{0,1} = S_{0,2} = 100$, $r = 5\%$, $\sigma_1 = 20\%$, $\sigma_2 = 30\%$, $\mu_1 = 15\%$, $\mu_2 = 20\%$, $\tau = 2/52$, $T = 1$, $K = 105$, and $y = 4$. By using the finite-difference approach with an extremely large sample size (10^9), we find that the true values of $\partial \Pr\{L \geqslant y\}/\partial \sigma_i$ are approximately -1.95 and -1.70 for $i = 1, 2$, respectively. We use these values as benchmarks to test the performances of the estimators.

We compare $\bar{M}_n$ to the central finite-difference (FD) estimator that uses common random number and a step size that minimizes the MSE based on a pilot simulation. Denote the FD estimator by $\bar{M}_n^{FD}$. The RRMSEs of $\bar{M}_n$ and $\bar{M}_n^{FD}$ for the probability sensitivities with respect to (w.r.t.) σ_1 and σ_2 are summarized in Table 4. From the table, we can see that $\bar{M}_n$ significantly outperforms the FD estimator. Furthermore, to obtain n observations of the FD estimator, we need to run $2n$ simulations. For the results reported in Table 4, because we need to find sensitivities with respect to two parameters, i.e., σ_1 and σ_2, we run a total of $4n$ simulations. Therefore, $\bar{M}_n$ performs much better than $\bar{M}_n^{FD}$ in terms of computational efficiency.

Table 4. Performance comparison of the estimators for the portfolio risk example.

	RRMSE (%, w.r.t. σ_1)		RRMSE (%, w.r.t. σ_2)	
n	$\bar{M}_n$	$\bar{M}_n^{FD}$	$\bar{M}_n$	$\bar{M}_n^{FD}$
2,000	4.0	9.5	6.0	10.4
5,000	3.0	6.6	4.5	7.4
100,000	1.2	2.2	1.4	2.1

7. Conclusions

In this paper, we study how to estimate probability sensitivities through simulations. We propose an estimator based on a result of Hong (2009). We show that the estimator is consistent and asymptotically normally distributed for both terminating and steady-state simulations. We also demonstrate how to use importance sampling to accelerate the rate of convergence of the estimator. Numerical experiments show that our estimators have desired properties and may perform better than the SPA and LR estimators.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. The differentiability of $h_\gamma(t)$ may be verified in similar ways as in Lemmas 1 and 2.

2. Typically, we do not know how to simulate directly from the steady-state distribution. Therefore, we often start the simulation with certain initial conditions, and delete the observations obtained from the warm-up period to remove the initial-condition bias. The observations obtained after the warm-up periods are used to represent the steady-state behaviors of the simulation (see, for example, Law and Kelton 2000).

3. Because all discrete-event computer simulations run on computers with finite memory, they can only have a finite number of states in practice.

4. In §5, we show that $\bar{M}_n^{IS}$ has a smaller variance than $\bar{M}_n$ when δ_ns are the same for both estimators. In this section, however, δ_ns are not the same. The δ_n for $\bar{M}_n$ is selected by the procedure in the electronic companion, and the δ_n for $\bar{M}_n^{IS}$ is chosen to be 10^{-10}.

5. The number of time steps k is related to the discretization error. Usually, the diffusion process cannot be

exactly generated, but approximated by some discretization schemes such as Euler scheme (Glasserman 2004), and the discretization error depends on $\Delta t = T/k$. This discretization error can be reduced by using large k, but not by using large sample size. If higher precision, and hence smaller discretization error, are required, we may use a larger number of time steps k. However, an OU process can be generated exactly without discretization errors. Thus, all reported estimators except $\overline{M}_n$ are unbiased for whatever k. We use OU process only because of its analytical tractability, which helps us to test the performances of the estimators.

Acknowledgments

The authors thank Shane Henderson, the associate editor, and two anonymous referees for their helpful comments and suggestions. This research is supported in part by the Hong Kong Research Grants Council under grants CERG 613706 and CERG 613907.

References

Billingsley, P. 1968. *Convergence of Probability Measures*. Wiley, New York.

Billingsley, P. 1995. *Probability and Measure*, 3rd ed. Wiley, New York.

Broadie, M., P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Sci.* **42** 269–285.

Cao, X. R. 1985. Convergence of parameter sensitivity estimates in a stochastic experiment. *IEEE Trans. Automatic Control* **30** 845–853.

Chien, C., D. Goldsman, B. Melamed. 1997. Large-sample results for batch means. *Management Sci.* **43** 1288–1295.

Duffie, D., K. J. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, Princeton, NJ.

Fu, M. C. 2006. Gradient estimation. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science: Simulation*. Elsevier, Amsterdam.

Fu, M. C., J.-Q. Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer, Norwell, MA.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Kluwer, Norwell, MA.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Glynn, P. W., D. L. Iglehart. 1985. Large-sample theory for standardized time series: An overview. *Proc. 1985 Winter Simulation Conf.*, 129–134.

Heidelberger, P., P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Oper. Res.* **32** 185–209.

Ho, Y. C., X. R. Cao. 1983. Perturbation analysis and optimization of queueing networks. *J. Optim. Theory Appl.* **40** 559–582.

Hong, L. J. 2009. Estimating quantile sensitivities. *Oper. Res.* **57**(1) 118–130.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, Singapore.

L'Ecuyer, P. 1991. An overview of derivative estimation. *Proc. 1991 Winter Simulation Conf.*, 207–217.

Lehmann, E. L. 1999. *Elements of Large-Sample Theory*. Springer-Verlag, New York.

Marsden, J. E., M. J. Hoffman. 1993. *Elementary Classical Analysis*, 2nd ed. Freeman, New York.

Ross, S. M. 1996. *Stochastic Processes*, 2nd ed. Wiley, New York.

Schruben, L. 1983. Confidence interval estimation using standardized time series. *Oper. Res.* **31** 1090–1108.

Utev, S. A. 1990. On the central limit theorem for ϕ-mixing arrays of random variables. *Theory Probab. Its Appl.* **35** 131–139.