# SIMULATING PROBABILITY SENSITIVITIES

L. Jeff Hong
Guangwu Liu

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, China

## ABSTRACT

Probabilities are important measures of random performances. They are widely used in financial and service industries. Probabilities are often estimated through running simulation experiments. Probability sensitivities provide information on how changes in the input parameters affect the output probabilities. They are important for understanding and controlling stochastic systems. In this paper, we show how to use the simulated data, which are used to estimate probability, to estimate probability sensitivities. Our estimator is consistent, asymptotically normally distributed, and works for both terminating and steady-state simulations.

## 1 INTRODUCTION

Probabilities are important measures of random performances. They are widely used in practice. In financial industry, for instance, default probabilities are important measures of credit risk. In service industries, for instance, service qualities are often measured by probability that the sojourn times or waiting times of customers are more than certain service standard. To understand or control the probability performance of a system, one needs to understand how changes in the input parameters affect the output probabilities. These effects are often called probability sensitivities. When parameters vary continuously, sensitivities are essentially partial derivatives. They play crucial roles in managing or controlling random performances. For instance, it is important for financial institutes to know how sensitive a firm's default probability is to changes in firm's return rate and volatility, and for service providers to know how service quality deteriorates with respect to the changes in the arrival rate or service rate.

Simulations are often used to evaluate probabilities when the models of the random performance are complicated. In this paper, we are interested in estimating probability sensitivities using the same simulation observations that are used to estimate probabilities. Then, in a single simulation run, one can obtain not only the estimate of probability but also the estimates of probability sensitivities. Moreover, we develop estimates that can be applied to both terminating and steady-state simulations.

The sensitivity estimators can also be used to solve chance-constrained optimization problems. Chance-constrained optimization, or probabilistic programming, refers to the optimization problems with probability constraints. It has been studied extensively in the stochastic programming literature (e.g., Prékopa 2003), and algorithms in the literature often require the random distributions to have certain special structures. When probabilities and their gradients can be estimated, we may use simulation to generate the estimates of both constraint values and constraint gradients and feed them into a gradient-based nonlinear optimization algorithm, e.g., quasi-Newton methods (Nocedal and Wright 1999), to solve the problem. This method allows fairly general random distributions, and solve many chance-constrained problems that are otherwise difficult to solve.

Estimating probability sensitivities through simulation has been studied in the literature. Since probability function is the expectation of an indicator function, estimating sensitivities of probability functions can be viewed as a special case of estimating sensitivities of expectation functions, which has been studied extensively in the simulation literature. The reader can refer to L'Ecuyer (1991) and Fu (2006) for comprehensive reviews. The typical methods include perturbation analysis (PA), likelihood ratio/score function (LR/SF) method and weak derivative (WD) method. In the PA family, since the indicator function is discontinuous, infinitesimal perturbation analysis (IPA) cannot be applied directly and smoothed perturbation analysis (SPA) is required. However, SPA uses conditional expectation and often requires to determine what to condition on, which is often difficult to determine for practical problems. Similarly, WD method needs to choose the WD representation of the indicator function, which is also difficult. LR/SF method is widely applicable. However, its estimator often has large variance, which significantly degrades its performance. Computing probability sensitivities has also been studied in the stochastic optimization literature. Marti (2005) introduces three methods: integral transformation, asymptotic expansions and orthogonal function series expansions. However, these methods often require the probability function to have certain structure. Hong (2006) studies the estimation of quantile sensitivity through simulation. He has an intermediate result which gives the closed form of probability sensitivity under restrict assumptions.

In this paper, we first show the closed form of Hong (2006) indeed holds under much weaker assumptions in Section 2. We then propose an estimator of probability sensitivity. We show that, for both i.i.d. and dependent sequences, the estimator is consistent and follows asymptotic normal distribution under various assumptions in Sections 3 and 4. We summarize the numerical experiments in Section 5, followed by the conclusions in Section 6.

## 2 CLOSED FORMS OF PROBABILITY SENSITIVITY

Let $L(\theta)$ denote the random performance we are interested in, where $\theta$ denote a vector of continuous parameters. In this paper, we assume that $\theta$ is of one-dimensional and it is taken from an open set $\Theta \in \Re$. If $\theta$ is of multi-dimensional, we may treat each dimension as a one-dimensional parameter while fixing other dimensions constants. Let $F(y;\theta) = \Pr\{L(\theta) \le y\}$ denote the cumulative distribution function (CDF) of $L(\theta)$. Then we may define a probability function $p_y(\theta) = F(y;\theta)$ for any $y \in \Re$. We are interested in estimating $p_y^{'}(\theta) = dp_y(\theta)/d\theta$.

Let $L^{'}(\theta) = dL(\theta)/d\theta$ be the sample-path derivative of $L(\theta)$. If $L(\theta) = h(\theta, X)$ with some function $h$ and random variable (or vector) $X$, then $L^{'}(\theta) = \partial h(\theta, X)/\partial\theta$. For instance, $L(\theta)$ may be the random return of a financial portfolio which has $\theta$ share of a stock with an annual return $X$. Then $L^{'}(\theta) = X$. When $L(\theta)$ cannot be represented by a closed-form function, $L^{'}(\theta)$ may still be evaluated through IPA in many situations(Glasserman 1991). For instance, $L(\theta)$ may be the sojourn time of a customer in a GI/G/1 queue and $\theta$ may be the mean service time. Though the closed form of $L(\theta)$ is not available, $L^{'}(\theta)$ may be obtained through IPA. In this paper, we assume that $L^{'}(\theta)$ is available for any $\theta \in \Theta$.

We make the following assumptions on $L(\theta)$ and $L^{'}(\theta)$.

**Assumption 1** *The pathwise derivative $L^{'}(\theta)$ exists w.p.1 for any $\theta \in \Theta$, and there exists a random variable $K$ with $\mathrm{E}(K) < \infty$, such that*

$$|L(\theta_2) - L(\theta_1)| \le K|\theta_2 - \theta_1|$$

*for all $\theta_1, \theta_2 \in \Theta$.*

**Assumption 2** *For any $\theta \in \Theta$, $L(\theta)$ has a continuous density $f(t;\theta)$ in a neighborhood of $y$, and $\partial F(t;\theta)/\partial\theta$ exists and it is continuous at $t = y$ and continuous with respect to $\theta$.*

**Assumption 3** *For any $\theta \in \Theta$, let*

$$g(t;\theta) = \mathrm{E}\left[ L^{'}(\theta) \middle| L(\theta) = t \right].$$

*Then $g(t;\theta)$ is continuous at $t = y$.*

Assumption 1 is a typical assumption used to in pathwise derivative estimation (see, for instance, Broadie and Glasserman 1996). It guarantees the validity of interchanging differentiation and expectation. Assumptions 2 and 3 ensure that $L(\theta)$ has good mathematical properties in the neighborhood of $y$. However, outside of the neighborhood, $L(\theta)$ may take any form. Moreover, note that $p_y(\theta) = F(y;\theta)$. Therefore, assuming that $F(y;\theta)$ is differentiable with respect to $\theta$ is equivalent to assuming that $p_y(\theta)$ is differentiable with respect to $\theta$.

Then we have the following theorem that gives the closed form of $p_y^{'}(\theta)$.

**Theorem 1** *Suppose that Assumptions 1 to 3 are satisfied, then*

$$p_y^{'}(\theta) = -f(y;\theta) \cdot \mathrm{E}\left[ L^{'}(\theta) \middle| L(\theta) = y \right]. \qquad (1)$$

Note that Theorem 1 has the same conclusion as Theorem 1 of Hong (2006). However, it requires much simpler assumptions. Let $q_\alpha(\theta)$ be the $\alpha$-quantile of $L(\theta)$. If Assumptions 1 to 3 holds at $q_\alpha(\theta)$ instead of $y$, then with the same approach used in the proof of Theorem 2 of Hong (2006), we may show that

$$q_\alpha^{'}(\theta) = \mathrm{E}\left[ L^{'}(\theta) \middle| L(\theta) = q_\alpha(\theta) \right].$$

It is the same result as Theorem 2 of Hong (2006). However, it requires much simpler assumptions and applies to much more general situations.

## 3 ESTIMATING PROBABILITY SENSITIVITY USING INDEPENDENT OBSERVATIONS

To simplify the notation, we let $L$ and $D$ denote $L(\theta)$ and $L^{'}(\theta)$. Furthermore, let $g(t)$ and $f(t)$ denote $g(t;\theta)$ and $h(t;\theta)$ respectively, since $\theta$ is fixed when estimating $p_y^{'}(\theta)$. Let

$$h(t) = \mathrm{E}\left[D \cdot 1_{\{L\le t\}}\right].$$

Then, by Assumption 2,

$$h^{'}(y) = f(y)g(y). \tag{2}$$

Therefore, by Theorem 1,

$$p_y^{'}(\theta) = -f(y)g(y) = -h^{'}(y).$$

We further extend the definitions of $g(t)$ and $h(t)$. Let $\gamma > 1$ be a constant, and let

$$\begin{aligned} g_\gamma(t) &= \mathrm{E}\left[|D|^\gamma \mid L = t\right], \\ h_\gamma(t) &= \mathrm{E}\left[|D|^\gamma \cdot 1_{\{L\le t\}}\right]. \end{aligned}$$

If $g_\gamma(t)$ is continuous at $y = t$, then similar to Equation (2), we can prove that

$$h_\gamma^{'}(y) = f(y)g_\gamma(y).$$

Let $(L_1, D_1), (L_2, D_2), \ldots, (L_n, D_n)$ be the observations from $(L, D)$. Let $\delta_n$, $n = 1, 2, \ldots$, be a sequence of positive constants such that $\delta_n \to 0$ as $n \to 0$, and let

$$\bar{M}_n = -\frac{1}{2n\delta_n}\sum_{i=1}^{n} D_i \cdot 1_{\{y-\delta_n \le L_i \le y+\delta_n\}}. \tag{3}$$

When $(L_1, D_1), (L_2, D_2), \ldots, (L_n, D_n)$ are i.i.d. observations, we have the following theorems.

**Theorem 2** *Suppose that Assumptions 1 to 3 are satisfied, and $g_2(t)$ is continuous at $t = y$. If $n\delta_n \to \infty$ as $n \to \infty$, then $\bar{M}_n \to p_y^{'}(\theta)$ in probability as $n \to \infty$.*

Theorem 2 shows that $\bar{M}_n$ is a consistent estimator of probability sensitivity for i.i.d. observations.

**Theorem 3** *Suppose that Assumptions 1 to 3 are satisfied, $g_2(t)$ and $g_{2+\gamma}(t)$ are continuous at $t = y$ for some $\gamma > 0$, and $h^{'''}(t)$ exists at $t = y$. If $n\delta_n \to \infty$ and $n\delta_n^5 \to c$ as $n \to \infty$, then*

$$\sqrt{2n\delta_n}\left[\bar{M}_n - p_y^{'}(\theta)\right] \Rightarrow -\frac{\sqrt{2c}}{6}h^{'''}(y) + \sqrt{f(y)g_2(y)} \cdot N(0,1)$$

*as $n \to \infty$. If $n\delta_n \to \infty$ and $n\delta_n^5 \to 0$ as $n \to \infty$, then*

$$\sqrt{2n\delta_n}\left[\bar{M}_n - p_y^{'}(\theta)\right] \Rightarrow \sqrt{f(y)g_2(y)} \cdot N(0,1) \tag{4}$$

*as $n \to \infty$.*

Theorem 3 shows that $\bar{M}_n$ follows an asymptotic normal distribution for i.i.d. observations.

Theorem 3 shows that the rate of convergence of $\bar{M}_n$ is $(n\delta_n)^{-1/2}$, which can be of $O(n^{-2/5})$ if $n\delta_n^5 \to c$. However, if $n\delta_n^5 \to c$, the asymptotic normal distribution has a non-zero mean. If the mean is difficult to estimate, then confidence intervals of $p_y^{'}(\theta)$ can be difficult to construct. When $n\delta_n^5 \to 0$, $(n\delta_n)^{-1/2}$ is of $o(n^{-2/5})$, which is not as good as $O(n^{-2/5})$. However, the asymptotic normal distribution has a zero mean, which helps in the construction of asymptotically valid confidence intervals.

Let $\sigma_\infty^2 = f(y)g_2(y)$. To construct confidence intervals using Equation (4), we need to know how to estimate $\sigma_\infty^2$. Note that

$$\sigma_\infty^2 = f(y)\mathrm{E}[D^2|L = y]$$

and

$$p_y'(\theta) = -f(y)E[D|L = y].$$

Then we may use the same approach that estimates $p_y'(\theta)$ to estimating $\sigma_\infty^2$. Let

$$\bar{V}_n = \frac{1}{2n\delta_n}\sum_{i=1}^{n} D_i^2 \cdot 1_{\{y-\delta_n \le L_i \le y+\delta_n\}}.$$

If we assume that $g_2(t)$, $g_4(t)$ are continuous at $t = y$, by Theorem 2, we can easily show that $\bar{V}_n$ is a consistent estimator of $\sigma_\infty^2$. Then an asymptotically valid $100(1-\beta)\%$ confidence interval of $p_y'(\theta)$ is

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

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

# 4 ESTIMATING PROBABILITY SENSITIVITY USING DEPENDENT OBSERVATIONS

In this section we consider the estimation of probability sensitivity using dependent data. We show that, under certain conditions on $(L_i, D_i)$, the estimator $\bar{M}_n$ of Equation (3) 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 $\sigma$-algebra generated by $\{(L_i, D_i), i = 1, 2, \ldots, k\}$ and $\mathcal{G}_k$ be the $\sigma$-algebra generated by $\{(L_i, D_i), i = k, k+1, \ldots\}$. Following Billingsley (1999), let

$$\phi(k) = \sup\left\{|\Pr(B|A) - \Pr(B)| : A \in \mathcal{F}_s, \Pr(A) > 0, B \in \mathcal{G}_{s+k}\right\}.$$

Then the sequence is $\phi$-mixing if $\phi(k) \to 0$ as $k \to \infty$. In this section, we make the following assumption on $\{(L_i, D_i), i = 1, 2, \ldots\}$.

**Assumption 4** *The sequence $\{(L_i, D_i), i = 1, 2, \ldots\}$ satisfies that $\sum_{k=1}^{\infty} \sqrt{\phi(k)} < \infty$.*

Note that Assumption 4 implies that the sequence is $\phi$-mixing. Indeed, many stochastic processes are $\phi$-mixing. For instance, $m$-dependent process and stationary Markov process with finite state space (Billingsley 1968) are $\phi$-mixing, and positive recurrent regenerative processes are also $\phi$-mixing (Glynn and Iglehart 1985). In the simulation literature, the $\phi$-mixing assumption has also been used to study the steady-state behaviors of simulation, e.g., Chien et al. (1997) on batch means and Heidelberger and Lewis (1984) on quantile estimators. Chien et al. (1997) assume that $\phi(k) = O(k^{-9})$ which clearly implies our assumption, and Heidelberger and Lewis (1984) assume precisely what we assume. Then we have following theorem on the consistency of $\bar{M}_n$.

**Theorem 4** *Suppose that Assumptions 1 to 4 are satisfied, and $g_2(t)$ is continuous at $t = y$. If $n\delta_n \to \infty$ as $n \to \infty$, then $\bar{M}_n \to p_y'(\theta)$ in probability as $n \to \infty$.*

To study the asymptotic normality of $\bar{M}_n$, we need the following additional assumption.

**Assumption 5** *For all $n = 1, 2, \ldots$, $\alpha_n > 0$. Furthermore, there exists a constant $\alpha_\infty > 0$ such that $\lim_{n\to\infty} \alpha_n = \alpha_\infty$.*

Then we have the following theorem on the asymptotic normality of $\bar{M}_n$.

**Theorem 5** *Suppose that Assumptions 1 to 5 are satisfied, $g_2(t)$ and $g_{2+\gamma}(t)$ are continuous at $t = y$ for some $\gamma > 0$, and $h'''(t)$ exists at $t = y$. If $n\delta_n \to \infty$ and $n\delta_n^5 \to c$ as $n \to \infty$, then*

$$\sqrt{2n\delta_n}\left[\bar{M}_n - p_y'(\theta)\right] \Rightarrow -\frac{\sqrt{2c}}{6}h'''(y) + \sqrt{\alpha_\infty f(y) g_2(y)} \cdot N(0,1)$$

*as $n \to \infty$. If $n\delta_n \to \infty$ and $n\delta_n^5 \to 0$ as $n \to \infty$, then*

$$\sqrt{2n\delta_n}\left[\bar{M}_n - p_y'(\theta)\right] \Rightarrow \sqrt{\alpha_\infty f(y) g_2(y)} \cdot N(0,1) \quad (5)$$

*as $n \to \infty$.*

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

To construct an asymptotically valid confidence interval of $p_y'(\theta)$ using a dependent sequence $\{(L_i, D_i), i = 1, 2, \ldots\}$, we may use Equation (5). However, to use the equation, we need an approach to consistently estimating $\sigma_\infty^2 = \alpha_\infty f(y) g_2(y)$. Since $\alpha_\infty$ is unknown and difficult to estimate, $\sigma_\infty^2$ of dependent sequences is more difficult to estimate than it of i.i.d. sequences.

We suggest to use batch means method to estimate $\sigma_\infty^2$. 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 instance, a reasonable choice may be $m_n = k_n = \sqrt{n}$. Let

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

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

$$\begin{aligned}\bar{V}_n &= \frac{2m_n\delta_n}{k_n - 1}\sum_{j=1}^{k_n}\left[\bar{M}_{m_n}^{(j)} - \frac{1}{k_n}\sum_{j=1}^{k_n}\bar{M}_{m_n}^{(j)}\right]^2 \\ &= \frac{2m_n\delta_n}{k_n - 1}\sum_{j=1}^{k_n}\left[\bar{M}_{m_n}^{(j)} - \bar{M}_n\right]^2.\end{aligned}$$

Then we have the following lemma on the consistency of $\bar{V}_n$ in estimating $\sigma_\infty^2$ for stationary sequence.

**Lemma 1** *Suppose that the conditions of Theorem 5 are satisfied. Moreover, suppose that $m_n\delta_n \to \infty$ and $k_n\delta_n \to \infty$ as $n \to \infty$, and* *the sequences* $\left\{\left[2n\delta_n\left(\bar{M}_n - E[\bar{M}_n]\right)^2\right]^2, n = 1, 2, \ldots\right\}$ *and* $\left\{\left[2m_n\delta_n\left(\bar{M}_0(m_n) - E[\bar{M}_n]\right)^2\right]^2, n = 1, 2, \ldots\right\}$ *are uniformly integrable. Then* $\bar{V}_n \to \sigma_\infty^2$ *in probability as* $n \to \infty$.

Then an asymptotically valid $100(1-\beta)\%$ confidence interval of $p_y'(\theta)$, when the sequence $(L_1, D_1), (L_2, D_2), \ldots, (L_n, D_n)$ are dependent, is

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

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

## 5 NUMERICAL EXPERIMENTS

We have conducted three experiments to study the performances of our estimators. In the first example, we estimate the sensitivity of default probability through terminating simulation; in the second example, we estimate the probability sensitivity of the sojourn time of an M/M/1 queue through steady-state simulations. In the last example, we use our sensitivity estimator to solve a portfolio optimization problem, which may be formulated as a chance-constrained optimization problem. In the first two examples, our point estimator and confidence interval have desired properties. In the last example, our sensitivity estimator greatly improves the performance of the optimization algorithm.

## 6 CONCLUSIONS

In this paper we study how to estimate probability sensitivities through running simulation models. We first prove that probability sensitivity can be written in a closed form. We then propose an estimator based on the closed form. We show that the estimator is consistent and asymptotically normally distributed for both terminating simulations and steady-state simulations.

## ACKNOWLEDGMENTS

This research was partially supported by Hong Kong Research Grants Council grants CERG 613305 and CERG 613706.

## REFERENCES

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

Billingsley, P. 1999. *Convergence of Probability Measures*, Second Edition. Wiley, New York.

Broadie, M. and P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science*, **42** 269-285.

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

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

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

Glynn, P. W. and D. L. Iglehart. 1985. Large-sample theory for standardized time series: an overview. *Proceedings of the 1985 Winter Simulation Conference*, 129-134.

Heidelberger, P. and P. A. W. Lewis. 1984. Quantile estimation in dependent sequences. *Operations Research*, **32** 185-209.

Hong, L. J. 2006. Estimating quantile sensitivities. Working paper, Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong, China.

L'Ecuyer. 1991. An overview of derivative estimation. *Proceedings of the 1991 Winter Simulation Conference*, 207-217.

Marti, K. 2005. *Stochastic Optimization Methods.* Springer, Berlin.

Nocedal, J. and S. J. Wright. 1999. *Numerical Optimization.* Springer-Verlag, New York.

Prékopa, A. 2003. Probabilistic programming. A. Ruszczynski and A. Shapiro, ed. *Handbooks in Operations Research and Management Science: Stochastic Programming*, Elsevier.

## AUTHOR BIOGRAPHIES

**L. JEFF HONG** is an assistant professor in industrial engineering and logistics management at The Hong Kong University of Science and Technology. His research interests include Monte-Carlo method, sensitivity analysis and simulation optimization. His e-mail address is `<hongl@ust.hk>`.

**GUANGWU LIU** is a Ph.D. student in the Department of Industrial Engineering and Logistics Management at The Hong Kong University of Science and Technology. He is currently working on estimating sensitivities for risk measures. His e-mail address is `<liugw@ust.hk>`.