# Variance Reduction for Generalized Likelihood Ratio Method By Conditional Monte Carlo and Randomized Quasi-Monte Carlo

Yijie Peng, Michael Fu, Jiaqiao Hu, Pierre l'Ecuyer, Bruno Tuffin

# Variance Reduction for Generalized Likelihood Ratio Method By Conditional Monte Carlo and Randomized Quasi-Monte Carlo


Yijie Peng
Peking Universtiy
Beijing 100871, China

Michael C. Fu
Universtity of Maryland
College Park, Maryland 20742, USA

Jiaqiao Hu
Stony Brook University
Stony Brook, NY 11794-3600, USA

Pierre L'Ecuyer
University of Montreal
Montreal, Quebec, Canada

Bruno Tuffin
INRIA, Univ Rennes, CNRS, IRISA
Campus de Beaulieu, 35042 Rennes Cedex, France



April 12, 2021



## Abstract

The generalized likelihood ratio (GLR) method is a recently introduced gradient estimation method for handling discontinuities for a wide scope of sample performances. We put the GLR methods from previous work into a single framework, simplify regularity conditions for justifying unbiasedness of GLR, and relax some of those conditions that are difficult to verify in practice. Moreover, we combine GLR with conditional Monte Carlo methods and randomized quasi-Monte Carlo methods to reduce the variance. Numerical experiments show that the variance reduction could be significant in various applications.


## 1 Introduction

Simulation is widely used for analyzing and optimizing complex stochastic systems [1]. Specifically, we first generate simple input random variables $X = (X_1, \dots, X_n)$, and then simulate a complex output random variable $Y$ by evaluating the performance function $\vartheta(\cdot)$ at the input random variables, i.e., $Y = \vartheta(X)$, which is also known as the sample performance. Depending on the application, $\vartheta(\cdot)$ could take complicated functional forms. In stochastic activity networks used in project management, the performance function is nonlinear, because we are interested in the duration of the critical path, i.e., the activity path taking the longest duration; in a queueing system, the performance function can be estimated using a recursive equation when we are interested in waiting times of customers following certain dynamic mechanism; in financial options, the performance may be discontinuous due to thresholds determining whether the option is in the money or not.

Stochastic gradient estimation has been an important topic studied in simulation for decades [7], because it can be used for sensitivity analysis and optimization in complex stochastic models. Specifically, suppose the input random variables $X(\theta)$ and performance function $\vartheta(\cdot;\theta)$ are parameterized by $\theta$, and we aim to estimate the derivative of the expectation of the output random variable, i.e, $\partial\mathbb{E}[Y(\theta)]/\partial\theta$. When $\theta$ is multi-dimensional, the gradient is a vector of derivatives with respect to parameters in each dimension. By the definition of derivative, $\partial\mathbb{E}[Y(\theta)]/\partial\theta = \lim_{\delta\to 0}\mathbb{E}[Y(\theta+\delta) - Y(\theta)]/\delta$, and the finite difference (FD) estimator, $(Y(\theta+\delta) - Y(\theta))/\delta$, is straightforward and always implementable, and it treats the simulation model as a black-box. However, FD requires simulating one more sample path for each dimension of the parameter, and it suffers from a bias-variance tradeoff, i.e., the balance between choosing large $\delta$ to reduce the variance and using small values of $\delta$ to make the bias low. These undesirable properties are especially apparent when the dimension of the parameters is high.

In contrast, single-run stochastic gradient estimation techniques may provide unbiased estimates for the gradient with respect to all parameters simultaneously in the process of estimating the sample performance. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two classic single-run unbiased stochastic gradient estimation techniques [24], [17], [44]. A measured-value differentiation can be viewed as a technique to reduce the variance of LR at a cost of extra simulations [14]. IPA is a pathwise derivative estimate obtained from

$$\frac{\partial\mathbb{E}[Y(\theta)]}{\partial\theta} = \mathbb{E}\left[\frac{\partial Y(\theta)}{\partial\theta}\right]$$
$$=\mathbb{E}\left[\left.\frac{\partial\vartheta(x;\theta)}{\partial\theta}\right|_{x=X} + \sum_{i=1}^{n}\left.\frac{\partial\vartheta(x;\theta)}{\partial x_i}\right|_{x=X}\frac{\partial X_i}{\partial\theta}\right].$$

The interchange of derivative and expectation is usually justified by the dominated convergence theorem, which typically requires continuity and almost sure differentiability of the sample performance $Y(\theta) = \vartheta(X;\theta)$ with respect to $\theta$. A detailed theoretical discussion on IPA can be found in [10]. The LR estimator on the other hand can be obtained from

$$\frac{\partial \mathbb{E}[Y(\theta)]}{\partial \theta} = \frac{\partial}{\partial \theta}\int \vartheta(x) f_X(x;\theta)dx$$
$$= \int \vartheta(x)\frac{\partial f_X(x;\theta)}{\partial \theta}dx = \mathbb{E}\left[\vartheta(X)\frac{\partial \log f_X(x;\theta)}{\partial \theta}\right],$$

where $f_X(\cdot;\theta)$ is the joint density of the input random variables. To justify unbiasedness of the estimator, LR does not require continuity for the sample performance, but it can only estimate derivatives with respect to a distributional parameter, i.e., a parameter in the distribution of input random variables, but not a structural parameter, i.e., a parameter directly appearing in function $\vartheta(\cdot;\theta)$. Detailed theoretical discussion on LR can be found in [25]. A common belief in the simulation literature is that IPA usually has smaller variance than LR when they both apply. This may not be true in general (see, e.g., [4] for a counterexample), but has been substantiated in [4] under a sufficient condition. A hybrid of IPA and LR (IPA-LR) in [24] can be obtained by

$$\frac{\partial \mathbb{E}[Y(\theta)]}{\partial \theta} = \frac{\partial}{\partial \theta}\int \vartheta(x;\theta) f_X(x;\theta)dx$$
$$=\mathbb{E}\left[\left.\frac{\partial \vartheta(x;\theta)}{\partial \theta}\right|_{x=X} + \vartheta(X;\theta)\frac{\partial \log f_X(x;\theta)}{\partial \theta}\right].$$

Structural parameter is allowed for IPA-LR, but it requires differentibility of function $\vartheta(\cdot;\theta)$ with respect to the structural parameter.

The single-run unbiased stochastic gradient estimation techniques are considered in traditional backgrounds including discrete event systems [2], and risk management in financial engineering [11], [18], [20], [36], [47], [3], [19], [22], [16], [12]. Recently, this topic has attracted attention in artificial intelligence and machine learning [38], [43], where the dimension of the parameter is typically extremely high, so that the single-run and unbiasedness properties are particularly helpful in gradient-based optimization. Backpropagation (BP), the most popular gradient estimation technique for training artificial neural networks (ANN), is shown in [43] to be pathwise equivalent to IPA, and backward propagation of errors can reduce computational complexity. In addition, [43] shows that an LR-based method can significantly improve the robustness of ANN.

In simulation, discontinuities often appear in the sample performance with respect to the structural parameter, so that IPA and LR fail to be unbiased. For example, the derivative of the distribution function $\partial F(z)/\partial z = \partial \mathbb{E}[\mathbf{1}\{Y \le z\}]/\partial z$ has a discontinuous sample performance with respect to $z$ and the derivative and expectation cannot be interchanged. Smoothed perturbation analysis (SPA) and push-out LR can be used to address the discontinuity issue [9], [44], but SPA has to choose what to be conditioned on and push-out LR requires an explicit transformation to push the parameter into the density of input distribution, which are problem dependent. Peng et al. [42] propose a generalized LR (GLR) method to systematically treat sensitivity analysis for a large class of sample performances with discontinuities. In contrast to SPA and push-out LR, the GLR estimator has an analytical form without the need for conditioning and transformation. In [42], the unbiasedness of GLR is justified by a set of conditions, including that the tails of the input distribution go smoothly to zero fast enough, which excludes exponentially distributed random variables and uniform random numbers, for example. This smoothness requirement is relaxed in [41], where the inputs of the stochastic model are uniform random numbers, which are the basic building blocks for generating other random variables. GLR is shown to generalize both LR and push-out LR.

In this work, we put the GLR methods in [42] and [41] under a common framework. Then, by adopting a similar technique developed in [41], we simplify the assumptions required for justifying the unbiasedness of GLR in [42] and further relax some of the conditions that are difficulty to verify in practice. Although the GLR method has broad applicability, previous work indicates that it may inherit some of the undesirable variance properties of the LR method. We address this issue by combining GLR with conditional Monte Carlo (CMC) methods and randomized quasi-Monte Carlo (RQMC) methods.

CMC methods can reduce the variance and smooth the performance function in simulation by conditioning on certain events or random variables and then integrating out the remaining randomness [1]. For an estimator $H(Z)$, we have

$$\mathbb{E}[H(Z)] = \mathbb{E}[\widehat{H}(\widehat{Z})], \tag{1}$$

where $\widehat{H}(\widehat{Z}) := \mathbb{E}[H(Z)|\widehat{Z}]$ with $\widehat{Z}$ being a part of input random variables in $Z$. The variance reduction for the conditional estimator $\widehat{H}(\widehat{Z})$ can be seen from the following variance decomposition formula:

$$\mathrm{Var}(H(Z)) = \mathrm{Var}(\mathbb{E}[H(Z)|\widehat{Z}]) + \mathbb{E}[\mathrm{Var}(H(Z)|\widehat{Z})]$$
$$\ge \mathrm{Var}(\widehat{H}(\widehat{Z})).$$

Typically, $\widehat{H}(\widehat{Z})$ is smoother than $H(Z)$, due to the integration taken in the conditional expectation. SPA uses CMC to smooth the sample performance, after which IPA is applied to differentiate the conditional expectation. GLR does not need smoothing to obtain an unbiased derivative estimator, but CMC can be applied afterward to reduce the variance for GLR. The relation between applications of CMC in SPA and GLR is illustrated in Figure 1.

RQMC methods replace the vectors of uniform random numbers that drive independent simulation runs by dependent vectors of uniform random numbers that cover the space more evenly. When estimating an expectation, they can provide an unbiased estimator with a variance that converges to zero at a faster rate than with Monte Carlo [35],

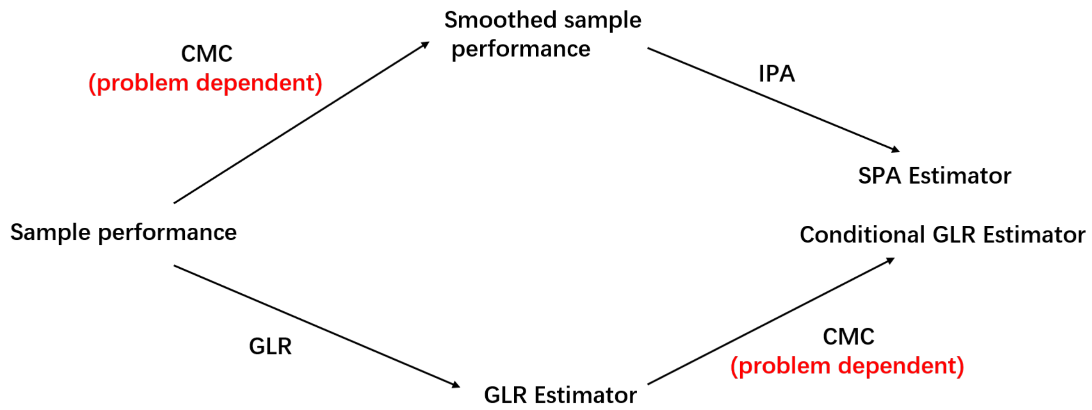



Figure 1: Applications of CMC to obtain SPA and conditional GLR.

[5]. Such a faster rate can be proved when the estimator inside the expectation is sufficiently smooth as a function of the underlying uniform random numbers. When the estimator is not smooth (e.g., discontinuous), the convergence rate may not be improved, but RQMC could still reduce the variance by a constant factor. We show, through numerical experiments, that the variance of the GLR estimator can be significantly reduced by appropriately combining GLR with CMC and RQMC. Similar type of combination of CMC and RQMC for reducing the variance of quantile estimation can be found in [39].

The rest of the paper is organized as follows. Section 2 introduces the GLR method. Variance reduction for GLR by CMC and RQMC is discussed in Section 3. Section 4 exemplifies the applications of the method to a stochastic activity network, a single-server queue, and a barrier option, with numerical experiments on them presented in Section 5. The last section offers conclusions.

## 2 Generalized Likelihood Ratio Method

We consider the sensitivity analysis problem of estimating

$$\frac{\partial}{\partial\theta}\mathbb{E}[\varphi(g(X;\theta))], \tag{2}$$

where $\varphi: \mathbb{R}^n \to \mathbb{R}$ is a measurable function not necessarily continuous, $g(x;\theta) = (g_1(x;\theta), \ldots, g_n(x;\theta))^T$ is a vector of functions with sufficient smoothness for $x \in \mathbb{R}^n$, and $X = (X_1, \ldots, X_n)$ is a vector of input random variables with a joint density $f_X(x;\theta)$ supported on $\Omega \subseteq \mathbb{R}^n$. From [42], the dimension of $g$ is required to be smaller or equal to the dimension of $x$ in deriving a GLR estimator. For simplicity of theoretical discussion, we require them to be the same. In application, $X$ does not have to include all the input random numbers given to a simulation model. Instead, $X$ can be only a subset of input random variables in simulation and we can condition on the remaining input random variables outside of X when deriving the GLR estimators. This leaves us freedom to select input random variables in $X$ for deriving GLR estimators as long as the conditions to ensure unbiasedness introduced later can be satisfied, and some choices of $X$ could lead to simpler forms and more desirable variance properties for GLR.

The Jacobian of $g(\cdot;\theta)$ is

$$J_g(x;\theta) := \begin{pmatrix} \frac{\partial g_1(x;\theta)}{\partial x_1} & \frac{\partial g_1(x;\theta)}{\partial x_2} & \cdots & \frac{\partial g_1(x;\theta)}{\partial x_n} \\ \frac{\partial g_2(x;\theta)}{\partial x_1} & \frac{\partial g_2(x;\theta)}{\partial x_2} & \cdots & \frac{\partial g_2(x;\theta)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(x;\theta)}{\partial x_1} & \frac{\partial g_n(x;\theta)}{\partial x_2} & \cdots & \frac{\partial g_n(x;\theta)}{\partial x_n} \end{pmatrix}.$$

In this work, we require the function $g(.;\theta)$ to be one-to-one (invertible) so that the Jacobian will be invertible. That is, $g$ must represent a valid multivariate change of variable. However, we only need the existence of the inversion or change of variable for $g$ rather than the explicit form. This requirement can be relaxed to the extent that that $g$ and its the Jacobian are only invertible locally [42]. Let $e_i$ be the $i$-th unit vector. We define

$$\begin{aligned} d(x;\theta) &:= -\mathrm{trace}\left(J_g^{-1}(x;\theta)\partial_\theta J_g(x;\theta)\right) \\ &+ \sum_{i=1}^{n} e_i^T J_g^{-1}(x;\theta)\left(\partial_{x_i} J_g(x;\theta)\right) J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) \\ &- \left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta)\right)^T \nabla_x \log f_X(x;\theta). \end{aligned}$$

From [42], when the tails of $f_X(\cdot;\theta)$ go smoothly to zero fast enough, then, under certain regularity conditions, it can be shown that an unbiased GLR estimator for (2) is given by

$$\begin{aligned} G(X;\theta) &:= \varphi(g(X;\theta))w(X;\theta), \\ w(x;\theta) &:= \frac{\partial \log f_X(x;\theta)}{\partial\theta} + d(x;\theta), \end{aligned} \tag{3}$$

where $\partial_y h(y)$ is the matrix obtained by differentiating $h$ with respect to $y$ element-wise.

Peng et al. [41] consider the case when $X$ is a vector of uniform random numbers $U = (U_1, \dots, U_n)$ such that $\Omega = (0,1)^n$ and $\partial \log f_X(x;\theta)/\partial\theta = 0$. Then under certain regularity conditions, we have the following unbiased GLR estimator for (2):

$$\widetilde{G}(U;\theta) := \varphi(g(U;\theta))d(U;\theta) + \sum_{i=1}^{n} \left[\varphi(g(\overline{U}_i;\theta))r_i(\overline{U}_i;\theta) - \varphi(g(\underline{U}_i;\theta))r_i(\underline{U}_i;\theta)\right] \quad (4)$$

where

$$\overline{U}_i := (U_1, \dots, \underbrace{1^-}_{i\text{th element}}, \dots, U_n),$$

$$\underline{U}_i := (U_1, \dots, \underbrace{0^+}_{i\text{th element}}, \dots, U_n),$$

and

$$r_i(x;\theta) := \left(J_g^{-1}(x;\theta)\, \partial_\theta g(x;\theta)\right)^T e_i, \quad i = 1, \dots, n.$$

In a special case when the input random variables are independent and non-parameterized with the tails of the distribution going smoothly to zero fast enough, from [41], the GLR estimator in [42] and the GLR estimator in [41] coincide even though the input random variables or sample point (in probability space) are interpreted differently in two GLR methods. This is in contrast to the unified IPA and LR framework in [24], where two different interpretations of the input random variables or sample point lead to two distinctive estimators, i.e, the IPA and LR estimators.

**Example 1.** *We use a simple density estimation problem to illustrate how to apply the two versions of the GLR method: We want to estimate*

$$\frac{\partial \mathbb{E}[\mathbf{1}\{X + U \le z\}]}{\partial z},$$

*where $X$ is a standard normal random variable, and $U$ is a uniform random variable $U(0,1)$ that is independent of $X$. Although there are two input random variables $X$ and $U$, we select one and condition on the other to apply GLR. In this example, $\varphi(\cdot) = \mathbf{1}\{\cdot \le 0\}$. This gives $n = 1$ in the preceding development. If we select $X$, then $g(x;z,u) = x + u - z$, $\partial g(x;z,u)/\partial x = 1$, $\partial g(x;z,u)/\partial z = -1$, $\nabla_x \log f_X(x) = -x$, and the other derivatives in (3) are zeros, so the GLR estimator can be given by $G(X;z,U) = -\mathbf{1}\{X+U \le z\}X$. If we select $U$, then $g(u;z,x) = x+u-z$, $\partial g(u;z,x)/\partial u = 1$, $\partial g(u;z,x)/\partial z = -1$, and other derivatives in (4) are zeros, so the GLR estimator can be given by $\widetilde{G}(U;z,X) = \mathbf{1}\{X \le z\} - \mathbf{1}\{X + 1 \le z\} = \mathbf{1}\{z - 1 < X \le z\}$, which coincides with an SPA estimator derived in [34] for the same problem. This example falls under the umbrella of distribution sensitivity estimation studied in both [40] and [41], which estimates the derivatives of the distribution function $F(z;\theta) = \mathbb{E}[\mathbf{1}\{g(X;\theta) \le z\}]$ with respect to both $\theta$ and $z$, with $g(\cdot;\theta)$ being a function with sufficient smoothness.*

*As a particular stochastic activity network (SAN) example, the output could be the maximum of the durations of activities on different paths. We then consider the following distribution function:*

$$\mathbb{E}\left[\mathbf{1}\left\{\max_{i=1,\dots,n} g_i(X;\theta) \le z\right\}\right] = \mathbb{E}\left[\prod_{i=1}^{n} \mathbf{1}\left\{g_i(X;\theta) - z \le 0\right\}\right],$$

*and the distribution sensitivities can be estimated by the GLR method using general frameworks (3) and (4) with*

$$\varphi(y) = \prod_{i=1}^{n} \mathbf{1}\{y_i \le 0\}.$$

*For the SAN example, $g_i$ would be the duration of activities on the $i$-th path, which will be analyzed in Section 4.*

**Example 2.** *For $0 \le \alpha \le 1$, the $\alpha$-quantile (also known as value-at-risk) of a random variable $g(U;\theta)$ with cdf $F(\cdot;\theta)$ is defined as*

$$q_\alpha(\theta) := \arg\min\{z : \; F(z;\theta) \ge \alpha\}.$$

*When $F(\cdot;\theta)$ is continuous, $q_\alpha(\theta) = F^{-1}(\alpha;\theta)$. Let $U^{(j)}$, $j = 1, \dots, m$, be i.i.d. realizations of $U \sim U(0,1)^d$, and $\hat{F}_m(\cdot)$ be the empirical distribution of $g(U^{(j)};\theta)$, $j = 1, \dots, m$. The empirical $\alpha$-quantile $\hat{F}_m^{-1}(\alpha)$ is the inverse of the empirical distribution evaluated at $\alpha$. This empirical quantile satisfies the following central limit theorem (46):*

$$\sqrt{m}\left(\hat{F}_m^{-1}(\alpha) - q_\alpha(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f(q_\alpha(\theta);\theta)}\right). \quad (5)$$

*We can estimate the asymptotic variance by*

$$\frac{m\alpha(1-\alpha)}{\sum_{j=1}^{m} G^{(j)}(z,\theta)\big|_{z=\hat{F}_m^{-1}(\alpha)}},$$

*where $G^{(j)}(z,\theta)$ is a GLR estimator for estimating the density using the same realizations of uniform random variables $U^{(j)}$ as in quantile estimator $\hat{F}_m^{-1}(\alpha)$. From [8], we can also apply distribution sensitivity estimators to estimate the quantile sensitivity*

$$\frac{dq_\alpha(\theta)}{d\theta} = -\frac{\partial F(z;\theta)}{\partial\theta}\Big|_{z=q_\alpha(\theta)} \Big/ f(q_\alpha(\theta);\theta),$$

*where the numerator and denominator can be estimated by GLR.*

We then simplify the regularity conditions in [42] and relax some of them that are difficult to verify in practice by adapting a similar technique as developed in [41]. In particular, we introduce the following conditions to justify the unbiasedness of the GLR estimation in the form (3).

(A.1) The function $g(\cdot;\theta)$ is invertible and the density $f_X(x;\theta)$ is supported on $\mathbb{R}^n$; $g(x;\theta)$ is twice continuously differentiable and $f_X(x;\theta)$ is continuously differentiable with respect to $(x,\theta) \in \mathbb{R}^n \times \Theta$, where $\Theta$ is a compact neighborhood of parameter $\theta$.

(A.2) The following limiting condition holds:

$$\lim_{x_i \to \pm\infty} \int_{\mathbb{R}^n} \sup_{\theta\in\Theta} \left| r_i(x;\theta) f_X(x;\theta) \right| \prod_{j\neq i} dx_j = 0, \quad i = 1,\ldots,n.$$

(A.3) The following integrability condition holds:

$$\int_{\mathbb{R}^n} \sup_{\theta\in\Theta} \left| \varphi(g(x;\theta)) w(x;\theta) f_X(x;\theta) \right| dx < \infty.$$

**Remark 1.** *Condition (A.1) requires that the function $g$ and density $f_X$ be sufficiently smooth. Condition (A.2) requires that the tails of $f_X(\cdot;\theta)$ go to zero fast enough. In contrast to the conditions in [42] for establishing unbiasedness of the GLR estimator, conditions (A.1)-(A.3) avoid certain integrability condition imposed on some intermediate smoothed function of $\varphi$ in the proof, which could be harder to verify in practice.*

**Theorem 1.** *Under conditions (A.1)-(A.3),*

$$\frac{\partial \mathbb{E}[\varphi(g(X;\theta))]}{\partial\theta} = \mathbb{E}[\varphi(g(X;\theta)) w(X;\theta)],$$

*where $w(\cdot)$ is defined by (3).*

The proof the theorem can be found in the appendix. The idea is to first smooth $\varphi(\cdot)$, which may not be continuous, then apply integration by parts to move differentiation from $\varphi(\cdot)$ to other smoother terms, and finally take limit to establish the unbiasedness of the final GLR estimator.

## 3 Variance Reduction

In this section, we discuss how to apply CMC and RQMC for reducing the variance of the GLR estimators.

### 3.1 Conditional Monte Carlo Method

The GLR estimator can be combined with CMC for reducing its variance. This will also be a key transformation to improve RQMC accuracy. We want to find an appropriate $\widehat{Z}$ to condition on such that the conditional expectation $\widehat{H}(\widehat{Z})$ in (1) becomes smoother and can be computed efficiently. To illustrate, we consider a special case of function $\varphi(\cdot)$:

$$\varphi(y) = \prod_{i=1}^{n} \mathbf{1}\{y_i \in \mathcal{Y}_i\} h(y),$$

where $h(\cdot)$ is a smooth function of $y = (y_1,\ldots,y_m)$, and $\mathcal{Y}_i$ is a set. This special case of $\varphi(\cdot)$ covers the examples of distribution sensitivities discussed later. Then the GLR estimator (3) can be written as

$$\prod_{i=1}^{n} \mathbf{1}\{g_i(X;\theta) \in \mathcal{Y}_i\} W(X;\theta),$$

where $W(x;\theta) := h(g(x;\theta))w(x;\theta)$, and the GLR estimator (4) can be written as

$$\sum_{i=1}^{n} \left[ \prod_{j=1}^{n} \mathbf{1}\{g_j(\overline{U}_i;\theta) \in \mathcal{Y}_i\} R_i(\overline{U}_i;\theta) - \prod_{j=1}^{n} \mathbf{1}\{g_j(\underline{U}_i;\theta) \in \mathcal{Y}_i\} R_i(\underline{U}_i;\theta) \right] + \prod_{i=1}^{n} \mathbf{1}\{g_i(U_i;\theta) \in \mathcal{Y}_i\} D(U;\theta),$$

where

$$R_i(u;\theta) := r_i(u;\theta) h(g(u;\theta)),$$
$$D(u;\theta) := d(u;\theta) h(g(u;\theta)).$$

Let $Z$ be a vector of generic random variables, which could either be $X$ or $U$, and $Q(\cdot)$ be a generic function which could either be $W(\cdot)$ or $D(\cdot)$. To smooth the sample performance, suppose there exists $i_* \in \{1,\ldots,n\}$ such that for $i = 1,\ldots,n$,

$$\mathbf{1}\{g_i(Z;\theta) \in \mathcal{Y}_i\} = \mathbf{1}\{Z_{i_*} \in \mathcal{Z}_i(Z_{-i_*};\theta)\},$$

where $z_{-j} = (z_1,\ldots,z_{j-1},z_{j+1},\ldots,z_n)$, and $\mathcal{Z}_i(\cdot)$ is a set depending on the argument. Here we only integrate out one random variable, and there are discussions on how to integrate more than one random variable for some applications in [34], which could lead to smaller variance. We have

$$\begin{aligned}
&\mathbb{E}\left[ \prod_{i=1}^{n} \mathbf{1}\{g_i(Z;\theta) \in \mathcal{Y}_i\} Q(Z;\theta) \right] \\
&= \mathbb{E}\left[ \mathbf{1}\{Z_{i_*} \in \cap_{i=1}^{n} \mathcal{Z}_i(Z_{-i_*})\} Q(Z;\theta) \right] \\
&= \mathbb{E}\left[ \int \mathbf{1}\{z_{i_*} \in \mathcal{Z}_i(Z_{-i_*};\theta)\} Q(z_{i_*}, Z_{-i_*}) f_{i_*}(z_{i_*}) dz_{i_*} \right] \\
&= \mathbb{E}\left[ \mathbb{P}\left( Z_{i_*} \in \cap_{i=1}^{n} \mathcal{Z}_i(Z_{-i_*}) \middle| Z_{-i_*} \right) \right. \\
&\qquad \left. \times \mathbb{E}\left[ Q(Z;\theta) \middle| Z_{i_*} \in \cap_{i=1}^{n} \mathcal{Z}_i(Z_{-i_*}), Z_{-i_*} \right] \right].
\end{aligned}$$

If, in addition, $Q(Z;\theta)$ is independent of $Z_{i*}$, then the equation above simplifies to

$$\mathbb{E}\left[ \mathbb{P}\left( Z_{i_*} \in \cap_{i=1}^{n} \mathcal{Z}_i(Z_{-i_*}) \middle| Z_{-i_*} \right) Q(Z;\theta) \right].$$

When $\mathcal{Y}_i$ is an interval $(a_i, b_i)$ and $g_i(\cdot)$ is strictly increasing with $z_{i_*}$, we have

$$\mathcal{Z}_i(z_{-i_*};\theta) = (g_i^{-1}(a_i; z_{-i_*}), g_i^{-1}(b_i; z_{-i_*})).$$

For the GLR estimator $-\mathbf{1}\{X + U \le z\}X$ in Example 1 of Section 2 for estimating the density, a conditional GLR estimator $(z - X)\mathbf{1}\{z - 1 < X < z\}$ can be obtained by

$$\begin{aligned}
&\mathbb{E}[-\mathbf{1}\{X + U \le z\}X] = \mathbb{E}[\mathbb{E}[-\mathbf{1}\{U \le z - X\}X|X]] \\
&= -\mathbb{E}[(z - X)\mathbf{1}\{z - 1 < X < z\}].
\end{aligned}$$

## 3.2 Randomized Quasi-Monte Carlo Method

Quasi-Monte Carlo (QMC) refers to a class of deterministic numerical integration methods in which the integrand is evaluated at a fixed set of $m$ points, and the average is used as an approximation. One limitation of the method is that it is very hard to estimate the approximation error in practice. RQMC takes the QMC points and randomizes them in a way that each point has the uniform distribution over $(0,1)^n$, so that each randomized point represents a proper realization of $U$ while the set of $m$ points still covers the unit hypercube $(0,1)^n$ more uniformly than typical independent random points (so the points are not independent) [27].

In general, for a given function $h$, RQMC estimates the integral $\mu = \int_{(0,1)^n} h(u)du$ by the average

$$\hat{\mu}_m := \frac{1}{m}\sum_{j=1}^{m} h(U^{(j)}),$$

where $U^{(1)},\ldots,U^{(m)}$ form an RQMC point set. The most common types of QMC point set constructions are lattice rules, polynomial lattices rules, and digital nets [5], [26]. For lattice rules, an appropriate randomization is a random shift modulo 1, which adds a single uniform random point to all the lattice points, and retains the shifted points that are in $(0,1)^n$ as the $m$ RQMC points. This randomization preserves the lattice structure, and there are explicit expressions for $\mathrm{Var}[\hat{\mu}_m]$ in terms of the Fourier coefficients of $h$, and computable bounds on this variance for certain classes of smooth functions [28], [30], [31]. When the mixed derivatives of $h$ are sufficiently smooth, the variance can converge at a faster rate than $\mathcal{O}(m^{-1})$, sometimes nearly $\mathcal{O}(m^{-2})$ and even faster in some cases. When $h$ is not smooth (e.g., discontinuous), these convergence rate results do not apply, although weaker results do apply [13], and even when the convergence rate is not improved, the variance is often reduced by a constant factor. For polynomial lattices rules and digital nets in general, which include Sobol' points, the random shift does not preserve the structure and net properties, but other appropriate randomizations do, including nested uniform scrambling, some affine scrambles, and random digital shifts. Variance bounds and convergence rate results are available for these, as well [5], [26].

We now discuss how to combine GLR with RQMC. Our model formulation (4) in terms of a function of independent $U(0,1)$ random variables makes it an obvious candidate for the application of RQMC, which is designed exactly for this type of formulation. In our setting, we can apply RQMC by taking $h$ as the GLR derivative estimator $\widetilde{G}(\,\cdot\,;\theta)$ to obtain the RQMC estimator

$$\bar{G}_m := \frac{1}{m}\sum_{j=1}^{m} \widetilde{G}\left(U^{(j)};\theta\right),$$

in which $\{U^{(1)},\ldots,U^{(m)}\}$ is an RQMC point set. For formulation (3), suppose $X$ can be generated by $\Gamma(U)$, and an RQMC estimator can be obtained by

$$\bar{G}_m := \frac{1}{m}\sum_{j=1}^{m} G\left(\Gamma(U^{(j)});\theta\right).$$

For example, when $X_1,\ldots,X_n$ are independent random variables with marginal distribution functions $F_1,\ldots,F_n$, then they can be generated by $X_i = F_i^{-1}(U_i)$, $i=1,\ldots,n$. Note that with RQMC, the terms in the sum of $\bar{G}_m$ are not independent, so one cannot estimate the variance of $\bar{G}_m$ through a straightforward application of the sample variance, as in standard Monte Carlo. To estimate the RQMC variance, one can make, say $l$, independent randomizations of the QMC point set, to obtain $l$ independent replicates of $\bar{G}_m$, and compute the sample variance of these $l$ replicates [29]. This could be used to compute a confidence interval on the true derivative, although one must be careful, because the distribution of $\bar{G}_m$ does not always converge to a normal distribution when $m\to\infty$ with RQMC [32]. The overall RQMC estimator in this case will be

$$\bar{G}_{m,l} := \frac{1}{l}\sum_{\ell=1}^{l} \bar{G}_m^{(\ell)},$$

where $\bar{G}_m^{(\ell)}$ is the $\ell$th independent replicate of $\bar{G}_m$, and $\mathrm{Var}[\bar{G}_{m,l}]$ is estimated by

$$\frac{1}{(l-1)l}\sum_{\ell=1}^{l}\left[\bar{G}_m^{(\ell)} - \bar{G}_{m,l}\right]^2.$$

For the $\alpha$-quantile sensitivity estimation discussed in Example 2 of Section 2, the cdf $F(\cdot;\theta)$ of $g(U;\theta)$ can be estimated by its empirical RQMC counterpart

$$\mathcal{F}_m(z) := \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\{g(U^{(j)};\theta)\le z\},$$

where $\{U^{(1)},\ldots,U^{(m)}\}$ is an RQMC point set, and the quantile $q_\alpha(\theta)$ can be estimated by the pseudo-inverse $\mathcal{F}_m^{-1}(\alpha)$. With $l$ independent randomizations of the RQMC points, we can average the $l$ independent randomizations of this empirical RQMC cdf,

$$\mathcal{F}_{m,\ell}(z) := \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\{g\left(U^{(j,\ell)};\theta\right)\le z\},$$

$$\bar{\mathcal{F}}_{m,l}(z) := \frac{1}{l}\sum_{\ell=1}^{l} \mathcal{F}_{m,\ell}(z),$$

and estimate $q_\alpha(\theta)$ by $\bar{\mathcal{F}}_{m,l}^{-1}(\alpha)$, which is a consistent estimator and satisfies the central limit theorem:

$$\sqrt{l}\left(\bar{\mathcal{F}}_{m,l}^{-1}(\alpha) - q_\alpha(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathrm{Var}\left(\mathcal{F}_{m,\ell}(q_\alpha(\theta))\right)}{f^2(q_\alpha(\theta))}\right).$$

The asymptotic variance can be estimated by

$$\left.\frac{\frac{1}{l-1}\sum_{\ell=1}^{l}\left(\mathcal{F}_{m,\ell}(z) - \bar{\mathcal{F}}_{m,l}(z)\right)^2}{\left(\bar{G}_{m,l}(z)\right)^2}\right|_{z=\bar{\mathcal{F}}_{m,l}^{-1}(\alpha)}$$

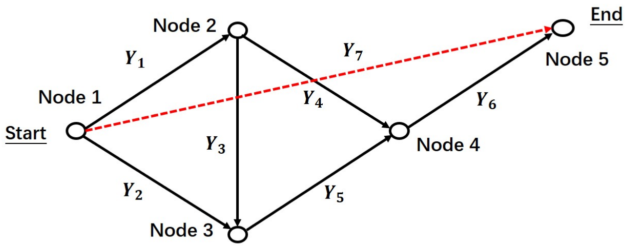



Figure 2: A SAN with seven activities.

where $\bar{G}_{m,l}(z)$ is a GLR estimator for density using RQMC. Kaplan et al. [23] point out that using $\frac{1}{l}\sum_{\ell=1}^{l}\mathcal{F}_{m,\ell}^{-1}(\alpha)$ instead of $\bar{\mathcal{F}}_{m,l}^{-1}(\alpha)$ would not provide a consistent estimator of $q_\alpha(\theta)$ as $l\to\infty$. Similarly, we can also estimate the quantile sensitivity by GLR using RQMC.

# 4 Applications

In this section, we discuss the applications of GLR and CMC to estimating distribution sensitivities for a SAN, single-server queue, and barrier option. Distribution sensitivities can also be applied to calibrate parameters and optimize quantile-based risk measures [40], [21].

## 4.1 Stochastic Activity Network

We estimate a simple SAN studied in [15], which is depicted in Figure 2. Distribution sensitivity estimation by SPA for a SAN with a different structure can be found in [6] and [34]. As in [15], the duration of the seventh activity is assumed to be deterministic. There are four different paths representing the tasks to reach the final stage of a project, i.e.,

$$(1,4,6),\quad (2,5,6),\quad (1,3,5,6),\quad (7).$$

The completion time for the entire project is $\max(Y_1+Y_4+Y_6, Y_2+Y_5+Y_6, Y_1+Y_3+Y_5+Y_6, Y_7)$, and the sample performance for the distribution function of completion time is

$$\begin{aligned}Y=&\mathbf{1}\{\max(Y_1+Y_4+Y_6,Y_2+Y_5+Y_6,\\&\quad Y_1+Y_3+Y_5+Y_6,Y_7)\le z\}\\=&\mathbf{1}\{Y_1+Y_4+Y_6\le z\}\mathbf{1}\{Y_2+Y_5+Y_6\le z\}\\&\times\mathbf{1}\{Y_1+Y_3+Y_5+Y_6\le z\}\mathbf{1}\{Y_7\le z\}.\end{aligned}$$

Unlike [15], where the durations of the first six activities all follow independent normal distributions, we assume that the first three activities follow independent exponential distributions: $Y_i=-\frac{1}{\lambda_i}\log(U_i)$, $i=1,2,3$, and the other three activities follow independent log-normal distributions, $Y_i=\exp(\mu_i+\sigma_i X_i)$, $i=4,5,6$. We note that $\mathbf{1}\{Y_1+Y_4+Y_6\le z\}\mathbf{1}\{Y_2+Y_5+Y_6\le z\}=\mathbf{1}\{Y_1+\max(Y_4,Y_3+Y_5)+Y_6\le z\}$ and $\mathbf{1}\{Y_2+Y_5+Y_6\le z\}\mathbf{1}\{Y_1+Y_3+Y_5+Y_6,Y_7\le z\}=\mathbf{1}\{Y_5+\max(Y_2,Y_1+Y_3)+Y_6\le z\}$.

For $z\le Y_7$, the distribution function $F(z)$ of the completion time $Y$ is equal to zero, and for $z>Y_7$, the distribution function is

$$\begin{aligned}F(z)=&\mathbb{E}[\mathbf{1}\{Y_1+\max(Y_4,Y_3+Y_5)+Y_6-z\le 0\}\\&\times\mathbf{1}\{Y_2+Y_5+Y_6-z\le 0\}]\\=&\mathbb{E}[\mathbf{1}\{Y_4+Y_1+Y_6-z\le 0\}\\&\times\mathbf{1}\{Y_5+\max(Y_2,Y_1+Y_3)+Y_6-z\le 0\}].\end{aligned}$$

To estimate the density $f(z)=\frac{\partial}{\partial z}F(z)$, we can view $(U_1,U_2)$ as the input random variables in the stochastic model $\varphi(g(U;\theta))$, and we have

$$\begin{aligned}&g_1(U_1,U_2;z)=-\frac{1}{\lambda_1}\log U_1+\max(Y_4,Y_3+Y_5)+Y_6-z,\\&g_2(U_1,U_2;z)=-\frac{1}{\lambda_2}\log U_2+Y_5+Y_6-z,\\&\partial_z g(u_1,u_2;z)=-(1,1)^T.\end{aligned}$$

The Jacobian matrix and its inverse are

$$\begin{aligned}&J_g(u_1,u_2;z)=-\begin{pmatrix}\frac{1}{\lambda_1 u_1}&0\\0&\frac{1}{\lambda_2 u_2}\end{pmatrix},\\&J_g^{-1}(u;z)=-\begin{pmatrix}\lambda_1 u_1&0\\0&\lambda_2 u_2\end{pmatrix},\end{aligned}$$

and

$$\begin{aligned}&r_1(u_1,u_2;z)=\lambda_1 u_1,\quad r_2(u_1,u_2;z)=\lambda_2 u_2,\\&d(u_1,u_2;z)=-\lambda_1-\lambda_2.\end{aligned}$$

We emphasize that the input random variables in the stochastic model $\varphi(g(U;\theta))$ must be selected carefully in a way that the Jacobian of $g$ is invertible. Also, in the example, the three paths are merged into two by some transformation, so we end up with a $2\times 2$ Jacobian. But we could also have kept the three paths and selected $(U_1,U_2,U_3)$ as the input, which is more direct. So there are many valid choices for what we take as "input" when applying the method. For the SAN, the authors of [34] select a minimal cut and compute the density on the completion time conditional on the other links (not in the cut) as a density estimator. For GLR, we can take the lengths of the links of the minimal cut as the input random variables. For $z>Y_7$, the GLR estimator of $f(z)$ is given by

$$\begin{aligned}&G(U_1,U_2;z)\\&=\lambda_1\mathbf{1}\{\max(Y_4+Y_6,Y_2+Y_5+Y_6,Y_3+Y_5+Y_6)\le z\}\\&+\lambda_2\mathbf{1}\{\max(Y_1+Y_4+Y_6,Y_5+Y_6,Y_3+Y_5+Y_6)\le z\}\\&-(\lambda_1+\lambda_2)\mathbf{1}\left\{\widetilde{Y}\le z\right\},\end{aligned}$$

where $\widetilde{Y}=\max(Y_1+Y_4,Y_2+Y_5,Y_1+Y_3+Y_5)$. Furthermore, we have

$$\begin{aligned}&\mathbb{E}[G(U_1,U_2;z)]=\mathbb{E}[\mathbb{E}[G(U_1,U_2;z)|Y_1,\ldots,Y_5]]\\&=\lambda_1\mathbb{E}\left[\Phi\left(\frac{1}{\sigma_6}\left(\log[(z-\max(Y_4,Y_2+Y_5,Y_3+Y_5))^+]-\mu_6\right)\right)\right]\\&+\lambda_2\mathbb{E}\left[\Phi\left(\frac{1}{\sigma_6}\left(\log[(z-\max(Y_1+Y_4,Y_5,Y_1+Y_3+Y_5))^+]-\mu_6\right)\right)\right]\\&-(\lambda_1+\lambda_2)\mathbb{E}\left[\Phi\left(\frac{1}{\sigma_6}\left(\log[(z-\widetilde{Y})^+]-\mu_6\right)\right)\right].\end{aligned}$$

On the other hand, we can also let $(X_4, X_5)$ be the input random variables in the stochastic model $\varphi(g(X;\theta))$, leading to

$$g_1(X_4, X_5; z) = \exp(\mu_4 + \sigma_4 X_4) + Y_1 + Y_6 - z,$$
$$g_2(X_4, X_5; z) = \exp(\mu_5 + \sigma_5 X_5) + \max(Y_2, Y_1 + Y_3) + Y_6 - z,$$
$$\partial_z g(x_4, x_5; z) = -\big(1, 1\big)^T,$$
$$\nabla \log f_{(X_4,X_5)}(x_4, x_5) = -\big(x_4, x_5\big)^T.$$

The Jacobian matrix and its inverse are

$$J_g(x_4, x_5; z) = \begin{pmatrix} \sigma_4 e^{\mu_4+\sigma_4 x_4} & 0 \\ 0 & \sigma_5 e^{\mu_5+\sigma_5 x_5} \end{pmatrix},$$
$$J_g^{-1}(x_4, x_5; z) = \begin{pmatrix} \frac{1}{\sigma_4} e^{-\mu_4-\sigma_4 x_4} & 0 \\ 0 & \frac{1}{\sigma_5} e^{-\mu_5-\sigma_5 x_5} \end{pmatrix},$$

and

$$d(x_4, x_5; z) = -\left(1 + \frac{x_4}{\sigma_4}\right) e^{-\mu_4-\sigma_4 x_4} - \left(1 + \frac{x_5}{\sigma_5}\right) e^{-\mu_5-\sigma_5 x_5}.$$

Again, the input random variables in the stochastic model $\varphi(g(X;\theta))$ must be selected carefully in a way that the Jacobian of $g$ is invertible. For $z > Y_7$, the GLR estimator of $f(z)$ is given by

$$G(X_4, X_5; z) = -\mathbf{1}\{Y \le z\}\left[\left(1 + \frac{X_4}{\sigma_4}\right)\frac{1}{Y_4} + \left(1 + \frac{X_5}{\sigma_5}\right)\frac{1}{Y_5}\right].$$

Furthermore, we have

$$\mathbb{E}[G(X_4, X_5; z)] = \mathbb{E}[\mathbb{E}[G(X_4, X_5; z)|Y_1, \ldots, Y_5]]$$
$$= -\,\mathbb{E}\left[\left(\frac{1}{Y_4} + \frac{X_4}{\sigma_4 Y_4} + \frac{1}{Y_5} + \frac{X_5}{\sigma_5 Y_5}\right) \times \Phi\left(\frac{1}{\sigma_6}\big(\log[(z - \widetilde{Y})^+] - \mu_6\big)\right)\right].$$

We can take the expression inside the expectation as a conditional GLR estimator.

## 4.2 Single-Server Queue

We consider distribution sensitivity estimation for the waiting time of the customers in a single-server first-come-first-served queue depicted in Figure 3. Density estimation for this single-server queue by SPA and RQMC can be found in [34]. When the $i$-th customer arrives, he/she may need to wait if the system time (waiting time plus service time), for the $(i-1)$-th customer is longer than the interarrival time between the $i$-th customer and $(i-1)$-th customer. Otherwise, the waiting time of the $i$-th customer is zero, i.e., the waiting time of customers follows the Lindley equation:

$$W_i = \max\{0, W_{i-1} + S_{i-1} - A_i\}, \quad i \ge 2,$$

where $W_i$ and $S_i$ are the waiting time and service time of the $i$-th customer, and $A_i$ is the interarrival time between

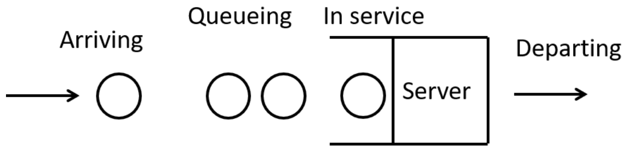



Figure 3: A single-server queue.

the $i$-th customer and the $(i-1)$-th customer. Suppose $S_i = \exp(\mu_1 + \sigma_1 X_i)$ and $A_i = \exp(\mu_2 + \sigma_2 X_i')$ follow independent log-normal distributions, where $X_i$ and $X_i'$ follow the standard normal distribution for $i \ge 1$. We can also represent the service time as $S_i = \exp\big(\mu_1 + \sigma_1 \Phi^{-1}(U_i)\big)$ so that uniform random numbers could be viewed as input random variables for deriving the GLR estimator. Let $\mu_1 = \theta$. The density $f(z;\theta)$ of $W_n$ over $(0,\infty)$ can be written as the derivative of the distribution function $F(z;\theta)$ of $W_n$:

$$\frac{\partial F(z;\theta)}{\partial z} = \frac{\partial}{\partial z}\mathbb{E}[\mathbf{1}\{W_n \le z\}] = \frac{\partial}{\partial z}\mathbb{E}[\mathbf{1}\{W_{n-1} + S_{n-1} - A_n \le z\}].$$

If we let

$$g(X_{n-1}, z) = W_{n-1} + \exp(\mu_1 + \sigma_1 X_{n-1}) - A_n - z,$$

then

$$\frac{\partial g(x_{n-1}, z)}{\partial x_{n-1}} = \sigma_1 \exp(\mu_1 + \sigma_1 x_{n-1}),$$
$$\frac{\partial^2 g(x_{n-1}, z)}{\partial x_{n-1}^2} = \sigma_1^2 \exp(\mu_1 + \sigma_1 x_{n-1}),$$
$$\frac{\partial \log f_{X_{n-1}}(x_{n-1})}{\partial x_{n-1}} = -x_{n-1},$$

and from [40], the GLR estimator for the density is

$$G_1(X_{n-1}; z) = -\mathbf{1}\{W_n \le z\}\frac{X_{n-1} + \sigma_1}{S_{n-1}\sigma_1}, \quad z > 0.$$

To estimate the derivative with respect to $\theta$, let

$$g_i(X_i;\theta) = \theta + \sigma_1 X_i, \quad i = 1, \ldots, n-1.$$

Then from [42], the GLR estimator for $\partial F(z;\theta)/\partial\theta$ is given by

$$G_2(X;\theta) = \frac{1}{\sigma_1}\mathbf{1}\{W_n \le z\}\sum_{i=1}^{n-1} X_i, \quad z > 0,$$

which coincides with the classic LR estimator. The GLR estimators in [42] and [41] also coincide for this example, because the input random variables are independent and non-parameterized, with the tails of the distribution going smoothly to zero fast enough (standard normal in this example). From [40], higher-order distribution sensitivities can also be obtained by the GLR method, e.g.,

$$\frac{\partial^2 F(z;\theta)}{\partial\theta\partial z} = \mathbb{E}\left[\mathbf{1}\{W_n \le z\}\frac{1 - \left(\sum_{i=1}^{n-1} X_i\right)(X_{n-1} + \sigma_1)}{S_{n-1}\sigma_1}\right].$$

By noticing that

$$\{W_n \le z\} = \left\{ X_n' \ge \frac{1}{\sigma_2}\left[\log\left[(W_{n-1}+S_{n-1}-z)^+\right] - \mu_2\right]\right\},$$

we have

$$\begin{aligned}
&\mathbb{E}\left[G_1(X_{n-1};z)\right] \\
&\quad = -\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{W_n \le z\}\frac{X_{n-1}+\sigma_1}{S_{n-1}\sigma_1^2}\middle| W_{n-1}, X_{n-1}\right]\right] \\
&\quad = -\mathbb{E}\left[\left(1-\Phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right)\right. \\
&\qquad\qquad \left.\times\frac{X_{n-1}+\sigma_1}{S_{n-1}\sigma_1^2}\right],
\end{aligned}$$

which offers a conditional GLR estimator for the density, and

$$\begin{aligned}
&\mathbb{E}\left[G_2(X;\theta)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{\sigma_1}\mathbf{1}\{W_n \le z\}\sum_{i=1}^{n-1}X_i\middle| W_{n-1}, X_{n-1},\dots,X_1\right]\right] \\
&= \frac{1}{\sigma_1}\mathbb{E}\left[\left(1-\Phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right)\right. \\
&\qquad\qquad \left.\times\sum_{i=1}^{n-1}X_i\right],
\end{aligned}$$

which offers a conditional GLR estimator for the distribution sensitivity with respect to $\theta$. Furthermore,

$$\begin{aligned}
&\frac{\partial^2 F(z;\theta)}{\partial z\partial\theta} \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{W_n \le z\}\frac{1-\left(\sum_{i=1}^{n-1}X_i\right)(X_{n-1}+\sigma_1)}{S_{n-1}\sigma_1}\right.\right. \\
&\qquad\qquad \left.\left.\middle| W_{n-1}, X_{n-1},\dots,X_1\right]\right] \\
&= \mathbb{E}\left[\left(1-\Phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right)\right. \\
&\qquad\qquad \left.\times\frac{1-\left(\sum_{i=1}^{n-1}X_i\right)(X_{n-1}+\sigma_1)}{S_{n-1}\sigma_1}\right].
\end{aligned}$$

The expression inside the expectation offers a conditional GLR estimator for the second-order distribution sensitivity.

### 4.3 Barrier Option

We consider the distribution sensitivity estimation for the payoff of a barrier option. An up-and-out (knockout) barrier option is worthless if the path of the underlying asset exceeds a barrier $L$. The event when the barrier option stays "alive" is $\{\max_{i=1,\dots,n} S_{t_i} < L\}$, where $S_t \in \mathbb{R}$ is the underlying asset price at time $t$. In Figure 4, the price curve on the top breaches the barrier prior to expiration $T$, so that its payoff is zero. Suppose $S_t = S_0\exp\{(r-\sigma^2/2)t + \sigma B_t + \sum_{j=1}^{N(t)} J_j\}$ follows a geometric jump-diffusion process, where $S_0$ is the initial underlying asset price, $r$ is the interest rate, $\sigma$ is the implied volatility, $\{N(t)\}$ is a counting process, and $J_i$, $i \in \mathbb{Z}^+$, are the jump sizes. Let $S_0 = \theta$. Assume discrete monitoring, and $\Delta$ is the step size of the discrete monitoring points $t_i = i\Delta$, $i = 1,\dots,n$, $T = n\Delta$. The barrier option would have a positive payoff if it stays "alive" and at $T$ the price of the underlying asset is above the strike price $K$. For a European barrier option, the payoff is

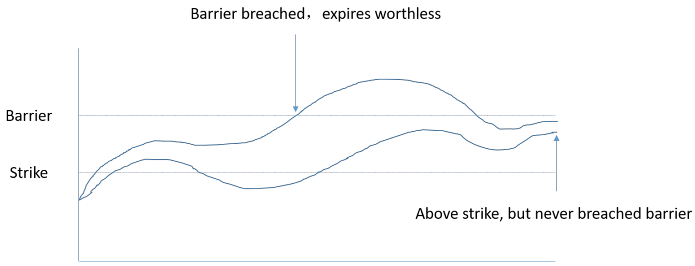



Figure 4: A European up-and-out barrier call option.

$$e^{-rn\Delta}(S_n - K)\prod_{i=1}^{n-1}\mathbf{1}\{S_i \le L\}\mathbf{1}\{K < S_n < L\},$$

where

$$S_i = \theta\exp\left(i\left(r-\frac{\sigma^2}{2}\right)\Delta + \sigma\sqrt{\Delta}\sum_{j=1}^{i}X_j + \sum_{\ell=1}^{N_i}J_{i,\ell}\right),$$

where $X_i = (B_{i\Delta} - B_{(i-1)\Delta})/\sqrt{\Delta}$, $i = 1,\dots,n$, which are i.i.d. standard normal random variables, $N_i = N(i\Delta) - N(i(\Delta-1))$, and $J_{i,\ell}$ is the $\ell$-th jump in the $i$th period.

For $0 \le z \le e^{-rn\Delta}(L-K)$, we can represent the distribution function by

$$\begin{aligned}
F(z;\theta) &= \mathbb{E}\left[\prod_{i=1}^{n-1}\mathbf{1}\{g_i(X;\theta)\le 0\}\mathbf{1}\{\tilde{g}_n(X;\theta,z)\le 0\}\right. \\
&\qquad \left.+\sum_{i=1}^{n}\prod_{j=1}^{i-1}\mathbf{1}\{g_j(X;\theta)\le 0\}\mathbf{1}\{g_i(X;\theta)>0\}\right] \\
&= 1-\mathbb{E}\left[\prod_{i=1}^{n-1}\mathbf{1}\{g_i(X;\theta)\le 0\}\left(\mathbf{1}\{g_n(X;\theta)\le 0\}\right.\right. \\
&\qquad \left.\left.-\mathbf{1}\{\tilde{g}_n(X;\theta,z)\le 0\}\right)\right],
\end{aligned}$$

where

$$\begin{aligned}
g_i(X;\theta) &= \log\theta + i\left(r-\frac{\sigma^2}{2}\right)\Delta + \sigma\sqrt{\Delta}\sum_{j=1}^{i}X_j \\
&\quad + \sum_{j=1}^{i}\sum_{\ell=1}^{N_j}J_{j,\ell} - \log L, \quad i = 1,\dots,n, \\
\tilde{g}_n(X;\theta,z) &= g_n(X;\theta) + \log\left(\frac{L}{e^{rn\Delta}z+K}\right).
\end{aligned}$$

From [42], the GLR estimators for $f(z;\theta)$ and $\partial F(z;\theta)/\partial\theta$ are given respectively by

$$G_1(X;z) = -\prod_{i=1}^{n-1} \mathbf{1}\{g_i(X;\theta)\le 0\}\mathbf{1}\left\{\tilde{g}_n(X;\theta,z)\le 0\right\} \times \frac{e^{rn\Delta}X_n}{\sigma\sqrt{\Delta}(e^{rn\Delta}z+K)},$$

and

$$G_2(X;\theta) = -\prod_{i=1}^{n-1} \mathbf{1}\{g_i(X;\theta)\le 0\} \times \left(\mathbf{1}\left\{g_n(X;\theta)\le 0\right\} - \mathbf{1}\left\{\tilde{g}_n(X;\theta,z)\le 0\right\}\right)\frac{X_1}{\theta\sigma\sqrt{\Delta}}.$$

Define

$$T_i^{(l)} = \frac{1}{\sigma\sqrt{\Delta}}\left[\log(L/\theta) - i\left(r-\frac{\sigma^2}{2}\right)\Delta - \sum_{j=1}^{i}\sum_{\ell=1}^{N_j} J_{j,\ell}\right] - \sum_{j=1,j\neq l}^{i} X_j, \quad i=1,\dots,n,$$

$$\widehat{T}_n^{(l)} = \frac{1}{\sigma\sqrt{\Delta}}\left[\log[(e^{rm\Delta}z+K)/\theta] - n\left(r-\frac{\sigma^2}{2}\right)\Delta - \sum_{j=1}^{n}\sum_{\ell=1}^{N_j} J_{j,\ell}\right] - \sum_{j=1,j\neq l}^{n} X_j.$$

We have

$$\mathbb{E}[\mathbb{E}[G_1(X;z)|X_2,\dots,X_n]] = -\frac{e^{rn\Delta}}{\sigma\sqrt{\Delta}(e^{rn\Delta}z+K)} \times\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X_1 \le \min\left(\widehat{T}_n^{(1)}, \min_{i=1,\dots,n-1} T_i^{(1)}\right)\right\}X_n\middle| X_2,\dots,X_n\right]\right] = -\frac{e^{rn\Delta}}{\sigma\sqrt{\Delta}(e^{rn\Delta}z+K)}\mathbb{E}\left[\Phi\left(\min\left(\widehat{T}_n^{(1)}, \min_{i=1,\dots,n-1} T_i^{(1)}\right)\right)X_n\right],$$

and

$$\begin{aligned}\mathbb{E}[\mathbb{E}[G_2(X;\theta)|X_2,\dots,X_n]] &= -\frac{1}{\theta\sigma\sqrt{\Delta}}\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{\min\left(\widehat{T}_n^{(1)}, \min_{i=1,\dots,n-1} T_i^{(1)}\right) < X_1 \le \min_{i=1,\dots,n} T_i^{(1)}\right\}X_1\middle| X_2,\dots,X_n\right]\right] \\ &= -\frac{1}{\theta\sigma\sqrt{\Delta}}\mathbb{E}\left[\int_{\min\left(\widehat{T}_n^{(1)},\min_{i=1,\dots,n-1} T_i^{(1)}\right)}^{\min_{j=1,\dots,n} T_j^{(1)}} x\phi(x)dx\right].\end{aligned}$$

To avoid integration in the estimator, we can also condition on other random variables, which gives a less smooth estimator than that derived above:

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[G_2(X;\theta)|X_1,X_3,\dots,X_n]] \\ &= -\frac{1}{\theta\sigma\sqrt{\Delta}}\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X_1\le T_1^{(1)}\right\} \times\mathbf{1}\left\{\min\left(\widehat{T}_n^{(2)}, \min_{i=2,\dots,n-1} T_i^{(2)}\right) < X_2 \le \min_{i=2,\dots,n} T_i^{(2)}\right\} \times X_1\middle| X_1,X_3,\dots,X_n\right]\right] \\ &= -\frac{1}{\theta\sigma\sqrt{\Delta}}\mathbb{E}\left[\mathbf{1}\left\{X_1\le T_1^{(1)}\right\}X_1 \times\left[\Phi\left(\min_{i=2,\dots,n-1} T_i^{(2)}\right) - \Phi\left(\min\left(\widehat{T}_n^{(2)}, \min_{i=2,\dots,n-1} T_i^{(2)}\right)\right)\right]\right].\end{aligned}$$

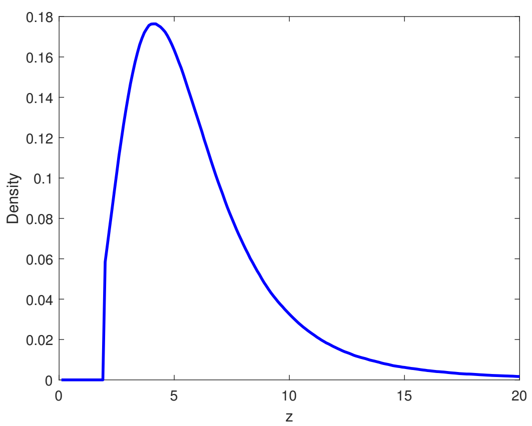
# 5 Numerical Experiments

We report numerical experiments on the three applications discussed in the previous section. GLR is compared with the finite difference method $(Y(\theta+\delta)-Y(\theta))/\delta$ using common random numbers (FDC($\delta$)) to generate $Y(\theta+\delta)$ and $Y(\theta)$. The GLR method together with CMC is called conditional GLR (CGLR), and CGLR together with RQMC is denoted as CGLRQ. Both CGLR and CGLRQ are compared with SPA and SPA when combined with RQMC (SPAQ). For RQMC, we use the Sobol sequence scrambled by the algorithm of [37] in Matlab. Derivations of SPA for the density estimation in three applications are similar to those in [34] and can be found in the appendix. We set the sample size as $m=2^{13}$ for the standard Monte Carlo and RQMC estimators, and their variances are estimated by $l=100$ independent experiments.

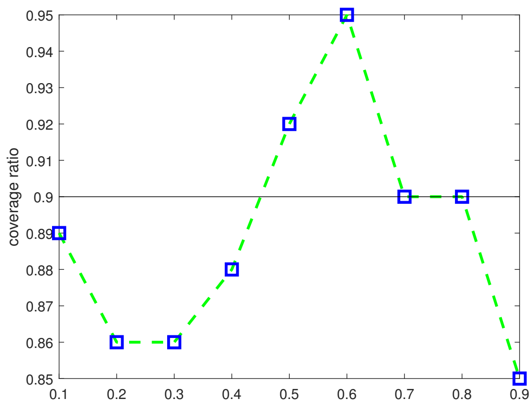
## 5.1 Stochastic Activity Network

The parameters for the stochastic models described in Section 4.1 are set as $\lambda_i=1$, $i=1,2,3$, $\mu_j=0$ and $\sigma_j=1$, $j=4,5,6$. We estimate the density function $\partial F(z)/\partial z$ for $z\in(0,20)$. Figure 5(a) shows the density curve estimated by GLR, and Figure 5(b) presents the coverage rate curve of the 90% confidence intervals based on the asymptotic normality result given by (5) for $\alpha$-quantiles with $\alpha=0.1\times i$, $i=1,\dots,9$, by $m=2^{13}$ samples and $l=100$ independent experiments.

Figure 6(a) shows a variance comparison between the two GLR estimators in Section 4.1. GLR-1 is derived using a stochastic model with uniform random numbers as inputs, and GLR-2 is derived using a stochastic model with normal random variables as inputs. Since the variance of GLR-1 is much smaller than GLR-2, GLR-1 is used for this example in the rest of the experiments, and we simply denote the method by GLR. Figure 6(b) shows that the variance of GLR is much smaller than FDC(0.01) and FDC(0.1), and the variance of FDC increases significantly when the perturbation size becomes smaller. The reason FDC works poorly is that the sample performance is discontinuous. If CMC is first applied to smooth the sample performance, then conditional FDC can achieve a comparable variance to SPA [33].

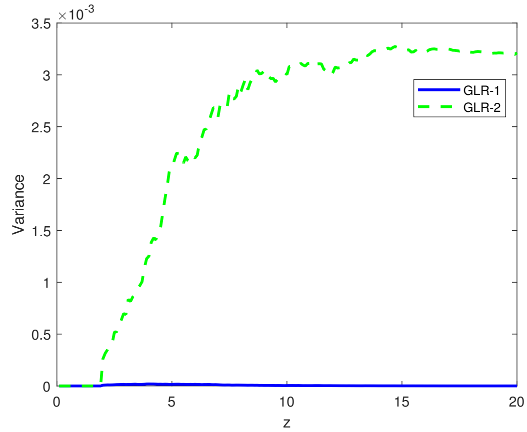



(a) Function curve of $\partial F(z)/\partial z$.

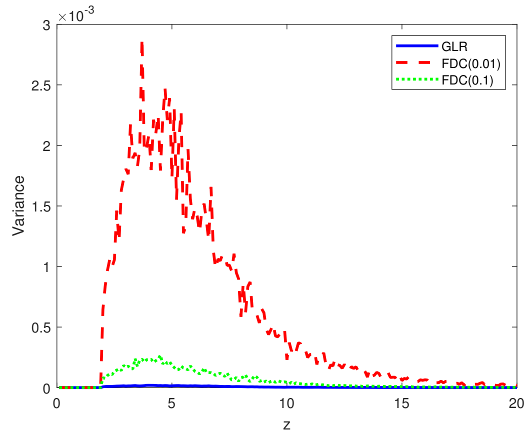



(b) Coverage rate curve.

Figure 5: Density estimation by GLR and coverage rates of 90% confidence intervals for $\alpha$-quantiles in the SAN example.




(a) Variance curves of GLR.




(b) Variance curves of GLR and FDC.

Figure 6: Variances of density estimation by GLR and FDC in the SAN example.

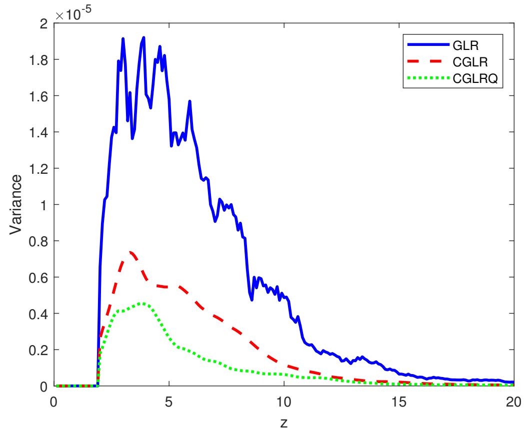
Figure 7(a) shows a variance comparison between GLR, CGLR, and CGLRQ. We can see that both CMC and RQMC substantially reduce the variance of GLR. From Figure 7(b), RQMC also reduces the variance of SPA, and CGLRQ and SPAQ achieve comparable magnitudes of variance.

## 5.2 Single-Server Queue

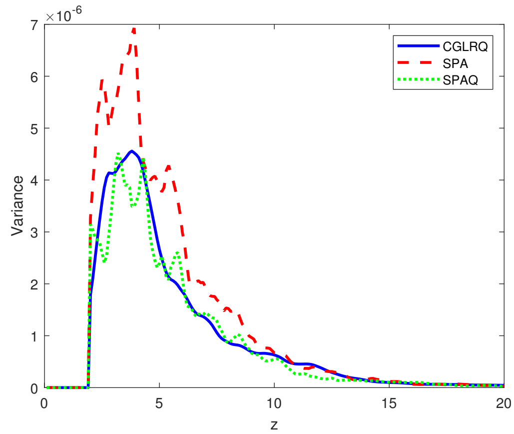
The parameters for the stochastic models described in Section 4.2 are set as $\mu_i = 0$ and $\sigma_i = 1$, $j = 4, 5, 6$, and $n = 10$. We estimate distribution sensitivities $\partial F(z;\theta)/\partial z$, $\partial F(z;\theta)/\partial\theta$, and $\partial^2 F(z;\theta)/\partial z\partial\theta$ for $z \in (0, 30)$. Figure 8 presents the three distribution sensitivity curves estimated by GLR.

From Figure 9(a), FDC(0.01) has a much larger variance than GLR for estimating the second-order distribution sensitivity $\partial^2 F(z;\theta)/\partial z\partial\theta$, and FDC suffers from a bias-variance tradeoff. In this example, the variances of GLR, CGLR, and CGLRQ cannot be distinguished statistically. Derivation of SPA for estimating the second-order distribution sensitivity is not straightforward due to the discontinuities in the SPA estimator for the first-order derivative (see the appendix).

From Figure 10(a), we can see that GLR and CGLR achieve comparable performance, whereas the variance of CGLRQ is much smaller than GLR and CGLR. The peak value on the variance curve of CGLRQ is about 1/6 of that of GLR and CGLR. Figure 10(b) shows that SPA and SPAQ have comparable variance, and the variance of CGLRQ is smaller than SPA and SPAQ when $z \leq 15$. The variance of CGLRQ becomes larger than SPA and SPAQ when $z > 15$, because the SPA and SPAQ estimators go to zero as $z \to \infty$, whereas the CGLRQ estimator converges to a random variable with a constant variance as $z \to \infty$.




(a) Variance curves of GLR.




(b) Variance curves of GLR and FDC.

Figure 7: Variances of density estimation by GLR, CGLR, CGLRQ, SPA, and SPAQ in the SAN example.

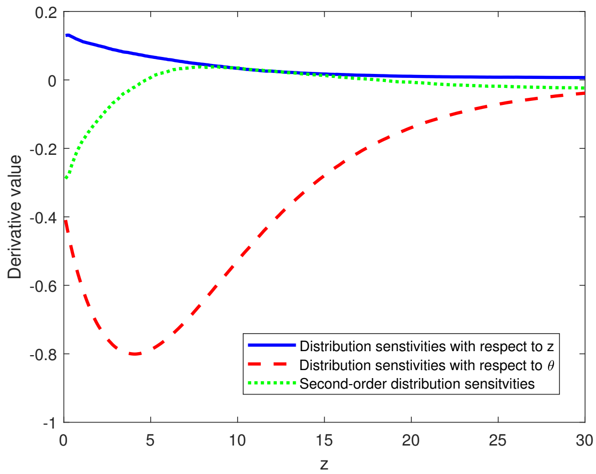



Figure 8: Estimation of distribution sensitivities $\partial F(z;\theta)/\partial z$, $\partial F(z;\theta)/\partial \theta$, and $\partial^2 F(z;\theta)/\partial z \partial \theta$ by GLR in the queueing example.

### 5.3 Barrier Option

The parameters for the stochastic models described in Section 4.3 are set as $T = 10$, $n = 10$, $\sigma = 1$, $r = 0.001$, $\Delta = 1$, $K = 100$, $L = 120$, and $S_0 = 100$. The counting process is assumed to be a Poisson process with intensity 1, and the jump size is assumed to follow a normal distribution with mean 0 and standard deviation 0.01. We estimate distribution sensitivities $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial \theta$ for $z \in (0, 30)$. To estimate $\partial F(z;\theta)/\partial \theta$, we apply the GLR estimator in Section 4.3 that does not involve integration. Figure 11 presents the two distribution sensitivity curves estimated by GLR.

Figure 12 presents the variance comparison between GLR and FDC for two first-order distribution sensitivities. FDC(0.01) has a much larger variance than GLR. The variances of GLR and FDC(0.1) are comparable, but GLR is unbiased and FDC(0.1) is biased. The bias estimate of FDC(0.01) and FDC(0.1) can be found in the appendix.

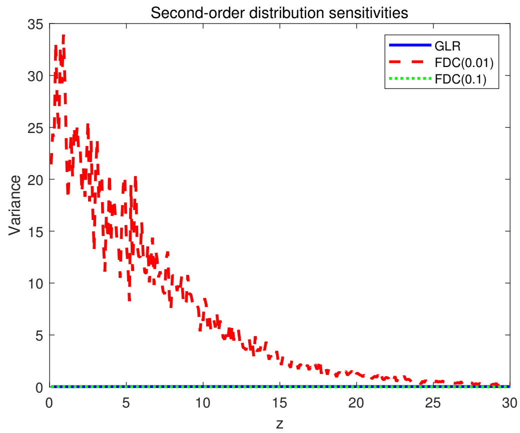
Figure 13 presents the variance comparison between GLR, CGLR, and CGLRQ for two first-order distribution sensitivities. The variance of CGLR is slightly smaller than that of GLR, whereas the variance of CGLRQ is substantially smaller than those of CGLR and GLR. The peak value on the variance curve of CGLRQ is about 1/3 of that of GLR and CGLR.

## 6 Conclusions

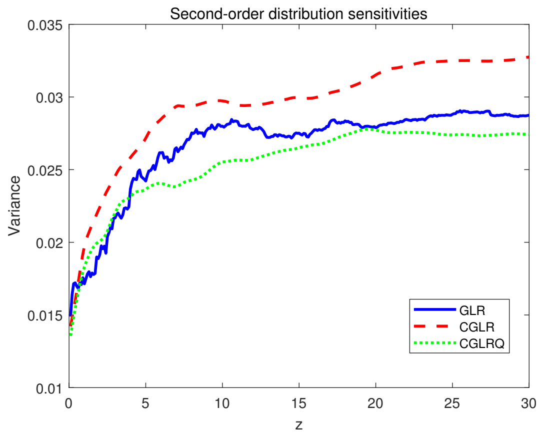
In this paper, we have studied the GLR methods in [42] and [41] under a single framework, simplified regularity conditions for justifying unbiasedness of GLR in [42] and relaxed some of the conditions that are difficult to verify in practice by adapting a similar technique developed in [41]. Moreover, we have discussed how to combine GLR with CMC and RQMC to reduce the variance of the resultant estimators, and applied them to estimate the distribution sensitivities for a SAN, a single-server queue, and a barrier option. Numerical results show that CMC and RQMC may reduce the variance of GLR substantially, and CGLRQ sometimes achieves even smaller variance than SPAQ.

## Acknowledgement

This document is the results of the research project funded by National Natural Science Foundation of China (NSFC) under Grants 71901003 and 72022001, by the Air Force Office of Scientific Research under Grant FA95502010211, by Discovery Grant RGPIN-2018-05795 from NSERC-Canada.




(a) Variance curves of GLR and FDC.




(b) Variance curves of GLR, CGLR, and CGLRQ.

Figure 9: Variances of estimation for $\partial^2 F(z;\theta)/\partial z\partial\theta$ by GLR, CGLR, CGLRQ and FDC in the queueing example.

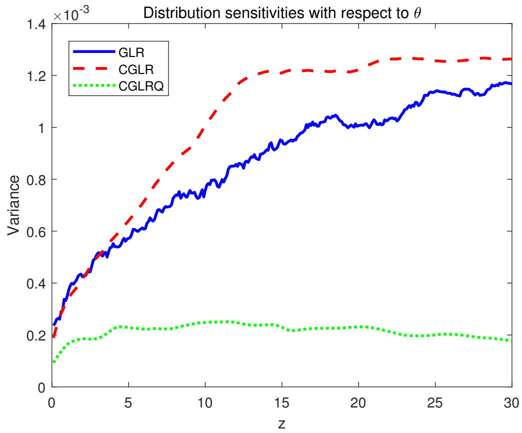
# A Appendix

**Proof of Theorem 1**

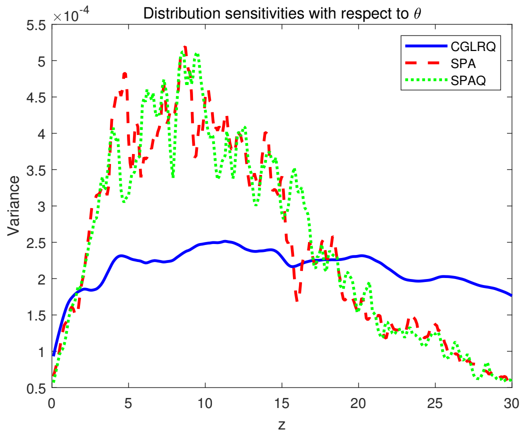
*Proof.* Proof. As in [42], define a sequence of bounded functions $\varphi_L(x) = \max\{\min\{\varphi(x), L\}, -L\}$, and then $|\varphi_L(x)| \leq \varphi(x)$ and $\lim_{L\to\infty}\varphi_L(x) = \varphi(x)$. From Theorem 1 in [42], there exists a sequence of bounded and smooth functions $\varphi_{\epsilon,L}(\cdot)$ such that

$$\lim_{L\to\infty} \|\varphi_{\epsilon,L} - \varphi_L\|_p = 0,$$

where $p > 1$, and $\|h\|_p := \left(\int_{\mathbb{R}^n} |h(x)|^p dx\right)^{1/p}$. Since $g(x;\theta)$ is an invertible function vector, its Jacobian is an invertible




(a) Variance curves of GLR and FDC.




(b) Variance curves of GLR, CGLR, and CGLRQ.

Figure 10: Variances of estimation for $\partial F(z;\theta)/\partial\theta$ by GLR, CGLR, CGLRQ, SPA and SPAQ in the queueing system example.

matrix. Then we have

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{[-M,M]^n} \varphi_{\epsilon,L}(g(x;\theta)) f_X(x;\theta)dx \\
=& \int_{[-M,M]^n}\Bigg(\varphi_{\epsilon,L}(g(x;\theta))\frac{\partial \log f(x;\theta)}{\partial\theta} \\
&\qquad +\nabla_y\varphi_{\epsilon,L}(y)|_{y=g(x;\theta)}\partial_\theta g(x;\theta)\Bigg) f_X(x;\theta)dx \\
=& \int_{[-M,M]^n}\Bigg(\varphi_{\epsilon,L}(g(x;\theta))\frac{\partial \log f(x;\theta)}{\partial\theta} \\
&\qquad +\nabla_x\varphi_{\epsilon,L}(g(x;\theta))J_g^{-1}(x;\theta)\partial_\theta g(x;\theta)\Bigg) f_X(x;\theta)dx.
\end{aligned}$$

The interchange of the differentiation and integration can be justified by the dominated convergence theory. Under condition (A.1), $f_X(x;\theta)$ is continuously differentiable and $g(x;\theta)$ is twice continuously differentiable in $\mathbb{R}^n\times\Theta$, so their function values and derivatives are bounded in a compact

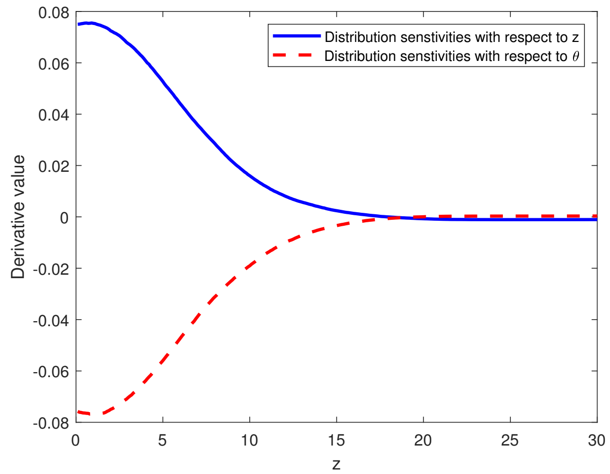



Figure 11: Estimation of distribution sensitivities $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial\theta$ by GLR in the barrier option example.

space $[-M,M]^n \times \Theta$. By the Gauss-Green Theorem,

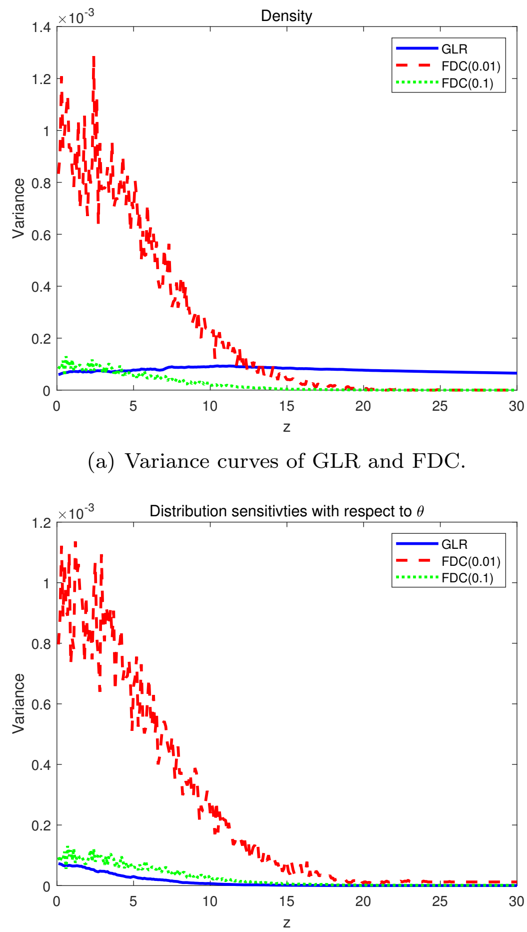
$$\begin{aligned}
&\int_{[-M,M]^n} \nabla_x \varphi_{\epsilon,L}(g(x;\theta)) J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) \\
&\qquad\qquad \times f_X(x;\theta)dx_1\cdots dx_n \\
=&\sum_{i=1}^n \int_{[-M,M]^{n-1}} \varphi_{\epsilon,L}(g(u;\theta)) \left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta)\right)^T \\
&\qquad\qquad \times e_i f_X(x;\theta) \prod_{j\neq i} dx_j \Big|_{x_i=-M}^{M} \\
&-\int_{[-M,M]^n} \varphi_{\epsilon,L}(g(x;\theta)) \\
&\times \operatorname{div}\left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) f_X(x;\theta)\right) dx_1\cdots dx_n,
\end{aligned}$$

where for $h(x) = (h_1(x),\ldots,h_n(x))$,

$$\operatorname{div}(h(x)) := \sum_{i=1}^n \frac{\partial h(x)}{\partial x_i}.$$

Then

$$\begin{aligned}
&\operatorname{div}\left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) f_X(x;\theta)\right) \\
&=\sum_{i=1}^n \frac{\partial}{\partial x_i} e_i^T J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) f_X(x;\theta) \\
&=\left(\sum_{i=1}^n e_i^T \left(\partial_{x_i} J_g^{-1}(x;\theta)\right)\partial_\theta g(x;\theta)\right. \\
&\quad \left.+\operatorname{trace}(J_g^{-1}(x;\theta)\partial_\theta J_g(x;\theta))\right) f_X(x;\theta) \\
&\quad +\sum_{i=1}^n \left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta)\right)^T e_i \frac{\partial f_X(x;\theta)}{\partial x_i}.
\end{aligned}$$




(a) Variance curves of GLR and FDC.

(b) Variance curves of GLR, CGLR, and CGLRQ.

Figure 12: Variances of estimation for $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial\theta$ by GLR and FDC in the barrier option example.

By differentiating equation $J_g^{-1}(x;\theta)J_g(x;\theta) = I$ with respect to $x_i$ on both sides, we have

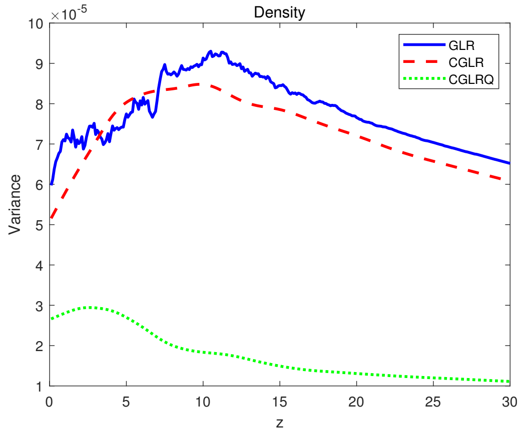
$$\begin{aligned}
0 =&\partial_{x_i}\left(J_g^{-1}(x;\theta)J_g(x;\theta)\right) \\
=&\partial_{x_i} J_g^{-1}(x;\theta)J_g(x;\theta) + J_g^{-1}(x;\theta)\partial_{x_i} J_g(x;\theta),
\end{aligned}$$

which leads to

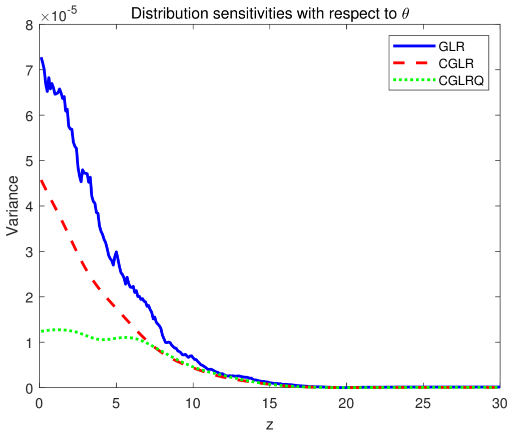
$$\partial_{x_i} J_g^{-1}(x;\theta) = -J_g^{-1}(x;\theta)\left(\partial_{x_i} J_g(x;\theta)\right) J_g^{-1}(x;\theta).$$

Therefore, we have

$$d(x;\theta) = -\operatorname{div}\left(J_g^{-1}(x;\theta)\partial_\theta g(x;\theta) f_X(x;\theta)\right)/f_X(x;\theta).$$




(a) Variance curves of GLR, CGLR, and CGLRQ.




(b) Variance curves of GLR, CGLR, and CGLRQ.

Figure 13: Variances of estimation for $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial\theta$ by GLR, CGLR, and CGLRQ in the barrier option example.

With the discussion above,

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{[-M,M]^n}\varphi_{\epsilon,L}(g(x;\theta))f_X(x;\theta)dx\\&=\sum_{i=1}^{n}\int_{[-M,M]^{n-1}}\varphi_{\epsilon,L}(g(x;\theta))r_i(x;\theta)f_X(x;\theta)\prod_{j\neq i}dx_j\Big|_{x_i=-M}^{M}\\&+\int_{[-M,M]^n}\varphi_\epsilon(g(x;\theta))w(x;\theta)f_X(x;\theta)dx.\end{aligned}$$

Under condition (A.2),

$$\begin{aligned}\lim_{\epsilon\to0}\lim_{M\to\infty}\sup_{\theta\in\Theta}\Bigg|\sum_{i=1}^{n}\int_{[-M,M]^{n-1}}&\varphi_{\epsilon,L}(g(x;\theta))r_i(x;\theta)\\&\times f_X(x;\theta)\prod_{j\neq i}dx_j\Big|_{x_i=-M}^{M}\Bigg|=0.\end{aligned}$$

By change of variables and Hölder's inequality,

$$\begin{aligned}&\lim_{\epsilon\to0}\sup_{\theta\in\Theta}\Bigg|\int_{[-M,M]^n}(\varphi_{\epsilon,L}(g(x;\theta))-\varphi_L(g(x;\theta)))\\&\qquad\times w(x;\theta)f_X(x;\theta)dx\Bigg|\\&=\lim_{\epsilon\to0}\sup_{\theta\in\Theta}\Bigg|\int_{\mathcal{S}_M}(\varphi_{\epsilon,L}(y)-\varphi_L(y))|\det(J_g(x;\theta))|w(x;\theta)\\&\qquad\times f_X(x;\theta)|_{x=g^{-1}(y;\theta)}dy\Bigg|\\&\le\lim_{\epsilon\to0}\|\varphi_{\epsilon,L}-\varphi_L\|_p\\&\times\sup_{\theta\in\Theta}\Bigg|\int_{\mathcal{S}_M}|\det(J_g(x;\theta))w(x;\theta)f_X(x;\theta)|^q_{x=g^{-1}(y;\theta)}dy\Bigg|^{1/q}=0,\end{aligned}\tag{6}$$

where

$$\mathcal{S}_M:=\{y\in\mathbb{R}^n:\ y=g(x;\theta),\ x\in[-M,M]^n\}.$$

With condition (A.3),

$$\begin{aligned}&\lim_{M\to\infty}\sup_{\theta\in\Theta}\Bigg|\int_{[-M,M]^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\\&\qquad-\int_{(0,1)^n}\varphi_L(g(u;\theta))w(x;\theta)f_X(x;\theta)dx\Bigg|\\&\le\lim_{M\to\infty}\int_{(0,1)^n\setminus[-M,M]^n}\sup_{\theta\in\Theta}\big|\varphi(g(x;\theta))w(x;\theta)f_X(x;\theta)\big|dx=0.\end{aligned}\tag{7}$$

Then we have

$$\begin{aligned}&\lim_{M\to\infty}\lim_{\epsilon\to0}\sup_{\theta\in\Theta}\Bigg|\int_{[\delta,1-\delta]^n}\varphi_{\epsilon,L}(g(u;\theta))w(x;\theta)f_X(x;\theta)dx\\&\qquad-\int_{\mathbb{R}^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\Bigg|\\&\le\lim_{M\to\infty}\lim_{\epsilon\to0}\sup_{\theta\in\Theta}\Bigg|\int_{[-M,M]^n}\varphi_{\epsilon,L}(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\\&\qquad-\int_{[-M,M]^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\Bigg|\\&+\lim_{M\to\infty}\sup_{\theta\in\Theta}\Bigg|\int_{[-M,M]^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\\&\qquad-\int_{\mathbb{R}^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\Bigg|=0,\end{aligned}$$

where the first term goes to zero because of (6) and the second term goes to zero because of (7). From [45], $\frac{d}{d\theta}\lim_{\epsilon\to0}h_\epsilon(\theta)=\lim_{\epsilon\to0}h'_\epsilon(\theta)$ holds if $h'_\epsilon(\theta)$ converges uniformly with respect to $\theta\in\Theta$ as $\epsilon\to0$. Therefore,

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}\varphi_L(g(x;\theta))f_X(x;\theta)dx\\&=\lim_{M\to\infty}\lim_{\epsilon\to0}\frac{\partial}{\partial\theta}\int_{[-M,M]^n}\varphi_L(g(x;\theta))f_X(x;\theta)dx\\&=\int_{\mathbb{R}^n}\varphi_L(g(u;\theta))w(x;\theta)f_X(x;\theta)du.\end{aligned}$$

With condition (A.3) and noticing that $|\varphi_L(x) - \varphi(x)| \leq \varphi(x)$,

$$\lim_{L\to\infty}\sup_{\theta\in\Theta}\left|\int_{\mathbb{R}^n}(\varphi_L(g(x;\theta)) - \varphi(g(x;\theta))) \times w(x;\theta)f_X(x;\theta)dx\right| = 0.$$

By this uniform convergence,

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}\varphi(g(x;\theta))f_X(x;\theta)dx\\ &= \frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}\lim_{L\to\infty}\varphi_L(g(x;\theta))f_X(x;\theta)dx\\ &= \lim_{L\to\infty}\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}\varphi_L(g(x;\theta))f_X(x;\theta)dx\\ &= \lim_{L\to\infty}\int_{\mathbb{R}^n}\varphi_L(g(x;\theta))w(x;\theta)f_X(x;\theta)dx\\ &= \int_{\mathbb{R}^n}\varphi(g(x;\theta))w(x;\theta)f_X(x;\theta)dx,\end{aligned}$$

which proves the theorem. □

**SPA for SAN**

There are many possible choices of conditioning random variables to derive SPA for density estimation, and we only use one possible choice. Certain conditions are needed to ensure unbiasedness for SPA. More general discussions of SPA for density estimation can be found in [34]. The SPA estimator can be obtained by

$$\begin{aligned}&\frac{\partial}{\partial z}\mathbb{E}[\mathbf{1}\{Y\leq z\}] = \frac{\partial}{\partial z}\mathbb{E}[\mathbb{E}[\mathbf{1}\{Y\leq z\}|Y_1,\dots,Y_5]]\\ &= \frac{\partial}{\partial z}\mathbb{E}\left[\Phi\left(\frac{1}{\sigma_6}\left(\log[(z-\widetilde{Y})^+] - \mu_6\right)\right)\right]\\ &= \mathbb{E}\left[\frac{\mathbf{1}\{z\geq\widetilde{Y}\}}{\sigma_6(z-\widetilde{Y})^+}\phi\left(\frac{1}{\sigma_6}\left(\log[(z-\widetilde{Y})^+]-\mu_6\right)\right)\right].\end{aligned}$$

**SPA for Single-Sever Queue**

The SPA estimator for the density is

$$\begin{aligned}\frac{\partial F(z;\theta)}{\partial z} &= \frac{\partial}{\partial z}\mathbb{E}[\mathbb{E}[\mathbf{1}\{W_n\leq z\}|W_{n-1},X_{n-1},\dots,X_1]]\\ &= \frac{\partial}{\partial z}\mathbb{E}\left[\left(1-\Phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right)\right]\\ &= \frac{1}{\sigma_2}\mathbb{E}\left[\frac{\mathbf{1}\{W_{n-1}+S_{n-1}\geq z\}}{(W_{n-1}+S_{n-1}-z)^+}\right.\\ &\quad\left.\times\phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right],\end{aligned}$$

and the SPA estimator for $\partial F(z;\theta)/\partial\theta$ is

$$\begin{aligned}&\frac{\partial F(z;\theta)}{\partial\theta}\\ &= \frac{\partial}{\partial\theta}\mathbb{E}\left[\left(1-\Phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\right)\right]\\ &= -\frac{1}{\sigma_2}\mathbb{E}\left[\mathbf{1}\{W_{n-1}+S_{n-1}\geq z\}\right.\\ &\left.\times\phi\left(\frac{1}{\sigma_2}\left(\log\left[(W_{n-1}+S_{n-1}-z)^+\right]-\mu_2\right)\right)\frac{\frac{\partial W_{n-1}}{\partial\theta}+\frac{\partial S_{n-1}}{\partial\theta}}{W_{n-1}+S_{n-1}-z}\right],\end{aligned}$$

where

$$\begin{aligned}\frac{\partial W_i}{\partial\theta} &= \mathbf{1}\{W_{i-1}+S_{i-1}\geq A_i\}\left(\frac{\partial W_{i-1}}{\partial\theta}+\frac{\partial S_{i-1}}{\partial\theta}\right), \quad i\geq 2,\\ \frac{\partial S_i}{\partial\theta} &= S_i, \quad i\geq 1.\end{aligned}$$

**SPA for Barrier Option**

To derive the SPA estimators for $f(z;\theta)$ and $\partial F(z;\theta)/\partial\theta$,

$$\begin{aligned}F(z;\theta) &= 1-\mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{n-1}\mathbf{1}\{g_i(X;\theta)\leq 0\}\right.\right.\\ &\left.\left.\times\left(\mathbf{1}\left\{g_n(X;\theta)\leq 0\right\}-\mathbf{1}\left\{\tilde{g}_n(X;\theta,z)\leq 0\right\}\right)|X_2,\dots,X_n\right]\right]\\ &= 1-\mathbb{E}\left[\Phi\left(\min_{i=1,\dots,n}T_i^{(1)}\right)-\Phi\left(\min\left(\widehat{T}_n^{(1)},\min_{i=1,\dots,n-1}T_i^{(1)}\right)\right)\right],\end{aligned}$$

and then we have

$$\begin{aligned}\frac{\partial F(z;\theta)}{\partial z} &= \frac{e^{rn\Delta}}{\sigma\sqrt{\Delta}(e^{rn\Delta}z+K)}\mathbb{E}\left[\phi\left(\min\left(\widehat{T}_n^{(1)},\min_{i=1,\dots,n-1}T_i^{(1)}\right)\right)\right.\\ &\quad\left.\times\mathbf{1}\left\{\widehat{T}_n^{(1)}<\min_{i=1,\dots,n-1}T_i^{(1)}\right\}\right],\\ \frac{\partial F(z;\theta)}{\partial\theta} &= \frac{1}{\theta\sigma\sqrt{\Delta}}\mathbb{E}\left[\phi\left(\min_{i=1,\dots,n}T_i^{(1)}\right)\right.\\ &\quad\left.-\phi\left(\min\left(\widehat{T}_n^{(1)},\min_{i=1,\dots,n-1}T_i^{(1)}\right)\right)\right].\end{aligned}$$

Figure 14 presents the variance comparison between CGLRQ, SPA, and SPAQ for estimating $\partial F(z;\theta)/\partial\theta$ in the barrier option example. We can see that SPA and SPAQ have a smaller variance than CGLRQ.

Figure 15 presents the bias of FDC(0.01) and FDC(0.1) for $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial\theta$ in the barrier option example. The true distribution sensitivities are estimated by SPAQ with the sample size as $m = 2^{13}$ and $l = 10^4$ independent experiments. Bias of FDC are also estimated with the sample size as $m = 2^{13}$ and $l = 10^4$ independent experiments. From Figure 15, we can see that FDC(0.01) has a smaller bias than FDC(0.1).

# References


[1] Asmussen, S., Glynn, P.W., 2007. Stochastic Simulation: Algorithms and Analysis. Springer.

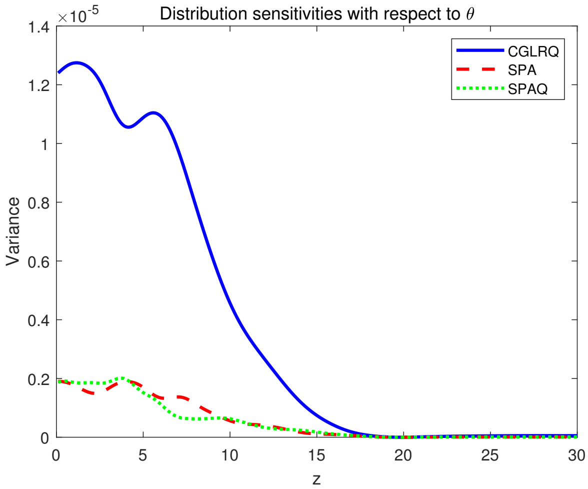
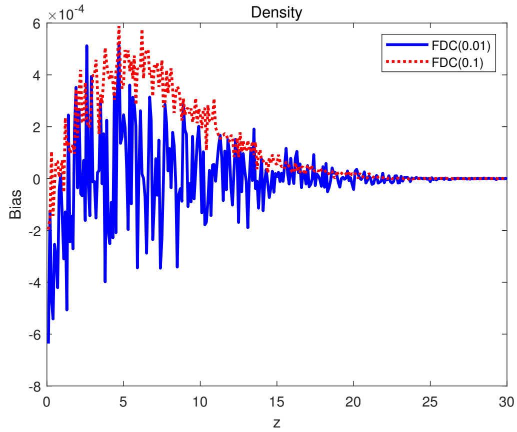
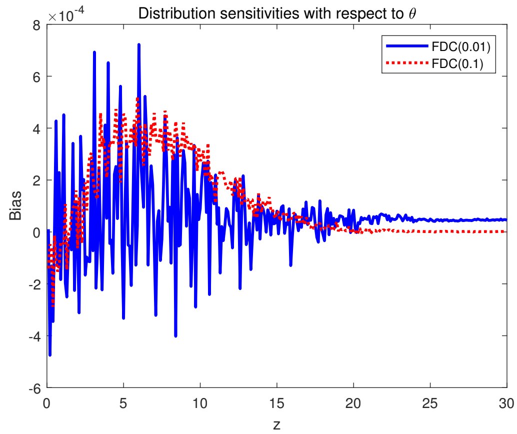



Figure 14: Variances of estimation for $\partial F(z;\theta)/\partial\theta$ by CGLRQ, SPA and SPAQ in the barrier option example.




(a) Bias curves of FDC for $\partial F(z;\theta)/\partial z$.




(b) Bias curves of FDC for $\partial F(z;\theta)/\partial\theta$.

Figure 15: Bias of estimation for $\partial F(z;\theta)/\partial z$ and $\partial F(z;\theta)/\partial\theta$ by FDC(0.01) and FDC(0.1) in the barrier option example.


[2] Cassandras, C.G., Lafortune, S., 2008. Introduction to Discrete Event Systems. Springer Science & Business Media.

[3] Chen, N., Liu, Y., 2014. American option sensitivities estimation via a generalized infinitesimal perturbation analysis approach. Operations Research 62, 616–632.

[4] Cui, Z., Fu, M.C., Hu, J.Q., Liu, Y., Peng, Y., Zhu, L., 2020. On the variance of single-run unbiased stochastic derivative estimators. INFORMS Journal on Computing 32, 390–407.

[5] Dick, J., Pillichshammer, F., 2010. Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration. Cambridge University Press.

[6] Fu, M.C., 2006. Sensitivity analysis in monte carlo simulation of stochastic activity networks, in: Perspectives in Operations Research. Springer, pp. 351–366.

[7] Fu, M.C., 2015. Stochastic gradient estimation, Fu, Michael C. ed., in: Handbook of Simulation Optimization, Springer. pp. 105–147. Chapter 5.

[8] Fu, M.C., Hong, L.J., Hu, J.Q., 2009. Conditional Monte Carlo estimation of quantile sensitivities. Management Science 55, 2019–2027.

[9] Fu, M.C., Hu, J.Q., 1997. Conditional Monte Carlo: Gradient Estimation and Optimization Applications. Kluwer Academic Publishers, Boston.

[10] Glasserman, P., 1991. Gradient Estimation via Perturbation Analysis. Kluwer Academic Publishers, Boston.

[11] Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer.

[12] Glynn, P.W., Peng, Y., Fu, M.C., Hu, J., 2021. Computing sensitivities for distortion risk measures. INFORMS Journal on Computing, forthcoming .

[13] He, Z., Wang, X., 2015. On the convergence rate of randomized quasi–Monte Carlo for discontinuous functions. SIAM Journal on Numerical Analysis 53, 2488–2503.

[14] Heidergott, B., Leahu, H., 2010. Weak differentiability of product measures. Mathematics of Operations Research 35, 27–51.

[15] Heidergott, B., Vázquez-Abad, F.J., Volk-Makarewicz, W., 2008. Sensitivity estimation for Gaussian systems. European Journal of Operational Research 187, 193–207.

[16] Heidergott, B., Volk-Makarewicz, W., 2016. A

measure-valued differentiation approach to sensitivity analysis of quantiles. Mathematics of Operations Research 41, 293–317.

[17] Ho, Y.C., Cao, X.R., 1991. Discrete Event Dynamic Systems and Perturbation Analysis. Kluwer Academic Publishers, Boston, MA.

[18] Hong, L.J., 2009. Estimating quantile sensitivities. Operations Research 57, 118–130.

[19] Hong, L.J., Juneja, S., Luo, J., 2014. Estimating sensitivities of portfolio credit risk using Monte Carlo. INFORMS Journal on Computing 26, 848–865.

[20] Hong, L.J., Liu, G., 2009. Simulating sensitivities of conditional value at risk. Management Science 55, 281–293.

[21] Hu, J., Peng, Y., Zhang, G., Zhang, Q., 2021. A stochastic approximation method for simulation-based quantile optimization. submitted to INFORMS Journal on Computing .

[22] Jiang, G., Fu, M.C., 2015. Technical note — On estimating quantile sensitivities via infinitesimal perturbation analysis. Operations Research 63, 435–441.

[23] Kaplan, Z.T., Li, Y., Nakayama, M.K., Tuffin, B., 2019. Randomized quasi-Monte Carlo for quantile estimation. Proceedings of Winter Simulation Conference , 428–439.

[24] L'Ecuyer, P., 1990. A unified view of the IPA, SF, and LR gradient estimation techniques. Management Science 36, 1364–1383.

[25] L'Ecuyer, P., 1995. Note: On the interchange of derivative and expectation for likelihood ratio derivative estimators. Management Science 41, 738–747.

[26] L'Ecuyer, P., 2009. Quasi-Monte Carlo methods with applications in finance. Finance and Stochastics 42, 926–938.

[27] L'Ecuyer, P., 2018. Randomized quasi-Monte Carlo: An introduction for practitioners. Monte Carlo and Quasi-Monte Carlo Methods: MCQMC 2016 (Glynn P. W., Owen A. B. eds.) , 29–52.

[28] L'Ecuyer, P., Lemieux, C., 2000. Variance reduction via lattice rules. Management Science 46, 1214–1235.

[29] L'Ecuyer, P., Lemieux, C., 2002. Recent advances in randomized quasi-Monte Carlo methods. Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications (M. Dror and P. L'Ecuyer and F. Szidarovszky ed.) , 419–474.

[30] L'Ecuyer, P., Munger, D., 2012. Randomized quasi-Monte Carlo: An introduction for practitioners. Monte Carlo and Quasi-Monte Carlo Methods 2010 (Plaskota L., Wozniakowski H. eds.) , 133–159.

[31] L'Ecuyer, P., Munger, D., 2016. Algorithm 958: Lattice builder: A general software tool for constructing rank-1 lattice rules. ACM Transactions on Mathematical Software 42, 1–30.

[32] L'Ecuyer, P., Munger, D., Tuffin, B., 2010. On the distribution of integration error by randomly-shifted lattice rules. Electronic Journal of Statistics 4, 950–993.

[33] L'Ecuyer, P., Perron, G., 1994. On the convergence rates of IPA and FDC derivative estimators. Operations Research 42, 643–656.

[34] L'Ecuyer, P., Puchhammer, F., Ben Abdellah, A., 2019. Monte Carlo and quasi-Monte Carlo density estimation via conditioning. arXiv preprint arXiv:1906.04607 .

[35] Lemieux, C., 2009. Monte Carlo and Quasi-Monte Carlo Method. New York: Springer.

[36] Liu, G., Hong, L.J., 2011. Kernel estimation of the Greeks for options with discontinuous payoffs. Operations Research 59, 96–108.

[37] Matousek, J., 1998. On the $l_2$-discrepancy for anchored boxes. Journal of Complexity 14, 527–556.

[38] Mohamed, S., Rosca, M., Figurnov, M., Mnih, A., 2020. Monte Carlo gradient estimation in machine learning. Journal of Machine Learning Research 21, 1–62.

[39] Nakayama, M.K., Kaplan, Z.T., Li, Y., Tuffin, B., L'Ecuyer, P., 2020. Quantile estimation via a combination of conditional monte carlo and randomized quasi-monte carlo. Proceedings of Winter Simulation Conference , 301–312.

[40] Peng, Y., Fu, M.C., Heidergott, B., Lam, H., 2020. Maximum likelihood estimation by Monte Carlo simulation: Towards data-driven stochastic modeling. Operations Research 68, 1896–1912.

[41] Peng, Y., Fu, M.C., Hu, J., L'Ecuyer, P., Tuffin, B., 2021a. Generalized likelihood ratio method for stochastic models with uniform random numbers as inputs. submittted to Operations Research, preprint in https://hal.inria.fr/hal-02652068/document .

[42] Peng, Y., Fu, M.C., Hu, J.Q., Heidergott, B., 2018. A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. Operations Research 66, 487–499.

[43] Peng, Y., Xiao, L., Heidergott, B., Hong, J., Lam, H., 2021b. A new likelihood ratio method for training artificial neural networks. submitted to INFORMS Journal on Computing, under third review .

[44] Rubinstein, R.Y., Shapiro, A., 1993. Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. Wiley, New York.

[45] Rudin, W., 1964. Principles of Mathematical Analysis. McGraw-Hill Education, New York.

[46] Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley.

[47] Wang, Y., Fu, M.C., Marcus, S.I., 2012. A new stochastic derivative estimator for discontinuous payoff functions with application to financial derivatives. Operations Research 60, 447–460.