Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs

Yijie Peng[a,b], Michael C. Fu[c], Jiaqiao Hu[d], Pierre L'Ecuyer[e], Bruno Tuffin[f]

[a] *Wuhan Institute for Artificial Intelligence, Guanghua School of Management, Peking University, Beijing, China*
[b] *Xijiang Laboratory, Changsha, China*
[c] *Robert H. Smith School of Business, University of Maryland, College Park, USA*
[d] *Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, USA*
[e] *Department of Computer Science and Operations Research, University of Montreal, Montreal, Canada*
[f] *INRIA, Univ Rennes, CNRS, IRISA, Campus de Beaulieu, Rennes Cedex, France*

Abstract

We propose a new unbiased stochastic gradient estimator for a family of stochastic models driven by uniform random numbers as inputs. Dropping the requirement that the tails of the density of the input random variables decay smoothly, the estimator extends the applicability of the generalized likelihood ratio (GLR) method. We demonstrate the new estimator for several general classes of input random variates, including independent inverse transform random variates and dependent input random variables governed by an Archimedean copula. We show how the new estimator works in settings such as density estimation, and we illustrate applications to credit risk derivatives. Numerical experiments substantiate broad applicability and flexibility in dealing with discontinuities in the sample performance.

Keywords: simulation, stochastic derivative estimation, discontinuous sample performance, uniform random numbers, generalized likelihood ratio method

1. Introduction

Simulation is a powerful technique for optimizing and analyzing complex stochastic systems [1]. Uniform random numbers are first generated by computer algorithms, and then are used as basic building blocks for generating other random variables, which in turn are fed into a simulation model to estimate output performance measures. In simulation, stochastic gradient estimation plays a central role in gradient-based optimization and sensitivity analysis. The finite difference (FD) method is

easily implementable, but its effective application often requires extra simulations and a careful balance of a bias-variance trade-off, especially when the parameter space is high-dimensional. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two well-established unbiased derivative estimation techniques [2, 3, 4, 5]. IPA typically leads to lower variance than LR [6, 7]. L'Ecuyer [6] provides a general framework unifying IPA and LR, under which the resulting estimator depends on the choice of what a sample point in a probability space represents. See [8] for a recent review.

In this paper, we study the stochastic gradient estimation problem for a stochastic model of the following form:

$$\varphi(g(U;\theta)), \tag{1}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a measurable function (not necessarily continuous), $g(\cdot;\theta) = (g_1(\cdot;\theta), \ldots, g_n(\cdot;\theta))$ is a vector of functions $g_i : (0,1)^n \to \mathbb{R}$ with certain smoothness properties to be made more precise in Section 3.2, and $U = (U_1, \ldots, U_n)$ is a vector of i.i.d. $\mathcal{U}(0,1)$ random variables (i.e., uniform over $(0,1)$). For this problem, a straightforward pathwise derivative estimator, i.e., IPA, obtained by directly interchanging derivative and expectation, may not apply because discontinuities in the sample performance of the stochastic model could be introduced by $\varphi(\cdot)$. On the other hand, the classic LR is also not applicable because of the existence of structural parameters in $g(U;\theta)$.

Stochastic gradient estimation research has traditionally focused on applications in discrete event dynamic systems (DEDS), including queueing systems [9, 10, 11], inventory management [12, 13], statistical quality control [14, 15], maintenance systems [16, 17], and financial engineering and risk management such as computing financial derivatives [18, 19, 20, 21, 22, 23, 24], value-at-risk (VaR) and conditional VaR (CVaR) [25, 26, 27, 28, 29]. Recently, stochastic gradient estimation techniques have attracted attention in machine learning and artificial intelligence (AI); see [30] and [31] for recent reviews.

IPA requires continuity in the sample performance and LR does not directly apply to structural parameters (parameters directly appearing in the sample performance), which significantly limit their applicability. Smoothed perturbation analysis (SPA) deals with discontinuous sample performances by using a conditioning technique [32, 33], but a good choice of conditioning is problem-dependent. Push-out LR addresses structural parameters by pushing the parameters out of the sample performance and into the density [4], which can be achieved alternatively with the IPA-LR in [6], which requires an explicit transformation. Recently, Peng et al. [34] proposed a generalized likelihood ratio (GLR) method that can deal with a large scope of discontinuous

sample performances with structural parameters in a unified framework. The method extends the application domain of IPA and LR and does not require conditioning and transformation techniques tailored to specific problem structures.

The GLR method has the virtue of handling many applications in a uniform manner, and it has been used to deal with discontinuities in financial options, statistical quality control, maintenance systems, and inventory systems [35, 34]. Distribution sensitivities — meaning the derivatives of the distribution function with respect to both the arguments and the parameters in the underlying stochastic model — lie at the center of many applications such as quantile sensitivity estimation, confidence interval construction for the quantile and quantile sensitivities, and statistical inference [36, 37]. Peng et al. [38] derive GLR estimators for any order of distribution sensitivities and apply them to maximum likelihood estimation for complex stochastic models without requiring analytical likelihoods. Glynn et al. [39] apply the GLR method to estimate sensitivity of a distortion risk measure, which is a Lebesgue-Stieltjes integral of quantile sensitivities and includes VaR and CVaR as special cases.

Although the existing GLR method has broad applicability, it requires knowing the explicit analytical form of the input distribution density and requires both tails of the density for each argument decay to zero smoothly at a sufficient rate, which may not be satisfied in some applications. This smoothness requirement can sometimes be circumvented through a change of variables [34], but such a transformation may lead to a GLR estimator with undesirable statistical properties. In this work, we relax this smoothness requirement and propose new unbiased GLR gradient estimators for stochastic models whose inputs are uniform random numbers, which are the basic building blocks in generating other random variables. Unlike in [34] which treats a setting where the surface integration contribution for the GLR estimator is zero, here we derive a new estimator, which includes the surface integration, called GLR with uniform random numbers as inputs (GLR-U). Moreover, in settings where the surface integration contribution is zero, we are able to relax certain integrability conditions given in [34].

We provide specific forms of the GLR-U estimators for several types of stochastic models and apply the GLR-U method to various problem settings, including distribution sensitivities and credit risk financial derivatives. The GLR-U estimator with independent parameterized input random variables generated from the inverse transform of uniform random numbers reduces to the classic LR estimator, which indicates that GLR-U is a generalization of LR from a different perspective than that of [34]. In the setting of [34], where independent input random variables are non-parameterized and the tails of their distribution go smoothly to zero at a sufficient

rate, the GLR-U estimator with input random variables generated from the inverse transform of uniform random numbers coincides with the GLR estimator in [34].

We also show how GLR-U can provide sensitivity estimators for models defined in terms of random vectors with given marginal distributions and whose dependence structures are specified by Archimedean copulas [40]. The Gaussian copula has been widely used due to its simplicity [41], and sensitivity analysis for portfolio credit risk derivatives with joint defaults governed by a Gaussian copula has been studied in [42] using LR and SPA. However, the Gaussian copula was widely criticized after the 2008 financial crisis, because it underestimates the probability of joint defaults. Archimedean copulas [40], not covered in [42], are relatively easy to simulate and can better capture the asymmetric tail dependence structure of the joint default data [43]. For sensitivity analysis of Archimedean copulas, IPA may fail for a discontinuous sample performance, whereas LR and GLR in [34] typically are not applicable.

Conditional Monte Carlo (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]. We will show in the numerical experiments that CMC can be applied to GLR-U for reducing the variance. Another effective variance reduction technique is randomized quasi-Monte Carlo (RQMC), which replaces the vectors of uniform random numbers that drive independent simulation runs by dependent vectors of uniform random numbers that cover the space more evenly [44]. When estimating an expectation, RQMC can provide an unbiased estimator whose variance converges at a faster rate than with Monte Carlo when the estimator inside the expectation is sufficiently smooth as a function of the underlying uniform random numbers. The GLR-U method developed in this work is compatible with RQMC, and we show in the numerical experiments that the variance of the GLR-U estimator can be reduced significantly by appropriately combining with CMC and RQMC.

The rest of the paper is organized as follows. Section 2 sets the framework. The new GLR-U estimator is presented in Section 3 with the specific forms of the estimators for three types of models. Applications are given in Section 4. Numerical experiments can be found in Section 5. The last section concludes. The technical proofs and additional numerical results can be found in the appendices.

2. Problem Formulation

Consider the stochastic model (1) with uniform random numbers as inputs. For simplicity, we take the dimension of U to be the same as that of g. This constraint can be relaxed to allow the dimension m of U to be no less than the dimension n

of g. When $m > n$, we can choose n components out of m components of vector U to reduce to the stochastic model (1), an example of which is presented later in the numerical experiments. The constraint $n \leq m$ is required for technical reasons, as described in more detail in Section 2.4 of [34]. When θ is a scalar, we consider the problem of estimating the following derivative:

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

When θ is a vector, each component of the gradient can be computed separately.

In Peng et al. [34], the stochastic model considered for the derivative estimation problem is $\varphi(g(X;\theta))$, where the density of $X = (X_1, \ldots, X_n)$ is assumed to be known and both tails decay smoothly at a sufficient rate for each component of the distribution of X when the other components are fixed. This assumption is not satisfied by discontinuous densities, such as the uniform and exponential distributions, and a primary objective of this work is to address this limitation. It is tempting to transform stochastic model (1) into $\varphi(g(F_1(Y_1), \ldots, F_n(Y_n);\theta))$, where Y_i has continuous cumulative distribution function (cdf) $F_i(\cdot), i = 1, \ldots, n$, with tails going smoothly to zero, so that the problem can be put into the same form as the original stochastic model in [34]. However, because of the cdfs of Y_i's appearing in the sample performance, we show in Appendix A that this simple transformation does not work as expected. Before deriving our GLR-U derivative estimator, we first introduce four important examples for generating random variables to illustrate potential applications of the stochastic model (1) and will return to these examples later.

Example 1. Independent Parameterized Inputs Generated via the Inverse Transform Method Suppose $X = (X_1, \ldots, X_n)$ is a vector of independent random variables, where each X_i has cdf $F_i(\cdot;\theta), i = 1, \ldots, n$, and is generated by (standard) inversion:

$$X_i = F_i^{-1}(U_i;\theta), \quad i = 1, \ldots, n,$$

with i.i.d. $U_i \sim \mathcal{U}(0,1)$. A stochastic model with i.i.d. $\mathcal{U}(0,1)$ random numbers as input can be written as $\varphi(g(U;\theta)) = \varphi(F_1^{-1}(U_1;\theta), \ldots, F_n^{-1}(U_n;\theta))$, where $g(u;\theta) = (F_1^{-1}(u_1;\theta), \ldots, F_n^{-1}(u_n;\theta))$.

Example 2. Independent Non-parameterized Inputs Generated via the Inverse Transform Method Suppose $X = (X_1, \ldots, X_n)$ is a vector of non-parameterized

independent random variables, and is generated by inversion. A stochastic model with i.i.d. $\mathcal{U}(0,1)$ random numbers as input is given by

$$\varphi(g(U;\theta)) = \varphi(h(F_1^{-1}(U_1),\ldots,F_n^{-1}(U_n);\theta)),$$

where $h(x;\theta) := (h_1(x;\theta),\ldots,h_n(x;\theta))$ and $g_i(u;\theta) := h_i(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n);\theta)$, $i = 1,\ldots,n$.

Example 3. Archimedean Copulas Copulas are a general way of representing the dependence in a multivariate distribution. A *copula* is a multivariate cdf whose one-dimensional marginals are all independent $\mathcal{U}(0,1)$. It can be defined by a function $C(\cdot;\theta) : [0,1]^n \to [0,1]$ that satisfies certain conditions required for C to be a valid cdf; (e.g., it cannot decrease, $C(u;\theta) = 0$ whenever one coordinate of u is 0, and $C(\mathbf{1};\theta) = 1$); see [40]. For any given copula and arbitrary marginal distributions with continuous cdf's $F_1(\cdot), F_2(\cdot), ..., F_n(\cdot)$ with densities $f_i(\cdot)$, $i = 1,\ldots,n$, one can define a multivariate distribution having exactly these marginals with joint cdf F_X given by $F_X(x) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n);\theta)$ for all $x := (x_1,\ldots,x_n)$. Nelson [40] shows that any multivariate distribution can be represented in this way. If $C(\cdot;\theta)$ is absolutely continuous, the density of the joint distribution is

$$f_X(x;\theta) = c\left(F_1(x_1),\ldots,F_n(x_n);\theta\right) \prod_{i=1}^{n} f_i(x_i),$$

where $c(v;\theta) = \frac{\partial^n C(v;\theta)}{\partial v_1 \cdots \partial v_n}$, $v = (v_1,\ldots,v_n)$, and the derivative is interpreted as a Radon-Nikodym derivative when $C(\cdot;\theta)$ is not nth-order differentiable.

To generate $X = (X_1,\ldots,X_n)$ from the joint cdf $F_X(\cdot)$, generate $V = (V_1,\ldots,V_n)$ from the copula and return $X_i = F_i^{-1}(V_i)$ for each i. Generating V from the copula is not always obvious, but there are classes of copulas for which this can be easily done, one of them being the *Archimedean copulas*. This important family of copulas can model strong forms of tail dependence using a single parameter, which makes them convenient to use. An Archimedean copula C_a is defined by

$$C_a(v;\theta) = \psi_\theta\left(\psi_\theta^{[-1]}(v_1) + \ldots + \psi_\theta^{[-1]}(v_n)\right),$$

where the generator function $\psi_\theta : [0,\infty) \to [0,1]$ is a strictly decreasing convex function such that $\lim_{x\to\infty}\psi_\theta(x) = 0$, $\theta \in [0,\infty)$ is a parameter governing the strength of dependence, and $\psi_\theta^{[-1]}$ is a pseudo-inverse defined by $\psi_\theta^{[-1]}(x) = \mathbf{1}\{0 \le x \le \psi_\theta(0)\}\psi_\theta^{-1}(x)$, with the convention that $\psi_\theta^{-1}(0) = \inf\{x : \psi_\theta(x) = 0\}$. Note that the

roles of notations ψ_θ and $\psi_\theta^{[-1]}$ are reversed from their definitions in [40]. Archimedean copulas are absolutely continuous, and their densities have the form:

$$c_a(v;\theta) = \mathbf{1}\left\{0 \le \sum_{i=1}^n \psi_\theta^{-1}(v_i) \le \psi_\theta^{-1}(0)\right\} \left.\frac{\partial^n \psi_\theta(x)}{\partial x^n}\right|_{x=\sum_{i=1}^n \psi_\theta^{-1}(v_i)} \prod_{i=1}^n \frac{\partial \psi_\theta^{-1}(v_i)}{\partial v_i},$$

assuming that the generator function $\psi_\theta(\cdot)$ is smooth. In general we do not have an analytical expression for $c_a(\cdot;\theta)$, which can be discontinuous in θ, so discontinuities may exist for the density f_X with the copula defined by $c_a(\cdot;\theta)$; thus, LR and GLR in Peng et al. [34] typically do not apply to sensitivity analysis with Archimedean copulas.

Marshall and Olkin [45] propose the following simple algorithm to generate V from an Archimedean copula with generator function $\psi_\theta(\cdot)$:

(i) Generate a random variable Y_θ from the distribution with Laplace transform $\psi_\theta(\cdot)$ (with at least one uniform random number as input).
(ii) For $i = 1, \ldots, n$, let $V_i = \psi_\theta\left(-(\log U_i)/Y_\theta\right)$ with i.i.d. $U_i \sim \mathcal{U}(0,1)$.

For a given Y_θ, this gives a stochastic model with uniform random numbers U_i as inputs:

$$\varphi(g(U;\theta)) = \phi\left(F_1^{-1}\left(\psi_\theta\left(-(\log U_1)/Y_\theta\right)\right), \ldots, F_n^{-1}\left(\psi_\theta\left(-(\log U_n)/Y_\theta\right)\right)\right),$$

where $\varphi(v_1, \ldots, v_n) = \phi(F_1^{-1}(v_1), \ldots, F_1^{-1}(v_n))$ and

$$g(u;\theta) = \left(\psi_\theta\left(-(\log u_1)/Y_\theta\right), \ldots, \psi_\theta\left(-(\log u_n)/Y_\theta\right)\right).$$

Since the inputs of the stochastic model (1) are uniform random numbers, the simulation of the stochastic model (1) and the GLR-U estimator based on the new model can be easily adapted to RQMC, which replaces the random seeds in Monte Carlo simulation with randomized low-discrepancy sequences for reducing the variance of GLR-U estimator [46, 47].

3. A Generalized Likelihood Ratio Method

In this section, we derive the GLR-U estimator for the derivative (2) of the expectation of stochastic model (1). We first provide an overview for the derivation of the new method. Then the general theory for GLR is derived, and it is applied to the three examples in the previous section.

3.1. Overview

To illustrate the main idea behind the derivation of GLR-U, we first consider a simple one-dimensional problem in the framework of Peng et al. [34], i.e., $\varphi(g(X;\theta))$, where X is a random variable with density $f(\cdot;\theta)$ supported on Ω. When $\varphi:\ \mathbb{R}\mapsto\mathbb{R}$ and $g:\ \Omega\times\mathbb{R}\mapsto\mathbb{R}$ are smooth (infinitely differentiable) functions, the IPA-LR framework in [6] applies, i.e., assuming that the derivative and expectation can be interchanged, we have

$$\begin{aligned}&\frac{\partial\mathbb{E}[\varphi(g(X;\theta))]}{\partial\theta}=\frac{\partial}{\partial\theta}\int_\Omega\varphi(g(x;\theta))f(x;\theta)dx\\&=\int_\Omega\underbrace{\varphi(g(x;\theta))\frac{\partial\log f(x;\theta)}{\partial\theta}}_{LR\ part}f(x;\theta)dx+\int_\Omega\underbrace{\left.\frac{\partial\varphi(y)}{\partial y}\right|_{y=g(x;\theta)}\frac{\partial g(x;\theta)}{\partial\theta}}_{IPA\ part}f(x;\theta)dx,\end{aligned}\tag{3}$$

where the indicated LR and IPA parts serve as unbiased estimators. However, in the case where φ is not smooth, the second term may not be well defined, e.g., when φ is discontinuous. To address this more general setting, we approximate φ by a smooth function φ_ϵ such that $\varphi_\epsilon\to\varphi$ as $\epsilon\to 0$ in a suitable manner, and by applying integration by parts (Theorem 37.2 in [48]), we can express the second term in the right-hand side of (3) as

$$\begin{aligned}&\int_\Omega\left.\frac{\partial\varphi_\epsilon(y)}{\partial y}\right|_{y=g(x;\theta)}\frac{\partial g(x;\theta)}{\partial\theta}f(x;\theta)dx=\int_\Omega\frac{\partial\varphi_\epsilon(g(x;\theta))}{\partial x}\left(\frac{\partial g(x;\theta)}{\partial x}\right)^{-1}\frac{\partial g(x;\theta)}{\partial\theta}f(x;\theta)dx\\&=\int_{\partial\Omega}\varphi_\epsilon(g(x;\theta))\left(\frac{\partial g(x;\theta)}{\partial x}\right)^{-1}\frac{\partial g(x;\theta)}{\partial\theta}f(x;\theta)\hat{v}ds\\&\qquad-\int_\Omega\varphi_\epsilon(g(x;\theta))\frac{\partial}{\partial x}\left\{\left(\frac{\partial g(x;\theta)}{\partial x}\right)^{-1}\frac{\partial g(x;\theta)}{\partial\theta}f(x;\theta)\right\}dx\\&=\int_{\partial\Omega}\varphi_\epsilon(g(x;\theta))r(x;\theta)f(x;\theta)\hat{v}ds-\int_\Omega\varphi_\epsilon(g(x;\theta))d(x;\theta)f(x;\theta)dx,\end{aligned}\tag{4}$$

where $\partial\Omega$ is the boundary of Ω which is assumed to be $(0,1)$ here and $\hat{v}$ is the unit normal vector pointing outward, ds is the surface measure,

$$r(x;\theta)=\left(\frac{\partial g(x;\theta)}{\partial x}\right)^{-1}\frac{\partial g(x;\theta)}{\partial\theta},$$

and

$$d(x;\theta)=\frac{1}{f(x;\theta)}\frac{\partial}{\partial x}\left\{\left(\frac{\partial g(x;\theta)}{\partial x}\right)^{-1}\frac{\partial g(x;\theta)}{\partial\theta}f(x;\theta)\right\}.$$

The derivation above can be extended straightforwardly to the case when g and x are multi-dimensional. In the remainder of the paper, Ω is the open unit hypercube in n dimensions. A generic measurable function $\varphi(\cdot)$ not necessarily continuous is first replaced by a smooth approximation before the derivation, and then after the derivation the limit is taken to recover unbiasedness under certain integrability conditions. Since the derivatives of $\varphi(\cdot)$ only appear in the intermediate steps but not in the final expression of (4), we only need the existence of a function that smooths $\varphi(\cdot)$, which can be established under a mild regularity condition. The condition and the proof can be found in Section 3.2 and Appendix A, respectively.

The first term of (3) and the second term of (4) can be estimated by simulation, but the first term of (4) is a surface integration that could be difficult to compute in general. The surface integration contribution in [34] is shown to be zero under certain conditions including the tails of the input densities decaying to zero smoothly and at a sufficient rate. In this work, we consider uniform random variables as inputs, so the support is an open cube $\Omega = (0,1)^n$, which has a simple boundary. Thus, the normal vector $\hat{v}$ to the surface is constant, and the surface measure ds is amenable for computation on each face of the cube, i.e., the surface integration contribution can be estimated by simulation. In addition, the input density $f(x;\theta) = 1$ for $x \in (0,1)^n$ and $\partial \log f(x;\theta)/\partial\theta = 0$, so that the first term vanishes in (3), leaving just the second term. Due to possible discontinuities in φ, this term is retransformed by integration by parts, which will be seen in Section 3.3, in the most basic case leads to an estimator that is not standard IPA but standard LR.

3.2. General Theory

When θ is a scalar, denote the Jacobian of g by

$$J_g(u;\theta) := \begin{pmatrix} \frac{\partial g_1(u;\theta)}{\partial u_1} & \frac{\partial g_1(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_1(u;\theta)}{\partial u_n} \\ \frac{\partial g_2(u;\theta)}{\partial u_1} & \frac{\partial g_2(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_2(u;\theta)}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(u;\theta)}{\partial u_1} & \frac{\partial g_n(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_n(u;\theta)}{\partial u_n} \end{pmatrix}, \qquad \text{and}$$

$$\partial_\theta g(u;\theta) := \left(\frac{\partial g_1(u;\theta)}{\partial \theta}, \ldots, \frac{\partial g_n(u;\theta)}{\partial \theta} \right)^T,$$

with the superscript T indicating vector transposition. In addition, we define two weight functions in the GLR estimator:

$$r_i(u;\theta) := \left(J_g^{-1}(u;\theta)\ \partial_\theta g(u;\theta) \right)^T e_i, \quad i = 1, \ldots, n, \tag{5}$$

$$d(u;\theta) := \sum_{i=1}^{n} e_i^T J_g^{-1}(u;\theta)\ (\partial_{u_i} J_g(u;\theta)) J_g^{-1}(u;\theta)\partial_\theta g(u;\theta) - \text{trace}(J_g^{-1}(u;\theta)\ \partial_\theta J_g(u;\theta)), \tag{6}$$

where e_i is the ith unit column vector and $\partial_z J_g$ is the matrix obtained by differentiating J_g with respect to z element-wise. When θ is a vector, we need to compute (5) and (6) for each coordinate. Notice that the computation of the Jacobian matrix $J_g(\cdot;\theta)$, its inverse, and its derivatives with respect to u_i, which necessitates significant computational burden for computing (5) and (6), needs to be done only once for all coordinates of θ. Therefore, the computational complexity of the GLR-U estimator for a d-dimensional θ is less than d times the computational complexity of the GLR-U estimator for a 1-dimensional θ.

Let x^- and x^+ be the limits taken from the left-hand side and right-hand side of x, respectively, and for a function $h(\cdot)$, denote $h(x^-) := \lim_{x\to x^-} h(x)$ and $h(x^+) := \lim_{x\to x^+} h(x)$. We introduce the following conditions for our analysis. The unbiasedness of the proposed GLR-U estimator can be established under two different subsets of these conditions.

(A.1) There exist a family of smooth functions (i.e., infinitely differentiable) $\varphi_\epsilon(\cdot)$ indexed by ϵ and $p > 1$ such that

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\int_{(0,1)^n} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))|^p\ du = 0,$$

and if $n > 1$, for any $i = 1,\ldots,n$,

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta, u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]}\int_{(0,1)^{n-1}} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))|^p\ du_{-i} = 0,$$

where $u_{-i} := (u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n)$, Θ is a compact neighborhood of the value of interest for parameter θ, while if $n = 1$,

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta, u\in(0,1)\setminus[\varepsilon,1-\varepsilon]} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))| = 0.$$

(A.2) The following integrability condition holds:

$$\int_{(0,1)^{n-1}}\sup_{\theta\in\Theta, u_i\in(0,1)} \left|\varphi(g(u;\theta))\ r_i(u;\theta)\right|\ du_{-i} < \infty, \quad i = 1,\ldots,n.$$

(A.3) The Jacobian $J_g(u;\theta)$ is invertible for almost every $u \in (0,1)^n$, and the performance function $g(u;\theta)$ is twice continuously differentiable with respect to $(u,\theta) \in [0,1]^n \times \Theta$.

(A.4) The following integrability condition holds:

$$\int_{(0,1)^n} \sup_{\theta\in\Theta} \left|\varphi(g(u;\theta))\; d(u;\theta)\right| \; du < \infty.$$

(A.5) The function $g(\cdot;\theta)$ is invertible for all $\theta \in \Theta$, and

$$\lim_{u_i\to 1^-} \sup_{\theta\in\Theta, u_{-i}\in(0,1)^{n-1}} |r_i(u;\theta)| = \lim_{u_i\to 0^+} \sup_{\theta\in\Theta, u_{-i}\in(0,1)^{n-1}} |r_i(u;\theta)| = 0, \quad i = 1,\ldots,n.$$

Remark 1. Condition (A.1) can be checked in certain settings when $\varphi_\epsilon(\cdot)$ can be explicitly constructed, e.g., $\varphi_\epsilon(x) = 0$ if $x \le -\epsilon$, $\varphi_\epsilon(x) = 1$ if $x \ge \epsilon$, $\varphi_\epsilon(x) = (x+\epsilon)/(2\epsilon)$ if $-\epsilon < x < \epsilon$, and $\varphi(x) = \mathbf{1}\{x \le 0\}$; see a more general case in Proposition 1. The invertibility of the Jacobian matrix in condition (A.3) justifies the local invertibility of function $g(\cdot;\theta)$, whereas global invertibility of $g(\cdot;\theta)$ in condition (A.5) is stronger, although much weaker than requiring an explicit inverse function for $g(\cdot;\theta)$ in deriving the push-out LR estimator [4]. In general, it is difficult to find an explicit inverse function for a nonlinear function $g(\cdot;\theta)$, but the existence of the inversion could be guaranteed by the inverse function theorem.

Unbiasedness of the new GLR-U estimator developed in this work is established under two sets of conditions in the following theorem.

Theorem 1. *Under conditions (A.1) – (A.4) or (A.3) – (A.5), for all $\theta \in \Theta$,*

$$\frac{\partial \mathbb{E}[\varphi(g(U;\theta))]}{\partial\theta} = \mathbb{E}[G(U;\theta)], \qquad \textit{where} \tag{7}$$

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

with $\overline{U}_i := (U_1,\ldots,\underbrace{1^-}_{i\text{th element}},\ldots,U_n)$, $\underline{U}_i := (U_1,\ldots,\underbrace{0^+}_{i\text{th element}},\ldots,U_n)$, *and* $r_i(\cdot)$ *and* $d(\cdot)$ *defined by (5) and (6), respectively.*

Remark 2. The proof of the theorem can be found in Appendix A. Even if $g(\overline{U}_i;\theta) = \infty$ or $g(\underline{U}_i;\theta) = \infty$, the GLR-U estimator could be well defined; see e.g., Sections 3.3 and 3.4. In the case where the surface integration part becomes zero, we can prove the result without assuming (A.1), and we also avoid the integrability condition in [34] on certain intermediate quantities (the smoothed function), which is difficult to verify in practice. To achieve this goal, the proof is obtained by first truncating the support of the input uniform random numbers, i.e., $(0,1)^n$, to a compact set $[\varepsilon, 1-\varepsilon]^n$ for $\varepsilon > 0$ and then appropriately expanding it to the whole space, i.e., letting $\varepsilon \to 0$. This is merely a technical step used in the proof rather than a treatment required in actual implementation. We caution that the conditions in the theorem may still be difficult to check when φ is too complicated to allow an explicit smoothed function or g comprises complex nonlinear functions. The conditions in Theorem 1 are sufficient for justifying unbiasedness of GLR-U, but they are not necessary conditions. Puchhammer and L'Ecuyer [49] compare the GLR-U with their push-out LR in the density estimation problem. They show examples in which the conditions in this work are not satisfied, whereas their conditions for justifying unbiasedness of push-out LR are satisfied in the density estimation problem, even though the GLR-U and their push-out LR give the same estimator. In Sections 3.3 and 3.4, we will prove that LR is a special case of GLR-U for a stochastic model with independent and continuous input random variables, and the GLR in [34] coincides with GLR-U in certain circumstance. Peng et al. [34] prove that push-out LR is a special case of GLR. Therefore, the more restrictive conditions for justifying unbiasedness of GLR-U are adopted to accommodate the generality of stochastic model (1), and can be simplified significantly for certain special forms of the stochastic model (1) such as in the density estimation problem [50].

We now examine the special case where $g(u;\theta) = (g_1(u_1;\theta), \ldots, g_n(u_n;\theta))$, which covers Examples 1, 2, and 3 in Section 2. For n independent uniform random numbers, we can take $g_i(u_i;\theta) = F_i^{-1}(u_i;\theta)$ for $i = 1, \ldots, n$, while for dependent variables governed by an Archimedean copula, conditional on $Y_\theta = y$, we have $g_i(u_i;\theta) = \psi_\theta(-\log u_i/y)$ for $i = 1, \ldots, n$. In this special case, the Jacobian becomes

$$J_g(u;\theta) = \begin{pmatrix} \frac{\partial g_1(u_1;\theta)}{\partial u_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial g_2(u_2;\theta)}{\partial u_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial g_n(u_n;\theta)}{\partial u_n} \end{pmatrix}.$$

Then we have

$$r_i(u;\theta) = \frac{\partial g_i(u_i;\theta)}{\partial \theta} \Big/ \frac{\partial g_i(u_i;\theta)}{\partial u_i}, \quad i = 1, \ldots, n, \qquad \text{and} \qquad d(u;\theta) = \sum_{i=1}^{n} d_i(u_i;\theta),$$

$$\text{where} \qquad d_i(u_i;\theta) := \frac{\partial g_i(u_i;\theta)}{\partial \theta} \frac{\partial^2 g_i(u_i;\theta)}{\partial u_i^2} \Big/ \left(\frac{\partial g_i(u_i;\theta)}{\partial u_i} \right)^2 - \frac{\partial^2 g_i(u_i;\theta)}{\partial \theta \partial u_i} \Big/ \frac{\partial g_i(u_i;\theta)}{\partial u_i}.$$

Moreover, condition (A.1) in Theorem 1 can be replaced by a set of a simpler assumptions when the performance function $\varphi(x)$ is a product of n indicators: $\varphi(x) = \prod_{i=1}^{n} \mathbf{1}\{x_i \le 0\}$, in which case a smoothed function $\varphi_\epsilon(\cdot)$ can be constructed explicitly. The performance function in the distribution sensitivities discussed in Section 4.1 is an indicator function. The payoffs of many financial derivatives such that in Section 5.2 contain the product of indicator functions. The proof of the following proposition can be found in Appendix A.

Proposition 1. *Consider the stochastic model*

$$\varphi(g(U;\theta)) = \prod_{i=1}^{n} \mathbf{1}\{g_i(U_i;\theta) \le 0\}.$$

Then condition (A.1) holds if for $i = 1, \ldots, n$ and there exists $0 < \varepsilon < 1/2$ such that

$$\inf_{\theta \in \Theta, u_i \in [\varepsilon, 1-\varepsilon]} \left| \frac{\partial g_i(u_i;\theta)}{\partial u_i} \right| > 0 \quad \textit{and} \quad \inf_{\theta \in \Theta, u_i \in (0,1) \setminus [\varepsilon, 1-\varepsilon]} \left| g_i(u_i;\theta) \right| > 0. \tag{8}$$

It is straightforward to extend the conclusion of Proposition 1 to the case $\varphi(g(U;\theta)) = \prod_{i=1}^{n} \mathbf{1}\{g_i(U_i;\theta) \le 0\} h(g(U;\theta))$, where $h(\cdot)$ is smooth and $h(g(U;\theta))$ satisfies certain moment conditions. Specifically, by applying Hölder's inequality,

$$\begin{aligned}
&\int_{(0,1)^n} \left| \left(\prod_{i=1}^{n} \mathbf{1}\{g_i(u_i;\theta) \le 0\} - \chi_\epsilon(g(u;\theta)) \right) h(g(u;\theta)) \right|^p \, du \\
&\le \left[\int_{(0,1)^n} \left| \left(\prod_{i=1}^{n} \mathbf{1}\{g_i(u_i;\theta) \le 0\} - \chi_\epsilon(g(u;\theta)) \right) \right|^{pp'} \, du \right]^{1/p'} \left[\int_{(0,1)^n} \left| h(g(u;\theta)) \right|^{pq'} \, du \right]^{1/q'},
\end{aligned}$$

where $1/p' + 1/q' = 1$ and χ_ϵ is the smoothed function of the product of indicator, making the first integral go to zero as $\epsilon \to 0$. The other integrals in condition (A.1) can be treated similarly by applying Hölder's inequality. Letting $\varphi_\epsilon(g(u;\theta)) =$

$\chi_\epsilon(g(u;\theta))h(g(u;\theta))$, we check condition (A.1) for stochastic model $\varphi(g(U;\theta))$. If the functions g_i can be decomposed as products of the form $g_i(u_i;\theta) = \xi_i(\theta)\eta_i(u_i)$ for $i = 1,\ldots,n$, then conditions (A.2), (A.4), (A.5), and (6) can be simplified. For example, an exponential random variable with mean θ can be generated by $-\log(U_i)/\theta$ where $U_i \sim \mathcal{U}(0,1)$. When this decomposition holds, we can write

$$r_i(u;\theta) = \frac{d\log\xi_i(\theta)}{d\theta} \Big/ \frac{d\log\eta_i(u_i)}{du_i}, \quad i = 1,\ldots,n,$$

and

$$d(u;\theta) = \sum_{i=1}^{n} d_i(u_i;\theta), \quad \text{with} \quad d_i(u_i;\theta) = \frac{d\log\xi_i(\theta)}{d\theta}\left[\frac{\eta_i(u_i)\eta_i''(u_i)}{(\eta_i'(u_i))^2} - 1\right].$$

(A.2') Boundedness condition on functions of input random numbers:

$$\inf_{u_i\in(0,1)} \left|\frac{d\log\eta_i(u_i)}{du_i}\right| > 0, \quad i = 1,\ldots,n.$$

(A.4') Integrability condition on functions of input random numbers:

$$\mathbb{E}\left[\frac{\left|\eta_i(U_i)\eta_i''(U_i)\right|}{(\eta_i'(U_i))^2}\right] < \infty, \quad i = 1,\ldots,n.$$

(A.5') Boundary condition on functions of input random numbers: functions $\eta_i(u_i)$, $i = 1,\ldots,n$, are monotone, and

$$\lim_{u_i\to 1^-}\left|\frac{d\log\eta_i(u_i)}{du_i}\right| = \lim_{u_i\to 0^+}\left|\frac{d\log\eta_i(u_i)}{du_i}\right| = \infty, \quad i = 1,\ldots,n.$$

Corollary 1. *Suppose that $g_i(u_i;\theta) = \xi_i(\theta)\eta_i(u_i)$, $i = 1,\ldots,n$, and $\varphi(\cdot)$ is bounded and*

$$\max_{i=1,\ldots,n}\sup_{\theta\in\Theta}\left|\frac{\partial\log\xi_i(\theta)}{\partial\theta}\right| < \infty.$$

Then conditions (A.2), (A.4), and (A.5) can be replaced by (A.2'), (A.4') and (A.5'), respectively, and condition (8) in Proposition 1 also simplifies to

$$\inf_{\theta\in\Theta}\left|\xi_i(\theta)\right| > 0, \quad \inf_{u_i\in[\varepsilon,1-\varepsilon]}\left|\eta_i'(u_i)\right| > 0, \quad \inf_{u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]}\left|\eta_i(u_i)\right| > 0, \quad i = 1,\ldots,n.$$

Unbiasedness of the GLR-U estimators in many examples of this paper can be justified by verifying these simplified conditions. The product form of the stochastic model is satisfied in most cases when the conditioning and push-out techniques can be done explicitly. Other assumptions in Corollary 1 ensure that the stochastic model $g_i(U_i;\theta)$ has a density. This requirement would not significantly limit applications, because the generality of φ enables the freedom to properly define the stochastic model g. Note that discrete random variables can also be generated by continuous random variables through certain functional transformations.

3.3. The Independent Parameterized Case

Let us return to the independent case of Example 1 and suppose that each X_i is continuous with density $f_i(\cdot;\theta)$. Our goal is to estimate

$$\frac{\partial \mathbb{E}[\varphi(F_1^{-1}(U_1;\theta),\ldots,F_n^{-1}(U_n;\theta))]}{\partial\theta},$$

for which the Jacobian is

$$J_g(u;\theta) = \begin{pmatrix} \frac{1}{f_1(X_1(u_1;\theta);\theta)} & 0 & \cdots & 0 \\ 0 & \frac{1}{f_2(X_2(u_2;\theta);\theta)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{f_n(X_n(u_n;\theta);\theta)} \end{pmatrix},$$

and

$$\partial_\theta g(u;\theta) = \left(\frac{\partial X_1(u_1;\theta)}{\partial\theta},\ldots,\frac{\partial X_n(u_n;\theta)}{\partial\theta}\right)^T,$$

where

$$X_i(u_i;\theta) := F_i^{-1}(u_i;\theta) \quad \text{and} \quad \frac{\partial X_i(u_i;\theta)}{\partial\theta} := -\frac{\partial F_i(x_i;\theta)}{\partial\theta} \Big/ f_i(x_i;\theta)\Big|_{x_i=X_i(u_i;\theta)}.$$

Then the weight functions in the GLR-U estimator are:

$$r_i(u;\theta) = -\frac{\partial F_i(x_i;\theta)}{\partial\theta}\bigg|_{x_i=X_i(u_i;\theta)} \quad \text{and} \quad d_i(u_i;\theta) = \frac{\partial \log f_i(x_i;\theta)}{\partial\theta}\bigg|_{x_i=X_i(u_i;\theta)},$$

so

$$\lim_{u_i\to 1^-} r_i(u;\theta) = \lim_{u_i\to 0^+} r_i(u;\theta) = 0.$$

Therefore,

$$\frac{\partial \mathbb{E}[\varphi(F_1^{-1}(U_1;\theta),\ldots,F_n^{-1}(U_n;\theta))]}{\partial\theta} = \mathbb{E}\Big[\varphi(F_1^{-1}(U_1;\theta),\ldots,F_n^{-1}(U_n;\theta))\ d(U;\theta)\ \Big]$$
$$= \mathbb{E}\Big[\varphi(X)\ \sum_{i=1}^{n}\frac{\partial \log f_i(X_i;\theta)}{\partial\theta}\ \Big].$$

The expression inside the last expectation coincides with the classic LR derivative estimator in the case where the LR method is applicable, i.e., when there are no structural parameters in the sample performance [2]. From this perspective, the GLR method generalizes the LR method by allowing the appearance of structural parameters.

3.4. The Independent Non-parameterized Case

We then return to the independent case of Example 2, and consider estimating derivative

$$\frac{\partial \mathbb{E}\big[\varphi(h(\beta(U);\theta))\big]}{\partial\theta} \quad \text{with} \quad \beta(u) = (F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)).$$

Assume that the Jacobian $J_h(\cdot;\theta)$ of $h(\cdot;\theta)$ is invertible almost everywhere, and densities $f_i(x_i)$, $i = 1,\ldots,n$, are differentiable and supported on the entire space, i.e, $\mathbb{R}$. The Jacobian can be expressed by $J_g(u;\theta) = J_h(x;\theta)\Lambda(x)\big|_{x=\beta(u)}$, where

$$J_h(x;\theta) := \begin{pmatrix} \frac{\partial h_1(x;\theta)}{\partial x_1} & \frac{\partial h_1(x;\theta)}{\partial x_2} & \cdots & \frac{\partial h_1(x;\theta)}{\partial x_n} \\ \frac{\partial h_2(x;\theta)}{\partial x_1} & \frac{\partial h_2(x;\theta)}{\partial x_2} & \cdots & \frac{\partial h_2(x;\theta)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n(x;\theta)}{\partial x_1} & \frac{\partial h_n(x;\theta)}{\partial x_2} & \cdots & \frac{\partial h_n(x;\theta)}{\partial x_n} \end{pmatrix}, \Lambda(x) := \begin{pmatrix} \frac{1}{f_1(x_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{f_2(x_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{f_n(x_n)} \end{pmatrix}.$$

Suppose $\lim_{x_i\to\pm\infty}\big(J_h^{-1}(x;\theta)\partial_\theta h(x;\theta)\big)^T\Lambda^{-1}(x)e_i = 0$, which implies the tail of $f_i(\cdot)$ decays to zero at a sufficient rate, and then $\lim_{u_i\to 0} r_i(u;\theta) = \lim_{u_i\to 1} r_i(u;\theta) = 0$, $i = 1,\ldots,n$. Notice that

$$\partial_{u_i} J_g(u;\theta) = \Lambda_i(x)\big(\partial_{x_i} J_h(x;\theta)\big)\Lambda(x)\big|_{x=\beta(u)} + J_h(x;\theta)\partial_{x_i}\Lambda(x)\big|_{x=\beta(u)} J_\beta(u),$$

where $\Lambda_i(x)$ is a diagonal matrix with $1/f_i(x_i)$ in all the diagonal entries. Plugging the expressions of $J_g(u;\theta)$ and $\partial_{u_i} J_g(u;\theta)$ above into (6), we have

$$\begin{aligned}
d(u;\theta) =& \sum_{i=1}^{n} e_i^T \Lambda^{-1}(x) J_h^{-1}(x;\theta)\Lambda_i(x)(\partial_{x_i} J_h(x;\theta))\Lambda(x)\Lambda^{-1}(x)J_h^{-1}(x;\theta)\partial_\theta h(x;\theta)\Big|_{x=\beta(u)} \\
&+ \sum_{i=1}^{n} e_i^T \Lambda^{-1}(x) J_h^{-1}(x;\theta) J_h(x;\theta)\big(\partial_{x_i}\Lambda(x)\big) J_\beta(u)\Lambda^{-1}(x)J_h^{-1}(x;\theta)\partial_\theta h(x;\theta)\Big|_{x=\beta(u)} \\
&- \mathrm{trace}\big(\Lambda^{-1}(x)J_h^{-1}(x;\theta)\big(\partial_\theta J_h(x;\theta)\big)\Lambda(x)\big)\Big|_{x=\beta(u)} \\
=& \sum_{i=1}^{n} e_i^T J_h^{-1}(x;\theta)(\partial_{x_i} J_h(x;\theta))J_h^{-1}(x;\theta)\partial_\theta h(x;\theta)\Big|_{x=\beta(u)} - \big(J_h^{-1}(x;\theta)\partial_\theta h(x;\theta)\big)^T \zeta(x)\Big|_{x=\beta(u)} \\
&- \mathrm{trace}\big(J_h^{-1}(x;\theta)\partial_\theta J_h(x;\theta)\big)\Big|_{x=\beta(u)},
\end{aligned}$$

where $\zeta(x) = \big(\frac{\partial \log f_1(x_1)}{\partial x_1}, \ldots, \frac{\partial \log f_n(x_n)}{\partial x_n}\big)^T$. Furthermore, $\widehat{d}(X;\theta) = d(F_1(X_1), \ldots, F_n(X_n);\theta)$ coincides with the GLR estimator for stochastic model $\varphi(h(X;\theta))$ in [34].

3.5. Archimedean Copulas

We consider estimating

$$\frac{\partial}{\partial\theta}\mathbb{E}\bigg[\varphi\bigg(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big), \ldots, \psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\bigg)\bigg],$$

where the expectation is with respect to both Y_θ and the independent U_i, $i = 1, \ldots, n$. By conditioning, we can use a mixture of LR and GLR:

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\mathbb{E}\bigg[\varphi\bigg(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big), \ldots, \psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\bigg)\bigg] \\
&= \mathbb{E}\bigg[\varphi\bigg(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big), \ldots, \psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\bigg)\frac{\partial \log f_Y(y;\theta)}{\partial\theta}\Big|_{y=Y_\theta}\bigg] \\
&\quad + \mathbb{E}\bigg[\frac{\partial\mathbb{E}\big[\varphi\big(\psi_\theta\big(-\frac{\log U_1}{y}\big), \ldots, \psi_\theta\big(-\frac{\log U_n}{y}\big)\big)\big]}{\partial\theta}\Big|_{y=Y_\theta}\bigg],
\end{aligned} \tag{9}$$

where $f_Y(\cdot;\theta)$ is the density function of Y_θ.

We now show how to use GLR-U to handle the second term on the right-hand side of (9) with Y_θ fixed and generated from other uniform random numbers. The Archimedean copula model falls into the special case where $g(u;\theta) =$

$(g_1(u_1;\theta),\ldots,g_n(u_n;\theta))$, discussed after Theorem 1. The Jacobian in this case is

$$J_g(u;\theta,y)=\begin{pmatrix} -\frac{1}{u_1 y}\psi'_\theta\big(-\frac{\log u_1}{y}\big) & 0 & \cdots & 0 \\ 0 & -\frac{1}{u_2 y}\psi'_\theta\big(-\frac{\log u_2}{Y_\theta}\big) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{u_n y}\psi'_\theta\big(-\frac{\log u_n}{y}\big) \end{pmatrix},$$

and

$$\partial_\theta g(u;\theta,y)=\left(\frac{\partial\psi_\theta(x_1)}{\partial\theta}\bigg|_{x_1=-\frac{\log u_1}{y}},\ldots,\frac{\partial\psi_\theta(x_n)}{\partial\theta}\bigg|_{x_n=-\frac{\log u_n}{y}}\right)^T.$$

The weight functions in the GLR-U estimator are

$$r_i(u;\theta,y)=-\frac{u_i y}{\psi'_\theta(x_i)}\frac{\partial\psi_\theta(x_i)}{\partial\theta}\bigg|_{x_i=-\frac{\log u_i}{y}},$$

$$d_i(u_i;\theta,y)=\left(-\frac{1}{\psi'_\theta(x_i)}\frac{\partial\psi'_\theta(x_i)}{\partial\theta}+\frac{\psi''_\theta(x_i)}{(\psi'_\theta(x_i))^2}\frac{\partial\psi_\theta(x_i)}{\partial\theta}+\frac{\partial\psi_\theta(x_i)}{\partial\theta}\frac{y}{\psi'_\theta(x_i)}\right)\bigg|_{x_i=-\frac{\log u_i}{y}}.$$

Example 4. The Clayton Copula The generator function for the Clayton copula is

$$\psi_\theta(x)=(1+x)^{-\frac{1}{\theta}},\qquad \theta\in(0,\infty).$$

Then

$$\frac{\partial\psi_\theta(x)}{\partial\theta}=\frac{1}{\theta^2}\log(1+x)(1+x)^{-\frac{1}{\theta}},\quad \frac{\partial\psi'_\theta(x)}{\partial\theta}=\frac{1}{\theta^2}(1+x)^{-\frac{1}{\theta}-1}\left[1-\frac{1}{\theta}\log(1+x)\right],$$

and

$$\psi'_\theta(x)=-\frac{1}{\theta}(1+x)^{-\frac{1}{\theta}-1},\quad \psi''_\theta(x)=\frac{1}{\theta}\left(\frac{1}{\theta}+1\right)(1+x)^{-\frac{1}{\theta}-2}.$$

By the inverse Laplace transformation, we find that $Y_\theta\sim\Gamma(1/\theta,1)$, the gamma distribution with density $f_Y(y;\theta)=\frac{y^{\frac{1}{\theta}-1}e^{-y}}{\Gamma(1/\theta)}$, where $\Gamma(s):=\int_0^\infty t^{s-1}e^{-t}dt$, and the LR term is

$$\frac{\partial\log f_Y(y;\theta)}{\partial\theta}=-\frac{d\log\Gamma(1/\theta)}{d\theta}-\frac{1}{\theta^2}\log y.$$

The weight functions in the GLR-U estimator are

$$r_i(u;\theta,y) = -\frac{1}{\theta}u_i y\left(1-\frac{\log u_i}{y}\right)\log\left(1-\frac{\log u_i}{y}\right),$$
$$d_i(u_i;\theta,y) = \frac{1}{\theta}\left[1+(1-(x_i+1)y)\log(1+x_i)\right]\Big|_{x_i=-\frac{\log u_i}{y}}.$$

In addition, we have $\lim_{u_i\to 0^+} r_i(u;\theta,y) = \lim_{u_i\to 1^-} r_i(u;\theta,y) = 0$.

GLR-U for Ali-Mikhail-Haq copulas can be found in Appendix A. For both the Clayton and Ali-Mikhail-Haq copulas, conditions (A.3) and (A.5) in Theorem 1 are satisfied. If $\varphi(\cdot)$ is bounded, condition (A.4) in Theorem 1 can also be verified straightforwardly for any $y = Y_\theta$.

4. Applications

We apply the GLR-U method to distribution sensitivity estimation, and estimate sensitivities for stopping time problems and credit risk derivatives, with specific forms for the function $\varphi(\cdot)$.

4.1. Distribution Sensitivities

Discontinuous sensitivity estimation can be applied to calibrate complex stochastic models by GLR methods. For $g(\cdot;\theta):(0,1)^n\to\mathbb{R}$, we estimate the two following first-order distribution sensitivities:

$$\frac{\partial F(z;\theta)}{\partial\theta} = \frac{\partial\mathbb{E}[\mathbf{1}\{g(U;\theta)-z\le 0\}]}{\partial\theta} = \mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial\theta}\right],$$
$$f(z;\theta) = \frac{\partial\mathbb{E}[\mathbf{1}\{g(U;\theta)-z\le 0\}]}{\partial z} = \mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial z}\right],$$

where $f(\cdot;\theta)$ is the density function of $Z(\theta)=g(U;\theta)$ and $U_{-i} := (U_1,\dots,U_{i-1},U_{i+1},\dots,U_n)$, $i=1,\dots,n$. By applying GLR-U, we obtain

$$\mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial\theta}\right] = \mathbb{E}[G_{1,i}(U;z,\theta)],$$

where

$$\begin{aligned}G_{1,i}(U;z,\theta) :=& \mathbf{1}\{g(\overline{U}_i;\theta)-z\le 0\}r_i(\overline{U}_i;\theta) - \mathbf{1}\{g(\underline{U}_i;\theta)-z\le 0\}r_i(\underline{U}_i;\theta)\\ &+\mathbf{1}\{g(U;\theta)-z\le 0\}d(U;\theta),\end{aligned}$$

$$r_i(u;\theta) = \left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1} \frac{\partial g(u;\theta)}{\partial \theta},$$

and

$$d(u;\theta) = \left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1} \left[\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1} \frac{\partial g(u;\theta)}{\partial \theta} \frac{\partial^2 g(u;\theta)}{\partial u_i^2} - \frac{\partial^2 g(u;\theta)}{\partial u_i \partial \theta}\right].$$

We also obtain

$$\mathbb{E}\left[\frac{\partial \mathbb{E}[\mathbf{1}\{g(U_i, U_{-i};\theta) - z \le 0\}|U_{-i}]}{\partial z}\right] = \mathbb{E}[G_{2,i}(U;z,\theta)],$$

where

$$\begin{aligned} G_{2,i}(U;z,\theta) :=& \mathbf{1}\{g(\overline{U}_i;\theta) - z \le 0\}\tilde{r}_i(\overline{U}_i;\theta) - \mathbf{1}\{g(\underline{U}_i;\theta) - z \le 0\}\tilde{r}_i(\underline{U}_i;\theta) \\ &+ \mathbf{1}\{g(U;\theta) - z \le 0\}\tilde{d}(U;\theta), \end{aligned}$$

with

$$\tilde{r}_i(u;\theta) = -\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1} \quad \text{and} \quad \tilde{d}(u;\theta) = -\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-2} \frac{\partial^2 g(u;\theta)}{\partial u_i^2}.$$

Sufficient conditions for justifying unbiasedness of the estimators $G_{1,i}$ and $G_{2,i}$ and an optimal linear combination $G_{k,1}, \ldots, G_{k,n}$, $k = 1.2$ can be found in Appendix A.

4.2. Credit Risk Derivatives

We consider two important types of credit risk derivatives: basket default swaps (BDSs) and collateralized debt obligations (CDOs) [42]. In a BDS contract, the buyer pays fixed premia $p_1, ..., p_k$ to the protection seller at dates $0 < T_1 < \ldots < T_k < T$, and if the ith default time $\tau_{(i)}$ occurs before T, i.e., $\tau_{(i)} < T$, these premium payments stop, and the seller undertakes the loss of the ith default and makes a payment to the buyer. Let L_i be the loss of the ith default. The discounted value of the ith default swap is the difference between the discounted payments made by the seller and those made by the buyer:

$$V_{\text{bds}}(\tau) = V_{\text{value}}(\tau) - V_{\text{prot}}(\tau),$$

where $V_{\text{prot}}(\tau)$ is the discounted premium paid by the buyer:

$$V_{\text{prot}}(\tau) = \begin{cases} \sum_{j=1}^{\ell} p_j \exp(-rT_j) + p_{\ell+1} \exp(-r\tau_{(i)}) \frac{\tau_{(i)} - T_\ell}{T_{\ell+1} - T_\ell}, & \text{if } T_\ell \le \tau_{(i)} \le T_{\ell+1}, \\ \sum_{j=1}^{k} p_i \exp(-rT_j), & \text{if } \tau_{(i)} > T, \end{cases}$$

and $V_{\text{value}}(X)$ the discounted payment made by the seller:

$$V_{\text{value}}(X) = L_{(i)} \exp(-r\tau_{(i)}) \mathbf{1}\{\tau_{(i)} < T\}.$$

In a CDO, the losses caused by the defaults of the assets in the portfolio are packaged together and then tranched. The tranches are ordered so that losses are absorbed sequentially following the order of the tranches. For example, a tranche of a CDO absorbs the loss above a threshold $\mathcal{L}_-$ and below a threshold $\mathcal{L}_+$, i.e.,

$$V_{\text{cdo}}(\tau) = (\mathcal{L} - \mathcal{L}_-) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_-\} - (\mathcal{L} - \mathcal{L}_+) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_+\},$$

where

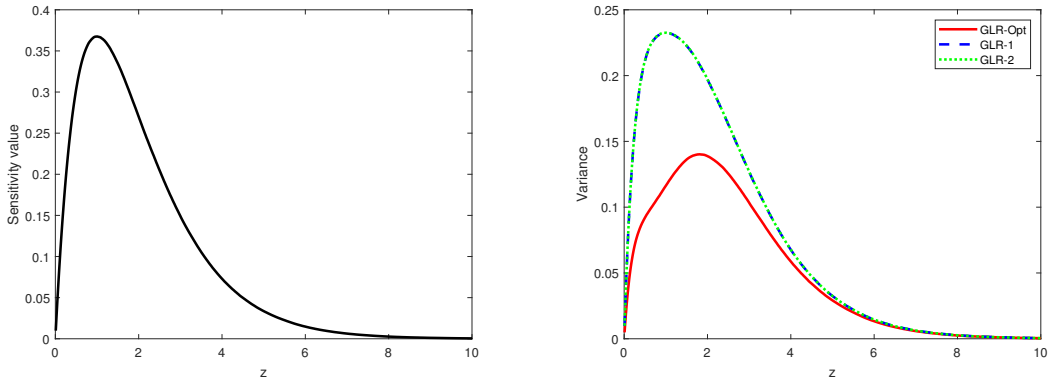
$$\mathcal{L} = \sum_{i=1}^{n} L_i \cdot \mathbf{1}\{\tau_i < T\}.$$

Suppose that the default times $(\tau_1, \ldots, \tau_n)$ have a joint distribution with marginal cdf's $F_i(\cdot)$, $i = 1, \ldots, n$, and a dependence structure modeled by an Archimedean copula. Then we can generate τ_i's by generating $V = (V_1, \ldots, V_n)$ from the copula, and putting $\tau_i = F_i^{-1}(V_i)$, $i = 1, \ldots, n$.

The sample performances $V_{\text{value}}(\tau)$, $V_{\text{prot}}(\tau)$, and $V_{\text{cdo}}(\tau)$ may be discontinuous with respect to the structural parameter θ in the copula model due to the presence of indicator functions and order indices. As a result, neither IPA nor LR can be applied directly for this model. On the other hand, $V_{\text{value}}(\tau)$, $V_{\text{prot}}(\tau)$, and $V_{\text{cdo}}(\tau)$ are all of the form $\varphi(g(U;\theta))$ that fits our framework, due to the generality of the measurable function $\varphi(\cdot)$. Unlike in [24] where a separate CMC technique needs to be derived for each type of cash flow, the GLR method in this work can estimate the derivative of the expectation for all three types of cash flows.

5. Numerical Experiments

In this section, we present numerical examples to demonstrate broad applicability and flexibility of the proposed GLR-U method to estimate sensitivities in various situations. The examples include an indicator function applied to a linear combination of two exponential random variables and a CDO model. The stochastic models in all examples have uniform random numbers as inputs, so that the GLR method in [34] does not apply. Therefore, for simplicity, we refer to GLR-U as GLR in this section. All experiments are implemented in Matlab on a laptop with an Intel i7-1065G7 CPU. The code for the numerical experiments in this paper can be found at https://github.com/pengyijie-pku/Generalized-Likelihood-Ratio.



(a) Distribution sensitivity curve.

(b) Variance curves.

Figure 1: Density curve estimated by GLR in the example of Section 5.1.

5.1. Distribution Sensitivities for a Linear Model

We estimate distribution sensitivities where $\varphi(\cdot)$ is an indicator function for a linear combination of two independent exponential random variables with means $1/\lambda_1$ and $1/\lambda_2$, i.e.,

$$\varphi(g(U;\theta)) = \mathbf{1}\{g(U;\theta) \le z\},$$

where

$$g(U;\theta) = -\frac{\theta}{\lambda_1}\log(U_1) - \frac{1}{\lambda_2}\log(U_2).$$

This sample performance falls into the special case stated in Corollary 1, and the conditions in Proposition 1 and the integrability condition on the weight function discussed in Section 4.1 can be checked straightforwardly. The specific forms of the estimators $G_{1,1}(\cdot), G_{1,2}(\cdot), G_{2,1}(\cdot), G_{2,2}(\cdot)$ for distribution sensitivities can be found in Appendix B.

For our numerical experiments, we take $\lambda_1 = 1$, $\lambda_2 = 1$, and $\theta = 1$, and we estimate the density of $g(U;\theta)$ at z, as a function of z from $z = 0.01$ to 10 with a step size of 0.01. We perform 10^6 independent simulation runs with GLR. The curves of the estimated density sensitivity and estimated variances as a function of z are given in Figure 1. We also estimate the sensitivity of $\mathbb{E}[\varphi(g(U;\theta))]$ with respect to θ at z, as a function of z from $z = 0.01$ to 10 with a step size of 0.01. The numerical observations are similar, and the details can be found in Appendix B.

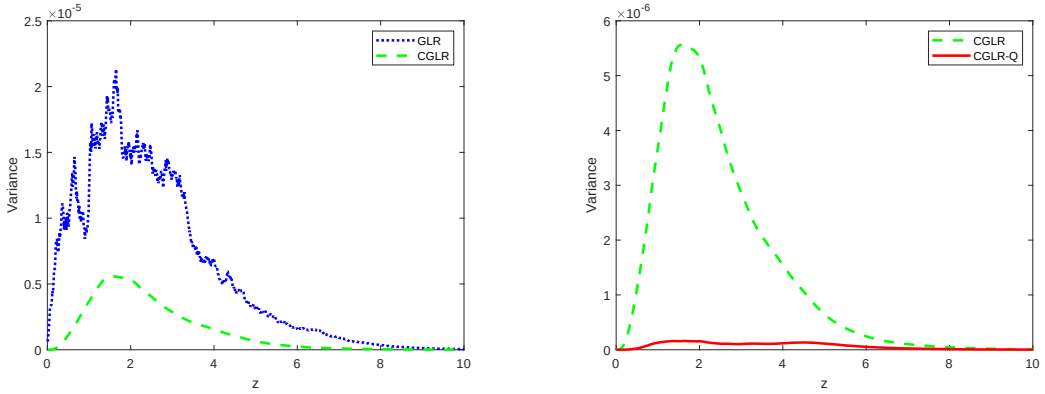
Figure 1(a) shows the estimated density sensitivity curve, and Figure 1(b) presents the sample variance curves of three distribution sensitivity estimators: $G_{2,1}(\cdot)$ (GLR-1), $G_{2,2}(\cdot)$ (GLR-2), and a combined GLR estimator that minimizes the variance of the linear combination of GLR-1 and GLR-2 with weights given by (GLR-Opt) in Appendix A. Due to space limitation, the variance comparison between GLR and the finite difference method with common random numbers (FDC) is relegated to Appendix B. The peak value of the variance curve of FDC(0.01) is about 200 times larger than that of GLR-Opt, and the variance of FDC(0.01) is about 10 times of that of FDC(0.1), which indicates that FDC suffers from a bias-variance tradeoff issue. The computational time of GLR is about 1.5 times of that of FDC.

The variance of GLR estimators can be reduced by applying CMC to these sensitivity estimators. CGLR is the conditional GLR method (GLR-Opt) with uniform random numbers as inputs, the specific form of which can be found in Appendix B. Then we apply RQMC to CGLR for estimating the density of $g(U;\theta)$, which basically replaces the input uniform random numbers with a RQMC point set to estimate the expectation. CGLR-Q is the CGLR method with the Sobol sequence scrambled by an algorithm of [51] in Matlab. We set 2^{13} for the size of the RQMC point set for CGLR-Q and the replications of input uniform random numbers for GLR and CGLR, and the variance of the estimators is estimated by 10^2 independent simulations. Figure 2(a) shows that the variance of CGLR is smaller than the variance of GLR over the entire range, and the peak value in the variance curve of GLR is about 4 times that of the variance curve of CGLR; Figure 2(b) shows that the variance of CGLR dominates the variance of CGLR-Q over the entire range, and the peak value in the variance curve of CGLR is around 40 times that of the variance of CGLR-Q. Similar comparison for estimating the derivative of $\mathbb{E}[g(U;\theta)]$ with respect to θ can be found in Appendix B.

5.2. Sensitivities of Collateralized Debt Obligations

We estimate the sensitivity with respect to the parameter θ that governs the dependence in the Clayton copula model for the expectation of the loss absorbed by the tranche that covers the first 30% of the total losses for 10 assets if there are defaults, i.e., $\mathcal{L}_- = 0$ and $\mathcal{L}_+ = 0.3 \times (\sum_{i=1}^{10} L_i)$. We set $r = 0.1$ and $T = 1$. The marginal distributions of the defaults are assumed to be exponential, so $\tau_i = -\frac{1}{\lambda_i}\log(X_i),\ i = 1,\ldots,10$. The parameters λ_i and loss L_i, $i = 1,\ldots,n$, are randomly generated from the uniform distribution over $(0,1)$ in the experiments.

We compare the GLR estimator with FDC(δ), where δ is the perturbation size. Due to the simplicity of the weight function of GLR, the computational time of GLR barely increases relative to that required to run the simulation model itself,



(a) Variance curves of GLR and CGLR. (b) Variance curves of CGLR and CGLR-Q.

Figure 2: Variance comparison between GLR, CGLR, and CGLR-Q estimators with an average of 2^{13} samples for estimating the density of $g(U;\theta)$ in the example of Section 5.1.

so the sensitivity estimate by GLR is almost a free byproduct that can be obtained simultaneously during the simulation. Table 1 shows the sensitivity estimates with sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$ under the Clayton copula, for $\theta = 0.5$. The variances of FDC(0.01) are about $10 - 50$ times those of GLR. For sample size $m = 10^6$, the estimate with FDC(0.1) lies outside of the 90% confidence interval of the GLR estimate, whereas the estimate with GLR lies in the 90% confidence interval of the FDC(0.01) estimate with the sample size $m = 10^7$, which is -0.179 ± 0.002. This indicates that FDC suffers from the bias-variance tradeoff, while GLR is accurate under a relatively small sample size. The computational time of GLR is about 5 times that of FDC. Numerical results for sensitivities of CDOs under the Ali-Mikhail-Haq copula and sensitivities of BDS under both the Clayton and Ali-Mikhail-Haq copulas can be found in Appendix B. The observations are similar to

	$m = 10^4$	$m = 10^5$	$m = 10^6$
GLR	-0.187 ± 0.01	-0.185 ± 0.004	-0.181 ± 0.002
FDC(0.1)	-0.171 ± 0.007	-0.177 ± 0.003	-0.176 ± 0.002
FDC(0.01)	-0.155 ± 0.07	-0.186 ± 0.02	-0.189 ± 0.006

Table 1: Sensitivity estimates of CDO with 10 assets governed by the Clayton copula with $\theta = 0.5$ based on 10^2 experiments (mean $\pm$ standard error).

those presented here.

6. Conclusions

In this paper, a GLR-U method is proposed for a family of stochastic models with uniform random numbers as inputs. The framework studied in this work covers a large range of discontinuities, and it includes many applications such as density estimation and credit risk financial derivatives. Since uniform random numbers are the basic building blocks for generating other random variables, our new method significantly relaxes the limitations on the input random variables in [34] and [38]. The technical conditions for justifying unbiasedness of GLR-U are relatively easy to satisfy in practice compared with previous conditions in [34] and [38], and we show how to verify them on illustrative examples. The variance of the GLR-U estimator can be reduced substantially (by a factor of more than 100 in one of our examples) by appropriately combining with CMC and RQMC. How to establish a general framework to apply CMC and RQMC to GLR for practical problems with different types of discontinuities in the sample performance is a good topic for future research.

7. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 72325007, 72250065 and 72022001, by the National Science Foundation (NSF) under Grants CMMI-1434419, CMMI-2027527, IIS-2123684, by the Air Force Office of Scientific Research under Grant FA95502010211, and by Discovery Grant RGPIN-2018-05795 from NSERC-Canada

References

[1] S. Asmussen, P. W. Glynn, Stochastic Simulation: Algorithms and Analysis, Vol. 57, Springer, 2007.

[2] P. W. Glynn, Likelihood ratio gradient estimation for stochastic systems, Communications of the ACM 33 (10) (1990) 75–84.

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

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

[5] P. W. Glynn, P. L'Ecuyer, Likelihood ratio gradient estimation for stochastic recursions, Advances in Applied Probability 27 (4) (1995) 1019–1053.

[6] P. L'Ecuyer, A unified view of the IPA, SF, and LR gradient estimation techniques, Management Science 36 (11) (1990) 1364–1383.

[7] Z. Cui, M. C. Fu, J.-Q. Hu, Y. Liu, Y. Peng, L. Zhu, On the variance of single-run unbiased stochastic derivative estimators, INFORMS Journal on Computing 32 (2) (2020) 390–407.

[8] M. C. Fu, Stochastic gradient estimation, Fu, Michael C. ed., in: Handbook of Simulation Optimization, Vol. 216, Springer, 2015, pp. 105–147, chapter 5.

[9] R. Suri, M. A. Zazanis, Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue, Management Science 34 (1) (1988) 39–64.

[10] M. C. Fu, J.-Q. Hu, Second derivative sample path estimators for the GI/G/m queue, Management Science 39 (3) (1993) 359–383.

[11] P. L'Ecuyer, P. W. Glynn, Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state, Management Science 40 (11) (1994) 1562–1578.

[12] M. C. Fu, Sample path derivatives for (s, S) inventory systems, Operations Research 42 (2) (1994) 351–364.

[13] S. Bashyam, M. C. Fu, Optimization of (s, S) inventory systems with random lead times and a service level constraint, Management Science 44 (12-part-2) (1998) S243–S256.

[14] M. C. Fu, J.-Q. Hu, Efficient design and sensitivity analysis of control charts using Monte Carlo simulation, Management Science 45 (3) (1999) 395–413.

[15] M. C. Fu, S. Lele, T. W. M. Vossen, Conditional Monte Carlo gradient estimation in economic design of control limits, Production and Operations Management 18 (1) (2009) 60–77.

[16] B. Heidergott, Optimisation of a single-component maintenance system: a smoothed perturbation analysis approach, European Journal of Operational Research 119 (1) (1999) 181–190.

[17] B. Heidergott, T. Farenhorst-Yuan, Gradient estimation for multicomponent maintenance systems with age-replacement policy, Operations Research 58 (3) (2010) 706–718.

[18] M. C. Fu, J.-Q. Hu, Sensitivity analysis for Monte Carlo simulation of option pricing, Probability in the Engineering and Informational Sciences 9 (3) (1995) 417–446.

[19] M. Broadie, P. Glasserman, Estimating security price derivatives using simulation, Management Science 42 (2) (1996) 269–285.

[20] G. Liu, L. J. Hong, Kernel estimation of the Greeks for options with discontinuous payoffs, Operations Research 59 (1) (2011) 96–108.

[21] Y. Wang, M. C. Fu, S. I. Marcus, A new stochastic derivative estimator for discontinuous payoff functions with application to financial derivatives, Operations Research 60 (2) (2012) 447–460.

[22] L. J. Hong, S. Juneja, J. Luo, Estimating sensitivities of portfolio credit risk using Monte Carlo, INFORMS Journal on Computing 26 (4) (2014) 848–865.

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

[24] L. Lei, Y. Peng, M. C. Fu, J.-Q. Hu, Copula sensitivity analysis for portfolio credit derivatives, European Journal of Operational Research 308 (2023) 455–466.

[25] L. J. Hong, Estimating quantile sensitivities, Operations Research 57 (1) (2009) 118–130.

[26] L. J. Hong, G. Liu, Simulating sensitivities of conditional value at risk, Management Science 55 (2) (2009) 281–293.

[27] M. C. Fu, L. J. Hong, J.-Q. Hu, Conditional Monte Carlo estimation of quantile sensitivities, Management Science 55 (12) (2009) 2019–2027.

[28] G. Jiang, M. C. Fu, Technical note — on estimating quantile sensitivities via infinitesimal perturbation analysis, Operations Research 63 (2) (2015) 435–441.

[29] B. Heidergott, W. Volk-Makarewicz, A measure-valued differentiation approach to sensitivity analysis of quantiles, Mathematics of Operations Research 41 (1) (2016) 293–317.

[30] S. Mohamed, M. Rosca, M. Figurnov, A. Mnih, Monte Carlo gradient estimation in machine learning, Journal of Machine Learning Research 21 (132) (2020) 1–62.

[31] Y. Peng, C.-H. Chen, M. Fu, Simulation optimization in the new era of AI, INFORMS TutORials in Operations Research (2023) 82–108.

[32] W.-B. Gong, Y.-C. Ho, Smoothed perturbation analysis of discrete event dynamical systems, IEEE Transactions on Automatic Control 32 (10) (1987) 858–866.

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

[34] Y. Peng, M. C. Fu, J.-Q. Hu, B. Heidergott, A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters, Operations Research 66 (2) (2018) 487–499.

[35] Y. Peng, M. C. Fu, J.-Q. Hu, On the regularity conditions and applications for generalized likelihood ratio method, in: Proceedings of the 2016 Winter Simulation Conference, IEEE Press, 2016, pp. 919–930.

[36] Y. Peng, M. C. Fu, P. W. Glynn, J.-Q. Hu, On the asymptotic analysis of quantile sensitivity estimation by Monte Carlo simulation, Proceedings of the 2017 Winter Simulation Conference (2017) 2336–2347.

[37] L. Lei, Y. Peng, M. C. Fu, J.-Q. Hu, Applications of generalized likelihood ratio method to distribution sensitivities and steady-state simulation, Journal of Discrete Event Dynamic Systems 28 (1) (2018) 109–125.

[38] Y. Peng, M. C. Fu, B. Heidergott, H. Lam, Maximum likelihood estimation by Monte Carlo simulation: Toward data-driven stochastic modeling, Operations Research 68 (6) (2020) 1896–1912.

[39] P. W. Glynn, Y. Peng, M. C. Fu, J. Hu, Computing sensitivities for distortion risk measures, INFORMS Journal on Computing 33 (4) (2021) 1520–1532.

[40] R. B. Nelsen, An Introduction to Copulas, 2nd Edition, Vol. 139, Springer Science & Business Media, 2006.

[41] D. Li, On default correlation: a copula approach, Journal of Fixed Income 9 (2000) 43–54.

[42] Z. Chen, P. Glasserman, Sensitivity estimates for portfolio credit derivatives using Monte Carlo, Finance and Stochastics 12 (4) (2008) 507–540.

[43] P. Embrechts, F. Lindskog, A. McNeil, Modelling dependence with copulas and applications to risk management, in: S. Rachev (Ed.), Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2003, pp. 329–384, chapter 8.

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

[45] A. W. Marshall, I. Olkin, Families of multivariate distributions, Journal of the American Statistical Association 83 (403) (1988) 834–841.

[46] C. Lemieux, Monte Carlo and Quasi-Monte Carlo Method, Series in Statitics. New York: Springer, 2009.

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

[48] R. C. Rogers, The calculus of several variables, Online [`http://www.math.nagoya-u.ac.jp/~{}richard/teaching/s2016/Ref2.pdf`] (2011).

[49] F. Puchhammer, P. L'Ecuyer, Likelihood ratio density estimation for simulation models, in: Proceedings of the 2022 Winter Simulation Conference, IEEE, 2022, pp. 109–120.

[50] P. L'Ecuyer, F. Puchhammer, A. Ben Abdellah, Monte Carlo and quasi–Monte Carlo density estimation via conditioning, INFORMS Journal on Computing 34 (3) (2022) 1729–1748.

[51] J. Matoušek, On the L_2-discrepancy for anchored boxes, Journal of Complexity 14 (4) (1998) 527–556.