

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

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

Yijie Peng

Department of Management Science and Information Systems, Guanghua School of Management, Peking University, Beijing, 100871, China, pengyijie@pku.edu.cn

Michael C. Fu

The Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, Maryland 20742-1815, USA, mfu@umd.edu

Jiaqiao Hu

Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600, USA, jqhu@ams.stonybrook.edu

Pierre L'Ecuyer

Department of Computer Science and Operations Research, University of Montreal, Montreal, Quebec, Canada, lecuyer@iro.umontreal.ca

Bruno Tuffin

INRIA, Univ Rennes, CNRS, IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, bruno.tuffin@inria.fr

We propose a new unbiased stochastic gradient estimator for a family of stochastic models with uniform random numbers, which are basic building blocks for stochastic simulation, as inputs. By extending the generalized likelihood ratio (GLR) method, the proposed estimator applies to discontinuous sample performances with structural parameters without requiring that the tails of the density of the input random variables go down to zero smoothly, an assumption made in previous work that precludes a direct formulation in terms of uniform random numbers as inputs. By overcoming this limitation, our new estimator greatly expands the applicability of the GLR method, which we demonstrate for several general classes of input random variates, including independent inverse transform random variates, random variables generated by acceptance-rejection methods and Markov chain Monte Carlo, and dependent input random variables governed by an Archimedean copula. We show how the new derivative estimator works in settings such as density estimation, distribution sensitivity for quantiles, and sensitivity analysis for Markov chain stopping time problems, which we illustrate with applications to stochastic activity networks and credit risk derivatives. Numerical experiments substantiate broad applicability and flexibility in dealing with discontinuities in sample performance.

Key words: simulation; stochastic derivative estimation; discontinuous sample performance; uniform random numbers, generalized likelihood ratio method.

MSC2000 subject classification:

OR/MS subject classification: Primary: ; secondary:

History:

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 performance function to simulate the mechanism of a system. 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 it must balance a bias-variance tradeoff and requires extra simulations. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two well-established unbiased derivative estimation techniques ([20], [30], [17], [58], [21]). IPA typically leads to lower variance than LR ([37, 6]), and the weak derivative method reduces the variance of LR at the cost of performing extra simulations ([55], [27]). [37] 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, and could in particular be a hybrid between IPA and LR. See [10] for a recent review.

Traditional applications of stochastic gradient estimation are in discrete event dynamic systems (DEDS), including queueing systems ([62], [12], [38]), inventory management ([9], [2]), statistical quality control ([15], [16]), maintenance systems ([25], [26]), and financial engineering and risk management such as computing financial derivatives ([13], [3], [43], [63], [32], [4], [41]), value-at-risk (VaR) and conditional VaR (CVaR) ([31], [33], [11], [35], [29]). Recently, stochastic gradient estimation techniques have attracted attention in machine learning and artificial intelligence; see [47] for a review paper written by a research team of Google's DeepMind. [54] show pathwise equivalence between IPA and backpropagation and how the computational complexity for estimating the gradient is reduced by propagating the errors backwardly along the artificial neural networks (ANN). An LR-based method is then proposed to train ANNs, which can improve the robustness in classifying images under both adversarial attacks and natural noise corruptions.

IPA requires continuity in the sample performance, whereas LR does not directly apply for 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 ([23], [14]), 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 ([58]), which can be achieved alternatively with the IPA-LR in [37], but it requires an explicit transformation. Recently, [53] proposed a generalized likelihood ratio (GLR) method that is capable of dealing 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 ([52], [53]). Distribution sensitivities, which mean 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 ([50], [40]). [51] derive GLR estimators for any order of distribution sensitivities and apply them to maximum likelihood estimation for complex stochastic models without requiring analytical likelihoods. [22] 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 that the density of the input distribution is known and that both tails of the density go down to zero smoothly and fast enough, which may not be satisfied in some applications, depending on what is interpreted as the input random variables. This smoothness requirement could sometimes be circumvented through a change of variables ([53]), however, such a transformation may lead to a GLR estimator with undesirable statistical properties, e.g., infinite variance; see the illustrative example on a stochastic activity network (SAN) in Appendix B. In this work, we relax this smoothness requirement and

establish the unbiasedness of 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 [53] where the surface integration part for the GLR estimator is zero, the surface integration part for the GLR estimator with uniform random numbers as inputs (GLR-U) in the present work is not necessarily zero but can be estimated by simulation. If the surface integration part is zero, we are able to relax certain integrability conditions given in [53] that are difficult to verify in practice.

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 is a generalization of LR from a different perspective than that of [53]. In a special case of [53], where independent input random variables are non-parameterized and the tails of their distribution go smoothly to zero fast enough, the GLR-U estimator with input random variables generated from the inverse transform of uniform random numbers coincides with the GLR estimator in [53]. Our GLR-U method can be applied to sensitivity analysis involving Markov chain Monte Carlo (MCMC) methods, often used for sampling from a complex probability distribution ([57]). The literature on sensitivity analysis for MCMC is sparse. An exception is [34], which provides an IPA estimator for a Gibbs-type MCMC method which requires to sample from conditional distributions of a joint distribution. Their method cannot handle more general Metropolis–Hastings algorithms which typically need to accept or reject samples, so that discontinuities and uniform random numbers naturally appear. 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 ([49]). The Gaussian copula has been widely used due to its simplicity ([42]), and sensitivity analysis for portfolio credit risk derivatives with joint defaults governed by a Gaussian copula has been studied in [5] 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 ([49]), not covered in [5], are relatively easy to simulate and can better capture the asymmetric tail dependence structure of the joint default data ([8]).

GLR-U can be used to estimate the distribution sensitivity functions, including the density and the quantile function together with confidence intervals, using a single batch of uniform random numbers. We formulate a stopping time problem with uniform random numbers as inputs in a Markov chain, and estimate its sensitivity by the GLR-U method. The acceptance-rejection method, a classic approach to sample from a general distribution ([7]), and control charts, an important technique in statistical quality control ([15]), fall into the framework of the stopping time problem. We provide numerical illustrations with various examples. Conditional Monte Carlo (CMC) methods can reduce the variance and smooth the performance function in simulation by conditioning on certain event 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 GLR estimator is presented in Section 3 with the specific forms of the estimators for four 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 a stochastic model of the following form:

$$\varphi(g(U; \theta)), \quad (1)$$

where $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a measurable function (not necessarily continuous), $g(\cdot; \theta) = (g_1(\cdot; \theta), \dots, g_n(\cdot; \theta))$ is a vector of functions $g_i : (0, 1)^n \rightarrow \mathbb{R}$ with certain smoothness properties to be made more precise shortly, and $U = (U_1, \dots, U_n)$ is a vector of i.i.d. $\mathcal{U}(0, 1)$ random variables (i.e., uniform over $(0, 1)$). For simplicity, we take θ as a scalar. When θ is a vector, each component of the gradient can be estimated separately using the method developed in this work. We consider the problem of estimating the following derivative:

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

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 $\varphi(\cdot)$. In [53], the stochastic model considered for the derivative estimation problem is $\varphi(g(X; \theta))$, where the density of $X = (X_1, \dots, X_n)$ is assumed to be known and both tails go down to zero smoothly and fast enough. This assumption is not satisfied by discontinuous densities, such as the uniform and exponential distributions, and we want to address this limitation. It is tempting to transform stochastic model (1) into $\varphi(g(F_1(Y_1), \dots, F_n(Y_n); \theta))$, where Y_i has continuous cumulative distribution function (cdf) $F_i(\cdot)$, $i = 1, \dots, n$, with tails going smoothly to zero, so that the problem can be put into the same form as the original stochastic model in [53]. 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 a GLR-U derivative estimator, we first introduce four important examples for generating random variables to illustrate potential applications of the stochastic model (1).

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

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

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 $\varphi(g(U; \theta)) = \varphi(F_1^{-1}(U_1; \theta), \dots, F_n^{-1}(U_n; \theta))$, where $g(u; \theta) = (F_1^{-1}(u_1; \theta), \dots, F_n^{-1}(u_n; \theta))$.

Example 2. Independent Non-parameterized Inputs Generated via the Inverse Transform Method Suppose $X = (X_1, \dots, X_n)$ 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 $\varphi(g(U; \theta)) = \varphi(h(F_1^{-1}(U_1), \dots, F_n^{-1}(U_n); \theta))$, where $h(x; \theta) := (h_1(x; \theta), \dots, h_n(x; \theta))$ and $g_i(u; \theta) := h_i(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n); \theta)$, $i = 1, \dots, n$.

Example 3. Markov Chain Monte Carlo MCMC methods are often used for sampling from a complex probability distribution. By constructing a Markov chain with a targeted equilibrium distribution, we can sample from a proposal distribution by recording states when the constructed

chain is mixed (“close” enough to the equilibrium distribution). In general, $\varphi_n(X)$ is a function of n states $X = (X_1, \dots, X_n)$ of a Markov chain defined by the following stochastic recurrence:

$$X_1 = g_1(U_1; \theta_1), \quad X_i = \kappa(X_{i-1}, g_i(U_i; \theta_i)), \quad i = 2, \dots, n, \quad (3)$$

where $g_i(\cdot)$ can depend on X_{i-1} . Here U_i can be a vector in general. For any parameter θ_i , the stochastic model $\varphi_n(X)$ falls into the framework of stochastic model $\varphi(g(U; \theta))$. GLR-U for sensitivity analysis of MCMC is based on transition function representation (3), which is similar to IPA for sensitivity analysis of MCMC. Classic LR for sensitivity analysis of MCMC is based on the transition probability representation of the Markov chain. LR usually leads to large variance compared with IPA ([6]), and the variance issue of LR would exacerbate in MCMC because the LR terms accumulate large variance as n grows large ([34])

The Metropolis–Hastings algorithm is the primary tool for simulating the Markov chain in MCMC ([57]). It works as follows. At step i , conditional on the current state X_{i-1} , it generates a proposed new state Y_i from a density $p_\theta(\cdot | X_{i-1})$, and the proposed new state is accepted with probability

$$\omega_\theta(X_{i-1}, Y_i) = \min \left\{ 1, \frac{\pi_\theta(Y_i) p_\theta(X_{i-1} | Y_i)}{\pi_\theta(X_{i-1}) p_\theta(Y_i | X_{i-1})} \right\},$$

where $\pi_\theta(\cdot)$ is the targeted equilibrium distribution. Then the transition function of the constructed Markov chain can be written as

$$X_i = X_{i-1} \mathbf{1} \{g_i(U_i; \theta) > 0\} + Y_i \mathbf{1} \{g_i(U_i; \theta) \leq 0\},$$

where $g_i(u_i; \theta) = u_i - \omega_\theta(X_{i-1}, Y_i)$. Discontinuities are introduced by the indicator function in the transition function, so IPA does not apply. The discontinuity issue cannot be directly handled by the method in [53], because the inputs are uniform random numbers U_i , $i = 1, \dots, n$. Transformation to meet the smoothness requirement of the input density in [53] would complicate function $g_i(\cdot)$, which is a linear function in our example, and this in turn may lead to high computational burden and undesirable statistical properties for the final estimators. In Appendix B, a classic LR estimator is derived for the Metropolis–Hastings algorithm based on a transition probability representation of the constructed Markov chain, but it leads to a much larger variance than GLR-U in an example of Section 5.3 with a moderately large n .

Example 4. Archimedean Copulas Copulas are a general way of representing the dependence in a multivariate distribution. A *copula* is any multivariate cdf whose one-dimensional marginals are all $\mathcal{U}(0, 1)$. It can be defined by a function $C(\cdot; \theta) : [0, 1]^n \rightarrow [0, 1]$ that satisfies certain conditions required for C to be a consistent cdf; (e.g., it cannot decrease, $C(u; \theta) = 0$ whenever one coordinate of u is 0, and $C(1; \theta) = 1$); see [49]. For any given copula and arbitrary marginal distributions with continuous cdf's $F_1(\cdot), F_2(\cdot), \dots, F_n(\cdot)$ with densities $f_i(\cdot)$, $i = 1, \dots, n$, one can define a multivariate distribution having exactly these marginals with joint cdf F_X given by $F_X(x) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n); \theta)$ for all $x := (x_1, \dots, x_n)$. [61] 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

$$f_X(x; \theta) = c(F_1(x_1), \dots, F_n(x_n); \theta) \prod_{i=1}^n f_i(x_i), \quad \text{where}$$

$$c(v; \theta) = \frac{\partial^n C(v; \theta)}{\partial v_1 \cdots \partial v_n},$$

$v = (v_1, \dots, v_n)$, and the derivative is interpreted as a Radon-Nikodym derivative when $C(\cdot; \theta)$ is not n th-order differentiable.

To generate $X = (X_1, \dots, X_n)$ from the joint cdf $F_X(\cdot)$, generate $V = (V_1, \dots, V_n)$ 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

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

where the generator function $\psi_\theta : [0, \infty) \rightarrow [0, 1]$ is a strictly decreasing convex function such that $\lim_{x \rightarrow \infty} \psi_\theta(x) = 0$, $\theta \in [0, \infty)$ is a parameter governing the strength of dependence, and $\psi_\theta^{[-1]}$ is a pseudo-inverse defined by $\psi_\theta^{[-1]}(x) = \mathbf{1}\{0 \leq x \leq \psi_\theta(0)\} \psi_\theta^{-1}(x)$, with the convention that $\psi_\theta^{-1}(0) = \inf\{x : \psi_\theta(x) = 0\}$. Archimedean copulas are absolutely continuous, and their densities have the form:

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

assuming that the generator function $\psi_\theta(\cdot)$ is smooth ([49]). In general we do not have an analytical expression for $c_a(\cdot; \theta)$, and it can be discontinuous in θ . Therefore, there is usually no analytical form, and discontinuities may exist for the density f_X with the copula defined by $c_a(\cdot; \theta)$, and LR and GLR in [53] typically do not apply to sensitivity analysis with Archimedean copulas.

[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 $i = 1, \dots, n$, let $V_i = \psi_\theta(-(\log U_i)/Y_\theta)$ 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:

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

where $\varphi(v_1, \dots, v_n) = \phi(F_1^{-1}(v_1), \dots, F_1^{-1}(v_n))$ and $g(u; \theta) = (\psi_\theta(-(\log u_1)/Y_\theta), \dots, \psi_\theta(-(\log u_n)/Y_\theta))$.

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 four 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 [53], i.e., $\varphi(g(X; \theta))$, where X is a random variable with density $f(\cdot; \theta)$ supported on Ω , and $\varphi : \mathbb{R} \mapsto \mathbb{R}$, $g : \Omega \times \mathbb{R} \mapsto \mathbb{R}$ are smooth functions. Assuming the derivative and expectation can be interchanged, we have

$$\begin{aligned} \frac{\partial \mathbb{E}[\varphi(g(X; \theta))]}{\partial \theta} &= \frac{\partial}{\partial \theta} \int_{\Omega} \varphi(g(x; \theta)) f(x; \theta) dx \\ &= \int_{\Omega} \underbrace{\varphi(g(x; \theta)) \frac{\partial \log f(x; \theta)}{\partial \theta}}_{LR \text{ part}} f(x; \theta) dx + \int_{\Omega} \underbrace{\frac{\partial \varphi(y)}{\partial y} \Big|_{y=g(x; \theta)} \frac{\partial g(x; \theta)}{\partial \theta}}_{IPA \text{ part}} f(x; \theta) dx, \end{aligned} \quad (4)$$

and the second term in the right-hand side of (4) can be further written as

$$\int_{\Omega} \frac{\partial \varphi(y)}{\partial y} \Big|_{y=g(x; \theta)} \frac{\partial g(x; \theta)}{\partial \theta} f(x; \theta) dx = \int_{\Omega} \frac{\partial \varphi(g(x; \theta))}{\partial x} \left(\frac{\partial g(x; \theta)}{\partial x} \right)^{-1} \frac{\partial g(x; \theta)}{\partial \theta} f(x; \theta) dx$$

$$\begin{aligned}
 &= \int_{\partial\Omega} \varphi(g(x;\theta)) \left(\frac{\partial g(x;\theta)}{\partial x} \right)^{-1} \frac{\partial g(x;\theta)}{\partial \theta} f(x;\theta) \hat{v} ds - \int_{\Omega} \varphi(g(x;\theta)) \frac{\partial}{\partial x} \left\{ \left(\frac{\partial g(x;\theta)}{\partial x} \right)^{-1} \frac{\partial g(x;\theta)}{\partial \theta} f(x;\theta) \right\} dx \\
 &= \int_{\partial\Omega} \varphi(g(x;\theta)) r(x;\theta) f(x;\theta) \hat{v} ds - \int_{\Omega} \varphi(g(x;\theta)) d(x;\theta) f(x;\theta) dx,
 \end{aligned} \tag{5}$$

where $\partial\Omega$ is the boundary of Ω , \hat{v} is the unit normal vector pointing outward, ds is the surface measure, the second equality holds by applying the Gauss-Green divergence theorem (Chapter 21 in [36]), and

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

The derivation above can be extended straightforwardly to the case when g and x are multi-dimensional. When $\varphi(\cdot)$ is smooth, (4) is essentially the IPA-LR framework in [37]. The difficulty lies in that $\varphi(\cdot)$ may be discontinuous, so that the IPA part in second term of (4) could be ill-defined. For a generic measurable function $\varphi(\cdot)$ not necessarily continuous, it is smoothed first in the derivation and then the limit is taken to retrieve unbiasedness under certain integrability conditions. Since the derivatives of $\varphi(\cdot)$ only appear in the intermediate steps but they are not in the final expression of (5), we only need the existence of a smoothed function for $\varphi(\cdot)$, which can be established under a mild regularity condition.

The first term of (4) and the second term of (5) can be estimated by simulation, but the first term of (5) is a surface integration that could be difficult to compute in general. The surface integration part in [53] is shown to be zero under certain conditions including that the tails of the input densities are required to go to zero smoothly and fast enough. In this work, we consider uniform random variables as inputs, so the support is a cube $\Omega = (0,1)^n$ whose boundary is simple. Thus, the surface normal vector \hat{v} is constant and the surface measure ds is amenable for computation on each face of the cube. Consequently, the surface integration part is included in the estimator and can be estimated by simulation. In addition, the input density $f(x;\theta) = 1$ for $x \in (0,1)^n$ and $\partial \log f(x;\theta) / \partial \theta = 0$. In (4), the first term vanishes and the second term is essentially the IPA estimator. Due to possible discontinuities in φ , this IPA part is retransformed into an LR-type estimator by integration by parts. In fact this can be seen in Section 3.3: in the most basic case, the estimator is not standard IPA but standard LR.

3.2. General Theory

Denote the Jacobian of g by

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

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

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

$$r_i(u;\theta) := \left(J_g^{-1}(u;\theta) \partial_\theta g(u;\theta) \right)^T e_i, \quad i = 1, \dots, n, \tag{6}$$

$$d(u; \theta) := \sum_{i=1}^n e_i^T J_g^{-1}(u; \theta) (\partial_{u_i} J_g(u; \theta)) J_g^{-1}(u; \theta) \partial_\theta g(u; \theta) - \text{trace}(J_g^{-1}(u; \theta) \partial_\theta J_g(u; \theta)), \quad (7)$$

where e_i is the i th 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 (6) and (7) for each dimension. Notice that the computation of the Jacobian matrix $J_g(\cdot; \theta)$, its inverse, and its derivatives with respect to u_i , which is the dominant computational burden for computing (6) and (7), needs to be done only once for all coordinates of θ . Therefore, the computational complexity of the GLR-U estimator is sublinear with respect to the number of coordinates of θ , so that it can be efficiently scaled up to multi-dimensional cases. 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 $h(x^-) := \lim_{x \rightarrow x^-} h(x)$ and $h(x^+) := \lim_{x \rightarrow x^+} h(x)$. 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 smooth function $\varphi_\epsilon(\cdot) \in \mathbb{C}^\infty$ and $p > 1$ such that

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

and if $n > 1$, for a fixed $\varepsilon > 0$ and any $u_i \in (0, 1) \setminus [\varepsilon, 1 - \varepsilon]$, $i = 1, \dots, n$,

$$\limsup_{\epsilon \rightarrow 0} \int_{\theta \in \Theta} \int_{(0,1)^{n-1}} |\varphi_\epsilon(g(u; \theta)) - \varphi(g(u; \theta))|^p du_{-i} = 0,$$

where $u_{-i} := (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n)$, and if $n = 1$, for any $u \in (0, 1) \setminus [\varepsilon, 1 - \varepsilon]$,

$$\limsup_{\epsilon \rightarrow 0} \int_{\theta \in \Theta} |\varphi_\epsilon(g(u; \theta)) - \varphi(g(u; \theta))| = 0.$$

(A.2) The following integrability condition holds:

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

(A.3) The Jacobian $J_g(u; \theta)$ is invertible almost everywhere (a.e.), and the performance function $g(u; \theta)$ is twice continuously differentiable with respect to $(u, \theta) \in (0, 1)^n \times \Theta$, where Θ is a compact neighborhood of the parameter θ of interest.

(A.4) The following integrability condition holds:

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

(A.5) Function $g(\cdot; \theta)$ is invertible, and

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

Remark 1. Condition (A.1) can be checked in certain settings when $\varphi_\epsilon(\cdot)$ can be explicitly constructed; see 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 ([58]). 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),*

$$\frac{\partial \mathbb{E}[\varphi(g(U; \theta))]}{\partial \theta} = \mathbb{E}[G(U; \theta)], \quad \text{where} \quad (8)$$

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

with $\bar{U}_i := (U_1, \dots, \underbrace{1^-}_{i\text{th element}}, \dots, U_n)$, $\underline{U}_i := (U_1, \dots, \underbrace{0^+}_{i\text{th element}}, \dots, U_n)$, and $r_i(\cdot)$ and $d(\cdot)$ defined by (6) and (7), respectively.

Remark 2. The proof of the theorem can be found in Appendix A. Even if $g(\bar{U}_i; \theta) = \infty$ 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 [53] on certain intermediate quantities (the smoothed function), which is difficult to verify in practice. 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 $\varepsilon \rightarrow 0$. We caution that the conditions in theorem may still be difficult to check when φ is too complicated to allow an explicit smoothed function or g comprises complex nonlinear functions.

We now examine the special case where $g(u; \theta) = (g_1(u_1; \theta), \dots, g_n(u_n; \theta))$, which covers Examples 1, 3, 4 in Section 2. For n independent uniform random numbers, we can take $g_i(u_i; \theta) = F_i^{-1}(u_i; \theta)$ for $i = 1, \dots, n$, 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 $i = 1, \dots, n$. In this special case, the Jacobian becomes

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

Then we have

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

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

Moreover, condition (A.1) in Theorem 1 can be replaced by a set of simpler assumptions when the performance function $\varphi(x)$ is a product of n indicators: $\varphi(x) = \prod_{i=1}^n \mathbf{1}\{x_i \leq 0\}$, 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 distribution sensitivities for the completion time in an SAN in Section 5.2 and the sensitivities involving acceptance-rejection and a control chart in Section 5.3 have performance functions which are products of n indicators. The proof of the following proposition can be found in Appendix A.

Proposition 1. *Consider the stochastic model*

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

Condition (A.1) holds if for $i = 1, \dots, n$ and a fixed $\varepsilon > 0$ in (9),

$$\inf_{\theta \in \Theta, u_i \in [\varepsilon, 1-\varepsilon]} \left| \frac{\partial g_i(u_i; \theta)}{\partial u_i} \right| > 0 \quad \text{and} \quad \inf_{\theta \in \Theta, u_i \in (0,1) \setminus [\varepsilon, 1-\varepsilon]} |g_i(u_i; \theta)| > 0. \quad (9)$$

Condition (9) implies no probability mass for $g_i(U_i; \theta)$. 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 $i = 1, \dots, n$, 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

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

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

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

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

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

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

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

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

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

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

Then conditions (A.2), (A.4), and (A.5) can be replaced by (A.2'), (A.4') and (A.5'), respectively, and condition (9) in Proposition 1 also simplifies to

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

Unbiasedness of the GLR-U estimators in many examples of this paper can be justified by verifying these simplified conditions.

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

$$\frac{\partial \mathbb{E}[\varphi(F_1^{-1}(U_1; \theta), \dots, F_n^{-1}(U_n; \theta))]}{\partial \theta}, \text{ for which the Jacobian is}$$

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

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

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

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

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

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

Therefore,

$$\begin{aligned} \frac{\partial \mathbb{E}[\varphi(F_1^{-1}(U_1; \theta), \dots, F_n^{-1}(U_n; \theta))]}{\partial \theta} &= \mathbb{E} \left[\varphi(F_1^{-1}(U_1; \theta), \dots, F_n^{-1}(U_n; \theta)) \frac{\partial \log f(U; \theta)}{\partial \theta} \right] \\ &= \mathbb{E} \left[\varphi(X) \sum_{i=1}^n \frac{\partial \log f_i(X_i; \theta)}{\partial \theta} \right]. \end{aligned}$$

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 ([20]). 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

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

The Jacobian can be expressed by $J_g(u; \theta) = J_h(x; \theta) \Lambda(x) \big|_{x=\beta(u)}$, where

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

assuming that the Jacobian $J_h(\cdot; \theta)$ of $h(\cdot; \theta)$ is invertible almost everywhere, and densities $f_i(x_i)$, $i = 1, \dots, n$, are differentiable and supported on the entire space, i.e. \mathbb{R} . Suppose $\lim_{x_i \rightarrow \pm\infty} (J_h^{-1}(x; \theta) \partial_\theta h(x; \theta))^T \Lambda^{-1}(x) e_i = 0$, which implies the tail of $f_i(\cdot)$ goes to zero fast enough, and then $\lim_{u_i \rightarrow 0} r_i(u; \theta) = \lim_{u_i \rightarrow 1} r_i(u; \theta) = 0$, $i = 1, \dots, n$. Notice that

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

where $\Lambda_i(x)$ is a diagonal matrix with $1/f_i(x_i)$ in the diagonal. Plugging the expressions of $J_g(u; \theta)$ and $\partial_{u_i} J_g(u; \theta)$ above into (7), we have

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

where $\zeta(x) = \left(\frac{\partial \log f_1(x_1)}{\partial x_1}, \dots, \frac{\partial \log f_n(x_n)}{\partial x_n} \right)^T$. Furthermore, $\widehat{d}(X; \theta) = d(F_1(X_1), \dots, F_n(X_n); \theta)$ coincides with the GLR estimator for stochastic model $\varphi(h(X; \theta))$ in [53].

3.5. Markov Chain Monte Carlo For stochastic model $\varphi_n(X)$ in Example 3 of Section 2, under similar conditions in Theorem 1,

$$\frac{\partial \mathbb{E}[\varphi_n(X)]}{\partial \theta_i} = \mathbb{E}[G_n^{(i)}(U; \theta_i)],$$

where $G_n^{(i)}(U; \theta) := \varphi_n(X_1^{+i}, \dots, X_n^{+i}) r_i(1^-; \theta) - \varphi_n(X_1^{-i}, \dots, X_n^{-i}) r_i(0^+; \theta) + \varphi_n(X) d_i(U_i; \theta)$,

with $\{X_j^{+i}, j \geq 1\}$ and $\{X_j^{-i}, j \geq 1\}$ being Markov chains generated by (3) with U_i being replaced by 1 and 0, respectively. In particular, for $\theta_i = \theta$, $i = 1, \dots, n$,

$$\frac{\partial \mathbb{E}[\varphi_n(X)]}{\partial \theta} = \sum_{i=1}^n \frac{\partial \mathbb{E}[\varphi_n(Z_1, \dots, Z_n)]}{\partial \theta_i} \Big|_{\theta_i = \theta} = \sum_{i=1}^n \mathbb{E}[G_n^{(i)}(U; \theta)].$$

Example 5. The Metropolis–Hastings Algorithm In this special case, $d_i(u_i; \theta) = 0$, and

$$r_i(u_i; \theta) = -\frac{\partial \omega_\theta(X_{i-1}, Y_i)}{\partial \theta}.$$

Markov chain $\{X_j^{+i}, j \geq 1\}$ is generated by (3) with the proposed state at step i rejected with certainty, and $\{X_j^{-i}, j \geq 1\}$ is a Markov chain generated by (3) with the proposed state at step i accepted with certainty.

3.6. Archimedean Copulas

We consider estimating

$$\frac{\partial}{\partial \theta} \mathbb{E} \left[\varphi \left(\psi_{\theta} \left(-\frac{\log U_1}{Y_{\theta}} \right), \dots, \psi_{\theta} \left(-\frac{\log U_n}{Y_{\theta}} \right) \right) \right],$$

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

$$\begin{aligned} & \frac{\partial}{\partial \theta} \mathbb{E} \left[\varphi \left(\psi_{\theta} \left(-\frac{\log U_1}{Y_{\theta}} \right), \dots, \psi_{\theta} \left(-\frac{\log U_n}{Y_{\theta}} \right) \right) \right] \\ &= \mathbb{E} \left[\varphi \left(\psi_{\theta} \left(-\frac{\log U_1}{Y_{\theta}} \right), \dots, \psi_{\theta} \left(-\frac{\log U_n}{Y_{\theta}} \right) \right) \frac{\partial \log f_Y(y; \theta)}{\partial \theta} \Big|_{y=Y_{\theta}} \right] \\ &+ \mathbb{E} \left[\frac{\partial \mathbb{E} [\varphi(\psi_{\theta}(-\frac{\log U_1}{y}), \dots, \psi_{\theta}(-\frac{\log U_n}{y}))]}{\partial \theta} \Big|_{y=Y_{\theta}} \right], \end{aligned} \quad (10)$$

where $f_Y(\cdot; \theta)$ is the density function of Y_{θ} . This equality follows from Theorem 1 of [37] with ω in that theorem being replaced by y and $h(\theta, \omega)$ being replaced by $h(y, \theta) := \mathbb{E}[\varphi(\psi_{\theta}(-\ln U_1/y), \dots, \psi_{\theta}(-\ln U_n/y))]$, under the assumption that for all y , this last expectation is continuous in θ and differentiable except perhaps on a countable set. Specifically, (10) can be rewritten as

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{\mathbb{R}} h(y, \theta) F_Y(dy; \theta) &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}} h(y, \theta) L(y, \theta, \theta_0) F_Y(dy; \theta_0) \\ &= \int_{\mathbb{R}} \left(h(y, \theta) \frac{\partial L(y, \theta, \theta_0)}{\partial \theta} + \frac{\partial h(y, \theta)}{\partial \theta} L(y, \theta, \theta_0) \right) F_Y(dy; \theta_0), \end{aligned}$$

where $L(y, \theta, \theta_0) := f_Y(y; \theta) / f_Y(y; \theta_0)$. The first term on the right-hand side of (10) can be dealt with by the LR method straightforwardly if $f_Y(\cdot; \theta)$ admits an analytical form. [18] show how to apply the LR method with only the Laplace transform $\psi_{\theta}(\cdot)$.

We now show how to use GLR-U to handle the second term on the right-hand side of (10) with Y_{θ} fixed and generated from other uniform random numbers. The Archimedean copula model falls into the special case where $g(u; \theta) = (g_1(u_1; \theta), \dots, g_n(u_n; \theta))$, discussed after Theorem 1. The Jacobian in this case is

$$J_g(u; \theta, y) = \begin{pmatrix} -\frac{1}{u_1 y} \psi'_{\theta} \left(-\frac{\log u_1}{y} \right) & 0 & \dots & 0 \\ 0 & -\frac{1}{u_2 y} \psi'_{\theta} \left(-\frac{\log u_2}{y} \right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\frac{1}{u_n y} \psi'_{\theta} \left(-\frac{\log u_n}{y} \right) \end{pmatrix}, \quad \text{and}$$

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

The weight functions in the GLR-U estimator are

$$\begin{aligned} r_i(u; \theta, y) &= -\frac{u_i y}{\psi'_{\theta}(x_i)} \frac{\partial \psi_{\theta}(x_i)}{\partial \theta} \Big|_{x_i = -\frac{\log u_i}{y}}, \\ d_i(u_i; \theta, y) &= \left(-\frac{1}{\psi'_{\theta}(x_i)} \frac{\partial \psi'_{\theta}(x_i)}{\partial \theta} + \frac{\psi''_{\theta}(x_i)}{(\psi'_{\theta}(x_i))^2} \frac{\partial \psi_{\theta}(x_i)}{\partial \theta} + \frac{\partial \psi_{\theta}(x_i)}{\partial \theta} \frac{y}{\psi'_{\theta}(x_i)} \right) \Big|_{x_i = -\frac{\log u_i}{y}}. \end{aligned}$$

Example 6. The Clayton Copula The generator function for the Clayton copula is

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

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

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

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

$$r_i(u; \theta, y) = -\frac{1}{\theta} u_i y \left(1 - \frac{\log u_i}{y} \right) \log \left(1 - \frac{\log u_i}{y} \right),$$

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

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

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 For $g(\cdot; \theta) : (0, 1)^n \rightarrow \mathbb{R}$, we estimate the two following first-order distribution sensitivities:

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

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

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

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

$$G_{1,i}(U; z, \theta) := \mathbf{1}\{g(\bar{U}_i; \theta) - z \leq 0\} r_i(\bar{U}_i; \theta) - \mathbf{1}\{g(\underline{U}_i; \theta) - z \leq 0\} r_i(\underline{U}_i; \theta) + \mathbf{1}\{g(U; \theta) - z \leq 0\} d(U; \theta),$$

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

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

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

$$G_{2,i}(U; z, \theta) := \mathbf{1}\{g(\bar{U}_i; \theta) - z \leq 0\} \tilde{r}_i(\bar{U}_i; \theta) - \mathbf{1}\{g(\underline{U}_i; \theta) - z \leq 0\} \tilde{r}_i(\underline{U}_i; \theta) \\ + \mathbf{1}\{g(U; \theta) - z \leq 0\} \tilde{d}(U; \theta), \quad \text{with}$$

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

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

Example 7. Distribution Sensitivities for Quantiles For $0 \leq \alpha \leq 1$, the α -VaR (or α -quantile) of a random variable $Z(\theta) = g(U; \theta)$ with cdf $F(\cdot; \theta)$ is defined as

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

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 $U \sim \mathcal{U}(0, 1)^d$, and $\hat{F}_m(\cdot)$ be the empirical distribution of $Z_j := g(U^{(j)}; \theta)$, $j = 1, \dots, m$. The empirical α -quantile $\hat{F}_m^{-1}(\alpha)$, which is the inverse of the empirical distribution evaluated at α , is simply $Z_{(\lceil \alpha m \rceil)}$, where $Z_{(1)} < \dots < Z_{(m)}$ are the realizations of Z_1, \dots, Z_m sorted in increasing order (the order statistics), and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x . This empirical quantile satisfies the following central limit theorem ([60]):

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

Traditionally, batching and sectioning techniques are used to estimate the asymptotic variance to construct a confidence interval on the empirical quantile, and these methods lead to subcanonical convergence rates ([48]). With the GLR-U density estimator, however, we can estimate the asymptotic variance by

$$\frac{m\alpha(1-\alpha)}{\sum_{j=1}^m G_{2,i}(U^{(j)}; z, \theta)|_{z=\hat{F}_m^{-1}(\alpha)}},$$

using the same realizations of the uniform random variables $U^{(j)}$ as in the quantile estimator $\hat{F}_m^{-1}(\alpha)$. It follows from [50] that this asymptotic variance estimator is consistent. Another application of distribution sensitivity estimation is to estimate quantile sensitivity $\partial q_\alpha(\theta) / \partial \theta = - \frac{\partial F(z; \theta)}{\partial \theta} \Big|_{z=q_\alpha(\theta)} / f(q_\alpha(\theta); \theta)$ ([11]).

4.2. Stopping Time Problems In this subsection, we consider estimating the derivative of the expectation of a sample performance that depends on a stopping time N :

$$\frac{\partial \mathbb{E}[h_N(X_1, \dots, X_N)]}{\partial \theta_i}, \tag{11}$$

where N is the first time the Markov chain $\{X_i : i \in \mathbb{Z}^+\}$ defined by (3) hits a set \mathcal{X} , i.e.,

$$N = \min \{n \in \mathbb{N} : X_n \in \mathcal{X}\} = \min \{n \in \mathbb{N} : \kappa(X_{n-1}, g_n(U_n; \theta_n)) \in \mathcal{X}\}.$$

By decomposing the probability space into disjoint events $\{N = n\} = \cap_{j=1}^{n-1} \{X_j \notin \mathcal{X}\} \cup \{X_n \in \mathcal{X}\}$, $n \in \mathbb{Z}^+$, the expectation of this sample performance can be rewritten as

$$\mathbb{E}[h_N(X_1, \dots, X_N)] = \sum_{n=1}^{\infty} \mathbb{E}[\varphi_n(X_1, \dots, X_n)], \quad (12)$$

$$\text{where } \varphi_n(X_1, \dots, X_n) = h_n(X_1, \dots, X_n) \prod_{j=1}^{n-1} \mathbf{1}\{X_j \notin \mathcal{X}\} \mathbf{1}\{X_n \in \mathcal{X}\}.$$

The derivative estimation problem of each expectation in the summation on the right-hand side of (12) has been discussed in Section 3.5, and we have

$$\frac{\partial \mathbb{E}[h_N(X_1, \dots, X_N)]}{\partial \theta_i} = \sum_{n=i}^{\infty} \mathbb{E}[G_n^{(i)}(U_1, \dots, U_n; \theta_i)].$$

We can use the randomization technique of [56] to obtain a single-run unbiased estimator of (11), which is given by

$$\frac{1}{q(N')} G_{N'}^{(i)}(U_1, \dots, U_{N'}; \theta_i), \quad (13)$$

where N' is a discrete random variable supported on integer set containing $\{i, i+1, \dots\}$ with probability mass function $q(\cdot)$. In particular, for $\theta_i = \theta$, $i \in \mathbb{Z}^+$, an unbiased estimator is $\sum_{i=1}^{N'} G_{N'}^{(i)}(U_1, \dots, U_{N'}; \theta) / q(N')$ with $q(\cdot)$ supported on \mathbb{Z}^+ .

This stopping time problem generalizes those in [19] and [28] by allowing the distribution of the stopping time N to depend on parameter θ . The classical IPA and LR methods do not cover this case.

Example 8. Acceptance-Rejection Acceptance-rejection is a classic approach to sample a random variable X from a general distribution with density $\pi_\theta(\cdot)$ ([7]). Suppose there is a constant c_θ such that $\pi_\theta(x) \leq p_\theta(x)/c_\theta$, $\forall x$. We first generate random variables $\{Y_i, i \in \mathbb{Z}^+\}$ from a proposal distribution with density $p_\theta(\cdot)$, and X is the first Y_i such that

$$U_i \leq \omega_\theta(Y_i), \quad \omega_\theta(Y_i) := c_\theta \pi_\theta(Y_i) / p_\theta(Y_i).$$

Any sample performance $h(X)$ can be generated by the acceptance-rejection method as

$$h(X) = \sum_{n=1}^{\infty} \prod_{i=1}^{n-1} \mathbf{1}\{g_i(U_i; \theta) > 0\} \mathbf{1}\{g_n(U_n; \theta) \leq 0\} h(Y_n),$$

where $g_i(u_i; \theta) = u_i - \omega_\theta(Y_i)$. Here again discontinuities are introduced by the indicator functions, and uniform random numbers are a natural choice of inputs. Conditional on the filtration generated by $\{Y_i, i \in \mathbb{Z}^+\}$, sensitivity analysis for the sample performance generated by the acceptance-rejection method falls into a special case of sensitivity analysis for the stopping time problem, and conditions (A.2)-(A.5) and the condition in Proposition 1 are satisfied.

Example 9. Controls Charts Control charts aim to detect (statistically) when a manufacturing or business process goes out of control. The system is assumed to output samples having different statistical distributions when in control versus when out of control. The systems goes out of control at the (unobservable) random time χ with cdf $F_0(\cdot)$. The output Z_i of the i th sample has cdf

$F_1(\cdot)$ when in control and $F_2(\cdot)$ when out of control. For an exponentially weighted moving average (EWMA) chart, the (observable) test statistic after the i th sample is $Y_i = \alpha Z_i + (1 - \alpha)Y_{i-1}$. The stopping time N is the time when the system is declared out of control: $N = \min\{i : Y_i \notin [\theta_1, \theta_2]\} = \min\{i : X_i \notin [0, 1]\}$, where θ_1 and θ_2 are fixed lower and upper control limits for the test statistic, and $X_i = \frac{Y_i - \theta_1}{\theta_2 - \theta_1}$. Note that $\{X_i, i \geq 1\}$ is a Markov chain that follows the recursion

$$X_i = (1 - \alpha)X_{i-1} + g_i(U_i; \theta), \quad \text{where} \\ g_i(U_i; \theta) = \frac{\alpha}{\theta_2 - \theta_1} [\mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(U_i) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(U_i) - \theta_1],$$

and Δ is the sampling period (the time between any two successive monitoring epochs). Here the density for distribution F_i , e.g., exponential distribution, may not satisfy the smoothness requirement in [53].

Here we consider sensitivity analysis with respect to $\theta = \theta_2$. This model falls into the special case of Corollary 1 where $g_i(u_i; \theta) = \xi_i(\theta)\eta_i(u_i)$, with $\xi_i(\theta) = \alpha/(\theta - \theta_1)$ and $\eta_i(u_i) = \mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(u_i) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(u_i) - \theta_1$, for $i = 1, \dots, n$. Then we have

$$r(u; \theta) = -\frac{1}{\theta_2 - \theta_1} [\mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(u_i) f_1(F_1^{-1}(u_i)) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(u_i) f_2(F_2^{-1}(u_i)) - \theta_1], \quad \text{and} \\ d(u; \theta) = \frac{1}{\theta_2 - \theta_1} \sum_{i=1}^n \left[\mathbf{1}\{i < \chi/\Delta\} \frac{F_1^{-1}(u_i)}{f_1(F_1^{-1}(u_i))} + \mathbf{1}\{i \geq \chi/\Delta\} \frac{F_2^{-1}(u_i)}{f_2(F_2^{-1}(u_i))} + 1 \right].$$

4.3. Credit Risk Derivatives We consider two important types of credit risk derivatives: basket default swaps (BDSs) and collateralized debt obligations (CDOs). In a BDS contract, the buyer pays fixed premia p_1, \dots, p_k to the protection seller at dates $0 < T_1 < \dots < T_k < T$, and if the i th default time $\tau_{(i)}$ occurs before T , i.e., $\tau_{(i)} < T$, these premium payments stop, and the seller undertakes the loss of the i th default and makes a payment to the buyer. Let L_i be the loss of the i th default. The discounted value of the i th 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:

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

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 tranching. 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.,

$$V_{\text{cdo}}(\tau) = (\mathcal{L} - \mathcal{L}_-) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_-\} - (\mathcal{L} - \mathcal{L}_+) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_+\}, \quad \text{where} \\ \mathcal{L} = \sum_{i=1}^n L_i \cdot \mathbf{1}\{\tau_i < T\}.$$

Suppose that the vector of default times (τ_1, \dots, τ_n) have a joint distribution with marginal cdf's $F_i(\cdot)$, $i = 1, \dots, n$, and a dependence structure modeled by an Archimedean copula. Then we can generate τ_i 's by generating $V = (V_1, \dots, V_n)$ from the copula, and putting $\tau_i = F_i^{-1}(V_i)$, $i = 1, \dots, n$.

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 [41] 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, SAN, MCMC, and a CDO model. Additional examples including acceptance-rejection and control charts can be found in Appendix B. The stochastic models in all examples have uniform random numbers as inputs, so that the GLR method in [53] 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>.

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.,

$$\varphi(g(U; \theta)) = \mathbf{1}\{g(U; \theta) \leq z\} \quad \text{where} \quad g(U; \theta) = -\frac{\theta}{\lambda_1} \log(U_1) - \frac{1}{\lambda_2} \log(U_2).$$

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 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 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 page limit, 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.

In Figure 2(a), the solid line in the center is the estimate for the α -quantiles q_α of $g(U; \theta)$, $\alpha = 0.1, \dots, 0.9$, using a batch of 10^4 independent simulation replications, and the upper and lower dotted lines are calculated respectively by adding and subtracting twice the estimated standard deviation to the estimated quantile values, which are estimated by GLR using the same batch of 10^4 independent simulation replications for estimating the quantiles. Figure 2(b) presents the coverage rates of the 90%-confidence intervals for quantiles q_α , $\alpha = 0.1, \dots, 0.9$, by 10^4 independent macro experiments. The true quantile values can be calculated by inverting a hypoexponential

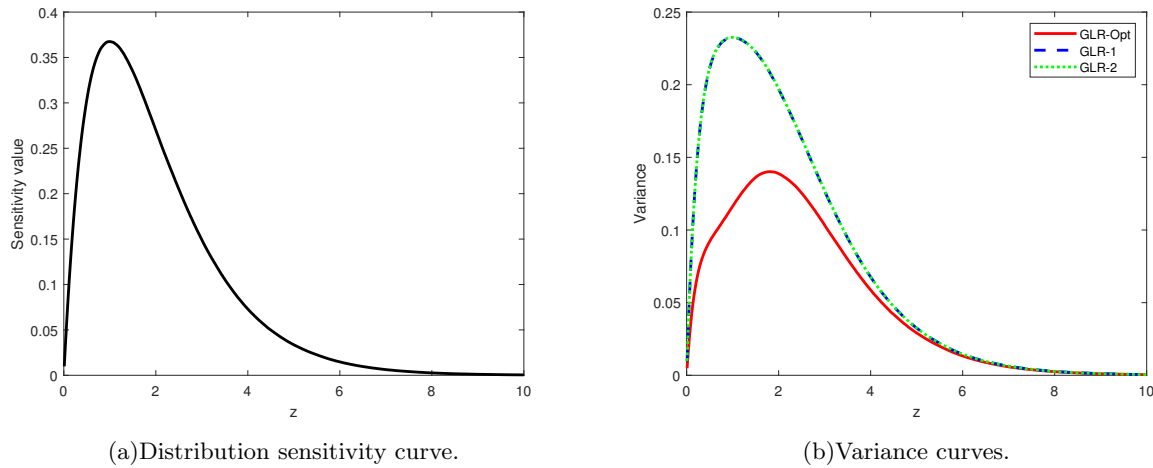


FIGURE 1. Density curve estimated by GLR in the example of Section 5.1.

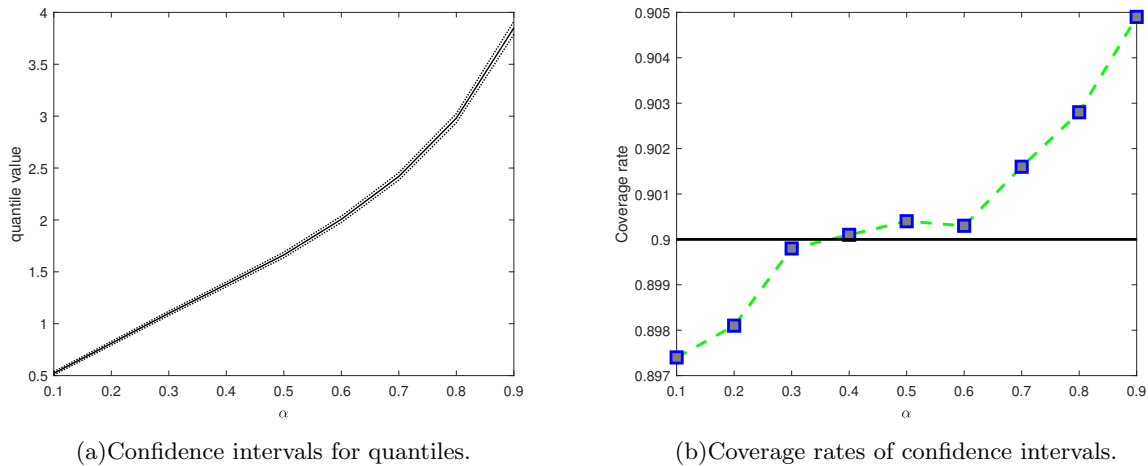


FIGURE 2. 90%-confidence intervals estimated by GLR for quantiles and coverage rates in example of Section 5.1.

distribution ([39, Section 4.6]). We can see the coverage rates of the 90%-confidence intervals match the target value statistically. As the quantile gets closer to the tail of the distribution, the variances of the quantile estimates and confidence interval estimates are larger because there are many fewer samples in the tail.

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 [46] 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 3(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 3(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.

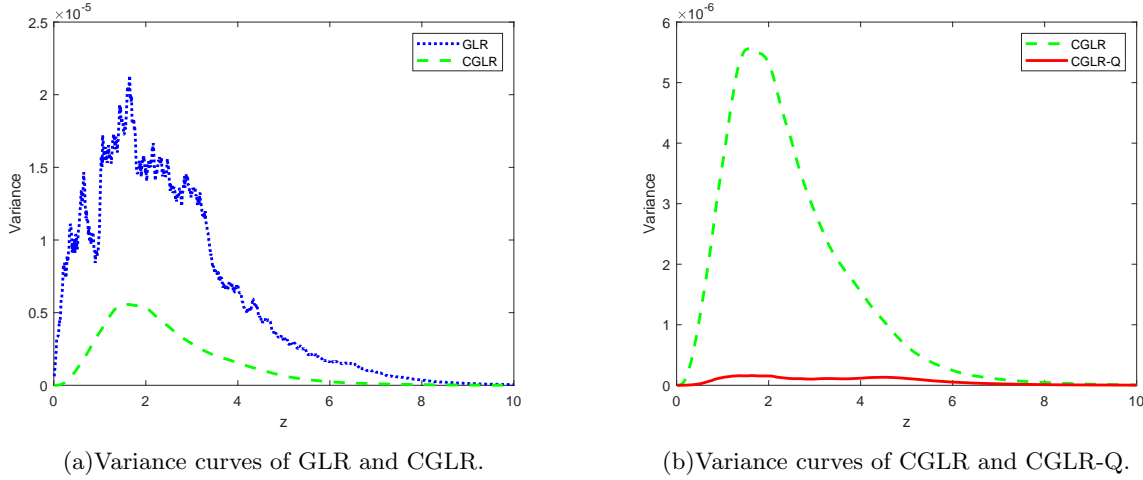


FIGURE 3. Variance comparison between GLR, CGLR, and CGLR-Q estimators with an average of 2^{13} samples for estimating the density of $g(U; \theta)$ in the example of Section 5.1.

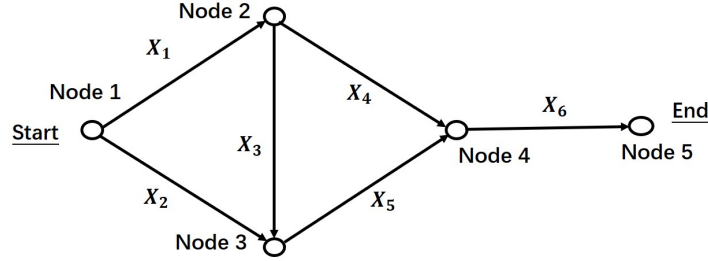


FIGURE 4. A SAN with six activities.

Similar comparison for estimating the derivative of $\mathbb{E}[g(U; \theta)]$ with respect to θ can be found in Appendix B.

5.2. Distribution Sensitivities for a Stochastic Activity Network We estimate distribution sensitivities for the sample performance of a small SAN depicted in Figure 4. There are five nodes representing different stages of activity. The nodes are connected by the arcs representing the activities in each stage. The durations of activities follow independent exponential distributions, i.e., $X_i = -\frac{1}{\lambda_i} \log(U_i)$, $i = 1, \dots, 6$. Let $\theta = \lambda_6$. There are three different paths representing the tasks to reach the final stage of a project, i.e., $\pi_1 = (1, 4, 6)$, $\pi_2 = (2, 5, 6)$, $\pi_3 = (1, 3, 5, 6)$, and the completion time for each path is additive, i.e., $\sum_{j \in \pi_i} X_j$, $i = 1, 2, 3$. The completion time for the entire project is $\max(X_1 + X_4 + X_6, X_2 + X_5 + X_6, X_1 + X_3 + X_5 + X_6)$, and the sample performance for the distribution function of completion time is

$$\begin{aligned} \varphi(g(U; \theta)) &= \mathbf{1} \{ \max(X_1 + X_4 + X_6, X_2 + X_5 + X_6, X_1 + X_3 + X_5 + X_6) \leq z \} \\ &= \mathbf{1} \{ X_1 + X_4 + X_6 \leq z \} \mathbf{1} \{ X_2 + X_5 + X_6 \leq z \} \mathbf{1} \{ X_1 + X_3 + X_5 + X_6 \leq z \}, \end{aligned}$$

where $\varphi(v_1, v_2, v_3) = \prod_{i=1}^3 \mathbf{1} \{ v_i \leq 0 \}$, and we have

$$\begin{aligned} g_i(u; \theta) &= - \sum_{j \in \pi_i} \frac{1}{\lambda_j} \log u_j - z, \quad i = 1, 2, 3, \\ \partial_z g(u; \theta) &= - \left(1, 1, 1 \right)^T \quad \text{and} \quad \partial_\theta g(u; \theta) = \frac{\log u_6}{\theta^2} \left(1, 1, 1 \right)^T. \end{aligned}$$

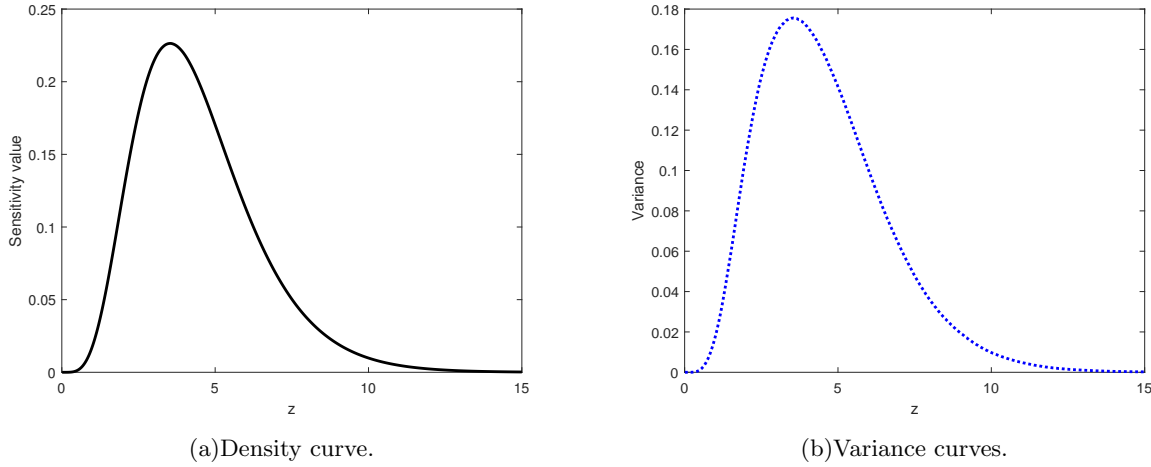


FIGURE 5. Density estimation by GLR in the SAN example of Section 5.2.

This sample performance goes beyond the setting in Section 4.1, but it can be put under the more general stochastic model (1). In the theory of GLR, the dimension of the vector of the input uniform random numbers is assumed to be the same as that of the argument vector of function g . For the SAN model, there are six input uniform random numbers, while the dimension of the argument vector of g is three. Therefore, we can arbitrarily choose three uniform random numbers to condition on and treat the remaining three as the inputs to the stochastic models for deriving GLR. We condition on (U_4, U_5, U_6) and treat (U_1, U_2, U_3) as the input uniform random numbers for deriving GLR. The Jacobian matrix is

$$J_g(u; \theta, z) = - \begin{pmatrix} \frac{1}{\lambda_1 u_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2 u_2} & 0 \\ \frac{1}{\lambda_1 u_1} & 0 & \frac{1}{\lambda_3 u_3} \end{pmatrix}.$$

The GLR estimator for $\mathbb{E}[\varphi(g(U; \theta))]$ with respect to θ can be obtained by multiplying the GLR estimator for the density by $-\log U_6/\theta^2$. Here we only report on the performance of the GLR estimators for the density of $g(U; \theta)$. Additional results for quantiles can be found in Appendix B. We have

$$r_1(u; \theta, z) = \lambda_1 u_1, \quad r_2(u; \theta, z) = \lambda_2 u_2, \quad r_3(u; \theta, z) = 0, \quad \text{and} \quad d(u; \theta, z) = -\lambda_1 - \lambda_2.$$

In this case, condition (A.1) in Theorem 1 can be justified by checking conditions in Proposition 1. Conditions (A.2) and (A.3) can also be checked straightforwardly.

In the experiments, we set $\lambda_i = 1$, $i = 1, \dots, 6$, in the stochastic model, and estimate distribution sensitivity curves from $z = 0.01$ to 15 with a step size of 0.01. Figure 5 presents the density and variance curves estimated by the GLR method using 10^6 independent simulation replications. The numerical comparisons among FDC, GLR, CGLR, CGLR-Q can be found in Appendix B.

5.3. Sensitivities for Metropolis–Hastings Algorithm For the Metropolis–Hastings algorithm, we choose the target distribution to be normal with density $\pi_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp(-(x - \theta)^2/2)$ and the proposal distribution to be normal with density $p(y|x) = \frac{1}{\sqrt{2\pi}\delta} \exp(-\frac{1}{2\delta^2}(y - x)^2)$. We estimate the derivative of the first two moments with respect to θ , which are 1 and 0. Then $\omega_\theta(X_{i-1}, Y_i) = \min \left\{ 1, \frac{\pi_\theta(Y_i)}{\pi_\theta(X_{i-1})} \right\}$ and

$$\frac{\partial \omega_\theta(X_{i-1}, Y_i)}{\partial \theta} = (Y_i - X_{i-1}) \exp \left(\frac{(X_{i-1} - \theta)^2}{2} - \frac{(Y_i - \theta)^2}{2} \right) \mathbf{1} \left\{ \exp \left(\frac{(X_{i-1} - \theta)^2}{2} - \frac{(Y_i - \theta)^2}{2} \right) < 1 \right\}.$$

	GLR	FDC(0.01)	FDC(0.1)
$\partial\mathbb{E}[X_n]/\partial\theta \approx 1$	0.999 ± 0.001	0.99 ± 0.02	0.988 ± 0.004
$\partial\mathbb{E}[X_n^2]/\partial\theta \approx 0$	0.002 ± 0.003	-0.05 ± 0.04	0.097 ± 0.008

TABLE 1. Sensitivities for Metropolis–Hastings algorithm with respect to parameter $\theta = 0$ in the targeted normal distribution, based on 10^6 independent replications (mean \pm standard error).

In the experiment, we set $\theta = 0$, $d = 1$, the initial state $X_1 = 1$, and the number of transitions $n = 100$. From the histogram reported in Appendix B, the Markov chain is mixed well after 100 steps. From Table 6, the variances of FDC(0.01) are 400 and 180 times those of GLR, and the true sensitivity values lie out of the 90% confidence interval of FDC(0.1). The computational time of GLR is about 30 times that of FDC.

5.4. Sensitivities of Collateralized Debt Obligations We estimate the sensitivity with respect to the parameter θ that governs the dependence in the 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 $\tau_i = -\frac{1}{\lambda_i} \log(X_i)$, $i = 1, \dots, 10$. The parameters λ_i and loss L_i , $i = 1, \dots, n$, 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, so the sensitivity estimate by GLR is almost a free byproduct that can be obtained simultaneously during the simulation. Table 3 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 those in this example.

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

	$m = 10^4$	$m = 10^5$	$m = 10^6$
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 2. 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).

random variables in [53] and [51]. The technical conditions for justifying unbiasedness of GLR-U are relatively easy to satisfy in practice, 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.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 72022001, 92146003, and 71901003, by the National Science Foundation (NSF) under Grant CMMI-1434419, by the Air Force Office of Scientific Research under Grant FA95502010211, by Discovery Grant RGPIN-2018-05795 from NSERC-Canada, and by the Key Research and Development Program of Beijing Municipal Science and Technology Commission.

Appendices

Appendix A. Theoretical Supplements

Proof of Theorem 1 under conditions (A.1)-(A.4)

Proof. We have

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) \, du &= \int_{[\delta, 1-\delta]^n} \nabla_x \varphi_\epsilon(x)|_{x=g(u; \theta)} \partial_\theta g(u; \theta) \, du \\ &= \int_{[\delta, 1-\delta]^n} \nabla_u \varphi_\epsilon(g(u; \theta)) J_g^{-1}(u; \theta) \partial_\theta g(u; \theta) \, du, \end{aligned}$$

where

$$\nabla \varphi_\epsilon(x) := \left(\frac{\partial \varphi_\epsilon(x)}{\partial x_1}, \dots, \frac{\partial \varphi_\epsilon(x)}{\partial x_n} \right).$$

The interchange of the differentiation and integration can be justified by uniform integrability:

$$\int_{[\delta, 1-\delta]^n} \sup_{\theta \in \Theta} |\nabla_x \varphi_\epsilon(x)|_{x=g(u; \theta)} \partial_\theta g(u; \theta) \, du < \infty.$$

Since $g(u; \theta)$ is twice continuously differentiable in $(0, 1)^n \times \Theta$ by condition (A.3), the elements in $\partial_\theta g(u; \theta)$, $J_g(u; \theta)$, and $J_g^{-1}(u; \theta)$ are bounded in a compact space $[\delta, 1 - \delta]^n \times \Theta$, and so are $\nabla_x \varphi_\epsilon(x)|_{x=g(u; \theta)}$, $r_i(u; \theta)$ and $d(u; \theta)$. By the Gauss-Green Theorem,

$$\begin{aligned} &\int_{[\delta, 1-\delta]^n} \nabla_u \varphi_\epsilon(g(u; \theta)) J_g^{-1}(u; \theta) \partial_\theta g(u; \theta) \, du_1 \cdots du_n \\ &= \sum_{i=1}^n \int_{[\delta, 1-\delta]^{n-1}} \varphi_\epsilon(g(u; \theta)) (J_g^{-1}(u; \theta) \partial_\theta g(u; \theta))^T e_i \prod_{j \neq i} du_j \Big|_{u_i=\delta}^{1-\delta} \\ &\quad - \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) \operatorname{div} (J_g^{-1}(u; \theta) \partial_\theta g(u; \theta)) \, du_1 \cdots du_n, \end{aligned}$$

where for $n = 1$, the integration in the first term on the right-hand side of the equation is absent, and this term becomes

$$\varphi_\epsilon(g(1 - \delta; \theta)) r_1(1 - \delta; \theta) - \varphi_\epsilon(g(\delta; \theta)) r_1(\delta; \theta), \tag{14}$$

and for $h(u) = (h_1(u), \dots, h_n(u))$,

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

Then

$$\begin{aligned} \operatorname{div}(J_g^{-1}(u; \theta) \partial_\theta g(u; \theta)) &= \sum_{i=1}^n \frac{\partial}{\partial u_i} e_i^T J_g^{-1}(u; \theta) \partial_\theta g(u; \theta) \\ &= \sum_{i=1}^n e_i^T \partial_{u_i} J_g^{-1}(u; \theta) \partial_\theta g(u; \theta) + \operatorname{trace}(J_g^{-1}(u; \theta) \partial_\theta J_g(u; \theta)). \end{aligned}$$

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

$$\begin{aligned} 0 &= \partial_{u_i} (J_g^{-1}(u; \theta) J_g(u; \theta)) \\ &= (\partial_{u_i} J_g^{-1}(u; \theta)) J_g(u; \theta) + J_g^{-1}(u; \theta) \partial_{u_i} J_g(u; \theta), \end{aligned}$$

which leads to

$$\partial_{u_i} J_g^{-1}(u; \theta) = -J_g^{-1}(u; \theta) (\partial_{u_i} J_g(u; \theta)) J_g^{-1}(u; \theta).$$

Therefore, we have

$$d(u; \theta) = -\operatorname{div}(J_g^{-1}(u; \theta) \partial_\theta g(u; \theta)).$$

With the discussion above,

$$\begin{aligned} &\frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) \, du \\ &= \sum_{i=1}^n \int_{[\delta, 1-\delta]^{n-1}} \varphi_\epsilon(g(u; \theta)) r_i(u; \theta) \prod_{j \neq i} du_j \Big|_{u_i=\delta}^{1-\delta} + \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) d(u; \theta) \, du. \end{aligned} \tag{15}$$

With condition (A.1),

$$\begin{aligned} &\limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} (\varphi_\epsilon(g(u; \theta)) - \varphi(g(u; \theta))) d(u; \theta) \, du \right| \\ &\leq \sup_{\theta \in \Theta} \left(\int_{[\delta, 1-\delta]^n} |\varphi_\epsilon(g(u; \theta)) - \varphi(g(u; \theta))|^p \, du \right)^{1/p} \left(\int_{[\delta, 1-\delta]^n} |d(u; \theta)|^q \, du \right)^{1/q} \\ &\leq C \lim_{\epsilon \rightarrow 0} \left(\int_{[\delta, 1-\delta]^n} |\varphi_\epsilon(g(u; \theta)) - \varphi(g(u; \theta))|^p \, du \right)^{1/p} = 0, \end{aligned}$$

where the first inequality holds by applying Hölder's inequality for $1/p + 1/q = 1$, and C is a positive constant. Similarly, we can show the uniform convergence of the first term on the right-hand side of (15) for $n > 1$, and

$$\lim_{\epsilon \rightarrow 0} \left| \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) \, du - \int_{[\delta, 1-\delta]^n} \varphi(g(u; \theta)) \, du \right| = 0.$$

For $n = 1$, the first term of the right side of (15) becomes (14), and condition (A.1) implies that

$$\limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} |(\varphi_\epsilon(g(1-\delta; \theta)) - \varphi(g(1-\delta; \theta))) r_1(1-\delta; \theta) - (\varphi_\epsilon(g(\delta; \theta)) - \varphi(g(\delta; \theta))) r_1(\delta; \theta)| = 0.$$

From [59], $\frac{d}{d\theta} \lim_{\epsilon \rightarrow 0} h_\epsilon(\theta) = \lim_{\epsilon \rightarrow 0} h'_\epsilon(\theta)$ holds if $h'_\epsilon(\theta)$ converges uniformly with respect to $\theta \in \Theta$ as $\epsilon \rightarrow 0$. Therefore,

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi(g(u; \theta)) du &= \frac{\partial}{\partial \theta} \lim_{\epsilon \rightarrow 0} \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) du = \lim_{\epsilon \rightarrow 0} \frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) du \\ &= \lim_{\epsilon \rightarrow 0} \sum_{i=1}^n \int_{[\delta, 1-\delta]^{n-1}} \varphi_\epsilon(g(u; \theta)) r_i(u; \theta) \prod_{j \neq i} du_j \Big|_{u_i=\delta}^{1-\delta} + \lim_{\epsilon \rightarrow 0} \int_{[\delta, 1-\delta]^n} \varphi_\epsilon(g(u; \theta)) d(u; \theta) du \\ &= \sum_{i=1}^n \int_{[\delta, 1-\delta]^{n-1}} \varphi(g(u; \theta)) r_i(u; \theta) \prod_{j \neq i} du_j \Big|_{u_i=\delta}^{1-\delta} + \int_{[\delta, 1-\delta]^n} \varphi(g(u; \theta)) d(u; \theta) du, \end{aligned} \quad (16)$$

where the interchange of limit with differentiation and integration can be justified by uniform convergence for the first and second terms in (15) established above.

With condition (A.4),

$$\begin{aligned} &\limsup_{\delta \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} \varphi(g(u; \theta)) d(u; \theta) du - \int_{(0,1)^n} \varphi(g(u; \theta)) d(u; \theta) du \right| \\ &\leq \lim_{\delta \rightarrow 0} \int_{(0,1)^n \setminus [\delta, 1-\delta]^n} \sup_{\theta \in \Theta} |\varphi(g(u; \theta)) d(u; \theta)| du = 0, \end{aligned} \quad (17)$$

and uniform convergence can be established similarly for the first term in (16) under condition (A.2). Therefore,

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{(0,1)^n} \varphi(g(u; \theta)) du &= \lim_{\delta \rightarrow 0} \frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi(g(u; \theta)) du \\ &= \sum_{i=1}^n \int_{(0,1)^{n-1}} \left[\lim_{u_i \rightarrow 1^-} \varphi(g(u; \theta)) r_i(u; \theta) - \lim_{u_i \rightarrow 0^+} \varphi(g(u; \theta)) r_i(u; \theta) \right] \prod_{j \neq i} du_j \\ &\quad + \int_{(0,1)^n} \varphi(g(u; \theta)) d(u; \theta) du, \end{aligned}$$

which completes the proof. \square

Proof of Theorem 1 under conditions (A.3)-(A.5)

Proof. As in [53], define a sequence of bounded functions $\varphi_L(x) = \max\{\min\{\varphi(x), L\}, -L\}$, and then $|\varphi_L(x)| \leq \varphi(x)$ and $\lim_{L \rightarrow \infty} \varphi_L(x) = \varphi(x)$. From Theorem 1 in [53], there exists a sequence of bounded and smooth functions $\varphi_{\epsilon, L}(\cdot)$ such that

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

where $p > 1$, and $\|h\|_p := (\int_{\mathbb{R}^n} |h(x)|^p dx)^{1/p}$. Except for replacing $\varphi_\epsilon(\cdot)$ with $\varphi_{\epsilon, L}(\cdot)$, the procedures before (15) are the same as in the proof for Theorem 1 under conditions (A.1)-(A.4). Under condition (A.5),

$$\lim_{\epsilon \rightarrow 0} \limsup_{\delta \rightarrow 0} \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \int_{[\delta, 1-\delta]^{n-1}} \varphi_{\epsilon, L}(g(u; \theta)) r_i(u; \theta) \prod_{j \neq i} du_j \Big|_{u_i=\delta}^{1-\delta} \right| = 0.$$

With condition (A.3), $g(u; \theta)$ and $\det(J_g(u; \theta))$ are bounded in $[\delta, 1 - \delta]^n \times \Theta$. By change of variables and Hölder's inequality,

$$\begin{aligned} & \limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} (\varphi_{\epsilon, L}(g(u; \theta)) - \varphi_L(g(u; \theta))) d(u; \theta) du \right| \\ &= \limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{\mathcal{S}^\delta} (\varphi_{\epsilon, L}(x) - \varphi_L(x)) |\det(J_g(u; \theta))| d(u; \theta)|_{u=g^{-1}(x; \theta)} dx \right| \\ &\leq \lim_{\epsilon \rightarrow 0} \|\varphi_{\epsilon, L} - \varphi_L\|_p \sup_{\theta \in \Theta} \left| \int_{\mathcal{S}^\delta} |\det(J_g(u; \theta))| d(u; \theta)|_{u=g^{-1}(x; \theta)}^q dx \right|^{1/q} = 0, \end{aligned} \quad (18)$$

where

$$\mathcal{S}^\delta := \{x \in \mathbb{R}^n : x = g(u; \theta), u \in [\delta, 1 - \delta]^n\}.$$

With condition (A.4), we have

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} \varphi_{\epsilon, L}(g(u; \theta)) d(u; \theta) du - \int_{(0,1)^n} \varphi_L(g(u; \theta)) d(u; \theta) du \right| \\ &\leq \lim_{\delta \rightarrow 0} \limsup_{\epsilon \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} \varphi_{\epsilon, L}(g(u; \theta)) d(u; \theta) du - \int_{[\delta, 1-\delta]^n} \varphi_L(g(u; \theta)) d(u; \theta) du \right| \\ &+ \limsup_{\delta \rightarrow 0} \sup_{\theta \in \Theta} \left| \int_{[\delta, 1-\delta]^n} \varphi_L(g(u; \theta)) d(u; \theta) du - \int_{(0,1)^n} \varphi_L(g(u; \theta)) d(u; \theta) du \right| = 0, \end{aligned}$$

where the first term goes to zero because of (18) and the second term goes to zero by a similar argument as in (17). Therefore,

$$\begin{aligned} & \frac{\partial}{\partial \theta} \int_{(0,1)^n} \varphi_L(g(u; \theta)) du \\ &= \lim_{\delta \rightarrow 0} \lim_{\epsilon \rightarrow 0} \frac{\partial}{\partial \theta} \int_{[\delta, 1-\delta]^n} \varphi_L(g(u; \theta)) du = \int_{(0,1)^n} \varphi_L(g(u; \theta)) d(u; \theta) du. \end{aligned}$$

With condition (A.4) and noticing that $|\varphi_L(x) - \varphi(x)| \leq \varphi(x)$,

$$\lim_{L \rightarrow \infty} \sup_{\theta \in \Theta} \left| \int_{(0,1)^n} (\varphi_L(g(u; \theta)) - \varphi(g(u; \theta))) d(u; \theta) du \right| = 0.$$

By this uniform convergence and a similar argument as in the first part of the proof for (16) above,

$$\begin{aligned} & \frac{\partial}{\partial \theta} \int_{(0,1)^n} \varphi(g(u; \theta)) du = \frac{\partial}{\partial \theta} \int_{(0,1)^n} \lim_{L \rightarrow \infty} \varphi_L(g(u; \theta)) du \\ &= \lim_{L \rightarrow \infty} \frac{\partial}{\partial \theta} \int_{(0,1)^n} \varphi_L(g(u; \theta)) du = \lim_{L \rightarrow \infty} \int_{(0,1)^n} \varphi_L(g(u; \theta)) d(u; \theta) du = \int_{(0,1)^n} \varphi(g(u; \theta)) d(u; \theta) du, \end{aligned}$$

which proves the theorem. \square

Proof of Proposition 1

Proof. Define

$$\tilde{\chi}_\epsilon(x) := \begin{cases} 1 & x < -\epsilon, \\ 1 - \frac{(x + \epsilon)}{2\epsilon} & -\epsilon \leq x \leq \epsilon, \\ 0 & x > \epsilon, \end{cases}$$

and

$$\chi_\epsilon(x) := \frac{1}{\epsilon} \int_{\mathbb{R}} \tilde{\varphi}_\epsilon(x-y) \phi(y/\epsilon) dz,$$

where $\phi(\cdot)$ is the density of the standard normal distribution. By construction, $\chi_\epsilon(\cdot)$ is smooth. From the condition of the proposition, $g_i(u_i; \theta)$ is strictly monotone with respect to u_i on $[\delta, 1-\delta]$. Without loss of generality, we assume $g_i(u_i; \theta)$ is strictly increasing with respect to u_i on $[\delta, 1-\delta]$. For any $p > 1$, then we have

$$\begin{aligned} & \int_0^1 |\chi_\epsilon(g_i(u_i; \theta)) - \mathbf{1}\{g_i(u_i; \theta) \leq 0\}|^p du_i \\ & \leq \frac{1}{\epsilon} \int_0^1 \left(\int_{\mathbb{R}} \mathbf{1}\{\min(y-\epsilon, 0) \leq g_i(u_i; \theta) \leq \max(y+\epsilon, 0)\} \phi(y/\epsilon) dy \right)^p du_i \\ & \leq \frac{1}{\epsilon} \int_0^1 \int_{\mathbb{R}} (\mathbf{1}\{\min(y-\epsilon, 0) \leq g_i(u_i; \theta) \leq \max(y+\epsilon, 0)\})^p \phi(y/\epsilon) dy du_i \\ & \leq \frac{1}{\epsilon} \int_{\mathbb{R} \setminus [-\sqrt{\epsilon}, \sqrt{\epsilon}]} \phi(y/\epsilon) dy + \int_0^1 \mathbf{1}\{-\epsilon \leq g_i(u_i; \theta) \leq \epsilon + \sqrt{\epsilon}\} du_i \\ & = \frac{1}{\epsilon} \int_{\mathbb{R} \setminus [-\sqrt{\epsilon}, \sqrt{\epsilon}]} \phi(y/\epsilon) dy + g_i^{-1}(\epsilon + \sqrt{\epsilon}; \theta) - g_i^{-1}(-\epsilon; \theta), \end{aligned}$$

where the second inequality holds by applying Jensen's inequality. We have

$$g_i^{-1}(\epsilon + \sqrt{\epsilon}; \theta) - g_i^{-1}(-\epsilon; \theta) = (2\epsilon + \sqrt{\epsilon}) \left(\frac{\partial g_i(u_i; \theta)}{\partial u_i} \right)^{-1} \Big|_{u_i = g_i^{-1}(\xi_\epsilon; \theta)},$$

where $\xi_\epsilon \in (-\epsilon, \epsilon + \sqrt{\epsilon})$. Since $\inf_{\theta \in \Theta, u_i \in (0,1) \setminus [\delta, 1-\delta]} |g_i(u_i; \theta)| > 0$, $g_i^{-1}(\xi_\epsilon; \theta) \in [\delta, 1-\delta]$ for a sufficiently small ϵ . Therefore,

$$\begin{aligned} & \limsup_{\epsilon \rightarrow 0} \int_0^1 |\varphi_\epsilon(g(u; \theta) - z) - \mathbf{1}\{g(u; \theta) \leq z\}|^p du_i \\ & \leq \limsup_{\epsilon \rightarrow 0} (2\epsilon + \sqrt{\epsilon}) \left| \frac{\partial g_i(u_i; \theta)}{\partial u_i} \right|_{u_i = g_i^{-1}(\xi_\epsilon; \theta)}^{-1} \\ & \leq \lim_{\epsilon \rightarrow 0} (2\epsilon + \sqrt{\epsilon}) \left(\inf_{\theta \in \Theta, u_i \in [\delta, 1-\delta]} \left| \frac{\partial g_i(u_i; \theta)}{\partial u_i} \right| \right)^{-1} = 0. \end{aligned}$$

From $\inf_{\theta \in \Theta, u_i \in (0,1) \setminus [\delta, 1-\delta]} |g_i(u_i; \theta)| > 0$, for $u_i \in (0, 1) \setminus [\delta, 1-\delta]$,

$$\limsup_{\epsilon \rightarrow 0} |\chi_\epsilon(g_i(u_i; \theta)) - \mathbf{1}\{g_i(u_i; \theta) \leq 0\}| = \lim_{\epsilon \rightarrow 0} \mathbf{1}\{|g_i(u_i; \theta)| \leq \epsilon\} = 0.$$

In addition,

$$\begin{aligned} & \int_0^1 \cdots \int_0^1 \left| \prod_{j=1}^n \chi_\epsilon(g_j(u_j; \theta)) - \prod_{j=1}^n \mathbf{1}\{g_j(u_j; \theta) \leq 0\} \right|^p du_1 \cdots du_n \\ & = \int_{(0,1)^n} \left| \sum_{i=1}^n \left(\prod_{j=1}^i \chi_\epsilon(g_j(u_j; \theta)) \prod_{j=i+1}^n \mathbf{1}\{g_j(u_j; \theta) \leq 0\} - \prod_{j=1}^{i-1} \chi_\epsilon(g_j(u_j; \theta)) \prod_{j=i}^n \mathbf{1}\{g_j(u_j; \theta) \leq 0\} \right) \right|^p du \\ & \leq n^{p-1} \sum_{i=1}^n \int_{(0,1)^n} \left| \prod_{j=1}^i \chi_\epsilon(g_j(u_j; \theta)) \prod_{j=i+1}^n \mathbf{1}\{g_j(u_j; \theta) \leq 0\} - \prod_{j=1}^{i-1} \chi_\epsilon(g_j(u_j; \theta)) \prod_{j=i}^n \mathbf{1}\{g_j(u_j; \theta) \leq 0\} \right| du \\ & \leq n^{p-1} \sum_{i=1}^n \int_{(0,1)} |\chi_\epsilon(g_i(u_i; \theta)) - \mathbf{1}\{g_i(u_i; \theta) \leq 0\}| du_i, \end{aligned}$$

where the first inequality holds by applying Jensen's inequality. Then the rest of the proof is straightforward. \square

Problem of Transformation.

Under the simple framework of Section 3.1, we illustrate the problem of transformation discussed at the beginning of Section 2. If we transform $\varphi(g(U; \theta))$ into $\varphi(g(F(X); \theta))$, where F is the cdf of X , then the right hand side of (5) becomes

$$\int_{\partial\Omega} \varphi(g(F(x); \theta)) r(x; \theta) f(x) \hat{v} ds - \int_{\Omega} \varphi(g(F(x); \theta)) d(x; \theta) f(x) dx, \quad \text{where,}$$

$$r(x; \theta) = \left(\frac{\partial g(u; \theta)}{\partial u} \Big|_{u=F(x)} f(x) \right)^{-1} \frac{\partial g(F(x); \theta)}{\partial \theta} = \frac{1}{f(x)} \left(\frac{\partial g(u; \theta)}{\partial u} \Big|_{u=F(x)} \right)^{-1} \frac{\partial g(F(x); \theta)}{\partial \theta},$$

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

Therefore, the right hand side of (5) can be rewritten as

$$\int_{\partial\Omega} \varphi(g(F(x); \theta)) \left(\frac{\partial g(u; \theta)}{\partial u} \Big|_{u=F(x)} \right)^{-1} \frac{\partial g(F(x); \theta)}{\partial \theta} \hat{v} ds$$

$$- \int_{\Omega} \varphi(g(F(x); \theta)) \frac{\partial}{\partial x} \left\{ \left(\frac{\partial g(u; \theta)}{\partial u} \Big|_{u=F(x)} \right)^{-1} \frac{\partial g(x; \theta)}{\partial \theta} \right\} dx$$

The surface integral typically would not be zero because the density f whose tails might go to zero has been canceled by the extra $1/f(x)$ term due to the transformation. For example, let $\varphi(z) = z$, $g(u; \theta) = \theta u$, $u \in (0, 1)$, X follows the standard normal distribution. Then notice that

$$\int_{\partial\Omega} \varphi(g(F(x); \theta)) \left(\frac{\partial g(u; \theta)}{\partial u} \Big|_{u=F(x)} \right)^{-1} \frac{\partial g(F(x); \theta)}{\partial \theta} \hat{v} ds = F^2(x) \Big|_{x=-\infty}^{\infty} = 1.$$

Therefore, this approach would not transform the problem into the framework of [53], and the same difficulty addressed in this work exists.

GLR-U for the Ali-Mikhail-Haq Copula.

The generator function for this copula is

$$\psi_{\theta}(x) = \frac{1 - \theta}{e^x - \theta}, \quad \theta \in [0, 1).$$

Then, we have

$$\frac{\partial \psi_{\theta}(x)}{\partial \theta} = -\frac{1}{e^x - \theta} + \frac{1 - \theta}{(e^x - \theta)^2}, \quad \frac{\partial \psi'_{\theta}(x)}{\partial \theta} = \frac{e^x}{(e^x - \theta)^2} - \frac{2e^x(1 - \theta)}{(e^x - \theta)^3},$$

and

$$\psi'_{\theta}(x) = -\frac{e^x(1 - \theta)}{(e^x - \theta)^2}, \quad \psi''_{\theta}(x) = \frac{2e^{2x}(1 - \theta)}{(e^x - \theta)^3} - \frac{e^x(1 - \theta)}{(e^x - \theta)^2}.$$

Y_θ has a geometric distribution with parameter θ , with probability mass function

$$(1 - \theta)\theta^{y-1} \quad \text{for } y = 1, 2, \dots,$$

so the LR term is

$$\frac{\partial \log f_Y(y; \theta)}{\partial \theta} = -\frac{1}{1 - \theta} + \frac{y - 1}{\theta}.$$

The weight functions in the GLR-U estimator are

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

and

$$d_i(u_i; \theta, y) = \frac{1 - u_i^{\frac{1}{y}}}{1 - \theta} \left(y - \frac{u_i^{-\frac{1}{y}} + \theta}{u_i^{-\frac{1}{y}} - \theta} \right) - \frac{2}{u_i^{-\frac{1}{y}} - \theta} + \frac{1}{1 - \theta}.$$

We have

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

GLR-U for Distribution Sensitivities.

To establish the unbiasedness of $G_{1,i}(U; z, \theta)$ and $G_{2,i}(U; z, \theta)$ conditional on U_{-i} , condition (A.1) in Theorem 1 in the main body of the paper can be justified by checking the conditions in Proposition 1 in the main body of the paper. Given this conditional unbiasedness, the unconditional unbiasedness of $G_{1,i}(U; z, \theta)$ and $G_{2,i}(U; z, \theta)$ will follow from

$$\mathbb{E} \left[\sup_{\theta \in \Theta} |\mathbb{E}[G_{1,i}(U; z, \theta) | U_{-i}]| \right] < \infty \quad \text{and} \quad \mathbb{E} \left[\sup_{z \in \mathcal{Z}} |\mathbb{E}[G_{2,i}(U; z, \theta) | U_{-i}]| \right] < \infty,$$

where \mathcal{Z} is a neighborhood of z . Since the indicator function is bounded, the conditions of Proposition 1 in the main body of the paper follow from the integrability condition on the weight functions: for $i = 1, \dots, n$,

$$\mathbb{E} \left[\sup_{\theta \in \Theta} |r_i(\overline{U}_i; \theta)| \right] < \infty, \quad \mathbb{E} \left[\sup_{\theta \in \Theta} |r_i(\underline{U}_i; \theta)| \right] < \infty, \quad \mathbb{E} \left[\sup_{\theta \in \Theta} |d(U; \theta)| \right] < \infty, \quad \text{and}$$

$$\mathbb{E} [|\tilde{r}_i(\overline{U}_i; \theta)|] < \infty, \quad \mathbb{E} [|\tilde{r}_i(\underline{U}_i; \theta)|] < \infty, \quad \mathbb{E} [|\tilde{d}(U; \theta)|] < \infty.$$

The GLR-U estimator for estimating the distribution sensitivities is not unique. We can consider the above GLR-U estimator for each i and construct the following linear combination of these n GLR-U estimators with real-valued weights w_i , as in Hammersley and Handscomb [24, p.19]:

$$\sum_{i=1}^n w_i G_{r,i}(U; z, \theta) \quad \text{subject to} \quad \sum_{i=1}^n w_i = 1, \quad r = 1, 2.$$

An optimal GLR-U estimator, which minimizes the variance, can be obtained by solving

$$\arg \min_{(w_1, \dots, w_n)} \text{Var} \left(\sum_{i=1}^n w_i G_{r,i}(U; z, \theta) \right) \quad \text{subject to} \quad \sum_{i=1}^n w_i = 1.$$

This leads to the optimal weights

$$w_i^* = \frac{e_i^T \Sigma^{-1} e}{e^T \Sigma^{-1} e}, \quad i = 1, \dots, n, \quad (\text{GLR-Opt})$$

where $e = (1, \dots, 1)^T$, e_i is a d -dimensional unit vector in i th direction, and $\Sigma = (\Sigma_{i'i})_{n \times n}$ is the covariance matrix of $(G_{r,1}(U; z, \theta), \dots, G_{r,n}(U; z, \theta))$. In practice, w_i^* 's must be estimated, and such estimators will be correlated with the $G_{r,i}$. This linear combination idea is equivalent to a control variate formulation.

Appendix B. Supplements of Section 5

Distribution Sensitivities for a Linear Model

For the components in the GLR estimator, we have

$$\begin{aligned} \frac{\partial g(u; \theta)}{\partial u_1} &= -\frac{\theta}{\lambda_1 u_1}, & \frac{\partial g(u; \theta)}{\partial u_2} &= -\frac{1}{\lambda_2 u_2}, & \frac{\partial g(u; \theta)}{\partial \theta} &= -\frac{1}{\lambda_1} \log u_1, & \text{and} \\ \frac{\partial^2 g(u; \theta)}{\partial u_1^2} &= \frac{\theta}{\lambda_1 u_1^2}, & \frac{\partial^2 g(u; \theta)}{\partial u_2^2} &= \frac{1}{\lambda_2 u_2^2}, & \frac{\partial^2 g(u; \theta)}{\partial \theta \partial u_1} &= -\frac{1}{\lambda_1 u_1}, & \frac{\partial^2 g(u; \theta)}{\partial \theta \partial u_2} &= 0. \end{aligned}$$

Then the GLR estimators for distribution sensitivities in Section 5.1 are

$$\begin{aligned} G_{1,1}(U; z, \theta) &= -\frac{1}{\theta} \mathbf{1}\{g(U; \theta) \leq z\} (\log U_1 + 1), \\ G_{1,2}(U; z, \theta) &= \frac{\lambda_2}{\lambda_1} \left[\mathbf{1} \left\{ -\frac{\theta}{\lambda_1} \log U_1 \leq z \right\} \log U_1 - \mathbf{1}\{g(U; \theta) \leq z\} \log U_1 \right], \\ G_{2,1}(U; z, \theta) &= \frac{\lambda_1}{\theta} \left[\mathbf{1} \left\{ -\frac{1}{\lambda_2} \log U_2 \leq z \right\} - \mathbf{1}\{g(U; \theta) \leq z\} \right], \\ G_{2,2}(U; z, \theta) &= \lambda_2 \left[\mathbf{1} \left\{ -\frac{\theta}{\lambda_1} \log U_1 \leq z \right\} - \mathbf{1}\{g(U; \theta) \leq z\} \right]. \end{aligned}$$

For the numerical experiments in Section 5.1 of the main body of the paper, we 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. We perform 10^6 independent simulation runs with GLR. Figure 6(a) shows the curves of the distribution sensitivities with respect to θ , and Figure 6(b) presents the sample variance curves of three distribution sensitivity estimators: $G_{1,1}(\cdot)$ (GLR-1), $G_{1,2}(\cdot)$ (GLR-2), and a combined GLR estimator minimizing variance in a family of linear combinations of GLR-1 and GLR-2 given by (GLR-Opt).

Figure 7 presents the variance curves for two distribution sensitivities estimated by FDC using a batch of 10^6 independent simulation replications. $\text{FDC}(\delta)$ denotes the FDC with perturbation size δ . Comparing Figure 7 with Figure 1(b) in the main body of the paper and Figure 6(b), we can see the variance of FDC is much larger than those of the GLR estimators, and the variance of $\text{FDC}(0.01)$ is about 10 times of that of $\text{FDC}(0.1)$, which indicates that FDC suffers from a bias-variance tradeoff issue.

We now illustrate how the variance of GLR estimators can be reduced by applying CMC to these sensitivity estimators for the example in Section 5.1. For $G_{1,1}(U; z, \theta)$ and $G_{2,1}(U; z, \theta)$, we condition

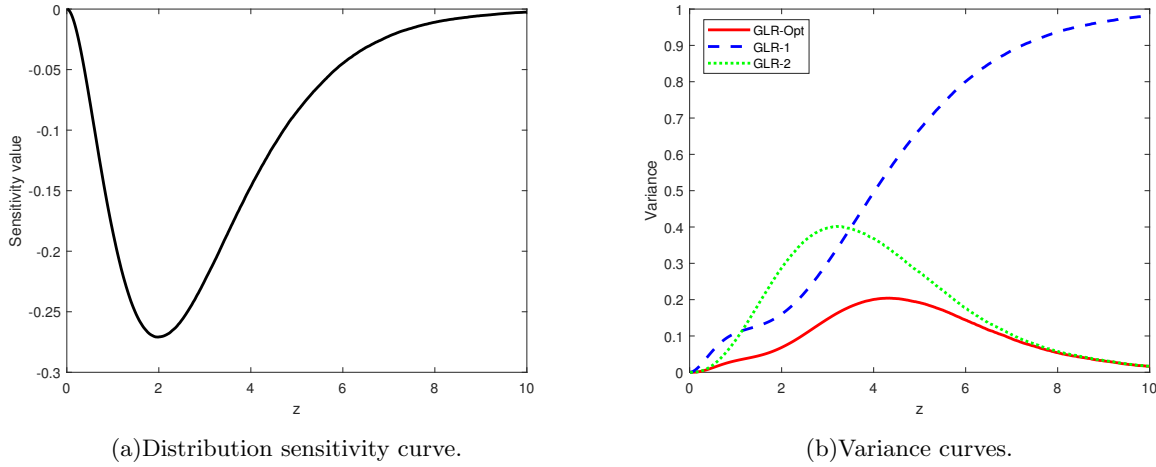


FIGURE 6. Distribution sensitivities with respect to θ estimated by GLR in the example of Section 5.1 in the main body of the paper.

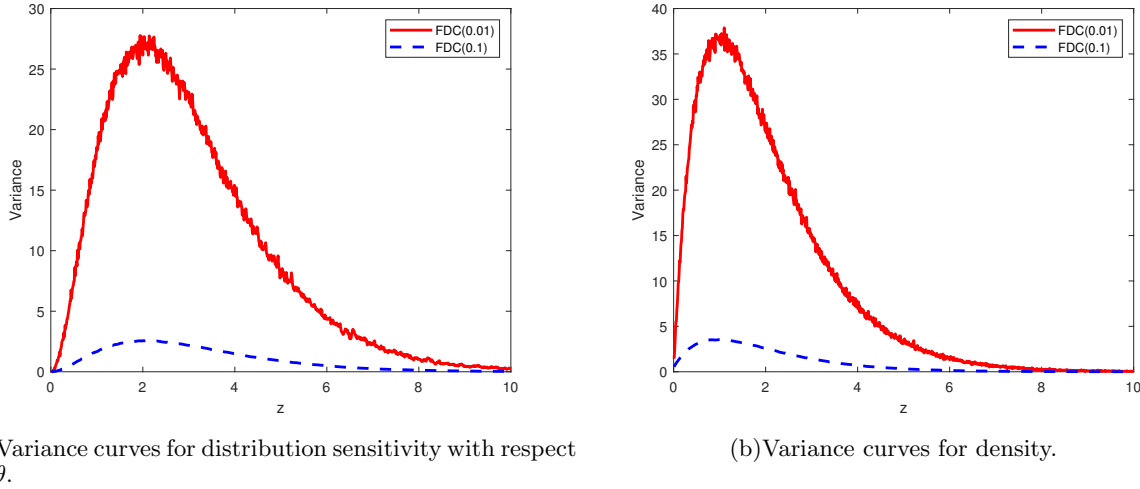


FIGURE 7. Variance curves for distribution sensitivities estimated by FDC in the example of Section 5.1 of the main body the paper.

on U_1 and integrate over U_2 for calculating the conditional expectations, and for $G_{1,2}(U; z, \theta)$ and $G_{2,2}(U; z, \theta)$, we condition on U_2 and integrate over U_1 for calculating the conditional expectations. The resulting conditional GLR estimators are given by

$$\hat{G}_{1,1}(U_1; z, \theta) = -\frac{1}{\theta} (\log U_1 + 1) \left[1 - \exp \left[-\lambda_2 \left(z + \frac{\theta \log U_1}{\lambda_1} \right) \mathbf{1} \left\{ z + \frac{\theta \log U_1}{\lambda_1} \geq 0 \right\} \right] \right],$$

$$\begin{aligned} \hat{G}_{1,2}(U_2; z, \theta) &= \lambda_2 \left(1 + \frac{\lambda_1 z}{\theta} \right) \exp \left(-\frac{\lambda_1 z}{\theta} \right) \\ &- \lambda_1 \left[1 + \frac{(\lambda_2 z + \log U_2)}{\theta} \mathbf{1} \left\{ z + \frac{\theta \log U_2}{\lambda_2} \geq 0 \right\} \right] \left[1 - \exp \left[-\lambda_1 \left(z + \frac{\theta \log U_2}{\lambda_2} \right) \mathbf{1} \left\{ z + \frac{\theta \log U_2}{\lambda_2} \geq 0 \right\} \right] \right], \end{aligned}$$

and

$$\hat{G}_{2,1}(U_1; z, \theta) = \frac{\lambda_1}{\theta} \left[\exp \left[-\lambda_2 \left(z + \frac{\theta \log U_1}{\lambda_1} \right) \mathbf{1} \left\{ z + \frac{\theta \log U_1}{\lambda_1} \geq 0 \right\} \right] - \exp(-\lambda_2 z) \right],$$

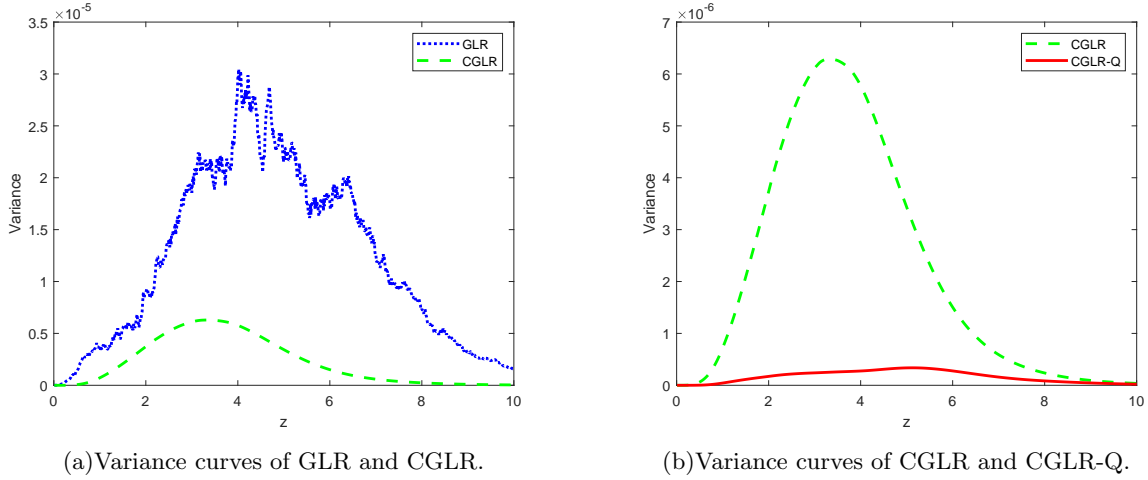


FIGURE 8. Variance comparison between GLR, CGLR, and CGLR-Q estimators with an average of 2^{13} samples for estimating distribution sensitivity with respect to θ in the example of Section 5.1 of the main body of the paper.

$$\widehat{G}_{2,2}(U_2; z, \theta) = \lambda_2 \left[\exp \left[-\frac{\lambda_1}{\theta} \left(z + \frac{\log U_2}{\lambda_2} \right) \mathbf{1} \left\{ z + \frac{\log U_2}{\lambda_2} \geq 0 \right\} \right] - \exp \left(-\frac{\lambda_1 z}{\theta} \right) \right].$$

The CGLR estimators become smoother than the original GLR estimators. Similarly, we can have an optimal CGLR estimators using (GLR-Opt).

To compare the variances of GLR, CGLR, and CGLR-Q for estimating the derivative $\mathbb{E}[g(U; \theta)]$ with respect to θ , we report the results of an experiment designed in the same way as in the main body of the paper. Similar observations can be found in the experimental results.

We can also directly apply RQMC to GLR estimators for estimating the two distribution sensitivities. GLR-Q is the GLR method with the Sobol sequence scrambled by an algorithm of [46] in Matlab. From Figure 9, we can see that although the variance of GLR-Q becomes smaller than that of GLR, the performance enhancement of GLR-Q relative to GLR is not as dramatic as that of CGLR-Q relative to CGLR. This observation corroborates that RQMC typically requires certain smoothness to achieve desirable variance reduction.

Distribution Sensitivities for a Stochastic Activity Network

An alternative GLR estimator different from that in the main body of the paper can be derived by conditioning on (U_1, U_5, U_6) and treating (U_4, U_2, U_3) as the input uniform random numbers for deriving GLR. The Jacobian matrix is

$$J_g(u; \theta, z) = - \begin{pmatrix} \frac{1}{\lambda_4 u_4} & 0 & 0 \\ 0 & \frac{1}{\lambda_2 u_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3 u_3} \end{pmatrix}.$$

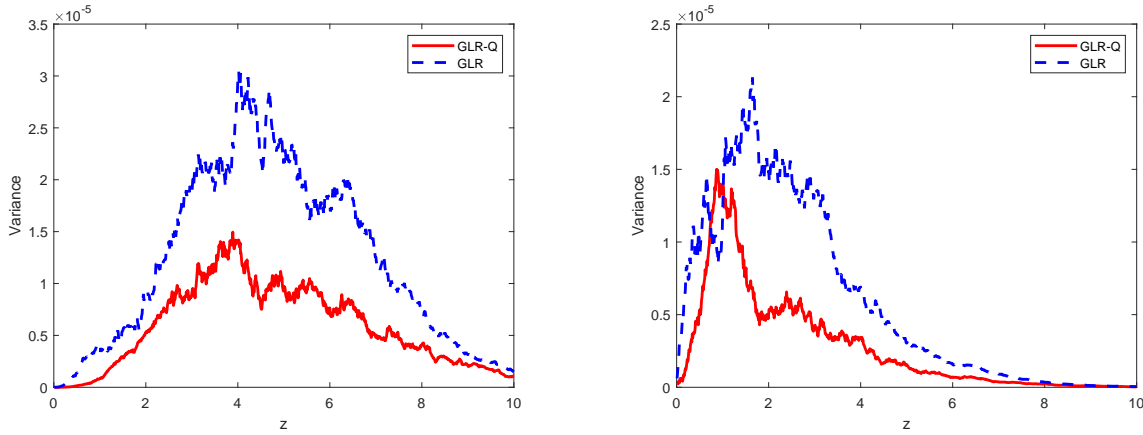
Then we have

$$r_1(u; \theta, z) = \lambda_4 u_4, \quad r_2(u; \theta, z) = \lambda_2 u_2, \quad r_3(u; \theta, z) = \lambda_3 u_3,$$

and

$$d(u; \theta, z) = -\lambda_2 - \lambda_3 - \lambda_4.$$

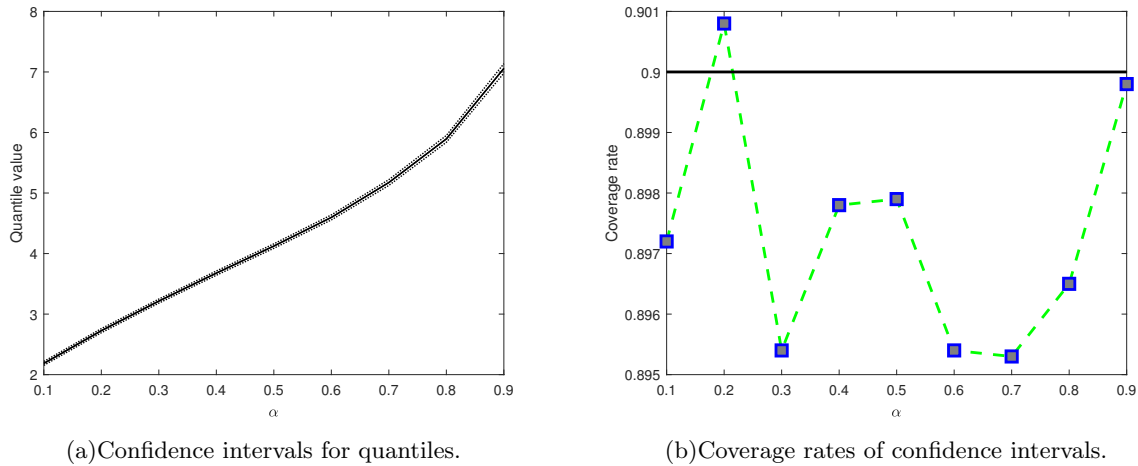
This alternative estimator leads to a comparable performance to the one in Section 4.2.



(a) Variance curves for distribution sensitivity with respect to θ .

(b) Variance curves for density.

FIGURE 9. Variance curves for distribution sensitivities estimated by GLR and GLR-Q estimators with an average of 2^{13} samples in the example of Section 5.1 of the main body of the paper.



(a) Confidence intervals for quantiles.

(b) Coverage rates of confidence intervals.

FIGURE 10. 90%-confidence intervals estimated by GLR for quantiles and coverage rates in example of Section 5.2 of the main body of the paper.

The setting of the problem is the same as that in Section 5.2 of the main body of the paper. Similar to Figure 2 in the example in Section 5.1 of the main body of the paper, Figure 10 presents confidence intervals using a batch of 10^4 independent simulation replications and the coverage rates of the 0.9-confidence intervals by 10^4 independent macro experiments for quantiles q_α , $\alpha = 0.1, \dots, 0.9$. Again the coverage ratios of the 0.9-confidence intervals match the target value statistically.

Figure 11 shows that the peak value of the variance curve of FDC(0.01) is over 100 times larger than that of GLR, and variances of FDC(0.01) are about 10 times of those of FDC(0.1) throughout the curve. The computational times of GLR is about 1.5 times that of FDC.

The GLR estimator for the example in Section 5.2 of the main body of the paper has the following analytical form:

$$G(U; \theta) = \lambda_1 \mathbf{1} \{ \max(X_4 + X_6, X_2 + X_5 + X_6, X_3 + X_5 + X_6) \leq z \} \\ + \lambda_2 \mathbf{1} \{ \max(X_1 + X_4 + X_6, X_5 + X_6, X_3 + X_5 + X_6) \leq z \}$$

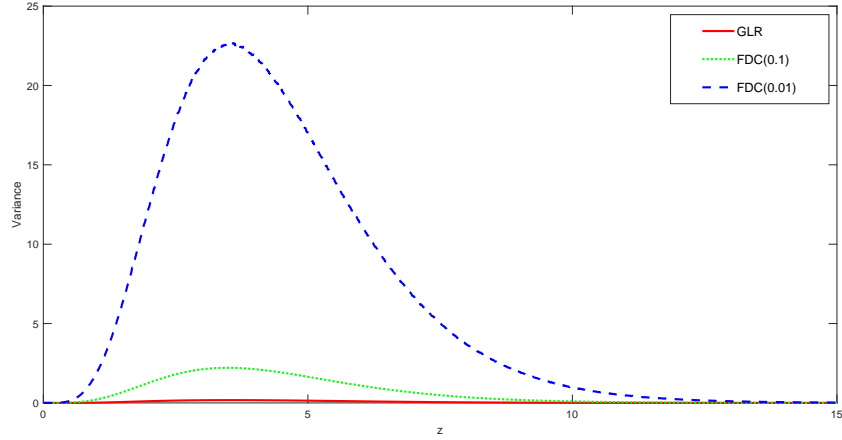


FIGURE 11. Variance curves for density estimates by FDC in the example of Section 5.2 of the main body of the paper.

$$-(\lambda_1 + \lambda_2) \mathbf{1} \{ \max(X_1 + X_4 + X_6, X_2 + X_5 + X_6, X_1 + X_3 + X_5 + X_6) \leq z \}.$$

By conditioning on $\hat{U} = (U_1, \dots, U_5)$ and integrating over U_6 , we have a conditional GLR estimator:

$$\begin{aligned} \hat{G}(\hat{U}; \theta) = & \lambda_1 [1 - \exp[-\lambda_6(z - \max(X_4, X_2 + X_5, X_3 + X_5))]] \\ & + \lambda_2 [1 - \exp[-\lambda_6(z - \max(X_1 + X_4, X_5, X_1 + X_3 + X_5))]] \\ & - (\lambda_1 + \lambda_2) [1 - \exp[-\lambda_6(z - \max(X_1 + X_4, X_2 + X_5, X_1 + X_3 + X_5))]]. \end{aligned}$$

For comparing the performance of GLR, CGLR, and CGLR-Q for the example in Section 5.2 of the main body of the paper, we set the number of samples in the estimators as 2^{13} and the number of samples for estimating variances as 100. From Figure 12, we can see that the variance of CGLR-Q dominates the variance of CGLR which dominates that of GLR over the entire range. The peak value in the variance curve of GLR is about 3 times that of the variance curve of CGLR, and the peak value in the variance curve of CGLR-Q is about 2/3 that of the variance curve of CGLR. In this example, the variance reduction of CGLR and CGLR-Q is not as dramatic as that in the example in Section 5.1 because the problem in this example has a larger number of input random numbers.

In the online appendix of [53], there is a discussion on how to transform a distribution not supported on the entire space to a distribution supported on the entire space by a change of variables. In this example, we show that this technique may not lead to a GLR estimator with desirable statistical properties. We consider a change of variables $Y_i = \log X_i$, $i = 1, \dots, 6$, and then the density of input random variable Y_i becomes $\lambda_i e^{-\lambda_i e^{y_i}} e^{y_i}$, $y_i \in (-\infty, \infty)$, which has a tail going down smoothly to zero. The sample performance can be written as

$$\mathbf{1} \{g_1(Y_1, Y_2, Y_3; Y_4, Y_5, Y_6, z) \leq 0\} \mathbf{1} \{g_2(Y_1, Y_2, Y_3; Y_4, Y_5, Y_6, z) \leq 0\} \mathbf{1} \{g_3(Y_1, Y_2, Y_3; Y_4, Y_5, Y_6, z) \leq 0\},$$

$$\begin{aligned} \text{where} \quad g_1(y_1, y_2, y_3; y_4, y_5, y_6, z) &= e^{y_1} + e^{y_4} + e^{y_6} - z, \\ g_2(y_1, y_2, y_3; y_4, y_5, y_6, z) &= e^{y_2} + e^{y_5} + e^{y_6} - z, \\ g_3(y_1, y_2, y_3; y_4, y_5, y_6, z) &= e^{y_1} + e^{y_3} + e^{y_5} + e^{y_6} - z. \end{aligned}$$

The Jacobian matrix of function $g = (g_1, g_2, g_3)$ with respect to arguments (y_1, y_2, y_3) and its matrix inversion are

$$J_g = \begin{pmatrix} e^{y_1} & 0 & 0 \\ 0 & e^{y_2} & 0 \\ e^{y_1} & 0 & e^{y_3} \end{pmatrix}, \quad J_g^{-1} = \begin{pmatrix} e^{-y_1} & 0 & 0 \\ 0 & e^{-y_2} & 0 \\ -e^{-y_3} & 0 & e^{-y_3} \end{pmatrix}.$$

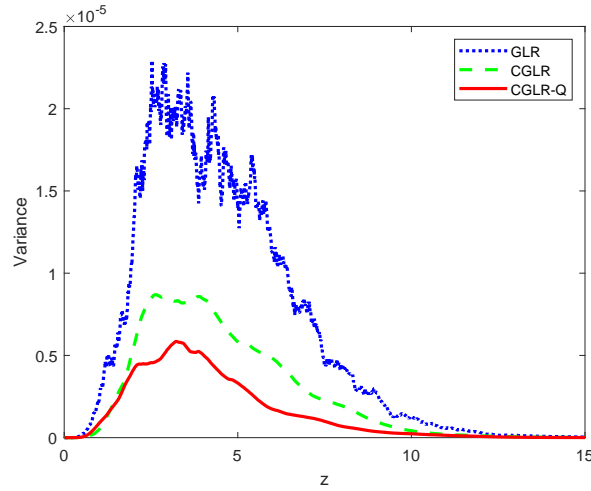


FIGURE 12. Variance comparison between GLR, CGLR, and CGLR-Q estimators with an average of 2^{13} samples in the example of Section 5.2 in the main body of the paper.

Furthermore, we have

$$\partial_{y_1} J_g = \begin{pmatrix} e^{y_1} & 0 & 0 \\ 0 & 0 & 0 \\ e^{y_1} & 0 & 0 \end{pmatrix}, \quad \partial_{y_2} J_g = \begin{pmatrix} 0 & 0 & 0 \\ 0 & e^{y_2} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \partial_{y_3} J_g = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & e^{y_3} \end{pmatrix},$$

$\partial_z g = -(1, 1, 1)^T$ and $\nabla \log f_Y(y) = (-\lambda_1 e^{y_1} + 1, -\lambda_2 e^{y_2} + 1, -\lambda_3 e^{y_3} + 1)^T$. For estimating the density, the weight function in the GLR estimator in [53] has the form

$$\sum_{i=1}^3 (J_g^{-1} \partial_{y_i} J_g J_g^{-1} e_i)^T \partial_z g - (\partial_z g)^T J_g^{-1} \nabla \log f_Y = -\sum_{i=1}^3 \lambda_i + \lambda_1 e^{Y_1 - Y_3} = -\sum_{i=1}^3 \lambda_i + \lambda_1 X_1 / X_3.$$

The term X_1/X_3 in the weight function given above could lead to a GLR estimator with infinite variance, which can be observed in Figure 13 presenting the variance of the GLR estimator as a function of z based on 10^6 independent experiments.

Sensitivities for Metropolis–Hastings Algorithm

Figure 14 presents the histogram of the state of Markov chain at 100th step in the Metropolis–Hastings algorithm based on 10000 independent experiments in the example of Section 5.3 of the main body of the paper. We can see that the distribution of the state of Markov chain is close to the equilibrium distribution, which is a standard normal distribution.

For the Metropolis–Hastings algorithm, the transition probability distribution of the constructed Markov chain is

$$w_\theta(x_{i-1}, x_i) p_\theta(x_{i-1} | x_i) dx_i + (1 - w_\theta(x_{i-1}, x_i)) \mathbf{1}\{x_i = x_{i-1}\}.$$

The LR estimator for $\partial \mathbb{E}[\varphi_n(X)] / \partial \theta$ is

$$\begin{aligned} \varphi_n(X) \sum_{i=1}^n \left[\left(\frac{\partial \log w_\theta(X_{i-1}, Y_i)}{\partial \theta} + \frac{\partial \log p_\theta(X_{i-1} | Y_i)}{\partial \theta} \right) \mathbf{1}\{U_i \leq w_\theta(X_{i-1}, Y_i)\} \right. \\ \left. - \frac{\partial \log w_\theta(X_{i-1}, Y_i)}{\partial \theta} \mathbf{1}\{U_i > w_\theta(X_{i-1}, Y_i)\} \right]. \end{aligned}$$

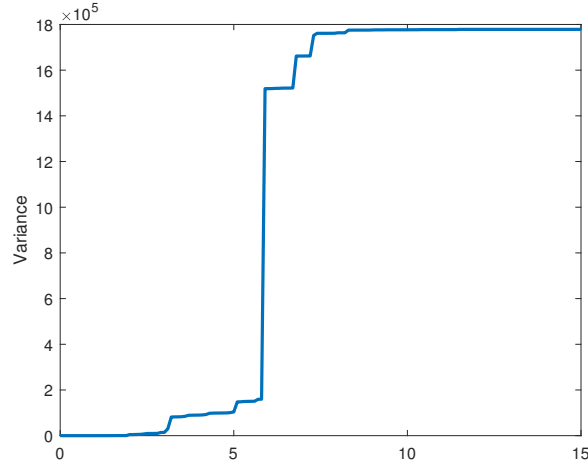


FIGURE 13. Variance curves for density estimates by GLR in [53] after a change of variables in the example of Section 5.2 of the main body of the paper based on 10^6 independent experiments.

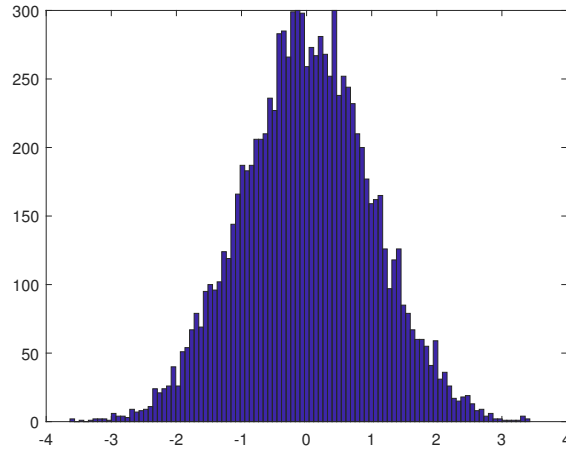


FIGURE 14. Histogram of the state of Markov Chain at 100th step in the Metropolis-Hastings Algorithm based on 10000 independent experiments in the example of Section 5.3 of the main body of the paper.

The setting of the experiment is set the same as that in Section 5.3. For estimating $\partial \mathbb{E}[X_n]/\partial \theta$, the standard deviation of the LR estimator is about 115, and for estimating $\partial \mathbb{E}[X_n^2]/\partial \theta$, the standard deviation of the LR estimator is about 162. The standard deviations of the LR estimators are about 100 and 40 times larger than those of the GLR-U estimators and are even much larger than those of FDC(0.01).

Sensitivities for Acceptance-Rejection Method

For the acceptance-rejection method, the target distribution is a beta distribution with density:

$$\pi_\theta(x) = \frac{[x(1-x)]^{\theta-1}}{\int_0^1 [x(1-x)]^{\theta-1} dx}, \quad x \in (0, 1),$$

and let $c_\theta = \sup_{x \in (0,1)} \pi_\theta(x) = (1/4)^{\theta-1} \int_0^1 [x(1-x)]^{\theta-1} dx$. The uniform distribution on $(0, 1)$ is chosen as the proposal distribution. Then we have $w_\theta(Y_i) = [4Y_i(1-Y_i)]^{\theta-1}$ and $\frac{\partial w_\theta(Y_i)}{\partial \theta} = \log[4Y_i(1-Y_i)] [4Y_i(1-Y_i)]^{\theta-1}$.

	GLR	FD(0.01)	FD(0.1)
$\partial\mathbb{E}[X]/\partial\theta = 0$	0.0001 ± 0.00007	0.001 ± 0.003	-0.0003 ± 0.0003
$\partial\mathbb{E}[X^2]/\partial\theta = -0.02$	-0.02 ± 0.00004	-0.016 ± 0.003	-0.0189 ± 0.0003

TABLE 3. Sensitivities for Acceptance-Rejection method with respect to parameter $\theta = 2$ in the targeted beta distribution, based on 10^8 independent replications (mean \pm standard error).

$Y_i)]w_\theta(Y_i)$. In the experiment, we set $\theta = 2$ and the randomization distribution in (13) of the main body of the paper for GLR is a geometric distribution with parameter 0.4. For the stopping time problem, FDC is not easy to implement, so we implement the finite difference (FD) method with independent random numbers. From Table 7, the variances of FD(0.01) are about 1800 and 5600 times those of GLR, and the true sensitivity value of $\partial\mathbb{E}[X^2]/\partial\theta$ lies out of the 90% confidence interval of FD(0.1). The computational time of GLR is about 5 times that of FD.

Sensitivities of Control Charts

Suppose we pay a cost of c per unit of time when the system is out of control and this is not yet detected, and a one-time cost C to fix the system when it is declared out of control. This model is a regenerative process, which regenerates each time we fix the system. The expected cost over one regenerative cycle is $c\mathbb{E}[(N - \lceil\chi/\Delta\rceil)^+] + C$, and therefore the average cost per unit of time over an infinite horizon is

$$\frac{c\mathbb{E}[(N - \lceil\chi/\Delta\rceil)^+] + C}{\mathbb{E}[N]}. \quad (19)$$

The goal might be to select the control limits θ_1 and θ_2 to minimize this average cost. Widening the gap $\theta_2 - \theta_1$ would reduce the frequency of intervention, so we would pay the fixed cost C less often, but then the penalty c would be paid over longer periods of time on average. The optimal control limits achieve an optimal balance between these two types of costs. The sample performance of the expectation in the numerator of (19) is different from those treated by SPA in [15] and [16], the development of which depends on the specific structure of the problem. The GLR method in our work provides unbiased derivative estimators for the expectations in both the numerator and denominator of (19).

We estimate sensitivities of the expectations in the numerator and denominator of (19) for control charts with respect to upper control limit $\theta = \theta_2$ discussed in Section 4.2. When the system is in control, the output sample is assumed to follow a uniform distribution on $[-1, 1]$ and we define $Z_i = 2U_i - 1$. When the system is out of control, the output sample is assumed to follow a uniform distribution on $[0, 2]$ and we define $Z_i = 2U_i$. Then the transition function of the Markov chain is

$$X_i = (1 - \alpha)X_{i-1} + \frac{\alpha}{\theta_2 - \theta_1} [\mathbf{1}\{i < \chi/\Delta\}(2U_i - 1) + \mathbf{1}\{i \geq \chi/\Delta\}2U_i - \theta_1],$$

and the weight functions in the GLR estimator are

$$r_i(u; \theta) = -\frac{1}{\theta_2 - \theta_1} \left[\mathbf{1}\{i < \chi/\Delta\} \left(u_i - \frac{1}{2} \right) + \mathbf{1}\{i \geq \chi/\Delta\} u_i - \frac{\theta_1}{2} \right],$$

$$d(u; \theta) = \frac{1}{\theta_2 - \theta_1}.$$

In the experiment, we set $\alpha = 1/2$, $\theta_1 = -1$, $\theta_2 = 1$, $\Delta = 1$, $X_0 = 0$, and assume $\chi = 1 + 3\log U$. The randomized horizon N' in (13) of the main body of the paper follows a geometric distribution with parameter 0.1. In the random horizon problem, it is not easy to synchronize the two sample paths for the FD method, because perturbing parameter θ affects the stopping time N , so it would require substantially more computational overhead to generate the sample paths using common

	GLR	FD(0.01)	FD(0.1)
$\partial \mathbb{E}[N]/\partial \theta$	8.6 ± 0.2	8.6 ± 0.4	10.5 ± 0.04
$\partial \mathbb{E}[(N - \lceil \chi/\Delta \rceil)^+]/\partial \theta$	8.8 ± 0.07	8.6 ± 0.4	10.5 ± 0.04

TABLE 4. Derivatives of control charts with respect to upper control limit $\theta = 1$, based on 10^6 independent replications (mean \pm standard error).

	$m = 10^4$	$m = 10^5$	$m = 10^6$
GLR	-0.188 ± 0.007	-0.189 ± 0.004	-0.191 ± 0.002
FDC(0.1)	-0.194 ± 0.006	-0.195 ± 0.02	-0.195 ± 0.002
FDC(0.01)	-0.110 ± 0.05	-0.202 ± 0.01	-0.192 ± 0.005

TABLE 5. Sensitivity estimates of CDO with 10 assets governed by the Ali-Mikhail-Haq copula with $\theta = 0.5$ based on 10^2 experiments (mean \pm standard error).

	$n = 10^4$	$n = 10^5$	$n = 10^6$
GLR	0.034 ± 0.003	0.035 ± 0.001	0.035 ± 0.0007
FDC(0.1)	0.033 ± 0.004	0.031 ± 0.001	0.031 ± 0.0007
FDC(0.01)	-0.002 ± 0.03	0.021 ± 0.009	0.031 ± 0.004

TABLE 6. Sensitivity estimates of BDS with 10 assets governed by the Clayton copula with $\theta = 0.5$ based on 10^2 experiments (mean \pm standard error).

random numbers to implement FDC. In general, decreasing the perturbation size δ would reduce the bias of $\text{FD}(\delta)$ but increase the variance. In Table 4, we can see that the sensitivity results estimated by GLR match those estimated by $\text{FD}(0.01)$, while the variances of $\text{FD}(0.01)$ are 4 – 30 times those of GLR, whereas the sensitivity results estimated by $\text{FD}(0.1)$ are significantly biased. The computational time of GLR is about 5 times that of FD.

Sensitivities of Credit Risk Derivatives

For CDOs, Table 5 shows the sensitivity estimates for sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$ under the Ali-Mikhail-Haq copula, with $\theta = 0.5$. The numerical observations are similar to those in Section 5.4 of the main body of the paper. Again for sample size $m = 10^6$, the estimate with FDC(0.1) falls 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 sample size $m = 10^7$, which is -0.193 ± 0.002 .

For BDS, we estimate the sensitivities of the expectation of the discounted payment to the fifth default of 10 assets made by the seller $V_{\text{value}}(\tau) = L_{(5)} \exp(-r\tau_{(5)}) \mathbf{1}\{\tau_{(5)} < T\}$. The marginal distributions of the defaults and the parameters are set in the same way as those in Section 5.1 in the main body of the paper. Tables 6 and 7 show the respective Clayton copula and Ali-Mikhail-Haq copula sensitivity estimates with $\theta = 0.5$ for sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$. We also implement FDC(0.01) with sample size $m = 10^7$, which leads to 0.034 ± 0.001 for the Clayton copula and 0.051 ± 0.002 for the Ali-Mikhail-Haq copula. The results are similar to those for estimating CDO sensitivities.

References

	$n = 10^4$	$n = 10^5$	$n = 10^6$
GLR	0.051 ± 0.002	0.050 ± 0.001	0.051 ± 0.0009
FDC(0.1)	0.048 ± 0.003	0.052 ± 0.001	0.051 ± 0.001
FDC(0.01)	0.074 ± 0.03	0.039 ± 0.01	0.049 ± 0.003

TABLE 7. Sensitivity estimates of BDS with 10 assets governed by the Ali-Mikhail-Haq copula with $\theta = 0.5$ based on 10^2 experiments (mean \pm standard error).

- [1] Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis*, volume 57 (Springer).
- [2] Bashyam S, Fu MC (1998) Optimization of (s, S) inventory systems with random lead times and a service level constraint. *Management Science* 44(12-part-2):S243–S256.
- [3] Broadie M, Glasserman P (1996) Estimating security price derivatives using simulation. *Management Science* 42(2):269–285.
- [4] Chen N, Liu Y (2014) American option sensitivities estimation via a generalized infinitesimal perturbation analysis approach. *Operations Research* 62(3):616–632.
- [5] Chen Z, Glasserman P (2008) Sensitivity estimates for portfolio credit derivatives using Monte Carlo. *Finance and Stochastics* 12(4):507–540.
- [6] Cui Z, Fu MC, Hu JQ, Liu Y, Peng Y, Zhu L (2020) On the variance of single-run unbiased stochastic derivative estimators. *INFORMS Journal on Computing* 32(2):390–407.
- [7] Devroye L (1986) *Non-Uniform Random Variate Generation* (Springer).
- [8] Embrechts P, Lindskog F, McNeil A (2003) Modelling dependence with copulas and applications to risk management. Rachev S, ed., *Handbook of Heavy Tailed Distributions in Finance*, 329–384 (Elsevier), chapter 8.
- [9] Fu MC (1994) Sample path derivatives for (s, S) inventory systems. *Operations Research* 42(2):351–364.
- [10] Fu MC (2015) Stochastic gradient estimation, Fu, Michael C. ed. *Handbook of Simulation Optimization*, volume 216, 105–147 (Springer), chapter 5.
- [11] Fu MC, Hong LJ, Hu JQ (2009) Conditional Monte Carlo estimation of quantile sensitivities. *Management Science* 55(12):2019–2027.
- [12] Fu MC, Hu JQ (1993) Second derivative sample path estimators for the GI/G/m queue. *Management Science* 39(3):359–383.
- [13] Fu MC, Hu JQ (1995) Sensitivity analysis for Monte Carlo simulation of option pricing. *Probability in the Engineering and Informational Sciences* 9(3):417–446.
- [14] Fu MC, Hu JQ (1997) *Conditional Monte Carlo: Gradient Estimation and Optimization Applications* (Kluwer Academic Publishers, Boston).
- [15] Fu MC, Hu JQ (1999) Efficient design and sensitivity analysis of control charts using Monte Carlo simulation. *Management Science* 45(3):395–413.
- [16] Fu MC, Lele S, Vossen TWM (2009) Conditional Monte Carlo gradient estimation in economic design of control limits. *Production and Operations Management* 18(1):60–77.
- [17] Glasserman P (1991) *Gradient Estimation via Perturbation Analysis* (Kluwer Academic Publishers, Boston).
- [18] Glasserman P, Liu Z (2010) Sensitivity estimates from characteristic functions. *Operations Research* 58(6):1611–1623.
- [19] Glynn PW (1987) Likelihood ratio gradient estimation: an overview. *Proceedings of the 1987 Winter Simulation Conference*, 366–375 (ACM).
- [20] Glynn PW (1990) Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM* 33(10):75–84.

- [21] Glynn PW, L'Ecuyer P (1995) Likelihood ratio gradient estimation for stochastic recursions. *Advances in applied probability* 27(4):1019–1053.
- [22] Glynn PW, Peng Y, Fu MC, Hu J (2021) Computing sensitivities for distortion risk measures. *INFORMS Journal on Computing* 33(4):1520–1532.
- [23] Gong WB, Ho YC (1987) Smoothed perturbation analysis of discrete event dynamical systems. *IEEE Transactions on Automatic Control* 32(10):858–866.
- [24] Hammersley JM, Handscomb DC (1964) *Monte Carlo Methods* (London: Methuen).
- [25] Heidergott B (1999) Optimisation of a single-component maintenance system: a smoothed perturbation analysis approach. *European Journal of Operational Research* 119(1):181–190.
- [26] Heidergott B, Farenhorst-Yuan T (2010) Gradient estimation for multicomponent maintenance systems with age-replacement policy. *Operations Research* 58(3):706–718.
- [27] Heidergott B, Leahu H (2010) Weak differentiability of product measures. *Mathematics of Operations Research* 35(1):27–51.
- [28] Heidergott B, Vazquez-Abad FJ (2009) Gradient estimation for a class of systems with bulk services (a problem in public transportation). *ACM Transactions on Modeling and Computer Simulation* 19(3):13.
- [29] Heidergott B, Volk-Makarewicz W (2016) A measure-valued differentiation approach to sensitivity analysis of quantiles. *Mathematics of Operations Research* 41(1):293–317.
- [30] Ho YC, Cao XR (1991) *Discrete Event Dynamic Systems and Perturbation Analysis* (Kluwer Academic Publishers, Boston, MA).
- [31] Hong LJ (2009) Estimating quantile sensitivities. *Operations Research* 57(1):118–130.
- [32] Hong LJ, Juneja S, Luo J (2014) Estimating sensitivities of portfolio credit risk using Monte Carlo. *INFORMS Journal on Computing* 26(4):848–865.
- [33] Hong LJ, Liu G (2009) Simulating sensitivities of conditional value at risk. *Management Science* 55(2):281–293.
- [34] Jacobi L, Joshi MS, Zhu D (2018) Automated sensitivity analysis for Bayesian inference via Markov chain Monte Carlo: Applications to Gibbs sampling. *Available at SSRN 2984054* .
- [35] Jiang G, Fu MC (2015) Technical note — On estimating quantile sensitivities via infinitesimal perturbation analysis. *Operations Research* 63(2):435–441.
- [36] Lang S (2013) *Undergraduate Analysis* (Springer Science & Business Media).
- [37] L'Ecuyer P (1990) A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science* 36(11):1364–1383.
- [38] L'Ecuyer P, Glynn PW (1994) Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state. *Management Science* 40(11):1562–1578.
- [39] L'Ecuyer P, Puchhammer F, Ben Abdellah A (2022) Monte Carlo and quasi-Monte Carlo density estimation via conditioning. *INFORMS Journal on Computing, forthcoming* .
- [40] Lei L, Peng Y, Fu MC, Hu JQ (2018) Applications of generalized likelihood ratio method to distribution sensitivities and steady-state simulation. *Journal of Discrete Event Dynamic Systems* 28(1):109–125.
- [41] Lei L, Peng Y, Fu MC, Hu JQ (2020) Sensitivity analysis of portfolio credit derivatives by conditional Monte Carlo simulation. *preprint in ResearchGate with DOI: 10.13140/RG.2.2.21951.97444* .
- [42] Li D (2000) On default correlation: a copula approach. *Journal of Fixed Income* 9:43–54.
- [43] Liu G, Hong LJ (2011) Kernel estimation of the Greeks for options with discontinuous payoffs. *Operations Research* 59(1):96–108.
- [44] 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 (Springer).
- [45] Marshall AW, Olkin I (1988) Families of multivariate distributions. *Journal of the American Statistical Association* 83(403):834–841.

- [46] Matoušek J (1998) On the L_2 -discrepancy for anchored boxes. *Journal of Complexity* 14(4):527–556.
- [47] Mohamed S, Rosca M, Figurnov M, Mnih A (2020) Monte Carlo gradient estimation in machine learning. *Journal of Machine Learning Research* 21(132):1–62.
- [48] Nakayama MK (2014) Confidence intervals for quantiles using sectioning when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation* 24(4):1–21.
- [49] Nelsen RB (2006) *An Introduction to Copulas*, volume 139 (Springer Science & Business Media), second edition.
- [50] Peng Y, Fu MC, Glynn PW, Hu JQ (2017) On the asymptotic analysis of quantile sensitivity estimation by Monte Carlo simulation. *Proceedings of Winter Simulation Conference* 2336–2347.
- [51] Peng Y, Fu MC, Heidergott B, Lam H (2020) Maximum likelihood estimation by Monte Carlo simulation: Towards data-driven stochastic modeling. *Operations Research* 68(6):1896–1912.
- [52] Peng Y, Fu MC, Hu JQ (2016) On the regularity conditions and applications for generalized likelihood ratio method. *Proceedings of Winter Simulation Conference*, 919–930 (IEEE Press).
- [53] Peng Y, Fu MC, Hu JQ, Heidergott B (2018) A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. *Operations Research* 66(2):487–499.
- [54] Peng Y, Xiao L, Heidergott B, Hong J, Lam H (2022) A new likelihood ratio method for training artificial neural networks. *INFORMS Journal on Computing*, forthcoming .
- [55] Pflug GC (1988) Derivatives of probability measures-concepts and applications to the optimization of stochastic systems. Varaiya P, Kurzhanski AB, eds., *Discrete Event Systems: Models and Applications*, 252–274 (Berlin, Heidelberg: Springer Berlin Heidelberg), ISBN 978-3-540-48045-7.
- [56] Rhee Ch, Glynn PW (2015) Unbiased estimation with square root convergence for SDE models. *Operations Research* 63(5):1026–1043.
- [57] Robert CP, Casella G (2004) *Monte Carlo Statistical Methods*, volume 319 (Springer).
- [58] Rubinstein RY, Shapiro A (1993) *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method* (Wiley, New York).
- [59] Rudin W (1964) *Principles of Mathematical Analysis* (McGraw-Hill Education, New York).
- [60] Serfling RJ (1980) *Approximation Theorems of Mathematical Statistics* (Wiley).
- [61] Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8:229–231.
- [62] Suri R, Zazanis MA (1988) Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Management Science* 34(1):39–64.
- [63] Wang Y, Fu MC, Marcus SI (2012) A new stochastic derivative estimator for discontinuous payoff functions with application to financial derivatives. *Operations Research* 60(2):447–460.