Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

LIKELIHOOD RATIO DENSITY ESTIMATION FOR SIMULATION MODELS

Florian Puchhammer

Department of Statistics and Actuarial Science
University of Waterloo
200 University Ave W
Waterloo, ON N2L 3G1, CANADA

Pierre L'Ecuyer

Département d'Informatique et
de Recherche Opérationnelle
Université de Montréal
C.P.6128, Succ. Centre-Ville
Montréal, QC H3C 3J7, CANADA

ABSTRACT

We consider the problem of estimating the density of a random variable X which is the output of a simulation model. We show how an unbiased density estimator can be constructed via the classical likelihood ratio derivative estimation method proposed over 35 years ago by Glynn, Rubinstein, and others. We then extend this density estimation method to cover situations where it does not apply directly. What we obtain is closely related to the generalized likelihood ratio method proposed recently by Peng and his co-authors, although the assumptions differ. We compare the methods and assumptions on some examples.

1 INTRODUCTION

We are interested in estimating the unknown density f of a continuous random variable X when realizations of X can be generated by simulation. Specifically, we assume that $X = h(\mathbf{Y})$ where the random vector $\mathbf{Y}$ has a known distribution from which we can sample exactly, and we know how to compute $h(\mathbf{y})$ for any given realization $\mathbf{y}$. There is a well-developed set of simulation methods to estimate the expectation $\mu = \mathbb{E}[X]$, reduce the variance of the estimators of μ, and compute a confidence interval on μ (Asmussen and Glynn 2007; Law 2014). But simulation experiments typically provide information for much more than estimating the mean. In particular, this information can be exploited to estimate the entire distribution of X instead of just its mean. This could be done simply by providing an estimate of the cumulative distribution function (cdf), but a density estimate actually provides a better visual insight on the distribution. A simple and widely used reasonable density estimator is a histogram. An improvement is a kernel density estimator (KDE) (Parzen 1962; Scott 2015), also popular but less obvious to construct and tune.

These density estimators are biased. To reduce the bias, one must reduce the width of the rectangles for the histogram or the bandwidth for the KDE, but in both cases this increases the variance. Therefore, a compromise must be made to minimize the mean square error (MSE), and as a result the MSE of the density estimator does not converge at the canonical rate of $\mathcal{O}(n^{-1})$ with n independent simulation runs. The best rate with the optimal bandwidth is $\mathcal{O}(n^{-2/3})$ for the histogram and $\mathcal{O}(n^{-4/5})$ for the KDE (Scott 2015). When the observations of X are given and assumed independent, the KDE is pretty much the best available method. But when the observations are generated by (Monte Carlo) simulation, there are opportunities to do better by manipulating the way the observations are obtained and collecting additional information.

Since the density is the derivative of the cdf: $f(x) = F'(x)$ where $F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y$, one may think of taking the derivative of an unbiased cdf estimator as a density estimator. Doing this with the empirical cdf does not work, because its derivative is zero almost everywhere. To make it work, one needs to construct a cdf estimator which is *continuous* in x and differentiable almost everywhere.

One approach that follows this path and was examined recently (Asmussen 2018; L'Ecuyer et al. 2022) uses *conditional Monte Carlo* (CMC) to construct a *conditional density estimator* (CDE). It consists in

978-1-6654-7661-4/22/$31.00 ©2022 IEEE

hiding part of the information in $\mathbf{Y}$ and computing the cdf and density of X conditional on the information that remain. This requires hiding enough information for the cdf of the resulting conditional distribution to be continuous (i.e., no probability mass) while keeping sufficient information for the conditional density to be sufficiently easy to compute. When this can be achieved, under mild additional conditions, the derivative of the sample cdf can be taken as an unbiased density estimator. This CDE can be very effective and it gives an $\mathscr{O}(n^{-1})$ convergence rate for the MSE. However, it is not always easy to construct a valid and effective CDE in this way.

The method developed in this paper provides an alternative. As for the CDE, the *likelihood ratio density estimator* (LRDE) is based on the idea of constructing a continuous cdf estimator and taking the derivative. But instead of replacing the cdf by a conditional expectation, we make a change of variable to obtain an expression that defines the cdf $F(x)$ as an integral whose bounds (ideally) do not depend on x, in which case we can differentiate this expression by taking the derivative inside the integral. The change of variable typically introduces a likelihood ratio in the integrand, hence the name LRDE. In this sense, it is based on the likelihood ratio (LR) derivative estimation method described and studied by Glynn (1987), L'Ecuyer (1990), L'Ecuyer (1995), Glynn and L'Ecuyer (1995), and Asmussen and Glynn (2007), among others. Our method generalizes the approach of Laub et al. (2019), designed for the special case where X is a sum of dependent random variables with known joint density, and it was originally inspired by that paper. In case the integral bounds still depend on x, things get a little more complicated and the differentiation yields additional terms that account for this change in the region of integration as a function of x, but we can still obtain an unbiased density estimator, under appropriate conditions.

A closely related density estimation approach was proposed by Lei et al. (2018) and Peng et al. (2020), based on a *generalized likelihood ratio* (GLR) method of Peng et al. (2018). The GLR is an extension of the classical LR derivative estimation method designed to deal with discontinuities that are not easily handled by ordinary LR. Lei et al. (2018) sketched out how to construct a density estimator via GLR, although their general formulas are not so easy to interpret and implement in applications. More convenient GLR density estimator formulas are given in Theorem 1 of Peng et al. (2020). Peng et al. (2022a) recently proposed GLR-U, a variant of GLR in which the stochastic model is defined directly in terms of independent uniform random variables over $(0,1)$, and which can handle a broad class of situations that were not covered by the original GLR setting of Peng et al. (2018). GRL-U can be adapted to estimate densities.

Our development in this paper starts with the ordinary LR derivative estimator instead, and is simpler. Our LRDE uses a change of variable followed by a change of density to construct a smooth cdf estimator whose derivative provides an unbiased density estimator, the LRDE. We give explicit conditions under which this LRDE is unbiased. Then we explore the situation in which the integration domain after the change of variable depends on x, and we show how to extend our plain vanilla LRDE to this setting. The resulting estimators are equivalent in many cases to those obtained via GLR or GLR-U, although our unbiasedness conditions differ from those given for those GLR methods. We illustrate this with examples in which our conditions hold whereas those given in the GLR papers do not hold. Our results therefore expand the range of applicability of LR and GLR for density estimation. L'Ecuyer and Puchhammer (2022) already introduced the plain vanilla LRDE, but without its extension to the more complicated setting examined here and without the detailed comparison with the GLR in terms of the given conditions for unbiasedness.

The remainder is organized as follows. In Section 2, we provide background on density estimation, define the general setting, develop our vanilla form of the LRDE, and provide some examples. We further extend our LRDE to the case where the integration domain after the change of variable depends on x, and we provide sufficient conditions for unbiasedness. In Section 3, we review density estimators that have been proposed based on the GLR approach, with the corresponding assumptions. For each case, we highlight certain assumptions that often do not hold, and we give specific illustrations in which this happens while all our assumptions from Section 3 hold. In Section 4, we illustrate and compare different LRDE variants on a stochastic activity network example. We show that some variants work well while others do not (the corresponding assumptions do hot hold), and this depends on the choice of change of variable. We also report the results of numerical experiments with this example. A conclusion is given in Section 5.

2 LIKELIHOOD RATIO DENSITY ESTIMATION

2.1 Setting

We assume a model of the form $X = h(\mathbf{Y})$, where $\mathbf{Y}$ is a random vector having a known density $f_{\mathbf{Y}}$ with respect to the Lebesgue measure, continuous over a measurable set (typically a rectangular box) $R \subseteq \mathbb{R}^d$, $h : R \to \mathbb{R}$ is a measurable function, and the random variable X has a continuous cdf F (it has a density f with respect to the Lebesgue measure, with no mass point). We further assume that we can sample realizations of $\mathbf{Y}$ from its exact distribution by Monte Carlo and that we know how to compute $X = h(\mathbf{Y})$ for any such realization. Thus, we can easily generate independent realizations of X. Our aim is to estimate the unknown density f of X over a finite interval $[a,b]$ from such realizations of X and any additional information that can be collected from the simulation.

We denote by $\hat{f}_n$ an estimator of f from a sample of size n, and we measure the quality of $\hat{f}_n$ over $[a,b]$ by the *mean integrated square error* (MISE), defined as

$$\text{MISE} = \text{MISE}(\hat{f}_n) = \int_a^b \mathbb{E}[\hat{f}_n(x) - f(x)]^2 \mathrm{d}x. \tag{1}$$

The MISE is the sum of the *integrated variance* (IV) and the *integrated square bias* (ISB):

$$\text{MISE} = \text{IV} + \text{ISB} = \int_a^b \mathbb{E}(\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)])^2 \mathrm{d}x + \int_a^b (\mathbb{E}[\hat{f}_n(x)] - f(x))^2 \mathrm{d}x.$$

For the situation where n independent observations of X are given, the two most popular density estimation methods are histograms and KDEs. They achieve a MISE of $\mathcal{O}(n^{-2/3})$ and $\mathcal{O}(n^{-4/5})$, respectively, when their parameters are selected optimally to find the sweet spot between overfitting and oversmoothing (Raykar and Duraiswami 2006; Scott 2015; L'Ecuyer and Puchhammer 2022). For the LRDE methods examined in this paper, under the given assumptions, the ISB is zero and the MISE is $\mathcal{O}(n^{-1})$.

2.2 The LRDE without boundary terms

The cdf of X can be written as

$$F(x) = \mathbb{P}[h(\mathbf{Y}) \le x] = \int_R \mathbb{I}[h(\mathbf{y}) \le x] f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}\mathbf{y}. \tag{2}$$

The aim is to obtain an expression for the derivative $F'(x)$ of the above integral with respect to x, which can be written as the expectation of a random variable that we know how to sample. We cannot just move the differentiation with respect to x inside the integral in (2), because the indicator function is discontinuous in x. To get around that, the idea is to introduce a change of variable $\mathbf{y} \mapsto \mathbf{z} = \mathbf{z}(x)$ of the form $\mathbf{y} = \varphi(\mathbf{z};x)$ for a family of one-to-one functions $\{\varphi(\cdot;x), x \in [a,b]\}$ such that the inequality $h(\varphi(\mathbf{z};x)) \le x$ reduces to $\tilde{h}(\mathbf{z}) \le 1$, for some function $\tilde{h}$ independent of x when $\mathbf{z}$ is given. The indicator term $\mathbb{I}[h(\mathbf{y}) \le x]$ can then be rewritten as $\mathbb{I}[\tilde{h}(\mathbf{z}) \le 1]$, which no longer depends on x. We assume for now that $\tilde{R} = \varphi^{-1}(R) \stackrel{\text{def}}{=} \varphi^{-1}(R;x)$ is independent of x, where φ^{-1} denotes the inverse of $\varphi(\mathbf{z};x)$ with respect to $\mathbf{z}$ for fixed x. This ensures that the integration domain remains independent of x after the change of variable. We will lift this condition in Section 2.3. The dependence on x is thus entirely moved into the distribution of $\mathbf{z}$, and we can rewrite

$$F(x) = \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le 1] f_{\mathbf{Y}}(\varphi(\mathbf{z};x)) |J_{\varphi}(\mathbf{z};x)| \mathrm{d}\mathbf{z},$$

where $|J_\varphi(\mathbf{z};x)|$ is the Jacobian of the transformation $\mathbf{y} = \varphi(\mathbf{z};x)$, defined as the determinant of the matrix of partial derivatives of the coordinates of $\varphi(\mathbf{z};x)$ with respect to the coordinates of $\mathbf{z}$ (for fixed x). This Jacobian is assumed to exist and be nonzero.

Suppose we want to estimate the derivative at $x = x_0$. In a small open neighborhood $\Upsilon(x_0)$ of x_0, the *likelihood ratio* between the density of $\mathbf{z}$ for x and for x_0 is

$$L(\mathbf{z};x,x_0) = \frac{f_{\mathbf{Y}}(\varphi(\mathbf{z};x))|J_\varphi(\mathbf{z};x)|}{f_{\mathbf{Y}}(\varphi(\mathbf{z};x_0))|J_\varphi(\mathbf{z};x_0)|}.$$

In this neighborhood, we have

$$F(x) = \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le 1] L(\mathbf{z};x,x_0) f_{\mathbf{Y}}(\varphi(\mathbf{z};x_0))|J_\varphi(\mathbf{z};x_0)| \mathrm{d}\mathbf{z}. \tag{3}$$

Under appropriate conditions to be specified below, we can take the derivative with respect to x inside the integral in (3), to obtain

$$\begin{aligned}
f(x) = F'(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le c] L(\mathbf{z};x,x_0) f_{\mathbf{Y}}(\varphi(\mathbf{z};x_0))|J_\varphi(\mathbf{z};x_0)| \mathrm{d}\mathbf{z} \\
&= \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le 1] \left(\frac{\mathrm{d}}{\mathrm{d}x} L(\mathbf{z};x,x_0)\right) f_{\mathbf{Y}}(\varphi(\mathbf{z};x_0))|J_\varphi(\mathbf{z};x_0)| \mathrm{d}\mathbf{z} \\
&= \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le 1] \left(\frac{\mathrm{d}}{\mathrm{d}x} L(\mathbf{z};x,x_0)\right) \frac{f_{\mathbf{Y}}(\varphi(\mathbf{z};x))|J_\varphi(\mathbf{z};x)|}{L(\mathbf{z};x,x_0)} \mathrm{d}\mathbf{z} \\
&= \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \le 1] \left(\frac{\mathrm{d}}{\mathrm{d}x} \ln L(\mathbf{z};x,x_0)\right) f_{\mathbf{Y}}(\varphi(\mathbf{z};x))|J_\varphi(\mathbf{z};x)| \mathrm{d}\mathbf{z} \\
&= \int_{R} \mathbb{I}[h(\mathbf{y}) \le x] S(\mathbf{y},x) f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}\mathbf{y} \\
&= \mathbb{E}[\mathbb{I}[h(\mathbf{Y}) \le x] S(\mathbf{Y},x)],
\end{aligned} \tag{4}$$

where

$$\begin{aligned}
S(\mathbf{y},x) &= S(\varphi(\mathbf{z};x),x)\big|_{\mathbf{z}=\varphi^{-1}(\mathbf{y};x)} = \frac{\mathrm{d}\ln L(\mathbf{z};x,x_0)}{\mathrm{d}x}\Big|_{\mathbf{z}=\varphi^{-1}(\mathbf{y};x)} = \frac{\mathrm{d}\ln(f_{\mathbf{Y}}(\varphi(\mathbf{z};x))|J_\varphi(\mathbf{z};x)|)}{\mathrm{d}x}\Big|_{\mathbf{z}=\varphi^{-1}(\mathbf{y};x)} \\
&= \left(\nabla(\ln f_{\mathbf{Y}})(\mathbf{y})^\top \nabla_x \varphi(\mathbf{z};x) + \frac{\mathrm{d}\ln|J_\varphi(\mathbf{z};x)|}{\mathrm{d}x}\right)\Big|_{\mathbf{z}=\varphi^{-1}(\mathbf{y};x)}
\end{aligned}$$

is the *score function* associated with L. This yields the unbiased LRDE (for one sample)

$$\hat{f}(x) = \mathbb{I}[h(\mathbf{Y}) \le x] S(\mathbf{Y},x), \tag{5}$$

where $\mathbf{Y}$ is generated from the density $f_{\mathbf{Y}}$.

We made the change of variable to be able to move the differentiation inside the integral, and after that we reversed the change of variable. Our approach can be seen as an application of the LR derivative estimation methodology (Glynn 1990; L'Ecuyer 1990), with an additional change of variable to be able to apply the method for density estimation. Formula (5) was already derived in L'Ecuyer and Puchhammer (2022) with a more specific change of variable.

Assumption 1 We assume that $\mathbf{Y}$ has a density $f_{\mathbf{Y}}$ with respect to the Lebesgue measure on R. There is a family of bijective functions $\{\varphi(\cdot;x) : \tilde{R} \to R,\ x \in [a,b]\}$ so that $h(\varphi(\mathbf{z};x)) \le x$ is equivalent to $\tilde{h}(\mathbf{z}) \le 1$ for some function $\tilde{h}$ independent of x. Furthermore, φ and $f_{\mathbf{Y}}$ satisfy

(A1) For all $x \in [a,b]$, the mapping $\varphi(\cdot;x)$ has continuous partial derivatives and $|J_\varphi(\cdot;x)| \neq 0$.

(A2) With probability 1 over the realizations of $\mathbf{Y} = \varphi(\mathbf{Z};x)$, $f_\mathbf{Y}(\varphi(\mathbf{Z};x))|J_\varphi(\mathbf{Z};x)|$ is a continuous function of x over $[a,b]$, and is differentiable except perhaps at a countable set of points $D(\mathbf{Y}) \subset [a,b]$. There is also a random variable Γ defined over the same probability space as $\mathbf{Y}$ with $\mathbb{E}[\Gamma] < \infty$ and

$$\sup_{x\in[a,b]\setminus D(\mathbf{Y})} |\mathbb{I}[h(\mathbf{Y}) \leq x] S(\mathbf{Y},x)| \leq \Gamma.$$

Condition (A1) guarantees that the change of variable with $\varphi(\mathbf{z};x)$ is justified and (A2) allows us to exchange the order of integration and differentiation in the second line of (4).

Theorem 1 Suppose $\tilde{R} = \varphi^{-1}(R;x)$ is independent of x. Under Assumption 1, the LRDE (5) is unbiased for the density $f(x)$ at x for all $x \in [a,b]$. If $\mathbb{E}[\Gamma^2] \leq K_\gamma$ for some constant $K_\gamma < \infty$, then its variance is also bounded by K_γ uniformly over the interval $[a,b]$, so its IV is bounded by $(b-a)K_\gamma$.

What is appropriate for $\varphi(\mathbf{z};x)$ depends entirely on the model $h(\mathbf{y})$, and we cannot provide a generic rule that always works for its selection. In several examples that we have examined (not all given here because of space limitation), we had good success with the following strategy: rewrite $\mathbf{y}$ as a function of $\mathbf{z}$ and x so that $h(\varphi(\mathbf{z};x))$ is either $\hat{h}(\mathbf{z})x$ or $\hat{h}(\mathbf{z})+x$ for some measurable function $\hat{h}$. In both cases, the indicator function in (2) becomes independent of x. We illustrate this in the following simple example.

Example 1 (A sum of random variables) Suppose that $R = \mathbb{R}^d$ or $R = [0,\infty)^d$, and that $\mathbf{Y} = (Y_1,\ldots,Y_d)^\top$ has a differentiable density $f_\mathbf{Y}$ over R. Let $X = h(\mathbf{Y}) = Y_1 + Y_2 + \cdots + Y_d = \mathbf{1}\cdot\mathbf{Y}$ (the scalar product), where $\mathbf{1}$ is the vector with all entries equal to 1. We want to estimate the density at some point $x > 0$.

We first take the change of variable

$$\mathbf{y} = \varphi(\mathbf{z};x) = x\mathbf{z}. \tag{6}$$

This gives $h(x\mathbf{z}) = xh(\mathbf{z}) \leq x$ iff $\tilde{h}(\mathbf{z}) \leq 1$ independent of x with $\tilde{h} = h$. For each case of R, this φ also gives $\varphi^{-1}(R;x) = R$, which does not depend on x. Moreover, $|J_\varphi(\mathbf{z};x)| = x^d$, $\nabla_x\varphi(\mathbf{z};x) = \mathbf{z}$, $S(\mathbf{y},x) = (d + (\nabla(\ln f_\mathbf{Y})(\mathbf{y}))\cdot\mathbf{y})/x$, and Theorem 1 applies. The resulting LRDE (5) is the special case considered in Laub et al. (2019); their Proposition 1 was already proving that $\mathbb{I}[\mathbf{1}\cdot\mathbf{Y} \leq x]S(\mathbf{Y},x)$ is an unbiased estimator of the density of X at x for this special case. If $Y_1,\ldots,Y_d$ are independent and Y_j has density f_j, we further have $\ln f_\mathbf{Y}(\mathbf{y}) = \sum_{j=1}^d \ln f_j(y_j)$. Then,

$$S(\mathbf{y},x) = \frac{d + (\nabla(\ln f_\mathbf{Y})(\mathbf{y}))\cdot\mathbf{y}}{x} = \frac{1}{x}\left(d + \sum_{j=1}^d y_j \frac{f_j'(y_j)}{f_j(y_j)}\right). \tag{7}$$

An alternative change of variable is

$$\mathbf{y} = \varphi(\mathbf{z};x) = \mathbf{z} + x\mathbf{e}_j, \tag{8}$$

where $\mathbf{e}_j$ is the jth unit vector in $\mathbb{R}^d$. We have $h(\varphi(\mathbf{z};x)) \leq x$ iff $h(\mathbf{z}) \leq 0$, so we can take $\tilde{h} = h+1$. Here, for $R = \mathbb{R}^d$, we have $\varphi^{-1}(R;x) = R$ independent of x, and we obtain an unbiased LRDE of the form (5). We then have $\nabla_x\varphi(\mathbf{z}) = \mathbf{e}_j$, $|J_\varphi(\mathbf{z};x)| = 1$, and if the Y_j are independent, $S(\mathbf{y},x) = f_j'(y_j)/f_j(y_j)$. Theorem 1 applies and in this case we only need the density f_j to be differentiable (not the other ones). A similar estimator was obtained by Peng et al. (2020) in a slightly less general setting. For $R = [0,\infty)^d$, on the other hand, $\varphi^{-1}(R;x) = [-x,\infty)\times[0,\infty)^{d-1}$ is *not* independent of x, so Theorem 1 does not apply. In fact, the LRDE (5) is biased in this case. □

2.3 The LRDE with boundary terms

Sometimes, the integration domain $\tilde{R}$ in (3) depends on x. We saw an example of that already at the end of Example 1, with $\varphi(\mathbf{z};x) = \mathbf{z} + x\mathbf{e}_j$ and $R = [0,\infty)^d$. In this case, we denote the domain by $\tilde{R}(x)$, and we have to account for the dependence by differentiating the region boundary with respect to x, assuming that the boundary moves smoothly as a function of x. More specifically, let $\partial\tilde{R}(x)$ denote the boundary of $\tilde{R}(x)$, and let $\mathbf{b}(\mathbf{z}(x),x)$ denote the rate of displacement of $\partial\tilde{R}(x)$ as a function of x in the direction normal to this boundary and pointing outward of $\tilde{R}(x)$, at the boundary point $\mathbf{z}=\mathbf{z}(x)\in\partial\tilde{R}(x)$. The infinitesimal displacement $\mathbf{b}(\mathbf{z}(x),x)\mathrm{d}x$ enlarges or shrinks the region $\tilde{R}(x)$ at point $\mathbf{z}(x)$, and this changes the integral at a rate given by $\mathbf{b}(\mathbf{z}(x),x)\mathrm{d}x$ times the value of the integrand at $\mathbf{z}(x)$, if we assume that this integrand ($[\cdots]$ below) is continuous in x in a neighborhood of this point. The derivative of (3) is then

$$F'(x)=\frac{\mathrm{d}}{\mathrm{d}x}\int_{\tilde{R}(x)}[\cdots]\mathrm{d}\mathbf{z}=\int_{\tilde{R}(x)}\frac{\mathrm{d}}{\mathrm{d}x}[\cdots]\mathrm{d}\mathbf{z}+\int_{\partial\tilde{R}(x)}[\cdots]\mathbf{b}(\mathbf{z}(x),x)\mathrm{d}\mathbf{z}. \tag{9}$$

This is Leibniz's integral rule (Wikipedia 2022). Formula (9) is also useful when $f_{\mathbf{Y}}$ has discontinuities in R, and we can partition R into disjoint subregions $R=R_1\cup R_2\cup\cdots$ in which $f_{\mathbf{Y}}$ is differentiable. We can then compute the integral over each subregion and add up. Even when $\varphi^{-1}(R;x)$ is independent of x, the boundaries of the subregions $\tilde{R}_\ell(x)=\varphi^{-1}(R_\ell;x)$ may depend on x, in which case the estimator will be a sum of terms of the form (9).

By undoing the change of variable, the last term in (9) can be rewritten as

$$\int_{\partial\tilde{R}(x)}[\cdots]\mathbf{b}(\mathbf{z}(x),x)\mathrm{d}\mathbf{z}=\int_{\partial R}\mathbb{I}[h(\mathbf{y})\le x]\mathbf{b}(\varphi^{-1}(\mathbf{y},x),x)f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y}.$$

We now derive an explicit form of the LRDE for the common situation in which the region R is a rectangular box $R=\prod_{j=1}^{d}(\alpha_j,\beta_j)$, where $\alpha_j<\beta_j$ are in $\mathbb{R}\cup\{\pm\infty\}$, and $\tilde{R}(x)$ depends on x. In the rest of this paper, any expression that contains $\beta_j=\infty$ or $\alpha_j=-\infty$ must be interpreted as a limit. Under our continuity assumptions, for each x, $\varphi(\cdot;x)$ is also a bijection between the boundary of $\tilde{R}(x)$ and the boundary of R. We also assume that $f_{\mathbf{Y}}$ is differentiable on R. In this case, each boundary panel of R lies on a hyperplane of the form $y_j=\alpha_j$ or $y_j=\beta_j$ for some index j. If $\mathbf{y}$ lies on the boundary panel with $y_j=\beta_j$, which we denote ∂R_j^+, and if $\mathbf{z}(x)=\varphi^{-1}(\mathbf{y};x)$, then the displacement rate of the corresponding boundary piece of $\tilde{R}(x)$ in the positive perpendicular direction is the derivative with respect to x of the jth coordinate of $\mathbf{z}(x)=\varphi^{-1}(\mathbf{y};x)$. That is, we have

$$\mathbf{b}(\mathbf{z}(x),x)=(\nabla_x\mathbf{z}(x))\cdot\mathbf{e}_j=(\nabla_x\varphi^{-1}(\mathbf{y};x))\cdot\mathbf{e}_j\stackrel{\text{def}}{=}r_j(\mathbf{y},x).$$

Let $R_{-j}=\prod_{k\ne j}(\alpha_k,\beta_k)$ and let $\mathbf{y}_{-j}$ be the vector $\mathbf{y}$ with its jth coordinate removed. We then have

$$\begin{aligned}\int_{\partial R_j^+}\mathbb{I}[h(\mathbf{y})\le x]\mathbf{b}(\varphi^{-1}(\mathbf{y},x),x)f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y} &= \int_{\partial R_j^+}\mathbb{I}[h(\mathbf{y})\le x]r_j(\mathbf{y},x)f_{\mathbf{Y}_{-j}}(\mathbf{y}_{-j})f_{Y_j|\mathbf{Y}_{-j}}(y_j\mid\mathbf{y}_{-j})\mathrm{d}\mathbf{y}\\ &= \int_{R_{-j}}\mathbb{I}[h(\mathbf{y})\le x]r_j(\mathbf{y},x)f_{\mathbf{Y}_{-j}}(\mathbf{y}_{-j})f_{Y_j|\mathbf{Y}_{-j}}(\beta_j\mid\mathbf{y}_{-j})\mathrm{d}\mathbf{y}_{-j}\\ &= \mathbb{E}[\mathbb{I}[h(\mathbf{Y}_{-j}^+)\le x]r_j(\mathbf{Y}_{-j}^+,x)f_{Y_j|\mathbf{Y}_{-j}}(\beta_j\mid\mathbf{Y}_{-j})]\end{aligned} \tag{10}$$

where $\mathbf{Y}_{-j}^+$ is the vector $\mathbf{Y}$ in which Y_j has been replaced by β_j.

For $\mathbf{y}$ on the boundary panel ∂R_j^- with $y_j=\alpha_j$ instead, we need to add a negative sign because the box is enlarged by a displacement in the negative direction, so we get $\mathbf{b}(\mathbf{z}(x),x)=-r_j(\mathbf{y},x)$, and the expectation in (10) becomes $\mathbb{E}[\mathbb{I}[h(\mathbf{Y}_{-j}^-)\le x]r_j(\mathbf{Y}_{-j}^-,x)f_{Y_j|\mathbf{Y}_{-j}}(\alpha_j\mid\mathbf{Y}_{-j})]$ where $\mathbf{Y}_{-j}^-$ is the vector $\mathbf{Y}$ in which Y_j has been replaced by α_j.

By summing over all the $2d$ boundary panels, we obtain that if

$$B(\mathbf{y},x)=\sum_{j=1}^{d}\Big(\mathbb{I}[h(\mathbf{y}_{-j}^+)\le x]r_j(\mathbf{y}_{-j}^+,x)f_{Y_j|\mathbf{Y}_{-j}}(\beta_j\mid\mathbf{y}_{-j})-\mathbb{I}[h(\mathbf{y}_{-j}^-)\le x]r_j(\mathbf{y}_{-j}^-,x)f_{Y_j|\mathbf{Y}_{-j}}(\alpha_j\mid\mathbf{y}_{-j})\Big) \tag{11}$$

(in which some of the terms might be 0), then $\mathbb{E}[B(\mathbf{Y},x)]=\int_{\partial R}\mathbb{I}[h(\mathbf{y})\le x]\mathbf{b}(\varphi^{-1}(\mathbf{y},x),x)f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y}$.

Assumption 2 Suppose Assumption 1 holds. Let $R = (\alpha, \beta)$ and suppose $f_{\mathbf{Y}}$ is continuous on the closure of R. We also assume that:

(A3) For each $\mathbf{y} \in R$, the inverse transformation $\varphi^{-1}(\mathbf{y};x)$ is differentiable in x on $[a,b]$, and each term in (11) is well-defined, w.p.1.

Theorem 2 Under Assumption 2, the following estimator is unbiased for $f(x)$ for all $x \in [a,b]$:

$$\hat{f}(x) = \mathbb{I}[h(\mathbf{Y}) \le x] S(\mathbf{Y},x) + B(\mathbf{Y};x). \tag{12}$$

Example 2 Consider the model from Example 1, $h(\mathbf{Y}) = \mathbf{1} \cdot \mathbf{Y}$. Let the Y_j be independent and exponentially distributed with rate $\lambda_j > 0$, and let f_j be the density of Y_j for all j. In this case, $f_j'(x)/f_j(x) = -\lambda_j$. If we use the change of variable $\mathbf{y} = x\mathbf{z}$ in (6), the LRDE is $\hat{f}(x) = \mathbb{I}[\mathbf{1} \cdot \mathbf{Y} \le x](d - Y_1\lambda_1 - \cdots - Y_d\lambda_d)/x$ and is unbiased for $f(x)$ for all $x > 0$ by Theorem 1.

If we use the change of variable $\mathbf{y} = \mathbf{z} + x\mathbf{e}_j$ in (8), the jth coordinate of $\varphi^{-1}([0,\infty)^d;x)$ (i.e., the left boundary panel ∂R_j^-) depends on x, so we need to add $B(\mathbf{y};x)$ in (11). We have $r_j(\mathbf{y},x) = (\nabla_x(\mathbf{y} - x\mathbf{e}_j)) \cdot \mathbf{e}_j = -1$ and the other r_k's are zero. Hence, $r_j(\mathbf{y}_{-j}^-;x) f_j(\alpha_j) = -f_j(0) = -\lambda_j$. Moreover, $S(\mathbf{y},x) = f_j'(y_j)/f_j(y_j) = \lambda_j$ by Example 1. By Theorem 2, the following LRDE is unbiased for $f(x)$:

$$\hat{f}(x) = \mathbb{I}[\mathbf{1} \cdot \mathbf{Y} \le x] S(\mathbf{Y},x) + B(\mathbf{Y},x) = -\mathbb{I}[\mathbf{1} \cdot \mathbf{Y} \le x]\lambda_j - \mathbb{I}[\mathbf{1} \cdot \mathbf{Y}_{-j} \le x](-\lambda_j) = \mathbb{I}[\mathbf{1} \cdot \mathbf{Y}_{-j} \le x < \mathbf{1} \cdot \mathbf{Y}]\lambda_j. \quad \square$$

3 DENSITY ESTIMATION WITH THE GLR

Peng et al. (2018) suggested using the GLR gradient estimator as a density estimator (GLRDE) and gave a general formula for it. Simpler formulas for specific settings were later proposed by Peng et al. (2020), Peng et al. (2022a). All these papers consider a representation of the cdf of X of the form

$$F(x) = \mathbb{E}[\mathbb{I}[h(\mathbf{Y}) \le x]] = \mathbb{E}[\psi(g(\mathbf{Y},x))], \tag{13}$$

where $g : \mathbb{R}^d \times [a,b] \to \mathbb{R}^d$ is sufficiently smooth and $\psi : \mathbb{R}^d \to \mathbb{R}$ is measurable and independent of x, but can be discontinuous. This means that the dependence on x is put entirely into g and all singularities into ψ. The function $g(\cdot,x)$ here represents a change of variable just like φ, but in the other direction; i.e., we have $\mathbf{z} = g(\mathbf{y},x) = \varphi^{-1}(\mathbf{y};x)$. We obtain our earlier setting by also taking $\psi(\mathbf{z}) = \mathbb{I}[\tilde{h}(\mathbf{z}) < 1]$. That is, the general form (13) covers our setting. We now look more closely at the GLRDEs proposed by Peng et al. (2018) and by Peng et al. (2022a), and the assumptions given in those papers.

3.1 The Original GLRDE

The original GLRDE form (Peng et al. 2018) was

$$\begin{aligned} \hat{f}(x) &= \mathbb{I}[h(\mathbf{Y}) \le x] S(\mathbf{Y},x) \quad \text{with} \qquad (14) \\ S(\mathbf{y},x) &= -\operatorname{trace}\left(J_g^{-1}(\mathbf{y},x)\nabla_x J_g(\mathbf{y},x)\right) + \sum_{j=1}^{d} \mathbf{e}_j^\top J_g^{-1}(\mathbf{y},x)\left(\nabla_{y_j} J_g\right) J_g^{-1}(\mathbf{y},x)\nabla_x g(\mathbf{y},x) \\ &\quad - \left(J_g^{-1}(\mathbf{y},x)\nabla_x g(\mathbf{y},x)\right)^\top \nabla \ln f_{\mathbf{Y}}(\mathbf{y}). \end{aligned}$$

This is the same estimator as our LRDE in (5), with a more detailed development of the score function $S(\mathbf{Y},x)$. Peng et al. (2018) gave the following sufficient unbiasedness conditions for this GRLDE.

Assumption 3

(B1) The density $f_{\mathbf{Y}}$ is strictly positive and continuously differentiable over all of $\mathbb{R}^d$.

(B2) The function $g(\mathbf{y},x)$ is invertible in $\mathbf{y}$, and g is twice continuously differentiable on $\mathbb{R}^d \times [a,b]$.
(B3) One has $\lim_{y_j \to \pm\infty} \int_{\mathbb{R}^{d-1}} \sup_{x\in[a,b]} |r_j(\mathbf{y};x)| f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}\mathbf{y}_{-j} = 0$ for all $1 \le j \le d$
(B4) The GLRDE $\hat{f}(x)$ in (14) satisfies $\int_{\mathbb{R}^d} \sup_{x\in[a,b]} |\hat{f}(x)| \mathrm{d}\mathbf{y} < \infty$.

The majority of standard distributions violate (B1). For example, the exponential density has a jump at 0 and is zero on the left. But we showed in Example 2 that it can satisfy our assumptions and can give an unbiased LRDE (5). Peng et al. (2018) mention that (B1) can often be satisfied through a change of variable, but this is not always easy. Condition (B3) also imposes strong assumptions on the tails of $f_{\mathbf{Y}}$.

3.2 The GLRDE with Uniform Input

To get around (B1), Peng et al. (2022a) proposed a setting in which $\mathbf{Y}$ is replaced by a vector $\mathbf{U}$ uniformly distributed over the unit hypercube $(0,1)^d$. This is essentially a special case of our setting of Section 2.3, with $R = (\alpha,\beta) = (0,1)^d$ and $f_{\mathbf{Y}}$ taken as the uniform distribution. This setting can still handle a more general $\mathbf{Y}$ by incorporating a transformation from $\mathbf{U}$ to $\mathbf{Y}$ into the function g. The proposed GLRDE is

$$\hat{f}(x) = \mathbb{I}[h(\mathbf{U}) \le x] S(\mathbf{U},x) + \sum_{j=1}^{d} \left(\lim_{u_j \uparrow 1} \psi(g(\mathbf{u},x)) r_j(\mathbf{u},x) - \lim_{u_j \downarrow 0} \psi(g(\mathbf{u},x)) r_j(\mathbf{u},x) \right), \quad \text{where}$$

$$r_j(\mathbf{u},x) = \left(J_g^{-1}(\mathbf{u};x) \nabla_x g(\mathbf{u},x) \right)^{\top} \mathbf{e}_j, \quad 1 \le j \le d.$$

Peng et al. (2022a) showed that this estimator is unbiased for $f(x)$ under the following conditions.

Assumption 4 We have $F(x) = \mathbb{E}[\psi(g(\mathbf{U},x))]$ with $g(\cdot,x) : (0,1)^d \times [a,b] \to \mathbb{R}$. Moreover:

(C1) The function $g(\mathbf{u},x)$ is invertible in $\mathbf{u}$ and twice continuously differentiable on $(0,1)^d \times [a,b]$. Furthermore, the matrix $J_g(\mathbf{u},x)$ is invertible for almost $\mathbf{u}$.
(C2) For all $1 \le j \le d$,

$$\lim_{u_j \uparrow 1} \sup_{x\in[a,b], \mathbf{u}_{-j} \in (0,1)^{d-1}} |r_j(\mathbf{u},x)| = \lim_{u_j \downarrow 0} \sup_{x\in[a,b], \mathbf{u}_{-j} \in (0,1)^{d-1}} |r_j(\mathbf{u},x)| = 0.$$

(C3) For $\hat{f}(x)$ defined in (14) we have $\int_{(0,1)^d} \sup_{x\in[a,b]} |\hat{f}(x)| \mathrm{d}\mathbf{u} < \infty$.

Assumption 4 allows $\mathbf{Y}$ to have a more general support than (B1), but (C2) still requires a nice behaviour of $f_{\mathbf{Y}}$ on the boundary of its support. To alleviate this, Peng et al. (2022a) introduce another set of conditions:

Assumption 5 Take Assumption 4 without (C2), and add:

(D1) There is a sequence of functions $\psi_\varepsilon : \mathbb{R}^d \to \mathbb{R}$ that are differentiable an infinite number of times, and a $p > 1$, such that

$$\lim_{\varepsilon \to 0} \sup_{x\in[a,b]} \int_{(0,1)^d} |\psi_\varepsilon(g(\mathbf{u},x)) - \psi(g(\mathbf{u},x))|^p \mathrm{d}\mathbf{u} = 0.$$

Furthermore, if $d \ge 2$, we have for some fixed $\varepsilon > 0$ and all $u_j \in (0,1) \setminus [\varepsilon, 1-\varepsilon]$, $1 \le j \le d$,

$$\lim_{\varepsilon \to 0} \sup_{x\in[a,b]} \int_{(0,1)^{d-1}} |\psi_\varepsilon(g(\mathbf{u},x)) - \psi(g(\mathbf{u},x))|^p \mathrm{d}\mathbf{u}_{-j} = 0.$$

If $d = 1$, the above condition holds without integration and with $p = 1$.
(D2) The functions r_j , $1 \le j \le d$, satisfy $\int_{(0,1)^{d-1}} \sup_{x\in[a,b], u_j\in(0,1)} |\psi(g(\mathbf{u},x)) r_j(\mathbf{u},x)| \mathrm{d}\mathbf{u}_{-j} < \infty$.

This avoids (C2), but condition (D1) is usually hard to verify. Peng et al. (2022a) simplify this for the special case in which

$$\mathbb{I}[h(\mathbf{y}) \le x] = \prod_{j=1}^{d} \mathbb{I}[g_j(u_j,x) \le 0]. \tag{15}$$

This decomposition can be applied for example if h is the maximum of several functions. Then, (D1) is satisfied if for a fixed $\varepsilon > 0$ and for all $1 \le j \le d$, we have

$$\inf_{x\in[a,b],u_j\in[\varepsilon,1-\varepsilon]} \left|\mathrm{d}g_j(u_j,x)/\mathrm{d}u_j\right| > 0 \text{ and } \inf_{x\in[a,b],u_j\in(0,1)\setminus[\varepsilon,1-\varepsilon]} \left|g_j(u_j,x)\right| > 0, \qquad \text{(D1')}$$

Another simplification works for the special case in which

$$g_j(u_j,x) = \xi_j(x)\eta_j(u_j) \quad \text{for all} \quad 1 \le j \le d. \qquad (16)$$

Conditions (C2), (C3), (D2), and (D1') can then be replaced respectively by

$$\lim_{u_j\uparrow 1} |\mathrm{d}\log\eta_j(u_j)/\mathrm{d}u_j| = \lim_{u_j\downarrow 0} |\mathrm{d}\log\eta_j(u_j)/\mathrm{d}u_j| = \infty, \qquad \text{(C2')}$$

$$\mathbb{E}\left[|\eta_j(U_j)\eta_j''(U_j)|/\left(\eta_j'(U_j)\right)^2\right] < \infty, \qquad \text{(C3')}$$

$$\inf_{u_j\in(0,1)} |\mathrm{d}\log\eta_j(u_j)/\mathrm{d}u_j| > 0, \qquad \text{(D2')}$$

$$\inf_{x\in[a,b]} |\xi_j(x)| > 0, \quad \inf_{u_j\in[\varepsilon,1-\varepsilon]} |\eta_j'(u_j)| > 0, \quad \inf_{u_j\in(0,1)\setminus[\varepsilon,1-\varepsilon]} |\eta_j(u_j)| > 0. \qquad \text{(D1'')}$$

However, the decompositions (15) and (16) hold simultaneously only in special cases.

The following example exhibits simple situations in which our assumptions hold, showing that the LRDE is unbiased, yet none of the Assumptions 3 through 5, including their simplifications, are satisfied. This example illustrates the limitations of the simplified conditions (C2') and (D1').

Example 3 Let U_1, U_2 be two independent uniforms on $(0,1)$, $h(\mathbf{U}) = \max\{U_1,U_2\}$, and $[a,b] \subset (0,1)$. For the LRDE, we set $R = (0,1)^2$ and consider the change of variable $\varphi(\mathbf{z};x) = x\mathbf{z}$. As $\varphi^{-1}(\mathbf{u};x) = \mathbf{u}/x$, $\tilde{R}(x) = \varphi^{-1}(R,x) = (0,1/x)^2$ depends on x, we construct the LRDE (12). We have $S(\mathbf{u},x) = 2/x$, as demonstrated in Example 1, and $r_j(\mathbf{u},x) = \nabla_x(\varphi^{-1}(\mathbf{u};x)) \cdot \mathbf{e}_j = -u_j/x^2$, $j = 1,2$. This is zero when $u_j = 0$, and $\mathbb{I}[h(\mathbf{u}) \le x] = 0$ for $u_j = 1$. Hence, $B(\mathbf{u},x) = 0$, and the LRDE is $\hat{f}(x) = 2\mathbb{I}[\max\{U_1,U_2\} \le x]/x$.

This estimator coincides with the GLRDE using $g_j(\mathbf{u},x) = u_j/x$. Clearly, (B1) from Assumption 3 does not hold. Since $r_j(\mathbf{u},x) = -u_j/x^2 \to -1/x^2 \neq 0$ for $u_j \to 1$, (C2) is also violated. We have a decomposition of the form $g_j(u_j,x) = \xi_j(x)\eta_j(u_j)$ (16), so we can check the alternative condition (C2'). Here, $\eta_j(u) = u$ and $\mathrm{d}\log\eta_j(u)/\mathrm{d}u = 1/u \to 1$ for $u \to 1$. Thus, Assumption 4 does not hold. Verifying (D1) is extremely hard, but we can rewrite $\mathbb{I}[\max\{u_1,u_2\} \le x] = \mathbb{I}[u_1/x - 1 \le 0]\mathbb{I}[u_2/x - 1 \le 0]$, see (15), so we can check (D1') instead, with $g_j(u,x) = u/x - 1$. However, any pair $u = x$ gives $g_j(u,x) = 0$ and the second infimum in (D1') cannot be positive. □

4 EXAMPLE: A STOCHASTIC ACTIVITY NETWORK

We consider a stochastic activity network (SAN) example, for which we compare various LRDE constructions with respect to their performance, and we demonstrate how changing model parameters can affect their unbiasedness. A SAN is a directed graph connecting a source with a sink, in which each edge has a random length Y_j, $1 \le j \le d$. We consider the SAN depicted in Figure 1, taken from L'Ecuyer and Puchhammer (2022) with different model parameters. This example has $d = 11$, but larger networks can usually be handled in the exact same way as presented here. SANs model a wide variety of problems with precedence relations between activities, such as maximum flow problems or the lifetimes of multi-component systems; see L'Ecuyer et al. (2022) and L'Ecuyer and Puchhammer (2022). Here, we look at the length X of the longest path from the source to the sink. We refer to the six possible paths by the indexes j of those Y_j that lie on them; $\Pi_1 = \{1,4,10\}$, $\Pi_2 = \{1,4,8,11\}$, $\Pi_3 = \{2,5,10\}$, $\Pi_4 = \{2,5,8,11\}$, $\Pi_5 = \{2,6,9,11\}$, $\Pi_6 = \{3,7,9,11\}$. Denoting by L_l the length of path l for $1 \le l \le 6$, we can write X as

$$X = h(\mathbf{Y}) = \max_{1\le l\le 6} L_l = \max_{1\le l\le 6} \sum_{j\in\Pi_l} Y_j.$$

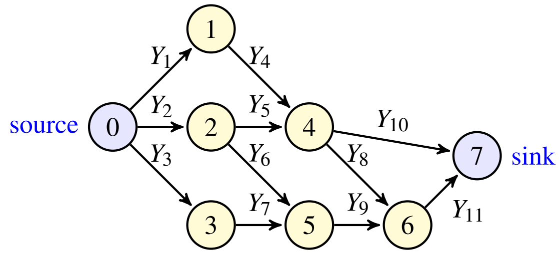

Figure 1: A stochastic activity network 8 nodes and 11 links.

Suppose that the Y_j, $1 \le j \le 11$, are independent and have a Weibull distribution with scale parameter $\lambda = 1$ and shape parameter $\alpha > 0$. I.e., the density of Y_j is $f_j(y) = \alpha y^{\alpha-1} \exp(-y^\alpha)$ for $y > 0$. Notice that f_j is unbounded at 0 when $\alpha < 1$.

We first construct a LRDE without boundary terms. We set $R = (0,\infty)^{11}$, fix $x > 0$, and define the change of variable $\mathbf{y} = \varphi_1(\mathbf{z};x) = x\mathbf{z}$ in (6). We have $h(\varphi_1(\mathbf{z};x)) = xh(\mathbf{z})$, so $h(\mathbf{y}) \le x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) \le 1$. Here, $\tilde{R}(x) = \varphi_1^{-1}(R;x) = (0,\infty)^{11}$ is independent of x, and we can construct the LRDE (5). We derived the score function for this change of variable in (7). Using $f_j'(y)/f_j(y) = (\alpha - 1 - \alpha y^\alpha)/y$, we get

$$\hat{f}_1(x) = \mathbb{I}[h(\mathbf{Y}) \le x] x^{-1} \left(11 + \sum_{j=1}^{11} Y_j f_j'(Y_j)/f_j(Y_j) \right) = \mathbb{I}[h(\mathbf{Y}) \le x] \alpha x^{-1} \left(11 - \sum_{j=1}^{11} Y_j^\alpha \right).$$

By Theorem 1, $\hat{f}_1(x)$ is unbiased for any $x > 0$ and all $\alpha > 0$.

We can derive alternate LRDEs as follows. We take a directed minimal cut, i.e., a set of links $\mathscr{C}$ so that each path from source to sink contains exactly one element of $\mathscr{C}$. In Figure 1, there are 14 possible choices for $\mathscr{C}$, e.g.,$\mathscr{C}_1 = \{10,11\}$, $\mathscr{C}_2 = \{1,2,3\}$, and $\mathscr{C}_3 = \{3,4,5,6\}$. By definition of $\mathscr{C}$, the change of variable $\varphi_{\mathscr{C}}(\mathbf{z};x) = \mathbf{z} + x\sum_{j\in\mathscr{C}} \mathbf{e}_j$ satisfies $h(\varphi_{\mathscr{C}}(\mathbf{z};x)) = h(\mathbf{z}) + x$. Hence, $h(\mathbf{y}) \le x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) + 1 \le 1$. Since $\nabla_x \varphi_{\mathscr{C}}(\mathbf{z};x) = \sum_{j\in\mathscr{C}} \mathbf{e}_j$ and $|J_\varphi| = 1$, the score function is $S(\mathbf{y};x) = \sum_{j\in\mathscr{C}} f_j'(y_j)/f_j(y_j)$. Here, $\varphi_{\mathscr{C}}^{-1}(\mathbf{y};x) = \mathbf{y} - x\sum_{j\in\mathscr{C}} \mathbf{e}_j$, so $\tilde{R}(x) = \varphi_{\mathscr{C}}^{-1}(R;x)$ depends on x in the dimensions $j \in \mathscr{C}$, and we need to consider boundary terms. We have $r_j = (\nabla_x \varphi_{\mathscr{C}}^{-1}(\mathbf{y};x)) \cdot \mathbf{e}_j = -1$ for $j \in \mathscr{C}$ and $r_j = 0$ otherwise, and $\mathbb{I}[h(\mathbf{Y}_{-j}^+) \le x] = 0$. Therefore, the LRDE (12) for an arbitrary directed minimal cut $\mathscr{C}$ is

$$\hat{f}_{\mathscr{C}}(x) = \mathbb{I}[h(\mathbf{Y}) \le x] \sum_{j\in\mathscr{C}} Y_j^{-1} \left(\alpha - 1 - \alpha Y_j^\alpha \right) + f_1(0) \sum_{j\in\mathscr{C}} \mathbb{I}[h(\mathbf{Y}_{-j}^-) \le x].$$

For $\alpha > 1$ and $x > 0$, this estimator is unbiased by Theorem 2. For $\alpha \in (0,1)$, the term $f_1(0)$ is infinite and Theorem 2 does not apply. In some cases, such a closed form formula for arbitrary $\mathscr{C}$ can be very convenient for selecting a $\mathscr{C}$ for which we expect the LRDE to have low variance.

The LRDE $\hat{f}_{\mathscr{C}}$ can also be obtained with the approach suggested in Peng et al. (2020). Their strategy, however, encompasses difficulties which do not appear with our approach. Moreover, it is hard to obtain a formula for a general $\mathscr{C}$. To apply their strategy, we set $\psi(\mathbf{z}) = \prod_{l=1}^{6} \mathbb{I}[z_l \le 0]$ and pick a subset of six variables $u_{j_1}, \ldots, u_{j_6}$ while considering the remaining ones fixed, so that the selection $g_l(u_{j_l},x) = \sum_{k\in\Pi_l} F_k^{-1}(u_k) - x$ yields an invertible J_g. This, however, will not work here, as $|J_g| = 0$ for any possible choice of variables. Indeed, a closer analysis of the network shows that $L_1 + L_4 = L_2 + L_3$. While this dependence is feasibly detectable in this toy model, it can substantially increase the computational overhead if the SAN comprises thousands of paths. L'Ecuyer and Puchhammer (2022) solve this dependence problem by discarding the path Π_4. We then put $(j_1, j_2, \ldots, j_5) = (5,7,8,10,11)$ and define $g_1(u_{10},x) = L_1 - x$, $g_2(u_8,u_{11},x) = L_2 - x$, $g_3(u_5,u_{10},x) = L_3 - x$, $g_4(u_{11},x) = L_5 - x$, and $g_5(u_7,u_{11},x) = L_6 - x$. This will yield the estimator $f_{\mathscr{C}_3}$. Notice that J_g is not diagonal or triangular, so it is virtually impossible to anticipate the structure of J_g^{-1} and this particular form of the estimator from the selected indexes $(j_1, j_2, \ldots, j_5)$ and the discarded Π_4.

For a numerical illustration, we consider five different scenarios, $\alpha \in \{0.5, 1, 1.5, 2, 3\}$. In the first one, $\alpha = 0.5$, $f_j(0)$ is infinite. For $\alpha = 1$, we get the exponential distribution with rate $\lambda = 1$, therefore $f_j(y)$ has a jump of height 1 at zero. For $\alpha = 1.5, 2, 3$, the $f_j(y)$ are continuous on $\mathbb{R}$, but their derivatives are infinite, positive, and zero, respectively, at $y = 0$.

The estimator $\hat{f}_1$ is unbiased for $\alpha > 0$. Similarly, the LRDE $f_{\mathscr{C}}$ is unbiased for the selected α, except for $\alpha = 0.5$. We will compare the estimators obtained with $\hat{f}_1$ and $\hat{f}_{\mathscr{C}}$ for $\mathscr{C} = \mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3$, as defined above. To do that, we take a and b as estimates of the 2.5 and 97.5 percentile of X, respectively, for each α.

We estimate the performance of each LRDE $\hat{f}$ by estimating the MISE (1) using random integration nodes x_k, $1 \le k \le 10^3$, which we sample independently and uniformly in $[a + (k-1)(b-a)/10^3, a + k(b-a)/10^3]$ (a stratification scheme). Since the estimator is unbiased, MISE $=$ IV, it suffices to estimate at each x_k the empirical variance $\widehat{\mathrm{Var}}_k$ of $n = 10^6$ realizations of the estimator $\hat{f}$. This gives the following unbiased estimator for the IV:

$$\widehat{\mathrm{IV}} = \frac{(b-a)}{10^3 n} \sum_{k=1}^{1000} \widehat{\mathrm{Var}}_k.$$

The results are summarized in Table 1. For $\alpha = 0.5$ the only unbiased LRDE is $\hat{f}_1$. For all other choices of α, $\hat{f}_1$ is the estimator with the smallest variance. This is because the score function of the $\hat{f}_{\mathscr{C}_j}$ has a larger variance and/or the presence of non-zero boundary terms for $\alpha = 1$. Among the $f_{\mathscr{C}_j}$, the variance seems to increase slightly with the cardinality of $\mathscr{C}_j$, causing more summands in the score function, with one exception at $\alpha = 1.5$.

Table 1: Estimated IV for various LRDEs under different distribution parameters α for the SAN.

Estimator	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 3.0$
$\hat{f}_1$	2.28E-7	9.08E-7	1.73E-6	2.62E-6	4.48E-6
$\hat{f}_{\mathscr{C}_1}$	-	4.31E-6	1.65E-3	7.57E-5	2.43E-5
$\hat{f}_{\mathscr{C}_2}$	-	9.69E-6	8.63E-4	7.88E-5	3.52E-5
$\hat{f}_{\mathscr{C}_3}$	-	1.72E-5	2.49E-3	1.25E-4	4.72E-5

5 CONCLUSION

We showed how an unbiased density estimator can be obtained by combining the LR derivative estimation method with a clever change of variable. We compared our approach with the recently proposed GLR, which can also be used to construct unbiased density estimators. We gave various examples in which the required assumptions for unbiasedness of the GLR are not satisfied whereas our assumptions hold. This often happens when the GLR estimators turn out to be the same as ours, which suggests that the assumptions made in the GLR papers are more restrictive than needed. In future work, we plan to explore how well these LRDEs can perform and how easily they can be applied in more elaborate applications. We also plan to further study their combination with quasi-Monte Carlo methods and variance-reduction techniques such as conditioning (Ben Abdellah et al. 2021; L'Ecuyer and Puchhammer 2022; Peng et al. 2022b).

ACKNOWLEDGMENTS

P. L'Ecuyer was supported by discovery grant RGPIN-110050 from NSERC-Canada and a Grant from the IVADO Fundamental Research Program. F. Puchhammer was additionally funded by NSERC through the Discovery Grants RGPAS-2020-00093, RGPIN-2020-04897, and RGPIN-238959, and the Department of Statistics and Actuarial Science of the University of Waterloo.

REFERENCES

Asmussen, S. 2018. "Conditional Monte Carlo for Sums, with Applications to Insurance and Finance". *Annals of Actuarial Science* 12(2):455–478.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation*. New York: Springer-Verlag.

Ben Abdellah, A., P. L'Ecuyer, A. Owen, and F. Puchhammer. 2021. "Density Estimation by Randomized Quasi-Monte Carlo". *SIAM Journal on Uncertainty Quantification* 9(1):280–301.

Glynn, P. W. 1987. "Likelihood Ratio Gradient Estimation: an Overview". In *Proceedings of the 1987 Winter Simulation Conference*, edited by A. Thesen, H. Grant, and W. D. Kelton, 366–375. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Glynn, P. W. 1990. "Likelihood Ratio Gradient Estimation for Stochastic Systems". *Communications of the ACM* 33(10):75–84.

Glynn, P. W., and P. L'Ecuyer. 1995. "Likelihood Ratio Gradient Estimation for Regenerative Stochastic Recursions". *Advances in Applied Probability* 27:1019–1053.

Laub, P. J., R. Salomone, and Z. I. Botev. 2019. "Monte Carlo Estimation of the Density of the Sum of Dependent Random Variables". *Mathematics and Computers in Simulation* 161:23–31.

Law, A. M. 2014. *Simulation Modeling and Analysis*. 5th ed. New York: McGraw-Hill.

L'Ecuyer, P. 1990. "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques". *Management Science* 36(11):1364–1383.

L'Ecuyer, P. 1995. "On the Interchange of Derivative and Expectation for Likelihood Ratio Derivative Estimators". *Management Science* 41(4):738–748.

L'Ecuyer, P., and F. Puchhammer. 2022. "Density Estimation by Monte Carlo and Quasi-Monte Carlo". In *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC 2020*, edited by A. Keller, 3–21. Berlin: Springer.

L'Ecuyer, P., F. Puchhammer, and A. Ben Abdellah. 2022. "Monte Carlo and Quasi-Monte Carlo Density Estimation via Conditioning". *INFORMS Journal on Computing* 34(3):1305–1840.

Lei, L., Y. Peng, M. C. Fu, and J.-Q. Hu. 2018. "Applications of Generalized Likelihood Ratio Method to Distribution Sensitivities and Steady-State Simulation". *Discrete Event Dynamic Systems* 28(1):109–125.

Parzen, E. 1962. "On Estimation of a Probability Density Function and Mode". *Annals of Mathematical Statistics* 33(3):1065–1076.

Peng, Y., M. C. Fu, B. Heidergott, and H. Lam. 2020. "Maximum Likelihood Estimation by Monte Carlo Simulation: Towards Data-Driven Stochastic Modeling". *Operations Research* 68(6):1896–1912.

Peng, Y., M. C. Fu, J.-Q. Hu, and B. Heidergott. 2018. "A New Unbiased Stochastic Derivative Estimator for Discontinuous Sample Performances with Structural Parameters". *Operations Research* 66(2):487–499.

Peng, Y., M. C. Fu, J.-Q. Hu, P. L'Ecuyer, and B. Tuffin. 2022a. "Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs". Submitted manuscript. https://hal.inria.fr/hal-02652068/document, accessed April 2, 2022.

Peng, Y., M. C. Fu, J.-Q. Hu, P. L'Ecuyer, and B. Tuffin. 2022b. "Variance Reduction for Generalized Likelihood Ratio Method By Conditional Monte Carlo and Randomized Quasi-Monte Carlo". *Journal of Management Science and Engineering* 7(4):550–577.

Raykar, V. C., and R. Duraiswami. 2006. "Fast Optimal Bandwidth Selection for Kernel Density Estimation". In *Proceedings of the 2006 SIAM International Conference on Data Mining*, edited by J. Ghosh, D. Lambert, D. Skillicorn, and J. Srivastava, 524–528. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Scott, D. W. 2015. *Multivariate Density Estimation*. 2nd ed. New York, New York: Wiley.

Wikipedia 2022. "Leibniz Integral Rule". https://en.wikipedia.org/wiki/Leibniz_integral_rule, accessed April 1, 2022.

AUTHOR BIOGRAPHIES

FLORIAN PUCHHAMMER is a postdoctoral fellow at the Department of Statistics and Actuarial Science at the University of Waterloo, Canada. His main research interests are quasi-Monte Carlo and randomized quasi-Monte Carlo methods, stochastic simulation, and discrepancy theory. Email: florian.puchhammer@gmail.com.

PIERRE L'ECUYER is a Professor in the Departement d'Informatique et de Recherche Opérationnelle, at the Université de Montréal, Canada. He is a member of the CIRRELT and GERAD research centers. His main research interests are random number generation, quasi-Monte Carlo methods, efficiency improvement via variance reduction, sensitivity analysis and optimization of discrete-event stochastic systems, and discrete-event simulation in general. He has published over 300 scientific articles, and has developed software libraries and systems for random number generation and stochastic simulation. He has been a referee for 168 different scientific journals. More information can be found on his web page: http://www.iro.umontreal.ca/~lecuyer. Email: lecuyer@iro.umontreal.ca.