
Central
SUPER-PRESTIGE

LIKELIHOOD RATIO DENSITY ESTIMATION FOR SIMULATION MODELS

Florian Puchhammer and Pierre L'Ecuyer

UNIVERSITY OF WATERLOO
Université de Montréal
GERAD GROUP FOR RESEARCH IN DECISION ANALYSIS
CIRRELT
IVADO

2022 Winter Simulation Conference, Singapore

Context

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of an output random variable X from a model, and compute a confidence interval on $\mathbb{E}[X]$.

But simulation can provides information to do much more! The output data can be used to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or (for better visualization) its density $f(x) = F'(x)$. Here we assume that X has a density f and we want to estimate f over a finite interval $[a, b] \subset \mathbb{R}$.

Context

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of an output random variable X from a model, and compute a confidence interval on $\mathbb{E}[X]$.

But simulation can provides information to do much more! The output data can be used to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or (for better visualization) its density $f(x) = F'(x)$. Here we assume that X has a density f and we want to estimate f over a finite interval $[a, b] \subset \mathbb{R}$.

Classical density estimation in statistics was developed in the context where $X_1, \ldots, X_n$ are **given independent observations of** X and one estimates the density f of X from that.
Leading methods: histogram: $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-2/3})$;
kernel density estimator (KDE); $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-4/5})$.

Context

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of an output random variable X from a model, and compute a confidence interval on $\mathbb{E}[X]$.

But simulation can provides information to do much more! The output data can be used to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or (for better visualization) its density $f(x) = F'(x)$. Here we assume that X has a density f and we want to estimate f over a finite interval $[a, b] \subset \mathbb{R}$.

Classical density estimation in statistics was developed in the context where $X_1, \ldots, X_n$ are **given independent observations of** X and one estimates the density f of X from that.
Leading methods: histogram: $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-2/3})$;
kernel density estimator (KDE); $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-4/5})$.

In this talk, we assume that $X_1, \ldots, X_n$ are **generated by simulation** from a model.
We can choose n and we have some freedom on how the simulation is performed.
Unbiased density estimators can then be defined, under certain conditions.

Density estimation

Want to estimate the density of $X = h(\mathbf{Y}) = h(Y_1, \ldots, Y_s)$, assuming we know how to get Monte Carlo samples of $\mathbf{Y}$ from its multivariate distribution.

Let $\hat{f}_n(x)$ denote the density estimator at x, with sample size n, for $x \in [a, b]$.

Density estimation

Want to estimate the density of $X = h(\mathbf{Y}) = h(Y_1, \ldots, Y_s)$, assuming we know how to get Monte Carlo samples of $\mathbf{Y}$ from its multivariate distribution.

Let $\hat{f}_n(x)$ denote the density estimator at x, with sample size n, for $x \in [a, b]$.

We use simple error measures:

$$\begin{aligned}
\text{MISE} &= \text{mean integrated squared error} = \int_a^b \mathbb{E}[(\hat{f}_n(x) - f(x))^2]\mathrm{d}x = \text{IV} + \text{ISB} \\
\text{IV} &= \text{integrated variance} = \int_a^b \mathrm{Var}[\hat{f}_n(x)]\mathrm{d}x \\
\text{ISB} &= \text{integrated squared bias} = \int_a^b (\mathbb{E}[\hat{f}_n(x)] - f(x))^2\mathrm{d}x
\end{aligned}$$

To minimize the MISE, we may need to balance the IV and ISB.

Conditional Monte Carlo (CMC) for density estimation

Estimate the density $f(x)$ by a **conditional density estimator (CDE)** $f(x \mid \mathcal{G})$, given partial information $\mathcal{G}$. This $\mathcal{G}$ must hide enough information so that $X \mid \mathcal{G}$ has a density $f(x \mid \mathcal{G})$ (no mass point) and yet contain enough information so $f(x \mid \mathcal{G})$ is easy to compute.
This corresponds to applying CMC + IPA to the cdf:

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}[\mathbb{P}[X \le x \mid \mathcal{G}]] \stackrel{?}{=} \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}[X \le x \mid \mathcal{G}]\right] = \mathbb{E}[f(x \mid \mathcal{G})].$$

Assumption CDE. For all realizations of $\mathcal{G}$, $F(x \mid \mathcal{G})$ is a continuous function of x over $[a, b]$, differentiable except perhaps over a denumerable set of points $D(\mathcal{G}) \subset [a, b]$, and $f(x \mid \mathcal{G}) = F'(x \mid \mathcal{G}) = \mathrm{d}F(x \mid \mathcal{G})/\mathrm{d}x$ (when it exists) is bounded uniformly in x by a random variable Γ such that $\mathbb{E}[\Gamma^2] < \infty$.

Proposition CDE: Under Ass. CDE, for $x \in [a, b]$, $\mathbb{E}[f(x \mid \mathcal{G})] = f(x)$ and $\mathrm{Var}[f(x \mid \mathcal{G})] < \mathbb{E}[\Gamma^2]$.

The CDE $f(x \mid \mathcal{G})$ is then unbiased with uniformly bounded variance, so MISE $= \mathcal{O}(n^{-1})$.

Conditional Monte Carlo (CMC) for density estimation

Estimate the density $f(x)$ by a **conditional density estimator (CDE)** $f(x \mid \mathcal{G})$, given partial information $\mathcal{G}$. This $\mathcal{G}$ must hide enough information so that $X \mid \mathcal{G}$ has a density $f(x \mid \mathcal{G})$ (no mass point) and yet contain enough information so $f(x \mid \mathcal{G})$ is easy to compute.
This corresponds to applying CMC + IPA to the cdf:

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}[\mathbb{P}[X \le x \mid \mathcal{G}]] \stackrel{?}{=} \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}[X \le x \mid \mathcal{G}]\right] = \mathbb{E}[f(x \mid \mathcal{G})].$$

Assumption CDE. For all realizations of $\mathcal{G}$, $F(x \mid \mathcal{G})$ is a continuous function of x over $[a, b]$, differentiable except perhaps over a denumerable set of points $D(\mathcal{G}) \subset [a, b]$, and $f(x \mid \mathcal{G}) = F'(x \mid \mathcal{G}) = \mathrm{d}F(x \mid \mathcal{G})/\mathrm{d}x$ (when it exists) is bounded uniformly in x by a random variable Γ such that $\mathbb{E}[\Gamma^2] < \infty$.

Proposition CDE: Under Ass. CDE, for $x \in [a, b]$, $\mathbb{E}[f(x \mid \mathcal{G})] = f(x)$ and $\mathrm{Var}[f(x \mid \mathcal{G})] < \mathbb{E}[\Gamma^2]$.

The CDE $f(x \mid \mathcal{G})$ is then unbiased with uniformly bounded variance, so MISE $= \mathcal{O}(n^{-1})$.

However, finding a $\mathcal{G}$ that satisfies the CDE conditions is not always easy.

A likelihood ratio density estimator (LRDE)

This is another way to estimate the derivative of the cdf F, based on usual LR (not GLR).
Suppose $X = h(\mathbf{Y})$ where $\mathbf{Y}$ has known density f_Y over $R \subseteq \mathbb{R}^d$. We have

$$F(x) = \mathbb{P}[h(\mathbf{Y}) \leq x] = \int_R \mathbb{I}[h(\mathbf{y}) \leq x]\, f_Y(\mathbf{y}) \mathrm{d}\mathbf{y}.$$

To differentiate, we want to change the integrand into a continuous function of x.

A likelihood ratio density estimator (LRDE)

This is another way to estimate the derivative of the cdf F, based on usual LR (not GLR). Suppose $X = h(\mathbf{Y})$ where $\mathbf{Y}$ has known density f_Y over $R \subseteq \mathbb{R}^d$. We have

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

To differentiate, we want to change the integrand into a continuous function of x.

Main idea: Make 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 $\mathbb{I}[h(\mathbf{y}) \le x] = \{h(\varphi(\mathbf{z}; x)) \le x\} \equiv \{\tilde{h}(\mathbf{z}) \le 1\}$ for a function $\tilde{h}$ independent of x when $\mathbf{z}$ is given.

In case $\tilde{R} = \varphi^{-1}(R) \stackrel{\text{def}}{=} \varphi^{-1}(R; x)$ is independent of x, the integration domain remains independent of x after the change of variable, and we can rewrite

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

where $|J_\varphi(\mathbf{z}; x)|$ is the Jacobian of the transformation $\mathbf{y} = \varphi(\mathbf{z}; x)$.

Take a given $x = x_0$. In a small open neighborhood of x_0, we can write

$$F(x) = \int_{\tilde{R}} \mathbb{I}[\tilde{h}(\mathbf{z}) \leq 1] \; L(\mathbf{z}; x, x_0) \; f_Y(\varphi(\mathbf{z}; x_0))|J_\varphi(\mathbf{z}; x_0)|\mathrm{d}\mathbf{z}$$

where

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

is the likelihood ratio between the density of $\mathbf{z}$ at x and at x_0. Under appropriate conditions:

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

where

$$S(\mathbf{y}, x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln L(\mathbf{z}; x, x_0) = \left((\nabla (\ln f_Y)(\mathbf{y}))^{\mathrm{t}} \cdot \nabla_x \varphi(\mathbf{z}; x) + \frac{\mathrm{d} \ln |J_\varphi(\mathbf{z}; x)|}{\mathrm{d}x} \right) \Big|_{z=\varphi^{-1}(y;x)}$$

is the score function associated with L.

This gives the unbiased likelihood ratio density estimator (LRDE)

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

where $\mathbf{Y} \sim f_Y$. Here, $\mathbf{Y}$ can have a multivariate distribution for which conditioning is hard.

where

$$S(\mathbf{y}, x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln L(\mathbf{z}; x, x_0) = \left((\nabla(\ln f_Y)(\mathbf{y}))^{\mathrm{t}} \cdot \nabla_x \varphi(\mathbf{z}; x) + \frac{\mathrm{d} \ln |J_\varphi(\mathbf{z}; x)|}{\mathrm{d}x} \right) \Big|_{z=\varphi^{-1}(y;x)}$$

is the score function associated with L.

This gives the unbiased likelihood ratio density estimator (LRDE)

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

where $\mathbf{Y} \sim f_Y$. Here, $\mathbf{Y}$ can have a multivariate distribution for which conditioning is hard.

This LR approach was proposed long ago to estimate the derivative of $\mathbb{E}[h(\mathbf{Y})]$ with respect to a parameter of the distribution of $\mathbf{Y}$; see, e.g., Glynn (1987), L'Ecuyer (1990), and Asmussen and Glynn (2007). Here we just add a change of variable.

Assumption LRDE. For all $x \in [a, b]$, the mapping $\varphi(\cdot; x)$ has continuous partial derivatives and $|J_\varphi(\cdot; x)| \neq 0$. With probability 1 over the realizations of $\mathbf{Y} = \varphi(\mathbf{Z}; x)$, $f_Y(\varphi(\mathbf{Z}; x))|J_\varphi(\mathbf{Z}; x)|$ is a continuous function of x over $[a, b]$, and is differentiable except perhaps at a countable set of points $D(\mathbf{Y})$. There is also a random variable Γ defined over the same probability space as $\mathbf{Y}$ with $\mathbb{E}[\Gamma^2] < \infty$ and

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

Theorem LRDE. Suppose $\tilde{R} = \varphi^{-1}(R; x)$ is independent of x. Under Assumption LRDE, the LRDE is unbiased for the density $f(x)$ at x, and its variance is uniformly bounded by $\mathbb{E}[\Gamma^2]$, for all $x \in [a, b]$.

The proof is a direct application of Theorem 1 of L'Ecuyer (1990).

Simple illustration: a sum of random variables

Let $R = \mathbb{R}^d$ or $R = [0, \infty)^d$, $\mathbf{Y} = (Y_1, \ldots, Y_d)^t$ have a differentiable density f_Y over R, and $X = h(\mathbf{Y}) = Y_1 + Y_2 + \cdots + Y_d$. We want to estimate the density of X at $x \in \mathbb{R}$.

Consider first the change of variable:

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

This gives $h(\mathbf{y}) = h(x\mathbf{z}) = x h(\mathbf{z}) \leq x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) \leq 1$.
Both $\tilde{h}$ and $\varphi^{-1}(R; x) = R$ do not depend on x, and Theorem LRDE applies.

We have $|J_\varphi(\mathbf{z}; x)| = x^d$, $\nabla_x \varphi(\mathbf{z}; x) = \mathbf{z}$, $S(\mathbf{y}, x) = \left(d + (\nabla(\ln f_Y)(\mathbf{y})) \cdot \mathbf{y}\right)/x$.

Simple illustration: a sum of random variables

Let $R = \mathbb{R}^d$ or $R = [0, \infty)^d$, $\mathbf{Y} = (Y_1, \ldots, Y_d)^t$ have a differentiable density f_Y over R, and $X = h(\mathbf{Y}) = Y_1 + Y_2 + \cdots + Y_d$. We want to estimate the density of X at $x \in \mathbb{R}$.

Consider first the change of variable:

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

This gives $h(\mathbf{y}) = h(x\mathbf{z}) = x h(\mathbf{z}) \leq x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) \leq 1$.
Both $\tilde{h}$ and $\varphi^{-1}(R; x) = R$ do not depend on x, and Theorem LRDE applies.

We have $|J_\varphi(\mathbf{z}; x)| = x^d$, $\nabla_x \varphi(\mathbf{z}; x) = \mathbf{z}$, $S(\mathbf{y}, x) = \left(d + (\nabla(\ln f_Y)(\mathbf{y})) \cdot \mathbf{y}\right)/x$.

If $Y_1, \ldots, Y_d$ are independent and Y_j has density f_j, then $\ln f_Y(\mathbf{y}) = \sum_{j=1}^d \ln f_j(y_j)$ and

$$S(\mathbf{y}, x) = \frac{1}{x}\left(d + \sum_{j=1}^d y_j \frac{f_j'(y_j)}{f_j(y_j)}\right).$$

An alternative change of variable:

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

Here, $h(\mathbf{y}) = h(\varphi(\mathbf{z}; x)) = h(\mathbf{z}) + x \le x$ iff $h(\mathbf{z}) \le 0$, so we can take $\tilde{h} = h + 1$.
We have $\nabla_x \varphi(\mathbf{z}) = \mathbf{e}_j$ and $|J_\varphi(\mathbf{z}; x)| = 1$.

If $R = \mathbb{R}^d$, then $\varphi^{-1}(R; x) = R$ independent of x, Theorem LRDE applies, and we have an unbiased LRDE.

If the Y_j are independent and f_j is differentiable, then $S(\mathbf{y}, x) = f_j'(y_j)/f_j(y_j)$.
A similar estimator was obtained by Peng et al. (2020) via GLR.

An alternative change of variable:

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

Here, $h(\mathbf{y}) = h(\varphi(\mathbf{z}; x)) = h(\mathbf{z}) + x \le x$ iff $h(\mathbf{z}) \le 0$, so we can take $\tilde{h} = h + 1$.
We have $\nabla_x \varphi(\mathbf{z}) = \mathbf{e}_j$ and $|J_\varphi(\mathbf{z}; x)| = 1$.

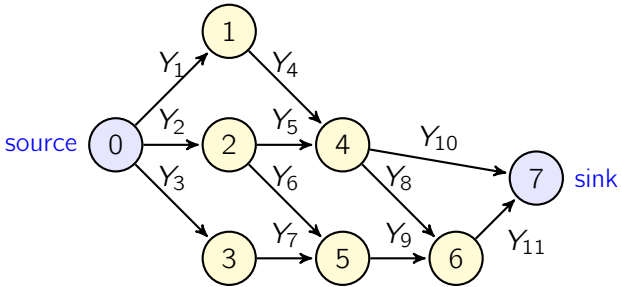
If $R = \mathbb{R}^d$, then $\varphi^{-1}(R; x) = R$ independent of x, Theorem LRDE applies, and we have an unbiased LRDE.

If the Y_j are independent and f_j is differentiable, then $S(\mathbf{y}, x) = f_j'(y_j)/f_j(y_j)$.
A similar estimator was obtained by Peng et al. (2020) via GLR.

But if $R = [0, \infty)^d$, then $\varphi^{-1}(R; x) = [-x, \infty) \times [0, \infty)^{d-1}$ depends on x, so Theorem LRDE does not apply. The LRDE is biased in this case.

Example: a stochastic activity network (SAN)

Activity k has random duration Y_k (length of arc k) with known cdf F_k.
Project duration $X =$ (random) length of longest path from source to sink.
In the network below, the are six paths, $\Pi_1 = \{1, 4, 10\}$, $\Pi_2 = \{1, 4, 8, 11\}, \ldots, \Pi_6$.
We want to estimate the density of $X = h(\mathbf{Y}) = \max_{1 \le l \le 6} \sum_{j \in \Pi_l} Y_j$.



Suppose the Y_j are independent with Weibull(α, λ) density f_j, with shape $\alpha > 0$ and scale $\lambda = 1$. For $\alpha < 1$, f_j is unbounded at 0.

LRDE for the SAN

Let $R = (0, \infty)^{11}$, fix $x > 0$, and take $\mathbf{y} = \varphi(\mathbf{z}; x) = x\mathbf{z}$.
Then $h(\mathbf{y}) = h(\varphi(\mathbf{z}; x)) = xh(\mathbf{z}) \leq x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) \leq 1$.
Here, $\tilde{R}(x) = \varphi^{-1}(R; x) = (0, \infty)^{11}$ is independent of x,

Since $f_j'(y)/f_j(y) = (\alpha - 1 - \alpha y^\alpha)/y$, we get

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

and $\mathbb{E}[\hat{f}(x)] = f(x)$ for any $x > 0$ and all $\alpha > 0$.

LRDE for the SAN

Let $R = (0, \infty)^{11}$, fix $x > 0$, and take $\mathbf{y} = \varphi(\mathbf{z}; x) = x\mathbf{z}$.
Then $h(\mathbf{y}) = h(\varphi(\mathbf{z}; x)) = xh(\mathbf{z}) \leq x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) \leq 1$.
Here, $\tilde{R}(x) = \varphi^{-1}(R; x) = (0, \infty)^{11}$ is independent of x,

Since $f_j'(y)/f_j(y) = (\alpha - 1 - \alpha y^\alpha)/y$, we get

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

and $\mathbb{E}[\hat{f}(x)] = f(x)$ for any $x > 0$ and all $\alpha > 0$.

Note that $\mathbf{y} = \varphi(\mathbf{z}; x) = \mathbf{z} + x\,\mathbf{e}_j$ is not a proper change of variable here (not one-to-one).

The LRDE with boundary terms

What if $\tilde{R}(x) = \varphi^{-1}(R; x)$ depends on x?

Let $\partial\tilde{R}(x)$ be the boundary of $\tilde{R}(x)$, and $\mathbf{b}(\mathbf{z}(x), x)$ the rate of displacement of $\partial\tilde{R}(x)$ as a function of x at $\mathbf{z}(x) \in \partial\tilde{R}(x)$, in the normal direction pointing outward of $\tilde{R}(x)$. We have

$$
\begin{aligned}
F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\int_{\tilde{R}(x)}[\cdots]\mathrm{d}\mathbf{z} &= \int_{\tilde{R}(x)}\frac{\mathrm{d}}{\mathrm{d}x}[\cdots]\mathrm{d}\mathbf{z} + \int_{\partial\tilde{R}(x)}[\cdots]\mathbf{b}(\mathbf{z}(x), x)\mathrm{d}\mathbf{z} \\
&= \int_{\tilde{R}(x)}\frac{\mathrm{d}}{\mathrm{d}x}[\cdots]\mathrm{d}\mathbf{z} + \int_{\partial R}\mathbb{I}[h(\mathbf{y}) \leq x]\mathbf{b}(\varphi^{-1}(\mathbf{y}, x), x) f_Y(\mathbf{y})\mathrm{d}\mathbf{y}.
\end{aligned}
$$

The LRDE with boundary terms

What if $\tilde{R}(x) = \varphi^{-1}(R; x)$ depends on x?

Let $\partial\tilde{R}(x)$ be the boundary of $\tilde{R}(x)$, and $\mathbf{b}(\mathbf{z}(x), x)$ the rate of displacement of $\partial\tilde{R}(x)$ as a function of x at $\mathbf{z}(x) \in \partial\tilde{R}(x)$, in the normal direction pointing outward of $\tilde{R}(x)$. We have

$$\begin{aligned} F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\int_{\tilde{R}(x)}[\cdots]\mathrm{d}\mathbf{z} &= \int_{\tilde{R}(x)}\frac{\mathrm{d}}{\mathrm{d}x}[\cdots]\mathrm{d}\mathbf{z} + \int_{\partial\tilde{R}(x)}[\cdots]\mathbf{b}(\mathbf{z}(x), x)\mathrm{d}\mathbf{z} \\ &= \int_{\tilde{R}(x)}\frac{\mathrm{d}}{\mathrm{d}x}[\cdots]\mathrm{d}\mathbf{z} + \int_{\partial R}\mathbb{I}[h(\mathbf{y}) \le x]\mathbf{b}(\varphi^{-1}(\mathbf{y}, x), x) f_Y(\mathbf{y})\mathrm{d}\mathbf{y}. \end{aligned}$$

Common special case: $R = \prod_{j=1}^{d}(\alpha_j, \beta_j)$, a rectangular box.

Let ∂R_j^- and ∂R_j^+ denote the boundary panels on which $y_j = \alpha_j$ and $y_j = \beta_j$, respectively.

The rate of outward displacement at $\mathbf{y} \in \partial R_j^+$ $[\mathbf{y} \in \partial R_j^-]$ is

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

Let $R_{-j} = \prod_{k \neq j}(\alpha_k, \beta_k)$, $\mathbf{y}_{-j}$ be $\mathbf{y}$ with its jth coordinate removed, and $\mathbf{Y}_{-j}^{-}$ [$\mathbf{Y}_{-j}^{+}$] be $\mathbf{Y}$ with Y_j replaced by α_j [β_j]. We then have

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

and similarly for ∂R_j^- with α_j and $\mathbf{Y}_{-j}^{-}$.

By summing over all boundary panels, we get
$\int_{\partial R} \mathbb{I}[h(\mathbf{y}) \leq x]\mathbf{b}(\varphi^{-1}(\mathbf{y}, x), x) f_Y(\mathbf{y}) \mathrm{d}\mathbf{y} = \mathbb{E}[B(\mathbf{Y}, x)]$ where

$$B(\mathbf{y}, x) = \sum_{j=1}^{d} \left(\mathbb{I}[h(\mathbf{y}_{-j}^{+}) \leq x] r_j(\mathbf{y}_{-j}^{+}, x) f_{Y_j|Y_{-j}}(\beta_j \mid \mathbf{y}_{-j}) - \mathbb{I}[h(\mathbf{y}_{-j}^{-}) \leq x] r_j(\mathbf{y}_{-j}^{-}, x) f_{Y_j|Y_{-j}}(\alpha_j \mid \mathbf{y}_{-j}) \right).$$

The LRDE with boundary terms

Assumption LRDE-B. Let Assumption LRDE hold. Let $R = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ and suppose f_Y is continuous on the closure of R. For each $\mathbf{y} \in R$, $\varphi^{-1}(\mathbf{y}; x)$ is differentiable in x on $[a, b]$, and each term in $B(\mathbf{y}, x)$ is well-defined, w.p.1.

Theorem LRDE-B. Under Assumption LRDE-B,

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

is and unbiased estimator for $f(x)$ for all $x \in [a, b]$.

Simple illustration with a sum of random variables.

Let $X = Y_1 + \cdots + Y_d = \mathbf{1} \cdot \mathbf{Y}$. We saw that if $\varphi(\mathbf{z}; x) = \mathbf{z} + x\,\mathbf{e}_j$ and $R = [0, \infty)^d$, then $\varphi^{-1}(R; x) = [-x, \infty) \times [0, \infty)^{d-1}$. That is, the left boundary panel ∂R_j^- depends on x. We have $r_j(\mathbf{y}, x) = -1$ for $\mathbf{y} \in \partial R_j^-$ and the other r_k's are zero.

Suppose the Y_j are independent exponential with rate λ_j.
Theorem LRDE-B applies and we obtain the following unbiased estimator for $f(x)$:

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

Alternate density estimator for the SAN. Select a directed minimal cut $\mathcal{C}$, such as $\mathcal{C}_1 = \{10, 11\}$, $\mathcal{C}_2 = \{1, 2, 3\}$, $\mathcal{C}_3 = \{3, 4, 5, 6\}$, etc. The change of variable $\mathbf{y} = \varphi_{\mathcal{C}}(\mathbf{z}; x) = \mathbf{z} + x \sum_{j \in \mathcal{C}} \mathbf{e}_j$ satisfies $h(\mathbf{y}) = h(\mathbf{z}) + x \leq x$ iff $\tilde{h}(\mathbf{z}) \stackrel{\text{def}}{=} h(\mathbf{z}) + 1 \leq 1$.

This gives $\nabla_x \varphi_{\mathcal{C}}(\mathbf{z}; x) = \sum_{j \in \mathcal{C}} \mathbf{e}_j$ and $|J_\varphi| = 1$, so $S(\mathbf{y}; x) = \sum_{j \in \mathcal{C}} f_j'(y_j)/f_j(y_j)$. Here, $\varphi_{\mathcal{C}}^{-1}(\mathbf{y}; x) = \mathbf{y} - x \sum_{j \in \mathcal{C}} \mathbf{e}_j$, so $\tilde{R}(x) = \varphi_{\mathcal{C}}^{-1}(R; x)$ depends on x in the dimensions $j \in \mathcal{C}$.

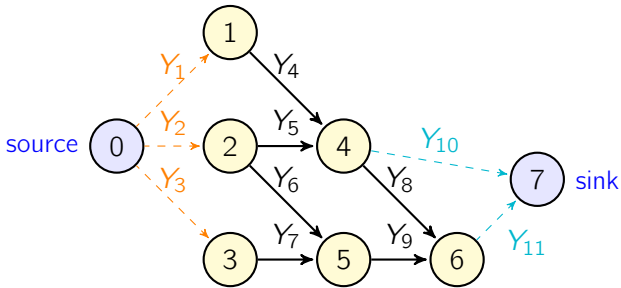
We have $r_j(\mathbf{y}, x) = -1$ for $j \in \mathcal{C}$ and $\mathbf{y} \in \partial R_j^-$, $r_j(\mathbf{y}, x) = 0$ otherwise.

The LRDE-B for an arbitrary directed minimal cut $\mathcal{C}$ and Weibull Y_j is then

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

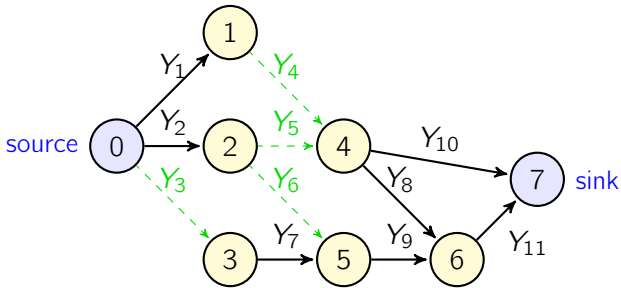
For $\alpha \geq 1$ and $x > 0$, this estimator is unbiased by Theorem LRDE-B. For $\alpha < 1$, the terms $f_j(0)$ are infinite and Theorem LRDE-B does not apply.

Minimal directed cuts in SAN



$\mathcal{C}_1 = \{10, 11\}$, $\mathcal{C}_2 = \{1, 2, 3\}$

Minimal directed cut in SAN



$\mathcal{C}_3 = \{3, 4, 5, 6\}$

Numerical illustration

Integrated variance (IV) of density estimator over $[a, b]$ which contains the central 95% of the density of X. Each Y_j is Weibull($\alpha, \lambda = 1$).

The LRDE $\hat{f}$ uses the change of variable $\mathbf{y} = \varphi(\mathbf{z}; x) = x\mathbf{z}$. This is the winner here.

Each LRDE-B $\hat{f}_{\mathcal{C}}$ uses $\varphi_{\mathcal{C}}(\mathbf{z}; x) = \mathbf{z} + x \sum_{j \in \mathcal{C}} \mathbf{e}_j$.
The variance tends to be larger when there are more arcs in the cut.

Estimator	$\mathcal{C}$	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 3.0$
LRDE $\hat{f}$		2.28E-7	9.08E-7	1.73E-6	2.62E-6	4.48E-6
LRDE-B $\hat{f}_{\mathcal{C}}$	$\{10, 11\}$	–	4.31E-6	1.65E-3	7.57E-5	2.43E-5
LRDE-B $\hat{f}_{\mathcal{C}}$	$\{1, 2, 3\}$	–	9.69E-6	8.63E-4	7.88E-5	3.52E-5
LRDE-B $\hat{f}_{\mathcal{C}}$	$\{3, 4, 5, 6\}$	–	1.72E-5	2.49E-3	1.25E-4	4.72E-5

GLR density estimators

Peng et al. (2020) proposed an adaptation of a generalized likelihood ratio (GLR) method of Peng et al. (2018) to density estimation. Peng et al. (2021) give an improved version.

The estimators are often similar to our LRDEs, but the assumptions are stronger and harder to verify.

GLR density estimators

Peng et al. (2020) proposed an adaptation of a generalized likelihood ratio (GLR) method of Peng et al. (2018) to density estimation. Peng et al. (2021) give an improved version.

The estimators are often similar to our LRDEs, but the assumptions are stronger and harder to verify.

The cdf of X is assumed to have the form

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

where $g : \mathbb{R}^d \times [a, b] \to \mathbb{R}^d$ is smooth and $\psi : \mathbb{R}^d \to \mathbb{R}$ is independent of x but can be discontinuous.

Our setting fits this framework by taking $g = \varphi^{-1}$ and $\psi(\mathbf{z}) = \mathbb{I}[\tilde{h}(\mathbf{z}) < 1]$.

Original GLRDE form Peng et al. (2018): $\hat{f}(x) = \mathbb{I}[h(\mathbf{Y}) \leq x] S(\mathbf{Y}, x)$ with

$$
\begin{aligned}
S(\mathbf{y}, x) \;=\; & -\operatorname{trace}\left(J_g^{-1}(\mathbf{y}, x) \nabla_x J_g(\mathbf{y}, x)\right) + \sum_{j=1}^{d} \mathbf{e}_j^{\mathrm{t}} J_g^{-1}(\mathbf{y}, x) \left(\nabla_{y_j} J_g\right) J_g^{-1}(\mathbf{y}, x) \nabla_x g(\mathbf{y}, x) \\
& - \left(J_g^{-1}(\mathbf{y}, x) \nabla_x g(\mathbf{y}, x)\right)^{\mathrm{t}} \nabla \ln f_Y(\mathbf{y}).
\end{aligned}
$$

Assumptions GLRDE.

1. f_Y is strictly positive and continuously differentiable over all of $\mathbb{R}^d$.
2. $g(\mathbf{y}, x)$ is invertible in $\mathbf{y}$, and g is twice continuously differentiable on $\mathbb{R}^d \times [a, b]$.
3. $\lim_{y_j \to \pm\infty} \int_{\mathbb{R}^{d-1}} \sup_{x \in [a,b]} |r_j(\mathbf{y}; x)| f_Y(\mathbf{y}) \mathrm{d}\mathbf{y}_{-j} = 0$ for all $1 \leq j \leq d$.
4. $\int_{\mathbb{R}^d} \sup_{x \in [a,b]} |\hat{f}(x)| \mathrm{d}\mathbf{y} < \infty$.

Our paper gives simple examples in which these assumptions do not hold whereas ours hold.

GLRDE with uniform inputs

Peng et al. (2021) proposed a variant of GLR in which $\mathbf{Y} = \mathbf{U}$ is uniform over $(0,1)^d$. For density estimation, this turns out to be a special case of our LRDE-B setting with f_Y uniform over $R = (\boldsymbol{\alpha}, \boldsymbol{\beta}) = (0,1)^d$.

The proposed GLRDE is similar to our LRDE-B:

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

where

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

There is a large set of assumptions that are not so easy to verify.

Conclusion

- The LR derivative estimation method popularized some 35 years ago can provide an unbiased density estimator, by using a simple change of variable.
- Sometimes, the boundaries of the integral that we have to differentiate w.r.t. x depend on x. Then we need to differentiate also with respect to the boundaries (Leibniz rule) and this adds an extra term.
- Extra: One can also apply conditional Monte Carlo to these LRDE estimators to make them smoother, and then combine them with RQMC. In separate experiments, we found that this can provide large variance reductions.

Some relevant references

- S. Asmussen. Conditional Monte Carlo for sums, with applications to insurance and finance, Annals of Actuarial Science, 12, 2: 455-478, 2018.
- P. W. Glynn. Likelihood ratio gradient estimation: an overview. Proceedings of the 1987 Winter Simulation Conference, 366–375.
- P. J. Laub, R. Salomone, Monte Carlo estimation of the density of the sum of dependent random variables. Mathematics and Computers in Simulation 161: 23–31, 2019.
- P. L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. Management Science 36: 1364–1383, 1990.
- P. L'Ecuyer and F. Puchhammer, "Density Estimation by Monte Carlo and Quasi-Monte Carlo," in Monte Carlo and Quasi Monte Carlo Methods: MCQMC 2020, A. Keller, Ed., Springer-Verlag, 3–21, 2022.
- P. L'Ecuyer, F. Puchhammer, and A. Ben Abdellah. Monte Carlo and Quasi-Monte Carlo Density Estimation via Conditioning. INFORMS Journal on Computing, 34 (3), 1729–1748, 2022.
- Y. Peng, M. C. Fu, B. Heidergott, H. Lam. Maximum likelihood estimation by Monte Carlo simulation: Towards data-driven stochastic modeling. Operations Research, 68, 1896–1912, 2020.
- Y. Peng, M. C. Fu, J. Q. Hu, B. Heidergott. A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. Operations Res. 66, 487–499, 2018.
- Y. Peng, M. C. Fu, J.-Q. Hu, P. L'Ecuyer, and B. Tuffin, "A Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs," submitted, revised in July 2022.
- F. Puchhammer and P. L'Ecuyer, "Likelihood Ratio Density Estimation for Simulation Models," Proceedings of the 2022 Winter Simulation Conference, 2022. (Published paper for this talk.)