Density estimation by Monte Carlo and quasi-Monte Carlo

Pierre L'Ecuyer

Joint work with **Amal Ben Abdellah and Florian Puchhammer**

Université de Montréal

GERAD
GROUP FOR RESEARCH IN DECISION ANALYSIS

CIRRELT

ICIAM 2023, Tokyo

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 usually provides information to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or 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 usually provides information to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or 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.
Most popular methods: histogram: $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-2/3})$;
kernel density estimator (KDE); $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-4/5})$. **Biased**.

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 usually provides information to estimate the entire distribution of X, e.g., the cumulative distribution function (cdf) $F(x) = \mathbb{P}[X \leq x]$, or 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.
Most popular methods: histogram: $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-2/3})$;
kernel density estimator (KDE); $\mathrm{MSE}[\hat{f}_n(x)] = \mathcal{O}(n^{-4/5})$. **Biased**.

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.
Under certain conditions, **unbiased** density estimators can be obtained and can also be combined with randomized quasi-Monte Carlo (RQMC).

This talk is mostly a review of recent papers listed on the last slide.

Density estimation

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

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.

Classical density estimators

For given observations $X_1, \ldots, X_n$:

Histogram: Partition $[a, b]$ in m intervals of size $h = (b - a)/m$ and define

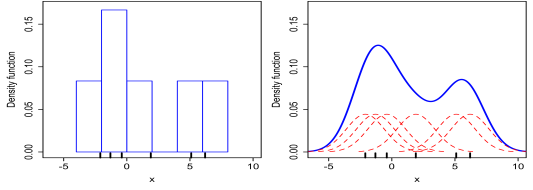
$$\hat{f}_n(x) = \frac{n_j}{nh} \quad \text{for } x \in I_j = [a + (j-1)h, a + jh), \quad j = 1, ..., m$$

where n_j is the number of observations X_i that fall in interval j.
MISE $= \mathcal{O}(n^{-2/3})$ in the best case.

Kernel Density Estimator (KDE) : Select kernel k (unimodal symmetric density centered at 0) and bandwidth $h > 0$ (horizontal stretching factor for the kernel).

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right).$$

MISE $= \mathcal{O}(n^{-4/5})$ in the best case.



Why not take the stochastic derivative of an estimator of F?

An unbiased estimator of the cdf F is given by the empirical cdf

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[X_i \leq x].$$

Can we estimate the density $f(x) = F'(x)$ by taking the derivative of an estimator of $F(x)$?

Not with the empirical cdf, because $\mathrm{d}\hat{F}_n(x)/\mathrm{d}x = 0$ almost everywhere.

One way to get a **smoother estimator of** F: replace $\mathbb{I}[X_i \leq x]$ by a conditional probability which is continuous in x and differentiable in x except perhaps at a denumerable set of points. Then take its derivative (a conditional density) as a density estimator.

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 $X \mid \mathcal{G}$ 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.

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

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 $X \mid \mathcal{G}$ 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.

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

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

The CDE is then unbiased with uniformly bounded variance, so $\mathrm{MISE}[\hat{f}_n] = \mathcal{O}(n^{-1})$.

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 $X \mid \mathcal{G}$ 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.

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

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

The CDE is then unbiased with uniformly bounded variance, so $\mathrm{MISE}[\hat{f}_n] = \mathcal{O}(n^{-1})$.

Finding a $\mathcal{G}$ that satisfies the CDE conditions is not always easy, but often doable.

A likelihood ratio density estimator (LRDE)

Another way to construct a smooth estimator of the cdf F, that we can differentiate.
Suppose $X = h(\boldsymbol{Y})$ where $\boldsymbol{Y}$ has known density $f_{\boldsymbol{Y}}$ over $R \subseteq \mathbb{R}^d$. We have

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

To differentiate w.r.t. x, we want to make the indicator part independent of x.

A likelihood ratio density estimator (LRDE)

Another way to construct a smooth estimator of the cdf F, that we can differentiate.
Suppose $X = h(\boldsymbol{Y})$ where $\boldsymbol{Y}$ has known density $f_{\boldsymbol{Y}}$ over $R \subseteq \mathbb{R}^d$. We have

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

To differentiate w.r.t. x, we want to make the indicator part independent of x.

Main idea: Move the x out of the indicator. Make a change of variable of the form $\boldsymbol{y} = \varphi(\boldsymbol{z}; x)$, for a family of bijections $\{\varphi(\cdot; x),\ x \in [a, b]\}$ such that $\{h(\boldsymbol{y}) \leq x\} = \{h(\varphi(\boldsymbol{z}; x)) \leq x\} \equiv \{\tilde{h}(\boldsymbol{z}) \leq 1\}$ where the function $\tilde{h}$ is independent of x when $\boldsymbol{z}$ is given.

A likelihood ratio density estimator (LRDE)

Another way to construct a smooth estimator of the cdf F, that we can differentiate.
Suppose $X = h(\boldsymbol{Y})$ where $\boldsymbol{Y}$ has known density $f_{\boldsymbol{Y}}$ over $R \subseteq \mathbb{R}^d$. We have

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

To differentiate w.r.t. x, we want to make the indicator part independent of x.

Main idea: Move the x out of the indicator. Make a change of variable of the form $\boldsymbol{y} = \varphi(\boldsymbol{z}; x)$, for a family of bijections $\{\varphi(\cdot; x),\ x \in [a, b]\}$ such that $\{h(\boldsymbol{y}) \leq x\} = \{h(\varphi(\boldsymbol{z}; x)) \leq x\} \equiv \{\tilde{h}(\boldsymbol{z}) \leq 1\}$ where the function $\tilde{h}$ is independent of x when $\boldsymbol{z}$ is given. We can then write

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

where $|J_\varphi(\boldsymbol{z}; x)|$ is the Jacobian of the transformation $\boldsymbol{y} = \varphi(\boldsymbol{z}; x)$, assuming that the new integration domain $\tilde{R} = \varphi^{-1}(R; x)$ is independent of x.

For x in a small open neighborhood of a given x_0, we can write

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

where

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

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

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

where

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

is the score function associated with L.

This gives the unbiased **likelihood ratio density estimator (LRDE)**

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

where $\boldsymbol{Y} \sim f_{\boldsymbol{Y}}$.

where

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

is the score function associated with L.

This gives the unbiased **likelihood ratio density estimator (LRDE)**

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

where $\boldsymbol{Y} \sim f_{\boldsymbol{Y}}$.

This LR approach is often used to estimate the derivative of $\mathbb{E}[h(\boldsymbol{Y})]$ with respect to a parameter of the distribution of $\boldsymbol{Y}$; see, e.g., Glynn (1987), L'Ecuyer (1990), or Asmussen and Glynn (2007). Here we add a change of variable to use it for density estimation.

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 $\boldsymbol{Y} = \varphi(\boldsymbol{Z}; x)$, $f_{\boldsymbol{Y}}(\varphi(\boldsymbol{Z}; x))|J_\varphi(\boldsymbol{Z}; x)|$ is a continuous function of x over $[a, b]$, and is differentiable except perhaps at a countable set of points $D(\boldsymbol{Y})$. There is also a random variable Γ defined over the same probability space as $\boldsymbol{Y}$ with $\mathbb{E}[\Gamma^2] < \infty$ and

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

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 $x \in [a, b]$. This LRDE is then unbiased with uniformly bounded variance, so again $\text{MISE}[\hat{f}_n] = \mathcal{O}(n^{-1})$.

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

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

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

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

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 $\boldsymbol{y} \in \partial R_j^+$ $[\boldsymbol{y} \in \partial R_j^-]$ is

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

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

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

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

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

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

The LRDE with boundary terms

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

Theorem LRDE-B. Under Assumption LRDE-B,

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

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

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 made in those papers are generally 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 made in those papers are generally stronger and harder to verify.

The cdf of X is assumed to have the form

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

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.

Quasi-Monte Carlo Error and Variance Bounds

Let $h(\boldsymbol{y}) = g(\boldsymbol{u})$ where $g : [0,1)^s \to \mathbb{R}$.
Integration error with point set $P_n = \{\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{n-1}\} \subset [0,1)^s$:

$$E_n = \frac{1}{n}\sum_{i=0}^{n-1} g(\boldsymbol{u}_i) - \int_{[0,1)^s} g(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\,.$$

For certain classes of sufficiently smooth functions g, there are explicit RQMC constructions for which $\mathbb{E}[E_n] = 0$ and $\mathrm{Var}[E_n] = \mathcal{O}(n^{-2+\epsilon}), \quad \forall \epsilon > 0$. Examples:
— randomly-shifted lattice rules;
— digital nets with left matrix scramble (LMS) + random digital shift.

With ordinary Monte Carlo (MC), one has $\mathrm{Var}[E_n] = \mathcal{O}(n^{-1})$.

Bounding the AIV under RQMC for a KDE

KDE density estimator at a single point x:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - g(\boldsymbol{U}_i)}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(\boldsymbol{U}_i).$$

With RQMC points $\boldsymbol{U}_i$, this is an RQMC estimator of $\mathbb{E}[\tilde{g}(\boldsymbol{U})] = \int_{[0,1)^s} \tilde{g}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mathbb{E}[\hat{f}_n(x)]$.

RQMC does not change the bias, but may reduce $\mathrm{Var}[\hat{f}_n(x)]$, and then the IV.

To get RQMC variance bounds, we need bounds on the variation of $\tilde{g}$.

Bounding the AIV under RQMC for a KDE

KDE density estimator at a single point x:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - g(\boldsymbol{U}_i)}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(\boldsymbol{U}_i).$$

With RQMC points $\boldsymbol{U}_i$, this is an RQMC estimator of $\mathbb{E}[\tilde{g}(\boldsymbol{U})] = \int_{[0,1)^s} \tilde{g}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mathbb{E}[\hat{f}_n(x)]$.

RQMC does not change the bias, but may reduce $\mathrm{Var}[\hat{f}_n(x)]$, and then the IV.

To get RQMC variance bounds, we need bounds on the variation of $\tilde{g}$.
Partial derivatives:

$$\frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} \tilde{g}(\boldsymbol{u}) = \frac{1}{h} \frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} k\left(\frac{x - g(\boldsymbol{u})}{h}\right).$$

We assume they exist and are uniformly bounded. E.g., Gaussian kernel k.

Bounding the AIV under RQMC for a KDE

KDE density estimator at a single point x:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - g(\boldsymbol{U}_i)}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(\boldsymbol{U}_i).$$

With RQMC points $\boldsymbol{U}_i$, this is an RQMC estimator of $\mathbb{E}[\tilde{g}(\boldsymbol{U})] = \int_{[0,1)^s} \tilde{g}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mathbb{E}[\hat{f}_n(x)]$.

RQMC does not change the bias, but may reduce $\mathrm{Var}[\hat{f}_n(x)]$, and then the IV.

To get RQMC variance bounds, we need bounds on the variation of $\tilde{g}$.
Partial derivatives:

$$\frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} \tilde{g}(\boldsymbol{u}) = \frac{1}{h}\frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} k\left(\frac{x - g(\boldsymbol{u})}{h}\right).$$

We assume they exist and are uniformly bounded. E.g., Gaussian kernel k.
By expanding via the chain rule, we obtain terms in h^{-j} for $j = 2, \ldots, |\mathfrak{v}| + 1$.
The term for $\mathfrak{v} = \mathcal{S}$ grows as $h^{-s-1}k^{(s)}\left((g(\boldsymbol{u}) - x)/h\right)\prod_{j=1}^{s} g_j(\boldsymbol{u}) = \mathcal{O}(h^{-s-1})$ when $h \to 0$.

An AIV upper bound proved in Ben Abdellah et al. (2021)

Assumptions. Let $g : [0,1]^s \to \mathbb{R}$ be piecewise monotone in each coordinate u_j when the other coordinates are fixed, with at most M_j pieces. Assume that all first-order partial derivatives of g are continuous and that $\|g_{\mathfrak{w}_1} g_{\mathfrak{w}_2} \cdots g_{\mathfrak{w}_\ell}\|_1 < \infty$ for all selections of non-empty, mutually disjoint index sets $\mathfrak{w}_1, \ldots, \mathfrak{w}_\ell \subseteq \mathcal{S} = \{1, \ldots, s\}$.

For each $j \in \mathcal{S}$, let $G_j = \left\| \prod_{\ell \in \mathcal{S} \setminus \{j\}} g_{\{\ell\}} \right\|_1$ and $c_j = M_j \, \|k^{(s)}\|_\infty \, (G_j + \mathbb{I}(s=2) \, \|g_{\{1,2\}}\|_1)$.

Proposition Then the Hardy-Krause variation of $\tilde{g}$ satisfies

$$V_{\mathrm{HK}}(\tilde{g}) \le c_j h^{-s} + \mathcal{O}(h^{-s+1}) \qquad \text{for each } j.$$

Corollary. With RQMC point sets having $D^*(P_n) = \mathcal{O}(n^{-1+\epsilon})$ for all $\epsilon > 0$ when $n \to \infty$, using KH and squaring gives the bound

$$\mathrm{AIV} = \mathcal{O}(n^{-2+\epsilon} h^{-2s}) \qquad \text{for all } \epsilon > 0.$$

By picking h to minimize the AMISE bound, we obtain $\mathrm{AMISE} = \mathcal{O}(n^{-4/(2+s)+\epsilon})$.

Worst than MC when $s \ge 3$. The factor h^{-2s} hurts, although this is only an upper bound.

Combining RQMC with the CDE or the LRDE

In these cases, we are applying RQMC to unbiased estimators.
No banwidth h to select, much more RQMC-friendly.

When the CDE or LRDE estimator is sufficiently smooth as a function of the underlying uniform random numbers, RQMC can improve the convergence rate beyond $\mathcal{O}(1/n)$.

Example: displacement of cantilever beam (Bingham 2017)

$$X = h(Y_1, Y_2, Y_3) = \frac{\kappa}{Y_1}\sqrt{\frac{{Y_2}^2}{w^4} + \frac{{Y_3}^2}{t^4}}$$

where $\kappa = 5 \times 10^5$, $w = 4$, $t = 2$, Y_1, Y_2, Y_3 independent normal, $Y_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$,

Description	Symbol	μ_j	σ_j
Young's modulus	Y_1	2.9×10^7	1.45×10^6
Horizontal load	Y_2	500	100
Vertical load	Y_3	1000	100

We estimate the density of X over [3.1707, 5.6675], which covers about 99% of the density (it clips 0.5% on each side).

CDE estimator for cantilever beam

Conditioning on $\mathcal{G}_{-1} = \{Y_2, Y_3\}$ means hiding Y_1. We have

$$X = \frac{\kappa}{Y_1}\sqrt{\frac{Y_2^2}{w^4} + \frac{Y_3^2}{t^4}} \leq x \quad \text{if and only if} \quad Y_1 \geq \frac{\kappa}{x}\sqrt{\frac{Y_2^2}{w^4} + \frac{Y_3^2}{t^4}} \stackrel{\text{def}}{=} W_1(x).$$

For $x > 0$,

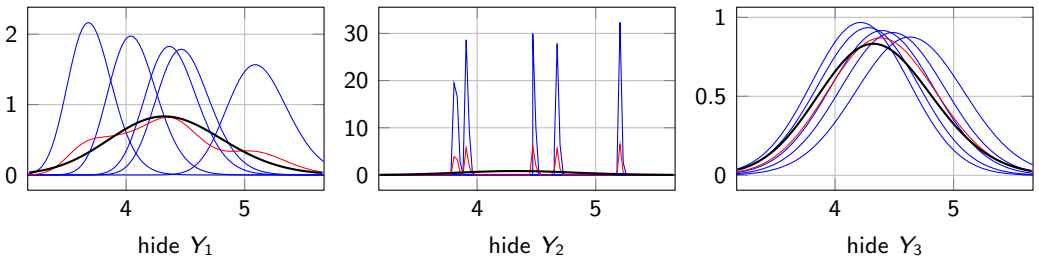
$$F(x \mid \mathcal{G}_{-1}) = \mathbb{P}[Y_1 \geq W_1(x) \mid W_1(x)] = 1 - \Phi((W_1(x) - \mu_1)/\sigma_1)$$

and

$$f(x \mid \mathcal{G}_{-1}) = F'(x \mid \mathcal{G}_{-1}) = -\frac{\phi((W_1(x) - \mu_1)/\sigma_1)W_1'(x)}{\sigma_1} = \frac{\phi((W_1(x) - \mu_1)/\sigma_1)W_1(x)}{x\sigma_1}.$$

The MISE rate improves from $\mathcal{O}(n^{-4/5})$ to $\mathcal{O}(n^{-1})$ with the CDE.

We can also hide Y_2 or Y_3 instead of Y_1.



Five realizations of conditional density $f(\cdot \mid \mathcal{G}_{-k})$,
their average (the estimator), and **true density**.
Hiding Y_3 gives the best CDE estimator.

We can also take a linear combination of the three estimators (CDE-c).

LRDE estimator: we can use for example $Y_1 = \varphi(Z_1; x) = Z_1/x$ and $(Y_2, Y_3) = (Z_2, Z_3)$.

$$X = \frac{\kappa}{Y_1}\sqrt{\frac{Y_2^2}{w^4} + \frac{Y_3^2}{t^4}} \leq x \quad \text{if and only if} \quad \frac{\kappa}{Z_1}\sqrt{\frac{Y_2^2}{w^4} + \frac{Y_3^2}{t^4}} \leq 1.$$

Cantilever beam

Estimated MISE $= Kn^{-\hat{\nu}}$. MISE with $n = 2^{19}$ is $2^{-\mathrm{e19}}$.

	$\hat{\nu}$				e19			
	KDE	$\mathcal{G}_{-3}$	CDE-c	LRDE	KDE	$\mathcal{G}_{-3}$	CDE-c	LRDE
MC	0.76	0.99	0.98	1.03	15.8	22.8	22.5	16.8
Lat+s	1.03	2.06	2.04	1.55	21.9	41.6	41.9	26.4
Lat+s+b	0.93	2.27	2.25	1.25	21.0	46.8	47.0	24.7
Sob+LMS	0.97	2.21	2.21	1.31	21.5	45.7	46.1	25.6

The MISE decreases roughly as $\mathcal{O}(n^{-2})$ or better for CDE+RQMC.

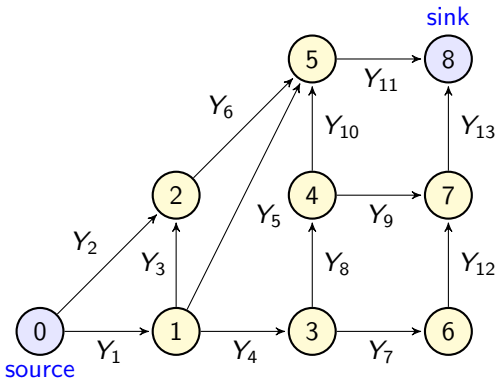
For $n = 2^{19}$, the MISE is about $\mathbf{2^{-15.8}}$ for KDE+MC and $\mathbf{2^{-47}}$ for CDE+RQMC; i.e., MISE is divided by more than $2^{31} \approx$ **2 billions**.

Example: a stochastic activity network (SAN)

Precedence relations between activities. Activity k has random duration Y_k (length of arc k) with known cdf $F_k(y) := \mathbb{P}[Y_k \leq y]$.

Project duration $X =$ (random) length of longest path from source to sink.

Specific case (Avramidis and Wilson 1998):
$Y_k \sim N(\mu_k, \sigma_k^2)$ for $k = 1, 2, 4, 11, 12$;
$Y_k \sim \text{Expon}(1/\mu_k)$ otherwise;
$\mu_1, \ldots, \mu_{13}$: 13.0, 5.5, 7.0, 5.2, 16.5, 14.7, 10.3, 6.0, 4.0, 20.0, 3.2, 3.2, 16.5;
$\sigma_k = \mu_k/4$.



CMC for the SAN Example

X = length of longest path

Hiding a single Y_j does not work: it does not make the conditional cdf of X continuous.

CMC for the SAN Example

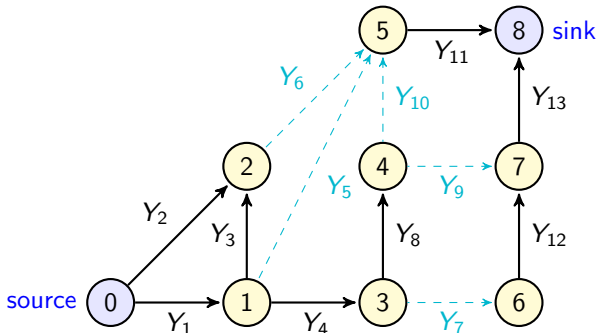
X = length of longest path

Hiding a single Y_j does not work: it does not make the conditional cdf of X continuous.

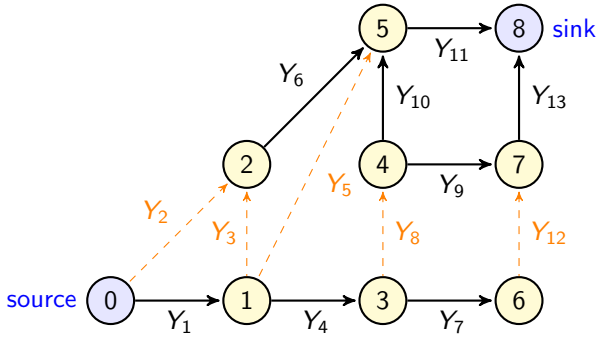
One working solution:

Pick a minimal cut $\mathcal{L}$ between source and sink, and let $F(x \mid \mathcal{G}) = \mathbb{P}[X \le x \mid \{Y_j, j \notin \mathcal{L}\}]$.

Ex.: $\mathcal{L} = \{5, 6, 7, 9, 10\}$ and $Y_j = F_j^{-1}(U_j)$. This estimator continuous in the U_j's and in x.



Another minimal cut: $\mathcal{L} = \{2, 3, 5, 8, 12\}$.



For each $j \in \mathcal{L}$, let P_j be the length of the longest path that goes through arc j when we exclude Y_j from that length. Then

$$F(x \mid \mathcal{G}) = \mathbb{P}[X < x \mid \{Y_j : j \notin \mathcal{L}\}] = \prod_{j \in \mathcal{L}} F_j(x - P_j)$$

and

$$f(x \mid \mathcal{G}) = \sum_{j \in \mathcal{L}} f_j(x - P_j) \prod_{l \in \mathcal{L}, l \neq j} F_l(x - P_j),$$

if f_j exists for all $j \in \mathcal{L}$.

Under this conditioning, the cdf of every path length is continuous in x, and so is $F(\cdot \mid \mathcal{G})$, and Assumption 1 holds, so $f(x \mid \mathcal{G})$ is an unbiased density estimator.

LRDE for the SAN

One possible change of variable (LRDE-1): $\boldsymbol{y} = \varphi(\boldsymbol{z}; x) = x\boldsymbol{z}$.

Then $h(\boldsymbol{y}) = h(\varphi(\boldsymbol{z}; x)) = xh(\boldsymbol{z}) \leq x$ iff $\tilde{h}(\boldsymbol{z}) \stackrel{\text{def}}{=} h(\boldsymbol{z}) \leq 1$.
Here, $\varphi^{-1}(R; x) = (0, \infty)^{11}$ is independent of x and this gives the unbiased estimator

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

LRDE for the SAN

One possible change of variable (LRDE-1): $\boldsymbol{y} = \varphi(\boldsymbol{z}; x) = x\boldsymbol{z}$.

Then $h(\boldsymbol{y}) = h(\varphi(\boldsymbol{z}; x)) = xh(\boldsymbol{z}) \leq x$ iff $\tilde{h}(\boldsymbol{z}) \stackrel{\text{def}}{=} h(\boldsymbol{z}) \leq 1$.
Here, $\varphi^{-1}(R; x) = (0, \infty)^{11}$ is independent of x and this gives the unbiased estimator

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

Another change of variable (LRDE-2): $\boldsymbol{y} = \varphi_{\mathcal{C}}(\boldsymbol{z}; x) = \boldsymbol{z} + x \sum_{j \in \mathcal{C}} \boldsymbol{e}_j$ where $\mathcal{C}$ is a minimal cut such as $\mathcal{C} = \{10, 11\}$, or $\mathcal{C} = \{3, 4, 5, 6\}$, etc.

We have $h(\boldsymbol{y}) = h(\boldsymbol{z}) + x \leq x$ iff $\tilde{h}(\boldsymbol{z}) \stackrel{\text{def}}{=} h(\boldsymbol{z}) + 1 \leq 1$.
This gives $\nabla_x \varphi_{\mathcal{C}}(\boldsymbol{z}; x) = \sum_{j \in \mathcal{C}} \boldsymbol{e}_j$ and $|J_\varphi| = 1$, so $S(\boldsymbol{y}; x) = \sum_{j \in \mathcal{C}} f_j'(y_j)/f_j(y_j)$. Here, $\varphi_{\mathcal{C}}^{-1}(\boldsymbol{y}; x) = \boldsymbol{y} - x \sum_{j \in \mathcal{C}} \boldsymbol{e}_j$, so $\tilde{R}(x) = \varphi_{\mathcal{C}}^{-1}(R; x)$ depends on x in the dimensions $j \in \mathcal{C}$.
We have $r_j(\boldsymbol{y}, x) = -1$ for $j \in \mathcal{C}$ and $\boldsymbol{y} \in \partial R_j^-$, $r_j(\boldsymbol{y}, x) = 0$ otherwise.

Results for SAN Example

		$\hat{\nu}$	e19
KDE	MC	0.78	20.9
	Lat+s	0.95	22.7
	Sob+LMS	0.74	21.9
CDE	MC	0.96	25.6
	Lat+s	1.31	30.9
	Sob+LMS	1.27	29.9
LRDE-1	MC	1.00	20.5
	Lat+s	1.22	23.5
	Sob+LMS	1.16	24.6

With CDE+RQMC, we observe a convergence rate near $\mathcal{O}(n^{-1.3})$ for the MISE.

For $n = 2^{19}$, by using CDE+RQMC rather than KDE+MC,
the MISE is divided by about **500** to **1000**.

LRDE does not perform as well as the CDE in this case.

Conclusion

- Combining a KDE with RQMC can reduce the MISE and sometimes improve its convergence rate, even though the MISE bounds converge faster only when the dimension is very small.
- The CDE is an unbiased density estimator with better convergence rate. Combining it with RQMC can provide an even better rate, and sometimes huge MISE reductions.
- When we cannot find $\mathcal{G}$ for which Assumption 1 holds and $f(x \mid \mathcal{G})$ is easy to compute, the LRDE may provide a good unbiased alternative.
- The LRDE is often discontinuous in $\boldsymbol{U}$, in which case it does not combine well with RQMC. Sometimes, one can apply CMC to the LRDE to make it smoother, and then combine with RQMC. We found that this can provide large variance reductions.
- Extensions: Density estimation for a function of the state of a Markov chain, using Array-RQMC. Generalization to multivariate output.

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.
- A. Ben Abdellah, P. L'Ecuyer, A. B. Owen, and F. Puchhammer. Density estimation by Randomized Quasi-Monte Carlo. *SIAM Journal on Uncertainty Quantification*, **9**, 1, 280–301, 2021.
- J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, Cambridge, U.K., 2010.
- P. W. Glynn. Likelihood ratio gradient estimation: an overview. Proceedings of the 1987 Winter Simulation Conference, 366–375.
- P. J. Laub, R. Salomone, Z. I. Botev. 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. Randomized quasi-Monte Carlo: An introduction for practitioners. In *Monte Carlo and Quasi-Monte Carlo Methods 2016*, 29–52, 2018.
- 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 Research, 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, 2022.
- F. Puchhammer and P. L'Ecuyer, "Likelihood Ratio Density Estimation for Simulation Models," Proceedings of the 2022 Winter Simulation Conference, 109–120, 2022.