

Monte Carlo and Randomized Quasi-Monte Carlo Density Estimation by Conditioning

Pierre L'Ecuyer

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



Optimization Days, Montreal, May 2019

What this talk is about

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of a random variable X and compute a confidence interval on $\mathbb{E}[X]$. $\text{MSE} = \text{Var}[\bar{X}_n] = \mathcal{O}(n^{-1})$.

What this talk is about

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of a random variable X and compute a confidence interval on $\mathbb{E}[X]$. $\text{MSE} = \text{Var}[\bar{X}_n] = \mathcal{O}(n^{-1})$.

But simulation usually 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 of X , defined by $F(x) = \mathbb{P}[X \leq x]$, or its density f defined by $f(x) = F'(x)$.

What this talk is about

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of a random variable X and compute a confidence interval on $\mathbb{E}[X]$. $\text{MSE} = \text{Var}[\bar{X}_n] = \mathcal{O}(n^{-1})$.

But simulation usually 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 of X , defined by $F(x) = \mathbb{P}[X \leq x]$, or its density f defined by $f(x) = F'(x)$.

If X_1, \dots, X_n are n indep. realizations of X , the empirical cdf

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

is unbiased for $F(x)$ at all x , and $\text{Var}[\hat{F}_n(x)] = \mathcal{O}(n^{-1})$.

What this talk is about

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of a random variable X and compute a confidence interval on $\mathbb{E}[X]$. $\text{MSE} = \text{Var}[\bar{X}_n] = \mathcal{O}(n^{-1})$.

But simulation usually 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 of X , defined by $F(x) = \mathbb{P}[X \leq x]$, or its density f defined by $f(x) = F'(x)$.

If X_1, \dots, X_n are n indep. realizations of X , the empirical cdf

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

is unbiased for $F(x)$ at all x , and $\text{Var}[\hat{F}_n(x)] = \mathcal{O}(n^{-1})$.

However, for a continuous r.v. X , the density f provides a better visual idea of the distribution. **Here we focus on estimating f over $[a, b] \subset \mathbb{R}$.**

What this talk is about

Monte Carlo (MC) simulation is widely used to estimate the expectation $\mathbb{E}[X]$ of a random variable X and compute a confidence interval on $\mathbb{E}[X]$. $\text{MSE} = \text{Var}[\bar{X}_n] = \mathcal{O}(n^{-1})$.

But simulation usually 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 of X , defined by $F(x) = \mathbb{P}[X \leq x]$, or its density f defined by $f(x) = F'(x)$.

If X_1, \dots, X_n are n indep. realizations of X , the empirical cdf

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

is unbiased for $F(x)$ at all x , and $\text{Var}[\hat{F}_n(x)] = \mathcal{O}(n^{-1})$.

However, for a continuous r.v. X , the density f provides a better visual idea of the distribution. **Here we focus on estimating f over $[a, b] \subset \mathbb{R}$.**

Can we have $\mathbb{E}[\hat{f}_n(x) - f(x)]^2 = \mathcal{O}(n^{-1})$???

Small example: A stochastic activity network

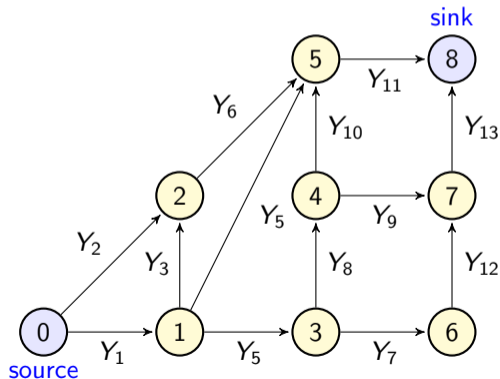
Gives precedence relations between activities. Activity k has random duration Y_k (also length of arc k) with known cdf $F_k(y) := \mathbb{P}[Y_k \leq y]$.

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

Can look at deterministic equivalent of X , $\mathbb{E}[X]$, cdf, density, ...

Want to estimate the density of X ,

$$f(x) = F'(x) = \frac{d}{dx} \mathbb{P}[X \leq x].$$



Small example: A stochastic activity network

Gives precedence relations between activities. Activity k has random duration Y_k (also 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.

Can look at deterministic equivalent of X , $\mathbb{E}[X]$, cdf, density, ...

The **sample cdf**

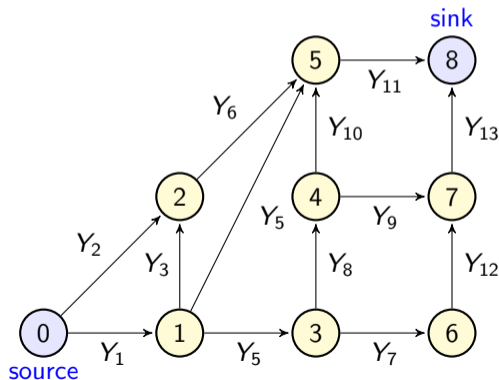
$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[X_i \leq x]$$

is an unbiased estimator of the **cdf**

$$F(x) = \mathbb{P}[X \leq x].$$

Want to estimate the **density of X** ,

$$f(x) = F'(x) = \frac{d}{dx} \mathbb{P}[X \leq x].$$



Small example: A stochastic activity network

Gives precedence relations between activities. Activity k has random duration Y_k (also 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.

Can look at deterministic equivalent of X , $\mathbb{E}[X]$, cdf, density, ...

The **sample cdf**

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

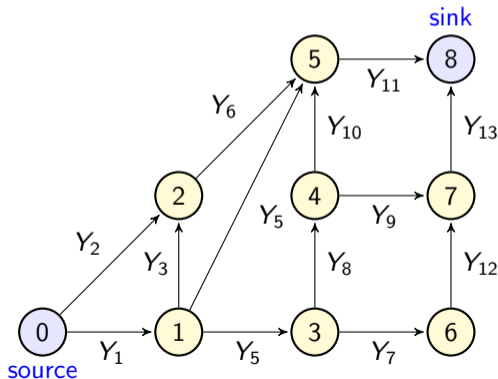
is an unbiased estimator of the **cdf**

$$F(x) = \mathbb{P}[X \leq x].$$

Want to estimate the **density of X** ,

$$f(x) = F'(x) = \frac{d}{dx} \mathbb{P}[X \leq x].$$

The sample derivative $\hat{F}'_n(x)$ is **useless** to estimate $f(x)$, because it is 0 almost everywhere.

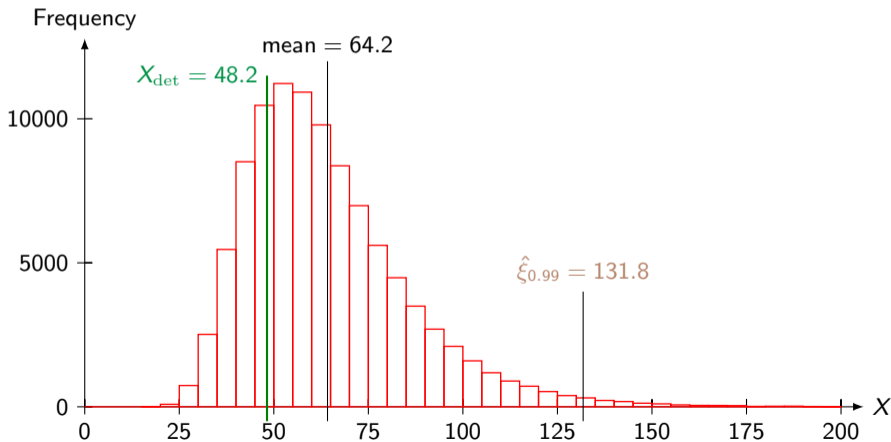


Numerical illustration from Elmaghraby (1977):

$Y_k \sim N(\mu_k, \sigma_k^2)$ for $k = 1, 2, 4, 11, 12$, and $Y_k \sim \text{Expon}(1/\mu_k)$ otherwise.

μ_1, \dots, μ_{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.

Results of an experiment with $n = 100\,000$. Note: X is not normal!



Density Estimation

Suppose we estimate the density f over a finite interval $[a, b]$.

Let $\hat{f}_n(x)$ denote the density estimator at x , with sample size n .

We use the following measures of error:

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

$$\text{IV} = \text{integrated variance} = \int_a^b \text{Var}[\hat{f}_n(x)] dx$$

$$\text{ISB} = \text{integrated squared bias} = \int_a^b (\mathbb{E}[\hat{f}_n(x)] - f(x))^2 dx$$

Density Estimation

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

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

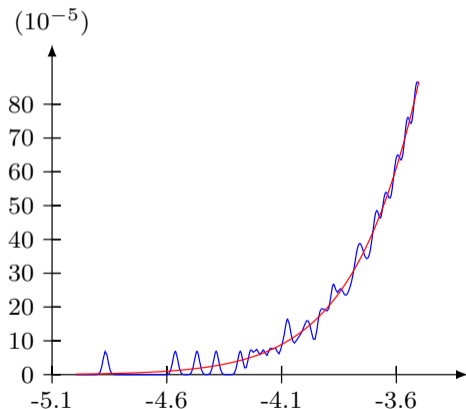
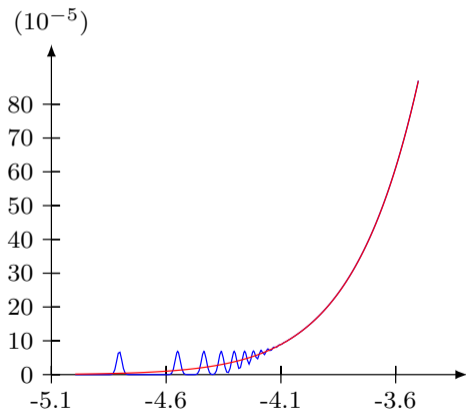
where n_j is the number of observations X_i that fall in interval j .

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

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

KDE bandwidth selection: an illustration in $s = 1$ dimension

KDE (blue) vs true density (red) with RQMC point sets with $n = 2^{19}$:
midpoint rule (left), Stratified sample of $U = F(X)$ (right)



KDE asymptotic convergence with Monte Carlo for smooth f

For any $g : \mathbb{R} \rightarrow \mathbb{R}$, define

$$R(g) = \int_a^b (g(x))^2 dx,$$

$$\mu_r(g) = \int_{-\infty}^{\infty} x^r g(x) dx, \quad \text{for } r = 0, 1, 2, \dots$$

For histograms and KDEs, when $n \rightarrow \infty$ and $h \rightarrow 0$:

$$\text{AMISE} = \text{AIV} + \text{AISB} \sim \frac{C}{nh} + Bh^\alpha.$$

	C	B	α
Histogram	1	$R(f') / 12$	2
KDE	$\mu_0(k^2)$	$(\mu_2(k))^2 R(f'') / 4$	4

The asymptotically optimal h is

$$h^* = \left(\frac{C}{B\alpha n} \right)^{1/(\alpha+1)}$$

and it gives $\text{AMISE} = Kn^{-\alpha/(1+\alpha)}$.

	C	B	α	h^*	AMISE
Histogram	1	$\frac{R(f')}{12}$	2	$(nR(f')/6)^{-1/3}$	$\mathcal{O}(n^{-2/3})$
KDE	$\mu_0(k^2)$	$\frac{(\mu_2(k))^2 R(f'')}{4}$	4	$\left(\frac{\mu_0(k^2)}{(\mu_2(k))^2 R(f'') n} \right)^{1/5}$	$\mathcal{O}(n^{-4/5})$

To estimate h^* , one can estimate $R(f')$ and $R(f'')$ via KDE (plugin).

This is under the simplifying assumption that h must be the same all over $[a, b]$.

Can we take the stochastic derivative of an estimator of F ?

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

A simple candidate cdf estimator is the empirical cdf

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

However $d\hat{F}_n(x)/dx = 0$ almost everywhere, so this cannot be a useful density estimator!

We need a smoother estimator of F .

Conditional Monte Carlo (CMC) for Derivative Estimation

Idea: Replace indicator $\mathbb{I}[X_i \leq x]$ by its conditional cdf given filtered information:

$$F(x | \mathcal{G}) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x | \mathcal{G}]$$

where the sigma-field \mathcal{G} contains not enough information to reveal X but enough to compute $F(x | \mathcal{G})$, and is chosen so that the following holds:

Conditional Monte Carlo (CMC) for Derivative Estimation

Idea: Replace indicator $\mathbb{I}[X_i \leq x]$ by its conditional cdf given filtered information:

$$F(x | \mathcal{G}) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x | \mathcal{G}]$$

where the sigma-field \mathcal{G} contains not enough information to reveal X but enough to compute $F(x | \mathcal{G})$, and is chosen so that the following holds:

Assumption 1. For all realizations of \mathcal{G} , $F(x | \mathcal{G})$ 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 for which $F'(x | \mathcal{G}) = dF(x | \mathcal{G})/dx$ (when it exists) is bounded uniformly in x by a random variable Γ such that $\mathbb{E}[\Gamma^2] \leq K_\gamma < \infty$.

Theorem 1: Under Assump. 1, for $x \in [a, b]$, $\mathbb{E}[F'(x | \mathcal{G})] = f(x)$ and $\text{Var}[F'(x | \mathcal{G})] < K_\gamma$.

Conditional Monte Carlo (CMC) for Derivative Estimation

Idea: Replace indicator $\mathbb{I}[X_i \leq x]$ by its conditional cdf given filtered information:

$$F(x | \mathcal{G}) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x | \mathcal{G}]$$

where the sigma-field \mathcal{G} contains not enough information to reveal X but enough to compute $F(x | \mathcal{G})$, and is chosen so that the following holds:

Assumption 1. For all realizations of \mathcal{G} , $F(x | \mathcal{G})$ 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 for which $F'(x | \mathcal{G}) = dF(x | \mathcal{G})/dx$ (when it exists) is bounded uniformly in x by a random variable Γ such that $\mathbb{E}[\Gamma^2] \leq K_\gamma < \infty$.

Theorem 1: Under Assump. 1, for $x \in [a, b]$, $\mathbb{E}[F'(x | \mathcal{G})] = f'(x)$ and $\text{Var}[F'(x | \mathcal{G})] < K_\gamma$.

Theorem 2: If $\mathcal{G} \subset \tilde{\mathcal{G}}$ both satisfy Assumption 1, then $\text{Var}[F'(x | \mathcal{G})] \leq \text{Var}[F'(x | \tilde{\mathcal{G}})]$.

Conditional Monte Carlo (CMC) for Derivative Estimation

Idea: Replace indicator $\mathbb{I}[X_i \leq x]$ by its conditional cdf given filtered information:

$$F(x | \mathcal{G}) \stackrel{\text{def}}{=} \mathbb{P}[X \leq x | \mathcal{G}]$$

where the sigma-field \mathcal{G} contains not enough information to reveal X but enough to compute $F(x | \mathcal{G})$, and is chosen so that the following holds:

Assumption 1. For all realizations of \mathcal{G} , $F(x | \mathcal{G})$ 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 for which $F'(x | \mathcal{G}) = dF(x | \mathcal{G})/dx$ (when it exists) is bounded uniformly in x by a random variable Γ such that $\mathbb{E}[\Gamma^2] \leq K_\gamma < \infty$.

Theorem 1: Under Assump. 1, for $x \in [a, b]$, $\mathbb{E}[F'(x | \mathcal{G})] = f(x)$ and $\text{Var}[F'(x | \mathcal{G})] < K_\gamma$.

Theorem 2: If $\mathcal{G} \subset \tilde{\mathcal{G}}$ both satisfy Assumption 1, then $\text{Var}[F'(x | \mathcal{G})] \leq \text{Var}[F'(x | \tilde{\mathcal{G}})]$.

Conditional density estimator (CDE) with sample size n : $\hat{f}_{\text{cde},n}(x) = \frac{1}{n} \sum_{i=1}^n F'(x | \mathcal{G}^{(i)})$

where $\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(n)}$ are n independent realizations of \mathcal{G} .

Example 1. A sum of independent random variables

$X = Y_1 + \dots + Y_d$, where the Y_j are independent and continuous with cdf F_j and density f_j , and \mathcal{G} is defined by hiding Y_k for an arbitrary k :

$$\mathcal{G} = \mathcal{G}_k = S_{-k} \stackrel{\text{def}}{=} Y_1 + \dots + Y_k + \dots + Y_d.$$

We have

$$F(x | \mathcal{G}_k) = \mathbb{P}[X \leq x | S_{-k}] = \mathbb{P}[Y_k \leq x - S_{-k}] = F_k(x - S_{-k})$$

and the density estimator becomes $F'(x | \mathcal{G}_k) = f_k(x - S_{-k})$.

Example 1. A sum of independent random variables

$X = Y_1 + \dots + Y_d$, where the Y_j are independent and continuous with cdf F_j and density f_j , and \mathcal{G} is defined by hiding Y_k for an arbitrary k :

$$\mathcal{G} = \mathcal{G}_k = S_{-k} \stackrel{\text{def}}{=} Y_1 + \dots + Y_k + \dots + Y_d.$$

We have

$$F(x | \mathcal{G}_k) = \mathbb{P}[X \leq x | S_{-k}] = \mathbb{P}[Y_k \leq x - S_{-k}] = F_k(x - S_{-k})$$

and the density estimator becomes $F'(x | \mathcal{G}_k) = f_k(x - S_{-k})$. Shifted density of Y_k .

The idea of using CMC for density estimation was introduced by Asmussen (2018) for this special case, with $k = d$ and same F_j for all j .

Example 2: generalization

Let $X = h(Y_1, \dots, Y_d)$ and define \mathcal{G}_k again by erasing a continuous Y_k ;
 $\mathcal{G}_k = (Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_d)$.

Example 2: generalization

Let $X = h(Y_1, \dots, Y_d)$ and define \mathcal{G}_k again by erasing a continuous Y_k ;
 $\mathcal{G}_k = (Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_d)$.

Exemple: $X = (Y_1 + Y_2^2)/Y_3$ where $Y_3 > 0$.

Example 2: generalization

Let $X = h(Y_1, \dots, Y_d)$ and define \mathcal{G}_k again by erasing a continuous Y_k ;
 $\mathcal{G}_k = (Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_d)$.

Exemple: $X = (Y_1 + Y_2^2)/Y_3$ where $Y_3 > 0$.

If $k = 3$, since $X \leq x$ if and only if $Y_3 \geq (Y_1 + Y_2^2)/x$,
 we have $F(x | \mathcal{G}_3) = \mathbb{P}(X \leq x | Y_1, Y_2) = 1 - F_3((Y_1 + Y_2^2)/x)$,

Example 2: generalization

Let $X = h(Y_1, \dots, Y_d)$ and define \mathcal{G}_k again by erasing a continuous Y_k ;
 $\mathcal{G}_k = (Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_d)$.

Exemple: $X = (Y_1 + Y_2^2)/Y_3$ where $Y_3 > 0$.

If $k = 3$, since $X \leq x$ if and only if $Y_3 \geq (Y_1 + Y_2^2)/x$,
 we have $F(x | \mathcal{G}_3) = \mathbb{P}(X \leq x | Y_1, Y_2) = 1 - F_3((Y_1 + Y_2^2)/x)$,
 and the density estimator at x is $F'(x | \mathcal{G}_3) = f_3((Y_1 + Y_2^2)/x)(Y_1 + Y_2^2)/x^2$.

Example 2: generalization

Let $X = h(Y_1, \dots, Y_d)$ and define \mathcal{G}_k again by erasing a continuous Y_k ;
 $\mathcal{G}_k = (Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_d)$.

Exemple: $X = (Y_1 + Y_2^2)/Y_3$ where $Y_3 > 0$.

If $k = 3$, since $X \leq x$ if and only if $Y_3 \geq (Y_1 + Y_2^2)/x$,
 we have $F(x | \mathcal{G}_3) = \mathbb{P}(X \leq x | Y_1, Y_2) = 1 - F_3((Y_1 + Y_2^2)/x)$,
 and the density estimator at x is $F'(x | \mathcal{G}_3) = f_3((Y_1 + Y_2^2)/x)(Y_1 + Y_2^2)/x^2$.

If $k = 2$, then $F(x | \mathcal{G}_2) = \mathbb{P}(X \leq x | Y_1, Y_3) = \mathbb{P}(|Y_2| \leq (Y_3x - Y_1)^{1/2}) = F_2(Z) - F_2(-Z)$
 where $Z = (Y_3x - Y_1)^{1/2}$, and the density estimator at x is
 $F'(x | \mathcal{G}_2) = (f_2(Z) + f_2(-Z))dZ/dx = (f_2(Z) - f_2(-Z))Y_3/(2Z)$.

This second estimator can be problematic if Z can take values near 0; this shows that a good choice of k can be crucial in general.

Example 3: discontinuity issues

Let $X = \max(Y_1, Y_2)$ where Y_1 and Y_2 are independent and continuous. With $\mathcal{G} = \mathcal{G}_2$ (we hide Y_2):

$$\mathbb{P}[X \leq x \mid Y_1 = y] = \begin{cases} \mathbb{P}[Y_2 \leq x \mid Y_1 = y] = F_2(x) & \text{if } x \leq y; \\ 0 & \text{if } x < y. \end{cases}$$

If $F_2(y) > 0$, this function is discontinuous at $x = y$, so Assumption 1 does not hold. The method does not work in this case.

Example 3: discontinuity issues

Let $X = \max(Y_1, Y_2)$ where Y_1 and Y_2 are independent and continuous. With $\mathcal{G} = \mathcal{G}_2$ (we hide Y_2):

$$\mathbb{P}[X \leq x \mid Y_1 = y] = \begin{cases} \mathbb{P}[Y_2 \leq x \mid Y_1 = y] = F_2(x) & \text{if } x \leq y; \\ 0 & \text{if } x < y. \end{cases}$$

If $F_2(y) > 0$, this function is discontinuous at $x = y$, so Assumption 1 does not hold. The method does not work in this case.

Same problem if $X = \min(Y_1, Y_2)$. With $\mathcal{G} = \mathcal{G}_2$, we have

$$\mathbb{P}[X \leq x \mid Y_1 = y] = \begin{cases} F_2(x) & \text{if } x \leq y; \\ 1 & \text{if } x > y. \end{cases}$$

If $F_2(y) < 1$, this function is also discontinuous at $x = y$.

Elementary quasi-Monte Carlo (QMC) Bounds (Recall)

Integration error for $g : [0, 1]^s \rightarrow \mathbb{R}$ with point set $P_n = \{\mathbf{u}_0, \dots, \mathbf{u}_{n-1}\} \subset [0, 1]^s$:

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

Koksma-Hlawka inequality: $|E_n| \leq V_{\text{HK}}(g) D^*(P_n)$ where

$$V_{\text{HK}}(g) = \sum_{\emptyset \neq \mathbf{v} \subseteq \mathcal{S}} \int_{[0,1]^s} \left| \frac{\partial^{|\mathbf{v}|} g}{\partial \mathbf{v}}(\mathbf{u}) \right| d\mathbf{u}, \quad (\text{Hardy-Krause (HK) variation})$$

$$D^*(P_n) = \sup_{\mathbf{u} \in [0,1]^s} \left| \text{vol}[\mathbf{0}, \mathbf{u}) - \frac{|P_n \cap [\mathbf{0}, \mathbf{u})|}{n} \right| \quad (\text{star-discrepancy}).$$

There are explicit point sets for which $D^*(P_n) = \mathcal{O}((\log n)^{s-1}/n) = \mathcal{O}(n^{-1+\epsilon})$, $\forall \epsilon > 0$.
Explicit RQMC constructions for which $\mathbb{E}[E_n] = 0$ and $\text{Var}[E_n] = \mathcal{O}(n^{-2+\epsilon})$, $\forall \epsilon > 0$.

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

Combining RQMC with the KDE

Done in Ben Abdellah, L'Ecuyer, Owen, Puchhammer (2019).

Difficulty:

The KDE has a very large variation when the bandwidth h is small (to reduce the bias). So unless the (effective) dimension is very small, RQMC reduces the MISE only modestly.

Applying RQMC to the CDE

To apply RQMC to the CDE, we must be able to write the density estimator as a function of $\mathbf{u} \in [0, 1)^s$:

$$\begin{aligned}F(x | \mathcal{G}) &= \tilde{g}(x, \mathbf{u}), \\F'(x | \mathcal{G}) &= \tilde{g}'(x, \mathbf{u}) = dg(x, \mathbf{u})/dx\end{aligned}$$

for some $\tilde{g} : [a, b] \times [0, 1)^s$ for which $\tilde{g}'(x, \cdot)$ has bounded HK variation for each x .

Applying RQMC to the CDE

To apply RQMC to the CDE, we must be able to write the density estimator as a function of $\mathbf{u} \in [0, 1)^s$:

$$\begin{aligned}F(x | \mathcal{G}) &= \tilde{g}(x, \mathbf{u}), \\F'(x | \mathcal{G}) &= \tilde{g}'(x, \mathbf{u}) = dg(x, \mathbf{u})/dx\end{aligned}$$

for some $\tilde{g} : [a, b] \times [0, 1)^s$ for which $\tilde{g}'(x, \cdot)$ has bounded HK variation for each x .

CDE sample: $\tilde{g}'(x, \mathbf{U}_1), \dots, \tilde{g}'(x, \mathbf{U}_n)$ where $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ is an RQMC point set over $[0, 1)^s$.

Applying RQMC to the CDE

To apply RQMC to the CDE, we must be able to write the density estimator as a function of $\mathbf{u} \in [0, 1]^s$:

$$\begin{aligned}F(x | \mathcal{G}) &= \tilde{g}(x, \mathbf{u}), \\F'(x | \mathcal{G}) &= \tilde{g}'(x, \mathbf{u}) = dg(x, \mathbf{u})/dx\end{aligned}$$

for some $\tilde{g} : [a, b] \times [0, 1]^s$ for which $\tilde{g}'(x, \cdot)$ has bounded HK variation for each x .

CDE sample: $\tilde{g}'(x, \mathbf{U}_1), \dots, \tilde{g}'(x, \mathbf{U}_n)$ where $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ is an RQMC point set over $[0, 1]^s$.

If $\tilde{g}'(x, \cdot)$ does not have bounded variation, RQMC can still be worthwhile, although there is no guarantee.

Example: sum of independent random variables (again)

$X = Y_1 + \cdots + Y_d$, where the Y_j are independent and continuous with cdf F_j and density f_j , and \mathcal{G} is defined by hiding Y_k for an arbitrary k :

$$\mathcal{G}_k = S_{-k} \stackrel{\text{def}}{=} Y_1 + \cdots + Y_k + \cdots + Y_d = F_1^{-1}(U_1) + \cdots + F_k^{-1}(U_k) + \cdots + F_d^{-1}(U_d).$$

We have $F(x | \mathcal{G}_k) = F_k(x - S_{-k}) = \tilde{g}(x, \cdot)$

and the density estimator is $F'(x | \mathcal{G}_k) = f_k(x - S_{-k}) = \tilde{g}'(x, \mathbf{U})$ where $\mathbf{U} = (U_1, \dots, U_d)$.

If $\tilde{g}'(x, \cdot)$ has bounded HK variation, then $\text{MISE} = \mathcal{O}(n^{-2+\epsilon})$.

Experimental setting for numerical experiments

We want to test the method on some examples. For each method and each n considered, we compute the CDE with n samples, evaluate it at a set of n_e evaluation points over $[a, b]$, repeat this n_r times, compute the variance at each evaluation point, and estimate the IV.

Experimental setting for numerical experiments

We want to test the method on some examples. For each method and each n considered, we compute the CDE with n samples, evaluate it at a set of n_e evaluation points over $[a, b]$, repeat this n_r times, compute the variance at each evaluation point, and estimate the IV.

We repeat this for $n = 2^{14}, \dots, 2^{19}$ and fit the model $IV = Kn^{-\nu}$ by linear regression:
 $\log_2 IV \approx \log_2 K - \nu \log_2 n$. We report $\hat{\nu}$ and also the IV for $n = 2^{19}$.

Experimental setting for numerical experiments

We want to test the method on some examples. For each method and each n considered, we compute the CDE with n samples, evaluate it at a set of n_e evaluation points over $[a, b]$, repeat this n_r times, compute the variance at each evaluation point, and estimate the IV.

We repeat this for $n = 2^{14}, \dots, 2^{19}$ and fit the model $IV = Kn^{-\nu}$ by linear regression: $\log_2 IV \approx \log_2 K - \nu \log_2 n$. We report $\hat{\nu}$ and also the IV for $n = 2^{19}$.

MC and RQMC Point sets:

- ▶ **MC**: Independent points (MC),
- ▶ **Lat+s**: lattice rule with a random shift modulo 1,
- ▶ **Lat+s+b**: lattice rule with a random shift modulo 1 + baker's transformation,
- ▶ **LMS**: Sobol' points with left matrix scrambling (LMS) + digital random shift.

Displacement of a cantilever beam (Bingham 2017)

Displacement X of a cantilever beam with horizontal load Y_2 and vertical load Y_3 :

$$X = h(Y_1, Y_2, Y_3) = \frac{\kappa}{Y_1} \sqrt{\frac{Y_2^2}{w^4} + \frac{Y_3^2}{t^4}} \quad (1)$$

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

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

Conditioning on $\mathcal{G}_1 = \{Y_2, Y_3\}$ means hiding Y_1 . We have

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

For $x > 0$,

$$F(x | \mathcal{G}_1) = \mathbb{P}[Y_1 \geq W_1 | W_1] = 1 - \Phi((W_1 - \mu_1)/\sigma_1)$$

and

$$F'(x | \mathcal{G}_1) = -\frac{\phi((W_1 - \mu_1)/\sigma_1)W_1'(x)}{\sigma_1} = \frac{\phi((W_1 - \mu_1)/\sigma_1)W_1(x)}{x\sigma_1}.$$

Suppose we condition on $\mathcal{G}_2 = \{Y_1, Y_3\}$ instead, i.e., hide Y_2 . We have

$$X \leq x \quad \text{if and only if} \quad Y_2^2 \leq w^4 \left((xY_1/\kappa)^2 - Y_3^2/t^4 \right) \stackrel{\text{def}}{=} W_2.$$

Suppose we condition on $\mathcal{G}_2 = \{Y_1, Y_3\}$ instead, i.e., hide Y_2 . We have

$$X \leq x \quad \text{if and only if} \quad Y_2^2 \leq w^4 \left((xY_1/\kappa)^2 - Y_3^2/t^4 \right) \stackrel{\text{def}}{=} W_2.$$

If $W_2 \leq 0$, then $F'(x | \mathcal{G}_2) = 0$. If $W_2 > 0$,

$$F(x | \mathcal{G}_2) = \mathbb{P}[-\sqrt{W_2} \leq Y_2 \leq \sqrt{W_2} | W_2] = \Phi((\sqrt{W_2} - \mu_2)/\sigma_2) - \Phi(-(\sqrt{W_2} + \mu_2)/\sigma_2)$$

and

$$F'(x | \mathcal{G}_2) = \frac{\phi((\sqrt{W_2} - \mu_2)/\sigma_2) + \phi(-(\sqrt{W_2} + \mu_2)/\sigma_2)}{w^4 x (Y_1/\kappa)^2 / (\sigma_2 \sqrt{W_2})} > 0.$$

Suppose we condition on $\mathcal{G}_2 = \{Y_1, Y_3\}$ instead, i.e., hide Y_2 . We have

$$X \leq x \quad \text{if and only if} \quad Y_2^2 \leq w^4 \left((xY_1/\kappa)^2 - Y_3^2/t^4 \right) \stackrel{\text{def}}{=} W_2.$$

If $W_2 \leq 0$, then $F'(x | \mathcal{G}_2) = 0$. If $W_2 > 0$,

$$F(x | \mathcal{G}_2) = \mathbb{P}[-\sqrt{W_2} \leq Y_2 \leq \sqrt{W_2} | W_2] = \Phi((\sqrt{W_2} - \mu_2)/\sigma_2) - \Phi(-(\sqrt{W_2} + \mu_2)/\sigma_2)$$

and

$$F'(x | \mathcal{G}_2) = \frac{\phi((\sqrt{W_2} - \mu_2)/\sigma_2) + \phi(-(\sqrt{W_2} + \mu_2)/\sigma_2)}{w^4 x (Y_1/\kappa)^2 / (\sigma_2 \sqrt{W_2})} > 0.$$

For conditioning on \mathcal{G}_3 , the analysis is the same as for \mathcal{G}_2 , by symmetry, and we get

$$F'(x | \mathcal{G}_3) = \frac{\phi((\sqrt{W_3} - \mu_3)/\sigma_3) + \phi(-(\sqrt{W_3} + \mu_3)/\sigma_3)}{t^4 x (Y_1/\kappa)^2 / (\sigma_3 \sqrt{W_3})} > 0.$$

for $W_3 > 0$, where W_3 is defined in a similar way as W_2 .

Instead of choosing a single conditioning k , we can take a convex combination:

$$\hat{f}(x) = \alpha_1 F'(x | \mathcal{G}_1) + \alpha_2 F'(x | \mathcal{G}_2) + \alpha_3 F'(x | \mathcal{G}_3),$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. This is equivalent to taking $F'(x | \mathcal{G}_1)$ as the main estimator and the other two as **control variates (CV)**. We can use CV theory to optimize the α_j 's.

Instead of choosing a single conditioning k , we can take a convex combination:

$$\hat{f}(x) = \alpha_1 F'(x | \mathcal{G}_1) + \alpha_2 F'(x | \mathcal{G}_2) + \alpha_3 F'(x | \mathcal{G}_3),$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. This is equivalent to taking $F'(x | \mathcal{G}_1)$ as the main estimator and the other two as **control variates (CV)**. We can use CV theory to optimize the α_j 's.

	\hat{v}					$-\log_2 \text{MISE} (n = 2^{19})$				
	KDE	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	comb.	KDE	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	comb.
MC	0.80	0.97	0.98	0.99	0.98	14.7	19.3	14.5	22.8	22.5
Lat+s	—	2.06	2.82	2.04	2.02	—	38.9	25.4	41.5	41.5
Lat+s+b	—	2.26	2.55	1.98	2.07	—	44.3	23.3	45.5	46.0
Sob+LMS	0.96	2.21	2.03	2.21	2.21	20.5	44.0	23.6	45.7	46.1

Instead of choosing a single conditioning k , we can take a convex combination:

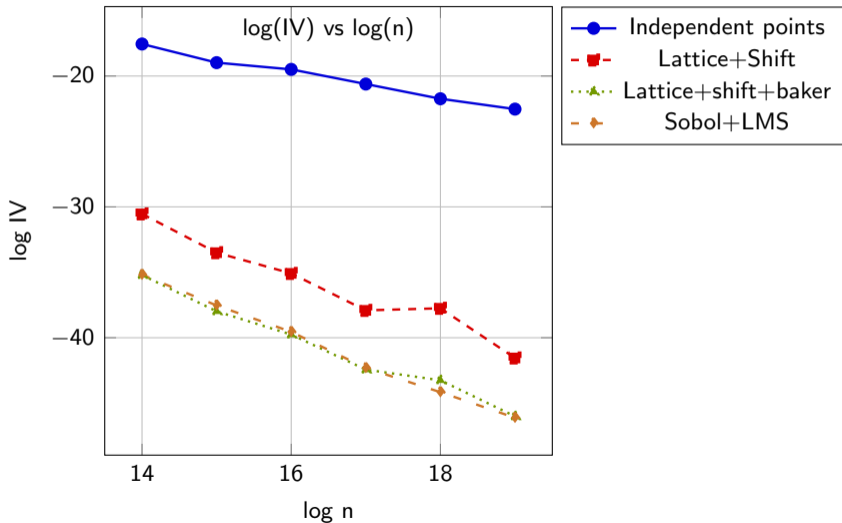
$$\hat{f}(x) = \alpha_1 F'(x | \mathcal{G}_1) + \alpha_2 F'(x | \mathcal{G}_2) + \alpha_3 F'(x | \mathcal{G}_3),$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. This is equivalent to taking $F'(x | \mathcal{G}_1)$ as the main estimator and the other two as **control variates (CV)**. We can use CV theory to optimize the α_j 's.

	\hat{v}					$-\log_2 \text{MISE} (n = 2^{19})$				
	KDE	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	comb.	KDE	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	comb.
MC	0.80	0.97	0.98	0.99	0.98	14.7	19.3	14.5	22.8	22.5
Lat+s	—	2.06	2.82	2.04	2.02	—	38.9	25.4	41.5	41.5
Lat+s+b	—	2.26	2.55	1.98	2.07	—	44.3	23.3	45.5	46.0
Sob+LMS	0.96	2.21	2.03	2.21	2.21	20.5	44.0	23.6	45.7	46.1

For $n = 2^{19}$, the MISE is about $2^{-14.7}$ for the usual KDE+MC and 2^{-46} for the new CDE+RQMC; i.e., **MISE is divided by more than $2^{31} \approx 2$ millions.**

Comparison for CDE with linear combination of 3 estimators, for cantilever.

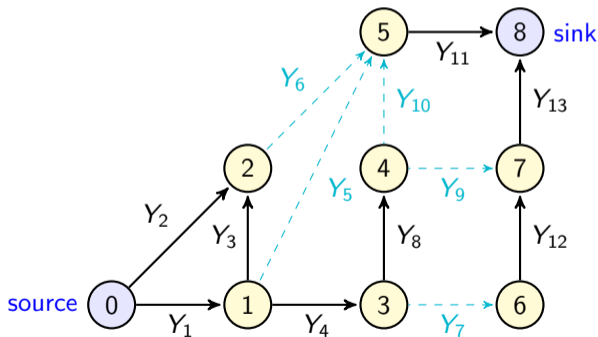


CMC for the SAN Example

Want to estimate the density of the longest path length X .

CMC estimator of $\mathbb{P}[X \leq x]$: $F(x | \mathcal{G}) = \mathbb{P}[X \leq x | \{Y_j, j \notin \mathcal{L}\}]$ for a minimal cut \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 .



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

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

and

$$F'(x | \mathcal{G}) = \sum_{j \in \mathcal{L}} f_j[x - P_j] \prod_{l \in \mathcal{L}, l \neq j} F_l[x - P_l],$$

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 $F(\cdot | \mathcal{G})$, and Assumption 1 holds, so $F'(x | \mathcal{G})$ is an unbiased density estimator.

Estimated $\text{MISE} = Kn^{-\nu}$, for KDE with CMC.

		$\hat{\nu}$	$-\log_2 \text{MISE} (n = 2^{19})$
KDE	MC	0.77	20.9
	Lat+s	0.75	22.0
	Sobol+LMS	0.76	22.0
CDE	MC	0.99	25.5
	Lat+s	1.26	29.9
	Sobol+LMS	1.25	29.9

With RQMC, we observe a convergence rate near $\mathcal{O}(n^{-1.25})$ for the IV and the MISE.

For $n = 2^{19}$, by using the new CDE+RQMC rather than the usual KDE+MC, the MISE is divided by about $2^9 \approx 500$.

Conclusion

- ▶ The CDE is an unbiased density estimator with better convergence rate for the IV and the MISE. Combining it with RQMC can provide an even better rate, and sometimes huge MISE reductions.
- ▶ **Future:** Density estimation for a function of the state of a Markov chain, using Array-RQMC.
- ▶ What if we cannot find \mathcal{G} for which Assumption 1 holds and $F'(x | \mathcal{G})$ is easy to compute?
Current work: density estimator based on likelihood ratio derivative estimation.
- ▶ Lots of potential applications.

- ▶ S. Asmussen. Conditional Monte Carlo for sums, with applications to insurance and finance, *Annals of Actuarial Science*, prepublication, 1–24, 2018.
- ▶ S. Asmussen and P. W. Glynn. *Stochastic Simulation*. Springer-Verlag, 2007.
- ▶ A. Ben Abdellah, P. L'Ecuyer, A. B. Owen, and F. Puchhammer. Density estimation by Randomized Quasi-Monte Carlo. Submitted, 2018.
- ▶ J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, Cambridge, U.K., 2010.
- ▶ P. L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science* 36: 1364–1383, 1990.
- ▶ P. L'Ecuyer. Quasi-Monte Carlo methods with applications in finance. *Finance and Stochastics*, 13(3):307–349, 2009.
- ▶ P. L'Ecuyer. Randomized quasi-Monte Carlo: An introduction for practitioners. In P. W. Glynn and A. B. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2016*, 2017.
- ▶ P. L'Ecuyer and G. Perron. On the Convergence Rates of IPA and FDC Derivative Estimators for Finite-Horizon Stochastic Simulations. *Operations Research*, 42 (4):643–656, 1994.
- ▶ D. W. Scott. *Multivariate Density Estimation*. Wiley, 2015.