On Density estimation by Randomized Quasi-Monte Carlo

Pierre L'Ecuyer

Joint work with **Amal Ben Abdellah, Art B. Owen, and Florian Puchhammer**

Université de Montréal DIRO GERAD GROUP FOR RESEARCH IN DECISION ANALYSIS CIRRELT Inria

Workshop at RICAM, Linz, Austria, December 2018

What this talk is about

Quasi-Monte Carlo (QMC) and randomized QMC (RQMC) methods have been studied extensively for estimating an integral, say $\mathbb{E}[X]$, or approximating a function from its evaluation at a finite number of points.

How can we use them to estimate the entire distribution of X?
Here we will focus on estimating the density of X over $[a, b] \subset \mathbb{R}$.

People often look at empirical distributions via histograms, for example.
More refined methods: kernel density estimators (KDEs).
Can RQMC improve such density estimators, and by how much?
Are there other types of density estimators than KDEs, that work better with RQMC?

Setting

Classical density estimation was developed in the context where independent observations $X_1, \ldots, X_n$ of X are **given** and one wishes to estimate the density f of X from that.

Here we assume that $X_1, \ldots, X_n$ are **generated by simulation** from a stochastic model. We can choose n and we have some freedom on how the simulation is performed.

The X_i's are realizations of a random variable $X = g(\mathbf{U}) \in \mathbb{R}$ with density f, where $\mathbf{U} = (U_1, \ldots, U_s) \sim U(0,1)^s$ and $g(\mathbf{u})$ can be computed easily for any $\mathbf{u} \in (0,1)^s$.

Can we obtain a better estimate of f with RQMC instead of MC? How much better?

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:

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

Density Estimation

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

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

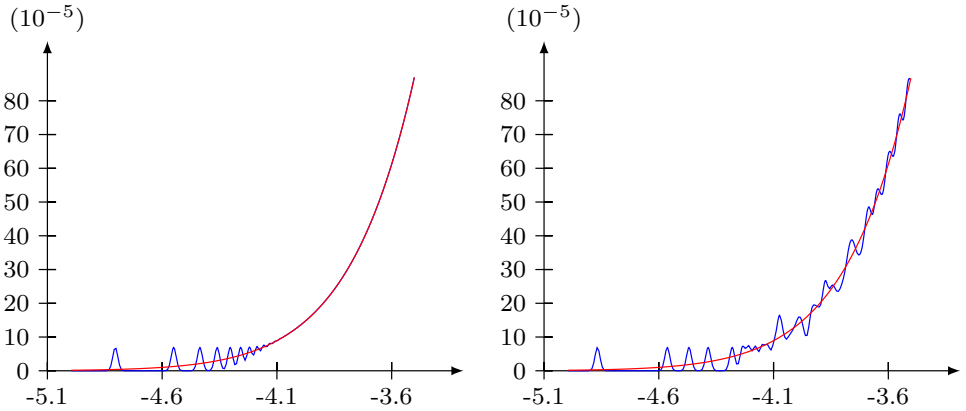
where n_j is the number of observations X_i that fall in interval j.

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 $n = 2^{19}$:
lattice + shift (left), Sobol + 31-bit scramble (right)



Asymptotic convergence with Monte Carlo for smooth f

For $g : \mathbb{R} \to \mathbb{R}$, define

$$\begin{aligned} R(g) &= \int_a^b (g(x))^2 \mathrm{d}x, \\ \mu_r(g) &= \int_{-\infty}^{\infty} x^r g(x) \mathrm{d}x, \qquad \text{for } r = 0, 1, 2, \ldots \end{aligned}$$

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

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

	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 $\mathrm{AMISE} = Kn^{-\alpha/(1+\alpha)}$.

	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).

Asymptotic convergence with RQMC for smooth f

Idea: Replace $\mathbf{U}_1, \ldots, \mathbf{U}_n$ by RQMC points.

RQMC does not change the bias.

For a KDE with smooth k, one could hope (perhaps) to get

$$\mathrm{AIV} = C' n^{-\beta} h^{-1} \quad \text{for } \beta > 1, \text{ instead of } C n^{-1} h^{-1}.$$

If the IV is reduced, the optimal h can be taken smaller to reduce the ISB as well (re-balance) and then reduce the MISE.

Asymptotic convergence with RQMC for smooth f

Idea: Replace $\mathbf{U}_1, \ldots, \mathbf{U}_n$ by RQMC points.

RQMC does not change the bias.

For a KDE with smooth k, one could hope (perhaps) to get

$$\mathrm{AIV} = C' n^{-\beta} h^{-1} \quad \text{for } \beta > 1, \text{ instead of } C n^{-1} h^{-1}.$$

If the IV is reduced, the optimal h can be taken smaller to reduce the ISB as well (re-balance) and then reduce the MISE.

Unfortunately, things are not so simple.

Roughly, decreasing h increases the variation of the function in the estimator. So we rather have something like

$$\mathrm{AIV} = C' n^{-\beta} h^{-\delta}$$

or $\mathrm{IV} \approx C' n^{-\beta} h^{-\delta}$ in some bounded region.

Elementary QMC Bounds (Recall)

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

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

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

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

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

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

Bounding the AIV under RQMC for a KDE

KDE density estimator at a single point x:

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

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

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:

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

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

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 $\tilde{g}$.

The partial derivatives are:

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

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 $j = 2, \ldots, |\mathfrak{v}| + 1$.
One of the term for $\mathfrak{v} = \mathcal{S}$ grows as $h^{-s-1} k^{(s)}\left((g(\mathbf{u}) - x)/h\right) \prod_{j=1}^{s} g_j(\mathbf{u}) = \mathcal{O}(h^{-s-1})$ when $h \to 0$, so this AIV bound grows in h as h^{-2s-2}. Not so good!

Improvement by a Change of Variable, in One Dimension

Suppose $g : [0,1] \to \mathbb{R}$ is monotone. Change of variable $w = (x - g(u))/h$.

In one dimension ($s = 1$), we have $\mathrm{d}w/\mathrm{d}u = -g'(u)/h$, so

$$V_{\mathrm{HK}}(\tilde{g}) = \frac{1}{h}\int_0^1 k'\left(\frac{x - g(u)}{h}\right)\left(\frac{-g'(u)}{h}\right)\mathrm{d}u = \frac{1}{h}\int_{-\infty}^{\infty} k'(w)\mathrm{d}w = \mathcal{O}(h^{-1}).$$

Then, if k and g are continuously differentiable, with RQMC points having $D^*(P_n) = \mathcal{O}(n^{-1+\epsilon})$, we obtain $\mathrm{AIV} = \mathcal{O}(n^{-2+\epsilon}h^{-2})$.

With $h = \Theta(n^{-1/3})$, this gives $\mathrm{AMISE} = \mathcal{O}(n^{-4/3})$.

A similar argument gives

$$V_2^2(\tilde{g}) = \frac{1}{h^2}\int_0^1 \left(k'\left(\frac{x - g(u)}{h}\right)\left(\frac{-g'(u)}{h}\right)\right)^2 \mathrm{d}u = \frac{1}{h^3}L_g\int_{-\infty}^{\infty}(k'(w))^2\mathrm{d}w = \mathcal{O}(h^{-3})$$

if $|g'| \le L_g$, and then with NUS: $\mathrm{AIV} = \mathcal{O}(n^{-3+\epsilon}h^{-3})$.

With $h = \Theta(n^{-3/7})$, this gives $\mathrm{AMISE} = \mathcal{O}(n^{-12/7})$.

Higher Dimensions

Let $s = 2$ and $\mathfrak{v} = \{1, 2\}$. With the change of variable $(u_1, u_2) \to (w, u_2)$, the Jacobian is $|\mathrm{d}w/\mathrm{d}u_1| = |g_1(u_1, u_2)/h|$, where $g_j = \partial g/\partial u_j$. If $|g_2|$ and $|g_{12}/g_1|$ are bounded by a constant L,

$$\begin{aligned}
\int_{[0,1)^2} \left| \frac{\partial^{|\mathfrak{v}|} \tilde{g}}{\partial \mathbf{u}_{\mathfrak{v}}} \right| \mathrm{d}\mathbf{u} &= \frac{1}{h} \int_{[0,1)^2} \left| \frac{\partial^2}{\partial u_1 \partial u_2} k\left(\frac{x - g(\mathbf{u})}{h} \right) \right| \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \frac{1}{h} \int_{[0,1)^2} \left| k''\left(\frac{x - g(\mathbf{u})}{h} \right) \frac{g_1(\mathbf{u})}{h} \frac{g_2(\mathbf{u})}{h} + k'\left(\frac{x - g(\mathbf{u})}{h} \right) \frac{g_{12}(\mathbf{u})}{h} \right| \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \frac{1}{h} \int_0^1 \int_{-\infty}^{\infty} \left| k''(w) \frac{g_2(\mathbf{u})}{h} + k'(w) \frac{g_{12}(\mathbf{u})}{g_1(\mathbf{u})} \right| \mathrm{d}w \, \mathrm{d}u_2 \\
&= \frac{L}{h} \left[\mu_0(k'')/h + \mu_0(k') \right] \;=\; \mathcal{O}(h^{-2}).
\end{aligned}$$

This provides a bound of $\mathcal{O}(h^{-2})$ for $V_{\mathrm{HK}}(\tilde{g})$, then $\mathrm{AIV} = \mathcal{O}(n^{-2+\epsilon} h^{-4})$.

Generalizing to $s \geq 2$ gives $V_{\mathrm{HK}}(\tilde{g}) = \mathcal{O}(h^{-s})$, $\mathrm{AIV} = \mathcal{O}(n^{-2+\epsilon} h^{-2s})$, $\mathrm{MISE} = \mathcal{O}(n^{-4/(2+s)})$.

Beats MC for $s < 3$, same rate for $s = 3$. Not very satisfactory. **But only a bound!**

Empirical Evaluation with Linear Model in a limited region

Regardless of the asymptotic bounds, the true IV may behave better than for MC for pairs (n, h) of interest. We consider the model

$$\text{MISE} = \text{IV} + \text{ISB} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha}.$$

This model is only for a limited region of interest, not for everywhere, not necessarily asymptotic. The optimal h for this model satisfies

$$h^{\alpha+\delta} = \frac{C\delta}{B\alpha} n^{-\beta}.$$

and it gives $\text{MISE} \approx Kn^{-\alpha\beta/(\alpha+\delta)}$.

We can take the asymptotic α (known) and B (estimated as for MC).

To estimate C, β, and δ, estimate the IV over a grid of values of (n, h), and fit a linear regression model:

$$\log \text{IV} \approx \log C - \beta \log n - \delta \log h.$$

Model estimation

For each (n, h), we estimate the IV by making n_r indep. replications of the RQMC density estimator, compute the variance at n_e evaluation points (stratified) over $[a, b]$, and multiply by $(b - a)/n$. We use logs in base 2, since n is a power of 2.

Validation

After estimating model parameters, we can test out-of-sample with independent simulation experiments at pairs (n, h) with $h = \hat{h}_*(n)$.

For test cases in which density is known, to assess what RQMC can achieve, we can compute a MISE estimate at those pairs (n, h), and obtain new parameter estimates $\tilde{K}$ and $\tilde{\nu}$ of model $\mathrm{MISE} \approx K n^{-\nu}$.

Numerical illustrations

For each example, we first estimate model parameters by regression using a grid of pairs (n, h) with $n = 2^{14}, 2^{15}, \ldots, 2^{19}$ and (for KDE) $h = h_0, \ldots, h_5$ with $h_j = h_0 2^{j/2} = 2^{-\ell_0 + j/2}$.

For each n and each RQMC method, we make $n_r = 100$ independent replications and take $n_e = 64$ evaluation points over bounded interval $[a, b]$. Also tried larger n_e.

RQMC Point sets:

- MC: Independent points (Crude Monte Carlo),
- Stratification: stratified unit cube,
- LMS: Sobol' points with left matrix scrambling (LMS) + digital random shift,
- NUS: Sobol' points with NUS.

Simple test example with standard normal density

Let $Z_1, \ldots, Z_s$ i.i.d. standard normal generated by inversion, and

$$X = \frac{Z_1 + \cdots + Z_s}{\sqrt{s}}.$$

Then $X \sim \mathcal{N}(0, 1)$.

Here we can estimate IV, ISB, and MISE accurately.
We can compute $\int_a^b f''(x)\mathrm{d}x$ exactly.

We take $[a, b] = [-2, 2]$.

Estimates of model parameters for KDE

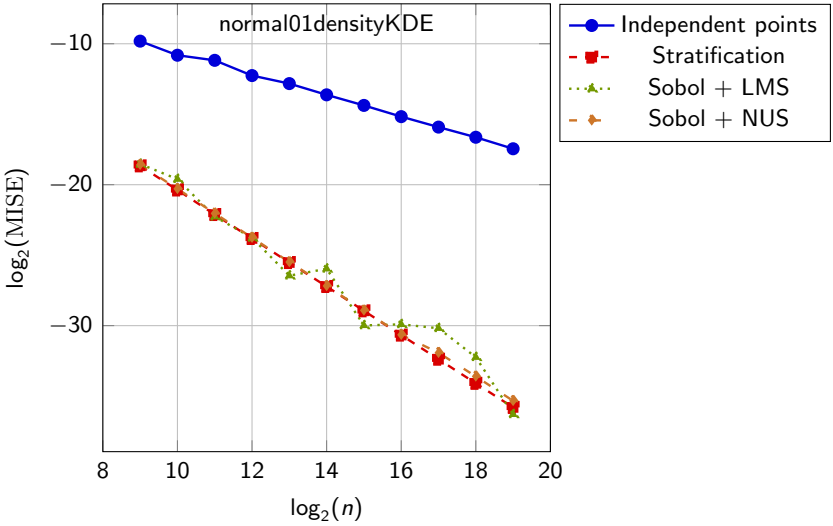
$$\mathrm{ISB} = Bh^{\alpha}, \qquad \mathrm{IV} \approx Cn^{-\beta}h^{-\delta}, \qquad \mathrm{MISE} \approx \kappa n^{-\nu}$$

We have $B = 0.04754$ with $\alpha = 4$.

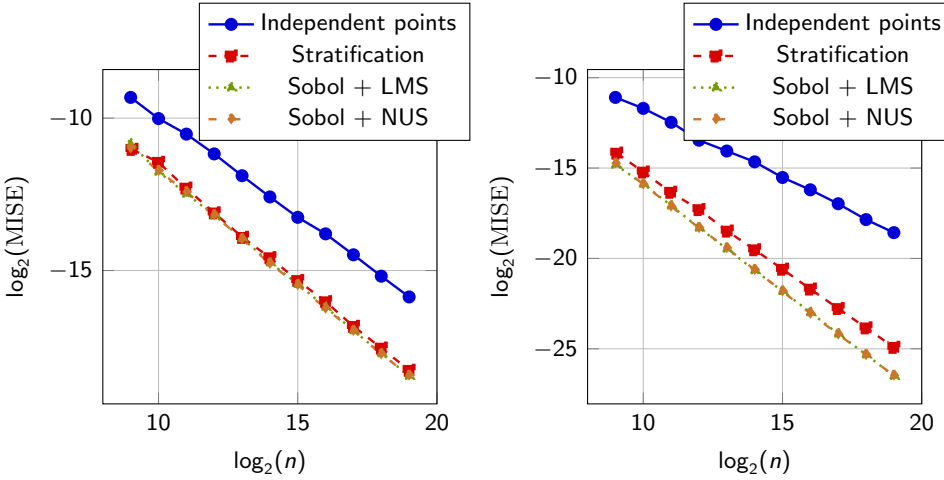
method	MC	Sobol + NUS				
s	1	1	2	3	5	20
R^2	0.999	0.999	1.000	0.995	0.979	0.993
β	1.017	2.787	2.110	1.788	1.288	1.026
δ	1.144	2.997	3.195	3.356	2.293	1.450
α	3.758	3.798	3.846	3.860	3.782	3.870
$\tilde{\nu}$	0.770	1.600	1.172	0.981	0.827	0.730
LGM	16.96	34.05	24.37	20.80	17.91	17.07

$\mathrm{LGM} = -\log_2(\mathrm{MISE})$ for $n = 2^{19}$.

Convergence of the MISE in log-log scale, for the one-dimensional example



Convergence of the MISE, for $s = 2$, for histograms (left) and KDE (right).



Displacement of a cantilever beam (Bingham 2017)

Displacement D of a cantilever beam with horizontal load X and vertical load Y:

$$D = \frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}}$$

where $L = 100$, $w = 4$, $t = 2$ (in inches), X, Y, and E are independent and normally distributed with means and standard deviations:

Description	Symbol	Mean	St. dev.
Young's modulus	E	2.9×10^7	1.45×10^6
Horizontal load	X	500	100
Vertical load	Y	1000	100

We want to estimate the density of D over $[a, b] = [0.336, 1.561]$ (about 99.5% of density).

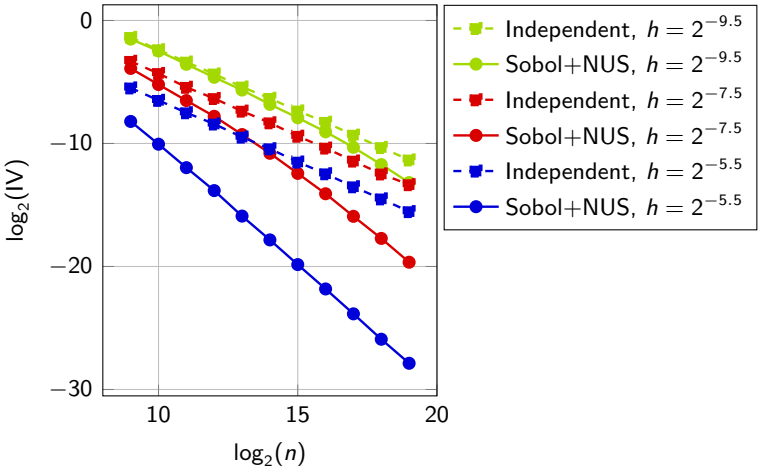
Parameter estimates of the linear regression model for IV and MISE:

$$\mathrm{IV} \approx C n^{-\beta} h^{-\delta}, \qquad \mathrm{MISE} \approx \kappa n^{-\nu}$$

Point set	$\hat{C}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\nu}$
KDE with Gaussian kernel, $\alpha = 4$				
Independent	0.210	0.993	1.037	0.789
Sobol+LMS	5.28E-4	1.619	2.949	0.932
Sobol+NUS	5.24E-4	1.621	2.955	0.932

Good fit: we have $R^2 > 0.99$ in all cases.

$\log_2(\mathrm{IV})$ vs $\log_2 n$ for cantilever with KDE.



A weighted sum of lognormals

$$X = \sum_{j=1}^{s} w_j \exp(Y_j)$$

where $\mathbf{Y} = (Y_1, \ldots, Y_s)^t \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

Let $\mathbf{C} = \mathbf{A}\mathbf{A}^t$. To generate $\mathbf{Y}$, generate $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and put $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$.

We will use principal component decomposition (PCA).

This has several applications. In one of them, with $w_j = s_0(s - j + 1)/s$, $e^{-\rho}\max(X - K, 0)$ is the payoff of a financial option based on an average price at s observation times, under a GBM process. Want to estimate density of positive payoffs.

A weighted sum of lognormals

$$X = \sum_{j=1}^{s} w_j \exp(Y_j)$$

where $\mathbf{Y} = (Y_1, \ldots, Y_s)^t \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

Let $\mathbf{C} = \mathbf{A}\mathbf{A}^t$. To generate $\mathbf{Y}$, generate $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and put $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$.

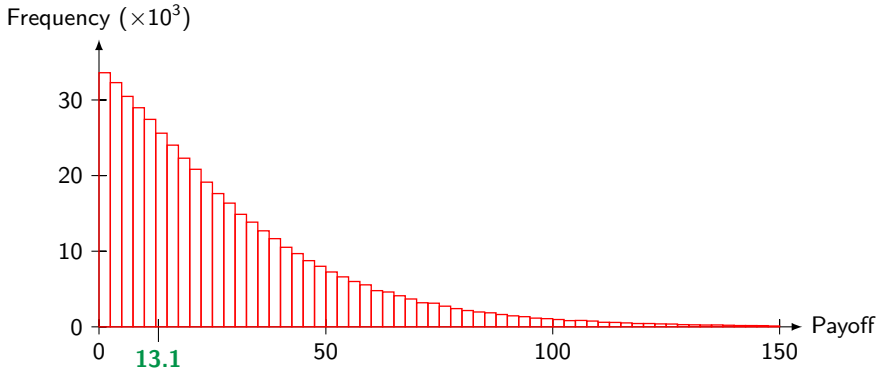
We will use principal component decomposition (PCA).

This has several applications. In one of them, with $w_j = s_0(s - j + 1)/s$, $e^{-\rho}\max(X - K, 0)$ is the payoff of a financial option based on an average price at s observation times, under a GBM process. Want to estimate density of positive payoffs.

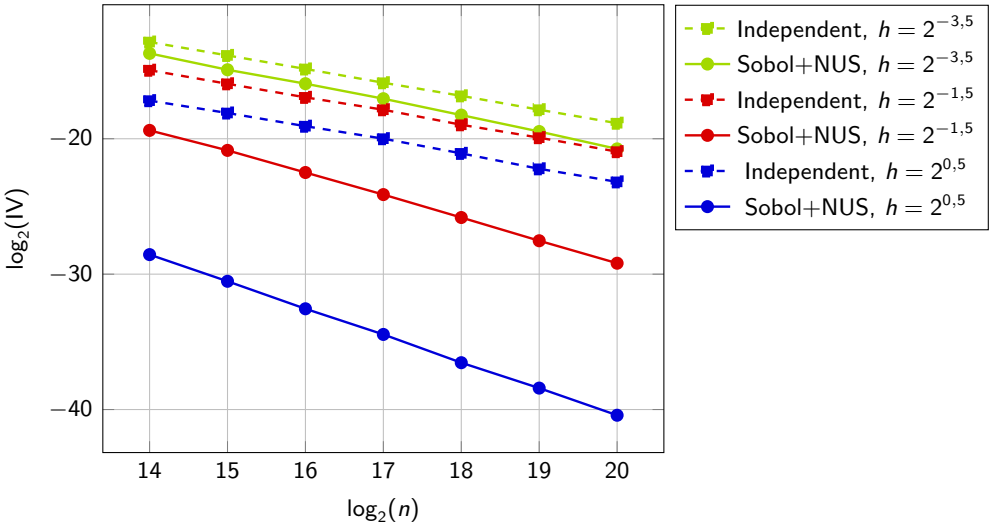
Numerical experiment: Take $s = 12$, $\rho = 0.037$, $s_0 = 100$, $K = 101$, and $\mathbf{C}$ defined indirectly via: $Y_j = Y_{j-1}(\mu - \sigma^2)j/s + \sigma B(j/s)$ where $Y_0 = 0$, $\sigma = 0.12136$, $\mu = 0.1$, and B a standard Brownian motion.

We will estimate the density of $e^{-\rho}(X - K)$ over $[a, b] = [0, 50]$.

Histogram of positive values from $n = 10^6$ independent simulation runs:



IV as a function of n for KDE.



The RQMC bound in $s = 12$ dimension gives a worst rate than MC, but we observe a better actual IV and MISE.

Stochastic derivative of an estimator of F?

The density is $f(x) = F'(x)$, so perhaps we can estimate $f(x)$ by the derivative w.r.t. x 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 $\mathrm{d}\hat{F}_n(x)/\mathrm{d}x = 0$ 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 the indicator $\mathbb{I}[X_i \leq x]$ by its conditional cdf given filtered (reduced) information $\mathcal{G}$:

$$F(x \mid \mathcal{G}) := \mathbb{P}[X \leq x \mid \mathcal{G}]$$

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

Assumption 1. For all realizations of $\mathcal{G}$, $F(x \mid \mathcal{G})$ is a continuous function of x which is differentiable except perhaps over a denumerable set of points, and for which the derivative $F'(x \mid \mathcal{G}) = \mathrm{d}F'(x \mid \mathcal{G})/\mathrm{d}x$ (when it exists) is bounded uniformly in x by some random variable Γ having finite variance.

Theorem: $F'(x \mid \mathcal{G})$ is an unbiased estimator of $f(x)$, with variance bounded uniformly in x.

Overall estimator with n iid replicates: $\hat{f}_{\mathrm{cmc},n}(x) = \frac{1}{n}\sum_{i=1}^{n} F'(x \mid \mathcal{G}_i)$.

With this estimator, $\mathrm{ISB} = 0$, so $\mathrm{MISE} = \mathrm{IV}$, which is rather easy to estimate.

Proposed by Asmussen (2018) for a sum of random variables.

Application to the normalized sum of standard normals

We had

$$X = \frac{Z_1 + Z_2 + \cdots + Z_s}{\sqrt{s}}$$

where each $Z_j \sim \mathcal{N}(0,1)$. For CMC, we can leave out Z_s, so $\mathcal{G} = (Z_1, \ldots, Z_{s-1})$ and

$$\begin{aligned} F(x \mid \mathcal{G}) &= \mathbb{P}[X \le x \mid Z_1, \ldots, Z_{s-1}] = \mathbb{P}[Z_s \le x\sqrt{s} - (Z_1 + Z_2 + \cdots + Z_{s-1})] \\ &= \Phi(x\sqrt{s} - (Z_1 + Z_2 + \cdots + Z_{s-1})). \end{aligned}$$

The resulting density estimator is

$$F'(x \mid \mathcal{G}) = \phi(x\sqrt{s} - (Z_1 + Z_2 + \cdots + Z_{s-1}))\ \sqrt{s}.$$

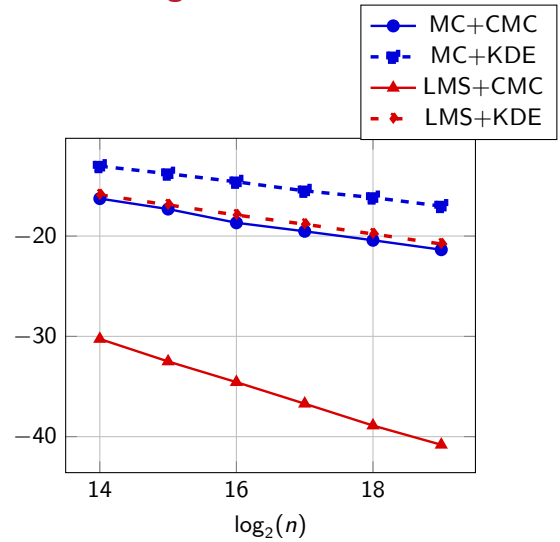
Assumption 1 is easily verified.

		$d = 3$			$d = 20$		
		MC	Sobol+LMS	Sobol+NUS	MC	Sobol+LMS	Sobol+NUS
CMC	ν	1.019	2.116	2.094	0.988	0.961	0.982
	LGM	21.36	40.81	40.65	19.27	19.58	19.54
KDE	ν	0.798	0.976	0.975	0.769	0.771	0.760
	LGM	17.01	20.79	20.80	17.00	17.10	17.07

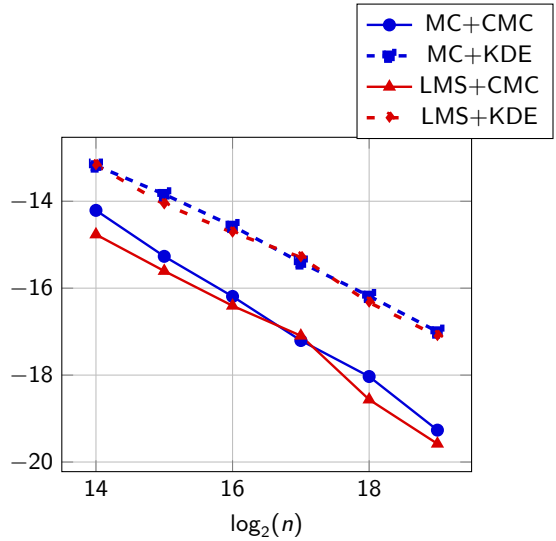
LGM $= -\log_2(\mathrm{MISE})$ for $n = 2^{19}$.

MISE convergence

$s = 3$



$s = 20$



Cantilever beam

$$D = \frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}}$$

where E, X, Y are normal r.v.'s. Want density over $[a, b] = [0.407, 1.515]$.

For CMC, we can leave out E, i.e., take $\mathcal{G} = (X, Y)$. Then,

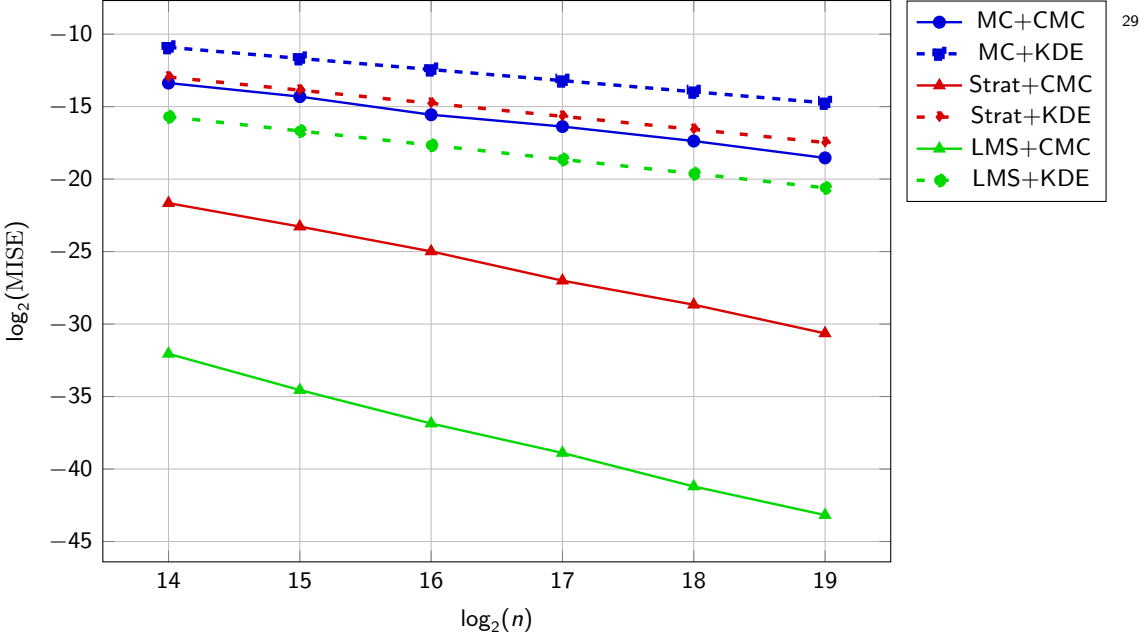
$$F(d \mid \mathcal{G}) = \mathbb{P}[D \le d \mid X, Y] = \mathbb{P}\left[\frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} \le d \mid X, Y\right] = 1 - \Phi\left(\frac{\frac{4L^3}{dwt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} - \mu_E}{\sigma_E}\right).$$

Taking the derivative w.r.t. d, we get

$$F'(d \mid \mathcal{G}) = \phi\left(\frac{\frac{4L^3}{dwt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} - \mu_E}{\sigma_E}\right) \times \frac{4L^3}{d^2 wt\sigma_E}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}}.$$

Estimated $\mathrm{MISE} = Kn^{-\nu}$ and $\mathrm{LGM} = -\log_2 \mathrm{MISE}$ for $n = 2^{19}$:

		MC	Strat	LMS	NUS
CMC	ν	1.02	1.80	2.22	2.16
	LGM	18.54	30.64	43.18	43.08
KDE	ν	0.81	0.89	0.94	0.96
	LGM	15.12	18.08	21.33	21.35

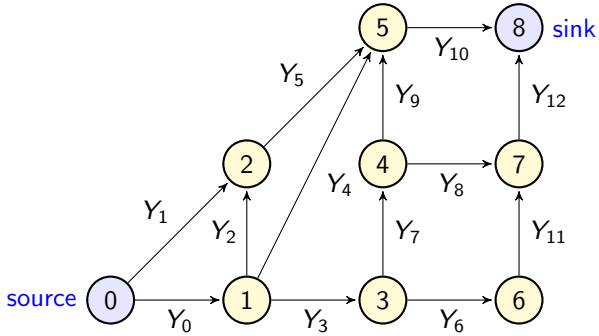

MC+CMC
MC+KDE
Strat+CMC
Strat+KDE
LMS+CMC
LMS+KDE
−10
−15
−20
−25
−30
−35
−40
−45
log2(MISE)
14
15
16
17
18
19
log2(n)

Example: A stochastic activity network

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

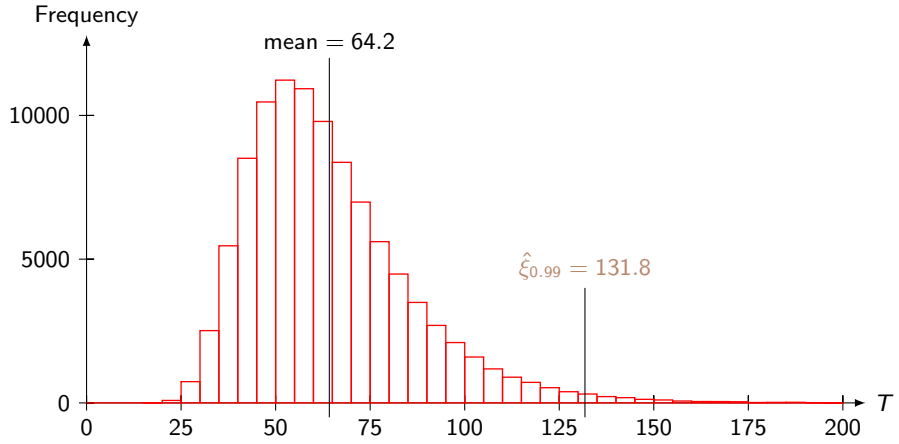
Project duration $T =$ (random) length of longest path from source to sink.

We want to estimate the density of T.



Numerical illustration from Elmaghraby (1977):
$Y_k \sim N(\mu_k, \sigma_k^2)$ for $k = 0, 1, 3, 10, 11$, and $V_k \sim \text{Expon}(1/\mu_k)$ otherwise.
$\mu_0, \ldots, \mu_{12}$: 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 $n = 100\,000$.



The SAN example, Sobol+NUS vs Independent points, summary for $n = 2^{19} = 524288$.

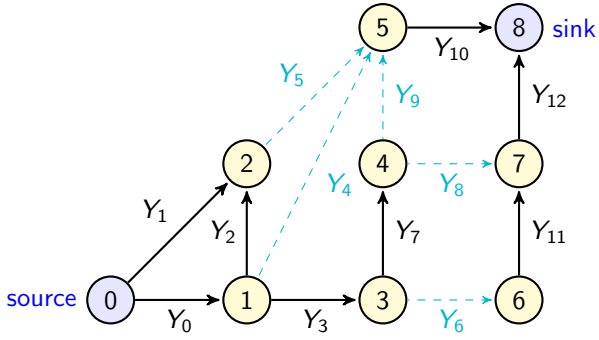
Density	m or h	Independent points		Sobol+NUS	
		$log_2 IV$	IV rate	$log_2 IV$	IV rate
	0.10	-16.64	-0.999	-16.71	-1.006
	0.18	-17.96	-0.999	-18.18	-1.015
	0.32	-19.33	-0.998	-19.79	-1.035
KDE	0.43	-19.99	-0.998	-20.71	-1.064

CMC for the SAN Example

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

CMC estimator of $\mathbb{P}[T \le t]$: $F(t \mid \mathcal{G}) = \mathbb{P}[T \le t \mid \{Y_j, j \notin \mathcal{L}\}]$

where $\mathcal{L} = \{4, 5, 6, 8, 9\}$ and $Y_j = F_j^{-1}(U_j)$. This estimator continuous in the U_j's and in t.



To compute $F(t \mid \mathcal{G})$: for each $l \in \mathcal{L}$, say from a_l to b_l, compute the length α_l of the longest path from 1 to a_l, and the length β_l of the longest path from b_l to the destination.

The longest path that passes through link l does not exceed t iff $\alpha_l + Y_l + \beta_l \leq t$, which occurs with probability $\mathbb{P}[Y_l \leq t - \alpha_l - \beta_l] = F_l[t - \alpha_l - \beta_l]$.

To compute $F(t \mid \mathcal{G})$: for each $l \in \mathcal{L}$, say from a_l to b_l, compute the length α_l of the longest path from 1 to a_l, and the length β_l of the longest path from b_l to the destination.

The longest path that passes through link l does not exceed t iff $\alpha_l + Y_l + \beta_l \leq t$, which occurs with probability $\mathbb{P}[Y_l \leq t - \alpha_l - \beta_l] = F_l[t - \alpha_l - \beta_l]$.

Since the Y_l are independent, we obtain

$$F(t \mid \mathcal{G}) = \prod_{l \in \mathcal{L}} F_l[t - \alpha_l - \beta_l].$$

To estimate the density of T, take the derivative w.r.t. t:

$$F'(t \mid \mathcal{G}) = \frac{\mathrm{d}}{\mathrm{d}t} F(t \mid \mathcal{G}) \overset{\text{w.p.1}}{=} \sum_{j \in \mathcal{L}} f_j[t - \alpha_j - \beta_j] \prod_{l \in \mathcal{L}, l \neq j} F_l[t - \alpha_l - \beta_l].$$

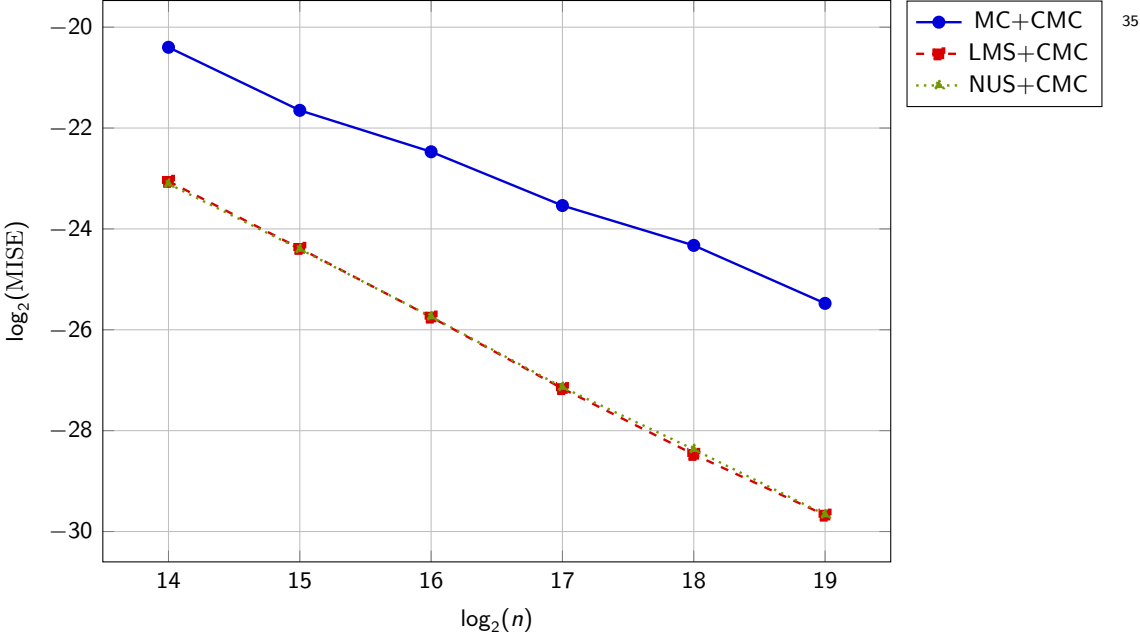
Assumption 1 holds if the F_j are smooth enough, and then $\mathbb{E}[F'(t \mid \mathcal{G})] = f_T(t)$.

With MC, the IV converges as $\mathcal{O}(1/n)$ and there is no bias, so MISE = IV.

With RQMC, we observe a convergence rate near $\mathcal{O}(n^{-4/3})$ for the IV and the MISE.

Estimated $\mathrm{MISE} = Kn^{-\nu}$ and $\mathrm{LGM} = -\log_2 \mathrm{MISE}$ for $n = 2^{19}$:

		MC	LMS	NUS
CMC	ν	0.99	1.34	1.32
	LGM	25.48	29.67	29.66


MC+CMC
LMS+CMC
NUS+CMC
−20
−22
−24
−26
−28
−30
log2(MISE)
14
15
16
17
18
19
log2(n)

Other examples

- Option pricing under GBM and under a VG process.
- Density estimation for a function of the state of a Markov chain, using Array-RQMC.
- More to come.

Conclusion

- Both CMC and RQMC can improve the convergence rate of the IV and MISE when estimating a density.
- With KDEs, the convergence rates observed in small examples are much better than the bounds proved from standard QMC theory.
  There are opportunities for QMC theoreticians here!
- The combination of CMC with RQMC for density estimation is very promising!
  Lots of potential applications!

Some references

- S. Asmussen. Conditional Monte Carlo for sums, with applications to insurance and finance, Annals of Actuarial Science, prepublication, 1–24, 2018.
- 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.
- A. B. Owen. Scrambled Net Variance for Integrals of Smooth Functions. Annals of Statistics, 25 (4):1541–1562, 1997.
- D. W. Scott. Multivariate Density Estimation. Wiley, 2015.