

# Density estimation by Randomized Quasi-Monte Carlo

Pierre L'Ecuyer



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

**SAMSI**, Raleigh-Durham, North-Carolina, May 2018

## What this talk is about

Quasi-Monte Carlo (QMC) and randomized QMC (RQMC) methods have been studied extensively for [estimating an integral](#),  $\mathbb{E}[X]$ .

Can they be useful for [estimating the entire distribution of  \$X\$](#) ?

E.g., estimating a density, a cdf, some quantiles, etc.

When hours or days of computing time are required to perform simulation runs, reporting only a confidence interval on the mean is a waste of information!

People routinely 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?

## Setting

Classical **density estimation** was developed in the context where independent observations  $X_1, \dots, X_n$  are **given** and one wishes to estimate the **density**  $f$  from which they come.

Here we assume that  $X_1, \dots, X_n$  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  $\mathbf{U} = (U_1, \dots, U_s) \sim U(0, 1)^s$  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:

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

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

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

## 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  $h > 0$  (serves as horizontal stretching factor for the kernel). The KDE is defined by

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

# Asymptotic convergence with Monte Carlo for smooth $f$

For  $g : \mathbb{R} \rightarrow \mathbb{R}$ , 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  $h \rightarrow 0$ :

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

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

	$C$	$B$	$\alpha$	$h^*$	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  $\mathbf{U}_1, \dots, \mathbf{U}_n$  by RQMC points.

RQMC does not change the bias.

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

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

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  $\mathbf{U}_1, \dots, \mathbf{U}_n$  by RQMC points.

RQMC does not change the bias.

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

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

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 may have something like

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

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

## Elementary 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

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

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

There are explicit point sets for which  $D^*(P_n) = \mathcal{O}((\log n)^{s-1}/n) = \mathcal{O}(n^{-1+\epsilon})$ .

Explicit RQMC constructions for which  $\mathbb{E}[E_n] = 0$  and  $\text{Var}[E_n] = \mathcal{O}(n^{-2+\epsilon})$ .

Also

$$|E_n| \leq V_2(g)D_2(P_n)$$

where

$$V_2^2(g) = \sum_{\emptyset \neq \mathbf{v} \subseteq \mathcal{S}} \int_{[0,1]^s} \left| \frac{\partial^{|\mathbf{v}|} g}{\partial \mathbf{v}} \right|^2 d\mathbf{u}, \quad (\text{square } L_2 \text{ variation}),$$

$$D_2^2(P_n) = \int_{\mathbf{u} \in [0,1]^s} \left| \text{vol}[\mathbf{0}, \mathbf{u}] - \frac{|P_n \cap [\mathbf{0}, \mathbf{u}]|}{n} \right|^2 d\mathbf{u} \quad (\text{square } L_2\text{-star-discrepancy}).$$

Explicit constructions for which  $D_2(P_n) = \mathcal{O}(n^{-1+\epsilon})$ .

Moreover, if  $P_n$  is a digital net randomized by a **nested uniform scramble** (NUS) and  $V_2(g) < \infty$ , then  $\mathbb{E}[E_n] = 0$  and  $\text{Var}[E_n] = \mathcal{O}(V_2^2(g)n^{-3+\epsilon})$  for all  $\epsilon > 0$ .

## 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  $\mathbb{E}[\tilde{g}(\mathbf{U})] = \int_{[0,1]^s} \tilde{g}(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f_n(x)]$ .

RQMC **does not change the bias**, but **may reduce**  $\text{Var}[\hat{f}_n(x)]$ , 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  $\mathbb{E}[\tilde{g}(\mathbf{U})] = \int_{[0,1]^s} \tilde{g}(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f_n(x)]$ .

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

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

The partial derivatives are:

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

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  $j = 2, \dots, |\mathbf{v}| + 1$ .

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

## Improvement by a Change of Variable, in One Dimension

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

In one dimension ( $s = 1$ ), we have  $dw/du = -g'(u)/h$ , so

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

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

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

A similar argument gives

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

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

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

## Higher Dimensions

Let  $s = 2$  and  $\mathbf{v} = \{1, 2\}$ . With the change of variable  $(u_1, u_2) \rightarrow (w, u_2)$ , the Jacobian is  $|dw/du_1| = |g_1(u_1, u_2)/h|$ , where  $g_j = \partial g/\partial u_j$ . If  $|g_2|$  and  $|g_{12}/g_1|$  are bounded by a constant  $L$ ,

$$\begin{aligned} \int_{[0,1]^2} \left| \frac{\partial^{|\mathbf{v}|} \tilde{g}}{\partial \mathbf{u}_{\mathbf{v}}} \right| d\mathbf{u} &= \frac{1}{h} \int_{[0,1]^2} \left| \frac{\partial^2}{\partial u_1 \partial u_2} k \left( \frac{x - g(\mathbf{u})}{h} \right) \right| du_1 du_2 \\ &= \frac{1}{h} \int_{[0,1]^2} \left| k'' \left( \frac{x - g(\mathbf{u})}{h} \right) \frac{g_1(\mathbf{u})}{h} \frac{g_2(\mathbf{u})}{h} + k' \left( \frac{x - g(\mathbf{u})}{h} \right) \frac{g_{12}(\mathbf{u})}{h} \right| du_1 du_2 \\ &= \frac{1}{h} \int_0^1 \int_{-\infty}^{\infty} \left| k''(w) \frac{g_2(\mathbf{u})}{h} + k'(w) \frac{g_{12}(\mathbf{u})}{g_1(\mathbf{u})} \right| dw du_2 \\ &= \frac{L}{h} [\mu_0(k'')/h + \mu_0(k')] = \mathcal{O}(h^{-2}). \end{aligned}$$

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

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

Beats MC for  $s < 3$ , same rate for  $s = 3$ . Not very satisfactory.

## 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

$$\text{MISE} = \text{IV} + \text{ISB} \approx C n^{-\beta} h^{-\delta} + B h^{\alpha}.$$

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  $\text{MISE} \approx K n^{-\alpha\beta/(\alpha+\delta)}$ .

We can take the asymptotic  $\alpha$  (known) and  $B$  (estimated as for MC).

To estimate  $C$ ,  $\beta$ , and  $\delta$ , estimate the IV over a grid of values of  $(n, h)$ , and fit a linear regression model:

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

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.

After estimating model parameters, 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, can compute a MISE estimate at each point, and obtain new parameter estimates  $\tilde{K}$  and  $\tilde{\nu}$  of model  $\text{MISE} \approx Kn^{-\nu}$ . Not useful to estimate an unknown density, but useful to assess what RQMC could achieve.

## Numerical illustrations

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

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:

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

## Simple test example with standard normal density

Let  $Z_1, \dots, Z_s$  i.i.d. standard normal generated by inversion, and  $X = (Z_1 + \dots + Z_s)/\sqrt{s}$ . Then  $X \sim \mathcal{N}(0, 1)$ .

Here we can estimate IV, ISB, and MISE accurately.

We can compute  $\int_a^b f''(x)dx$  exactly.

Take  $(a, b) = (-2, 2)$ . We have  $B = 0.04754$  with  $\alpha = 4$  for KDE.

## Estimates of model parameters for KDE

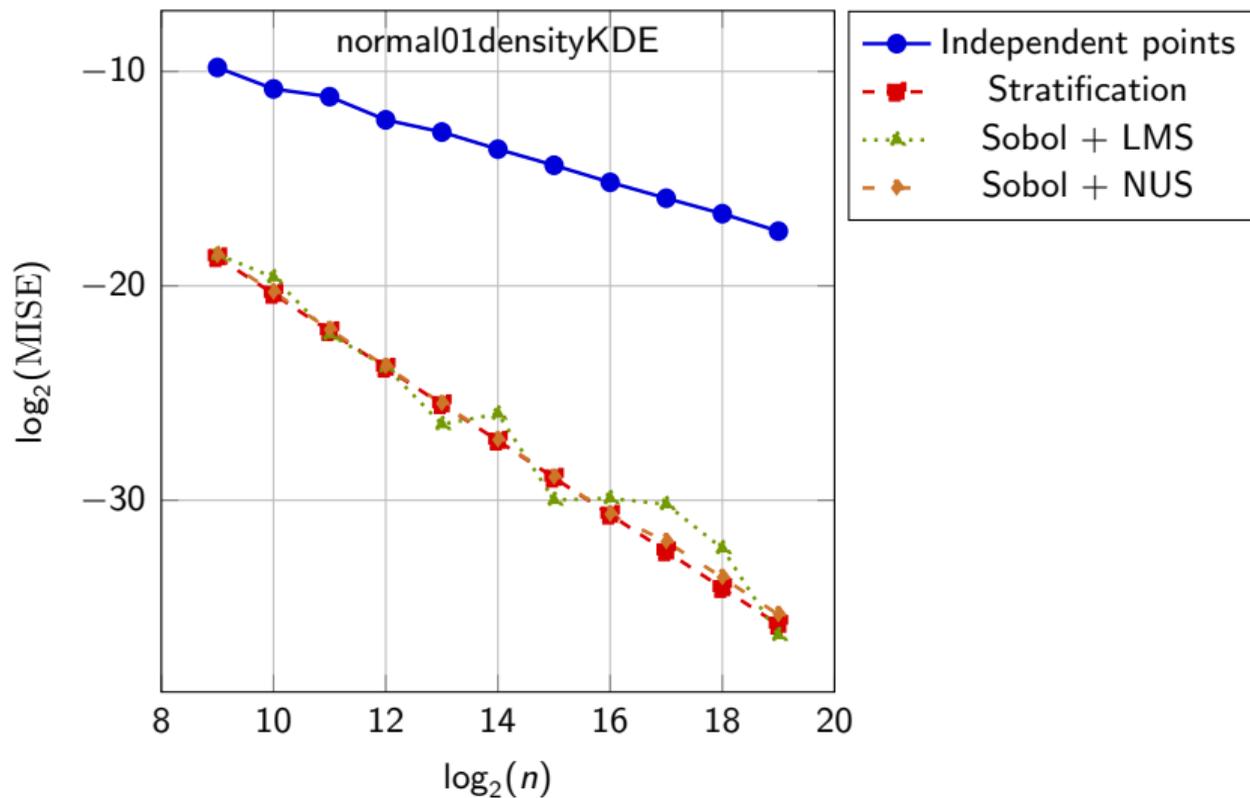
$$IV \approx Cn^{-\beta}h^{-\delta},$$

$$MISE \approx \kappa n^{-\nu}$$

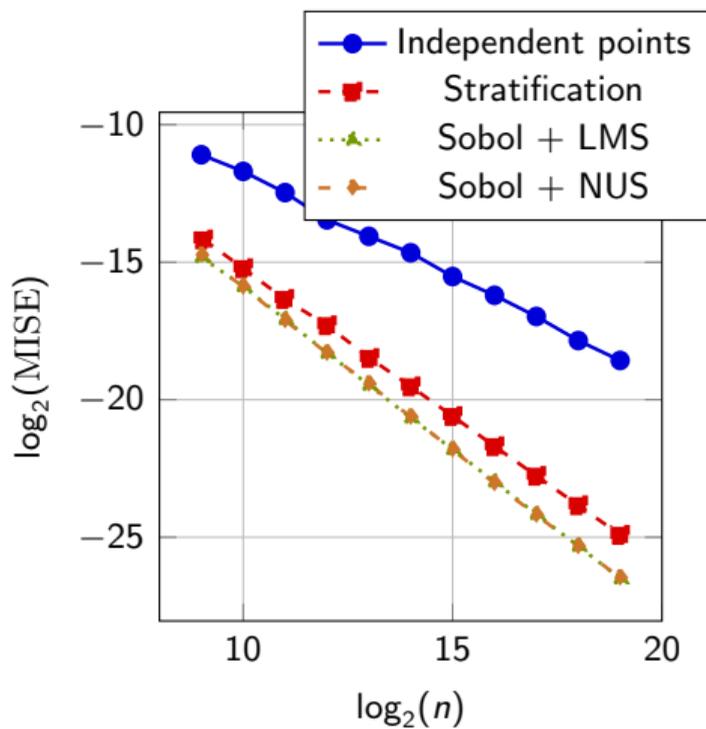
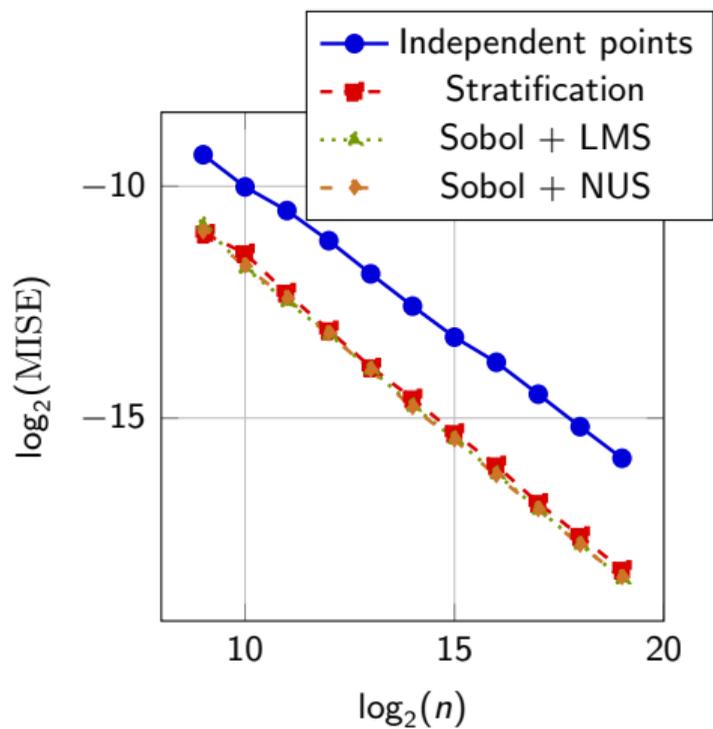
method	MC	NUS				
s	1	1	2	3	5	20
$R^2$	0.999	0.999	1.000	0.995	0.979	0.993
$\beta$	1.017	2.787	2.110	1.788	1.288	1.026
$\delta$	1.144	2.997	3.195	3.356	2.293	1.450
$\alpha$	3.758	3.798	3.846	3.860	3.782	3.870
$\tilde{\nu}$	0.770	1.600	1.172	0.981	0.827	0.730
LGM	16.96	34.05	24.37	20.80	17.91	17.07

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

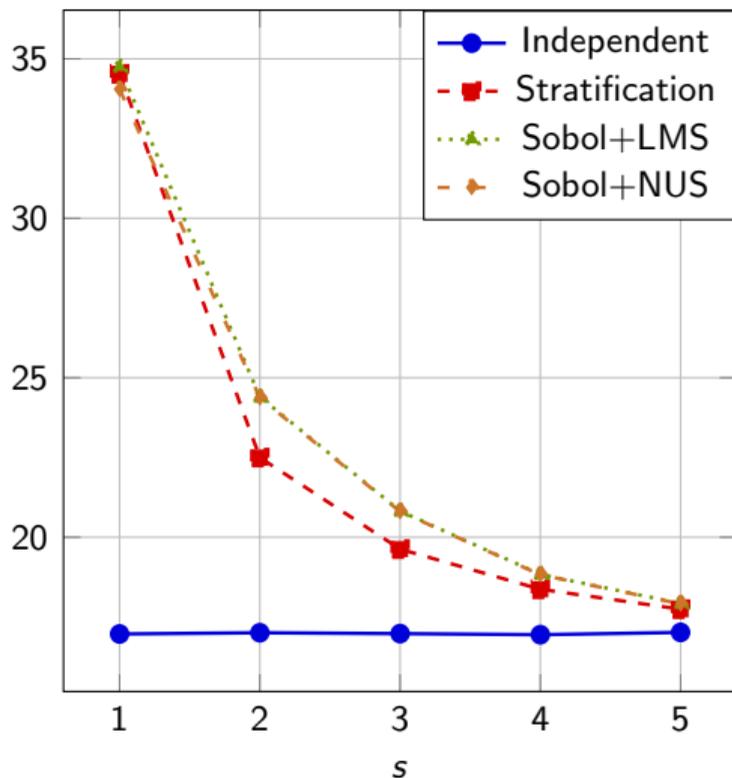
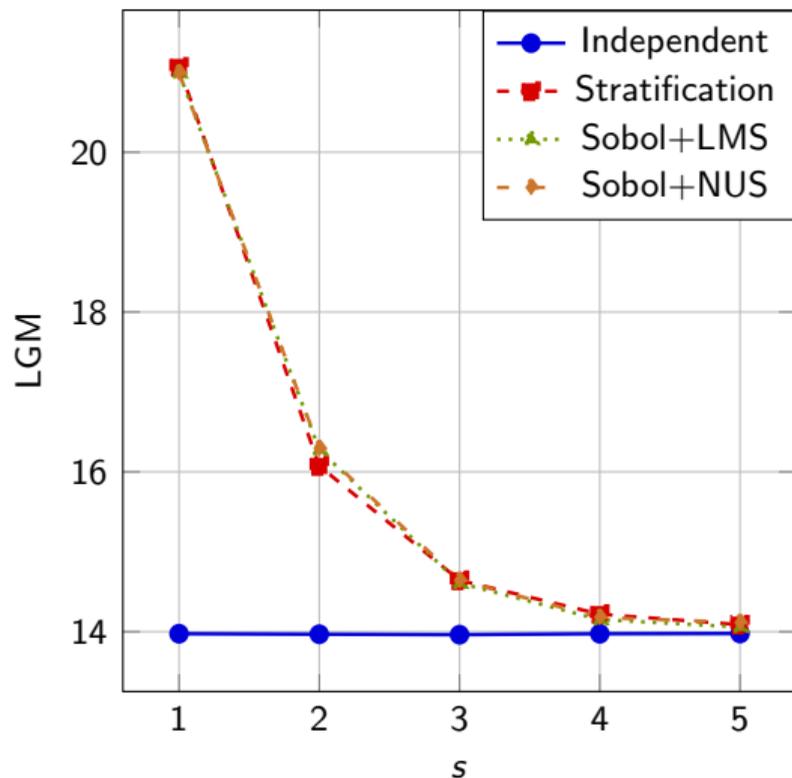
## Convergence of the MISE in log-log scale, for the one-dimensional example



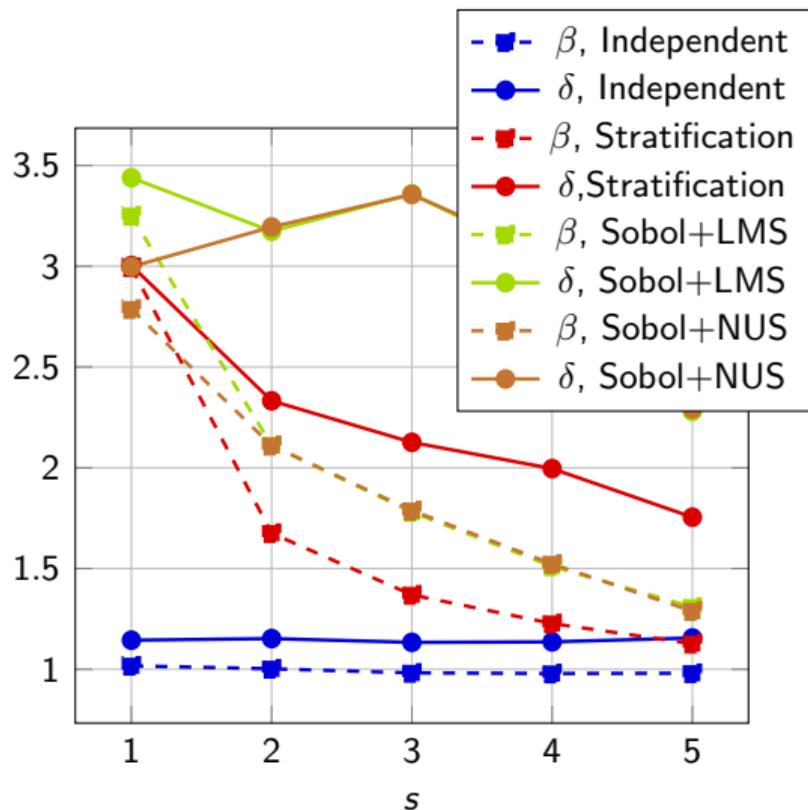
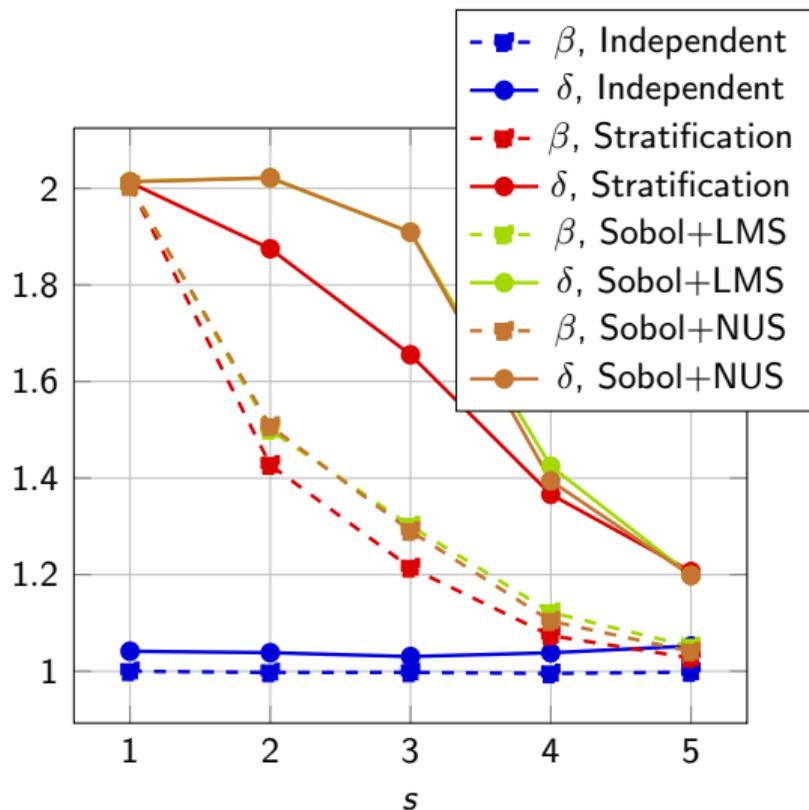
Convergence of the MISE, for  $s = 2$ , for histograms (left) and KDE (right).



LGM ( $n = 2^{19}$ ) for histogram (left) and KDE (right) for estimation over  $(-2, 2)$ .



Estimated parameters with histogram (left) and KDE (right) over  $(-2, 2)$ .

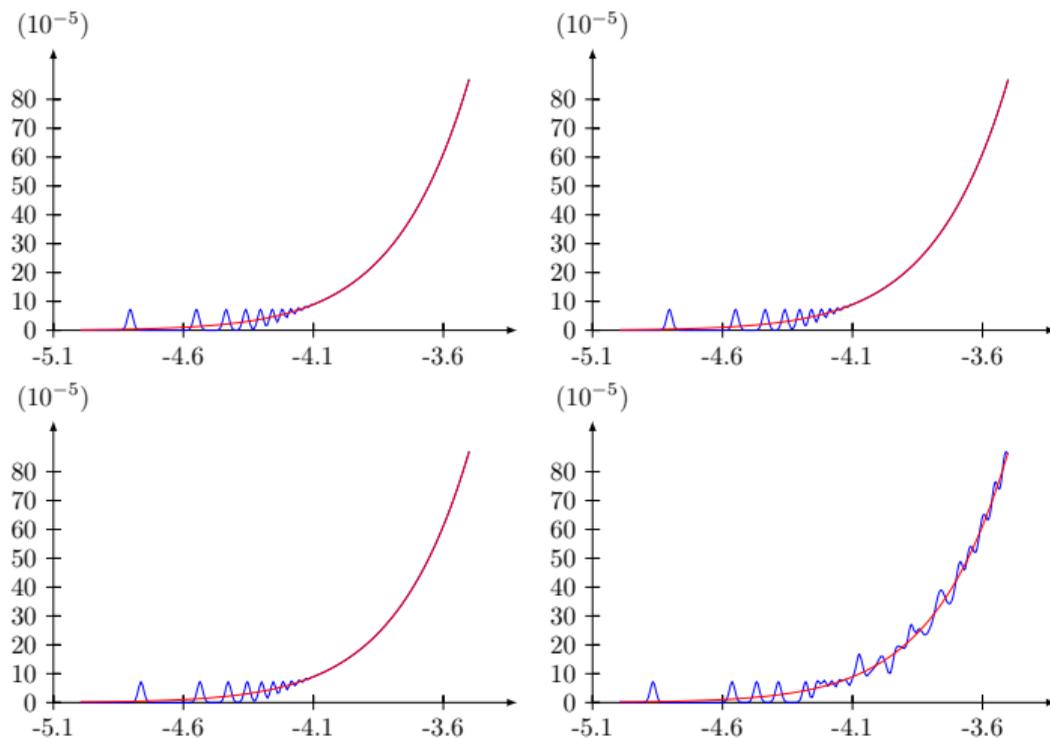


# KDE Estimate over $(-4, 4)$

method	MC	NUS				
		1	2	3	5	20
$s$	1	1	2	3	5	20
$\beta$	1.020	2.434	1.999	1.728	1.272	1.006
$\delta$	1.138	2.432	2.972	3.168	2.256	1.464
$\tilde{\nu}$	0.772	1.514	1.138	0.980	0.817	0.767
LGM	16.89	30.07	23.68	20.52	17.72	16.95

## KDE Estimate in the tail

KDE (blue) vs true density (red) with RQMC point sets with  $n = 2^{19}$ :  
 lattice + shift, Sobol + digital shift, Sobol + LMS-19bits + shift, Sobol + LMS-31bits + shift



## Displacement of a cantilevel beam

Displacement  $D$  of a cantilever beam with horizontal and vertical loads:

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

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	$E$	$2.9 \times 10^7$	$1.45 \times 10^6$
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).

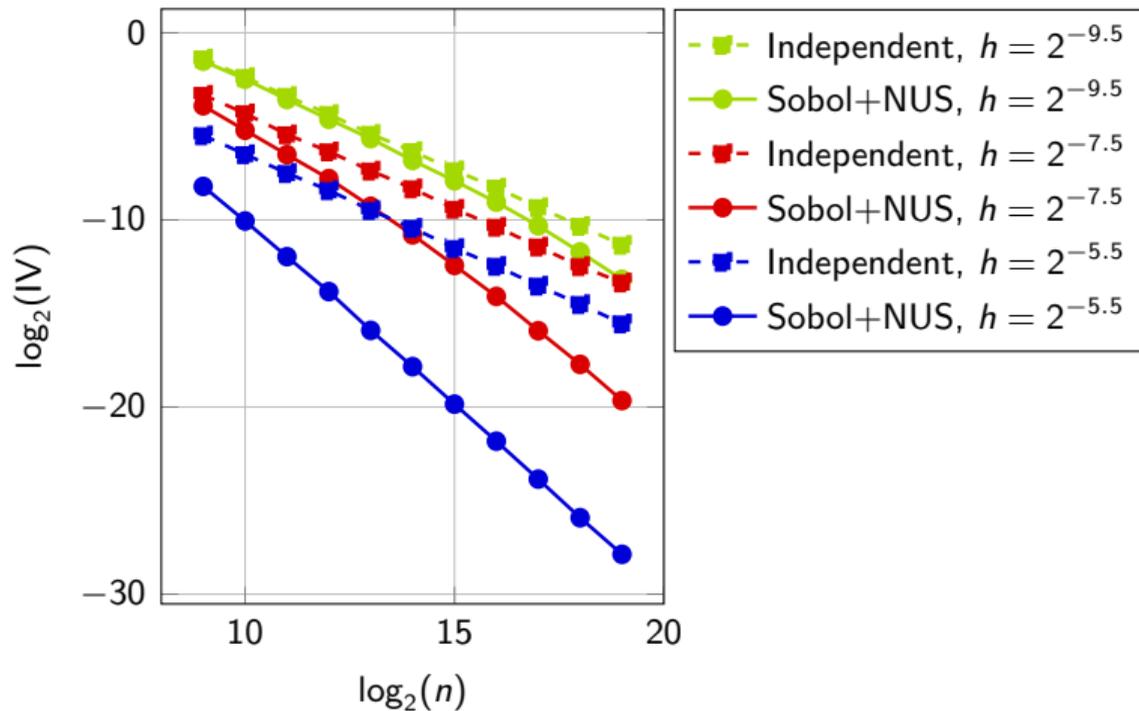
Parameter estimates of the linear regression model for IV and MISE:

$$IV \approx Cn^{-\beta}h^{-\delta}, \quad MISE \approx \kappa n^{-\nu}$$

Point set	$\hat{C}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\nu}$
Histogram, $\alpha = 2$				
Independent	0.888	1.001	1.021	0.797
Sobol+LMS	0.134	1.196	1.641	0.848
Sobol+NUS	0.136	1.194	1.633	0.848
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.

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



## A weighted sum of lognormals

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

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

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

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

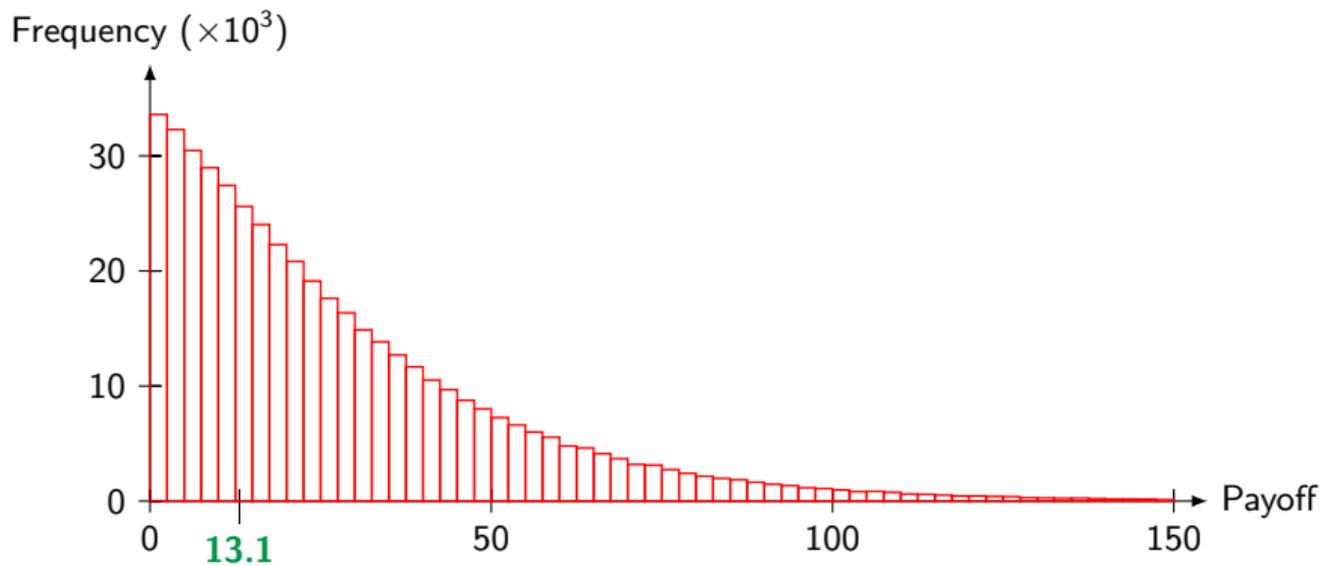
We will use principal component decomposition (PCA).

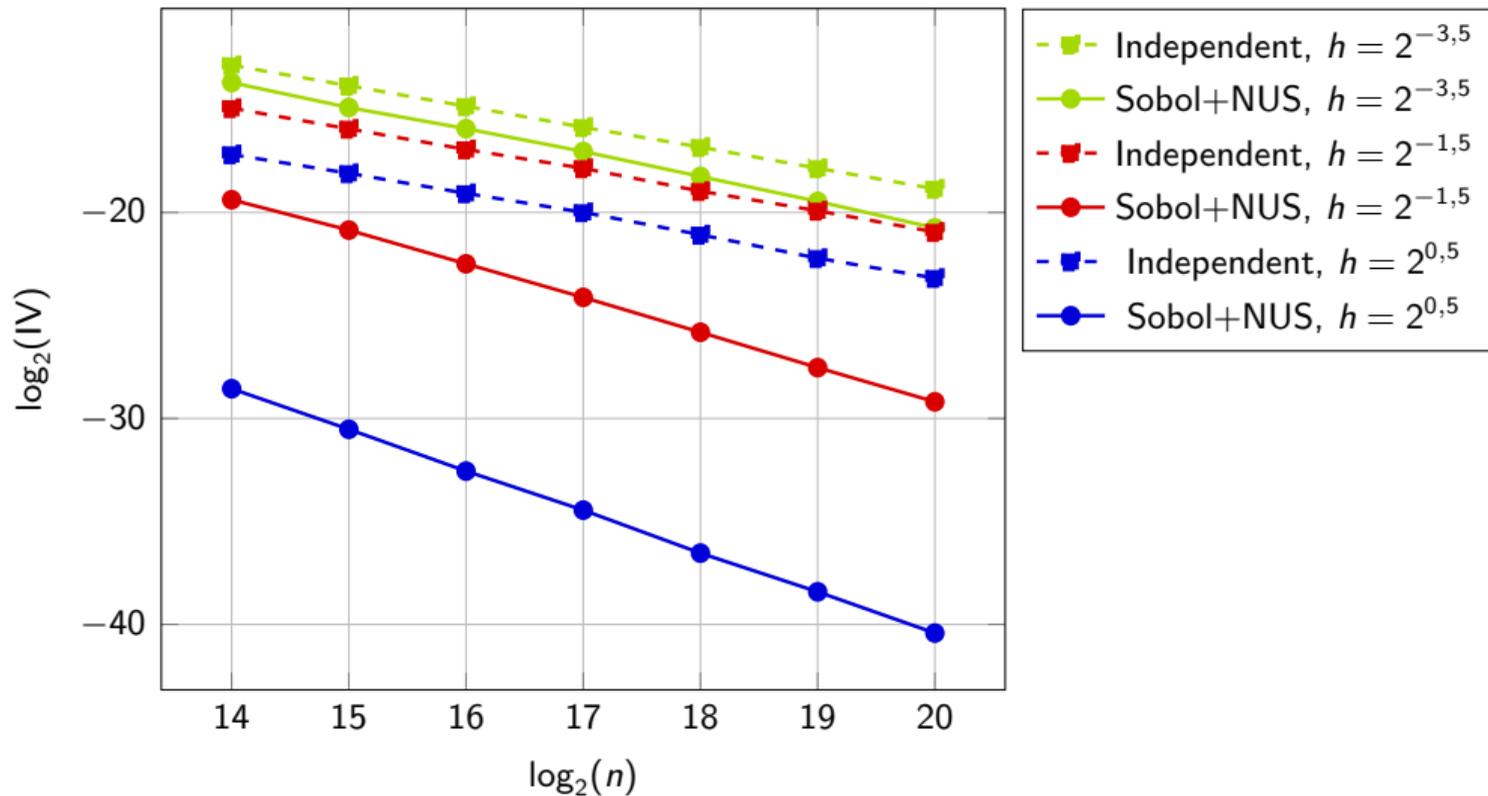
This has several applications. In one of them, with  $w_j = s_0(s - j + 1)/s$ ,  $e^{-\rho} \max(X - K, 0)$  is the [payoff of a financial option](#) based on an average price at  $s$  observation times, under a GBM process. Want to estimate density of positive payoffs.

**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:



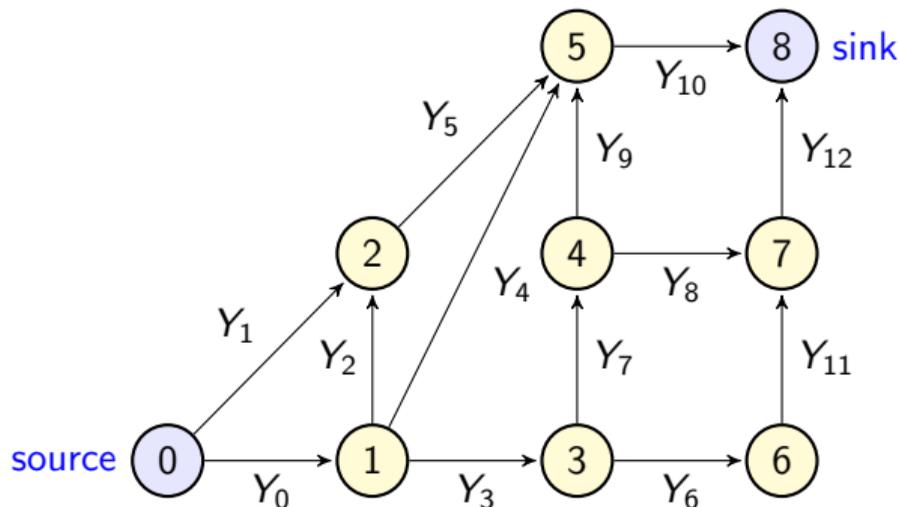


## Example: A stochastic activity network

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.

May want to estimate  $\mathbb{E}[T]$ ,  $\mathbb{P}[T > x]$ , a quantile, density of  $T$ , etc.



## Simulation

**Algorithm: to generate  $T$ :**

for  $k = 0, \dots, 12$  do

    Generate  $U_k \sim U(0, 1)$  and let  $Y_k = F_k^{-1}(U_k)$

    Compute  $X = T = h(Y_0, \dots, Y_{12}) = f(U_0, \dots, U_{12})$

**Monte Carlo:** Repeat  $n$  times independently to obtain  $n$  realizations  $X_1, \dots, X_n$  of  $T$ .

Estimate  $\mathbb{E}[T] = \int_{(0,1)^s} f(\mathbf{u})d\mathbf{u}$  by  $\bar{X}_n = \frac{1}{n} \sum_{i=0}^{n-1} X_i$ .

To estimate  $\mathbb{P}(T > x)$ , take  $X = \mathbb{I}[T > x]$  instead.

**RQMC:** Replace the  $n$  independent points by an RQMC point set of size  $n$ .

**Numerical illustration** from Elmaghraby (1977):

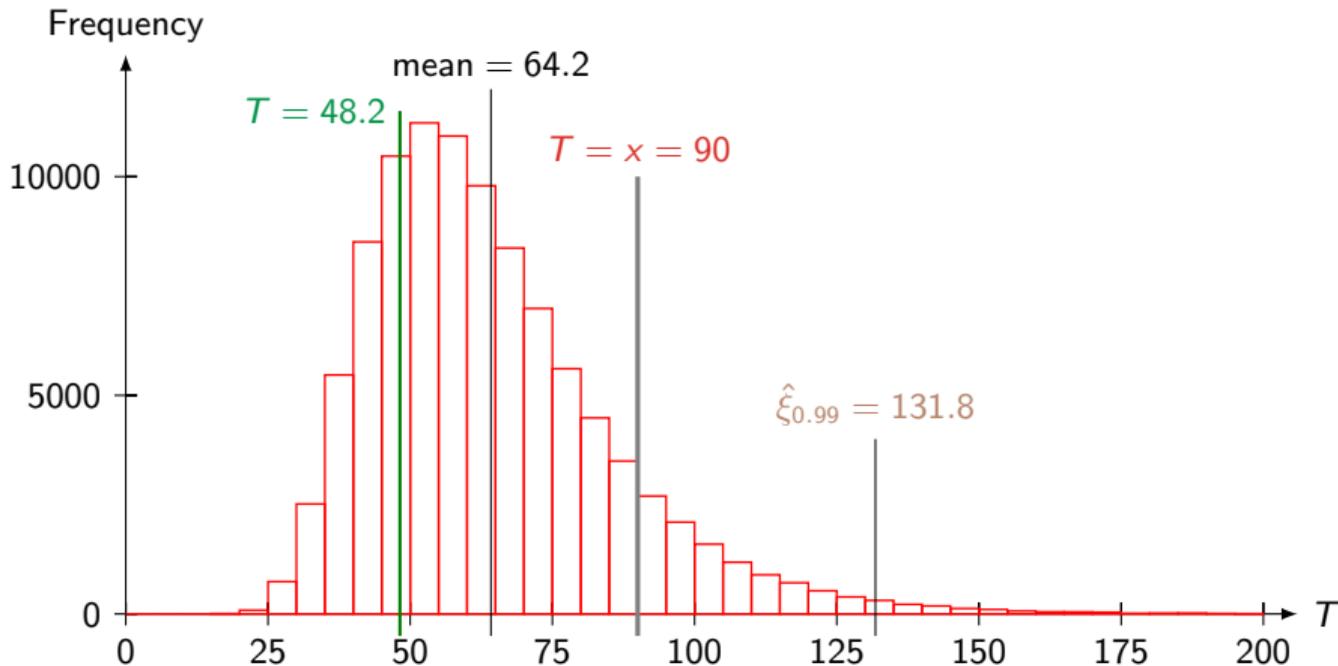
$Y_k \sim N(\mu_k, \sigma_k^2)$  for  $k = 0, 1, 3, 10, 11$ , and  $V_k \sim \text{Expon}(1/\mu_k)$  otherwise.

$\mu_0, \dots, \mu_{12}$ : 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.

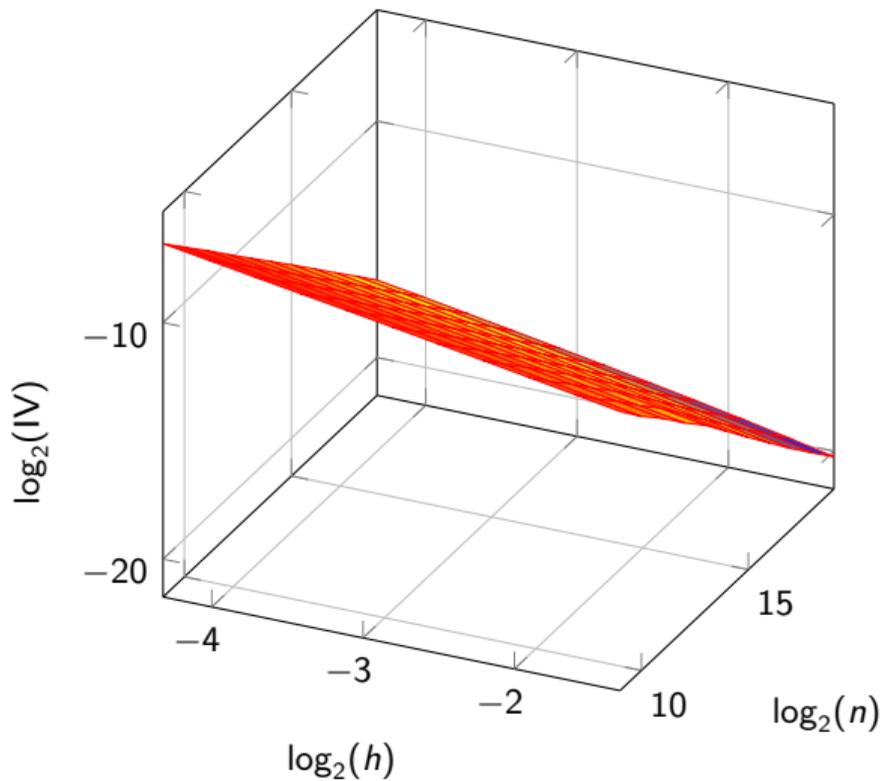
Naive idea: replace each  $Y_k$  by its expectation. Gives  $T = 48.2$ .

Results of an experiment with  $n = 100\,000$ .

Histogram of values of  $T$  is a density estimator that gives more information than a confidence interval on  $\mathbb{E}[T]$  or  $\mathbb{P}[T > x]$ .



$\log_2(\text{IV})$  with KDE, Sobol+NUS, for the SAN



## Results for SAN Network

Parameter estimates of the regression models for the SAN network,  $n = 2^{19}$ .

Point set	$C$	$\beta$	$\delta$	$\hat{\gamma}_*$	$\hat{\nu}^*$
Histogram, $\alpha = 2$					
Independent	0.892	0.999	1.005	0.333	0.665
Stratif.	0.897	1.001	1.006	0.333	0.666
Lattice+shift	0.841	0.988	0.988	0.331	0.662
Sobol+LMS	0.894	1.006	1.024	0.333	0.665
Sobol+NUS	0.888	1.005	1.026	0.332	0.665
KDE with Gaussian kernel, $\alpha = 4$					
Independent	0.254	1.001	1.004	0.199	0.799
Stratif.	0.248	1.001	1.012	0.199	0.799
Lattice+shift	0.222	0.969	0.947	0.196	0.784
Sobol+LMS	0.242	1.021	1.089	0.201	0.803
Sobol+NUS	0.246	1.023	1.087	0.201	0.804

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
Histogram	64	-19.32	-1.003	-19.78	-1.039
	256	-17.28	-0.999	-17.40	-1.011
	1024	-15.27	-1.001	-15.30	-1.003
	4096	-13.27	-0.998	-13.27	-1.000
Kernel	0.10	-16.64	-0.999	-16.71	-1.006
	0.13	-17.31	-0.999	-17.42	-1.007
	0.18	-17.96	-0.999	-18.18	-1.015
	0.24	-18.64	-0.999	-18.92	-1.029
	0.32	-19.33	-0.998	-19.79	-1.035
	0.43	-19.99	-0.998	-20.71	-1.064

## Conditional Monte Carlo for Derivative Estimation

The density is  $f(x) = F'(x)$ , so could we just take the derivative of a cdf estimator?

The derivative of the empirical cdf  $\hat{F}_n(x)$  is zero almost everywhere, ... does not work!

We need a **smooth cdf estimator**.

## Conditional Monte Carlo for Derivative Estimation

The density is  $f(x) = F'(x)$ , so could we just take the derivative of a cdf estimator?

The derivative of the empirical cdf  $\hat{F}_n(x)$  is zero almost everywhere, ... does not work!

We need a **smooth cdf estimator**.

Let  $X = X(\theta, \omega)$  with parameter  $\theta \in \mathbb{R}$ . Want to estimate  $g'(\theta) = \frac{d}{d\theta} \mathbb{E}[X(\theta, \omega)]$ .

When  $X(\cdot, \omega)$  is not continuous in  $\theta$  (+ other conditions), we cannot interchange the derivative and expectation, and **cannot** take  $\frac{d}{d\theta} X(\theta, \omega)$  as an estimator of  $g'(\theta)$ .

## Conditional Monte Carlo for Derivative Estimation

The density is  $f(x) = F'(x)$ , so could we just take the derivative of a cdf estimator?

The derivative of the empirical cdf  $\hat{F}_n(x)$  is zero almost everywhere, ... does not work!

We need a **smooth cdf estimator**.

Let  $X = X(\theta, \omega)$  with parameter  $\theta \in \mathbb{R}$ . Want to estimate  $g'(\theta) = \frac{d}{d\theta} \mathbb{E}[X(\theta, \omega)]$ .

When  $X(\cdot, \omega)$  is not continuous in  $\theta$  (+ other conditions), we cannot interchange the derivative and expectation, and **cannot** take  $\frac{d}{d\theta} X(\theta, \omega)$  as an estimator of  $g'(\theta)$ .

Often possible to replace  $X$  by a conditional expectation  $X_e = \mathbb{E}[X \mid \mathcal{G}]$  where  $\mathcal{G}$  contains partial information, not enough to reveal  $X$ , but enough to compute  $X_e$ .

If  $X_e$  is smooth enough in  $\theta$ , we may have

$$g'(\theta) = \mathbb{E} \left[ \frac{d}{d\theta} X_e(\theta, \omega) \right] = \mathbb{E}[X'_e].$$

This is gradient estimation by IPA + CMC. L'Ecuyer and Perron (1994), Asmussen (2017).

Then, we can simulate  $X'_e$  with RQMC instead of MC.

## Application to the SAN Example

We want a smooth estimate of  $\mathbb{P}[T \leq t]$ , whose sample derivative will be an unbiased estimate of the density at  $t$ .

**Naive estimator:** Generate  $T$  and compute  $X = \mathbb{I}[T \leq t]$ .

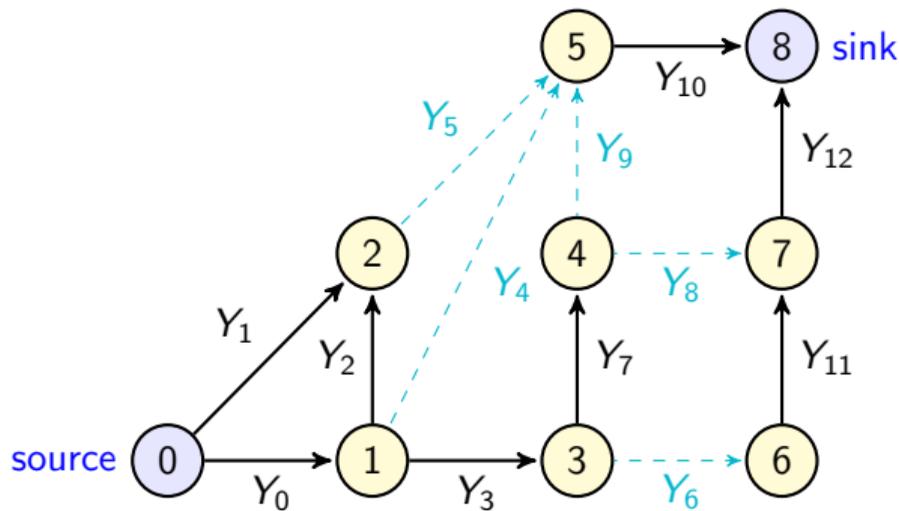
Repeat  $n$  times and average.

## Application to the SAN Example

We want a smooth estimate of  $\mathbb{P}[T \leq t]$ , whose sample derivative will be an unbiased estimate of the density at  $t$ .

**Naive estimator:** Generate  $T$  and compute  $X = \mathbb{I}[T \leq t]$ .

Repeat  $n$  times and average.



**Conditional Monte Carlo estimator of  $\mathbb{P}[T \leq t]$ .** Generate the  $Y_j$ 's only for the 8 arcs that **do not** belong to the cut  $\mathcal{L} = \{4, 5, 6, 8, 9\}$ , and replace  $\mathbb{I}[T \leq t]$  by its **conditional expectation** given those  $Y_j$ 's,

$$X_e = \mathbb{P}[T \leq t \mid \{Y_j, j \notin \mathcal{L}\}].$$

This makes the integrand **continuous in the  $U_j$ 's and in  $t$ .**

**Conditional Monte Carlo estimator of  $\mathbb{P}[T \leq t]$ .** Generate the  $Y_j$ 's only for the 8 arcs that **do not** belong to the cut  $\mathcal{L} = \{4, 5, 6, 8, 9\}$ , and replace  $\mathbb{I}[T \leq t]$  by its **conditional expectation** given those  $Y_j$ 's,

$$X_e = \mathbb{P}[T \leq t \mid \{Y_j, j \notin \mathcal{L}\}].$$

This makes the integrand **continuous in the  $U_j$ 's and in  $t$ .**

**To compute  $X_e$ :** for each  $l \in \mathcal{L}$ , say from  $a_l$  to  $b_l$ , compute the length  $\alpha_l$  of the longest path from 1 to  $a_l$ , and the length  $\beta_l$  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  $\mathbb{P}[Y_l \leq t - \alpha_l - \beta_l] = F_l[t - \alpha_l - \beta_l]$ .

**Conditional Monte Carlo estimator of  $\mathbb{P}[T \leq t]$ .** Generate the  $Y_j$ 's only for the 8 arcs that **do not** belong to the cut  $\mathcal{L} = \{4, 5, 6, 8, 9\}$ , and replace  $\mathbb{I}[T \leq t]$  by its **conditional expectation** given those  $Y_j$ 's,

$$X_e = \mathbb{P}[T \leq t \mid \{Y_j, j \notin \mathcal{L}\}].$$

This makes the integrand **continuous in the  $U_j$ 's and in  $t$** .

**To compute  $X_e$ :** for each  $l \in \mathcal{L}$ , say from  $a_l$  to  $b_l$ , compute the length  $\alpha_l$  of the longest path from 1 to  $a_l$ , and the length  $\beta_l$  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

$$X_e = \prod_{l \in \mathcal{L}} F_l[t - \alpha_l - \beta_l].$$

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

$$X'_e = \frac{d}{dt} X_e(t, \omega) \stackrel{\text{w.p.1}}{=} \sum_{j \in \mathcal{L}} f_j[t - \alpha_j - \beta_j] \prod_{l \in \mathcal{L}, l \neq j} F_l[t - \alpha_l - \beta_l].$$

One can prove that

$$\mathbb{E}[X'_e] = \frac{d}{dt} \mathbb{E}[X_e] = \frac{d}{dt} \mathbb{P}[T \leq t] = f_T(t)$$

via the dominated convergence theorem. See L'Ecuyer (1990).

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

Now, we can apply RQMC to simulate  $X'_e$ . It is a smooth function of the uniforms if each inverse cdf  $F_j^{-1}$  and density  $f_j$  are smooth.

Then we can get a convergence rate near  $\mathcal{O}(n^{-2})$  for the IV and the MISE.

## Conclusion

- ▶ We saw that RQMC can improve the convergence rate of the IV and MISE when estimating a density.
- ▶ With histograms and KDEs, the convergence rates observed in small examples are much better than those that we have proved based on standard QMC theory. There are opportunities for QMC theoreticians here.
- ▶ This also applies in the context of Array-RQMC for Markov chains.
- ▶ The combination of CMC with QMC for density estimation is very promising! Lots of potential applications! We are working on this.

- ▶ 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. In preparation, 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, C. Lécot, and B. Tuffin. A randomized quasi-Monte Carlo simulation method for Markov chains. *Operations Research*, 56(4):958–975, 2008.
- ▶ 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<sup>5</sup> (4):1541–1562, 1997.
- ▶ V. C. Raykar and R. Duraiswami. Fast optimal bandwidth selection for kernel density estimation. *Proceedings of the 2006 SIAM International Conference on Data Mining*. 524–528, 2006.
- ▶ D. W. Scott. *Multivariate Density Estimation*. Wiley, 2015.