DENSITY ESTIMATION BY RQMC

Florian Puchhammer
July 02, 2018
Université de Montréal, Canada

Joint work with **Amal Ben Abdellah, Pierre L'Ecuyer, and Art B. Owen**

Are we wasting information?

Traditionally, (R)QMC is used to estimate an integral $\mathbb{E}[X]$ and to provide a confidence interval.

Question: With all the collected sample data, isn't it a waste of information to "only" estimate the mean?

Answer: It is! We can even estimate the distribution of X.

Crude MC: Several methods are already known (e.g. Histogram, kernel density estimator, conditional density estimator).

Goal: Find a way to apply RQMC in density estimation so that

- the benefits of RQMC are incorporated,
- it outperforms current MC strategies.

Setting

Classical: Independent observations $X_1, \ldots, X_n$ of X are given.

Here: The random variable X is given by $X = g(\boldsymbol{U})$ with

- $g : (0,1)^d \to \mathbb{R}$,
- $\boldsymbol{U} = (U_1, \ldots, U_d) \sim \mathcal{U}(0,1)^d$,
- $g(\boldsymbol{u})$ is easy to compute,
- $X_1, \ldots, X_n$ are generated by simulation.

Estimators

We will always estimate the density f of the real r.v. X over a finite interval $[a, b]$ by a density estimator $\hat{f}_n$.

Histogram: Partition $[a, b]$ into m bins $I_1, \ldots, I_m$ of size $h = (b - a)/m$ and set

$$\hat{f}_n(x) = \frac{n_j}{nh}, \qquad x \in I_j, 1 \leq j \leq m.$$

where n_j is the number of observations X_i that fall into I_j.

Kernel density estimator (KDE): Select kernel function k (Gaussian) and bandwidth h and define

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

Error measures

We consider the mean integrated square error

$$\text{MISE} = \int_a^b \mathbb{E}[\hat{f}_n(x) - f(x)]^2 \, \mathrm{d}x.$$

The MISE can be decomposed into

$$\text{MISE} = \text{IV} + \text{ISB},$$

where IV is the integrated variance

$$\text{IV} = \int_a^b \text{Var}[\hat{f}_n(x)] \, \mathrm{d}x,$$

and ISB is the integrated square bias

$$\text{ISB} = \int_a^b \Big(\mathbb{E}[\hat{f}_n(x)] - f(x) \Big)^2 \, \mathrm{d}x.$$

Asymptotics for Monte Carlo

Idea: Use MC sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n \sim \mathcal{U}(0,1)^d$ to compute $\hat{f}_n$.

When $n \to \infty$, $h \to 0$, and $nh \to \infty$ we can write

$$\mathrm{AMISE} = \mathrm{AIV} + \mathrm{AISB} \approx \frac{C}{nh} + Bh^{\alpha}, \qquad \alpha > 0.$$

$\rightarrow$ **variance-bias tradeoff.**

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

We have (Scott 2015):

- C depends on k,
- $\alpha = 2$ for a histogram and $\alpha = 4$ for a KDE, and
- B can be estimated (plug-in-methods, see Raykar, Duraiswami 2006.)

The optimal bandwidth h_* is

$$h_* = h_*(n) = (C/B\alpha n)^{1/(\alpha+1)}\,.$$

Consequently,

$$\mathrm{AMISE} \approx K n^{-\alpha/(1+\alpha)},$$

i.e. $\mathcal{O}(n^{-2/3})$ for histogram, $\mathcal{O}(n^{-4/5})$ for KDE.

Asymptotics for RQMC

Idea: Replace $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n \sim \mathcal{U}(0,1)^d$ by RQMC points.

The bias does not change,

$$\text{AISB} = Bh^{\alpha}.$$

Hope: Strength of RQMC is variance reduction. For KDE with smooth k we hope to get

$$\text{AIV} = Cn^{-\beta}h^{-1}, \qquad \beta > 1.$$

$\longrightarrow$ can choose larger $h_*(n)$ in variance-bias tradeoff.

Reality: Not exactly...

In theory (and reality) the power of h decreases too:

$$\mathrm{AIV} = \int_a^b \mathrm{Var}(\hat{f}_n(x))\,\mathrm{d}x \approx C n^{-\beta} h^{-\delta}, \qquad \beta > 1, \delta > 1.$$

Strategy: Rewrite the KDE as

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

Idea: Use Koksma–Hlawka type inequalities to bound

$$\mathrm{Var}(\hat{f}_n(x)) \le \mathbb{E}[D_n^2(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_d)]\, V^2(\tilde{g}(\cdot, x)).$$

Problem: faster convergence of discrepancy $\rightarrow$ larger variation $\rightarrow$ larger δ.

Let V_{HK} be the Hardy–Krause variation. Naive bound: $V_{HK}^2(\tilde{g}) = \mathcal{O}(h^{-2d-2})$.

Theorem (Asymptotics for KDE)

Under reasonably mild conditions on g and k we have

$$\int_a^b V_{HK}^2(\tilde{g}(\cdot, x))\,\mathrm{d}x = \mathcal{O}(h^{-2d})$$

If the star discrepancy $D_n^*(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n) = \mathcal{O}(n^{-1+\varepsilon})$ then

$$\mathrm{AIV} = \mathcal{O}(n^{-2+\varepsilon} h^{-2d}).$$

Moreover, $h_* = \Theta(n^{-1/(2+d)})$ and, consequently,

$$\mathrm{AMISE} = \mathcal{O}(n^{-4/(2+d)}).$$

Remarks (KDE)

For D_n^* and V_{HK} we can get

$$\mathrm{AIV} = \mathcal{O}(n^{-2+\varepsilon} h^{-2d}), \qquad \mathrm{AMISE} = \mathcal{O}(n^{-4/(2+d)}).$$

- In $d = 1$ with nested uniform scrambling (NUS) we can get

$$\mathrm{AIV} = \mathcal{O}(n^{-3+\varepsilon} h^{-3}), \qquad \mathrm{AMISE} = \mathcal{O}(n^{-12/7}).$$

- With stratified sampling we can get rid of d in the exponent of h.

$$\mathrm{AIV} = \mathcal{O}(d n^{-(d+1)/d} h^{-3}).$$

Theorem: RQMC beats MC only for $d \leq 3$.

Hope: Reality is not as bad $\rightarrow$ empirical tests in limited region of interest.

The empirical model

For pairs (n, h) in a region of interest we consider the model

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

The key issue is to find a good bin-/bandwidth $h_*(n)$.

Suppose the parameters in red are known, then

$$h_*^{\alpha+\delta} = \frac{C\delta}{B\alpha} n^{-\beta}$$

and $\mathrm{MISE} \approx K n^{-\alpha\beta/(\alpha+\delta)} = K n^{-\nu}$.

The bias is the same as with MC → use same methods.

Estimation of the IV-parameters

Observe that we have a linear model

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

For each pair (n, h) of some sample grid we

1. construct n_r independent replications of the RQMC density estimator,
2. estimate the IV by numerically integrating the empirical variance over $[a, b]$ using n_e evaluation points.

After that, we can obtain C, β, and δ by linear regression.

Numerical illustrations

- As testing regions (n, h) we choose sample sizes $n = 2^{14}, 2^{15}, \ldots, 2^{19}$ and bin-/bandwidths $h_j = 2^{-\ell_0+j/2}$, $0 \leq j \leq 5$.
- For each n and each RQMC method we simulate $n_r = 100$ times independently.
- Integrals over $[a, b]$ are computed with $n_e = 1024$ stratified evaluation points.
- Main figure of interest is $\mathrm{LGM} = -\log_2(\mathrm{MISE})$ for $n = 2^{19}$.

Note: In the entire scheme only choosing ℓ_0 requires human intervention (easy and quick to get).

Point sets used:

- MC: crude Monte Carlo.
- Stratification: stratified unit cube.
- Sobol+LMS: Sobol’ points with left matrix scrambling + digital shift.
- Sobol+NUS: Sobol’ points with nested uniform scrambling.

Normalized sum of standard normals

Let $Z_1, \ldots, Z_d$ i.i.d. standard normals generated by inversion. We estimate

$$X = \frac{Z_1 + Z_2 + \cdots + Z_d}{\sqrt{d}} \sim \mathcal{N}(0, 1)$$

over $[a, b] = [-2, 2]$.

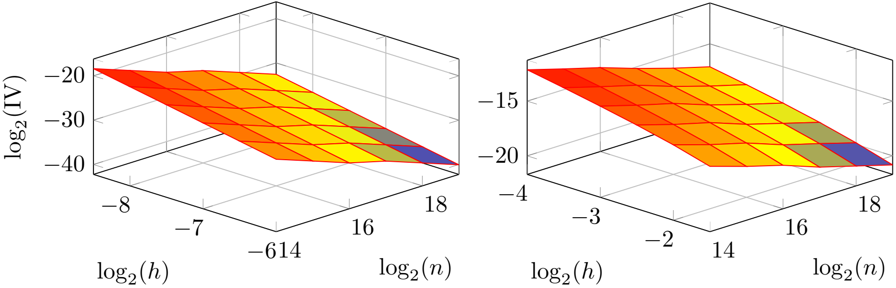
- Easy example to see impact of dimension d.
- We know exact target density $\rightarrow$ can test out of sample with $h_*(n)$ $\rightarrow$ results very reliable.

Estimated parameters with KDE

	MC	Sobol+NUS				
d	1	1	2	3	5	100
R^2	0.999	0.999	1.000	0.995	0.978	0.996

- Linear model for the IV fits very well.

KDE with Sobol+NUS, $d = 1$ (left) and $d = 100$ (right).



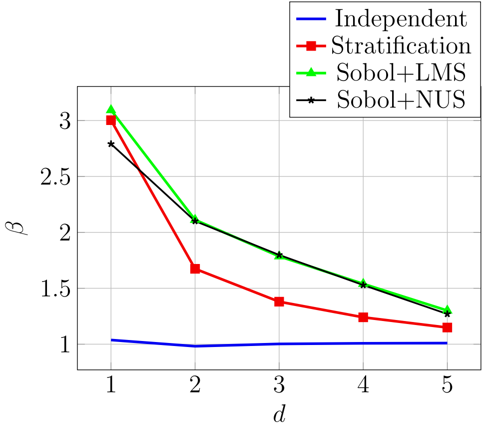
Estimated parameters with KDE

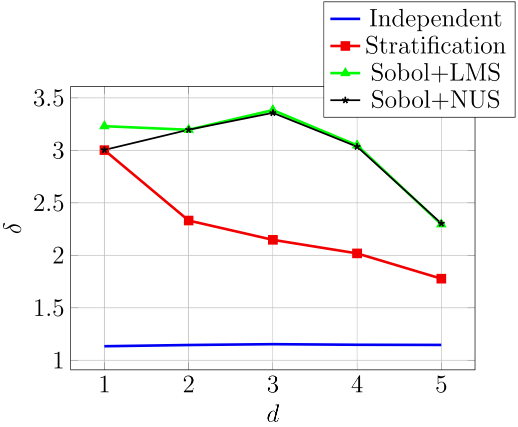
$$\mathrm{MISE} = \mathrm{IV} + \mathrm{ISB} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha}$$

	MC	Sobol+NUS				
d	1	1	2	3	5	100
R^2	0.999	0.999	1.000	0.995	0.978	0.996
β	1.038	2.791	2.101	1.798	1.270	1.010
δ	1.134	3.004	3.196	3.357	2.303	1.463

- β larger with RQMC, but so is δ $\rightarrow$ inhibits MISE-reduction.
- β and δ decrease for large d (approach MC values).

Parameters β (left) and δ (right) for the KDE





Estimated parameters with KDE

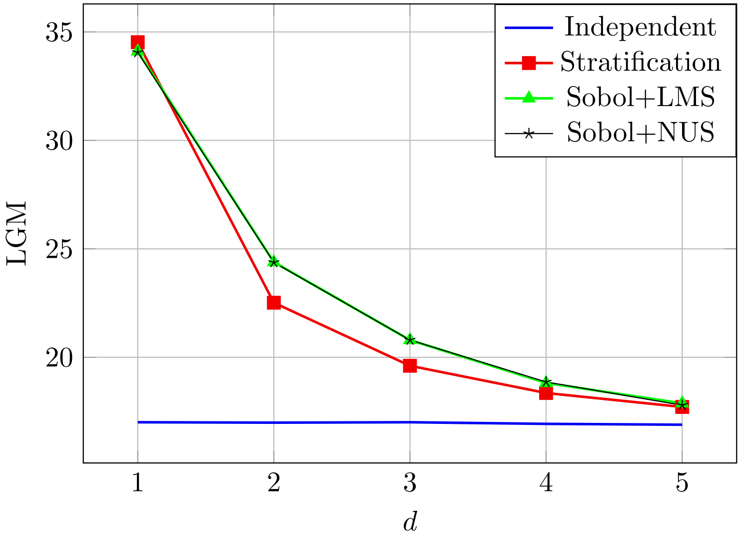
$$\text{MISE} = \text{IV} + \text{ISB} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha} = Kn^{-\nu}$$

	MC	Sobol+NUS				
d	1	1	2	3	5	100
R^2	0.999	0.999	1.000	0.995	0.978	0.996
β	1.038	2.791	2.101	1.798	1.270	1.010
δ	1.134	3.004	3.196	3.357	2.303	1.463
ν	0.781	1.595	1.169	0.975	0.806	0.774
LGM	17.01	34.06	24.38	20.80	17.79	17.05

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

- MISE rates are great in $d = 1$, deteriorate towards MC.
- In $d = 1$ with $n = 2^{19}$ the MISE is $2^{17} \approx 130{,}000$ times smaller with RQMC.
  LGM better with RQMC up to $d = 20$.

LGM for the KDE



Displacement of a cantilever beam

Displacement D of a cantilever beam with horizontal and vertical loads (Bingham):

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

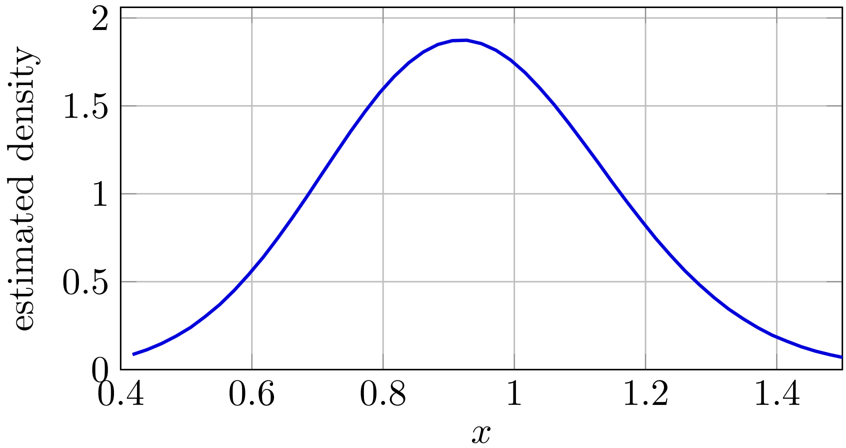
where $L = 100$, $w = 4$, $t = 2$ inches.

$X, Y,$ and E are independent normal r.v.'s $(d = 3)$ with

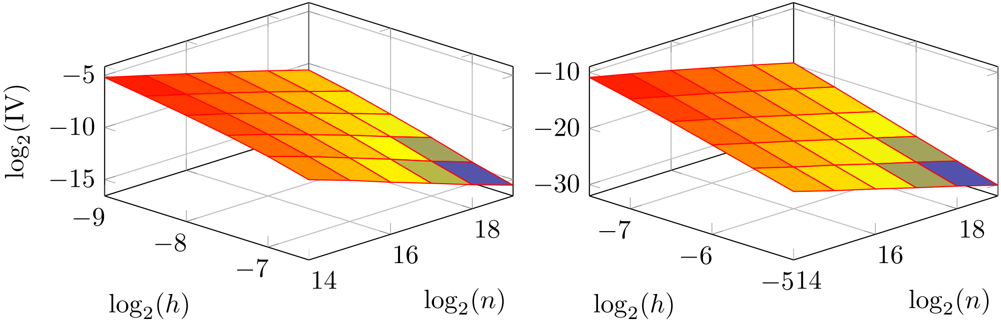
Description	Symbol	Mean	Std. dev.
Young's modulus	E	2.9×10^7	1.5×10^6
Horizontal load	X	500	100
Vertical load	Y	1000	100

We estimate the density of D over $[a, b] = [0.407, 1.515]$ ($\approx$ 99 % of the entire mass).

Estimated density of D with KDE and Sobol+NUS



Linear model for the Histogram (left) and the KDE (right)



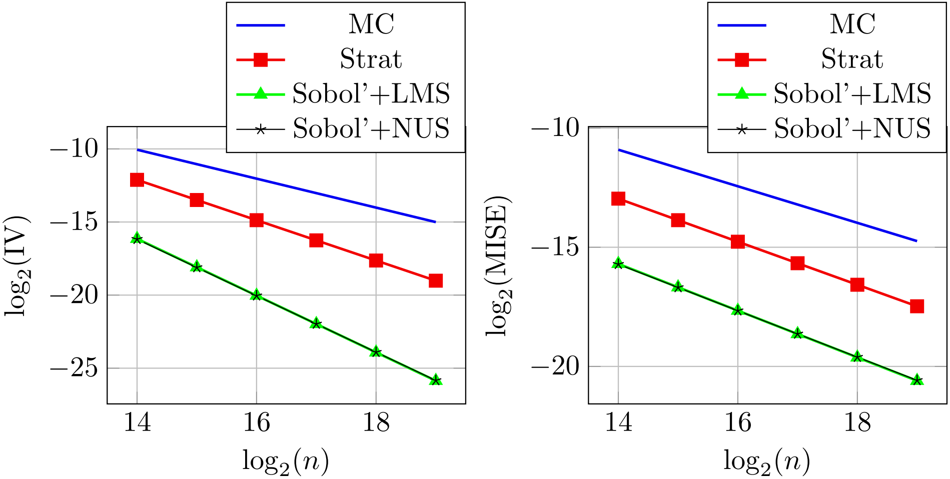
In all cases $R^2 > 0.99$, i.e. the linear model for the IV fits very well.

Estimated parameters for D

$\text{MISE} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha} = Kn^{-\nu}$, $\quad$ $\text{LGM} = -\log_2 \text{MISE}$ for $n = 2^{19}$.

	Histogram ($\alpha = 2$)			KDE ($\alpha = 4$)		
	MC	LMS	NUS	MC	LMS	NUS
β	0.992	1.234	1.232	0.991	1.943	1.932
δ	1.010	1.309	1.733	1.744	3.922	3.933
ν	0.659	0.661	0.658	0.767	0.981	0.974
LGM	11.70	13.03	13.03	14.74	20.60	20.58

- RQMC increases β significantly.
- δ increases even more $\longrightarrow$ inhibits MISE-reduction.
- Still, RQMC outperforms MC (by factor 4 with histogram, by factor 64 with KDE).



left: IV vs. n with fixed $h = 2^{-6} = 1/64$ for KDE.

right: estimated MISE with optimal $h = h_*(n)$ for KDE.

A weighted sum of lognormals

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

where $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

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

We use principal component decomposition in this example.

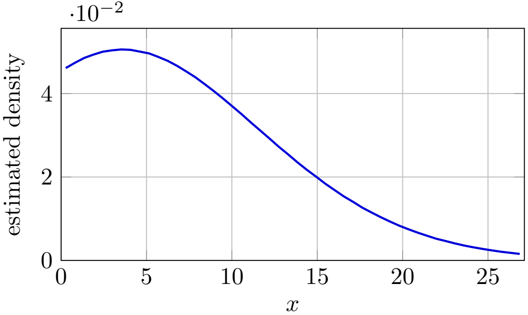
One application is with $w_j = s_0(d - j + 1)/d$, then $e^{-\rho} \max(X - K, 0)$ is the payoff of a financial option based on an average price at d observation times, under a GBM process.

Note: For the KDE we cannot discard realizations of X below K, as they contribute to the KDE above K.

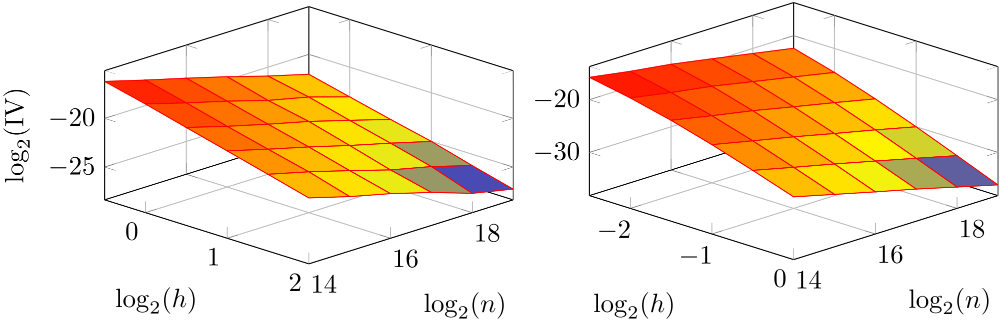
Density of positive payoffs

Take $d = 12$, $s_0 = 100$, $K = 101$. $\mathbf{\Sigma}$ is defined indirectly via $Y_j = Y_{j-1}(\mu - \sigma^2)j/d + \sigma B(j/s)$ with $Y_0 = 0$, $\sigma = 0.12136$, $\mu = 0.1$, and B a standard Brownian motion.

For simplicity, we ignore the discount factor $e^{-\rho}$ and estimate the density of $X - K$ over $[a, b] = [K, K + 27.13]$ (cuts off 29.5% on the left and 0.5% on the right).



Linear model for the histogram (left) and the KDE (right)



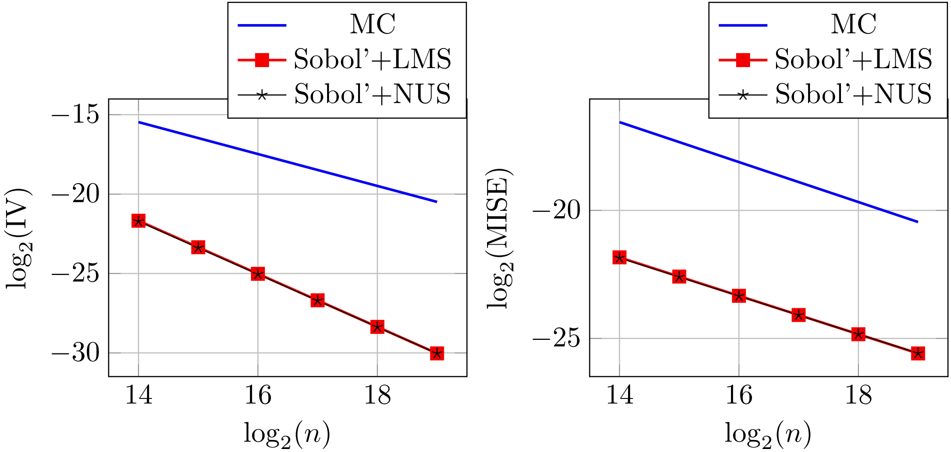
Again, the linear model for the IV fits almost perfectly ($R^2 > 0.99$ in all cases).

Estimated parameters for $X - K$

$\text{MISE} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha} = Kn^{-\nu}$, $\text{LGM} = -\log_2 \text{MISE}$ for $n = 2^{19}$.

	Histogram ($\alpha = 2$)			KDE ($\alpha = 4$)		
	MC	LMS	NUS	MC	LMS	NUS
β	1.015	1.140	1.146	1.005	1.671	1.663
δ	1.168	2.105	2.133	1.151	4.907	4.930
ν	0.641	0.556	0.554	0.780	0.750	0.745
LGM	17.53	18.63	18.61	20.45	25.59	25.58

- Again, RQMC increases β significantly, but δ increases even more.
- MISE is still a lot smaller with RQMC (by factor 2 for histogram and by factor 32 for KDE).



left: IV vs. n with fixed $h = 2^{-6} = 1/64$ for KDE.

right: estimated MISE with optimal $h = h_*(n)$ for KDE.

Conclusions

We studied different density estimators and investigated, if RQMC outperforms MC.

Histogram and KDE: $\mathrm{MISE} = \mathrm{IV} + \mathrm{ISB} \approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha}$

- We benefit from variance reduction, but not as easily as we hoped.
- We saw that RQMC can reduce IV and MISE.
- Observed IV and MISE are much better than in our theoretic results $\rightarrow$ opportunities for theorists.

References

Histogram and KDE

- 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. *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.

- A. B. Owen. *Scrambled Net Variance for Integrals of Smooth Functions.* Annals of Statistics, 25 (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.