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 $X = g(U)$ with

- $g : (0, 1)^d \rightarrow \mathbb{R}$,
- $U = (U_1, \ldots, U_d) \sim \mathcal{U}(0, 1)^d$,
- $g(u)$ 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

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

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

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

The MISE can be decomposed into

\[
MISE = IV + ISB,
\]

where \( IV \) is the integrated variance

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

and \( ISB \) is the integrated square bias

\[
ISB = \int_a^b \left(\mathbb{E}[\hat{f}_n(x)] - f(x)\right)^2 \, dx.
\]
Asymptotics for Monte Carlo

Idea: Use MC sample $U_1, \ldots, U_n \sim \mathcal{U}(0, 1)^d$ to compute $\hat{f}_n$.

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

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

$\rightarrow$ variance-bias tradeoff.
\[ \text{AMISE} = \text{AIV} + \text{AISB} \approx \frac{C}{nh} + Bh^\alpha, \]

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,

\[ \text{AMISE} \approx Kn^{-\alpha/(1+\alpha)}, \]

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

**Idea:** Replace $U_1, \ldots, U_n \sim \mathcal{U}(0,1)^d$ 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

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

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

**Reality:** Not exactly...
In theory (and reality) the power of \( h \) decreases too:

\[
AIV = \int_a^b \text{Var}(\hat{f}_n(x)) \, dx \approx C n^{-\beta} h^{-\delta}, \quad \beta > 1, \delta > 1.
\]

**Strategy:** Rewrite the KDE as

\[
\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{x - g(U_i)}{h} \right) = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}(U_i, x).
\]

**Idea:** Use Koksma–Hlawka type inequalities to bound

\[
\text{Var}(\hat{f}_n(x)) \leq \mathbb{E}[D_n^2(U_1, \ldots, U_d)] V^2(\tilde{g}(\cdot, x)).
\]

**Problem:** faster convergence of discrepancy \( \rightarrow \) larger variation \( \rightarrow \) larger \( \delta \).
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

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

If the star discrepancy $D_n^*(U_1, \ldots, U_n) = \mathcal{O}(n^{-1+\varepsilon})$ then

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

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

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

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

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

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

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

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

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

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

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

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

Suppose the parameters in red are known, then

\[
h_\alpha^{\alpha+\delta} = \frac{C \delta}{B \alpha} n^{-\beta}
\]

and \(\text{MISE} \approx Kn^{-\alpha\beta/(\alpha+\delta)} = Kn^{-\nu}\).

The bias is the same as with MC \(\rightarrow\) use same methods.
Estimation of the IV-parameters

Observe that we have a linear model

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

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 \(C, \beta, \) and \(\delta\) 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 \(\text{LGM} = -\log_2(\text{MISE})\) for \(n = 2^{19}\).

Note: In the entire scheme only choosing \(\ell_0\) 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

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>Sobol+NUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>1</td>
<td>1  2  3  5 100</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.999</td>
<td>0.999 1.000 0.995 0.978 0.996</td>
</tr>
</tbody>
</table>

Linear model for the IV fits very well.

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

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

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>Sobol+NUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>1</td>
<td>1 2 3 5 100</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.999</td>
<td>0.999 1.000 0.995 0.978 0.996</td>
</tr>
<tr>
<td>(\beta)</td>
<td>1.038</td>
<td>2.791 2.101 1.798 1.270 1.010</td>
</tr>
<tr>
<td>(\delta)</td>
<td>1.134</td>
<td>3.004 3.196 3.357 2.303 1.463</td>
</tr>
</tbody>
</table>

- \(\beta\) larger with RQMC, but so is \(\delta\) \(\rightarrow\) inhibits MISE-reduction.
- \(\beta\) and \(\delta\) decrease for large \(d\) (approach MC values).
Parameters $\beta$ (left) and $\delta$ (right) for the KDE

\begin{align*}
\beta &\quad \begin{array}{c}
1 \\
1.5 \\
2 \\
2.5 \\
3 \\
\end{array} \\
\delta &\quad \begin{array}{c}
1 \\
1.5 \\
2 \\
2.5 \\
3 \\
\end{array}
\end{align*}

\begin{align*}
d &\quad \begin{array}{c}
1 \\
2 \\
3 \\
4 \\
5 \\
\end{array} \\
\end{align*}
Estimated parameters with KDE

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

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>Sobol+NUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>(\beta)</td>
<td>1.038</td>
<td>2.791</td>
</tr>
<tr>
<td>(\delta)</td>
<td>1.134</td>
<td>3.004</td>
</tr>
<tr>
<td>(\nu)</td>
<td>0.781</td>
<td>1.595</td>
</tr>
<tr>
<td>LGM</td>
<td>17.01</td>
<td>34.06</td>
</tr>
</tbody>
</table>

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

- 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

![Graph showing LGM for different stratification methods: Independent, Stratification, Sobol+LMS, Sobol+NUS. The graph plots LGM against d, with distinct markers for each method at d values of 1, 2, 3, 4, and 5.](image-url)
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

<table>
<thead>
<tr>
<th>Description</th>
<th>Symbol</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young’s modulus</td>
<td>$E$</td>
<td>$2.9 \times 10^7$</td>
<td>$1.5 \times 10^6$</td>
</tr>
<tr>
<td>Horizontal load</td>
<td>$X$</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>Vertical load</td>
<td>$Y$</td>
<td>1000</td>
<td>100</td>
</tr>
</tbody>
</table>

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

\[ \text{MISE} \approx C n^{-\beta} h^{-\delta} + B h^\alpha = Kn^{-\nu}, \quad \text{LGM} = -\log_2 \text{MISE} \quad \text{for } n = 2^{19}. \]

<table>
<thead>
<tr>
<th></th>
<th>Histogram ($\alpha = 2$)</th>
<th></th>
<th>KDE ($\alpha = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MC</td>
<td>LMS</td>
<td>NUS</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.992</td>
<td>1.234</td>
<td>1.232</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.010</td>
<td>1.309</td>
<td>1.733</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.659</td>
<td>0.661</td>
<td>0.658</td>
</tr>
<tr>
<td>LGM</td>
<td>11.70</td>
<td>13.03</td>
<td>13.03</td>
</tr>
</tbody>
</table>

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

24/33
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 \( Y = (Y_1, \ldots, Y_d)^\top \sim \mathcal{N}(\mu, \Sigma). \)

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

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 $d = 12$, $s_0 = 100$, $K = 101$. $\Sigma$ 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).
Again, the linear model for the IV fits almost perfectly ($R^2 > 0.99$ in all cases).
Estimated parameters for $X - K$

\[
\text{MISE} \approx C n^{-\beta} h^{-\delta} + B h^\alpha = K n^{-\nu}, \quad \text{LGM} = -\log_2 \text{MISE} \quad \text{for } n = 2^{19}.
\]

<table>
<thead>
<tr>
<th></th>
<th>Histogram ($\alpha = 2$)</th>
<th>KDE ($\alpha = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MC</td>
<td>LMS</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.015</td>
<td>1.140</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.168</td>
<td>2.105</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.641</td>
<td>0.556</td>
</tr>
<tr>
<td>LGM</td>
<td>17.53</td>
<td>18.63</td>
</tr>
</tbody>
</table>

- Again, RQMC increases $\beta$ significantly, but $\delta$ 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:** $\text{MISE} = \text{IV} + \text{ISB} \approx C n^{-\beta} h^{-\delta} + B h^{\alpha}$

- 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 → opportunities for theorists.
Histogram and KDE


