Density estimation by Randomized Quasi-Monte Carlo*

Amal Ben Abdellah †, Pierre L'Ecuyer†, Art B. Owen ‡, and Florian Puchhammer†

Abstract. We consider the problem of estimating the density of a random variable X that can be sampled exactly by Monte Carlo (MC) simulation. We investigate the effectiveness of replacing MC by randomized quasi Monte Carlo (RQMC) to reduce the integrated variance (IV) and the mean integrated square error (MISE) for histograms and kernel density estimators (KDEs). We show both theoretically and empirically that RQMC estimators can achieve large IV and MISE reductions and even faster convergence rates than MC in some situations, while leaving the bias unchanged. Typically, RQMC provides a larger IV (and MISE) reduction with KDEs than with histograms. We also find that if RQMC is much more effective than MC to estimate the mean of X for a given application, it does not imply that it is much better than MC to estimate the density of X for the same application. Density estimation involves a well known bias-variance tradeoff in the choice of a bandwidth parameter h. RQMC improves the convergence at any h, although the gains diminish when h is reduced to control bias.

Key words. Density estimation, quasi-Monte Carlo, variance reduction, histogram, kernel density, simulation

AMS subject classifications. 62G07, 62G20, 65C05,

1. Introduction.

We want to estimate by simulation the density of a random variable $X = g(\mathbf{U})$ where $\mathbf{U} = (U_1, \ldots, U_s) \sim U(0,1)^s$ (uniform over the unit hypercube) and $g : (0,1)^s \to \mathbb{R}$. We assume that $g(\mathbf{u})$ can be computed easily for any $\mathbf{u} \in (0,1)^s$, that X has density f (with respect to the Lebesgue measure) over $\mathbb{R}$ and our goal is to estimate f over some bounded interval $[a,b]$. This interval may be a small region of greatest scientific interest or it may be the central 95% or 99% of the support. When the support of X is bounded then $[a,b]$ might include all of it. Estimating the density accurately in areas where it is very small (like far in a tail) requires specialized methods and is beyond the scope of this paper.

Our motivation comes from uncertainty quantification (UQ). In UQ, g may be a black box function of random inputs and $X = g(\mathbf{U})$ a quantity of interest. This X may be the breaking strength of an industrial part, a radiation exposure level, or the voltage in a circuit. Sampling methods can be used to find the mean of X, and the variance of X gives some idea of how close to the mean X will be. However, in applications like these, deviations above and below the mean have different consequences and estimating the density $f(x)$ brings additional information. The simulation literature has long looked at quantities such as the maximum duration of a contruction project or the waiting time in a call center or the future value of a financial holding and in these settings too, there is value in going beyond just the first two

*Submitted to the editors July 6, 2018.

Funding: This work has been supported by a Canada Research Chair, an Inria International Chair, an IVADO Grant, and NSERC Discovery Grant number RGPIN-110050 to P. L'Ecuyer. A. B. Owen was supported by the US National Science Foundation under Grants DMS-1521145 and DMS-1407397.

†DIRO, University of Montreal, 2920 Chemin de La Tour, Pavillon Aisenstadt, Montreal, QC, H3T 1N8, Canada (amalbenabdellah@gmail.com, lecuyer@iro.umontreal.ca, florian.puchhammer@umontreal.ca).

‡Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, CA, 94305-4065, USA (owen@stanford.edu).

moments of X.

We will look at histograms and kernel density estimates of f. Our main contribution is to show how randomized quasi-Monte Carlo methods can be used. Density estimation methods were developed for the context where an independent sample $X_1, \ldots, X_n$ from the unknown density f is given. Here we assume that we can generate a sample $X_1, \ldots, X_n$ of arbitrary size by choosing where to sample. With *crude Monte Carlo* (MC), we would estimate the density from n *independent* realizations of X, say $X_1, \ldots, X_n$, by simulation. Then the analysis is the same as if the data was collected from the real world. But with RQMC methods, the error behaves differently and the analysis is not the same.

We denote by $\hat{f}_n$ a density estimator based on a sample of size n, and we measure the quality of the estimator over $[a, b]$ by the *mean integrated square error* (MISE), defined as

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

which we want to minimize. The MISE can be decomposed as the sum of the *integrated variance* (IV) and the *integrated square bias* (ISB):

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

The simplest and most popular density estimator is certainly a *histogram*. It is very easy to construct and to understand. We partition $[a, b)$ into m equal parts of size $h = (b - a)/m$, and the density estimator is

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

where n_j is the number of observations that fall in this interval and n is the total number of observations. Here we write the intervals closed on one side and open on the other side so they form exactly a partition of $[a, b)$, and the boundary b is left out, but since X is assumed to have a density, the probability that it falls on a boundary is zero, so we can effectively ignore these distinctions.

A *kernel density estimator* (KDE) is defined by selecting a *kernel density* k, and a constant $h > 0$ called the *bandwidth*, which serves as a horizontal stretching factor for the kernel. Given a sample $X_1, \ldots, X_n$, the KDE is defined by

$$\hat{f}_n(x) = \hat{f}_{\mathrm{k},n}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) \tag{1.1}$$

for $x \in \mathbb{R}$. The kernels considered in this paper are probability densities that are symmetric about 0 and have positive bounded even moments. They are also nondecreasing on $(-\infty, 0]$ and hence nonincreasing on $[0, \infty)$, with a finite mode, $k(0) < \infty$. In our numerical experiments, we use the Gaussian kernel, which is the standard normal density function $k(x) = \exp(-x^2/2)/\sqrt{2\pi}$. This kernel is infinitely differentiable and has bounded derivatives of all orders. Although KDEs have better asymptotic theoretical properties than histograms

(see Section 2), the latter are more widely used, mainly because of their simplicity and ease of interpretation, and because they are much faster to compute and plot when n is large.

The aim of this paper is to examine how the asymptotic convergence rates of the IV and MISE can be improved by using RQMC instead of MC (in theory), and to assess empirically the improvements in the IV and MISE that are achieved for reasonable finite sample sizes n, on some examples.

We adopt the usual $\Theta(\cdot)$ notation for the *exact order*: $h(n) = \Theta(\varphi(n))$ means that there is some n_0 and constants $c_2 > c_1 > 0$ such that for all $n \geq n_0$, $c_1 \leq h(n)/\varphi(n) \leq c_2$. This is less restrictive than $h(n) \propto \varphi(n)$. Also, $c(n,h) = \mathcal{O}(\varphi(n,h))$ means that there is a constant $K > 0$ such that for all integers $n \geq 1$ and all $h \in (0,1]$, $c(n,h) \leq K\varphi(n,h)$. In limiting arguments we let $n \to \infty$ to model increasing sample size,

With MC, under mild assumptions, the IV is $\Theta(n^{-1}h^{-1})$ for both the histogram and the KDE when $nh \to \infty$, while the rate of the ISB is $\Theta(h^2)$ for the histogram and $\Theta(h^4)$ for the KDE (see Section 2). Replacing MC by RQMC with the same n and h may reduce the IV and perhaps its convergence rate, but it will not change the bias, because for all density estimators that we consider, for any $x \in [a,b]$, $\hat{f}_n(x)$ is an average which has the same expectation with RQMC as with MC. We shall therefore focus on how RQMC can reduce the IV and improve its convergence rate. The IV is easy to estimate in practice, either with MC or RQMC, but the ISB is harder to estimate. Techniques have been developed for estimating the ISB of a KDE with MC [18] and they can be used to estimate the ISB under RQMC as well, because the ISB is the same.

With RQMC, one might hope for an asymptotic IV of $\mathcal{O}(n^{-\beta}h^{-1})$ for some $\beta > 1$ instead of $\Theta(n^{-1}h^{-1})$, because RQMC is known to reduce the variance rate from $\Theta(n^{-1})$ to $\mathcal{O}(n^{-\beta})$ for $\beta > 1$ when estimating the integral of a smooth function. We could then reduce both the IV and ISB by using a somewhat smaller h than for MC. Unfortunately, decreasing h increases the variation of the integrand that corresponds to the density estimator at a given x. As a result, the exponent of h in the denominator of the IV bound typically becomes larger than 1. Thus, our asymptotic bound for the IV will have the form $\mathcal{O}(n^{-\beta}h^{-\delta})$ for some positive real numbers β and δ that depend on the setting. This bound converges in an asymptotic regime in which $n^{\beta}h^{\delta} \to \infty$ and $h \to 0$. Using the standard Koksma-Hlawka inequality, we prove under certain conditions that $\mathrm{IV} = \mathcal{O}(n^{-2+\epsilon}h^{-2s})$ for any $\epsilon > 0$. This bound degrades quickly with the dimension s, but we emphasize that this is only an asymptotic upper bound. The true IV may behave much better than the bound and does not necessarily increase at the same rate as a function of s and $1/h$.

In addition to developing asymptotic bounds, we study empirically the local behavior of the IV as a function of n and h in a limited bounded region of interest that contains the pairs (n,h) that we are likely to use. To approximate the IV in this selected region, we will use a model of the form

$$\mathrm{IV} \approx Cn^{-\beta}h^{-\delta} \tag{1.2}$$

for some parameters $C > 0$, $\beta > 0$, and $\delta > 0$ that depend on the problem and the RQMC method, and that we can estimate by regression. This model is expected to give a reasonably accurate approximation only for (n,h) in the selected region of interest and not necessarily everywhere, and the coefficients that we estimate may differ from the asymptotic ones.

For the examples we tried, the model turned out to give a reasonable approximation. We need such a model to estimate at what rate we should decrease h as a function of n. We estimate C, β, and δ by linear regression in log-log scale, based on experiments on a grid of values of n and h. With this regression model, we sometimes find $\delta \ll 2s$ and $\beta > 1$ for examples in large dimension s. But in other cases, even when RQMC reduces the variance considerably when estimating the mean $\mathbb{E}[X]$, it does not necessarily provide as much improvement on the MISE when estimating the density. When the dimension s gets very large, we typically obtain $\beta \approx \delta \approx 1$; that is, RQMC behaves pretty much like MC.

The remainder is organized as follows. In Section 2, we recall the definitions and basic properties of histograms and KDE estimators. In Section 3, we recall some variance bounds for RQMC integration. In Section 4, we examine how these bounds can be used to study the convergence rate of the IV and MISE for a KDE under RQMC. In Section 5 we consider stratified sampling. In Section 6, we explain how we estimate the model parameters over a limited region by linear regression. In Section 7, we provide numerical illustrations. We give our conclusions in Section 8.

We end this section with a remark about unit cubes. The uniform distributions on $(0,1)^s$, $[0,1]^s$ and $[0,1)^s$ give the same density f because they are exactly the same distribution. A countable sample size gives us probabilty zero of distinguishing between any two of them. We define our function g over $[0,1]^s$ because that is the natural domain for bounded variation and continuity on the closed cube implies uniform continuity and boundedness. When we sample by quasi-Monte Carlo we take points in $[0,1)^s$ because that is a standard assumption there, due in part to the ease with which that cube can be partitioned into subcubes.

2. Histograms and kernel density estimators with MC.

We recall asymptotic properties of the histogram and KDE when $nh \to \infty$ and $h \to 0$ simultaneously. The details can be found in [6], [18], and [20], for example. The asymptotic MISE, IV, and ISB in this regime are denoted AMISE, AIV, and AISB, respectively. Writing $\mathrm{AIV} = \tilde{g}(n,h)$ for some function $\tilde{g}$ means that $\lim_{n\to\infty,\,\tilde{g}(n,h)\to 0} \mathrm{IV}(n,h)/\tilde{g}(n,h) = 1$ where $\mathrm{IV}(n,h)$ is the IV for the given (n,h), and similarly for the AMISE and AISB. For measurable functions $\psi : \mathbb{R} \to \mathbb{R}$, we define the roughness functional $R(\psi) = \int_a^b (\psi(x))^2 \mathrm{d}x$ and the "moments" $\mu_r(\psi) = \int_{-\infty}^{\infty} x^r \psi(x) \mathrm{d}x$, for integers $r \geq 0$. Because k is a symmetric probability density function, we have $\mu_0(k) = 1$, $\mu_1(k) = 0$ and we have also assumed that $0 < \mu_2(k) < \infty$. Letting $f^{(r)}$ be the r'th derivative of f, we will assume that $R(f^{(r)}) < \infty$ for $r = 1, 2$ for the histogram estimator, and for $r \leq 4$ for the KDE.

With MC, for a histogram with bin (rectangle) width h, we have $\mathrm{AIV} = n^{-1}h^{-1}$. So if $h = \kappa n^{-\gamma}$, then $\mathrm{AIV} = n^{-1+\gamma}/\kappa$. On the other hand, $\mathrm{AISB} = h^2 R(f')/12$. The AMISE is minimized by taking $h = (nR(f')/6)^{-1/3}$, and therefore

$$\mathrm{AMISE} = \Big(\frac{9R(f')}{16}\Big)^{1/3} n^{-2/3},$$

as shown by Scott in [17].

For the KDE in (1.1), again with MC, we have $\mathrm{AIV} = n^{-1}h^{-1}\mu_0(k^2)$ and $\mathrm{AISB} =$

$(\mu_2(k))^2 R(f'')h^4/4$. The AMISE is minimized by taking

$$h^5 = Q/n, \quad \text{where} \quad Q := \frac{\mu_0(k^2)}{(\mu_2(k))^2 R(f'')},$$

if Q is well-defined and finite. This gives

$$\text{AMISE} = \frac{5}{4} Q^{-1/5} \mu_0(k^2) n^{-4/5}.$$

The above formulas tell us exactly how the asymptotically optimal h depends on k and f. Finding a good h amounts to finding a good approximation of $R(f')$ for the histogram and of $R(f'')$ for the KDE. This appears to be a circular problem, because f is precisely the unknown function that we want to estimate. Perhaps surprisingly, a successful approach for selecting a good h for the KDE is to estimate $R(f'')$ by estimating f'' also via KDE, integrating its square over $[a, b]$, and plugging this estimate into the formula for the optimal h [3, 16, 6, 18]. To do that, one needs to select a good h to estimate f'' by KDE. The asymptotically optimal h depends in turn on $R(f^{(4)})$ which we can estimate where $f^{(4)}$ is the fourth derivative of the unknown density. Then $R(f^{(4)})$ can be estimated by integrating the KDE estimator of $f^{(4)}$ and this goes on ad infinitum. In practice, one can select an integer $r_0 \geq 1$, get a rough estimate of $R(f^{(r_0+2)})$, and start from there. One simple way of doing this is to pretend that f is a normal density with a mean and variance equal to the sample mean $\hat{\mu}$ and variance $\hat{\sigma}^2$ of the data, and then compute $R(f^{(r_0+2)})$ for this normal density. To estimate the rth derivative $f^{(r)}$, one can take the sample derivative of the KDE with a smooth kernel k, yielding

$$\hat{f}_n^{(r)}(x) \approx \frac{1}{nh^{r+1}} \sum_{i=0}^{n-1} k^{(r)}\left(\frac{x - X_i}{h}\right). \tag{2.1}$$

The asymptotically optimal h to use in this KDE estimator is

$$h_*^{(r)} = \left(\frac{(2r+1)\mu_0((k^{(r)})^2)}{\mu_2^2(k)^2 R(f^{(r+2)})n}\right)^{1/(2r+5)}. \tag{2.2}$$

This strategy can be used as well to estimate $R(f')$ for the histogram, using a KDE to estimate f'. We will use this strategy to estimate a good h in our experiments with MC and RQMC, with a Gaussian kernel, with $r_0 = 1$ for the histogram and $r_0 = 2$ for the KDE.

In this paper we always take h to be the same for all $x \in [a, b]$. It is possible to improve upon kernel density estimation by using a locally varying bandwidth $h(x) > 0$. For instance, it is advantageous to have a larger $h = h(x)$ where $f(x)$ is smaller. The interested reader is referred to [19, 18].

3. Error and variance bounds for RQMC integration.

We recall some classical error and variance bounds for RQMC integration. They can be found in [5], [8], [11], and [13], for example. We will use them to obtain bounds on the AIV for the KDE.

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

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

Let $\mathfrak{v}$ denote a subset of coordinates, $\mathfrak{v} \subseteq \mathcal{S} := \{1, \ldots, s\}$, and let $\partial^{|\mathfrak{v}|} g / \partial \mathbf{u}_{\mathfrak{v}}$ denote the partial derivative of g with respect to each of the coordinates in $\mathfrak{v}$. When $\partial^{|\mathfrak{v}|} g / \partial \mathbf{u}_{\mathfrak{v}}$ exists and is continuous for $\mathfrak{v} = \mathcal{S}$, the *Hardy-Krause (HK) variation* of g is

$$V_{\mathrm{HK}}(g) = \sum_{\emptyset \neq \mathfrak{v} \subseteq \mathcal{S}} \int_{[0,1)^s} \left| \frac{\partial^{|\mathfrak{v}|} g}{\partial \mathbf{u}_{\mathfrak{v}}} \right| \mathrm{d}\mathbf{u}. \tag{3.1}$$

The *star-discrepancy* of P_n is

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

where $\mathrm{vol}[\mathbf{0}, \mathbf{u})$ is the volume of the box $[\mathbf{0}, \mathbf{u})$. The *Koksma-Hlawka inequality* states that

$$|E_n| \leq V_{\mathrm{HK}}(g) \cdot D^*(P_n). \tag{3.2}$$

Several known construction methods give P_n with $D^*(P_n) = \mathcal{O}((\log n)^{s-1}/n) = \mathcal{O}(n^{-1+\epsilon})$ for all $\epsilon > 0$. They include lattice rules and digital nets. Therefore, if $V_{\mathrm{HK}}(g) < \infty$, it is possible to achieve $|E_n| = \mathcal{O}(n^{-1+\epsilon})$ for the worst-case error. It is also known how to randomize the points of these constructions so that for the randomized points, $\mathbb{E}[E_n] = 0$ and $\mathrm{Var}[E_n] = \mathbb{E}[E_n^2] = \mathcal{O}(n^{-2+\epsilon})$ (e.g., by a random shift modulo 1 for a lattice rule and by a digital random shift for a digital net).

With the (square) L_2-star discrepancy

$$D_2^2(P_n) = \int_{\mathbf{u} \in [0,1)^s} \left| \mathrm{vol}[\mathbf{0}, \mathbf{u}) - \frac{|P_n \cap [\mathbf{0}, \mathbf{u})|}{n} \right|^2 \mathrm{d}\mathbf{u}$$

and the corresponding (square) variation

$$V_2^2(g) = \sum_{\emptyset \neq \mathfrak{v} \subseteq \mathcal{S}} \int_{[0,1)^s} \left| \frac{\partial^{|\mathfrak{v}|} g}{\partial \mathbf{u}_{\mathfrak{v}}} \right|^2 \mathrm{d}\mathbf{u}, \tag{3.3}$$

we obtain the (slightly different) inequality

$$|E_n| \leq V_2(g) \cdot D_2(P_n). \tag{3.4}$$

One always has $D_2(P_n) \leq D^*(P_n)$, and therefore we also know how to construct points sets P_n 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) [12, 13] and $V_2(g) < \infty$, then $\mathbb{E}[E_n] = 0$ and $\mathrm{Var}[E_n] = \mathbb{E}[E_n^2] = \mathcal{O}(V_2^2(g) n^{-3} (\log n)^{s-1}) = \mathcal{O}(V_2^2(g) n^{-3+\epsilon})$ for all $\epsilon > 0$.

4. RQMC to improve the convergence rate of the AIV for a KDE. Replacing MC by RQMC changes only the variance, not the bias. We now examine how the inequalities of Section 3 can be used to bound the AIV for a KDE with RQMC. For this, we bound the variance of $\hat{f}_n(x)$ at an arbitrary point $x \in [a, b]$ and we integrate the bound. With $X = g(\mathbf{U})$, the KDE at x is

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - g(\mathbf{U}_i)}{h} \right) = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}(\mathbf{U}_i) \tag{4.1}$$

where $\tilde{g}(\mathbf{U}_i) := h^{-1}k((x - g(\mathbf{U}_i))/h)$. Thus, $\hat{f}_n(x)$ can be seen as an RQMC estimator of $\mathbb{E}[\tilde{g}(\mathbf{U})] = \int_{[0,1)^s} \tilde{g}(\mathbf{u})\mathrm{d}\mathbf{u}$. To apply the variance bounds of Section 3 to this estimator, we need to bound the variation of $\tilde{g}$, by bounding each term of the sum in (3.1) or (3.3). This requires at least an s-fold derivative of $\tilde{g}$, and therefore s-fold derivatives of k and g. Note that the Gaussian kernel k is infinitely differentiable and has bounded derivatives of all orders, whereas some other popular kernels (e.g., Epanechnikov, for which $k(x) = (3/4)(1 - x^2)$ for $-1 \le x \le 1$) are not differentiable. It is common that the function g is also smooth.

The partial derivative of $\tilde{g}$ with respect to the coordinates in $\mathfrak{v}$ can be written

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

By expanding all the partial derivatives in this expression via the chain rule, we obtain terms in h^{-j} for $j = 2, \ldots, |\mathfrak{v}| + 1$. Then

$$\|\partial^{|\mathfrak{v}|}\tilde{g}/\partial \mathbf{u}_{\mathfrak{v}}\|_\infty = \mathcal{O}(h^{-|\mathfrak{v}|-1}). \tag{4.2}$$

So even if k and g are very smooth, this gives an AIV bound that grows as $\Theta(h^{-2s-2})$ for fixed n. Such a bound can be too conservative however because the variation depends on $\|\partial^{|\mathfrak{v}|}\tilde{g}/\partial \mathbf{u}_{\mathfrak{v}}\|_1$ not $\|\partial^{|\mathfrak{v}|}\tilde{g}/\partial \mathbf{u}_{\mathfrak{v}}\|_\infty$ and only values $g(\mathbf{u})$ within $\mathcal{O}(h)$ of x make a nonnegligible contribution to the L_1 norm.

We can study this localization precisely when $s = 1$ and show that it removes a factor of h^{-1} from the variation. Suppose that $g(u)$ is a nondecreasing function on $[0, 1]$. Then $\tilde{g}(u) = k((x - g(u))/h)/h$ is nonincreasing over the u with $g(u) \le x$ and nondecreasing over u with $g(u) \ge x$. In that case $V_{\mathrm{HK}}(\tilde{g})$ is the ordinary one dimensional total variation, and then

$$V_{\mathrm{HK}}(\tilde{g}) \le \left|\frac{1}{h}k(0) - \frac{1}{h}k\left(\frac{x - g(0)}{h}\right)\right| + \left|\frac{1}{h}k(0) - \frac{1}{h}k\left(\frac{x - g(1)}{h}\right)\right| \le \frac{2k(0)}{h}, \tag{4.3}$$

which has one less power of h^{-1} than the bound obtained via (4.2). The same bound holds for nonincreasing functions g. More generally, if g is monotone within each of M intervals that partition the domain $[0, 1]$ then $V_{\mathrm{HK}}(\tilde{g}) \le 2Mk(0)/h$. The factor of 2 is necessary because k is potentially increasing and then decreasing within each of those intervals. For k with Lipschitz constant L_k we can easily show that $V_{\mathrm{HK}}(\tilde{g}) \le L_k V_{\mathrm{HK}}(g)/h^2$. The more pessimistic power of h in this result could arise for g that oscillate frequently around a value near x, though we do not expect such functions in our motivating use cases. See [7] for general results on composing univariate functions of bounded variation.

Now suppose that k and g are continuously differentiable and $[0, 1]$ can be partitioned into M subintervals with g monotone over each subinterval. Then with RQMC points that achieve $D^*(P_n) = \mathcal{O}(n^{-1})$, we obtain $\mathrm{IV} = \mathcal{O}(n^{-2}h^{-2})$. With $h = \Theta(n^{-1/3})$, this gives $\mathrm{AMISE} = \mathcal{O}(n^{-4/3})$.

To bound the square variation $V_2^2(\tilde{g})$ with $s = 1$, suppose that g is continuously differentiable, Lipschitz with constant L_g, and is monotone over $[0, 1]$. Making the change of variable $w = (x - g(u))/h$, and noting that $\mathrm{d}w/\mathrm{d}u = -g'(u)/h$ with $|g'(u)| \le L_g$, we obtain

$$V_2^2(\tilde{g}) = \frac{1}{h^2}\int_0^1 \left(k'\left(\frac{x - g(u)}{h}\right)\left(\frac{-g'(u)}{h}\right)\right)^2 \mathrm{d}u \le \frac{L_g}{h^3}\int_{-\infty}^{\infty} (k'(w))^2 \mathrm{d}w = \mathcal{O}(h^{-3}). \tag{4.4}$$

If g is monotone over each of M intervals $S_1, \dots, S_M$ that partition $[0,1]$, then the above argument yields $V_2^2(\tilde{g}) = \mathcal{O}(Mh^{-3})$. Since M is constant in our limit, we obtain AIV $= \mathcal{O}(n^{-3}h^{-3})$. Because ISB $= \mathcal{O}(h^4)$ we can take $h = \Theta(n^{-3/7})$ and get AMISE $= \mathcal{O}(n^{-12/7})$.

For $s > 1$, the contribution to $\partial^{|\mathfrak{v}|}\tilde{g}/\partial \mathbf{u}_{\mathfrak{v}}$ of highest order in h^{-1} is

$$h^{-1-|\mathfrak{v}|}k^{(|\mathfrak{v}|)}\Big(\frac{x-g(\mathbf{U})}{h}\Big)(-1)^{|\mathfrak{v}|}\prod_{j\in\mathfrak{v}} g_j(\mathbf{U}),$$

where g_j is the partial derivative of g with respect to U_j. This power is also highest for the case $\mathfrak{v} = \{1, 2, \dots, s\}$ which we now assume. We think it is reasonable that one power of h^{-1} can commonly be removed here just as for $s = 1$. While it was simple to partition $[0,1]$ into M pieces, the multivariate analogue is more cumbersome. Similarly, composition of bounded variation functions is more complicated for $s > 1$ than for $s = 1$. For instance, Basu and Owen [2] give a Lipschitz continuous function f from $\mathbb{R}^2$ to $\mathbb{R}$ and a function τ from $[0,1]^2$ to $[0,1]^2$ with both components having finite variation in the sense of Hardy and Krause, where nonetheless, $f \circ \tau$ has infinite variation. Because of these two complications, we sketch the argument. We will proceed by sequentially partitioning the domain $[0,1]^s$ into subsets where $\tilde{g}$ is regular. Our conditions require only mild regularity.

Suppose that we can partition $[0,1]^d$ into a set of subdomains within which each g_j is either nonnegative or nonpositive but no g_j takes both positive and negative values. Further suppose that $|g_j| \le B_j < \infty$ over $[0,1]^s$. Let us now partition each of those subdomains into a finite number of regions over which $k^{(s)}$ is either nonnegative or nonpositive. For instance, the first derivative of the Gaussian density is increasing then decreasing then increasing again as its argument ranges from $-\infty$ to ∞, the second derivative has 4 regions, and higher derivatives have a finite number of such intervals defined through zeroes of Hermite polynomials. Now let $w = (x - g(\mathbf{u}))/h$ and suppose that each region can be partitioned once more into subsets S where either $\mathbf{u} \to (w, u_2, \dots, u_s)$ is a smooth and invertible mapping to an image $I(S)$, or the same holds with a different u_j than u_1 being displaced by w. The Jacobian of this mapping is $-g_j(\mathbf{u})/h$ which we further assume is nonzero almost everywhere in S. Now, for the case with $j = 1$,

$$\begin{aligned}\int_S \Big| h^{-1-s}k^{(s)}\left(\frac{x-g(\mathbf{u})}{h}\right)(-1)^s\prod_{\ell=1}^{s} g_\ell(\mathbf{u})\Big|\mathrm{d}\mathbf{u} &= h^{-s}\int_{I(S)}\Big|k^{(s)}(w)\prod_{\ell=2}^{s} g_\ell(\mathbf{u})\Big|\mathrm{d}w\prod_{\ell=2}^{s}\mathrm{d}u_\ell \\ &\le h^{-s}\int_{-\infty}^{\infty}|k^{(s)}(w)|\mathrm{d}w\prod_{\ell=2}^{s} B_\ell. \end{aligned} \tag{4.5}$$

The Hardy-Krause variation in dimension s is $\mathcal{O}(h^{-s-1})$ by (4.2) but we might realistically expect it to actually be $\mathcal{O}(h^{-s})$ by (4.5). Then AIV $= \mathcal{O}(n^{-2+\epsilon}h^{-2s})$ under appropriate conditions on the partial derivatives of g, from which AMISE $= \mathcal{O}(n^{-4/(2+s)})$ by taking $h = \Theta(n^{-1/(2+s)})$. This bound has a better rate than for MC if $s < 3$ and the same rate for $s = 3$. Of course, this is only an upper bound. The actual AIV may converge at a faster rate, and/or be smaller for the relevant values of (n, h), so RQMC may still bring an improvement over MC for large s.

5. Stratified sampling of $[0,1)^s$. Here we show that plain stratified sampling of the unit cube can bring an improvement over Monte Carlo. The randomized QMC methods we favor have some but not necessarily all of the properties that make stratification effective. We make the following assumption about g.

Assumption 1. *The function $g : [0,1]^s \to \mathbb{R}$ is nondecreasing in each coordinate and it is Lipschitz continuous with constant $K_g < \infty$.*

If g is nonincreasing in coordinate j we can redefine g by replacing u_j by $1 - u_j$ to satisfy Assumption 1 while preserving the density function of $g(\mathbf{U})$. We make the following assumption about the sampling of points $\mathbf{U}_i$.

Assumption 2. *The cube $[0,1)^s$ is partitioned into $n = b^s$ congruent cubical cells for some integer $b \geq 2$. We sample one point $\mathbf{U}_i$ uniformly within each cell for $i = 1, \ldots, n$ such that for any $x \in [0,1]$ and any $1 \leq i < j \leq n$, $\mathrm{Cov}[I_i(x), I_j(x)] \leq 0$.*

Stratified sampling independently within each cell satisfies Assumption 2 with a covariance of zero. If we use a digital net with NUS instead, the covariance is no longer zero in general, although we suspect (but do not prove) that it is never negative.

Let F be the cdf of $X = g(\mathbf{U})$ for $\mathbf{U} \sim U[0,1]^s$, and $V_i = F(g(\mathbf{U}_i))$ for each i. The KDE estimator (4.1) can be rewritten as

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - F^{-1}(V_i)}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} \check{g}(V_i) \tag{5.1}$$

which can be interpreted as a KDE estimator based on inversion of F with the one-dimensional uniform points V_i; that is, $g(\mathbf{U}_i)$ and $\tilde{g}(\mathbf{U}_i)$ are replaced by $F^{-1}(V_i)$ and $\check{g}(V_i)$. Then, given an upper bound on the discrepancy of the V_i, we obtain a variance bound for the KDE estimator (5.1) with $s = 1$. We will prove this for the mean square L_2-discrepancy under Assumptions 1 and 2. For any $x \in [0,1]$, define $I_i(x) = \mathbb{I}(V_i \leq x) = \mathbb{I}(g(\mathbf{U}_i) \leq F^{-1}(x))$. We have $\mathbb{E}[I_i(x)] = \mathbb{P}[F(g(\mathbf{U})) \leq x] = x$ and $\mathrm{Var}[I_i(x)] = x(1-x) \leq 1/4$. For any $x \in [0,1]$, let $\mathcal{S}(x)$ be the manifold that separates the region in which $F(g(\mathbf{u})) \leq x$ from that in which $F(g(\mathbf{u})) > x$. Under Assumption 1 we can write $\mathcal{S}(x) = \{\mathbf{u} \in [0,1]^s \mid F(g(\mathbf{u})) = x\}$. We include cases where $\mathcal{S}(x) = \emptyset$. We let $\mathcal{B}(x)$ be the set of subcubes whose interior intersects $\mathcal{S}(x)$.

Lemma 5.1. *Let g satisfy Assumption 1 and let $\mathbf{U}_i$ satisfy Assumption 2. Then the cardinality of $\mathcal{B}(x)$ is at most $sb^{s-1} = sn^{(s-1)/s}$.*

Proof. For each $\mathbf{i} = (i_1, i_2, \ldots, i_s) \in \{0, 1, \ldots, b-1\}^s$, define the cell $C_{\mathbf{i}} = \prod_{j=1}^{d} [i_j/b, (i_j + 1)/b)$. We define the 'diagonal string' of cells through $C_{\mathbf{i}}$ to be all cells $C_{\mathbf{i}'}$ with $\mathbf{i}' = (i_1 + k, i_2 + k, \ldots, i_s + k)$ for $k = -\min_j i_j$ to $k = s - \max_j i_j$. The manifold $\mathcal{S}(x)$ intersects at most one cell in this string because $F(g(\cdot))$ is nondecreasing. If we follow any string 'up and to the right' until it reaches its largest value of k then one of the $i_j + k$ will equal $b - 1$. There are s choices for the critical index j and for each of those there are b^{s-1} boundary cells determined by indices $\ell \neq j$. ∎

The squared L_2-star discrepancy of $V_1, \ldots, V_n$ is

$$D^2_{\mathrm{v},2} = D^2_{\mathrm{v},2}(V_1, \ldots, V_n) = \int_0^1 \left(\frac{1}{n} \sum_{i=1}^n I_i(x) - x \right)^2 \mathrm{d}x.$$

Our next proposition provides an upper bound for it.

Proposition 5.2. *Under Assumptions 1 and 2,*

$$\mathbb{E}[(D^2_{\mathrm{v},2})] \le \frac{s}{4} n^{-(s+1)/s}.$$

Proof. We have

$$\mathbb{E}[D^2_{\mathrm{v},2}] = \int_0^1 \mathrm{Var}\left[\frac{1}{n} \sum_{i=1}^n I_i(x) \right] \mathrm{d}x \le \frac{1}{n^2} \int_0^1 \sum_{i \in \mathcal{B}(x)} \mathrm{Var}[I_i(x)] \mathrm{d}x \le \frac{1}{4n^2} \int_0^1 |\mathcal{B}(x)| \mathrm{d}x$$
$$\le \frac{s}{4} n^{-(s+1)/s}.$$

∎

Corollary 5.3. *Under Assumptions 1 and 2, $V_2^2(\check{g}) = \mathcal{O}(h^{-3})$ and $\mathrm{AIV} = \mathcal{O}(s n^{-(s+1)/s} h^{-3})$.*

Proof. The function $\check{g}$ is continuously differentiable, Lipschitz and monotone on $[0,1]$ and so (4.4) applies to it, yielding $V_2^2(\check{g}) = \mathcal{O}(h^{-3})$. The rest follows from Lemma 5.1 and Proposition 5.2. ∎

This corollary tells us that that when the dimension s increases, the variance with stratified sampling does not grow as badly as $\Theta(h^{-2s})$ as a function of h for fixed n which could have been suggested by the bounds discused at the end of Section 4.

It is not necessary for the strata to be cubical. We could instead partition $[0,1)^s$ into $n = \prod_{j=1}^s m_j$ cells congruent to $\prod_{j=1}^d [0, 1/m_j)$, subject to the condition $\max_j m_j \le \lambda \min_j m_j$ for some $\lambda < \infty$. Then the number of strata intersected by the boundary manifold $\mathcal{S}$ is at most $\sum_j n/m_j \le \sum_j \lambda^{s-1} m_j^{s-1} \le s\lambda^{s-1} n^{1-1/s}$. The aspect ratio λ affects the constant but not the rate for the AIV bound.

6. A regression model for the local behavior of the IV and MISE. The analysis in the previous section was in terms of asymptotic bounds. But what matters most is how things behave in the range of values of n and h that one wants to use. In this section, we propose simple regression models to approximate the true IV, ISB, and MISE in a limited range of values of n and h. In this limited region, we shall use the approximations

$$\mathrm{IV} \approx C n^{-\beta} h^{-\delta} \quad \text{and} \quad \mathrm{ISB} \approx B h^{\alpha}, \tag{6.1}$$

for positive constants C, β, δ, and B, that can be estimated as explained below, and we take $\alpha = 2$ for the histogram and $\alpha = 4$ for the KDE. After that, we can use the model as follows to estimate the optimal h, which minimizes the MISE for the n that we have selected. The model in (6.1) gives

$$\mathrm{MISE} \approx C n^{-\beta} h^{-\delta} + B h^{\alpha}. \tag{6.2}$$

This estimate of MISE is a convex function of h for our setting where $\alpha \geq 1$ and $\delta > 0$. Taking the derivative with respect to h and setting it to zero yields

$$h^{\alpha+\delta} = \frac{C\delta}{B\alpha} n^{-\beta}. \tag{6.3}$$

Thus, if we take $h = \kappa n^{-\gamma}$, the constants κ and γ that minimize the MISE (based on our model) are $\kappa = \kappa_* := (C\delta/B\alpha)^{1/(\alpha+\delta)}$ and $\gamma = \gamma_* := \beta/(\alpha+\delta)$. Plugging them into the MISE expression gives

$$\text{MISE} \approx K n^{-\nu} \tag{6.4}$$

with $K = K_* := C\kappa_*^{-\delta} + B\kappa_*^{\alpha}$ and $\nu = \nu_* := \alpha\beta/(\alpha+\delta)$. If h is taken too small (e.g., by taking $\kappa < \kappa_*$ or $\gamma > \gamma_*$ in the formula for h), the IV will be too large and will dominate the MISE, so we will observe a MISE that decreases just like the IV. The opposite happens if h is too large: the ISB dominates the MISE.

We now look at how we can estimate the model parameters. Taking the log in the IV expression in (6.1), we obtain the linear model,

$$\log(\text{IV}) \approx \log C - \beta \log n - \delta \log h, \tag{6.5}$$

whose three parameters C, β, and δ, can be estimated by linear regression. We select a set of pairs (n, h) in some area of interest, we estimate the IV at each pair by using n_r independent replications of the density estimator at that pair, and we fit the linear model (6.5) to these IV estimates. In this paper, all the logarithms are in base 2, which is natural because we take powers of 2 for n. To approximate the integral that defines the IV, we take a stratified sample of n_e evaluation points over the interval $[a, b]$, compute the empirical variance of the density estimator at each of these points, based on the n_r replications, and take the average multiplied by $(b-a)$. In place of the stratification, one could also use a mid-point rule, with the n_e evaluation points $x_j = a + (b-a)(j-1/2)/n_e$ for $j = 1, \ldots, n_e$. We did not do that to protect ourselves in case the error with RQMC could have a periodic behavior, perhaps partially due to the regularity of the RQMC points.

To select the value of h used for each n in the experimental strategy just described, we can make some preliminary experiments with various h and look at the smoothness of the estimated density to evaluate a range of reasonable values of h. In these pilot runs, we can also estimate the IV very crudely with small n_r and n_e and use this together with an estimate of B (see the next paragraph) to get a rough estimate of the optimal h. Then we make our design based on values of h that are around this estimate, and do the experiments with a larger n_e to estimate the linear model. This general methodology can be used whenever we want to fit this type of regression model. We will use it later for our experiments.

For the ISB approximation in (6.1), we suggest to fix α to the known asymptotic value (2 for the histogram and 4 for the KDE). The constant B is not easy to estimate because we cannot really compute the ISB empirically when we do not know the exact density. However, since we know that the ISB is the same for RQMC as for MC, we can estimate the AISB using any of the techniques developed for MC, for example the plug-in method discussed in Section 2, and use this AISB estimate as an approximation of the ISB in our model. This

means taking $B = R(f')/12$ for the histogram and $B = (\mu_2(k))^2 R(f'')/4$ for the KDE. The only unknown ingredient in this expression is the integral $R(f'')$, which can be estimated as explained in Section 2, eventually using RQMC instead of MC to improve the accuracy. For a Gaussian kernel, $\mu_2(k) = 1$, so we take $B = R(f'')/4$ for the KDE. This estimate of B can then be used in (6.3) to select h. By adding the IV and ISB estimates, we have an estimator of the MISE.

We will test this methodology empirically, firstly on simple examples for which the density of X is known exactly. For these simple examples, we will compute the empirical square bias at the same number of evaluation points n_e as for the IV, by squaring the difference between the exact density and the average of the n_r density estimates at each point, then compute $(b-a)$ times the average of these n_e squared bias values to estimate the ISB. Fitting a linear model to these ISB estimates (in log scale) can provide alternative (direct) estimators of α and B over the region of interest, and we can compare them to the known theoretical asymptotic α and the estimate of B obtained by estimating (or computing) $R(f'')$.

Once we have estimates $\hat{\kappa}_*$ and $\hat{\gamma}_*$ of κ_* and γ_*, we can test the model out-of-sample by making an independent set of simulation experiments with pairs (n, h) that satisfy $h = \hat{h}_*(n) := \hat{\kappa}_* n^{-\hat{\gamma}_*}$ for a series of values of n. At each of these pairs (n, h), we sample n_r independent replicates of the RQMC density estimator and compute the IV estimate, as well as the MISE estimate in the simple examples where the density is known. In the latter case, we fit again the linear regression model for log(MISE) vs $\log n$ to re-estimate the parameters K and ν in (6.4) and assess the goodness-of-fit. In our results, we will denote these new estimates by $\tilde{K}$ and $\tilde{\nu}$, and the corresponding R^2 by R^2_{MISE}. We do similar regression for $\log_2$(IV) and $\log_2$(ISB) vs $\log_2(n)$ and denote the R^2's by R^2_{IV} and R^2_{ISB}.

The model testing steps described in the last two paragraphs are mainly for us to assess how good are our convergence rate estimates with RQMC. If the goal is only to estimate the density f and not the convergence properties, then these testing steps can be left out.

7. Numerical illustrations. In this section, we summarize the results of our experimental study with various examples, with a histogram and a KDE with Gaussian kernel. In all cases, we took $m = (b-a)/h$ bins for the histograms, where h was always selected so that m is an integer. We did this with independent points (MC) and with the following RQMC point sets: (1) stratification of the unit cube (Stratif); (2) a Sobol' point set with left random matrix scrambling and random digital shift (Sobol'+LMS); and (3) a Sobol' point set with NUS (Sobol'+NUS) [13, 14]. These point sets and randomizations are implemented in SSJ [9], which we used for our experiments. The short names in parentheses are used in the plots and tables.

To estimate the model parameters by regression we used a grid of 36 pairs (n, h) with $n = 2^{14}, \ldots, 2^{19}$ and $h = h_0, \ldots, h_5$ where $h_j = h_0 2^{j/2} = 2^{-\ell_0 + j/2}$. For the KDE, $2\ell_0$ is always an integer. For the histogram, the h_j's are approximately of this form but not exactly, because m must be an integer. To select ℓ_0 for each example and sampling method, we first made pilot runs to obtain a crude estimate $\tilde{h}(n)$ of the optimal h for each considered n as explained earlier, with only $n_e = 32$ evaluation points for the integrals. Then we selected ℓ_0 such that $h_0 = 2^{-\ell_0}$ (the smallest h) was near $\tilde{h}(2^{19})$ while h_5 (the largest h) was near $\tilde{h}(2^{14})$, so that $\tilde{h}(n)$ was in the range of the considered values of h for each n. This selection of ℓ_0 is

the only step that requires human intervention. For each n and each point set, we generate a sample of size n, sort the sample, and then compute each selected density estimator for each h, for this sample. That is, we use the same sample for all estimation methods and all h. We take $n_r = 100$ independent replications and $n_e = 1024$ evaluation points. We also tried larger values of n_e such as 4096, 20,357 (a prime number) in preliminary experiments with some of the examples, and the estimated IV and ISB were essentially the same, although computing the KDE and estimating the integrals $R(\cdot)$ involved in the estimation of B took a lot more time. We estimated B as explained in Sections 2 and 6, with $r_0 = 1$ for the histogram and $r_0 = 2$ for the KDE. In the end, we compare the efficiencies of different methods for the same example by comparing their estimated MISE for $n = 2^{19}$ with the h recommended by our model. We will use LGM as an abbreviation of $-\log_2(\text{MISE})$ for $n = 2^{19}$. The efficiency gain of RQMC vs MC can be assessed by comparing their LGMs.

7.1. A normalized sum of standard normals. We will use a technique from [15] to create test functions in which we can vary the dimension s without changing the answer. We let $Z_1, \ldots, Z_s$ be s independent standard normal random variables generated by inversion and put $X = (Z_1 + \cdots + Z_s)/\sqrt{s}$. Then X is standard normal so that $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$ for any s.

We use this example as a first empirical test case of how RQMC can improve the density estimators, and in what way the performance depends on s, h, and n. We will estimate the density over the interval $(-b, b)$ for $b = 2$ and $b = 4$. Since we know the exact density in this case, we can compute unbiased estimators of the IV, ISB, and MISE in all cases. We can also compute $R(f'')$ exactly in this example, which means we can compute B for the AISB and the asymptotically optimal h for the AMISE. However, we will first make experiments as if we did not know this B and have to estimate it, and then compare our estimates with the exact B. Here, g is a monotone increasing function, so Corollary 5.3 applies when we use RQMC point sets that satisfy Assumption 2.

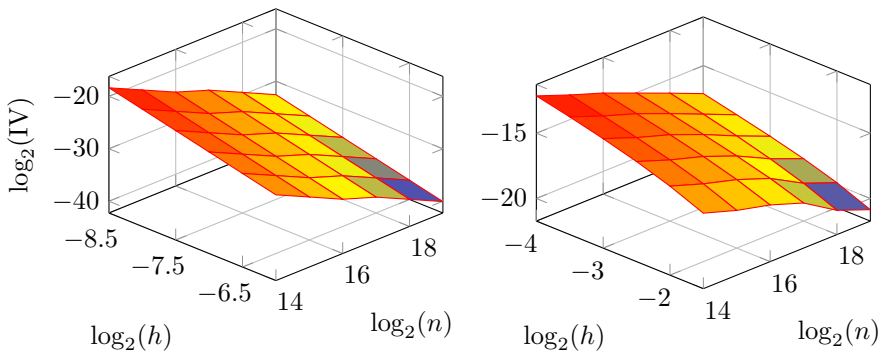

Figure 1. $\log_2(\text{IV})$ *for the KDE with Sobol'+NUS for* $s = 1$ *(left) and* $s = 20$ *(right).*

Tables 1 and 2 summarize the results for $b = 2$, when B is estimated. For MC, our estimates given in the first column are based on experiments made with $s = 1$, but are valid for all s, because the IV and ISB do not depend on s. The estimated values for MC agree with the theory: the exact asymptotic values are $\gamma = 0.4$ and $\nu = 2/3$ for the histogram,

Table 1

Parameter estimates for the histogram estimator, for a sum of normals, over $(-2, 2)$.

	MC	NUS	LMS	NUS	LMS	NUS	LMS	NUS	NUS	NUS	NUS
s		1	2	2	3	3	5	5	10	20	100
ℓ_0	5.0	8.5	6.0	6.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
C	0.813	1.316	1.222	1.299	0.769	0.744	0.948	1.057	0.697	0.772	0.741
β	1.000	2.007	1.502	1.504	1.274	1.274	1.058	1.071	1.004	1.007	0.996
δ	1.041	2.012	2.022	2.012	1.893	1.906	1.181	1.196	1.094	1.077	1.051
R^2	1.000	1.000	1.000	1.000	0.995	0.995	0.998	0.998	1.000	1.000	1.000
B	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
$\hat{\kappa}_*$	3.339	3.315	3.249	3.304	2.949	2.918	3.459	3.571	3.162	3.275	3.237
$\hat{\gamma}_*$	0.329	0.500	0.373	0.375	0.327	0.326	0.333	0.335	0.324	0.327	0.326
ℓ_*	4.507	7.775	5.395	5.398	4.659	4.653	4.529	4.532	4.503	4.509	4.509
$\hat{K}_*$	0.352	0.237	0.227	0.235	0.193	0.189	0.348	0.368	0.306	0.331	0.329
$\hat{\nu}_*$	0.658	1.000	0.747	0.750	0.655	0.652	0.665	0.670	0.649	0.655	0.653
$\tilde{\nu}$	0.667	1.018	0.747	0.752	0.659	0.663	0.678	0.662	0.652	0.661	0.652
R^2_{IV}	0.998	1.000	0.998	1.000	0.962	0.959	0.993	0.997	0.999	0.998	0.999
R^2_{ISB}	0.993	0.997	0.996	0.997	0.994	0.994	0.995	0.993	0.993	0.994	0.994
R^2_{MISE}	0.998	0.999	0.997	0.999	0.978	0.977	0.992	0.995	0.998	0.995	0.998
LGM	13.98	21.10	16.29	16.29	14.64	14.66	14.08	14.07	14.00	13.99	13.99

Table 2

Parameter estimates for the KDE, for a sum of normals, over $(-2, 2)$.

	MC	NUS	LMS	NUS	LMS	NUS	LMS	NUS	NUS	NUS	NUS
s		1	2	2	3	3	5	5	10	20	100
ℓ_0	4.5	8.5	6.0	6.0	5.0	5.0	4.5	4.5	4.0	4.0	4.0
C	0.265	0.032	0.243	0.212	0.144	0.180	0.140	0.096	0.029	0.078	0.079
β	1.038	2.791	2.112	2.101	1.786	1.798	1.301	1.270	1.011	0.996	1.010
δ	1.134	3.004	3.196	3.196	3.383	3.357	2.295	2.303	1.811	1.421	1.463
R^2	0.999	0.999	1.000	1.000	0.995	0.995	0.979	0.978	0.990	0.991	0.996
B	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
$\hat{\kappa}_*$	1.121	0.925	1.238	1.215	1.156	1.191	1.109	1.045	0.820	0.925	0.934
$\hat{\gamma}_*$	0.202	0.398	0.293	0.292	0.242	0.244	0.207	0.201	0.174	0.184	0.185
ℓ_*	3.675	7.682	5.268	5.266	4.386	4.391	3.776	3.765	3.590	3.604	3.612
$\hat{K}_*$	0.299	0.071	0.221	0.205	0.163	0.184	0.173	0.137	0.061	0.117	0.119
$\hat{\nu}_*$	0.808	1.594	1.174	1.168	0.967	0.978	0.826	0.806	0.696	0.735	0.740
$\tilde{\nu}$	0.781	1.595	1.176	1.169	0.976	0.975	0.832	0.806	0.744	0.764	0.774
R^2_{IV}	0.997	0.995	1.000	1.000	0.977	0.978	0.909	0.917	0.983	0.993	0.983
R^2_{ISB}	0.969	0.995	0.991	0.991	0.988	0.989	0.985	0.985	0.972	0.983	0.986
R^2_{MISE}	0.994	0.997	0.997	0.998	0.980	0.981	0.937	0.942	0.981	0.991	0.983
LGM	17.01	34.06	24.39	24.38	20.79	20.80	17.88	17.79	17.28	17.07	17.05

$\gamma = 0.2$ and $\nu = 4/5$ for the KDE, and $\beta = \delta = 1$ for both methods. The other columns give some results for Sobol'+LMS and Sobol'+NUS, for selected values of s. For all $s > 1$ that we have tried, LMS and NUS have a similar behavior. The first rows give the dimension s, the ℓ_0 found by pilot runs and used to fit the IV model, the estimated parameters C, β, and δ of

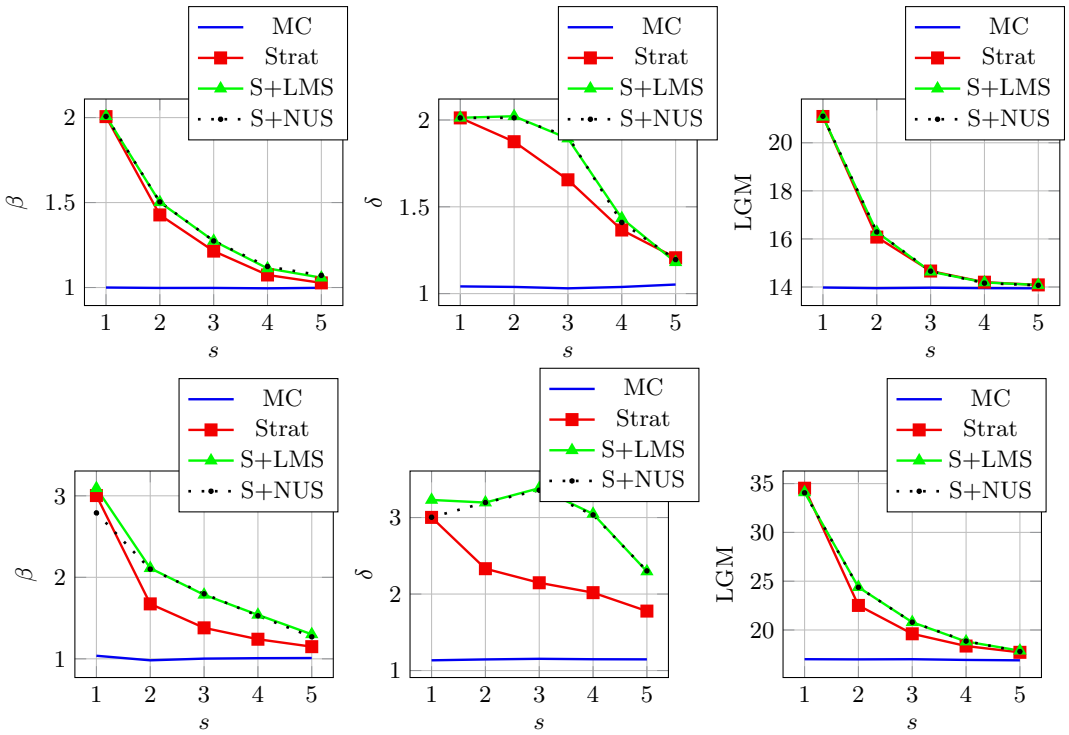

Figure 2. *Estimated β, δ, and LGM ($= -\log_2(\text{MISE})$ for $n = 2^{19}$) for the histogram (above) and the KDE (below) over $(-2, 2)$, with Monte Carlo, Stratification, Sobol'+LMS, and Sobol'+NUS.*

the IV model, the fraction R^2 of variance explained by this model, and the estimated B. All the other quantities are defined at the end of Section 6, except for $\ell_* = -\log_2 \hat{h}_*(2^{19})$. Recall that the rates $\tilde{\nu}$ and the LGM were obtained from a second-stage experiment, by using the estimated $\hat{h}_*(n)$ from the model in the first stage. All the R^2 coefficients are pretty close to 1, which means that the log-log linear model is reasonably good in the area considered. The estimate of B given in the tables turns out to be the same for all s and all RQMC methods, up to three decimal digits: it is $B \approx 0.01081$ for the histogram and $B \approx 0.0418$ for the KDE. Thus, the estimator of B has very little variance. The MISE reduction of RQMC vs MC can be assessed by comparing their LGMs given in the last row, for a given s. For example, with the KDE for $s = 1$, the MISE for $n = 2^{19}$ is approximately 2^{-34} for Sobol'+NUS compared to 2^{-17} for MC, i.e., about $2^{17} \approx 125{,}000$ times smaller. For $s = 2$, for both LMS and NUS, the MISE is about $2^{-24.3}$, which is about 150 times smaller than for MC.

Figure 1 gives a visual assessment of the fit of the linear model for $\log_2(\text{IV})$ in the selected region, with the KDE, for two values of s. We made similar plots for all s and for the histogram as well. The linear approximation turns out to be reasonable, better for the histogram than the KDE, and also better for stratification (not shown in the table) than for RQMC methods.

Figure 2 shows the estimated β, δ, and LGM for the two density estimators, for $s = 1, \ldots, 5$. Stratification, shown in these plots and not in the tables, is exactly equivalent to Sobol'+NUS for $s = 1$, and somewhat less effective for $s > 1$.

One important observation from the plots and the last row of the tables (the LGM) is that for all s, the RQMC methods never have a larger MISE than MC. Their MISE is much smaller for very small s, and becomes almost the same as for MC when s gets large. The MISE rate $\tilde{\nu}$ behaves similarly. RQMC brings more MISE reduction for the KDE than for the histogram, as we would expect based on the smoothness of the KDE integrand and lack thereof for the histogram. Another important observation is that the coefficients β and δ in the IV model (which are both 1 with MC) are *both* larger than 1 with RQMC. For small values of s, with RQMC, β is significantly larger than $\tilde{\nu}$, which means that the IV converges must faster as a function of n when h is fixed than when h is selected to optimize the MISE. This is explained by the large values of δ, sometimes even larger than 3, which indicate that reducing h to reduce the ISB increases the IV pretty fast, and this limits the MISE reduction that we can achieve. Recall that our asymptotic upper bound for the one-dimensional case with NUS in Section 4 gave $\beta = \delta = 3$ and $\nu_* = 12/7 \approx 1.714$. The values in the table are not far from these numbers.

Here f is the standard normal density with $R(f') = (-be^{-b^2} + \int_0^b e^{-x^2}\mathrm{d}x)/2\pi$ and $R(f'') = (-b\left(2b^2-1\right)e^{-b^2} + 3\int_0^b e^{-x^2}\mathrm{d}x)/4\pi$. In particular, for $b = 2$ we have $R(f') \approx 0.13456$ and $B = R(f')/12 \approx 0.01121$ for the histogram, whereas $R(f'') \approx 0.19018$ and $B = R(f'')/4 \approx 0.04754$ for the KDE. The values of B obtained by estimating the derivative of the density (given in the tables) are 0.0108 and 0.0418, respectively. The difference is not due to noise, because the estimated values are the same for all methods, They are due to the bias in the estimation of $R(f')$ and $R(f'')$ via KDE with finite n. We verified empirically that when we estimate these quantities with a larger n, the bias decreases slowly and goes to 0 when $n \to \infty$.

We repeated the density estimation experiment by using the exact values of B instead of the estimated ones to choose h, and the results were very close for all s. In particular, the MISE rates $\tilde{\nu}$ and the MISE values for $n = 2^{19}$ (the LGM) were almost the same. For example, for MC with the exact B, the LGM was 13.91 for the histogram and 16.82 for the KDE (compared with 13.97 and 17.01 with the estimated B). This is not surprising, because the values of B and h do not change much. With the exact B, the regression models for the ISB and MISE fit better (the coefficients R^2_{ISB} and R^2_{MISE} are closer to 1) but they were already near 1 with the estimated B.

So far, the density estimator was evaluated only over the interval $[-2, 2]$. If we go further in the tails, we get into areas in which there will be very few data points, so the density estimator will inevitably be poor, even if the data points are very regularly spaced with spacings proportional to the inverse density (which is arguably the best we can hope for). RQMC is not designed to solve this low-density (or rare-event) problem. To illustrate this, we redid the experiment above with the interval $[-4, 4]$ instead. Only 0.003% of the normal density is outside this interval. For this experiment, we fixed B to its asymptotically optimal value, $B \approx 0.1175$ for the histogram and $B \approx 0.05289$ for the KDE. The interesting parts of the results are in Tables 3 and 4

A key observation is that the gain from RQMC degrades when we get further into the tails. The parameters β and δ estimated by regression also change. For example, with Sobol'+NUS and $s = 1$, the LGM goes from 34.06 to 29.95, i.e., the MISE increases by a factor of 16.

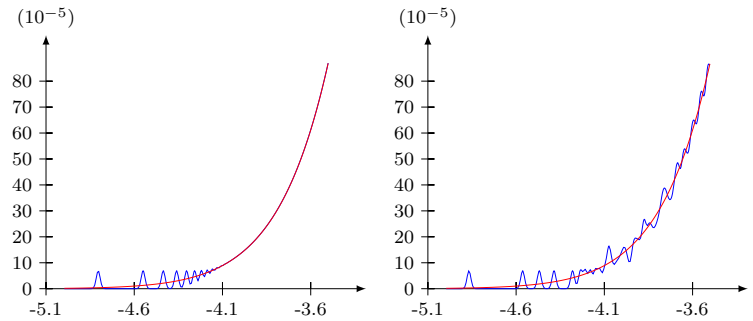

Figure 3. *KDE (blue) vs true density (red) with RQMC point sets with $n = 2^{19}$: Sobol' + digital shift (left) and Sobol' + 31-bit-LMS + shift (right)*

For the interval $[-2, 2]$ we had $\beta = 2.791$, $\delta = 3.004$, and $\hat{h}_*(2^{19}) \approx 2^{-7.7}$, and we now have $\beta = 2.429$, $\delta = 2.463$, and $\hat{h}_*(2^{19}) \approx 2^{-6.8}$. These differences disappear when s increases.

To estimate the density in an area where it is very small, one should use a technique like importance sampling to increase the number of samples in that area. After that, RQMC can bring additional gain.

Table 3

Parameter estimates of the regression model with a histogram estimator over $(-4, 4)$, $B \approx 0.1175$.

	MC	NUS	LMS	NUS	LMS	NUS	LMS	NUS	NUS	NUS
s		1	2	2	3	3	5	5	10	20
ℓ_0	5.0	8.0	6.0	6.0	5.5	5.5	5.0	5.0	5.0	5.0
C	0.833	1.875	1.281	1.343	0.811	0.820	1.023	1.169	0.747	0.822
β	0.997	1.963	1.481	1.487	1.223	1.223	1.055	1.073	1.000	1.004
δ	1.041	1.963	1.982	1.986	1.694	1.690	1.163	1.186	1.075	1.061
R^2	1.000	1.000	1.000	1.000	0.993	0.993	0.998	0.998	1.000	1.000
$\hat{\nu}_*$	0.656	0.991	0.744	0.746	0.662	0.663	0.667	0.673	0.651	0.656
$\tilde{\nu}$	0.667	0.985	0.743	0.753	0.663	0.666	0.674	0.658	0.654	0.659
LGM	13.88	20.50	16.10	16.16	14.55	14.55	13.99	13.99	13.93	13.93

Table 4

Parameter estimates of the regression model with a KDE over $(-4, 4)$, $B \approx 0.05289$.

	MC	NUS	LMS	NUS	LMS	NUS	LMS	NUS	NUS	NUS
s		1	2	2	3	3	5	5	10	20
ℓ_0	4.5	8.5	5.5	5.5	5.0	5.0	4.5	4.5	4.0	4.0
C	0.277	0.433	0.285	0.277	0.150	0.188	0.142	0.010	0.033	0.079
β	1.035	2.429	2.008	2.005	1.719	1.734	1.287	1.258	1.013	0.993
δ	1.127	2.463	2.982	2.979	3.183	3.163	2.253	2.258	1.799	1.416
R^2	0.999	0.998	1.000	1.000	0.995	0.996	0.980	0.980	0.991	0.992
$\hat{\nu}_*$	0.808	1.503	1.150	1.149	0.958	0.968	0.823	0.804	0.699	0.733
$\tilde{\nu}$	0.783	1.528	1.146	1.142	0.966	0.967	0.830	0.805	0.751	0.761
LGM	16.91	29.95	23.67	23.67	20.49	20.50	17.75	17.67	17.20	16.99

To better visualize what happens when the density gets too small, we illustrate in Figure 3 the true density (in red) vs the estimated density (in blue) in the left tail, for Sobol' points with a digital shift only (left) and with a 31-bit LMS + shift (right), for $s = 1$ and $n = 2^{19}$, with $h = 2^{-6.5}$. Note that if we apply the LMS only to the first 19 bits, we obtain the same result as in the left figure, because then the LMS only permutes the points. A random shift modulo 1 with either Sobol' points or a lattice rule also gives the same result. On the other hand, Sobol'+NUS gives similar results as in the right panel. In the left plot, we see some blue bumps on the left of -4.1, but on the right of -4.1 we cannot distinguish the estimated density from the exact one. Each bump corresponds to a data point. The uniform points U_i are regularly spaced at distance exactly $1/n$, so the data points $X_i = \Phi^{-1}(U_i)$ are spaced at a distance that varies proportionally to the inverse of the density. When this distance is small enough, the KDE estimator is very smooth and almost equal to the exact density, and a slight shift of the points makes almost no difference, so the variance with respect to the random shift is almost zero. This is what we see on the right of -4.1. When we get further into the tail, on the other hand, the data points become too sparse, and the KDE has sparse bumps around the data points, as we can see on the left side. Note that with a randomly-shifted lattice rule, we have exactly the same behavior. Here the shift (modulo $1/n$) is approximately 7.766×10^{-7}, but the behavior is the same regardless of the shift. The right plot shows what happens with Sobol'+LMS applied to the first 31 bits. Here the bumps are no longer regularly-spaced, because the points have different shifts. As a result, the density estimator on the right of -4.1 varies much more. To illustrate how small the variance of the density estimator can be in the smooth part of the left plot, we computed the empirical variance at $x = -3, -2, -1, 0$, for the two cases. Table 5 gives $\log_2$ of the variance. We see that the variance with Sobol'+shift is extremely small at all x where the density is not too small, due to the regular spacing of the U_i. When looking at the convergence rate of the IV in terms of $\log n$ for fixed h (not shown here), we find that in terms of the IV model in (6.1), it is not the rate β that is large, but the constant C that is extremely small. With Sobol'+LMS, and also Sobol'+NUS which behaves similarly, the variance is not as small.

Table 5
$\log_2$ *of the variance with Sobol'+shift and Sobol'+LMS, for* $s = 1$ *and* $n = 2^{19}$.

Point set	$x = -3$	$x = -2$	$x = -1$	$x = 0$
Sobol'+Shift	-94.35	-93.55	-96.78	-93.33
Sobol'+LMS	-36.46	-37.01	-40.79	-41.44

7.2. Displacement of a cantilevel beam. Bingham [4] gives the following simple model of the displacement D of a cantilever beam with horizontal and vertical loads:

$$D = \frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} \tag{7.1}$$

in which L is the length of the beam, fixed to 100 inches, w and t are the width and thickness of the cross-section, taken as 4 and 2 inches, while X, Y, and E are assumed independent and normally distributed with means and standard deviations given as follows (in inches):

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

Here, the exact density is unknown, so we will not be able to compute the ISB and the MISE, but we will estimate the AISB as we did in the previous example, and then use it to estimate the optimal h and the MISE. A plot of the estimated density, obtained with a KDE with Sobol'+NUS and $n = 2^{19}$ points, is given in Figure 4. For the experiments reported here, we estimate the density over the interval $(0.407, 1.515)$, which covers about 99% of the density (it excludes roughly 0.5 % on each side). We also tried the shorter interval $(0.590, 1.293)$, which excludes 5% of the density on each side, and the results were very similar, except that the estimated constant B was about 17% smaller for the histogram and 25% smaller for the KDE.

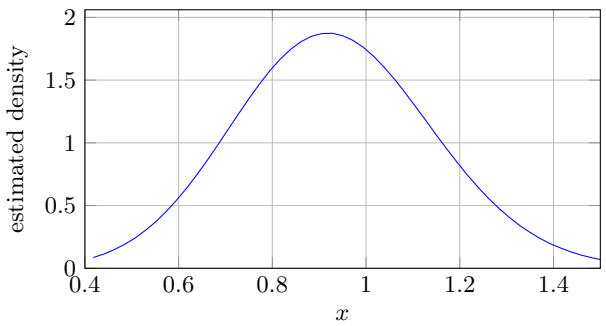

Figure 4. *Estimated density of the displacement of a cantilever beam.*

Table 6 gives the parameter estimates from our experiment, with the histogram and KDE estimators. RQMC increases the rate β significantly. For the KDE with Sobol' points, it goes from 1 to about 2. However, δ increases even more, from 1 to about 4. This means that although the variance decreases much faster than for MC as a function of n for fixed h, we cannot afford to decrease h very much to decrease the bias, so the MISE reduction is more limited than the IV reduction. The R^2 coefficient is very close to 1, showing that the linear model for $\log_2(\text{IV})$ fits very well. Figure 5 confirms this. It is reassuring to see that the estimate of B is about the same for all point sets, for both the histogram and the KDE. The estimated convergence rate of the MISE, $\hat{\nu}_*$, is improved by RQMC for the KDE but not for the histogram. However, RQMC reduces significantly the constant K in the MISE model for both the histogram and the KDE.

Figure 6 shows the estimated MISE as a function of n (with the estimated optimal h), as well as the estimated IV as a function of n, all in log scale, for the histogram and the KDE. The results for Sobol'+LMS and Sobol'+NUS are practically indistinguishable in those plots. We can see that although the MISE rate (the slope) is not improved much by RQMC, the MISE is nevertheless reduced by a significant factor. For example, with $n = 2^{19}$ and a KDE, the MISE is almost $2^6 = 64$ times smaller with Sobol'+LMS than with MC. For fixed h,

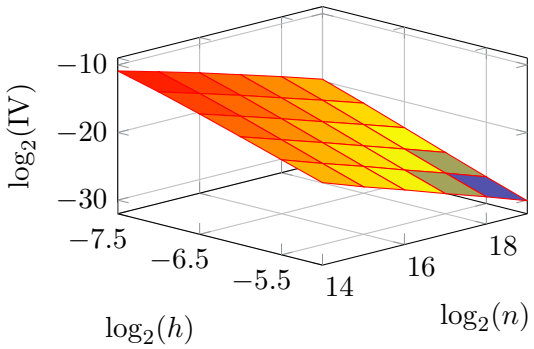

Figure 5. $\log_2(\mathrm{IV})$ *vs* $\log_2(n)$ *and* $\log_2(h)$ *for the KDE with Sobol'+NUS.*

the estimated IV converges at a faster rate with RQMC than with MC, as shown in the lower part of the figure.

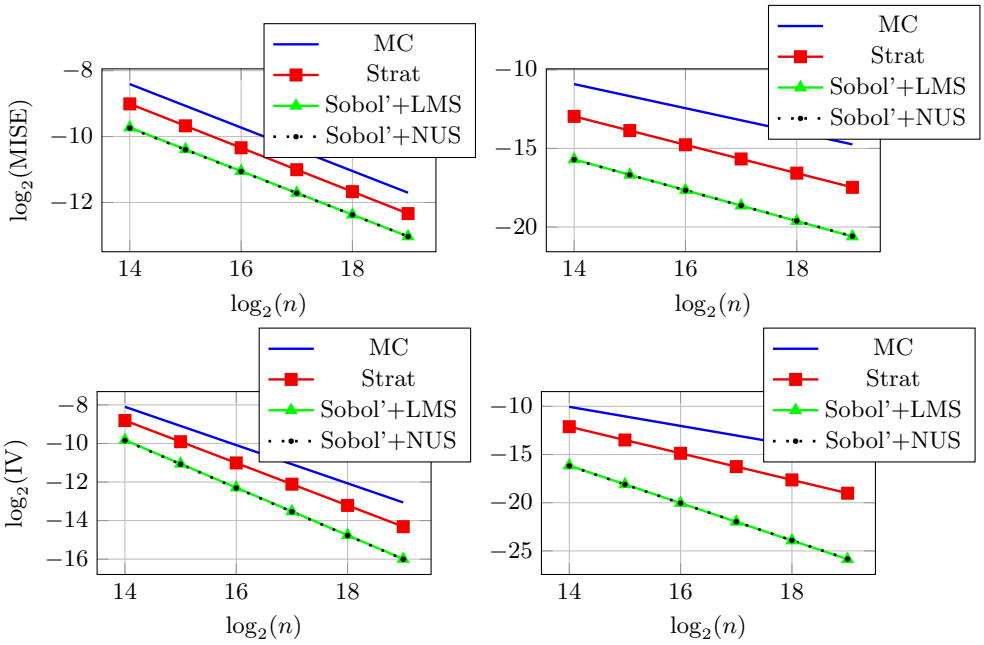

Figure 6. *Above: Estimated MISE as a function of* n *for the cantilever example for the histogram (left) and the KDE (right). Below: Estimated IV as a function of* n *for fixed* h*; we took* $h = 2^{-6} = 1/64$ *for both, the histogram (left) and for the KDE (right).*

7.3. A weighted sum of lognormals. In this example, we estimate the density of a weighted sum of lognormal random variables

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

Table 6

Experimental results for the density estimation of the displacement of a cantilever beam, over the interval $(0.407, 1.515)$.

	Histogram ($\alpha = 2$)				KDE ($\alpha = 4$)			
	MC	Strat	LMS	NUS	MC	Strat	LMS	NUS
C	0.831	0.424	0.130	0.119	0.109	0.022	1.8E-4	1.5E-4
β	0.992	1.102	1.234	1.232	0.991	1.380	1.943	1.932
δ	1.010	1.309	1.733	1.744	1.168	2.113	3.922	3.933
R^2	1.000	0.997	0.993	0.993	0.999	0.999	0.999	0.999
B	1.177	1.178	1.178	1.177	107.4	107.2	107.1	107.1
$\hat{\kappa}_*$	0.710	0.646	0.533	0.523	0.208	0.225	0.186	0.182
$\hat{\gamma}_*$	0.330	0.333	0.331	0.329	0.192	0.226	0.245	0.244
ℓ_*	6.758	6.962	7.188	7.186	5.909	6.443	7.090	7.085
$\hat{K}_*$	1.768	1.243	0.721	0.692	0.885	0.800	0.256	0.237
$\hat{\nu}_*$	0.659	0.666	0.661	0.658	0.767	0.903	0.981	0.974
LGM	11.70	12.34	13.03	13.03	14.74	17.48	20.60	20.58

where $\mathbf{Y} = (Y_1, \ldots, Y_s)^\mathsf{t}$ has a multinormal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Let $\mathbf{C} = \mathbf{A}\mathbf{A}^\mathsf{t}$ be a decomposition of $\mathbf{C}$. To generate $\mathbf{Y}$, we generate $\mathbf{Z}$ a vector of s independent standard normals by inversion, then put $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$. For MC, the choice of decomposition does not matter, but for RQMC it does, and here we take the decomposition used in principal component analysis (PCA) [1, 8]. We also tried sequential sampling (SS) and Brownian bridge sampling (BBS) but with them, RQMC did not improve the IV significantly as we will see with PCA.

This model has several applications. In one of them, for some positive constants ρ and s_0, by taking $w_j = s_0(s - j + 1)/s$, $e^{-\rho}\max(X - K, 0)$ is the payoff of a financial option based on the average value of a stock or commodity price at s observation times, under a geometric Brownian motion process. Estimating the density of this random payoff in its positive part is equivalent to estimating the density of X over the interval (K, ∞) (for simplicity we ignore the scaling factor $e^{-\rho}$). Note that when we estimate the KDE here, the realizations of X that are smaller than K are not discarded; they contribute to the KDE slightly above K. Discarding them would introduce a significant bias in the KDE due to a boundary effect at K.

For our numerical experiment, we take this special case with the same parameters as in [10]: $s = 12$, $s_0 = 100$, and $K = 101$. The matrix $\mathbf{C}$ is defined indirectly as follows. We have

$$Y_j = Y_{j-1}(\mu - \sigma^2)j/s + \sigma B(j/s)$$

where $Y_0 = 0$, $\sigma = 0.12136$, $\mu = 0.1$, and B is a standard Brownian motion. We will estimate the density of X over the interval $[a, b] = [K, K + 27.13]$. Approximately 0.5% of the density lies on the right of this interval and 29.05% lies on the left (this is when the option brings no payoff). Figure 7 shows a plot of the estimated density of $X - K$ obtained with a KDE with Sobol'+NUS and $n = 2^{19}$ points.

Table 7 summarizes the results of our experiments. Again, the linear model for the IV fits quite well in the selected area. With the KDE, RQMC improves β from 1 to about 5/3, which is significant, but at the same time δ increases (unfortunately) from about 1.1 to nearly

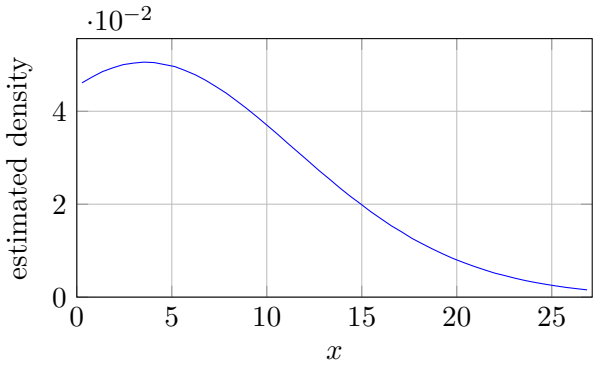

Figure 7. *Estimated density of the option payoff $X - K$.*

Table 7
Experimental results for the density estimation of the option payoff over the interval $(0, 27.13)$.

	Histogram ($\alpha = 2$)			KDE ($\alpha = 4$)		
	MC	LMS	NUS	MC	LMS	NUS
C	0.765	0.398	0.429	0.171	0.110	0.097
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
R^2	0.998	0.998	0.999	0.999	0.990	0.990
B	1.1E-5	1.1E-5	1.1E-5	1.1E-6	1.1E-6	1.1E-6
$\hat{\kappa}_*$	28.26	12.97	13.03	7.953	3.717	3.657
$\hat{\gamma}_*$	0.320	0.278	0.277	0.195	0.188	0.186
ℓ_*	1.269	1.581	1.565	0.715	1.670	1.668
$\hat{K}_*$	0.024	0.004	0.004	0.020	3.9E-4	3.6E-4
$\hat{\nu}_*$	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

5. This means we are very limited in how much we can decrease h to reduce the bias. The estimate of B is about the same for all point sets, which is reassuring. Somewhat surprisingly, the estimated MISE rate $\hat{\nu}_*$ is a bit worse for RQMC than for MC, due to the large δ. But the MISE is nevertheless significantly smaller for RQMC than for MC in the range of interest (about $2^5 = 32$ times smaller for the KDE), as shown in the upper panel of Figure 8, for which h was taken as the estimated optimal h from our model, as a function of n. That is, RQMC is truly beneficial for estimating the payoff density in this example. In the lower panel, we see that the estimated IV for fixed h converges faster with RQMC than with MC.

For comparison, when estimating the mean $\mathbb{E}[X]$ instead of the density, with Sobol'+LMS, the variance converges approximately as $\mathcal{O}(n^{-1.9})$ compared with $\mathcal{O}(n^{-1})$ for MC, and the variance is divided by a factor of about two millions compared with MC for $n = 2^{20}$. See [10], Table 3.

8. Conclusion. We explored RQMC combined with histograms and KDEs to estimate a density by simulation. RQMC can improve the IV and the MISE, sometimes by large factors, in situations in which the (effective) dimension is small. The improvement is usually more

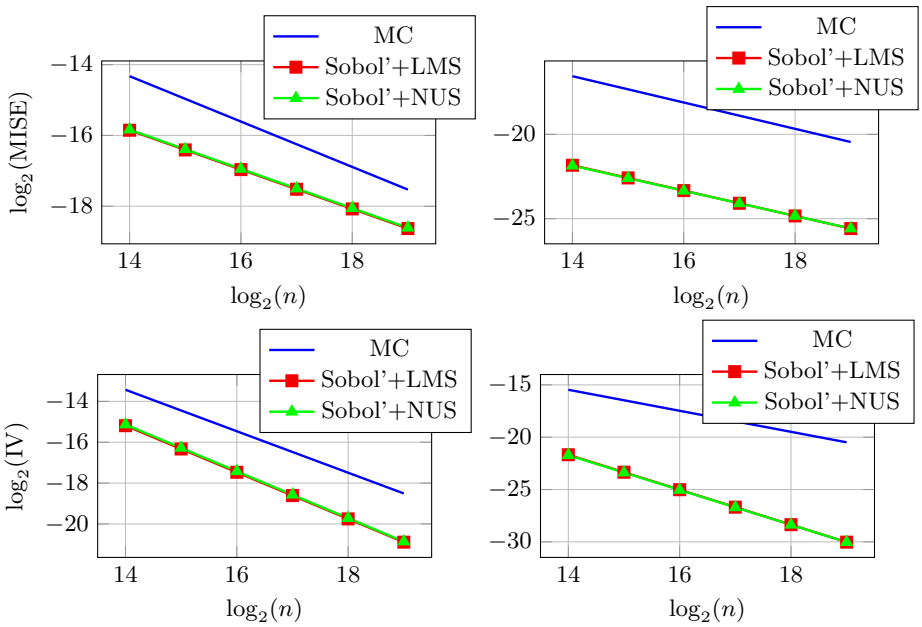

Figure 8. *Above: Estimated MISE as a function of n for the option payoff example for the histogram (left) and the KDE (right). Below: Estimated IV as a function of n for $h = 1/2$ for the histogram (left) and the KDE (right).*

limited when the dimension is large. We also found that the IV improvement degrades quickly as a function of h when $h \to 0$. In our empirical experiments, the IV is never significantly larger with KDE+RQMC than with KDE+MC, and it is often much smaller.

Acknowledgments. The idea of this work started during a workshop at the Banff International Research Station (BIRS) in October 2015. Most of the research was accomplished within a research program on quasi-Monte Carlo sampling methods at the Statistical and Applied Mathematical Sciences Institute (SAMSI), in North Carolina, in 2017–2018.

REFERENCES

[1] P. Acworth, M. Broadie, and P. Glasserman, *A comparison of some Monte Carlo and quasi-Monte Carlo techniques for option pricing*, in Monte Carlo and Quasi-Monte Carlo Methods 1996, P. Hellekalek, G. Larcher, H. Niederreiter, and P. Zinterhof, eds., vol. 127 of Lecture Notes in Statistics, Springer-Verlag, New York, 1998, pp. 1–18.

[2] K. Basu and A. B. Owen, *Transformations and Hardy–Krause variation*, SIAM Journal on Numerical Analysis, 54 (2016), pp. 1946–1966.

[3] A. Berlinet and L. Devroye, *A comparison of kernel density estimates*, Publications de l'Institut de Statistique de l'Université de Paris, 38 (1994), pp. 3–59. available at http://cgm.cs.mcgill.ca/~luc/np.html.

[4] D. Bingham, *Virtual library of simulation experiments*, 2017, https://www.sfu.ca/~ssurjano/canti.html.

[5] J. Dick and F. Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge, U.K., 2010.

[6] M. C. Jones, J. S. Marron, and S. J. Sheather, *A brief survey of bandwidth selection for density estimation*, Journal of the American Statistical Association, 91 (1996), pp. 401–407.

[7] M. Josephy, *Composing functions of bounded variation*, Proceedings of the American Mathematical Society, 83 (1981), pp. 354–356.
[8] P. L'Ecuyer, *Quasi-Monte Carlo methods with applications in finance*, Finance and Stochastics, 13 (2009), pp. 307–349.
[9] P. L'Ecuyer, *SSJ: Stochastic simulation in Java*. http://simul.iro.umontreal.ca/ssj/, 2016.
[10] P. L'Ecuyer, *Randomized quasi-Monte Carlo: An introduction for practitioners*, in Monte Carlo and Quasi-Monte Carlo Methods 2016, P. W. Glynn and A. B. Owen, eds., Berlin, 2017, Springer-Verlag. to appear.
[11] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, vol. 63 of SIAM CBMS-NSF Reg. Conf. Series in Applied Mathematics, SIAM, 1992.
[12] A. B. Owen, *Randomly permuted* (t, m, s)*-nets and* (t, s)*-sequences*, in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, H. Niederreiter and P. J.-S. Shiue, eds., vol. 106 of Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 299–317.
[13] A. B. Owen, *Scrambled net variance for integrals of smooth functions*, Annals of Statistics, 25 (1997), pp. 1541–1562.
[14] A. B. Owen, *Variance with alternative scramblings of digital nets*, ACM Transactions on Modeling and Computer Simulation, 13 (2003), pp. 363–378.
[15] A. B. Owen, *A randomized Halton algorithm in R*, tech. report, Stanford University, 2017. arXiv:1706.02808.
[16] V. C. Raykar and R. Duraiswami, *Fast optimal bandwidth selection for kernel density estimation*, in Proceedings of the 2006 SIAM International Conference on Data Mining, 2006, pp. 524–528.
[17] D. W. Scott, *On optimal and data-based histograms*, Biometrika, 66 (1979), pp. 605–610.
[18] D. W. Scott, *Multivariate Density Estimation*, Wiley, 2015.
[19] G. R. Terrell and D. W. Scott, *Variable kernel density estimation*, The Annals of Statistics, 20 (1992), pp. 1236–1265.
[20] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, 1995.