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). We investigate the effectiveness of replacing MC by randomized quasi Monte Carlo (RQMC) or by stratified sampling over the unit cube, to reduce the integrated variance (IV) and the mean integrated square error (MISE) for kernel density estimators. We show theoretically and empirically that the RQMC and stratified estimators can achieve substantial reductions of the IV and the MISE, and even faster convergence rates than MC in some situations, while leaving the bias unchanged. We also show that the variance bounds obtained via a traditional Koksma-Hlawka-type inequality for RQMC are much too loose to be useful when the dimension of the problem exceeds a few units. We describe an alternative way to estimate the IV, a good bandwidth, and the MISE, under RQMC or stratification, and we show empirically that in some situations, the MISE can be reduced significantly even in high-dimensional settings.

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

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

1. Introduction.

We are interested in estimating 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 we want to estimate f over some bounded interval $[a,b]$. A flurry of stochastic simulation applications fit this framework; see [1, 9], for example. The vector $\mathbf{U}$ represents the independent uniform random numbers that drive the simulation.

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):

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

*First version submitted to the editors July 6, 2018; Second version submitted April 29, 2019

Funding: This work has been supported by a Canada Research Chair, an Inria International Chair, an IVADO Ressearch 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 IIS-1837931, DMS-1521145 and DMS-1407397. The collaboration was also supported by the NSF Grant DMS-1638521 to SAMSI.

†DIRO, University of Montreal, 2920 Chemin de La Tour, Pavillon Aisenstadt, Montreal, QC, H3T 1N8, Canada (amal.ben.abdellah@umontreal.ca, 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).

Minimizing the MISE generally involves a bias-variance tradeoff.

The density is often estimated by a histogram for visualization, but one can do better with more refined techniques, such as a *kernel density estimator* (KDE), defined as follows. One selects a *kernel* $k : \mathbb{R} \to \mathbb{R}$, and a constant $h > 0$ called the *bandwidth*, which acts as a horizontal stretching factor for the kernel. The kernels considered here are smooth probability densities that are symmetric about 0. In our experiments, we will use the Gaussian kernel, which is the standard normal density. Given a sample $X_1, \ldots, X_n$, the KDE at $x \in \mathbb{R}$ is

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

Density estimation methods such as KDEs 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 of arbitrary size by choosing where to sample. With *crude Monte Carlo* (MC), we would estimate the density from a sample $X_1, \ldots, X_n$ of n *independent* realizations of X, obtained by simulation. Then the analysis is the same as if the data was collected from the real world, and the standard KDE methodology would apply [23]. In that context, the IV is $\mathcal{O}(1/nh)$ and the ISB is $\mathcal{O}(h^4)$, so the MISE is $\mathcal{O}(n^{-4/5})$ if h is chosen optimally. This is slower than the $\mathcal{O}(n^{-1})$ canonical rate for the variance when estimating the mean.

Our aim in this paper is to study if, when, and how using *randomized quasi-Monte Carlo* (RQMC) or *stratification* can provide a KDE with a smaller MISE than with crude MC. It is well known that when we estimate the mean $\mathbb{E}[X]$ by the average $\bar{X}_n = (X_1 + \cdots + X_n)/n$, under appropriate conditions, using RQMC provides an unbiased estimator whose variance converges at a faster rate (in n) than the MC variance [5, 12, 14, 17, 18]. This variance bound is easily proved by squaring a worst-case deterministic error bound obtained via a version of the *Koksma-Hlawka* (KH) inequality, which is a Hölder-type inequality that bounds the worst-case integration error by a product of the variation of g and the discrepancy of the set of points $\mathbf{U}$ at which g is evaluated. Hundreds of papers have studied this. Of course, the faster rate is an asymptotic property and the KH bound may hide a large constant factor, so it could happen that this bound is larger than the MC variance for a given n. But in applications, the true RQMC variance is often much smaller than both the bound and the MC variance, even for moderate sample sizes. The bottom line is that RQMC is practically useful in many applications, when estimating the mean by an average. Stratification of the unit hypercube also provably reduces the variance of $\bar{X}_n$, although its applicability degrades quickly with the dimension, and it is typically dominated by RQMC when the dimension exceeds 1 or 2 [12].

Since the KDE (1.1) at any given point is an average just like the estimator of an expectation, it seems natural to use RQMC to estimate a density as well, and to derive variance bounds via the same methods as for the mean estimator. This was the starting point of this paper. At first, we thought that the KH inequality would provide bounds on the IV of the KDE that converge faster for RQMC than for MC, and that a faster convergence rate of the MISE would follow. But things are not so simple. The best upper bound on the IV that KH gave us is $\mathcal{O}(n^{-2+\epsilon}h^{-2s})$ for any $\epsilon > 0$, while the ISB remains $\mathcal{O}(h^4)$ as with MC. This gives a bound of $\mathcal{O}(n^{-4/(2+s)+\epsilon})$ on the MISE if we select h to minimize this bound. The unwelcome h^{-2s} factor in the IV bound comes from the increase of the Hardy-Krause variation of each

summand in (1.1) as a function of the underlying uniforms when h decreases. This effect grows exponentially in s. To exploit the smaller power of n in the IV bound to reduce the MISE bound, one must simultaneously decrease the ISB. One can achieve this by taking a smaller h, which in turn drastically increases the IV bound. This limits seriously the rate at which the MISE bound can converge. The resulting rate for the bound beats the MC rate only for $s < 3$. For a special type of RQMC method, namely a digital net with a nested uniform scramble, we also prove that the IV and MISE rates are never worse than for MC.

For the KDE combined with a stratification of the unit hypercube into subcubes, which could be seen as a weak form of RQMC, we obtain bounds that converge as $\mathcal{O}(n^{-(s+1)/s}h^{-2})$ for the IV and $\mathcal{O}(n^{-(2/3)(s+1)/s})$ for the MISE. The latter beats the MC rate for all $s < 5$. These bounds are proved using arguments that do not involve KH and they are tight. We show examples where the IV and the MISE with stratification behave just like the bounds.

These results do not imply that stratification works better than RQMC, or that RQMC does not beat MC in more than two dimensions. The KH bounds are *only upper bounds* and nothing precludes that the true IV and MISE can be significantly smaller with RQMC than with MC or stratification, even if the RQMC variance bound is larger and converges more slowly. At a minimum, we should test empirically how the KDE really behaves in terms of IV and MISE under RQMC and under stratification. We also need a procedure to choose a good bandwidth h for the KDE with these sampling methods, since it will generally differ from a good h with MC. We do that in the second half of the paper. Our aim is to assess empirically the improvements achieved for reasonable sample sizes n in actual simulations. We use a regression model in log scale to estimate the IV and the MISE as functions of h and n, and the optimal h as a function of n. We find that RQMC often reduces the IV and the MISE significantly, even in more than 3 dimensions, and that it performs better than stratification. Sometimes, the convergence rate of the MISE is not improved but there is a significant gain in the constant and in the actual MISE. In all our experiments, the MISE was never larger with RQMC or stratification than with MC. We prove that this always holds for stratification. But for RQMC, we think that proving the observed gains in theorems would be very hard, hence the importance of testing with diverse numerical examples.

The remainder is organized as follows. In Section 2, we recall the definitions and basic properties of KDEs, including a strategy to find a good h under MC. In Section 3, we recall classical error and variance bounds for RQMC integration. In Section 4, we use classical QMC theory to derive KH bounds on the IV and the MISE for a KDE under RQMC, under reasonable assumptions. In Section 5 we derive IV and MISE bounds for a KDE combined with stratification. We have bounds that converge at a faster rate than for MC when the dimension is small. We also show that stratification never increases the IV or MISE compared with MC. In Section 6, we report on numerical experiments in which we estimate and compare the true IV and MISE of the KDE with MC, RQMC, and stratification, for various examples. We also provide a method to find a good bandwidth h, which is necessary for their effective implementation, and we use a regression model to capture how the IV and the MISE really behave in the examples. We give our conclusions in Section 7.

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 $h \in (0, 1]$, $c(n, h) \leq K\varphi(n, h)$.

2. Kernel density estimators with MC. We recall asymptotic properties of the KDE with MC when $nh \to \infty$ and $h \to 0$ together. The details can be found in [8, 23, 26], for example. The asymptotic MISE, IV, and ISB in this regime are denoted AMISE, AIV, and AISB, respectively. If $\mathrm{IV}(n, h)$ denotes the IV for a given (n, h), writing $\mathrm{AIV} = \tilde{g}(n, h)$ for some function $\tilde{g}$ means that $\lim_{n\to\infty, \tilde{g}(n,h)\to 0} \mathrm{IV}(n, h)/\tilde{g}(n, h) = 1$ 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 $r \geq 0$. We make the following assumptions in the rest of the paper.

Assumption 1. *The kernel k is a probability density function which is symmetric about 0, nondecreasing on $(-\infty, 0]$ and nonincreasing on $[0, \infty)$, has a finite mode $k(0) < \infty$ and its second moment is strictly positive and finite. Thus, $\mu_0(k) = 1$, $\mu_1(k) = 0$, and $0 < \mu_2(k) < \infty$.*

Assumption 2. *The density f is at least four times differentiable over $[a, b]$ (including at the boundaries) and $R(f^{(r)}) < \infty$ for $r \leq 4$, where $f^{(r)}$ is the rth derivative of f.*

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 $h^5 = Q/n$ where $Q := \mu_0(k^2)/[(\mu_2(k))^2 R(f'')]$, if Q is well-defined and finite. This gives

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

Thus, finding a good h amounts to finding a good approximation of $R(f'')$. But since f is precisely the unknown function that we want to estimate, this seems to be a circular problem. However, perhaps surprisingly, a viable approach is to estimate $R(f'')$ by estimating f'' also via a KDE, integrating its square over $[a, b]$, and plugging this estimate into the formula for the optimal h [2, 8, 22, 23]. To do that, one needs to select a good h to estimate f'' by a KDE. The asymptotically optimal h depends in turn on $R(f^{(4)})$ where $f^{(4)}$ is the fourth derivative of f. Then $R(f^{(4)})$ can be estimated by integrating the KDE 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 is

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

We will use this strategy to estimate a good h in our experiments with MC and RQMC, with a Gaussian kernel, with $r_0 = 2$.

In this paper we always take h to be the same for all $x \in [a,b]$. It is possible to improve the 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 [23, 25].

3. Error and variance bounds for RQMC integration. We recall classical error and variance bounds for RQMC integration. They can be found in [5], [10], [15], and [17], 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 $P_n = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subset [0,1]^s$ is

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

Let $\mathfrak{v}$ denote a subset of coordinates, $\mathfrak{v} \subseteq \mathcal{S} := \{1, \ldots, s\}$. For any $\mathbf{u} = (u_1, \ldots, u_s) \in [0,1]^s$ we denote by $\mathbf{u}_\mathfrak{v}$ the projection of $\mathbf{u}$ on the coordinates in $\mathfrak{v}$ and by $(\mathbf{u}_\mathfrak{v}, \mathbf{1})$ the point $\mathbf{u}$ in which u_j has been replaced by 1 for each $j \notin \mathfrak{v}$. We write $g_\mathfrak{v} := \partial^{|\mathfrak{v}|} g / \partial \mathbf{u}_\mathfrak{v}$ for the partial derivative of g with respect to each of the coordinates in $\mathfrak{v}$. When $g_\mathfrak{v}$ exists and is continuous for $\mathfrak{v} = \mathcal{S}$, the *Hardy-Krause variation* of g is

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

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

$$\mathrm{Var}[E_n] = \mathbb{E}[E_n^2] = \mathcal{O}(n^{-2+\epsilon}). \tag{3.3}$$

4. Bounding the convergence rate of the AIV for a KDE with RQMC. Replacing MC by RQMC does not affect the bias, because $\hat{f}_n(x)$ has the same expectation for both, but it can change the variance. Before trying to bound the variance under RQMC, it is instructive to recall how it is bounded under MC. To compute (or bound) the IV over an interval $[a,b]$, we compute (or bound) $\mathrm{Var}[\hat{f}_n(x)]$ at an arbitrary point $x \in [a,b]$ and integrate this bound over x. Since $\hat{f}_n(x)$ is an average of n independent realizations of $Y(x) = k((x-X)/h)/h$,

it suffices to compute $\mathrm{Var}[Y(x)]$ and we have $\mathrm{Var}[\hat{f}_n(x)] = \mathrm{Var}[Y(x)]/n$. With the change of variable $w = (x-v)/h$, we obtain [23, page 143]:

$$\begin{aligned}\mathrm{Var}[Y(x)] &= \frac{1}{h^2}\int_{-\infty}^{\infty} k^2\left(\frac{x-v}{h}\right) f(v)\mathrm{d}v - \mathbb{E}^2[Y(x)] \\ &= \frac{1}{h}\int_{-\infty}^{\infty} k^2(w) f(x-hw)\mathrm{d}w - \mathbb{E}^2[Y(x)] = \frac{f(x)}{h}\int_{-\infty}^{\infty} k^2(w)\mathrm{d}w \; - f^2(x) + \mathcal{O}(h).\end{aligned}$$

Integrating over $x \in [a,b]$, gives $\mathrm{IV} = p_0\,\mu_0(k^2)/(nh) - R(f)/n + \mathcal{O}(h/n)$ where $p_0 = \int_a^b f(x)\mathrm{d}x \le 1$.

With RQMC, this also holds for a single RQMC point $\mathbf{U}_i$ and $X = X_i = g(\mathbf{U}_i)$, but to obtain $\mathrm{Var}[\hat{f}_n(x)]$, we can no longer just divide $\mathrm{Var}[Y(x)]$ by n, because the n realizations of $Y(x)$ are not independent. RQMC is effective if and only if these realizations are negatively correlated, in the sense that if $Y_i = h^{-1}k((x-X_i)/h)$, then $\sum_{i\neq j}\mathrm{Cov}(Y_i,Y_j) \le 0$. This would imply that RQMC can never be worse than MC, but this seems hard to prove.

We now take a different path, in which we examine how the KH inequality (3.2) can be used to bound $\mathrm{Var}[\hat{f}_n(x)]$. With $X = g(\mathbf{U})$, we can write

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\tilde{g}(x,\mathbf{U}_i) \quad \text{where} \quad \tilde{g}(x,\mathbf{U}_i) := Y_i = \frac{1}{h}k\left(\frac{x-g(\mathbf{U}_i)}{h}\right). \tag{4.1}$$

Thus, $\hat{f}_n(x)$ can be interpreted as an RQMC estimator of $\mathbb{E}[\tilde{g}(x,\mathbf{U})] = \int_{[0,1]^s}\tilde{g}(x,\mathbf{u})\mathrm{d}\mathbf{u}$. To apply the bound in (3.3) to this estimator, we need to bound the variation of $\tilde{g}(x,\cdot)$, by bounding each term of the sum in (3.1).

To provide insight, we first examine the special case where $s=1$ and g is nondecreasing over $[0,1]$, under Assumption 1. Then $\tilde{g}(x,u) = k((x-g(u))/h)/h$ is nonincreasing over the u with $g(u) \le x$ and nondecreasing over u with $g(u) \ge x$. In that case $V_{\mathrm{HK}}(\tilde{g}(x,\cdot))$ is the ordinary one-dimensional total variation, and then

$$V_{\mathrm{HK}}(\tilde{g}(x,\cdot)) \le \left|\frac{1}{h}k(0) - \frac{1}{h}k\left(\frac{x-g(0)}{h}\right)\right| + \left|\frac{1}{h}k(0) - \frac{1}{h}k\left(\frac{x-g(1)}{h}\right)\right| \le \frac{2k(0)}{h}. \tag{4.2}$$

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 $V_{\mathrm{HK}}(\tilde{g}(x,\cdot)) \le 2Mk(0)/h$. The factor of 2 is necessary because k is potentially increasing and then decreasing within each of those intervals. Note that there are one-dimensional point sets P_n with $D^*(P_n) \le 1/n$. With such point sets, we obtain $\mathrm{Var}[\hat{f}_n(x)] \le (2Mk(0))^2/(nh)^2$, so $\mathrm{AIV} = \mathcal{O}((nh)^{-2})$. With $h = \Theta(n^{-1/3})$, this gives $\mathrm{AMISE} = \mathcal{O}(n^{-4/3})$.

We now consider the general case $s \ge 1$. To bound $V_{\mathrm{HK}}(\tilde{g}(x,\cdot))$ we will make a similar change of variables as for the IV under MC. We need additional assumptions on k and g.

Assumption 3. *The kernel function $k : \mathbb{R} \to \mathbb{R}$ is s times continuously differentiable and its derivatives up to order s are integrable and uniformly bounded over $\mathbb{R}$.*

Assumption 4. *Let $g : [0,1]^s \to \mathbb{R}$ be piecewise monotone in each coordinate u_j when the other coordinates are fixed, with a number of monotone pieces (which is 1 plus the number of*

times that the function switches from strictly decreasing to strictly increasing or vice-versa, in u_j) that is bounded uniformly in $\mathbf{u}$ by an integer M_j. We also assume that the first-order partial derivatives of g are continuous and that $\|g_{\mathfrak{v}}\|_\infty < \infty$ for all $\mathfrak{v} \subseteq \mathcal{S}$. This implies that any product of partial derivatives of g of order at most one in each variable is integrable.

Because the Hardy–Krause variation (3.1) involves mixed partial derivatives of $\tilde{g}(x,\cdot)$ of order up to s, things unfortunately become considerably more complicated than for MC. Roughly speaking, every derivative causes an additional factor h^{-1}, while we may dispose of only one such factor through a change of variables. This is reflected in Proposition 4.1 below. Similar to the one-dimensional case, we need to take into account how often g changes its monotonicity direction, and this is captured by the M_j's. For each $j \in \mathcal{S}$, let

$$G_j = \left\| \prod_{\ell \in \mathcal{S}\setminus\{j\}} g_{\{\ell\}} \right\|_\infty,$$
$$c_j = M_j \cdot \left(\|k^{(s)}\|_1 \cdot G_j + \mathbb{I}(s=2) \cdot \|k^{(s)}\|_\infty \cdot \|g_{\{1,2\}}\|_1 \right) < \infty,$$

where $\mathbb{I}(\cdot)$ is the indicator function, so the expression for c_j contains an extra term when $s = 2$. The source of this extra term is that for $s = 2$, the only partition of $\{1,2\}$ which contains no singletons is $\{1,2\}$ itself, and it gives a term in $h^{-2} = h^{-s}$, whereas for $s > 2$, all the extra terms are $\mathcal{O}(h^{-s+1})$. Our main result of this section is:

Proposition 4.1. *Let k and g satisfy Assumptions 1, 3 and 4, and $c = \min_{j\in\mathcal{S}} c_j$. Then the Hardy-Krause variation of $\tilde{g}(x,\mathbf{u}) = h^{-1}k((x - g(\mathbf{u}))h^{-1})$ (as a function of $\mathbf{u}$) satisfies*

$$V_{\mathrm{HK}}(\tilde{g}(x,\cdot)) \le ch^{-s} + \mathcal{O}(h^{-s+1}). \tag{4.3}$$

Note that the constants c_j depend on g via M_j and G_j (plus an extra term when $s = 2$). A large M_j means that g changes the sign of its slope many times in the direction of coordinate j. A large G_j means that the product of the slopes in the directions of the other coordinates can attain large values. When we have both, then c_j is large. Observe that the bound (4.3) uses the *smallest* c_j. In case g is constant with respect to one coordinate $\ell \neq j$, then $G_j = 0$, $c_j = 0$, and $c = 0$. That is, the term in h^{-s} disappears from (4.3) and the bound becomes $\mathcal{O}(h^{-s+1})$. This agrees with the fact that g is then effectively a $(s-1)$-dimensional function. Likewise, if g is almost constant (has very little variation) with respect to one or more coordinate(s) $\ell \neq j$, then G_j and therefore c_j will be small, unless the other terms in the product are very large. Before proving this proposition, we state a corollary that bounds the AIV and the AMISE rates under RQMC.

Corollary 4.2. *Let k and g satisfy Assumptions 1 to 4. For a KDE with kernel k, with the underlying observations obtained via sets P_n of n RQMC points for which $D^*(P_n) = \mathcal{O}(n^{-1+\epsilon})$ for all $\epsilon > 0$ when $n \to \infty$, by combining (4.3) with (3.2) and squaring, we find that* $\mathrm{AIV} = \mathcal{O}(n^{-2+\epsilon}h^{-2s})$ *for all $\epsilon > 0$. Then, by taking $h = \Theta(n^{-1/(2+s)})$, we obtain that* $\mathrm{AMISE} = \mathcal{O}(n^{-4/(2+s)+\epsilon})$ *for all $\epsilon > 0$. The exponent of n in this AMISE bound beats the MC rate for $s < 3$ and is almost equal to the MC rate for $s = 3$.*

Let $\Pi(\mathfrak{v})$ denote the set of all partitions of a set of coordinate indices $\mathfrak{v} \subseteq \mathcal{S}$, and let $\Pi_1(\mathfrak{v})$ denote the subset of all partitions that contain at least one singleton. For each partition $P \in \Pi_1(\mathfrak{v})$, we select a particular singleton and denote it by $\{j(P)\}$. Removing that singleton from P yields a partition of $\mathfrak{v}^* = \mathfrak{v} \setminus \{j(P)\}$ which we denote by P^*.

The proof of the proposition will use the following lemma, which describes when the aforementioned change of variable works and how it removes a factor $1/h$ from the bound.

Lemma 4.3. *Let Assumptions 1, 3 and 4 hold, let $h > 0$, $\mathfrak{v} \subseteq \mathcal{S}$, and $P \in \Pi_1(\mathfrak{v})$. Then*

$$\int_{[0,1]^{|\mathfrak{v}|}} \left| k^{(|P|)}\left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \cdot \prod_{\mathfrak{w} \in P} g_\mathfrak{w}(\mathbf{u}_\mathfrak{v}, \mathbf{1}) \right| \mathrm{d}\mathbf{u}_\mathfrak{v} \;\le\; h \cdot M_{j(P)} \cdot \left\| \prod_{\mathfrak{w} \in P^*} g_\mathfrak{w} \right\|_\infty \cdot \|k^{(|P|)}\|_1.$$

Proof. We assume without loss of generality that $1 \in \mathfrak{v}$ and $j(P) = 1$. We make the change of variables

$$u_1 \mapsto w = (x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1}))/h. \tag{4.4}$$

For any $\mathbf{u}_\mathfrak{v} \in [0,1]^{|\mathfrak{v}|}$ with fixed $\mathbf{u}_{\mathfrak{v}^*} \in [0,1]^{|\mathfrak{v}|-1}$, we partition $[0,1]$ into a part $\mathcal{N}(\mathbf{u}_{\mathfrak{v}^*})$ where $g(\mathbf{u}_\mathfrak{v}, \mathbf{1})$ is constant in u_1 and into sets $\mathcal{D}_l(\mathbf{u}_{\mathfrak{v}^*})$, $1 \le l \le L(\mathbf{u}_{\mathfrak{v}^*}) \le M_1$, on which it is either strictly decreasing or strictly increasing in u_1. Since $g_{\{1\}}$ is continuous by assumption, each of these sets is measurable. The restriction of $g_{\{1\}}$ to $\mathcal{N}(\mathbf{u}_{\mathfrak{v}^*})$ equals 0 identically. In all the other sets $\mathcal{D}_l(\mathbf{u}_{\mathfrak{v}^*})$ we apply the change of variables (4.4). Considering this in the left-hand side of the claim leads to

$$\begin{aligned}
&\int_{[0,1]^{|\mathfrak{v}|}} \left| k^{(|P|)}\left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \prod_{\mathfrak{w} \in P} g_\mathfrak{w}(\mathbf{u}_\mathfrak{v}, \mathbf{1}) \right| \mathrm{d}\mathbf{u}_\mathfrak{v} \\
&\quad= \int_{[0,1]^{|\mathfrak{v}|-1}} \sum_{l=1}^{L(\mathbf{u}_{\mathfrak{v}^*})} \int_{\mathcal{D}_l(\mathbf{u}_{\mathfrak{v}^*})} \left| k^{(|P|)}\left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \prod_{\mathfrak{w} \in P} g_\mathfrak{w}(\mathbf{u}_\mathfrak{v}, \mathbf{1}) \right| \mathrm{d}u_1 \mathrm{d}\mathbf{u}_{\mathfrak{v}^*} \\
&\quad\le h \int_{[0,1]^{|\mathfrak{v}|-1}} L(\mathbf{u}_{\mathfrak{v}^*}) \int_{-\infty}^{\infty} \left| k^{(|P|)}(w) \prod_{\mathfrak{w} \in P^*} g_\mathfrak{w}(\mathbf{u}_\mathfrak{v}, \mathbf{1}) \right| \mathrm{d}w \mathrm{d}\mathbf{u}_{\mathfrak{v}^*} \\
&\quad\le h \cdot M_1 \cdot \|k^{(|P|)}\|_1 \cdot \left\| \prod_{\mathfrak{w} \in P^*} g_\mathfrak{w} \right\|_\infty,
\end{aligned}$$

where we used Hölder's inequality in the last step. ∎

Proof of Proposition 4.1. We rewrite each summand w.r.t. $\mathfrak{v} \subseteq \mathcal{S}$ in (3.1) with the help of Faà di Bruno's formula (see [7, Proposition 1]) as follows

$$\begin{aligned}
\int_{[0,1]^{|\mathfrak{v}|}} |\tilde{g}_\mathfrak{v}(x, \mathbf{u}_\mathfrak{v}, \mathbf{1})| \, \mathrm{d}\mathbf{u}_\mathfrak{v} &= \frac{1}{h} \int_{[0,1]^{|\mathfrak{v}|}} \left| \sum_{P \in \Pi(\mathfrak{v})} k^{(|P|)}\left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \prod_{\mathfrak{w} \in P} \frac{\partial^{|\mathfrak{w}|}}{\partial \mathbf{u}_\mathfrak{w}} \left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \right| \mathrm{d}\mathbf{u}_\mathfrak{v} \\
&\le \sum_{P \in \Pi(\mathfrak{v})} \frac{1}{h^{|P|+1}} \int_{[0,1]^{|\mathfrak{v}|}} \left| k^{(|P|)}\left(\frac{x - g(\mathbf{u}_\mathfrak{v}, \mathbf{1})}{h}\right) \prod_{\mathfrak{w} \in P} g_\mathfrak{w}(\mathbf{u}_\mathfrak{v}, \mathbf{1}) \right| \mathrm{d}\mathbf{u}_\mathfrak{v}.
\end{aligned} \tag{4.5}$$

If $P \in \Pi_1(\mathfrak{v})$, we bound the corresponding summand in (4.5) via Lemma 4.3. If $P \notin \Pi_1(\mathfrak{v})$, we apply Hölder's inequality to obtain the upper bound

$$\frac{1}{h^{|P|+1}} \cdot \|k^{(|P|)}\|_\infty \cdot \Big\| \prod_{\mathfrak{w} \in P} g_{\mathfrak{w}} \Big\|_1 .$$

Furthermore, we observe that each element of $P \in \Pi(\mathfrak{v}) \setminus \Pi_1(\mathfrak{v})$ has a cardinality of at least 2. Therefore, P can contain at most $\lfloor |\mathfrak{v}|/2 \rfloor$ elements. This gives the following bound on $V_{\mathrm{HK}}(\tilde{g}(x,\cdot))$, which holds for any $j \in \mathcal{S}$:

$$\begin{aligned} V_{\mathrm{HK}}(\tilde{g}(x,\cdot)) &\le \sum_{\emptyset \ne \mathfrak{v} \subseteq \mathcal{S}} \left[\sum_{P \in \Pi_1(\mathfrak{v})} \frac{M_{j(P)}}{h^{|P|}} \|k^{(|P|)}\|_1 \cdot \Big\| \prod_{\mathfrak{w} \in P^*} g_{\mathfrak{w}} \Big\|_\infty + \sum_{P \in \Pi(\mathfrak{v}) \setminus \Pi_1(\mathfrak{v})} \frac{1}{h^{|P|+1}} \|k^{(|P|)}\|_\infty \cdot \Big\| \prod_{\mathfrak{w} \in P} g_{\mathfrak{w}} \Big\|_1 \right] \\ &\le h^{-s} M_j G_j \, \|k^{(s)}\|_\infty + \mathcal{O}(h^{-s+1}) + \sum_{\emptyset \ne \mathfrak{v} \subseteq \mathcal{S}} h^{-\lfloor |\mathfrak{v}|/2 \rfloor - 1} \sum_{P \in \Pi(\mathfrak{v}) \setminus \Pi_1(\mathfrak{v})} \|k^{(|P|)}\|_\infty \cdot \Big\| \prod_{\mathfrak{w} \in P} g_{\mathfrak{w}} \Big\|_1 . \end{aligned}$$

For $s = 1$ this already proves the claim. For $s = 2$, the only partition of $\mathcal{S}$ that contains no singleton is $\mathcal{S}$ itself and the result follows. For $s \ge 3$ we have $\lfloor s/2 \rfloor + 1 \le s - 1$, and then

$$\sum_{\emptyset \ne \mathfrak{v} \subseteq \mathcal{S}} h^{-\lfloor |\mathfrak{v}|/2 \rfloor - 1} \sum_{P \in \Pi(\mathfrak{v}) \setminus \Pi_1(\mathfrak{v})} \|k^{(|P|)}\|_\infty \cdot \Big\| \prod_{\mathfrak{w} \in P} g_{\mathfrak{w}} \Big\|_1 = \mathcal{O}(h^{-s+1}).$$

■

The bound of Proposition 4.1 suggests that the IV could converge at a much worse rate with RQMC than with MC when s is large. However, the next proposition, based on a result of [19], provides a different bound that does not grow as h^{-2s} when h decreases, for a particular type of RQMC point set, namely a (t, m, s)-net in base 2 randomized by a nested uniform scramble (NUS). This type of point set contains 2^m points in s dimensions, the t parameter measures the uniformity in some sense (the smaller the better) [5, 15], and the NUS shuffles the points in some particular way [16, 17]. We state the following result for base $b = 2$, but it can be extended to a general prime base $b \ge 2$.

Proposition 4.4. *Let P_n be a (t, m, s)-net in base 2 randomized by NUS, and let Assumption 1 hold. Then the IV of $\hat{f}_n$ satisfies*

$$\mathrm{IV} \le 2^t 3^s \mu_0(k^2)/(nh) + \mathcal{O}(h/n).$$

Moreover for any fixed $s \ge 1$, there is a fixed $t \ge 0$ for which we know how to construct a (t, m, s)-net P_n in base 2 for any integer $m \ge 1$. By using such a sequence of nets with NUS, we get $\mathrm{IV} = \mathcal{O}(1/(nh))$, and then by taking $h = \Theta(n^{-1/5})$, we obtain $\mathrm{MISE} = \mathcal{O}(n^{-4/5})$. That is, the MISE never converges at a worse rate than with plain MC.

Proof. Let $\mathrm{Var}_{\mathrm{MC}}$ and $\mathrm{Var}_{\mathrm{NUS}}$ denote the variance under MC and under NUS, respectively. Likewise, for any given pair (n, h), let $\mathrm{IV}_{\mathrm{MC}}(n,h)$ and $\mathrm{IV}_{\mathrm{NUS}}(n,h)$ denote the IV under MC and under NUS, respectively, and similarly for the MISE. Under Assumption 1, $\tilde{g}(x,\cdot)$ is square-integrable over $[0,1]^s$ for any $x \in [a,b]$, so we can apply Theorem 1 of [19], which tells us that

$$\mathrm{Var}_{\mathrm{NUS}}[\tilde{g}(x,\mathbf{U})] \;\le\; 2^t 3^s \mathrm{Var}_{\mathrm{MC}}[\tilde{g}(x,\mathbf{U})].$$

By integrating, we obtain $\mathrm{IV}_{\mathrm{NUS}}(n,h) \le 2^t 3^s \mathrm{IV}_{\mathrm{MC}}(n,h)$. We saw earlier that $\mathrm{IV}_{\mathrm{MC}}(n,h) \le \mu_0(k^2)/(nh) - R(f)/n + \mathcal{O}(h/n)$, and this proves the displayed inequality.

For the second part, for any $s \ge 1$, there is a fixed $t \ge 0$ for which we know how to construct a (t,s)-sequence in base 2; see [24] and [15, Section 4.5], for example. For any integer m, the first 2^m points of such a sequence form a (t,m,s)-net in base 2. Thus, t can be assumed to be bounded uniformly in m, and therefore $\mathrm{IV}_{\mathrm{NUS}}(n,h) = \mathcal{O}(\mathrm{IV}_{\mathrm{MC}}(n,h)) = \mathcal{O}(1/(nh))$. Since MC and NUS give the same ISB, we also have $\mathrm{MISE}_{\mathrm{NUS}}(n,h) \le 2^t 3^s \mathrm{MISE}_{\mathrm{MC}}(n,h) = \mathcal{O}(1/(nh) + h^4) = \mathcal{O}(h^{-4/5})$ if we take $h = \Theta(n^{-1/5})$. ∎

5. Stratified sampling of $[0,1)^s$. In this section, we examine how plain stratified sampling of the unit hypercube can reduce the IV of the KDE compared with MC. We consider point sets P_n constructed as in Assumption 5 below. This type of stratified sampling can never increase the IV compared with MC. We prove this via a standard variance decomposition argument. Then, under the additional condition that $g(\mathbf{u})$ is monotone with respect to each coordinate of $\mathbf{u}$, we prove an IV bound that converges at a faster rate than the IV under MC when $s < 5$. The KH inequality and the variation of g are not involved in the IV bound developed here; we work directly with the variance. For this reason, the bound will not contain the annoying factor h^{-2s} as in Proposition 4.1. On the other hand, the exponent of n will not be as good. Our main results are Propositions 5.1 and 5.2, and Corollary 5.3.

Assumption 5. *The hypercube $[0,1)^s$ is partitioned into $n = q^s$ congruent cubic cells $S_{\mathbf{i}} := \prod_{j=1}^s [i_j/q, (i_j+1)/q)$, $\mathbf{i} \in \mathbf{I} = \{\mathbf{i} = (i_1, i_2, \ldots, i_s) : 0 \le i_j < q \text{ for each } j\}$, for some integer $q \ge 2$. We construct $P_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ by sampling one point uniformly in each subcube $S_{\mathbf{i}}$, independently across the subcubes, and put $X_i = g(\mathbf{U}_i)$ for $i = 1, \ldots, n$.*

Proposition 5.1. *Under Assumptions 1 and 5, the IV of a KDE $\hat{f}_n$ with kernel k never exceeds the IV of the same estimator under standard MC, which satisfies* $\mathrm{IV} \le \mu_0(k^2)/(nh) - R(f)/n + \mathcal{O}(h/n)$.

Proof. We can decompose the variance under MC as

$$\begin{aligned}\mathrm{Var}[\tilde{g}(x,\mathbf{U})] &= \mathbb{E}[\mathrm{Var}[\tilde{g}(x,\mathbf{U}) \mid \mathbf{U} \in S_{\mathbf{i}}] + \mathrm{Var}[\mathbb{E}[\tilde{g}(x,\mathbf{U}) \mid \mathbf{U} \in S_{\mathbf{i}}] \\ &= \frac{1}{n}\sum_{\mathbf{i}\in\mathbf{I}} \mathrm{Var}[\tilde{g}(x,\mathbf{U}) \mid \mathbf{U} \in S_{\mathbf{i}}] + \frac{1}{n}\sum_{\mathbf{i}\in\mathbf{I}} (\mu_{\mathbf{i}} - \mu)^2,\end{aligned}$$

where $\mu = \mathbb{E}[\tilde{g}(x,\mathbf{U})]$ and $\mu_{\mathbf{i}} = \mathbb{E}[\tilde{g}(x,\mathbf{U}) \mid \mathbf{U} \in S_{\mathbf{i}}]$. By sampling exactly one point in each cell $S_{\mathbf{i}}$, the stratified sampling removes the second term, and the first term remains the same. Therefore, stratification never increases $\mathrm{Var}[\hat{f}_n(x)]$. The second term also indicates how the amount of variance reduction depends on how the $\mu_{\mathbf{i}}$ vary between boxes. ∎

Now we know that stratification can do no harm. To show that it can also improve the convergence rate of the MISE, we need additional conditions.

Assumption 6. *For each $j \in \mathcal{S}$, the function $g : [0,1)^s \to \mathbb{R}$ is* monotone *in u_j when the other $s-1$ coordinates are fixed, and the direction of monotonicity in u_j (nondecreasing or nonincreasing) is the same for all values of the other coordinates. Without loss of generality, we will assume in the rest of the paper that it is nondecreasing in each coordinate. (If it is*

nonincreasing in u_j, one can simply replace u_j by $1-u_j$ in the definition of g and this does not change the distribution of $X = g(\mathbf{U})$.)

Proposition 5.2. *Let Assumptions 1, 5 and 6 hold and let $\hat{f}_n$ be a KDE with kernel k obtained from $X_1, X_2, \ldots, X_n$. Then the IV of $\hat{f}_n$ satisfies*

$$\mathrm{IV} \le (b-a)s \cdot k^2(0) \cdot h^{-2} n^{-(s+1)/s}.$$

Corollary 5.3. *Under Assumptions 1, 2, 5 and 6, the AMISE bound is minimized by taking $h = \kappa n^{-(s+1)/(6s)}$ with $\kappa^6 = [(b-a)s \cdot k^2(0)]/[(\mu_2(k))^2 R(f'')/2]$, and this gives* AMISE $= Kn^{-\nu}$ *with $\nu = (2/3)(s+1)/s$ and $K = (b-a)s \cdot k^2(0)\kappa^{-2} + (\mu_2(k))^2 R(f'')\kappa^4/4$. The exponent of n in this bound beats the MC rate for all $s < 5$ and is equal to the MC rate for $s = 5$.*

The corollary is straightforward to prove. It suffices to minimize with respect to h the sum of the AISB (given in Section 2) and the IV bound.

Our proof of Proposition 5.2 is inspired by [13, Proposition 6]. It combines three lemmas that we state and prove before proving the proposition. These lemmas could be useful in other contexts as well. The first lemma bounds the variance of the KDE in terms of a uniform bound K_n on the variance of the empirical cdf estimator. The other lemmas bound K_n.

Let F be the cdf of the random variable X. For any $x \in \mathbb{R}$, the empirical cdf of a sample $X_1, X_2 \ldots, X_n$ evaluated at x is $\hat{F}_n(x) := n^{-1} \sum_{i=1}^n \mathbb{I}[X_i \le x]$. We denote the difference by $\Delta_n(x) = \hat{F}_n(x) - F(x)$. Let $K_n := \sup_{x \in \mathbb{R}} \mathrm{Var}[\Delta_n(x)] = \sup_{x \in \mathbb{R}} \mathrm{Var}[\hat{F}_n(x)]$.

Lemma 5.4. *Under Assumption 1, for all $x \in \mathbb{R}$, we have*

$$\mathrm{Var}[\hat{f}_n(x)] \le 2h^{-2} k(0) K_n.$$

Proof. We will prove the inequality

$$\mathrm{Var}[\hat{f}_n(x)] \le K_n h^{-4} \left(\int_{\mathbb{R}} k'\left(\frac{x-z}{h}\right) \mathrm{d}z \right)^2. \tag{5.1}$$

The result follows from this inequality by making the change of variable $w = (x-z)/h$ in the integral. To prove (5.1), we rewrite the variance of $\hat{f}_n(x)$ using integration by parts as follows:

$$\begin{aligned}
\mathbb{E}\left[\left|\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\right|^2\right] &= h^{-2}\,\mathbb{E}\left[\left|\int_{\mathbb{R}} k\left(\frac{x-z}{h}\right) \mathrm{d}\hat{F}_n(z) - \int_{\mathbb{R}} k\left(\frac{x-z}{h}\right) \mathrm{d}F(z)\right|^2\right] \\
&= h^{-4}\,\mathbb{E}\left[\left|\int_{\mathbb{R}} \left(\hat{F}_n(z) - F(z)\right) k'\left(\frac{x-z}{h}\right) \mathrm{d}z\right|^2\right] \\
&= h^{-4}\,\mathbb{E}\left[\int_{\mathbb{R}}\int_{\mathbb{R}} \Delta_n(y)\Delta_n(z) k'\left(\frac{x-y}{h}\right) k'\left(\frac{x-z}{h}\right) \mathrm{d}y\mathrm{d}z\right]. \qquad (5.2)
\end{aligned}$$

Observe that $\mathbb{E}[\Delta_n(z)] = 0$ since $\mathbb{E}[\hat{F}_n(z)] = F(z)$. Consequently,

$$\mathbb{E}[\Delta_n(y)\Delta_n(z)] = \mathrm{Cov}[\Delta_n(y), \Delta_n(z)] \le K_n.$$

Finally, (5.1) follows by interchanging the expectation and the integrals in (5.2). ∎

For all $x \in \mathbb{R}$, define $\mathbf{H}(x) = \{\mathbf{u} \in [0,1)^s : g(\mathbf{u}) \le x\}$ and its complement $\overline{\mathbf{H}}(x) = [0,1)^s \setminus \mathbf{H}(x)$. Under Assumption 5, let $\mathcal{B}(x)$ be the set of subcubes $S_{\mathbf{i}}$ that have a nonempty intersection with both $\mathbf{H}(x)$ and $\overline{\mathbf{H}}(x)$. The next lemma bounds the cardinality of $\mathcal{B}(x)$ when g is nondecreasing.

Lemma 5.5. *Under Assumptions 5 and 6, for all $x \in \mathbb{R}$, $|\mathcal{B}(x)| \le sn^{(s-1)/s}$.*

Proof. Let $\mathbf{I}_0 = \{\mathbf{i} = (i_1, i_2, \ldots, i_s) \in \mathbf{I}$ such that $\min_j i_j = 0\}$, which is the set of indices $\mathbf{i} \in \mathbf{I}$ for which at least one face of $S_{\mathbf{i}}$ lies on a face of $[0,1)^s$ that contains the origin. Each of the s faces of $[0,1)^s$ touches at most b^{s-1} elements of $\mathbf{I}_0$ and therefore $|\mathbf{I}_0| \le sq^{s-1} = sn^{(s-1)/s}$. Now, for any $\mathbf{i} = (i_1, i_2, \ldots, i_s) \in \mathbf{I}_0$, consider the *diagonal string* of subcubes $S_{\mathbf{i}'(k)}$ with $\mathbf{i}'(k) = (i_1 + k, i_2 + k, \ldots, i_s + k)$ for $0 \le k < q - \max_j i_j$. We argue that for any $x \in \mathbb{R}$, at most one subcube in this diagonal string can belong to $\mathcal{B}(x)$. Indeed, suppose that two distinct subcubes in the string belong to $\mathcal{B}(x)$, say $S_{\mathbf{i}'(k_1)}$ and $S_{\mathbf{i}'(k_2)}$ for $k_1 < k_2$. Since both subcubes contain points from $\mathbf{H}(x)$ and $\overline{\mathbf{H}}(x)$, there must be two points $\mathbf{u}_1 \in S_{\mathbf{i}'(k_1)} \cap \overline{\mathbf{H}}(x)$ and $\mathbf{u}_2 \in S_{\mathbf{i}'(k_2)} \cap \mathbf{H}(x)$. This implies that $g(\mathbf{u}_2) \le x < g(\mathbf{u}_1)$ while $\mathbf{u}_1 < \mathbf{u}_2$ coordinatewise, which contradicts the assumption that g is nondecreasing. Since there are no more than $sn^{(s-1)/s}$ diagonal strings and each contains at most one element of $\mathcal{B}(x)$, the result follows. ∎

Lemma 5.6. *Under Assumptions 5 and 6, $K_n \le (s/4)h^{-2}n^{-(s+1)/s}$.*

Proof. For each $\mathbf{i} \in \mathbf{I}$, consider the random variables

$$\delta_{\mathbf{i}}(x) = |P_n \cap \mathbf{H}(x) \cap S_{\mathbf{i}}| - n \operatorname{vol}(\mathbf{H}(x) \cap S_{\mathbf{i}}).$$

We make three observations. Firstly, $S_{\mathbf{i}}$ contains exactly one point of P_n by Assumption 5. Consequently, each $\delta_{\mathbf{i}}(x)$ is a Bernoulli random variable (with parameter $p = n \operatorname{vol}(\mathbf{H}(x) \cap S_{\mathbf{i}})$) minus its mean p and, therefore, $\operatorname{Var}[\delta_{\mathbf{i}}(x)] = p(1-p) \le 1/4$. Secondly, for each $\mathbf{i}$ for which $S_{\mathbf{i}} \notin \mathcal{B}(x)$, $\delta_{\mathbf{i}}(x) = 0$, so $\operatorname{Var}[\delta_{\mathbf{i}}(x)] = 0$. Thirdly, for any two distinct subcubes, the positions of the points of P_n in these subcubes are independent. As a consequence of these three observations we see that

$$\operatorname{Var}[\Delta_n(x)] = \operatorname{Var}\left[\frac{1}{n}\sum_{\mathbf{i}\in\mathbf{I}} \delta_{\mathbf{i}}(x)\right] = \frac{1}{n^2} \sum_{\mathbf{i}: S_{\mathbf{i}} \in \mathcal{B}(x)} \operatorname{Var}[\delta_{\mathbf{i}}(x)] \le \frac{1}{4n^2} sn^{-(s+1)/s}.$$

By applying Lemmas 5.4 and 5.5 we then obtain

$$\operatorname{Var}[\hat{f}_n(x)] \;\le\; 2h^{-2}k(0)K_n \;\le\; \frac{k(0)}{2(hn)^2}|\mathcal{B}(x)| \;\le\; \frac{sk(0)}{2}h^{-2}n^{-(s+1)/s}.$$

∎

Proof of Proposition 5.2. Combining Lemmas 5.4 and 5.6 and integrating the variance bound with respect to x over $[a,b]$ yields the result. ∎

In the above arguments, we assumed that the strata were cubic, but this is not necessary. We could instead partition $[0,1)^s$ into $n = \prod_{j=1}^s q_j$ cells congruent to $\prod_{j=1}^d [0, 1/q_j)$, subject to the condition $\max_j q_j \le \lambda \min_j q_j$ for some $\lambda < \infty$. Then one can bound the cardinality of $\mathcal{B}(x)$ and $\operatorname{Var}[\hat{F}_n(x)]$ in a similar way. Non-cubic strata make sense if g varies more in

some directions than in others. Finally, our bounds are proved under the assumption that g is monotone, but this assumption is *not necessary* for stratification to improve the MISE and/or its convergence rate.

6. Empirical Study. Our analysis in the previous sections was in terms of (asymptotic) bounds. Here, we study the IV and MISE behavior from a different viewpoint: our goal is to estimate empirically how they really behave in a range of values of n and h that one is likely to use. For this, we use a simple regression model to approximate the true IV and MISE in the region of interest. For some examples, we estimate the model parameters from simulated data, test the goodness of fit of the regression models in-sample and out-of-sample, and show how the model permits one to estimate the optimal h as a function of n, as well as the resulting MISE and its convergence rate, under RQMC.

6.1. Experimental setting and regression models for the local behavior of the IV, ISB, and MISE. We will use the following models to approximate the true IV and ISB in a limited range of values of n and h of interest:

$$\text{IV} \approx Cn^{-\beta}h^{-\delta} \quad \text{and} \quad \text{ISB} \approx Bh^{\alpha}, \tag{6.1}$$

for positive constants C, β, δ, and B, that can be estimated as explained below, and with $\alpha = 4$. This gives MISE $\approx Cn^{-\beta}h^{-\delta} + Bh^{\alpha}$. The bounds derived in the previous sections have this form, and this motivates our model, but here we want to estimate the true values, which generally differ from the bounds. Once the parameters are estimated, we can estimate the optimal h by minimizing the MISE estimate for any given n in the selected range. In our setting, this estimated MISE is a convex function of h. Taking the derivative with respect to h and setting it to zero yields $h^{\alpha+\delta} = [C\delta/(B\alpha)]n^{-\beta}$. 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

$$\text{MISE} \approx Kn^{-\nu} \tag{6.2}$$

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.

To estimate the model parameters for IV, we take the log to obtain the linear model

$$\log(\text{IV}) \approx \log C - \beta \log n - \delta \log h, \tag{6.3}$$

and we estimate the parameters C, β, and δ by linear regression. Since n is always a power of 2 for our RQMC points, we take all the logarithms in base 2. In our experiments, we selected a set of 36 pairs (n, h) with $n = 2^{14}, \ldots, 2^{19}$ and $h = h_0, \ldots, h_5$ where $h_j = h_0 2^{j/2} = 2^{-\ell_0 + j/2}$ and $2\ell_0$ is an integer selected from pilot runs. This selection of ℓ_0 is the only step that requires human intervention.

For each n and each point set (MC, Stratification or RQMC), we generate a sample of size n, sort the sample, and then compute the density estimator for each h, for this sample.

That is, we use the same sample for all estimation methods and all h. We make $n_r = 100$ independent replications of this procedure, which gives us independent replicates of the density estimator for the selected pairs (n, h). To obtain an unbiased estimator of the integral that defines the IV, we take a stratified sample of $n_e = 1024$ evaluation points over the interval $[a, b]$, compute the empirical variance of the KDE at each point, based on the n_r replications, and take the average multiplied by $(b-a)$. Larger values of n_e gave about the same estimates.

We approximate the ISB in (6.1) by the AISB, for which $\alpha = 4$ and $B = (\mu_2(k))^2 R(f'')/4$. For the Gaussian kernel, used in all our experiments, Assumptions 1 and 3 are satisfied, and $\mu_2(k) = 1$. We estimate the integral $R(f'')$ as explained in Section 2, using RQMC instead of MC to improve the accuracy.

Once we have the estimates $\hat{\kappa}_*$ and $\hat{\gamma}_*$ of κ_* and γ_*, we test the models out-of-sample by making an independent set of simulation experiments with pairs (n, h) that satisfy $h = \hat{h}_*(n) := \hat{\kappa}_* n^{-\hat{\gamma}_*}$ (the estimated optimal h) for a series of values of n. At each of these pairs (n, h), we sample n_r fresh 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 $\log(\text{MISE})$ vs $\log n$ to re-estimate the parameters K and ν in (6.2) and assess the goodness-of-fit. In our results, we denote these new estimates by $\tilde{K}$ and $\tilde{\nu}$. Of course, these model testing steps are not needed if one wishes to only estimate the density f and not to study the convergence properties.

In the end, we also compare the efficiencies of different methods for the same example by comparing their estimated MISE for $n = 2^{19}$ with the h recommended by the model. We denote by e19 the value of $-\log_2(\text{MISE})$ for $n = 2^{19}$; that is, we have $\text{MISE} = 2^{-e19}$. The efficiency gain of RQMC vs MC can be assessed by comparing their e19 values.

The point sets considered in our experiments were: (1) independent points (MC); (2) stratification of the unit cube (Stratif); (3) a Sobol' point set with a left random matrix scrambling and random digital shift (Sobol'+LMS); and (4) a Sobol' point set with nested uniform scrambling (Sobol'+NUS). The last two are well-known RQMC point sets [12, 17, 20] and we view stratification as a weak form of RQMC. The short names in parentheses are used in the plots and tables. These point sets and randomizations are implemented in SSJ [11], which we used for our experiments.

6.2. A normalized sum of standard normals. As in [21], we construct a set of test functions with arbitrary dimension s and for which the density f of X is always the standard normal, $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$ for any s. For this, let $Z_1, \ldots, Z_s$ be s independent standard normal random variables generated by inversion and put $X = (a_1 Z_1 + \cdots + a_s Z_s)/\sigma$, where $\sigma^2 = a_1^2 + \cdots + a_s^2$. For this simple example, the density is already known, so there is no need to estimate it, but this is convenient for testing the methodology, since it permits us to compute and compare unbiased estimators of the IV, ISB, and MISE for both MC and RQMC. For MC, these quantities do not depend on s, but for RQMC, the IV and MISE do depend on s, and we want to see in what way.

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 stratification. Assumption 4 holds only if we truncate the normal distributions of the Z_j, but it makes no significant difference on our empirical results if the truncated range contains the interval $[-8, 8]$, for example, so from the practical viewpoint, we can ignore it.

Table 1

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
$\hat{\kappa}_*$	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
e19	17.01	34.06	24.39	24.38	20.79	20.80	17.88	17.79	17.28	17.07	17.05

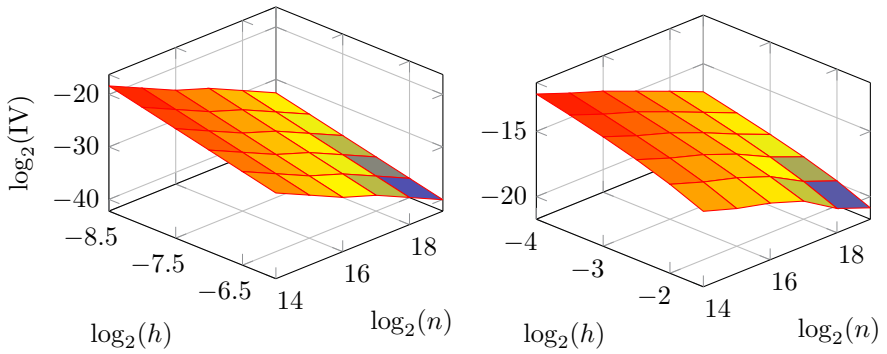

Figure 1. $\log_2(\mathrm{IV})$ *for the KDE with Sobol'+NUS for* $s = 1$ *(left) and* $s = 20$ *(right).*

We estimate the density over $[a, b] = [-b, b] = [-2, 2]$. In our first experiment, we take $a_1 = \cdots = a_s = 1$, so all the coordinates have the same importance (which is disadvantageous for RQMC). Later, we will consider varying coefficients a_j. Table 1 summarizes the results 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 $\gamma = 0.2$, $\nu = 0.8$, and $\beta = \delta = 1$. 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 give almost the same values. 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 the IV model, the fraction R^2 of variance explained by this model, and the estimated B. The other quantities are defined in Subsection 6.1, except

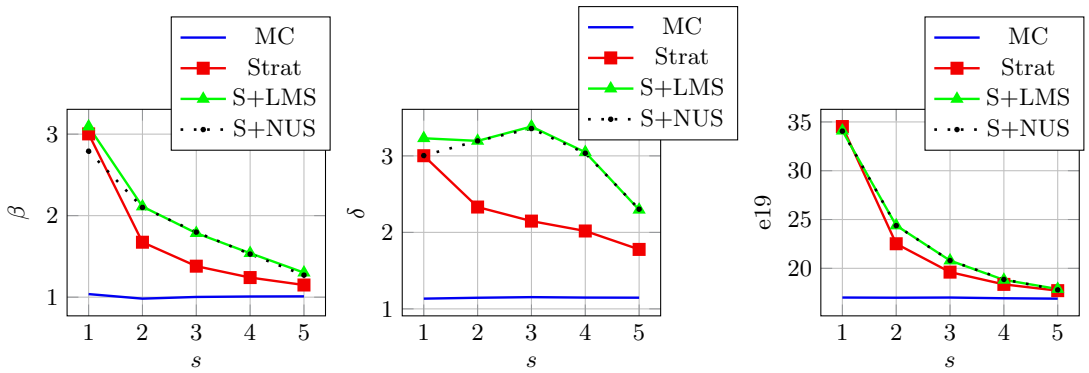

Figure 2. *Estimated β, δ, and e19 with MC, Stratification, Sobol'+LMS, and Sobol'+NUS.*

for $\ell_* = -\log_2 \hat{h}_*(2^{19})$, which gives an idea of the optimal h for $n = 2^{19}$.

Recall that the rates $\tilde{\nu}$ and e19 are 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 close to 1, which means that the log-log linear model is reasonably good in the area considered. The estimate of B is $B \approx 0.0418$ (same first three digits) for all s and all RQMC methods. Thus, this estimator of B has very little variance. The MISE reduction of RQMC vs MC can be assessed by comparing their values of e19 given in the last row. 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, for two values of s. We made similar plots for several $s > 1$ and all point sets, and the linear approximation looked reasonable in all cases. Figure 2 shows the estimated β, δ, and e19, for $s = 1, \ldots, 5$, for various point sets. Stratification, shown here and not in the table, 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 table (e19) is that for all s, the RQMC methods never have a larger MISE than MC. Their MISE is much smaller for small s, and becomes almost the same as for MC when s gets large. The MISE rate $\tilde{\nu}$ behaves similarly. 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 s, with RQMC, β is significantly larger than $\tilde{\nu}$, which means that the IV converges much faster as a function of n when h is fixed than when h varies with n 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 rapidly, and this limits the MISE reduction that we can achieve.

Here f is the standard normal density and $R(f'') = [-b\left(2b^2 - 1\right)e^{-b^2} + 3\int_0^b e^{-x^2}\mathrm{d}x]/4\pi$. For $b = 2$, this gives $R(f'') \approx 0.19018$, so the true constant B in the AISB is $B = R(f'')/4 \approx 0.04754$, whereas our estimate was 0.0418 for all s and all point sets. The difference is not due to noise, but is a bias coming from the fact that we estimated $R(f'')$ via KDE with finite n. We verified empirically that when we estimate these quantities with a larger n, the bias decreases slowly and appears to converge 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 values of e19 were almost the same.

We now take different coefficients (weights) a_j in the linear combination of the Z_j that defines X. Our purpose is to illustrate that there are situations where RQMC can perform very well with the KDE even when the dimension s is large. This can occur for example if the effective dimension is not large; i.e., when $g(\mathbf{u})$ depends mostly on just a few coordinates of $\mathbf{u}$, and does not vary much with respect to the other coordinates [4, 10]. To illustrate this, we take $a_j = 2^{-j}$ for $j = 1, \ldots, s$, and we repeat the same set of experiments as we did for equal weights, to estimate the density over $[-2, 2]$.

Table 2

Parameter estimates for the KDE under Sobol'+LMS, for a weighted sum of normals with $a_j = 2^{-j}$.

s	MC	2	4	10	20	50	100
C	0.171	0.173	0.038	6.7E-3	8.0E-3	7.3E-3	7.9E-3
β	1.000	2.100	1.650	1.420	1.427	1.425	1.429
δ	1.137	3.189	3.745	3.626	3.582	3.604	3.603
$\hat{K}_*$	0.213	0.183	0.080	0.032	0.035	0.033	0.035
$\hat{\nu}_*$	0.779	1.168	0.852	0.745	0.753	0.750	0.752
$\tilde{\nu}$	0.774	1.176	0.892	0.750	0.730	0.758	0.752
e19	16.96	24.76	19.71	18.96	18.98	18.99	19.04

Table 2 summarizes our findings for Sobol'+LMS, for s up to 100. The results with Sobol'+NUS are very similar. For $s = 1$, the results are obviously the same as for our previous setting, but they diverge when we increase s. For example, in the previous setting, the MISE estimate with $n = 2^{19}$ for $s = 2$, 10, and 100, was $2^{-24.38}$, $2^{-17.28}$, and $2^{-17.05}$, respectively, whereas with the new weights, it is $2^{-24.76}$, $2^{-18.96}$, and $2^{-19.04}$, respectively. For $s = 100$, in particular, the MISE with RQMC and $n = 2^{19}$ was about the same as for MC in the previous setting, and it is reduced by a factor of 4 in the present setting. We also see from the table that in 10 or more dimensions, the convergence rate of the MISE is not improved, but the constant is improved (empirically). As expected, when s increases beyond about 10, all the model parameters appear to stabilize as a function of s. In the previous setting, they were stabilizing around the MC values, but now they stabilize to different values. For example, in $s = 100$ dimensions, β was near the MC value of 1, and now it is about 1.4.

6.3. Displacement of a cantilevel beam. Bingham [3] gives the following simple model of the displacement D of a cantilever beam with horizontal and vertical loads:

$$D = \frac{4L^3}{Ewt}\sqrt{\frac{Y^2}{t^4} + \frac{X^2}{w^4}} \tag{6.4}$$

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

We want to estimate the density of the relative displacement $\tilde{X} = D/D_0 - 1$, where $D_0 = 2.2535$ inches. Here, the exact density is unknown, so unbiased estimators of the ISB and the MISE are not available, but we can estimate the AISB as in the previous example, and 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 3. For the experiments reported here, we estimate the density of $\tilde{X}$ over the interval $[0.407, 1.515]$, which covers about 99% of the density (it excludes roughly 0.5 % on each side).

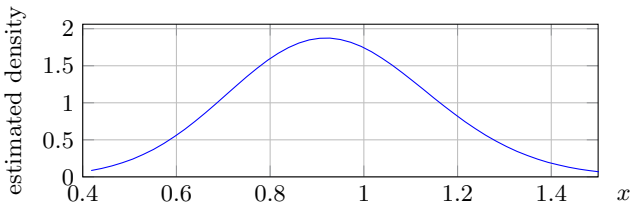

Figure 3. *Estimated density of $\tilde{X}$, the relative displacement of a cantilever beam.*

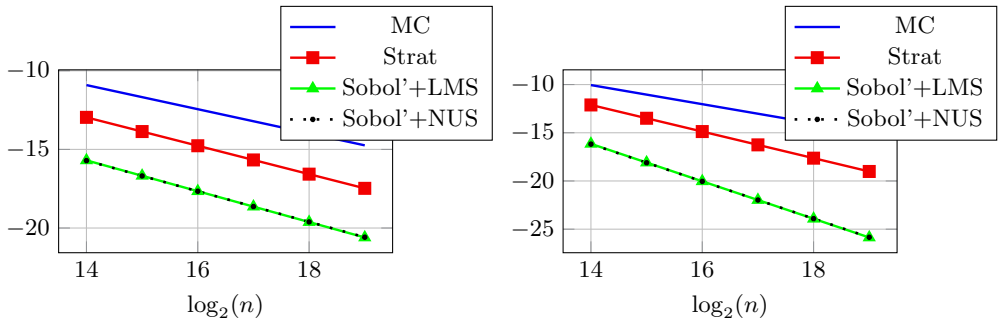

Figure 4. *Estimated MISE (left) and IV (right) as a function of n for $h = 2^{-6}$, for the cantilever example.*

Table 3 gives the parameter estimates from our experiment. RQMC increases the rate β 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 limited. RQMC improves both the estimated rate $\hat{\nu}_*$ and the constant K in the MISE model.

Figure 4 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. The results for Sobol'+LMS and Sobol'+NUS are practically indistinguishable in those plots. We see that although the MISE rate (slope) is not improved much by RQMC, the MISE is nevertheless reduced by a significant factor. With $n = 2^{19}$, the MISE is almost $2^6 = 64$ times smaller with Sobol'+LMS

Table 3

Experimental results for the KDE, for the displacement of a cantilever beam, over the interval $[0.407, 1.515]$.

	MC	Strat	LMS	NUS
C	0.109	0.022	1.8E-4	1.5E-4
β	0.991	1.380	1.943	1.932
δ	1.168	2.113	3.922	3.933
R^2	0.999	0.999	0.999	0.999
B	107.4	107.2	107.1	107.1
$\hat{\kappa}_*$	0.208	0.225	0.186	0.182
$\hat{\gamma}_*$	0.192	0.226	0.245	0.244
ℓ_*	5.909	6.443	7.090	7.085
$\hat{K}_*$	0.885	0.800	0.256	0.237
$\hat{\nu}_*$	0.767	0.903	0.981	0.974
e19	14.74	17.48	20.60	20.58

than with MC. For fixed h, the IV converges at a faster rate with RQMC than with MC.

Here, g is strictly decreasing in E and strictly increasing in both X and Y. Therefore, Corollary 5.3 applies. The asymptotic parameter values are $\beta = 4/3$, $\delta = 2$, and $\nu = 0.889$, which are very close to what we found empirically for stratification (see Table 3).

6.4. A weighted sum of lognormals. In this example, we estimate the density of a weighted sum of lognormals: $X = \sum_{j=1}^{s} w_j \exp(Y_j)$ 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) [6, 10]. 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 $e^{-\rho}$). When we compute the KDE here, the realizations of X 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 [12]: $s = 12$, $s_0 = 100$, 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 $B(\cdot)$ is a standard Brownian motion. We estimate the density of $\tilde{X} = X - K$ over the interval $[a, b] = [0, 27.13]$. 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 5 shows a plot of the estimated density of $\tilde{X} = X - K$ obtained from a KDE with Sobol'+NUS and $n = 2^{19}$ points.

Table 4 summarizes the results of our experiments. Again, the linear model for the IV fits extremely well in the selected area. RQMC improves β from 1 to about 5/3, which is

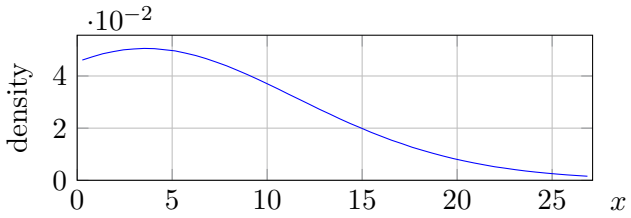

Figure 5. *Estimated density of the option payoff $X - K$.*

Table 4
Experimental results for the density estimation of the option payoff over the interval [0, 27.13].

	MC	LMS	NUS
C	0.171	0.110	0.097
β	1.005	1.671	1.663
δ	1.151	4.907	4.930
R^2	0.999	0.990	0.990
B	1.1E-6	1.1E-6	1.1E-6
$\hat{\kappa}_*$	7.953	3.717	3.657
$\hat{\gamma}_*$	0.195	0.188	0.186
ℓ_*	0.715	1.670	1.668
$\hat{K}_*$	0.020	3.9E-4	3.6E-4
$\hat{\nu}_*$	0.780	0.750	0.745
e19	20.45	25.59	25.58

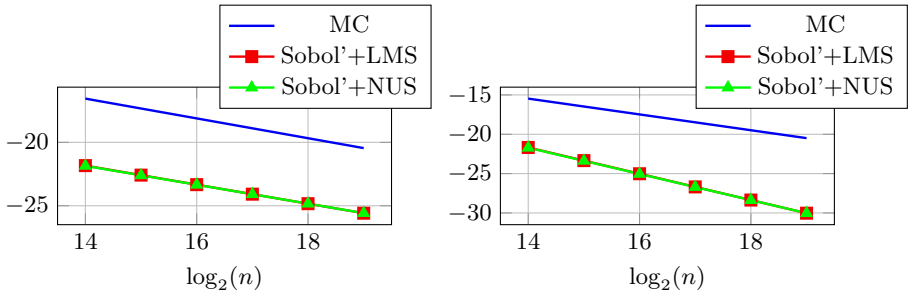

Figure 6. *Estimated MISE as a function of n (left) and estimated IV as a function of n for $h = 1/2$ (right).*

significant, but at the same time δ increases from about 1.1 to nearly 5. This means we are very limited in how much we can decrease h to reduce the bias. On the other hand, this empirical δ is not as bad as the one in the AIV bound of Corollary 4.2, which gives $\delta = 2s = 24$. The estimate of B is again about the same for all point sets. Somewhat surprisingly, in the region considered, the estimated MISE rate $\hat{\nu}_*$ is not better for RQMC than for MC, due to the large δ, but the MISE is nevertheless about 32 times smaller for RQMC than for MC in the range of interest, as shown in Figure 6, 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 12-dimensional example. In the lower panel, we see that the estimated IV for fixed h converges faster with RQMC than with MC.

7. Conclusion. We explored RQMC combined with 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 more limited when the effective 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 was never larger with KDE+RQMC than with KDE+MC, and was often much smaller.

Acknowledgments. The idea of this work started during a workshop at the Banff International Research Station (BIRS) in October 2015. Part 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. We thank Ilse Ipsen for her support in organizing this program.

REFERENCES

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation*, Springer-Verlag, New York, 2007.
[2] 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.
[3] D. Bingham, *Virtual library of simulation experiments*, 2017, https://www.sfu.ca/~ssurjano/canti.html.
[4] R. E. Caflisch, W. Morokoff, and A. Owen, *Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension*, J. of Computational Finance, 1 (1997), pp. 27–46.
[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] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York, 2004.
[7] M. Hardy, *Combinatorics of partial derivatives*, Electronic Journal of Combinatorics, 13 (2006), pp. Research Paper 1, 13.
[8] 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.
[9] A. M. Law, *Simulation Modeling and Analysis*, McGraw-Hill, New York, fifth ed., 2014.
[10] P. L'Ecuyer, *Quasi-Monte Carlo methods with applications in finance*, Finance and Stochastics, 13 (2009), pp. 307–349.
[11] P. L'Ecuyer, *SSJ: Stochastic simulation in Java*. http://simul.iro.umontreal.ca/ssj/, 2016.
[12] P. L'Ecuyer, *Randomized quasi-Monte Carlo: An introduction for practitioners*, in Monte Carlo and Quasi-Monte Carlo Methods: MCQMC 2016, P. W. Glynn and A. B. Owen, eds., Berlin, 2018, Springer, pp. 29–52.
[13] P. L'Ecuyer, C. Lécot, and B. Tuffin, *A randomized quasi-Monte Carlo simulation method for Markov chains*, Operations Research, 56 (2008), pp. 958–975.
[14] P. L'Ecuyer and C. Lemieux, *Recent advances in randomized quasi-Monte Carlo methods*, in Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications, M. Dror, P. L'Ecuyer, and F. Szidarovszky, eds., Kluwer Academic, Boston, 2002, pp. 419–474.
[15] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, vol. 63 of SIAM CBMS-NSF Reg. Conf. Series in Applied Mathematics, SIAM, 1992.
[16] 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.
[17] A. B. Owen, *Scrambled net variance for integrals of smooth functions*, Annals of Statistics, 25 (1997), pp. 1541–1562.
[18] A. B. Owen, *Latin supercube sampling for very high-dimensional simulations*, ACM Transactions on Modeling and Computer Simulation, 8 (1998), pp. 71–102.
[19] A. B. Owen, *Scrambling Sobol and Niederreiter-Xing points*, Journal of Complexity, 14 (1998), pp. 466–489.

[20] A. B. Owen, *Variance with alternative scramblings of digital nets*, ACM Transactions on Modeling and Computer Simulation, 13 (2003), pp. 363–378.
[21] A. B. Owen, *A randomized Halton algorithm in R*, tech. report, Stanford University, 2017. arXiv:1706.02808.
[22] 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.
[23] D. W. Scott, *Multivariate Density Estimation*, Wiley, second ed., 2015.
[24] I. M. Sobol', *The distribution of points in a cube and the approximate evaluation of integrals*, U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 86–112.
[25] G. R. Terrell and D. W. Scott, *Variable kernel density estimation*, The Annals of Statistics, 20 (1992), pp. 1236–1265.
[26] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, 1995.