MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Negative dependence, scrambled nets, and variance bounds

Christiane Lemieux
Department of Statistics and Actuarial Science, University of Waterloo, clemieux@uwaterloo.ca

In this paper, we provide a framework to study the dependence structure of sampling schemes such as those produced by randomized quasi-Monte Carlo methods. The main goal of this new framework is to determine conditions under which the negative dependence structure of a sampling scheme enables the construction of estimators with reduced variance compared to Monte Carlo estimators. To do so, we establish a generalization of the well-known Hoeffding's lemma—expressing the covariance of two random variables as an integral of the difference between their joint distribution function and the product of their marginal distribution functions—that is particularly well suited to study such sampling schemes. We also provide explicit formulas for the joint distribution of pairs of points randomly chosen from a scrambled $(0, m, s)$-net. In addition, we provide variance bounds establishing the superiority of dependent sampling schemes over Monte Carlo in a few different setups. In particular, we show that a scrambled $(0, m, 2)$-net yields an estimator with variance no larger than a Monte Carlo estimator for functions monotone in each variable.

Key words: negative orthant dependence, monotone functions, variance reduction, scrambled nets, Hoeffding's Lemma
MSC2000 subject classification: 11K38, 11K06
OR/MS subject classification: Primary: sampling, distributions, efficiency; secondary: simulation

---

1. Introduction. Monte Carlo (MC) methods are widely used in a variety of areas, in large part due to their simplicity and intuitive nature. They are especially attractive when one is faced with a problem where one or more aspects of a complicated probability distribution needs to be determined. However, the random sampling mechanism at the basis of these methods also has the drawback of producing estimators whose error decreases too slowly as the sampling size increases. For this reason, early on in the development of MC methods, alternative sampling mechanisms were proposed in order to alleviate this problem. One of them is *antithetic variates* (AV), originally proposed in [9]. Another form of improved sampling, coming from the field of experimental design, is *Latin Hypercube Sampling* (LHS), proposed in [15].

Both the AV and LHS methods use a form of *dependent sampling* to construct point sets $P_n = \{\mathbf{U}_i, i = 1, \ldots, n\} \subseteq [0,1]^s$ yielding estimators of the form

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{U}_i)$$

to estimate the integral

$$I(f) = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u}, \tag{1}$$

where f is a Borel-measurable square-integrable function defined over $[0,1]^s$ and taking values in the extended real numbers. That is, with AV and LHS the $\mathbf{U}_i$'s forming P_n are not independent, by contrast with the iid sampling scheme used by the MC method. For both methods, it can be proved that for certain types of functions, the corresponding estimator of $I(f)$ has a variance no larger than the naive MC estimator based on a sample of n iid points. The key ingredient to prove this result is that both AV and LHS are based on sampling schemes that satisfy the *negative quadrant dependence* (NQD) property introduced in [12]. The fact that this property is preserved under monotone transformations, combined with the very useful *Hoeffding's Lemma* [11, 12], is then used to show that these methods reduce the variance compared to the MC method for any function f that is monotone in each of its arguments.

Randomized quasi-Monte Carlo (RQMC) methods also use dependent sampling in an attempt to improve upon MC. In their case, the sampling schemes are obtained by applying an appropriate randomization to a *low-discrepancy point set* $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq [0,1]^s$, which is a deterministic construction designed so that its corresponding empirical distribution is close (in some sense to be discussed later) to the uniform distribution over $[0,1]^s$. The distance between these two distributions is precisely what is captured by the concept of *discrepancy*. The point set $\tilde{P}_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ obtained after applying the randomization is such that each $\mathbf{U}_i$ is randomly and uniformly distributed over $[0,1]^s$, but the $\mathbf{U}_i$'s have a dependence structure induced by the original low-discrepancy point set. The variance of estimators based on $\tilde{P}_n$ is typically analyzed using a series expansion for f that allows the variance under study to be decomposed into a sum of terms corresponding to each basis function. For instance, an Haar wavelet expansion is used to obtain several results on the variance of *scrambled nets* in [18, 19].

Our goal in this work is to explore a new approach to analyze RQMC methods, based on dependence concepts such as those used to study the variance of the AV and LHS estimators. Our purpose is not to provide an alternative approach to obtain known results, for example those establishing how smoothness assumptions on f can lead to improved convergence rates for the variance of RQMC estimators. Instead our goal is to provide conditions on $\tilde{P}_n$ and f so that we can guarantee that RQMC will do *no worse* than MC, building on the concepts of negative dependence to formulate the conditions on $\tilde{P}_n$ while at the same time keeping the conditions on f as weak as possible. By contrast, existing results that hold for a very wide range of functions can only guarantee that the RQMC variance is no larger than the MC variance up to a constant larger than one [18, 19]. Our goal is to get that constant down to 1 and see what minimal conditions on f we need in return.

The main results we provide to support this approach are as follows: first, thanks to a carefully designed generalization of Hoeffding's lemma, we are able to provide a general expression for the variance of an estimator based on a dependent sampling scheme. Secondly, this expression makes use of the joint distribution of pairs of points, for which we provide an explicit formula in the case of scrambled $(0, m, s)$-nets. Thirdly, and building on the previous results, we provide variance bounds establishing the superiority of dependent sampling schemes over MC in a few different setups, including the case where scrambled nets in two dimensions are used on functions that are monotone in each coordinate.

Unlike methods based on series expansion, our approach can be described as a "variance counterpart" to the error analysis approach based on the so-called Hlawka-Zaremba identity. The latter uses integration by parts to decompose the integration error of QMC-based approximations of the form $\hat{\mu}_n$ into an integral of a product of two terms, one of which strictly depends on P_n, and the other one strictly on f. In turn, this decomposition provides bounds of the form

$$|I(f) - \hat{\mu}_n| \leq D(P_n)V(f), \tag{2}$$

where $D(P_n)$ measures the *discrepancy* of P_n, and $V(f)$ measures the variability/smoothness of f [1, 31]. With our approach, which is detailed in Proposition 2 and Remark 7, we similarly use

integration by parts to decompose a covariance term capturing the difference between the RQMC and the MC variances. We then get an expression for this covariance term where the discrepancy of the RQMC point set $\tilde{P}_n$ is measured by the distance between the survival function it induces on pairs of distinct points and that of a pair of independent uniform points. The final expression is obtained by integrating this distance with respect to a measure induced by f. This approach thus allows us to get conditions based on the overall behavior of f to guarantee that the RQMC variance cannot exceed the MC variance. With the series expansion approach, conditions on f must instead be given at the much finer level of the basis functions used in the expansion, and are thus typically harder to verify.

The rest of the paper is organized as follows. In Section 2, we provide a detailed review of the different concepts used in this paper, as well as some notation. Section 3 contains results on multivariate integration by parts, with new elements including generalizations of known results and a discussion of the connection between them. Our generalization of Hoeffding's lemma is presented in Section 4, along with results showing its use to analyze the variance of estimators based on dependent sampling schemes. We then focus on the particular case of scrambled nets and provide an explicit formula for the joint distribution of two randomly chosen points from such sets in Section 5. We show that in the two-dimensional case, the resulting distribution has a form of negative dependence that is then used in Section 6 to show the corresponding nets' variance is no larger than the MC variance for functions monotone in each coordinate. We also examine two cases where s-dimensional dependent sampling schemes yield estimators with variance no larger than the MC estimator's variance. We conclude in Section 7 with a discussion of our findings and some ideas for future work.

2. Background and notation.

In this section, we introduce some notation and background information on the different topics covered in this paper. Since we rely on many different concepts to derive our main results, we have divided this rather long background section into four different sub-sections.

2.1. Concepts of negative dependence.

To assess the dependence structure of scrambled nets and possibly other sampling methods, we make use of the concepts of "negative upper/lower orthant dependence" (NUOD/NLOD) [16], which are multivariate extensions of the concept of negative quadrant dependence (NQD) [12]. We say X and Y are NQD if

$$P(X \le x, Y \le y) \le P(X \le x)P(Y \le y).$$

A vector $\mathbf{X} = (X_1, \dots, X_s)$ is NLOD if

$$P(X_1 \le x_1, \dots, X_s \le x_s) \le \prod_{\ell=1}^{s} P(X_\ell \le x_\ell),$$

and it is NUOD if

$$P(X_1 \ge x_1, \dots, X_s \ge x_s) \le \prod_{\ell=1}^{s} P(X_\ell \ge x_\ell).$$

Note that when $s = 2$, the NLOD and NUOD properties are equivalent (and both correspond to NQD), but it is not necessarily the case when $s \ge 3$.

Consider a sampling scheme $\tilde{P}_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ designed to construct an unbiased estimator of the form

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{U}_i)$$

for $I(f)$, where we assume each $\mathbf{U}_i$ is uniformly distributed over $[0,1]^s$ with a possible dependence structure between the $\mathbf{U}_i$'s. To assess this dependence, a key quantity of interest is

$$H(\mathbf{u},\mathbf{v};\tilde{P}_n) := \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j>i} P(\mathbf{U}_i \le \mathbf{u}, \mathbf{U}_j \le \mathbf{v}). \tag{3}$$

We can think of $H(\mathbf{u},\mathbf{v};\tilde{P}_n)$ as the joint distribution of a pair of (distinct) points $(\mathbf{U}_I,\mathbf{U}_J)$ randomly chosen in $\tilde{P}_n$. (Here, we use capital letters for the indices I and J to make it clear the points are randomly selected.) Note also that since $\mathbf{U}$ and $\mathbf{V}$ are each uniformly distributed over $[0,1]^s$, the joint distribution $H(\mathbf{u},\mathbf{v};P_n)$ is a *copula* defined over $[0,1]^{2s}$ [16].

There are two main types of dependence that we are interested in. First, if $H(\mathbf{u},\mathbf{v};\tilde{P}_n) \le \prod_{\ell=1}^s u_\ell v_\ell$ for all $0 \le u_\ell, v_\ell \le 1$, $\ell = 1,\dots,s$, then we say $\tilde{P}_n$ is an *NLOD sampling scheme*. Second, if the s pairs $(U_{I,\ell}, U_{J,\ell})$ for $\ell = 1,\dots,s$ are iid and

$$H(u,v;\tilde{P}_n(\{1\})) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j>i} P(U_{i,1} \le u, U_{j,1} \le v) \le uv, \text{ for all } 0 \le u,v \le 1, \tag{4}$$

where $\tilde{P}_n(\{1\}) = \{U_{i,1}, i = 1,\dots,n\}$, then we refer to $\tilde{P}_n$ as a *coordinate-wise independent NQD sampling scheme*. An example of this type of scheme is LHS sampling.

We are also interested in the quantity

$$T(\mathbf{u},\mathbf{v};\tilde{P}_n) := \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j>i} P(\mathbf{U}_i \ge \mathbf{u}, \mathbf{U}_j \ge \mathbf{v}), \tag{5}$$

and say that $\tilde{P}_n$ is an *NUOD sampling scheme* if $T(\mathbf{u},\mathbf{v};\tilde{P}_n) \ge \prod_{\ell=1}^s (1-u_\ell)(1-v_\ell)$ for all $0 \le u_\ell, v_\ell \le 1$, $\ell = 1,\dots,s$.

There is a close parallel between the above "average CDF" $H(\mathbf{u},\mathbf{v};\tilde{P}_n)$ and the variance expression for $\hat{\mu}_n$. Indeed, we have

$$\mathrm{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n} + \frac{n-1}{n}\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J)), \tag{6}$$

where $\sigma^2 = \mathrm{Var}(f(\mathbf{U}_i))$ and

$$\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J)) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j>i}\mathrm{Cov}(f(\mathbf{U}_i), f(\mathbf{U}_j)).$$

Hence we have that $\mathrm{Var}(\hat{\mu}_n) \le \mathrm{Var}(\hat{\mu}_{mc,n}) = \sigma^2/n$ if and only if $\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J)) \le 0$, where $\mathbf{U}_I, \mathbf{U}_J$ have a joint CDF given by (3) (or equivalently, a joint survival function given by (5)), and $\hat{\mu}_{mc,n}$ is the MC estimator based on n iid points.

A useful tool to analyze the covariance of two random variables (X,Y) is Hoeffding's lemma [11, 12], which we now recall.

LEMMA 1 **(Hoeffding)**. *If X and Y are two random variables with joint cdf $H(x,y)$ and respective marginal distributions $F(x)$ and $G(y)$, then*

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[H(x,y) - F(x)G(y)]dxdy,$$

assuming all expectations exist and are finite.

Rewriting the covariance in this way (which is based on integration by parts) allows us to see very clearly that if (X,Y) are NQD, then $\mathrm{Cov}(X,Y) \le 0$. In order to study the covariance $\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J))$ induced by an s-dimensional sampling scheme as described above, we need a generalization of Lemma 1, which will be the topic of Section 4. In the special case where we have a coordinate-wise independent NQD sampling scheme, Hoeffding's lemma can be used directly to study the corresponding covariance term, as done by Lehmann in a key result from [12] that we now recall.

THEOREM 1 **(Theorem 1(ii) from [12]).** *Let $(X_1,Y_1),\dots,(X_s,Y_s)$ be independent pairs of random variables. Let f and g be functions of s variables and let $X = f(X_1,\dots,X_s)$, $Y = g(Y_1,\dots,Y_s)$. Then (X,Y) is NQD if for each j, (X_j,Y_j) is NQD and f, g are concordant in the jth coordinate, for $j = 1,\dots,s$. That is, for any j, f and g are either both non-decreasing or both non-increasing in the jth coordinate.*

REMARK 1. In this work we focus on dependent sampling schemes obtained from a point set $\tilde{P}_n$, which is then used to define the estimator $\hat{\mu}_n$ and the joint distribution $H(\mathbf{u},\mathbf{v};\tilde{P}_n)$. An alternative approach would be to first choose a dependence structure via a copula $H(\mathbf{u},\mathbf{v})$ over $[0,1]^{2s}$ (where all marginal distributions over subsets $\mathcal{I}$ that satisfy either $\mathcal{I} \subseteq \{1,\dots,s\}$ or $\mathcal{I} \subseteq \{s+1,\dots,2s\}$ correspond to the independence copula) and then sample $n/2$ pairs $(\mathbf{U}_i,\mathbf{V}_i)$ that have this distribution. For this purpose, an obvious choice would be to use the "most negative" copula $H_{neg}(\mathbf{u},\mathbf{v}) := \prod_{j=1}^{s} \max(0, u_j + v_j - 1)$, which amounts precisely to the AV method. The distribution based on the corresponding point set $\tilde{P}_{av,n} = \{\mathbf{U}_i, (1-\mathbf{U}_i), i = 1,\dots,n/2\}$ in this case is given by

$$H_{av}(\mathbf{u},\mathbf{v};\tilde{P}_{av,n}) = \frac{n-2}{n-1}\prod_{j=1}^{s} u_j v_j + \frac{1}{n-1} H_{neg}(\mathbf{u},\mathbf{v}),$$

since $(n-1)/n$ of the $\binom{n}{2}$ pairs in $\tilde{P}_{av,n}$ are independent, while exactly $n/2$ pairs have the form $(\mathbf{U}_i, 1-\mathbf{U}_i)$. It is easy to see that $\tilde{P}_{av,n}$ is an NLOD sampling scheme since $H_{neg}(\mathbf{u},\mathbf{v}) \le \prod_{j=1}^{s} u_j v_j$. Note also that $X_j = U_j$ and $Y_j = 1 - U_j$ with $g = f$ satisfy the conditions of Theorem 1. Another interesting choice for $H(\mathbf{u},\mathbf{v})$ would be to use the extreme negative dependence construction in [29].

If f is linear in each of its variables, then the estimator $\hat{\mu}_{av,n}$ based on $\tilde{P}_{av,n}$ has zero variance and is thus optimal. However, it is possible to construct functions for which a dependent sampling scheme induced by a low-discrepancy point set $\tilde{P}_n$ is better than the AV estimator. This is illustrated in the following example.

EXAMPLE 1. Consider the function $f(u) = u^2$. Let $W \sim U(0,1)$ and $\tilde{P}_n = \{\left(\frac{i-1}{n} + W\right) \bmod 1, i = 1,\dots,n\}$, with corresponding estimator $\hat{\mu}_n$. Then it can be shown (after some easy but tedious calculations) that:

1. The joint distribution $H(u,v;\tilde{P}_n)$ associated with $\tilde{P}_n$ is given by

$$H(u,v;\tilde{P}_n) = F_{i,j}(u,v) \text{ for } u \in \left[\frac{i-1}{n}, \frac{i}{n}\right), v \in \left[\frac{j-1}{n}, \frac{j}{n}\right), i,j = 1,\dots,n,$$

where

$$F_{i,j}(u,v) = \begin{cases} \frac{1}{n-1}\left(\sum_{k=1}^{i-1} \min(u, v-\theta_k) + \sum_{k=1}^{j-1} \min(u-\theta_k, v)\right) & \text{if } i+j \le n \\ \frac{1}{n-1}\left(\sum_{k=1}^{n-i} \min(u, v-\theta_k) + \sum_{k=1}^{n-j} \min(u-\theta_k, v)\right. & \\ \qquad \left. +(i+j-n-1)(u+v-1)\right) & \text{if } i+j > n, \end{cases}$$

and $\theta_k = k/n$.

2. $\tilde{P}_n$ is not an NQD sampling scheme but it can be shown that

$$\int_0^1 \int_0^1 (H(u,v;\tilde{P}_n) - uv)dudv = -1/12n \leq 0$$

and thus we can say that it is NQD "on average".

3. We have that

$$\mathrm{Var}(\hat{\mu}_n) = \frac{1}{n}\mathrm{Var}\left(\hat{\mu}_{mc,n}\right) - \left(1 - \frac{1}{n}\right)\frac{n+1}{180n^3},$$

while

$$\mathrm{Var}(\hat{\mu}_{av,n}) = \frac{1}{8}\mathrm{Var}\left(\hat{\mu}_{mc,n}\right).$$

Hence we see that for the function $f(u) = u^2$, the AV estimator reduces the MC estimator's variance by a factor of 8, while the "shifted grid" estimator $\hat{\mu}_n$ reduces it by a factor larger than n. This is of course a very simple example on a one-dimensional function for which the integral has an analytic formula. But it highlights the fact that there are (easy-to-find) problems where an n-point dependent sampling scheme that is NQD only on average can do much better than one based on iid pairs having the most negative dependence.

2.2. Scrambled nets. A *digital net in base b* (for b prime) [5, 17] is a point set $P_n = \{\mathbf{V}_1, \ldots, \mathbf{V}_n\} \subseteq [0,1]^s$ with $n = b^m$ that is constructed via s generating matrices $C_1, \ldots, C_s$ of size $m \times m$ with entries in $\mathbb{F}_b$, in the following way: for $0 \leq i < b^m$ we write $i = \sum_{r=0}^{m-1} i_r b^r$, then $\mathbf{V}_i = (V_{i,1}, \ldots, V_{i,s})$ is obtained as $V_{i,\ell} = \sum_{r=1}^{m} V_{i,\ell,r} b^{-(r+1)}$, and $V_{i,\ell,r} = \sum_{p=1}^{m} C_{\ell,r,p} i_{p-1}$, where $C_{\ell,r,p}$ is the element on the rth row and pth column of C_ℓ.

The goal is to construct a net that is very uniformly distributed, and one way to assess this is via the concept of $(q_1, \ldots, q_s)$-equidistribution, where we say that P_n with $n = b^m$ is $(q_1, \ldots, q_s)$-*equidistributed in base b* if every *elementary interval* of the form

$$\prod_{\ell=1}^{s} \left[\frac{a_\ell}{b^{q_\ell}}, \frac{a_\ell + 1}{b^{q_\ell}}\right)$$

for $0 \leq a_\ell < b^{q_\ell}$ contains exactly $b^{m-q_1-\cdots-q_s}$ points from P_n, assuming $m \geq q_1 + \ldots + q_s$. We say that a digital net in base b has a *quality parameter* t if P_n is $(q_1, \ldots, q_s)$-equidistributed for all s-dimensional vectors of non-negative integers $(q_1, \ldots, q_s)$ such that $q_1 + \ldots + q_s \leq m - t$. We then refer to P_n as a digital (t, m, s)-net in base b. So the lower is t, the more uniform P_n is [17]. The construction proposed by Faure in [7] provides $(0, m, s)$-nets in base $b \geq s$. The widely used Sobol' sequences [25] provide (t, m, s)-nets in base 2 with $t = 0$ when $s = 2$ and $t > 0$ otherwise. Information on newer constructions can be found in [4, 5].

A *scrambled digital net in base b* is a randomized point set $\tilde{P}_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ with $n = b^m$ which we assume has the following two properties [14, 20]. Let $U_{i,\ell} = \sum_{r=1}^{\infty} U_{i,\ell,r} b^{-(r+1)}$, that is, $U_{i,\ell,r}$ represents the rth digit in the base b expansion of the ℓth coordinate of the ith point $\mathbf{U}_i$. Then we must have:

1. Each $\mathbf{U}_i \sim U([0,1]^s)$;
2. For two distinct points $\mathbf{U}_i, \mathbf{U}_j$ and in each dimension ℓ, if the two deterministic points $\mathbf{V}_i, \mathbf{V}_j$ (before scrambling is applied) have the same first r digits and differ on the $(r+1)$th digit, then (i) the scrambled points $(U_{i,\ell}, U_{j,\ell})$ also have the same first r digits, and the pair $(U_{i,\ell,r+1}, U_{j,\ell,r+1})$ is uniformly distributed over $\{(k_1, k_2), 0 \leq k_1 \neq k_2 < b\}$; (ii) the pairs $(U_{i,\ell,v}, U_{j,\ell,v})$ for $v > r+1$ are independent and uniformly distributed over $\{(k_1, k_2), 0 \leq k_1, k_2 < b\}$.

One way to scramble a digital net P_n so that the scrambled net $\tilde{P}_n$ has these properties is to multiply from the left each generating matrix C_ℓ by a randomly chosen NLT matrix S_ℓ (i.e., with entries on the diagonal uniformly chosen in $\{1,\dots,b-1\}$, and entries below the diagonal uniformly chosen in $\{0,\dots,b-1\}$, with the other entries set to 0). Also note that if P_n is a (t,m,s)-net then the scrambled net $\tilde{P}_n$ is a (t,m,s)-net as well. We refer the reader to [20] for further information on scrambling methods for digital nets.

2.3. Functional ANOVA decomposition. To study the performance of (R)QMC methods, the concept of ANOVA decomposition and effective dimension have proven to be very useful [3, 10, 18, 26, 27]. We make use of this decomposition for one of our results in Section 6. The functional ANOVA decomposition of a square-integrable function f defined over $[0,1]^s$ amounts to write f as

$$f(\mathbf{u}) = \sum_{\mathcal{I}\subseteq\{1,\dots,s\}} f_{\mathcal{I}}(\mathbf{u}),$$

where for nonempty $\mathcal{I}$ we have

$$f_{\mathcal{I}}(\mathbf{u}) = \int_{[0,1)^{s-j}} f(\mathbf{u})d\mathbf{u}_{-\mathcal{I}} - \sum_{\mathcal{J}\subset\mathcal{I}} f_{\mathcal{J}}(\mathbf{u}),$$

where $j=|\mathcal{I}|$, $\mathbf{u}_{-\mathcal{I}} = (u_\ell)_{\ell\notin\mathcal{I}}$ is a $(s-j)$-dimensional vector, and the inclusion $\mathcal{J}\subset\mathcal{I}$ is strict. In particular, this means $f_{\mathcal{I}}(\mathbf{u})$ is a function of the variables u_ℓ with $\ell\in\mathcal{I}$ only. Hence we see that

$$\int_{[0,1)^j} f_{\mathcal{I}}(\mathbf{u})d\mathbf{u}_{\mathcal{I}} = 0$$

for nonempty $\mathcal{I}$, while we set $f_\emptyset(\mathbf{u}) = I(f)$. These ANOVA components $f_{\mathcal{I}}$ are orthogonal, i.e.,

$$\int_{[0,1)^d} f_{\mathcal{I}}(\mathbf{u})f_{\mathcal{J}}(\mathbf{u})d\mathbf{u} = 0 \qquad \text{for } \mathcal{I}\neq\mathcal{J}.$$

Therefore the variance of f decomposes as $\sigma^2 = \sum_{\mathcal{I}\subseteq\{1,\dots,s\}}\sigma^2_{\mathcal{I}}$, where $\sigma^2_{\mathcal{I}} = \mathrm{Var}(f_{\mathcal{I}}(\mathbf{U}))$.

The function f is said to have *effective dimension d_T in the truncation sense* [3, 10] at the level p if d_T is the smallest integer k such that

$$\sum_{\mathcal{I}\subseteq\{1,\dots,k\}} \sigma^2_I \geq p\sigma^2,$$

with p taking a value close to 1 such as 0.95 or 0.99. This means that $100p\%$ of the variance of f is explained by its first k variables.

Alternatively, we may be interested in how well we can approximate f by using ANOVA components depending on at most k variables. This leads to the notion of *effective dimension in the superposition sense* (at the level p), which is the smallest integer k such that

$$\sum_{\mathcal{I}:|\mathcal{I}|\leq k} \sigma^2_{\mathcal{I}} \geq p\sigma^2.$$

2.4. Notation. We conclude this section by introducing notation that will be used when we analyze the dependence structure of scrambled nets in Section 5. We start by a definition:

Definition 1. Let $0\leq u,v<1$. Then $\gamma_b(u,v)$ is the smallest non-negative integer α such that

$$\lfloor b^\alpha u\rfloor = \lfloor b^\alpha v\rfloor \qquad \text{but} \qquad \lfloor b^{\alpha+1}u\rfloor \neq \lfloor b^{\alpha+1}v\rfloor.$$

In other words, $\gamma_b(u,v)$ represents the exact number of initial digits shared by u and v in their base b expansion.

Next, for a given integer r such that $1 \le r \le m$, we partition the unit square as

$$[0,1]^2 = D_0 \cup D_1 \cup \ldots \cup D_{r-1} \cup C_r,$$

where

$$D_\ell = \{(x,y) \in [0,1]^2 : \gamma_b(x,y) = \ell\}, \qquad \ell \ge 0$$
$$C_r = \{(x,y) \in [0,1]^2 : \gamma_b(x,y) \ge r\}.$$

While the obtained partition depends on b, we do not indicate this in the notation for $D_0, \ldots, D_{r-1}, C_r$ so as to not overburden it. On Figure 1, for $b = 2$, we show the partition of the unit square into D_0, D_1 and C_2 and into D_0 and C_1.

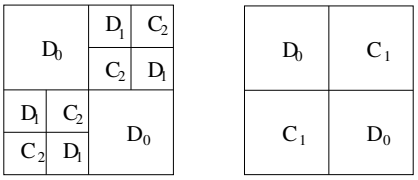

FIGURE 1. Partition of $[0,1]^2$ into $D_0 \cup D_1 \cup C_2$ (left) and into $D_0 \cup C_1$ (right)

Finally we point out that throughout the paper (and as done so far) indices I and J are used to denote randomly chosen indices for points taken from $\tilde{P}_n$, while $\mathcal{I}$ and $\mathcal{J}$ denote subsets of indices from $\{1, \ldots, s\}$.

3. Multivariate integration by parts. A key ingredient to obtain our generalization of Hoeffding's lemma is the use of an appropriate formula for multivariate integration by parts, involving integrals with respect to a general measure. For functions of *bounded variation* (in the sense of Hardy and Krause), the multivariate integration by parts formula we need has been obtained about 100 years ago by Young in [30], and is presented later in this section. Along with this result, we also discuss one of its variants—also from [30]—that has been used in [2] to obtain a generalization of Hoeffding's lemma different from ours. Finally, we present yet another multivariate integration by parts formula that is very closely related to the bivariate version presented in [13], and has the important advantage of not requiring f to be of bounded variation.

While formally, only the last of those formulas is used in Section 4, presenting the others allows us to highlight the similarities and differences between our approach and others who have worked with (a generalized) Hoeffding's lemma for the purpose of covariance study [2, 12], and also with a recent paper by Dick and Aistleitner [1] that uses measures induced by functions of bounded variation, as we also do here. The main references used to cover this material are [1, 13, 21, 30].

We first need to give the definition of variations that are needed. To do so, we use the following notation: for $\mathcal{I} \subseteq \{1, \ldots, s\}$ we write $-\mathcal{I} := \{1, \ldots, s\} \setminus \mathcal{I}$. We also write $f(\mathbf{a}^{\mathcal{I}}; \mathbf{b}^{-\mathcal{I}})$ to represent the function evaluation $f(x_1, \ldots, x_s)$ where $x_j = a_j$ if $j \in \mathcal{I}$ and $x_j = b_j$ otherwise.

DEFINITION 2. Consider a function f defined over $[0,1]^s$. Its *quasi-volume* or *increment* over an interval of the form $A = [\mathbf{a}, \mathbf{b}] = \prod_{j=1}^s [a_j, b_j] \subseteq [0,1]^s$ is given by

$$\Delta^{(s)}(f; A) = \sum_{\mathcal{I} \subseteq \{1,\ldots,s\}} (-1)^{|\mathcal{I}|} f(\mathbf{a}^{\mathcal{I}}; \mathbf{b}^{-\mathcal{I}}).$$

DEFINITION 3. If $\Delta^{(d)}(f;A) \geq 0$ for all closed axis-parallel boxes $A = [\mathbf{a}, \mathbf{b}) \subseteq [0,1]^s$ of arbitrary dimension $1 \leq d \leq s$, then f is said to be *quasi-monotone* or *completely monotone*. (The dimension d of A refers to the value $d = \sum_{j=1}^{s} \mathbf{1}_{a_j < b_j}$.)

Next we consider partitions $\mathcal{P}$ of $[0,1]^s$ defined by the product of one-dimensional partitions $0 = x_0^j < x_1^j < \ldots < x_{m_j}^j = 1$, for $j = 1, \ldots, s$, as

$$\mathcal{P} = \left\{ \left[x_{\ell_1}^1, x_{\ell_1+1}^1\right) \times \cdots \times \left[x_{\ell_s}^s, x_{\ell_s+1}^s\right), \ell_j = 0, \ldots, m_j - 1, j = 1, \ldots, s \right\}.$$

DEFINITION 4. The *variation in the sense of Vitali of a function f over* $[0,1]^s$ is given by

$$V^{(s)}(f;[0,1]^s) = \sup_{\mathcal{P}} \sum_{A \in \mathcal{P}} |\Delta^{(s)}(f;A)|.$$

The variation in the sense of Hardy and Krause differs from the above in that it also considers the quasi-volumes of lower dimension induced by f. More precisely, for a nonempty subset $\mathcal{I} \subseteq \{1, \ldots, s\}$, let $V_{\mathbf{1}}^{(|\mathcal{I}|)}(f;\mathcal{I};[0,1]^s)$ be the variation in the sense of Vitali defined over the subspace

$$\{(u_1, \ldots, u_s) \in [0,1]^s : u_j = 1 \text{ if } j \notin \mathcal{I}\}.$$

DEFINITION 5. The *variation in the sense of Hardy and Krause of a function f over* $[0,1]^s$ is given by

$$V_{hk}(f;[0,1]^s) = \sum_{j=1}^{s} \sum_{1 \leq i_1 < \ldots < i_j \leq s} V_{\mathbf{1}}^{(j)}(f;\{i_1, \ldots, i_j\};[0,1]^s).$$

It is well-known that right-continuous functions of bounded variation in the sense of Hardy and Krause can be used to define (finite) signed Borel measures. This can in fact be done in more than one way. A first approach to define a measure μ_f on $[0,1]^s$ based on f is to use the notion of quasi-volume introduced above, i.e., for $A \subseteq [0,1]^s$ a closed axis-parallel box of the form $A = [\mathbf{0}, \mathbf{a}) = \prod_{j=1}^{s} [0, a_j)$, we set

$$\mu_f(A) = \Delta^{(s)}(f;A). \tag{7}$$

This is the approach used in [13, Eq. (3.4)]. Note that when f has a continuous mixed partial derivative $\partial^{1:s} f(\mathbf{u}) = \partial^s f(u_1, \ldots, u_s) / \prod_{j=1}^{s} \partial u_j$, then we have

$$\int g(\mathbf{u}) d\mu_f(\mathbf{u}) = \int g(\mathbf{u}) \partial^{1:s} f(\mathbf{u}) d\mathbf{u}.$$

See [8, Sect. 6] and [21, Sect. 9] and the references therein for more details on this identity.

We present another definition of measure induced by f, denoted ν_f, which is used in [13, (3.8)] and [1, Thm. 3]. It is defined so that for any $A = [\mathbf{0}, \mathbf{a}) \subseteq [0,1]^s$, we have

$$\nu_f(A) = f(\mathbf{a}). \tag{8}$$

At first sight, this measure appears to be more simply defined than μ_f. However the evaluation of $\nu_f(A)$ gets more complicated for intervals $A = [\mathbf{a}, \mathbf{b})$ not anchored at the origin, while for μ_f we still have $\mu_f(A) = \Delta^{(s)}(f;A)$ for such A, as shown in the next lemma, whose proof is in the appendix. Note that the relation (10) given below to evaluate $\nu_f(A)$ for these more general intervals generalizes the $s = 2$ case described in [13, Eqs.(3.4)–(3.8)].

LEMMA 2. *(i) For $A = [\mathbf{a}, \mathbf{b}) \subseteq [0,1]^s$, we have that*

$$\mu_f(A) = \Delta^{(s)}(f;A) \tag{9}$$

and

$$\nu_f(A) = \sum_{\mathcal{I}\subseteq\{1,\dots,s\}} \mu_{f,\mathcal{I},\mathbf{0}}(A), \tag{10}$$

where

$$\mu_{f,\mathcal{I},\mathbf{0}}(A) = \Delta_0^{(|\mathcal{I}|)}(f;\mathcal{I};A), \tag{11}$$

and for $\mathcal{I}\subseteq\{1,\dots,s\}$*, we have*

$$\Delta_0^{(|\mathcal{I}|)}(f;\mathcal{I};A) = \sum_{\mathcal{J}\subseteq\mathcal{I}} (-1)^{|\mathcal{J}|} f(\mathbf{a}^{\mathcal{J}\cap\mathcal{I}};\mathbf{b}^{(-\mathcal{J})\cap\mathcal{I}};\mathbf{0}^{-\mathcal{I}});$$

(ii) If f *is quasi-monotone, then* μ_f *and* ν_f *are positive Borel measures.*

Using the two measures μ_f and ν_f induced by a right-continuous function f of bounded variation in the sense of Hardy and Krause, we can now provide a first integration by parts formula that is based on [13, Thm 1] and [30, Eqs. (5) and (18)].

THEOREM 2 **(Based on [13, 30])**. *Let* f *and* F *be right-continuous functions of bounded variation in the sense of Hardy and Krause over* $[0,1]^s$*. Then we have*

$$\int_{[0,1]^s} f(\mathbf{u})d\mu_F(\mathbf{u}) = \int_{[0,1]^s} \mu_F([\mathbf{u},\mathbf{1}])d\nu_f(\mathbf{u}). \tag{12}$$

REMARK 2. As described very insightfully in [30], there are several ways to obtain a formula for multivariate integration by parts. In particular, in [2] a generalization of Hoeffding's lemma is presented, which is instead based on the variant

$$\int_{[0,1]^s} f(\mathbf{u})d\mu_F(\mathbf{u}) = \int_{[0,1]^s} \sum_{I\subseteq\{1,\dots,s\}} (-1)^{|\mathcal{I}|}\mu_F([\mathbf{0},\mathbf{u}])d\mu_{f,\mathcal{I},\mathbf{1}}(\mathbf{u}), \tag{13}$$

where $\mu_{f,\mathcal{I},\mathbf{1}}$ is defined similarly to $\mu_{f,\mathcal{I},\mathbf{0}}$, but using the function $f_{\mathcal{I},\mathbf{1}}$ where variables u_j with $j\notin\mathcal{I}$ are fixed to 1 rather than 0. This corresponds to the relations (6) and (18) in [30] for $s=2$ and $s>2$, respectively, although we should mention that in [2] the relation (13) is obtained using a slightly more general setup than in [30].

REMARK 3. In [1] an integral identity that resembles (12) is established, using integration with respect to a "reflected" measure $\hat{\nu}_f$ corresponding to $\nu_h(1-A)$ in our notation, where h is defined via $h(\mathbf{u}) = f(1-\mathbf{u})$ and $1-A = \{\mathbf{u}\in[0,1]^s : (1-u_1,\dots,1-u_s)\in A\}$. More precisely, they use a formula of the form [1, Sect. 4]

$$\int_{[0,1]^s} f(\mathbf{u})d\mu_F(\mathbf{u}) = \int_{[0,1]^s} \mu_F([\mathbf{0},\mathbf{u}])d\hat{\nu}_f(\mathbf{u}). \tag{14}$$

Compared to (14) and (12), (13) is less compact and includes alternating signs, which complicates the derivation of conditions under which the corresponding integral is positive or negative. The reason why the variant (12) works well for us is that in our setup, μ_F is a probability measure generated by a point set and having it anchored at 1 simply means we work with probabilities of the form $P(U_1>u_1,\dots,U_s>u_s)$ instead of joint CDFs. In addition and as seen in Lemma 2, the sign of integrals with respect to the measure ν_f can be analyzed more easily since sufficient conditions under which ν_f is a positive Borel measure are known.

We now present the multivariate integration by parts formula that will be used in Section 4. As mentioned earlier, it is closely related to [13, Lemma 1], but slightly extends the latter by generalizing to $s\geq 2$ and allowing for a signed measure within the representation that is used.

THEOREM 3 **(Based on [13, Lemma 1])**. *Let f be a function defined over $[0,1]^s$. Let μ_F be a finite signed measure over $[0,1]^s$. Assume that f is integrable with respect to μ_F over $[0,1)^s$ and that it admits the representation*

$$f(\mathbf{u}) = \int_{A(\mathbf{u})} g(\mathbf{v}) d\eta_f(\mathbf{v}), \tag{15}$$

for every $\mathbf{u} \in [0,1)^s$, where $A(\mathbf{u}) = \{\mathbf{v} \in [0,1)^s : 0 \le v_j \le u_j, j = 1,\dots,s\}$, η_f is a finite signed Borel measure over $[0,1]^s$, and g is a Borel-measurable function that is integrable with respect to $\mu_F \times \eta$ over $A = \{(\mathbf{u},\mathbf{v}) : \mathbf{u} \in [0,1)^s, v_j \le u_j, j = 1,\dots,s\}$. Then we have

$$\int_{[0,1)^s} f(\mathbf{u}) d\mu_F(\mathbf{u}) = \int_{[0,1)^s} g(\mathbf{u}) \mu_F([\mathbf{u},\mathbf{1}]) d\eta_f(\mathbf{u}).$$

Proof. We get

$$\int_{[0,1)^s} f(\mathbf{u}) d\mu_F(\mathbf{u}) = \int_{[0,1)^s} \int_{[0,1)^s} \mathbf{1}_{\mathbf{v}\le\mathbf{u}} g(\mathbf{v}) d\eta_f(\mathbf{v}) d\mu_F(\mathbf{u}) = \int_{[0,1)^s} g(\mathbf{v}) \mu_F([\mathbf{v},\mathbf{1}]) d\eta_f(\mathbf{v}), \tag{16}$$

where the second equality is obtained using Fubini's theorem (which remains true when using finite signed measures, see [6, p. 193]), which applies because of our assumptions on g. □

REMARK 4. If f is of bounded variation (in the sense of Hardy and Krause) over $[0,1]^s$, then we have the representation

$$f(\mathbf{u}) = \int_{A(\mathbf{u})} d\nu_f(\mathbf{v}),$$

where ν_f is the finite (signed) Borel measure defined in (8). That is, in this case $g = 1$ and $\eta_f = \nu_f$ in (15) and we recover the integration by parts given in Theorem 2 above. We note that if $f(\mathbf{0}) \neq 0$ then the measure ν_f has a discontinuity (wrt to the Lebesgue measure) at $\mathbf{0}$ and thus the integral on the RHS of the above equation when $\mathbf{u} = \mathbf{0}$ is $\nu_f(\mathbf{0}) = f(\mathbf{0})$, as required.

REMARK 5. Going further, it is clear that requiring f to have the integral representation (15) is in fact less restrictive than asking for f to be of bounded variation. In the QMC literature, typical reasons causing a function to not have bounded variation over $[0,1]^s$ are either that f is unbounded (usually on or approaching its boundary), or f or its partial derivatives have discontinuities on hyperplanes that are not parallel to the hypercube's axes. In the latter case, the measure ν_f induced by f may still be finite (and signed), in which case the representation (15) with $g = 1$ and $\eta_f = \nu_f$ holds as in the previous remark. Example 2 below illustrates this possibility. As for unbounded functions, the next remark addresses this point in the case $s = 1$. The more general question of what are necessary and sufficient conditions for f to satisfy the representation (15) is something we plan to study in future work.

REMARK 6. Using Theorem 3, in Proposition 3 of the next section we extend Lehmann's covariance result, which was recalled in Section 2 (see Theorem 1). Our result and his require the same conditions on the point set $\tilde{P}_n$ when $s = 1$. Ideally the conditions on f should also be the same so that our result truly coincides with his in that case. The next proposition shows that ours only requires an additional mild assumption, which we say is to have f *one-side bounded*. This means $f(u) < \infty$ for $0 \le u < 1$ (resp. $0 < u \le 1$) if f is non-decreasing (resp. non-increasing). Note that this assumption is weaker than asking for bounded variation, since it does not require f to be bounded over $[0,1]$.

PROPOSITION 1. *Let f be a one-side bounded monotone function over $[0,1]$. Then f admits the representation* (15).

Proof. First, we can assume wlog that f is right-continuous. Indeed, if some of the jumps are left-continuous, we can replace them by right-continuous jumps as long as the atoms of μ_F (if they exist, and in our setup they won't since μ_F corresponds to a continuous CDF) and the location of the jumps of f do not coincide. That is, this substitution will not affect the value of the integral $\int f d\mu_F$. We can also assume wlog that f is monotone non-decreasing. If f is unbounded, then the only point where $f(u) = \infty$ must be at $u = 1$. Also, by adapting an argument used in the proof of Lemma 1.6.31 in [28], we can show that because of its monotonicity, f has at most countably many discontinuities in $[0,1]$. So we let $D_c = \{x_1, x_2, \ldots\} \subseteq [0,1]$ where x_i is the ith smallest value where f has a discontinuity.

Consider the function

$$\tilde{f}(v) = v + \sum_{i=1}^{\infty} 2^{-i} \mathbf{1}_{x_i \leq v}, 0 \leq v \leq 1.$$

We have that $\tilde{f}(v) \leq 2$ for all $0 \leq v \leq 1$. Hence the Lebesgue-Stieltjes measure generated by $\tilde{f}$ is finite, and will henceforth be denoted by η_f. It is easy to see that the measure ν_f induced by f (which is not finite and thus cannot play the role of η_f in (15)) is absolutely continuous with respect to η_f. Hence we can use the Radon-Nikodym Theorem [24] to write

$$f(u) = \nu_f([0,u]) = \int_{A(u)} g(v) d\eta_f(v)$$

for some function g (the Radon-Nikodym derivative), as required.

If f is bounded then we can take $g = 1$ and $\eta_f = \nu_f$ to obtain the representation (15). □

Example 2. Consider the function $f(u_1, u_2, u_3) = \max(u_1 + u_2 + u_3 - 1, 0)$. It is proved in [21] that f has unbounded variation in the sense of Hardy and Krause. However, it is easy to see that the measure ν_f induced by f is a signed finite measure. Therefore the representation (15) with $g = 1$ and $\eta_f = \nu_f$ holds. In fact, we can go one step further and construct an alternative representation based on a probability measure η_f such that ν_f is absolutely continuous wrt η_f and then apply the Radon-Nikodym Theorem. The idea is to define η_f as the measure corresponding to a certain sampling scheme producing U_1, U_2, U_3 such that $U_1 + U_2 + U_3 = 1$. Details are omitted as this falls outside the scope of this paper.

4. Variance bounds for NUOD sampling schemes. In this section we present a general result giving sufficient conditions for an estimator based on an NUOD sampling scheme to have a variance no larger than the MC estimator's variance. The result relies on a generalization of Hoeffding's lemma which, in the form given below and to the best of our knowledge, has not been shown elsewhere. As mentioned in Section 3 a related result is given in [2], but the form we provide in the following lemma makes it much easier to find sufficient conditions for the aforementioned variance bound to hold and also requires weaker conditions on f. Other papers generalizing Hoeffding's lemma are [23] and [22]. In [23], the authors consider two-dimensional extensions for what they call quasi-monotone functions (but they use a weaker definition than the one presented in Section 3 since they seem to consider only full s-dimensional volumes) and in [22] the authors focus on a very specific class of functions and thus get a result that is not applicable to the more general setup considered here.

Lemma 3 **(Generalization of Hoeffding's lemma).** *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be* s*-dimensional random vectors over* $[0,1]^s$ *with joint survival function* $T(\mathbf{u}, \mathbf{v})$ *and marginal survival functions* $R(\mathbf{u})$ *and* $S(\mathbf{v})$*, respectively. Let* f *be a function defined over* $[0,1]^s$ *satisfying the representation* (15) *with function* g *and measure* η_f*. Then*

$$\mathrm{Cov}(f(\mathbf{U}), f(\mathbf{V})) = \int_{[0,1]^{2s}} (T(\mathbf{u}, \mathbf{v}) - R(\mathbf{u})S(\mathbf{v})) g(\mathbf{u}) g(\mathbf{v}) d\eta_f(\mathbf{u}) d\eta_f(\mathbf{v}), \tag{17}$$

assuming f *satisfies all integrability conditions required for this covariance term to be well defined.*

Proof. Let $H(\mathbf{u},\mathbf{v})$ be the joint cdf of $\mathbf{U}$ and $\mathbf{V}$, and let $F(\mathbf{u})$ and $G(\mathbf{v})$ represent their marginal CDF. We write

$$\mathrm{Cov}(f(\mathbf{U}),f(\mathbf{V})) = \int_{[0,1]^{2s}} f(\mathbf{u})f(\mathbf{v})d\mu_H(\mathbf{u},\mathbf{v}) - \int_{[0,1]^s}\int_{[0,1]^s} f(\mathbf{u})f(\mathbf{v})d\mu_F(\mathbf{u})d\mu_G(\mathbf{v}).$$

Using Theorem 3 on both terms, we get

$$\mathrm{Cov}(f(\mathbf{U}),f(\mathbf{V})) = \int_{[0,1]^{2s}} T(\mathbf{u},\mathbf{v})\tilde{g}(\mathbf{u},\mathbf{v})d\tilde{\eta}_{\tilde{f}}(\mathbf{u},\mathbf{v}) - \int_{[0,1]^s}\int_{[0,1]^s} R(\mathbf{u})S(\mathbf{v})g(\mathbf{u})g(\mathbf{v})d\eta_f(\mathbf{u})d\eta_f(\mathbf{v}),$$

where $\tilde{f}(\mathbf{u},\mathbf{v}) = f(\mathbf{u})f(\mathbf{v})$, and we used the fact that if f satisfies the representation (15) then $\tilde{f}$ satisfies it as well with $\tilde{g}(\mathbf{u},\mathbf{v}) = g(\mathbf{u})g(\mathbf{v})$ and $\tilde{\eta}_f = \eta_f \times \eta_f$. Hence we get

$$\mathrm{Cov}(f(\mathbf{U}),f(\mathbf{V})) = \int_{[0,1]^{2s}} (T(\mathbf{u},\mathbf{v}) - R(\mathbf{u})S(\mathbf{v}))g(\mathbf{u})g(\mathbf{v})d\eta_f(\mathbf{u})d\eta_f(\mathbf{v}). \quad \square$$

Next we specialize this result to the study of the covariance term associated with a given sampling scheme.

Proposition 2. *Let f be a function defined over $[0,1]^s$ satisfying the representation (15) with function g and measure η_f. Let $\tilde{P}_n$ be a sampling scheme over $[0,1]^s$ and let $\mathbf{U}_I, \mathbf{U}_J$ be two distinct points randomly chosen from it. Then*

$$\mathrm{Cov}(f(\mathbf{U}_I),f(\mathbf{U}_J)) = \int_{[0,1]^s}\int_{[0,1]^s} [T(\mathbf{u},\mathbf{v};\tilde{P}_n) - \prod_{i=1}^s(1-u_i)\prod_{j=1}^s(1-v_j)]g(\mathbf{u})g(\mathbf{v})d\eta_f(\mathbf{u})d\eta_f(\mathbf{v}). \quad (18)$$

Proof. This is a simple application of Lemma 3, where the integrability conditions on f are verified because of our ongoing assumption that f is square-integrable.

Remark 7. The error decomposition approach for QMC methods based on the Hlawka-Zaremba identity makes use of the discrepancy function $E(\tilde{P}_n;\mathbf{u}) := A(\tilde{P}_n;\mathbf{u})/n - \prod_{j=1}^s u_j$ to measure the distance between the empirical distribution induced by $\tilde{P}_n$ and the uniform distribution, as $A(\tilde{P}_n;\mathbf{u})$ counts the number of points from $\tilde{P}_n$ that are in $[\mathbf{0},\mathbf{u}]$. It writes the error as $\hat{\mu}_n - I(f) = \int E(\tilde{P}_n;\mathbf{u})d\hat{\nu}_f(\mathbf{u})$ (see, e.g., [1, Proof of Thm 1]). As mentioned in the introduction, our approach can be seen as the variance counterpart of the above. With that in mind, we can say that in (18) the difference $T(\mathbf{u},\mathbf{v};\tilde{P}_n) - \prod_{i=1}^s(1-u_i)\prod_{j=1}^s(1-v_j)$ of survival functions plays the role of $E(\tilde{P}_n;\mathbf{u})$. In both cases, integration by parts is used to rewrite the error/variance so that the influence of the point set $\tilde{P}_n$ and the influence of the integrand can be disentangled.

We can now provide sufficient conditions under which the estimator based on $\tilde{P}_n$ has a variance no larger than the MC variance.

Proposition 3. *Let f be a function defined over $[0,1]^s$ satisfying the representation (15) with each of g and η_f not changing sign over their domain. If $\tilde{P}_n$ is an NUOD sampling scheme with corresponding estimator $\hat{\mu}_n$ for $I(f)$, then $\mathrm{Var}(\hat{\mu}_n) = \mathrm{Var}(\hat{\mu}_{mc,n}) + (n-1)\sigma_{I,J}/n \le \mathrm{Var}(\hat{\mu}_{mc,n})$, where the covariance term $\sigma_{I,J}$ is given by (18).*

Proof. The expression for $\mathrm{Var}(\hat{\mu}_n)$ comes from (6). So we need to show that the covariance $\sigma_{I,J}$ between two points randomly chosen from $\tilde{P}_n$ is non-positive, using expression (18) given in Proposition 2. Now, the term in square brackets in (18) is non-positive because $\tilde{P}_n$ is an NUOD sampling scheme, and since g and η_f do not change sign over their respective domain, overall the integral (18) is no larger than 0, as required. $\square$

REMARK 8. If f is bounded and f or $-f$ is quasi-monotone, then we can take $g=1$ and $\eta_f=\nu_f$ in (15), and the conditions of Proposition 3 are satisfied in this case. We note that these assumptions on f also imply it is of bounded variation in the sense of Hardy and Krause. We speculate that if f is unbounded on its boundary (e.g., when $u_j=1$ for at least one j) but f or $-f$ is quasi-monotone, it may still be possible to satisfy the conditions of Proposition 3 by using a finite (positive) measure η_f in (15) wrt which the unbounded measure ν_f is absolutely continuous, but this, along with our discussion in Remark 5, is something we leave for future work.

We end this section by stating a one-dimensional version of Lemma 3 that also includes conditions under which the covariance under study is non-positive. Compared to Lehmann's result recalled in Theorem 1, the one below gives an expression for the covariance using integration by parts, instead of simply proving that the covariance is no larger than 0.

PROPOSITION 4. *Let f be a monotone function over $[0,1]$ that is one-side bounded. Let U and V be random variables with joint CDF $H(u,v)$ and marginal CDFs $F(u)$ and $G(v)$, respectively. Then there exists a function g and a finite (possibly signed) measure η_f such that*

$$\mathrm{Cov}(f(U),f(V))=\int\int[H(u,v)-F(u)G(v)]g(u)g(v)d\eta_f(u)d\eta_f(v).$$

Furthermore, if U and V are NQD, then $\mathrm{Cov}(f(U),f(V))\le 0$.

Proof. Using Proposition 1, we know that f satisfies the representation (15). Thus we can apply Lemma 3 and get

$$\begin{aligned}\mathrm{Cov}(f(U),f(V))&=\int\int[1-F(u)-G(v)+H(u,v)-(1-F(u)(1-G(v))]g(u)g(v)d\eta_f(u)d\eta_f(v)\\&=\int\int[H(u,v)-F(u)G(v)]g(u)g(v)d\eta_f(u)d\eta_f(v).\end{aligned}\tag{19}$$

The fact that f is monotone implies that g and η_f never change sign over their domain. Indeed, if f is monotone increasing, then if f is bounded we use $g=1$ and $\eta_f=\nu_f$ which is positive; if $f(1)=\infty$ then from the proof of Proposition 1 we see that the measure η_f constructed there is positive, thus g is positive as well since ν_f is positive and g is the corresponding Radon-Nikodym derivative. Similarly, the case of f non-decreasing means either g or η_f is always negative over its domain.

Combining this with the NQD property of U and V, it is clear that (19) is no larger than 0. □

5. Dependence structure for $(0,m,s)$-nets. The main results of this section are to first derive the joint distribution of two randomly chosen distinct points from a scrambled $(0,m,s)$-net in base b, and then to show that in the case $s=2$, this distribution is NQD. We also discuss the case $s=3$ and show that the approach used for $s=2$ fails when trying to demonstrate that pairs of points are NLOD when $s=3$.

Before presenting these results, we start by considering the case $s=1$. This simple case allows us to highlight the proof's key elements that need to be generalized to higher dimensions, which is what we do in the two subsequent subsections.

5.1. One-dimensional case. Based on the properties of scrambled nets recalled in Section 2.2, when choosing two points randomly from a scrambled $(0,m,1)$-net $\tilde{P}_n$, they can lie anywhere in the unit square $[0,1]^2$ except in C_m, since the two points cannot lie in the same interval $[\ell b^{-m},(\ell+1)b^{-m})$. Furthermore, we can prove that the joint pdf of these two points is uniform in the remaining part of the unit square. More precisely, we have:

PROPOSITION 5. *Let $\psi_m(x,y)$ be the joint pdf of two distinct points (U_I, U_J) randomly chosen from a scrambled $(0,m,1)$-net in base b. Then*

$$\psi_m(x,y) = \begin{cases} \frac{b^m}{b^m-1} & \text{if } (x,y) \in D_0 \cup \ldots \cup D_{m-1} \\ 0 & \text{if } (x,y) \in C_m. \end{cases}$$

Proof. The fact that $\psi_m(x,y) = 0$ when $(x,y) \in C_m$ has already been explained. Consider D_α, for $0 \le \alpha < m$. The point (U_I, U_J) is in D_α if and only if $\gamma_b(U_I, U_J) = \alpha$. Hence its joint pdf over D_α is given by

$$\frac{N_\alpha}{N} \frac{1}{\mathrm{Vol}(D_\alpha)},$$

where $N = b^m(b^m-1)/2$ is the number of pairs of points in the net, and N_α is the number of pairs (U_i, U_j) with $\gamma_b(U_i, U_j) = \alpha$. Based on the properties of a $(0,m,1)$-net, we have that

$$N_\alpha = b^\alpha \frac{b(b-1)}{2} b^{2(m-\alpha-1)}.$$

Indeed, in each of the b^α intervals of the form $I_\alpha(r) := [rb^{-\alpha}, (r+1)b^{-\alpha})$, $0 \le r < b^\alpha$, two points in $I_\alpha(r)$ are such that $\gamma_b(U_i, U_j) = \alpha$ if there are two distinct integers $s, s' \in \{0, \ldots, b-1\}$ such that $U_i \in I_{\alpha+1}(br+s) = [(br+s)b^{-\alpha-1}, (br+s+1)b^{-\alpha-1})$ and $U_j \in I_{\alpha+1}(br+s')$. Since there are $b^{m-\alpha-1}$ points in each interval of the form $I_{\alpha+1}(br+s)$ and there are $b(b-1)/2$ possible choices for (s,s'), overall we get $b^{2(m-\alpha-1)}b(b-1)/2$ pairs with the required property in each $I_\alpha(r)$. By multiplying this quantity by the number b^α of intervals $I_\alpha(r)$, we obtain N_α.

It should also be clear from the properties of a scrambled $(0,m,1)$-net (namely, from 2(i)-(ii) in Section 2.2) that the joint pdf of U_I and U_J is uniform over D_α. Finally, we have that

$$\mathrm{Vol}(D_\alpha) = b^\alpha b(b-1)b^{-2(\alpha+1)} = (b-1)b^{-\alpha-1}. \tag{20}$$

Hence we get that over D_α, the density of (U_I, U_J) is given by

$$\frac{b^\alpha b(b-1)b^{2(m-\alpha-1)}}{b^m(b^m-1)} \times \frac{b^{\alpha+1}}{b-1} = \frac{b^m}{b^m-1},$$

which does not depend on α, and is thus constant over $[0,1]^2 \backslash C_m = D_0 \cup \ldots \cup D_{m-1}$. □

Now that we have derived the joint pdf of two points from the scrambled net, we can prove they are NQD.

PROPOSITION 6. *Let U_I and U_J be two randomly chosen distinct points from a scrambled $(0,m,1)$-net in base b. Let $0 \le u, v \le 1$ and let $\alpha = \gamma_b(u,v)$. Then*

$$H(u,v;\tilde{P}_n) = \begin{cases} \frac{b^m}{b^m-1}\left(uv - \frac{\min(u,v)}{b^m}\right) & \text{if } \alpha < m \\ \frac{b^m}{b^m-1}\left[\frac{\lfloor b^m u \rfloor}{b^m}(u + v - (\lfloor b^m u \rfloor + 1)b^{-m})\right] & \text{if } \alpha \ge m. \end{cases}$$

Furthermore, U_I and U_J are NQD. That is, $H(u,v;\tilde{P}_n) \le uv$ for any $0 \le u, v \le 1$.

Proof. The idea is to integrate the pdf given in Proposition 5. In the case where $\alpha < m$, as illustrated in Figure 2 we have

$$H(u,v;\tilde{P}_n) = \frac{b^m}{b^m-1}\left(uv - b^{-m}\min(u,v)\right) \le uv,$$

since $uv \le \min(u,v)$.

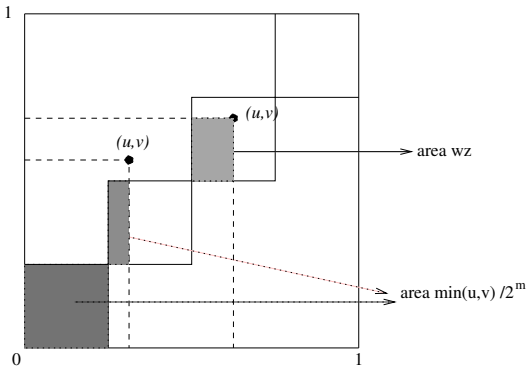

FIGURE 2. Computing the area corresponding to $H(u,v;\tilde{P}_n)$ for the two cases $\alpha < m$ (leftmost point) and $\alpha \geq m$ (rightmost point), when $b=2$ and $m=2$.

In the case where $\alpha \geq m$, let ℓ be such that $\frac{\ell}{b^m} \leq u,v < \frac{(\ell+1)}{b^m}$ (i.e., $\ell = \lfloor b^m u \rfloor = \lfloor b^m v \rfloor$). From Figure 2, it is easy to see that

$$H(u,v;\tilde{P}_n) = \int_0^u \int_0^v \psi_m(x,y)dxdy = \frac{b^m}{b^m-1}(uv - \ell b^{-2m} - wz), \tag{21}$$

where $w = (u - \ell b^{-m})$ and $z = v - \ell b^{-m}$. After some simplifications, we get

$$H(u,v;\tilde{P}_n) = \frac{1}{b^m-1}\left[\ell(u+v-(\ell+1)b^{-m}))\right].$$

We want to show that

$$\begin{aligned} H(u,v;\tilde{P}_n) \leq uv &\Leftrightarrow \ell(u+v-(\ell+1)b^{-m}) \leq (b^m-1)uv \\ &\Leftrightarrow B(u,v) := (b^m-1)uv - \ell(u+v) + \ell(\ell+1)b^{-m} \geq 0. \end{aligned}$$

The case $\ell = 0$ is trivial so we omit it. We consider two sub-cases: 1) assume $\ell = b^m - 1$. Then

$$B(u,v) = (b^m-1)(uv-(u+v)+1)$$

and since $u+v-1 \leq uv$, we have that $B(u,v) \geq 0$; 2) if $1 \leq \ell < b^m - 1$, then let $w = u - \ell b^{-m}$ and $z = v - \ell b^{-m}$, as before. We can write

$$\begin{aligned} B(u,v) &= b^m wz - uv + \ell b^{-m} \\ &= b^m wz - (w + \ell b^{-m})(z + \ell b^{-m}) + \ell b^{-m} \\ &= (b^m-1)wz - (w+z)\ell b^{-m} + \ell b^{-m}(b^m - \ell)/b^m, \end{aligned}$$

and thus $B(u,v) \geq 0$ if $w + z \leq (b^m - \ell)/b^m$, which holds since $w + z < 2/b^m$ and $\ell \leq b^m - 2$ and therefore $b^m - \ell \geq 2$. □

5.2. Volume loss & density inflation argument. When comparing the probability density function of pairs coming from a scrambled $(0,m,1)$-net in base b with those from a random sample, we can view the difference between them as arising from two phenomena, which we call "volume loss" and "density inflation". The *volume loss* refers to the fact that some areas of the unit square (namely C_m) have a density of 0 for the scrambled net, while the *density inflation* refers to the fact that for the rest of the unit square, the pair from the net has a density that is larger than what we get with random sampling (by a factor of $b^m/(b^m-1)$).

This type of analysis continues to be valid in higher dimensions, when comparing the conditional pdf of $(U_{I,\ell}, U_{J,\ell})$ given $U_{I,k} = x_k, U_{J,k} = y_k$, for $k = 1,\ldots,\ell-1$, with the uniform distribution over

$[0,1]^\ell$. The idea is to break down the integrals of the respective densities into pieces corresponding to the regions D_r and then apply the volume loss/density inflation argument. More precisely, for a given $(u,v) \in [0,1]^2$, and for $0 \le \ell \le m$, let

$$V_\ell = \int_{D_\ell \cap R(u,v)} dxdy \qquad \text{and} \qquad \tilde{V}_\ell = \int_{C_\ell \cap R(u,v)} dxdy, \tag{22}$$

where $R(u,v) = \{(x,y) \in [0,1]^2 : 0 \le x \le u, 0 \le y \le v\}$. We note that although the surface areas V_ℓ and $\tilde{V}_\ell$ both depend on (u,v) and on b, we do not indicate this in the notation so as to keep it light. Also, it should be clear that $V_\ell \le \tilde{V}_\ell$ since $D_\ell \subset C_\ell$.

The conditional pdf's of interest in higher dimensions are defined over $[0,1]^2$ and are of the form

$$\xi_r^{\mathbf{a}}(x,y) = \begin{cases} a_\ell & \text{if } (x,y) \in D_\ell, 0 \le \ell \le r-1 \\ 0 & \text{if } (x,y) \in C_r \end{cases} \tag{23}$$

with corresponding cdf given by

$$G(u,v) = \int_{R(u,v)} \xi_r^{\mathbf{a}}(x,y) dxdy = \sum_{\ell=0}^{r-1} a_\ell V_\ell. \tag{24}$$

In this form, they can be easily compared with the CDF associated with random uniform sampling, which we can write as

$$uv = \int_{R(u,v)} dxdy = \sum_{\ell=0}^{r-2} V_\ell + \tilde{V}_{r-1}. \tag{25}$$

The volume loss is captured by the term V_{r-1} in (24) rather than having $\tilde{V}_{r-1}$ in (25), and the density inflation is captured by the fact that at least some of the a_ℓ's in (24) must be larger than 1. To compare (24) with (25) in more detail, we must study further the areas V_ℓ and $\tilde{V}_\ell$. Their properties are summarized in the following proposition.

PROPOSITION 7. *Let V_ℓ and $\tilde{V}_\ell$ be defined as in (22), for a given $(u,v) \in [0,1]^2$ and $b \ge 2$. Then we have*

1. $V_\ell \le (b-1)\tilde{V}_\ell / b$ *for* $\ell \ge 0$;
2. $\tilde{V}_\ell \le b\tilde{V}_{\ell+1}$ *with equality when* $\ell > \gamma_b(u,v)$;
3. $V_\ell \le (b-1)\tilde{V}_{\ell+1}$ *for* $\ell \ge 0$.

Proof. To prove the first statement we rewrite the result of Proposition 6 for $n = b^\ell$ using our volume loss/density inflation framework and get

$$H(u,v;\tilde{P}_n) = \frac{b^\ell}{b^\ell - 1}(V_0 + V_1 + \ldots + V_{\ell-1}) = \frac{b^\ell}{b^\ell - 1}(\tilde{V}_0 - \tilde{V}_\ell).$$

Noting that $\tilde{V}_0 = uv$, it is easy to see from Proposition 6 that

$$\tilde{V}_\ell = \begin{cases} \frac{\min(u,v)}{b^\ell} & \text{if } \gamma_b(u,v) < \ell \\ uv - \frac{k_\ell}{b^\ell}\left(u + v - \frac{k_\ell+1}{b^\ell}\right) & \text{if } \gamma_b(u,v) \ge \ell, \end{cases} \tag{26}$$

where $k_\ell = \lfloor b^\ell \min(u,v) \rfloor$. (Note that if $\gamma_b(u,v) \ge \ell$ then $k_\ell = \lfloor b^\ell u \rfloor = \lfloor b^\ell v \rfloor$.)

On the other hand, since by definition, $V_\ell + \tilde{V}_{\ell+1} = \tilde{V}_\ell$ for $\ell \ge 0$, using the expression (26) for $\tilde{V}_\ell$ we get

$$V_\ell = \begin{cases} \frac{b-1}{b}\frac{\min(u,v)}{b^\ell} & \text{if } \gamma_b(u,v) < \ell \\ uv - \frac{k_\ell}{b^\ell}\left(u + v - \frac{k_\ell+1}{b^\ell}\right) - \frac{\min(u,v)}{b^{\ell+1}} & \text{if } \gamma_b(u,v) = \ell \\ \frac{k_{\ell+1}}{b^{\ell+1}}\left(u + v - \frac{k_{\ell+1}+1}{b^{\ell+1}}\right) - \frac{k_\ell}{b^\ell}\left(u + v - \frac{k_\ell+1}{b^\ell}\right) & \text{if } \gamma_b(u,v) > \ell. \end{cases}$$

When $\gamma_b(u,v) < \ell$, we thus have $V_\ell = ((b-1)/b)\tilde{V}_\ell$, $\tilde{V}_\ell = b\tilde{V}_{\ell+1}$ and $V_\ell = bV_{\ell+1}$. When $\gamma_b(u,v) = \ell$, then

$$V_\ell = \tilde{V}_\ell - \frac{\min(u,v)}{b^{\ell+1}}$$

and thus $V_\ell \le ((b-1)/b)\tilde{V}_\ell$ if and only if

$$\tilde{V}_\ell \le \frac{\min(u,v)}{b^\ell} \Leftrightarrow \left(u - \frac{k_\ell}{b^\ell}\right)\left(v - \frac{k_\ell}{b^\ell}\right) \le \frac{\min(u,v) - k_\ell/b^\ell}{b^\ell},$$

which holds since $(\max(u,v) - k_\ell/b^\ell) \le 1/b^\ell$.

When $\gamma_b(u,v) > \ell$, then showing that $V_\ell \le ((b-1)/b)\tilde{V}_\ell$, or equivalently, that $\tilde{V}_\ell/b \le \tilde{V}_{\ell+1}$, is equivalent to showing that

$$\left(u - \frac{k_{\ell+1}}{b^{\ell+1}}\right)\left(v - \frac{k_{\ell+1}}{b^{\ell+1}}\right) + \frac{k_{\ell+1}}{b^{\ell+1}}\frac{1}{b^{\ell+1}} \ge \frac{1}{b}\left[\left(u - \frac{k_\ell}{b^\ell}\right)\left(v - \frac{k_\ell}{b^\ell}\right) + \frac{k_\ell}{b^\ell}\frac{1}{b^\ell}\right]. \tag{27}$$

Now, we can write $k_{\ell+1} = bk_\ell + r$, where $0 \le r < b$ and then use the relation

$$\frac{k_{\ell+1}}{b^{\ell+1}} = \frac{k_\ell}{b^\ell} + \frac{r}{b^{\ell+1}}$$

to simplify (27) to

$$\left(u - \frac{k_{\ell+1}}{b^{\ell+1}}\right)\left(v - \frac{k_{\ell+1}}{b^{\ell+1}}\right) + \frac{r}{b^{\ell+1}}\frac{1}{b^{\ell+1}} \ge \frac{1}{b}\left(u - \frac{k_\ell}{b^\ell}\right)\left(v - \frac{k_\ell}{b^\ell}\right). \tag{28}$$

If we let $\tilde{u} = b^\ell(u - \frac{k_\ell}{b^\ell})$ and $\tilde{v} = b^\ell(v - \frac{k_\ell}{b^\ell})$, then we can rewrite (28) as

$$\left(\tilde{u} - \frac{r}{b}\right)\left(\tilde{v} - \frac{r}{b}\right) + \frac{r}{b^2} \ge \frac{1}{b}\tilde{u}\tilde{v}. \tag{29}$$

Since $\gamma_b(\tilde{u},\tilde{v}) \ge 1$ and it can be verified that $r = \lfloor b\min(\tilde{u},\tilde{v})\rfloor$, we can apply Proposition 6, whose proof (adapted to the notation used here and for the case $m = 1$) shows that when $\gamma_b(\tilde{u},\tilde{v}) \ge 1$, then (see (21))

$$\frac{b}{b-1}\left(\tilde{u}\tilde{v} - \frac{r}{b^2} - \left(\tilde{u} - \frac{r}{b}\right)\left(\tilde{v} - \frac{r}{b}\right)\right) \le \tilde{u}\tilde{v},$$

which shows that (29) holds.

For the second statement, in the above steps we saw that $\tilde{V}_\ell = b\tilde{V}_{\ell+1}$ when $\ell > \gamma_b(u,v)$. More generally, since $V_\ell = \tilde{V}_\ell - \tilde{V}_{\ell+1}$, the first statement implies that $\tilde{V}_\ell \le b\tilde{V}_{\ell+1}$ for all $\ell \ge 0$. Finally, it should be clear that together, the first and second statements imply the third one. □

5.3. Probability density function associated with scrambled $(0,m,s)$-nets. The other key element in our study of the one-dimensional case was to determine the pdf for pairs of points from a scrambled $(0,m,1)$-net in base b over D_α for a given $\alpha \in \{0,\dots,m-1\}$. To generalize this result to scrambled $(0,m,s)$-nets of arbitrary dimension $s \ge 1$, we study the conditional pdf of pairs of points $(U_{I,j}, U_{J,j})$ given that $\gamma_b(U_{I,k}, U_{J,k}) = \alpha_k$ for $k = 1,\dots,j-1$, and for $2 \le j \le s$. This conditional pdf is of the form $\xi^{\mathbf{a}}_{m-A_{j-1}}$ as given in (23), with $A_{j-1} = \alpha_1 + \ldots + \alpha_{j-1}$, and with the specific values of a_l determined in Proposition 8 below. Before we present this result, we need to introduce a counting function and a technical lemma describing how this function can be evaluated. Its proof and more information on this result are given in the appendix.

DEFINITION 6. Let P_n be a digital $(0,m,s)$-net in base b. We define $M(m,s)$ as the number of points in P_n that are in

$$\bigcup_{1\le r_j<b, j=1,\ldots,s} I(r_1,\ldots,r_s) \qquad \text{where } I(r_1,\ldots,r_s)=\prod_{j=1}^{s}\left[\frac{r_j}{b},\frac{r_j+1}{b}\right).$$

LEMMA 4. *Let P_n be a digital $(0,m,s)$-net in base b. (i) When $s\ge 1$ and $m\ge s$ we have*

$$M(m,s)=(b-1)^s b^{m-s}.$$

(ii) For $s>m$, $M(m,s)$ can be computed using the initial values

$$\begin{aligned} M(1,s)&=(b-1) && s>1,\\ M(2,s)&=(b-1)(b-(s-1)) && s>2, \end{aligned}$$

and the relation

$$M(m,s)=(b-1)^{m-1}(b-(s-m+1))+\sum_{k=2}^{m_s^*}(k-1)N(k,m,s),$$

which holds for $s\ge 3$ and $2\le m<s$, and where $m_s^*=\min(m-1,s-m+1)$, with

$$N(k,m,s)=M(m-k,m-1)\binom{s-m+1}{k}-\sum_{i=k+1}^{m_s^*}N(i,m,s)\binom{i}{k}$$

for $2\le k\le m_s^*$, $3\le m\le s-1$, and $s\ge 4$.

We are now ready to present the main result of this section, whose proof is in the appendix.

PROPOSITION 8. *Let $(\mathbf{U}_I,\mathbf{U}_J)$ be a randomly chosen pair of distinct points from a scrambled $(0,m,s)$-net in base b. Then the conditional pdf of $(U_{I,j},U_{J,j})$ given that $\gamma_b(U_{I,k},U_{J,k})=\alpha_k$ for $k=1,\ldots,j-1$, and $2\le j\le s$, has the form $\xi^{\mathbf{a}}_{m-A_{j-1}}$ given in (23), where $A_{j-1}=\alpha_1+\ldots+\alpha_{j-1}$, and with*

$$a_\ell=\frac{M(m-A_{j-1}-\ell,j)}{M(m-A_{j-1},j-1)}\frac{b^{\ell+1}}{b-1}$$

for $\ell=0,\ldots,m-A_{j-1}-1$.

REMARK 9. Note that the conditional density associated with a scrambled $(0,m,s)$-net $\tilde{P}_n$ is constant over the regions D_α and these regions are invariant under the transformation $\tau(\mathbf{u})=1-\mathbf{u}$ for $\mathbf{u}\in[0,1]^2$. Hence the (unconditional) pdf of two randomly chosen points $(\mathbf{U}_I,\mathbf{U}_J)$ from $\tilde{P}_n$ is invariant under the transformation $\tau(\mathbf{x})=1-\mathbf{x}$ for $\mathbf{x}\in[0,1]^{2s}$, which means that the NUOD and NLOD properties are equivalent for this type of sampling scheme.

5.4. Case where $s=2$. We now study the dependence structure of scrambled nets in the two-dimensional case, using the results from the previous section. More precisely, we determine the conditional joint pdf of $(U_{I,2},U_{J,2})$ given $\alpha_1=\gamma_b(u_1,v_1)$, where $U_{I,1}=u_1, U_{J,1}=v_1$. We will see that its form is different than the one we found in the one-dimensional case, but is still NQD.

Before we present this result, we first introduce some notation to denote a special type of pdf over $[0,1]^2$ that will be used to analyze the $s=2$ case.

DEFINITION 7. Let $b \geq 2$ be an integer. For $r \geq 1$, the pdf $\tilde{\psi}_r(x,y)$ is defined as

$$\tilde{\psi}_r(x,y) = \begin{cases} 1 & \text{if } (x,y) \in D_0 \cup D_1 \cup \ldots \cup D_{r-2} \\ \frac{b}{b-1} & \text{if } (x,y) \in D_{r-1} \\ 0 & \text{if } (x,y) \in C_r. \end{cases}$$

That is, $\tilde{\psi}_r(x,y) = \xi_r^{\mathbf{a}}(x,y)$ with $a_\ell = 1$ for $0 \leq \ell \leq r-2$, and $a_{r-1} = b/(b-1)$.

LEMMA 5. *For $r \geq 1$, if (U,V) has pdf $\tilde{\psi}_r(x,y)$ over $[0,1]^2$, then U and V are NQD.*

Proof. We first note that the corresponding univariate marginal distributions of $\tilde{\psi}$ are both given by

$$\sum_{\ell=0}^{r-1} a_\ell (b-1) b^{-(\ell+1)} = \frac{b-1}{b}\left(\frac{1-b^{-(r-1)}}{1-b^{-1}}\right) + \frac{b}{b-1}\frac{b-1}{b^r} = (1-b^{-(r-1)}) + b^{-(r-1)} = 1$$

over $[0,1]$ and thus $U[0,1]$. Next we note that if $r=1$, then $\tilde{\psi}_1(x,y) = \psi_1(x,y)$ which was shown to be NQD in Proposition 6. So assume $r > 1$ and let $(u,v) \in [0,1]^2$. Define $\ell_1 = \lfloor b^{r-1}u \rfloor$ and $\ell_2 = \lfloor b^{r-1}v \rfloor$. Let $G_r(u,v)$ be the cdf corresponding to $\tilde{\psi}_r(x,y)$. Then

$$G_r(u,v) = V_0 + \ldots + V_{r-2} + \frac{b}{b-1} V_{r-1}$$

while $uv = V_0 + \ldots + V_{r-2} + \tilde{V}_{r-1}$. Therefore we have

$$G_r(u,v) \leq uv \Leftrightarrow \frac{b}{b-1} V_{r-1} \leq \tilde{V}_{r-1},$$

which follows directly from Item 1 in Proposition 7. □

PROPOSITION 9. *Let $(\mathbf{U}_I, \mathbf{U}_J)$ be a randomly chosen pair of distinct points from a scrambled $(0,m,2)$-net in base b. Then its pdf (evaluated at $\mathbf{U}_I = (x_1,x_2)$ and $\mathbf{U}_J = (y_1,y_2)$) is given by*

$$\phi_m(x_1,x_2,y_1,y_2) = \psi_m(x_1,y_1)\tilde{\psi}_{m-\alpha_1}(x_2,y_2), \qquad 0 \leq x_i, y_i < 1, i = 1,2,$$

where $\alpha_1 = \gamma_b(x_1,y_1)$ and where $\psi_m(\cdot,\cdot)$ was introduced in Proposition 5.

Proof. Since $\{(U_{i,1}, i = 1,\ldots,b^m\}$ is a scrambled $(0,m,1)$-net, we can use Proposition 5 to establish that the marginal joint pdf for $(U_{I,1}, U_{J,1})$ is given by ψ_m and so it suffices to show that the joint pdf of $(U_{I,2}, U_{J,2})$ given $U_{I,1} = x_1, U_{J,1} = y_1$ and evaluated at x_2, y_2 is given by $\tilde{\psi}_{m-\alpha_1}(x_2,y_2)$.

Given α_1, we know that $\gamma_b(U_{I,2}, U_{J,2}) \leq m - \alpha_1 - 1$. Using Proposition 8, we have that the conditional pdf of $(U_{I,2}, U_{J,2})$ given that $\gamma_b(U_{I,1}, U_{j,1}) = \alpha_1$ is of the form $\xi_{m-\alpha_1}^{\mathbf{a}}$ with a_ℓ given by

$$a_\ell = \frac{M(m-\alpha_1-\ell,2)}{M(m-\alpha_1,1)} \frac{b^{\ell+1}}{b-1}, \qquad \ell = 0,\ldots,m-\alpha_1-1.$$

Recall that a formula for $\mathrm{Vol}(D_\alpha)$ was established in (20). We have two cases to consider: first, if $m-\alpha_1-\ell \geq 2$, then

$$a_\ell = \frac{(b-1)^2 b^{m-\alpha_1-\ell-2}}{(b-1)b^{m-\alpha_1-1}} \frac{b^{\ell+1}}{b-1} = 1.$$

Second, if $m-\alpha_1-\ell \leq 1$, then in fact we must have $\ell = m-\alpha_1-1$, and in that case

$$a_\ell = \frac{M(1,2)}{M(m-\alpha_1,1)} \frac{b^{m-\alpha_1}}{b-1} = \frac{b-1}{(b-1)b^{m-\alpha_1-1}} \frac{b^{m-\alpha_1}}{b-1} = \frac{b}{b-1},$$

as required. □

COROLLARY 1. *A scrambled* $(0,m,2)$*-net in base* b *is an NLOD sampling scheme and an NUOD sampling scheme.*

Proof. Combining Proposition 9 and Lemma 5 and using the notation $\alpha_1 = \gamma_b(x_1,y_1)$, we get

$$H(\mathbf{u},\mathbf{v};\tilde{P}_n) = \int_0^{u_1}\int_0^{v_1} \psi_m(x_1,y_1)\int_0^{u_2}\int_0^{v_2} \tilde{\psi}_{m-\alpha_1}(x_2,y_2)dx_2dy_2dx_1dy_1 \le u_1v_1u_2v_2.$$

This implies the NLOD property; the NUOD property follows from Remark 9. □

REMARK 10. For a scrambled net, the pairs of coordinates $(U_{I,1},U_{J,1}),\dots,(U_{I,s},U_{J,s})$ are not independent. Proposition 9 demonstrates this in the two-dimensional case. This is an important difference with the AV and LHS sampling schemes mentioned in the introduction. When pairs of coordinates are independent and identically distributed, sufficient conditions to get a variance reduction compared to MC is to have an NQD distribution for each such pair, i.e., a coordinate-wise independent NQD sampling scheme, and a function monotone in each coordinate. This is what is established in Lehmann's result that was recalled in Theorem 1, using an inductive approach. When pairs are not independent, we can still follow Lehmann's inductive approach if each conditional distribution for the pairs are NQD, as we see is the case here for scrambled $(0,m,2)$-nets. This is the approach considered in Proposition 12 and Corollary 2 of Section 6. When this does not hold, one can resort to the general covariance result from Proposition 3 in Section 4, as long as the sampling scheme can be shown to be NUOD. However the conditions on f are then stronger.

5.5. Case where $s=3$. We now move on to the three-dimensional case, first providing an expression for the conditional distribution of the pair $(U_{I,3},U_{J,3})$ given α_1 and α_2, and then showing that this conditional pdf is not NQD.

PROPOSITION 10. *Let* $(\mathbf{U}_I,\mathbf{U}_J)$ *be a randomly chosen pair of distinct points coming from a scrambled* $(0,m,3)$*-net in base* b*. Then the joint pdf of* $(U_{I,3},U_{J,3})$ *evaluated at* (x_3,y_3) *and given* $\gamma_b(U_{I,j},U_{J,j}) = \alpha_j$ *for* $j=1,2$ *is of the form* $\xi^{\mathbf{a}}_{m-\alpha_1-\alpha_2}(x_3,y_3)$ *with*

$$a_\ell = \begin{cases} 1 & 0 \le \ell \le m-\alpha_1-\alpha_2-3 \\ \frac{b(b-2)}{(b-1)^2} & \ell = m-\alpha_1-\alpha_2-2 \\ \frac{b^2}{(b-1)^2} & \ell = m-\alpha_1-\alpha_2-1. \end{cases}$$

if $\alpha_1+\alpha_2 \le m-3$*; by*

$$a_\ell = \begin{cases} \frac{b(b-2)}{(b-1)^2} & \ell = 0 \\ \frac{b^2}{(b-1)^2} & \ell = 1 \\ 0 & \ell > 1 \end{cases}$$

if $\alpha_1+\alpha_2 = m-2$*; and by*

$$a_\ell = \begin{cases} \frac{b}{b-1} & \ell = 0 \\ 0 & \ell > 0 \end{cases}$$

if $\alpha_1+\alpha_2 = m-1$.

Proof. From Proposition 8, we have that

$$a_\ell = \frac{M(m-A_2-\ell,3)}{M(m-A_2,2)}\frac{b^{\ell+1}}{b-1}.$$

Here we have three cases to consider, and in each case Lemma 4 is used to get an expression for $M(r,k)$ as a function of b, r, k, α_1, and α_2:

Case 1 ($m - A_2 - \ell \geq 3$): then we have

$$a_\ell = \frac{(b-1)^3 b^{m-A_2-\ell-3}}{(b-1)^2 b^{m-A_2-2}} \frac{b^{\ell+1}}{b-1} = 1.$$

Case 2 ($m - A_2 - \ell \leq 2$ and $m - A_2 \geq 2$): first, we note that in this case there are only two possible values for ℓ, which are $m - A_2 - 2$ or $m - A_2 - 1$, respectively corresponding to values of 2 and 1 for $m - A_2 - \ell$. In the first case, we get

$$a_{m-A_2-2} = \frac{(b-1)(b-2)}{(b-1)^2 b^{m-A_2-2}} \frac{b^{\ell+1}}{b-1} = \frac{b(b-2)}{(b-1)^2},$$

while in the second we get

$$a_{m-A_2-1} = \frac{b-1}{(b-1)^2 b^{m-A_2-2}} \frac{b^{\ell+1}}{b-1} = \frac{b^2}{(b-1)^2}.$$

Case 3 ($m - A_2 = 1$): in this case we must have $\ell = 0$ and we then get $a_0 = (b-1)(b/(b-1))/(b-1) = b/(b-1)$. Reorganizing the results according to the possible range of values for $A_2 = \alpha_1 + \alpha_2$, we get the expression given in the statement of the proposition. □

PROPOSITION 11. *The conditional distribution $H_m(u_3, v_3 | u_1, v_1, u_2, v_2)$ of a randomly chosen pair of distinct points coming from a scrambled $(0, m, 3)$-net in base b is not NQD.*

Proof. Consider the case where $\alpha_1 + \alpha_2 = m - 2$. We then have

$$H_m(u_3, v_3 | u_1, v_1, u_2, v_2) = \frac{b(b-2)}{(b-1)^2} V_0 + \left(\frac{b}{b-1}\right)^2 V_1.$$

In order for this to be NQD, it must be bounded by $V_0 + \tilde{V}_1$ for all pairs (u_3, v_3). But consider the case $u_3 = v_3 = 1/b$. Then $V_0 = 0$, $V_1 = (b-1)/b^3$ and $\tilde{V}_1 = 1/b^2$. So $H_m(u_3, v_3 | u_1, v_1, u_2, v_2) = 1/(b(b-1)) > V_0 + \tilde{V}_1 = 1/b^2$. □

Using our volume loss/density inflation argument, we see that the problem here is that we can have an inflation factor of $(b/(b-1))^2$ that is too large to be offset by the corresponding volume loss of $(b-1)/b$.

REMARK 11. Proposition 11 implies that when $s = 3$ the conditional distribution of the pairs of coordinates of $\mathbf{U}_I, \mathbf{U}_J$ are not all NQD. Of course, this does not mean that the unconditional (6-dimensional) distribution of $(\mathbf{U}_I, \mathbf{U}_J)$ is not NUOD. However, so far we have not been able to show that it was. To do so, one needs to establish bounds of the form

$$\sum_{\ell_2=0}^{m-1-\alpha_1} a_{2,\ell_2} V_{2,\ell_2} \sum_{\ell_3=0}^{m-1-\alpha_1-\ell_2} a_{3,\ell_3} V_{3,\ell_3} \leq \tilde{V}_{2,0} \tilde{V}_{3,0} \tag{30}$$

for every $\alpha_1 = 0, \ldots, m-1$ and $(u_2, v_2, u_3, v_3) \in [0,1)^4$, and where V_{j,ℓ_j} is associated with the pair (u_j, v_j) for $j = 2, 3$,

$$a_{2,\ell_2} = \begin{cases} 1 & \text{if } 0 \leq \ell_2 \leq m - \alpha_1 - 2 \\ \frac{b}{b-1} & \text{if } \ell_2 = m - \alpha_1 - 1 \\ 0 & \text{otherwise,} \end{cases}$$

and a_{3,ℓ_3} is specified as in Proposition 10. The RHS of (30) can be rewritten using the identities

$$\tilde{V}_{2,0} = \sum_{\ell_2=0}^{m-1-\alpha_1} V_{2,\ell_2} + \tilde{V}_{2,m-1-\alpha_1} \qquad \tilde{V}_{3,0} = \sum_{\ell_3=0}^{m-1-\alpha_1-\ell_2} V_{3,\ell_3} + \tilde{V}_{3,m-1-\alpha_1-\ell_2}$$

so as to match more closely the LHS of (30).

6. Variance bounds based on negative quadrant dependence. Now that we have shown that a randomly chosen pair of distinct points $(\mathbf{U}_I, \mathbf{U}_J)$ from a scrambled net is NUOD for $s=2$ and that the conditional distribution of $(\mathbf{U}_{I,2}, \mathbf{U}_{J,2})$ given $U_{I,1}, U_{J,1}$ is NQD, we can use these properties to derive an interesting result on the variance of the estimator for $I(f)$ based on a scrambled $(0,m,2)$-net in base b.

PROPOSITION 12. *Let f be a function defined over $[0,1]^2$ that is monotone in each coordinate. Let $\hat{\mu}_n$ be the estimator for $I(f)$ based on a scrambled $(0,m,2)$-net in base b, and let $\hat{\mu}_{mc,n}$ be the MC estimator for $I(f)$ based on $n=b^m$ points. Then*

$$\mathrm{Var}(\hat{\mu}_n) \le \mathrm{Var}(\hat{\mu}_{mc,n}).$$

Before we give the proof of this result, we present yet another extension (and a corollary) of Lehmann's key result recalled in Theorem 1.

PROPOSITION 13. *Let $(X_1,\ldots,X_s)$ and $(Y_1,\ldots,Y_s)$ be vectors of random variables with the property that the distribution of (X_1,Y_1) and the conditional distribution of (X_j,Y_j) given (X_1,Y_1), $\ldots,(X_{j-1},Y_{j-1})$ for each $j=2,\ldots,s$ are NQD. Let f and g be functions of s variables and let $X=f(X_1,\ldots,X_s)$, $Y=g(Y_1,\ldots,Y_s)$. Then $\mathrm{Cov}(X,Y)\le 0$ if f and g are concordant in the jth coordinate, for $j=1,\ldots,s$.*

Proof. We employ the same inductive approach as the one used in [12] to prove Theorem 1. So the result clearly holds for $s=1$ as a consequence of Theorem 1. Suppose it holds for $s-1$. For the case of s variables, we have

$$\mathrm{E}(XY) = \mathrm{E}[\mathrm{E}(f(X_1,\ldots,X_s)g(Y_1,\ldots,Y_s)|X_1=x_1,Y_1=y_1,\ldots,X_{s-1}=x_{s-1},Y_{s-1}=y_{s-1})]. \quad (31)$$

Now for fixed $x_1,y_1,\ldots,x_{s-1},y_{s-1}$, $f(x_1,\ldots,x_{s-1},x_s)$ and $g(y_1,\ldots,y_{s-1},y_s)$ are one-dimensional functions that are concordant (in x_s and y_s, respectively), and since the inner expectation in (31) is taken with respect to the conditional distribution of (X_s,Y_s) given $(X_1,Y_1),\ldots,(X_{s-1},Y_{s-1})$, which is NQD, we can apply again Theorem 1 (when $s=1$) to show that this inner expectation is bounded from above by $\mathrm{E}(f(X_1,\ldots,X_s)|X_i=x_i,i=1,\ldots,s-1)\mathrm{E}(g(Y_1,\ldots,Y_s)|Y_i=y_i,i=1,\ldots,s-1)$. Next we define

$$\begin{aligned}\tilde{f}(x_1,\ldots,x_{s-1}) &= \mathrm{E}(f(X_1,\ldots,X_s)|X_1=x_1,\ldots,X_{s-1}=x_{s-1})\\ \tilde{g}(x_1,\ldots,x_{s-1}) &= \mathrm{E}(g(Y_1,\ldots,Y_s)|Y_1=y_1,\ldots,Y_{s-1}=y_{s-1})\end{aligned}$$

and argue as in [12] that $\tilde{f}$ and $\tilde{g}$ are concordant in coordinate i for $i=1,\ldots,s-1$. Hence we can apply the induction hypothesis to handle the outer expectation of (31) and thus get

$$\mathrm{E}(XY) \le \mathrm{E}(\tilde{f}(X_1,\ldots,X_{s-1}))\mathrm{E}(\tilde{g}(Y_1,\ldots,Y_{s-1})) = \mathrm{E}(f(X_1,\ldots,X_s))\mathrm{E}(g(Y_1,\ldots,Y_s)),$$

as required. □

COROLLARY 2. *Let $\tilde{P}_n$ be a sampling scheme such that $U_{I,1},U_{J,1}$ are NQD and given $(U_{I,l},U_{J,l})$, $l=1,\ldots,j-1$, $(U_{J,j},U_{J,j})$ are NQD, for $j=2,\ldots,s$. Let $\hat{\mu}_n$ be the estimator of $I(f)$ corresponding to $\tilde{P}_n$. If f is monotone in each variable, then $\mathrm{Var}(\hat{\mu}_n)\le\mathrm{Var}(\hat{\mu}_{mc,n})$.*

Proof of Proposition 12. As seen in (6), we have that

$$\mathrm{Var}(\hat{\mu}_n) = \mathrm{Var}(\hat{\mu}_{mc,n}) + \frac{2}{n(n-1)}\sum_{i<j}\mathrm{Cov}(f(\mathbf{U}_i),(\mathbf{U}_j)),$$

where

$$\frac{2}{n(n-1)}\sum_{i<j}\mathrm{Cov}(f(\mathbf{U}_i),(\mathbf{U}_j)) = \mathrm{Cov}(f(\mathbf{U}_I),f(\mathbf{U}_J)),$$

and thus we simply need to show that $\mathrm{Cov}(f(\mathbf{U}_I),f(\mathbf{U}_J)) \le 0$. But this follows directly from Corollary 2, whose conditions on the distribution of $(\mathbf{U}_I,\mathbf{U}_J)$ are shown to hold via Lemma 5, Proposition 6 and Proposition 9. □

REMARK 12. If f satisfies the representation (15) with g and η_f that do not change sign over their domain, then instead of using Corollary 2 to prove Proposition 12, we can use Corollary 1 (establishing the NUOD property for scrambled $(0,m,2)$-nets) and Proposition 3. The latter requires stronger conditions on f but has the advantage of providing an explicit expression for the covariance instead of simply proving it is no larger than 0.

REMARK 13. In [18] (see also [19]), the variance bound $(b/(b-1))^{s-1}\mathrm{Var}(\hat{\mu}_{mc,n})$ for scrambled $(0,m,s)$-nets in base b is shown to hold for all square-integrable functions. Our result does not apply as widely — in particular because we can only prove it for $s=2$ for now — but it holds with a constant of 1 for any function that is monotone in each coordinate, rather than the constant 2 we get from the general result of Owen mentioned above, in the case $s=b=2$.

We end this section with a result extending Proposition 12 to the case of s-dimensional functions that can be written as a sum of bivariate ANOVA components.

PROPOSITION 14. *Let f be a function defined over $[0,1]^s$ that is monotone in each coordinate and has an effective dimension of 2 in the superposition sense (in proportion 1). Let $\hat{\mu}_n$ be the estimator for $I(f)$ based on a scrambled $(0,m,s)$-net in base b. Then*

$$\mathrm{Var}(\hat{\mu}_n) \le \mathrm{Var}(\hat{\mu}_{mc,n}).$$

Proof. Although it is not stated explicitly in [18], if we denote by $\hat{\mu}_n(f)$ the scrambled $(0,m,s)$-net estimator for a function f, then from the properties of the ANOVA components and the Haar wavelet expansion introduced by Owen to study the variance of scrambled nets in [18], we have

$$\mathrm{Var}(\hat{\mu}_n(f)) = \sum_{\emptyset\neq\mathcal{I}\subseteq\{1,\dots,s\}} \mathrm{Var}(\hat{\mu}_n(f_{\mathcal{I}})). \tag{32}$$

Since each two-dimensional projection of a scrambled $(0,m,s)$-net is a scrambled $(0,m,2)$-net and our assumption on f implies that $f_{\mathcal{I}}=0$ if $|\mathcal{I}|>2$, we simply apply Proposition 12 to the RHS of (32) to get

$$\mathrm{Var}(\hat{\mu}_n(f)) \le \sum_{\mathcal{I}:|\mathcal{I}|\le 2} \sigma^2_{\mathcal{I}}/n = \mathrm{Var}(\hat{\mu}_{mc,n}). \quad \square$$

7. Conclusion. In this paper we have provided a new framework to study the variance of estimators based on dependent sample points, such as those obtained from RQMC methods. The framework includes a useful representation (Proposition 2) for the covariance term characterizing a dependent sampling scheme which makes use of our generalization of Hoeffding's lemma (Lemma 3), a result we believe is in itself an original contribution. Building on this, we provided a general variance reduction result in Proposition 3 giving sufficient conditions on f to guarantee that an NUOD sampling scheme will do now worse than MC, and providing an explicit expression for the difference between the two variances. Another contribution of this paper is to have provided an explicit formula (Proposition 8) to describe the joint distribution of pairs of points randomly chosen from a scrambled $(0,m,s)$-net. In turn, this allowed us to prove that the estimator based on a scrambled $(0,m,2)$-net is no larger than the MC estimator variance for two-dimensional functions that are monotone in each coordinate (Proposition 12). The general bound proved by Owen for scrambled nets holds within a factor of two of the MC variance, and thus for this special case

we were able to improve this result. We have also provided two different sets of conditions where an s-dimensional dependent sampling scheme has a variance no larger than the MC variance; in one case (Proposition 14), we consider scrambled $(0, m, s)$-nets and require functions that are essentially the sum of one or two-dimensional functions, and in the other case (Corollary 2) we require functions that are monotone in each coordinate, along with a sampling scheme for which the nested conditional distributions of its pairs are all NQD.

For future work, we plan to study further the dependence properties of scrambled (t, m, s)-nets and other RQMC sampling schemes for dimensions $s \geq 2$, and find out whether they are NUOD or not. We also plan to look into the possibility of designing sampling schemes having the NUOD, the coordinate-wise independent NQD property, or the conditional NQD property from Corollary 2. In terms of the general variance bound for NUOD sampling schemes given in Proposition 3, we would like to relax the conditions on the function g and measure η_f representing f that are required for this result. Determining necessary and sufficient conditions on f for the representation (15) to hold would also be of interest. In addition, we plan to take a closer look at the covariance decomposition (18) and get a better understanding of the type of functions that minimize this quantity, in light of what we could call our "discrepancy function" $T(\mathbf{u}, \mathbf{v}; \tilde{P}_n) - \prod_{j=1}^{s}(1-u_j)(1-v_j)$. Finally, it is clear that our study of the dependence structure of scrambled nets makes use of quantities that also appear in the variance formula based on Haar wavelets developed by Owen in [18] and other references. Developing a better understanding of these connections would be another goal for future research.

Appendix

Proof of Lemma 2. For part (i), both results make use of the following simple fact: consider $\mathcal{J}$ a strict subset of $\{1, \ldots, s\}$ and $\mathcal{I} \subseteq \{1, \ldots, s\}$. Then it is clear that

$$\sum_{\substack{\mathcal{L} \subseteq \{1,\ldots,s\} \\ \mathcal{L} \cap \mathcal{J} = \mathcal{I} \cap \mathcal{J}}} (-1)^{|\mathcal{L}|} = 0, \tag{33}$$

since

$$\sum_{\substack{\mathcal{L} \subseteq \{1,\ldots,s\} \\ \mathcal{L} \cap \mathcal{J} = \mathcal{I} \cap \mathcal{J}}} (-1)^{|\mathcal{L}|} = \sum_{j=0}^{s-|\mathcal{J}|} \binom{s-|\mathcal{J}|}{j} (-1)^j = 0$$

and $0 \leq |\mathcal{J}| < s$.

We also use the principle of inclusion-exclusion to get, for a measure μ,

$$\mu\left(\prod_{i=1}^{s}[a_i, b_i]\right) = \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k \mu\left(\prod_{i=1}^{s}[0, b_i + \mathbf{1}_{i\in\mathcal{I}}(a_i - b_i)]\right). \tag{34}$$

So first we write

$$\begin{aligned}
\mu_f(A) &= \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k \mu_f\Big(\prod_{i=1}^{s}[0, b_i + \mathbf{1}_{i\in\mathcal{I}}(a_i - b_i)]\Big) \\
&= \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k \Delta^{(s)}\Big(f; \prod_{i=1}^{s}[0, b_i + \mathbf{1}_{i\in\mathcal{I}}(a_i - b_i)]\Big) \\
&= \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} \sum_{\mathcal{J}} (-1)^{k+|\mathcal{J}|} f(\mathbf{a}^{\mathcal{I}\cap\mathcal{J}^c}; \mathbf{b}^{\mathcal{I}^c\cap\mathcal{J}^c}; \mathbf{0}^{\mathcal{J}}),
\end{aligned}$$

where we used the notation $f(\mathbf{a}^{\mathcal{U}}, \mathbf{b}^{\mathcal{V}}, \mathbf{0}^{\mathcal{U}^c \cup \mathcal{V}^c})$ with $\mathcal{U} \cap \mathcal{V} = \emptyset$ to represent $f(x_1, \ldots, x_s)$ where

$$x_j = \begin{cases} a_j & \text{if } j \in \mathcal{U} \\ b_j & \text{if } j \in \mathcal{V} \\ 0 & \text{otherwise.} \end{cases}$$

Hence to prove that $\mu_f(A) = \Delta^{(s)}(f; A)$ it suffices to show that

$$\sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} \sum_{\mathcal{J}} (-1)^{k+|\mathcal{J}|} f(\mathbf{a}^{\mathcal{I}\cap\mathcal{J}^c}; \mathbf{b}^{\mathcal{I}^c\cap\mathcal{J}^c}; \mathbf{0}^{\mathcal{J}}) = \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k f(\mathbf{a}^{\mathcal{I}}; \mathbf{b}^{\mathcal{I}^c}) \tag{35}$$

But since for a given non-empty $\mathcal{I}$ we have that

$$f(\mathbf{a}^{\mathcal{I}^c\cap\mathcal{J}_1}; \mathbf{b}^{\mathcal{I}^c\cap\mathcal{J}_1^c}; \mathbf{0}^{\mathcal{I}}) = f(\mathbf{a}^{\mathcal{I}^c\cap\mathcal{J}_2}; \mathbf{b}^{\mathcal{I}^c\cap\mathcal{J}_2^c}; \mathbf{0}^{\mathcal{I}}) \tag{36}$$

if $\mathcal{J}_1$ and $\mathcal{J}_2$ are such that $\mathcal{J}_1 \cap \mathcal{I}^c = \mathcal{J}_2 \cap \mathcal{I}^c$, then from (33) we have that the inner sum on the LHS of (35) is zero except when $\mathcal{J}$ is empty, as required.

To prove (10), we proceed similarly and write

$$\begin{aligned} \nu_f(A) &= \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k \nu_f\Big(\prod_{i=1}^{s} [0, b_i + \mathbf{1}_{i\in\mathcal{I}}(a_i - b_i)]\Big) \\ &= \sum_{k=0}^{s} \sum_{\mathcal{I}:|\mathcal{I}|=k} (-1)^k f(\mathbf{a}^{\mathcal{I}}; \mathbf{b}^{\mathcal{I}^c}). \end{aligned} \tag{37}$$

We also have

$$\sum_{\mathcal{I}} \mu_{f,\mathcal{I},\mathbf{0}}(A) = \sum_{\mathcal{I}} \sum_{\mathcal{J}} (-1)^{|\mathcal{J}|} f(\mathbf{a}^{\mathcal{I}\cap\mathcal{J}}; \mathbf{b}^{\mathcal{I}\cap\mathcal{J}^c}; \mathbf{0}^{\mathcal{I}^c}). \tag{38}$$

Here again, we use the fact that for a given non-empty $\mathcal{I}$ and $\mathcal{J}_1$ and $\mathcal{J}_2$ such that $\mathcal{J}_1 \cap \mathcal{I}^c = \mathcal{J}_2 \cap \mathcal{I}^c$, (36) holds and therefore using (33) we get that the RHS of (38) simplifies to

$$\sum_{\mathcal{J}} (-1)^{|\mathcal{J}|} f(\mathbf{a}^{\mathcal{J}}; \mathbf{b}^{\mathcal{J}^c}).$$

Combining this with (37), we get that (10) holds.

Part (ii) follows from the definition of quasi-monotonicity and from (9) and (10). We note that the requirement that quasi-volumes of any dimension d be non-negative in the definition of quasi-monotonicity implies the positivity of the measures $\mu_{f,\mathcal{I},\mathbf{0}}$; if our definition of quasi-monotonicity was restricted to quasi-volumes of dimension s as in [2], this implication would not hold. □

Proof of Lemma 4. For (i), from the properties of a $(0, m, s)$-net, there are exactly b^{m-s} points in each elementary interval $I(r_1, \ldots, r_s)$, and there are $(b-1)^s$ intervals $I(r_1, \ldots, r_s)$ with $1 \le r_j \le b-1$.

For (ii), the initial values given for $M(m, s)$ are proved as follows: to get the expression for $M(1, s)$, first note that the point $(0, \ldots, 0)$ is in P_n and thus there is exactly one point in $I(0, \ldots, 0)$. Also, since $n = b$ and each one-dimensional projection $P_n(\{j\})$ of the net is a $(0, 1, 1)$-net, the $b-1$ remaining points must fall in an interval of the form $I(r_1, \ldots, r_s)$ with $r_\ell \neq 0$ for $\ell = 1, \ldots, s$, as otherwise there would be two of the b points from $P_n(\{j\})$ in a common interval $[0, 1/b)$ for one coordinate $j \in \{1, \ldots, s\}$.

To get the expression for $M(2, s)$, wlog we identify each point of P_n to the first two coordinates (r_1, r_2) of the interval $I(r_1, r_2, r_3, \ldots, r_s)$ to which it belongs. To be counted toward $M(2, s)$ we must have $r_1 \neq 0$, leaving $b-1$ possibilities for r_1. As for r_2, we must have $r_2 \neq 0$. But in addition, we must consider the $s-2$ points that lie in an interval $I(r_1, \ldots, r_s)$ with $r_j = 0$ for some $j > 2$.

These $s-2$ points are distinct, i.e., a given point cannot have more than one $r_j = 0$, as otherwise we would have two points in an elementary interval of the form $[0,1/b)^2$, since the point $(0,\dots,0)$ is already in there. Hence overall, there are $s-2+1$ values of r_2 that must be avoided (the point with $r_2 = 0$ and the $s-2$ points with one $r_j = 0$ for $j > 2$), thus leaving $(b-(s-1))$ possibilities.

To establish the relation for $M(m,s)$ that holds when $s \geq 3$ and $2 \leq m < s$, we must generalize the reasoning used to prove the expression for $M(2,s)$. Namely, each point in P_n is identified with the first m coordinates $r_1,\dots,r_m$ of the interval $I(r_1,r_2,r_3,\dots,r_s)$ to which it belongs. For each of the first $m-1$ integers $r_1,\dots,r_{m-1}$, we have $(b-1)$ possible values to choose from. For r_m, as we did to get $M(2,s)$, we first argue that we must not only remove the choice $r_m = 0$, but also take into account points that are in an elementary interval of the form $I(r_1,\dots,r_m,r_{m+1},\dots,r_s)$ with $r_1,\dots,r_m \neq 0$ and one (or more) $r_\ell = 0$ for $\ell > m$. Since the b points of the $(0,m,s)$-net that are in an elementary interval of the form $\prod_{\ell=1}^{m-1}[\frac{r_\ell}{b},\frac{r_\ell+1}{b})$ for given $r_\ell \neq 0$, $\ell = 1,\dots,m-1$, form a $(0,1,s)$-net (after proper rescaling, see [17, p. 48]), it implies that for each of the $s-m+1$ remaining coordinates, there is exactly one of those b points in $[0,1/b)$. If these points were all distinct, the number of allowable choices for r_m would be $b-(s-m+1)$. However, we must also take into account the fact that some points may have k coordinates among $m, m+1,\dots,s$ belonging to $[0,1/b)$, where $2 \leq k \leq \min(m-1, s-m+1)$. A point with k such coordinates will reduce by $k-1$ the number of uneligible points, i.e., contribute an increase of $k-1$ to $M(m,s)$, and we have $N(k,m,s)$ that many points overall. Finally, the reason why we have $k \leq m-1$ is because for any m-tuple of coordinates $\{i_1,\dots,i_m\}$, we already have the point $(0,\dots,0)$ whose projection over $\{i_1,\dots,i_m\}$ is in $[0,1/b)^m$ and a $(0,m,s)$-net cannot have more than one point in an elementary interval of volume b^{-m}.

To get the recursive relation for $N(k,m,s)$, we first note that for each choice of $(r_1,\dots,r_{m-k})$ with $r_\ell \neq 0, \ell = 1,\dots,m-k$, the b^k points falling in the elementary interval $\prod_{l=1}^{m-k}\left[\frac{r_\ell}{b},\frac{r_\ell+1}{b}\right)$ form a $(0,k,k)$-net over any set of k coordinates $\{i_1,\dots,i_k\}$ with $i_\ell \geq m$. Hence, there is exactly one point in this $(0,k,k)$-net that is in $[0,1/b)^k$ for any subset $\{i_1,\dots,i_k\}$. In addition, we argue that there are $M(m-k,m-1)$ possibilities for choices of $(r_1,\dots,r_{m-k})$ with $r_\ell \neq 0, \ell = 1,\dots,m-k$ that also have $r_\ell \neq 0$ for $m-k < \ell \leq m-1$. For each such choice, we have $\binom{s-m+1}{k}$ points with a k-tuple of zeros if such points are all distinct; but here as well, we must take into account the fact that they may not be. This is why from the term $M(m-k,m-1)\binom{s-m+1}{k}$ we must remove all the cases where a k-tuple of zeros appear in i-tuples of zeros with $i > k$. For a given i, there are $N(i,m,s)\binom{i}{k}$ such cases, where i goes from $k+1$ to m_s^*; the restriction that $i \leq m-1$ is based on the same reasoning as the restriction $k \leq m-1$ explained at the end of the proof of the relation for $M(m,s)$. □

Remark 14. At first sight, the definitions of $M(m,s)$ for $m < s$ and $N(k,m,s)$ used in Lemma 4 may look like they are circular. A closer look reveals that it is not the case. In particular, note that $N(l-1,\ell,k) = M(1,\ell-1)\binom{k-\ell+1}{\ell-1}$ for any ℓ with $3 \leq \ell < k$. So the idea is to first compute $M(1,k)$ and $M(2,k)$. To get the values $M(3,k)$, we first compute $N(2,3,\ell)$ for $4 \leq \ell \leq s$. Similarly, to get the values $M(4,k)$, we first compute $N(3,4,\ell)$ for $5 \leq \ell \leq s$, and then $N(2,4,\ell)$, which requires values of the form $M(2,\cdot)$ and $N(3,4,\cdot)$, which have already been computed. More generally, the values $M(m,\cdot)$ can be obtained by first computing $N(\min(m-1,\cdot-m+1),m,\cdot)$ down to $N(2,m,\cdot)$. To compute $N(m-\ell,m,\cdot)$, we need the values of $N(m-\ell+1,m,\cdot),\dots,N(m-1,m,\cdot)$ and also $M(\ell,\cdot)$ for $\ell \leq m-2$, which have all been previously computed. The following example illustrates how these computations are done.

Example 3. Consider the case where $b = s = 7$. Since $M(m,k)$ is easily computed via item (i) of Lemma 4 for $m \geq k$, here the goal is to compute $M(m,k)$ for $1 \leq m < k \leq 7$. The results are shown in Table 1. We have $M(1,k) = 6$ and $M(2,k) = 6 \times (7-k+1)$.

Next we compute $N(2,3,k)$, using $N(2,3,k) = M(1,2)\binom{k-2}{2}$ and thus getting 6, 18, 36, and 60 respectively for $k = 4,\dots,7$. We can then compute $M(3,k)$ using $M(3,k) = 36(9-k) + N(2,3,k)$.

Next, we compute $N(3,4,6) = M(1,3) = 6$ and $N(3,4,7) = 4M(1,3) = 24$ and then $N(2,4,k)$ for $k = 5,6,7$, first getting $N(2,4,5) = M(2,3) = 30$ and then $N(2,4,k) = M(2,3)\binom{k-3}{2} + N(3,4,k)\binom{3}{2}$

TABLE 1. Values of $M(m,k)$ for $1 \le m < k \le s$ in the case $b = s = 7$.

$m\backslash k$	1	2	3	4	5	6	7
1	6	6	6	6	6	6	6
2			30	24	18	12	6
3				186	172	144	132
4					1110	984	804
5						6666	5718
6							39990

so that $N(2,4,6) = 90 + 3N(3,4,6) = 108$ and $N(2,4,7) = 180 - 3N(3,4,7) = 108$. We can then compute $M(4,k) = 6^3(10-k) + \sum_{i=2}^{\min(3,k-3)}(i-1)N(i,4,k)$.

Similarly, we compute $N(3,5,7) = M(2,4) = 24$, $N(2,5,6) = M(3,4) = 186$, $N(2,5,7) = M(3,4)\binom{3}{2} - N(3,5,7)\binom{3}{2} = 486$ and then $M(5,k) = 6^4(11-k) + \sum_{i=2}^{k-4}(i-1)N(i,5,k)$.

Finally we compute $N(2,6,7) = M(4,5) = 1110$ and $M(6,7) = 5 \times 6^5 + N(2,6,7) = 39,990$.

Proof of Proposition 8. Using the properties of a scrambled net, we have that

$$a_\ell = \frac{N^{j-1}_{\alpha_1,\dots,\alpha_{j-1},\ell}}{N^{j-1}_{\alpha_1,\dots,\alpha_{j-1}}} \frac{1}{\mathrm{Vol}(D_\ell)},$$

where $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1}}$ is the number of pairs of points $(\mathbf{U}_i, \mathbf{U}_{i'})$ from the $(0,m,s)$-net in an elementary interval of the form $I_{\alpha_1,\dots,\alpha_{j-1}}(r_1,\dots,r_{j-1}) = \prod_{k=1}^{j-1}[r_k b^{-\alpha_k}, (r_k+1)b^{-\alpha_k})$ that are such that $\gamma_b(U_{i,k}, U_{i',k}) = \alpha_k$ for $k = 1,\dots,j-1$, while $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1},\ell}$ is the number of pairs of points $(\mathbf{U}_i, \mathbf{U}_{i'})$ in an elementary interval of the form $I_{\alpha_1,\dots,\alpha_{j-1}}(r_1,\dots,r_{j-1})$ that are such that $\gamma_b(U_{i,k}, U_{i',k}) = \alpha_k$ for $k = 1,\dots,j-1$ and $\gamma_b(U_{i,j}, U_{i',j}) = \ell$. Furthermore, as stated in (20), we have that $\mathrm{Vol}(D_\ell) = (b-1)b^{-\ell-1}$.

Next, we observe that the $b^{m-A_{j-1}}$ points from the $(0,m,s)$-net in $I_{\alpha_1,\dots,\alpha_{j-1}}(r_1,\dots,r_{j-1})$ form a scrambled $(0, m-A_{j-1}, s)$-net for any choice of $r_1,\dots,r_{j-1}$ (after rescaling), and thus the conditional pdf under study can be determined by focusing on a specific choice of r_k's. In particular, wlog we can choose $r_k = 0$ for $k = 1,\dots,j$. We then need to establish that the formulas given for $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1}}$ and $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1},\ell}$ are valid in this case. Furthermore, to count the number of pairs $(\mathbf{U}_i, \mathbf{U}_{i'})$ that add up to each of these two quantities, we can fix $\mathbf{U}_i$ and then count how many possibilities there are for $\mathbf{U}_{i'}$. The total number of pairs is then obtained by multiplying this quantity by the number of possible choices for $\mathbf{U}_i$, and dividing by two to not overcount. In particular, we can choose $\mathbf{U}_i$ to be $(0,\dots,0)$.

Starting with $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1}}$, after rescaling of the $b^{m-A_{j-1}}$ points that are in $I_{\alpha_1,\dots,\alpha_{j-1}}(0,\dots,0)$ and considering only their first $j-1$ coordinates, we have a $(0, m-A_{j-1}, j-1)$-net and by definition, exactly $M(m-A_{j-1}, j-1)$ points $\mathbf{U}_{i'}$ are such that $\gamma_b(0, U_{i',k}) = \alpha_k$ for $k = 1,\dots,j-1$. We also note that there are $b^{m-A_{j-1}}$ choices for $\mathbf{U}_i$. So overall the total number of pairs is $M(m-A_{j-1}, j-1)b^{m-A_{j-1}}/2$.

For $N^{j-1}_{\alpha_1,\dots,\alpha_{j-1},\ell}$, here as well we start by rescaling the $b^{m-A_{j-1}-\ell}$ points that are in $I_{\alpha_1,\dots,\alpha_{j-1},\ell}(0,\dots,0)$ and then consider their first j coordinates, thus working with a $(0, m-A_{j-1}-\ell, j)$-net, which has $M(m-A_{j-1}-\ell, j)$ points $\mathbf{U}_{i'}$ such that $\gamma_b(0, U_{i',k}) = \alpha_k$ for $k = 1,\dots,j-1$ and $\gamma_b(0, U_{i',j}) = \ell$. We have $b^{m-A_{j-1}-\ell}$ choices for the first point and then b^ℓ distinct intervals $I_{\alpha_1,\dots,\alpha_{j-1},\ell}(0,\dots,0)$ within $I_{\alpha_1,\dots,\alpha_{j-1}}(0,\dots,0)$, which gives a total number of pairs of

$$M(m-A_{j-1}-\ell, j)b^{m-A_{j-1}-\ell}b^\ell/2 = M(m-A_{j-1}-\ell, j)b^{m-A_{j-1}}/2. \quad \square$$

Acknowledgements. I wish to thank Ruodu Wang and Bin Wang for helpful discussions on this work. In particular, these discussions led to the idea of working directly with the joint density of pairs of points rather than their CDF, which simplified the analysis of the dependence structure of scrambled nets. I also want to thank the Associate Editor and anonymous referee for their helpful comments, which greatly helped improving the paper.

References

[1] Aistleitner C, Dick J (2015) Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arithmetica* 167:143–171.

[2] Beare B (2009) A generalization of Hoeffding's lemma, and a new class of covariance inequalities. *Statistics and Probability Letters* 79:637–642.

[3] Caflisch RE, Morokoff W, Owen AB (1997) Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance* 1(1):27–46.

[4] Dick J, Kuo FY, Sloan IH (2013) High-dimensional integration: the quasi-Monte Carlo way. *Acta Numerica* 22:133–288.

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

[6] Dunford N, Schwartz J (1957) *Linear Operators (Part I)* (New York: Wiley Interscience).

[7] Faure H (1982) Discrépance des suites associées à un système de numération (en dimension s). *Acta Arithmetica* 41:337–351.

[8] Fréchet M (1910) Extension au cas des intégrales multiples d'une définition de l'intégrale due à Stieltjes. *Nouvelles annales de mathématiques, 4è série* 10:241–256.

[9] Hammersley JM, Handscomb DC (1956) A new Monte Carlo technique: Antithetic variates. *Proceedings of the Cambridge Philosophical Society* 52:449–475.

[10] Hickernell FJ (1998) Lattice rules: How well do they measure up? Hellekalek P, Larcher G, eds., *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, 109–166 (New York: Springer).

[11] Hoeffding W (1940) Masstabinvariante korrelations-theorie. *Schriften Math. Inst. Univ. Berlin* 5:181–233.

[12] Lehmann EL (1966) Some concepts of dependence. *Annals of Mathematical Statistics* 37:1137–1153.

[13] Mardia K, Thompson J (1972) Unified treatment of moment-formulae. *Sankhya Series A* 34:121–132.

[14] Matoušěk J (1998) On the L_2-discrepancy for anchored boxes. *Journal of Complexity* 14:527–556.

[15] Mckay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245.

[16] Nelsen R (2006) *An Introduction to Copulas*. Springer Series in Statistics (Springer), second edition.

[17] Niederreiter H (1992) *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics* (Philadelphia: SIAM).

[18] Owen AB (1997) Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis* 34(5):1884–1910.

[19] Owen AB (1998) Scrambling Sobol and Niederreiter-Xing points. *Journal of Complexity* 14:466–489.

[20] Owen AB (2003) Variance and discrepancy with alternative scramblings. *ACM Transactions on Modeling and Computer Simulation* 13:363–378.

[21] Owen AB (2005) Multidimensional variation for quasi-Monte Carlo. Fan J, Li G, eds., *International Conference on Statistics in honour of Professor Kai-Tai Fang's 65th birthday*, 49–74 (Hackensack, NJ: World Scientific Publications).

[22] Prakasa Rao B (1998) Hoeffding identity, multivariance and multicorrelation. *Statistics* 32(1).

[23] Quesada-Molina J (1992) A generalization of an identity of Hoeffding and some applications. *J. Ital. Statist. Soc.* 3:405–411.

[24] Royden H (1968) *Real Analysis* (New York: Macmillan), second edition.

[25] Sobol' IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* 7:86–112.

[26] Sobol' IM (1969) *Multidimensional Quadrature Formulas and Haar Functions* (Moskow: Nauka), in Russian.

[27] Sobol' IM (1993) Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computer Experiments* 1:407–414, published in Russian in 1990.

[28] Tao T (2011) *An Introduction to Measure Theory*, volume 126 of *Graduate Texts in Mathematics* (Springer).

[29] Wang R, Wang B (2015) Extreme negative dependence and risk aggregation. *Journal of Multivariate Analysis* 136:12–25.

[30] Young W (1917) On multiple integration by parts and the second theorem of the mean. *Proc. London Math. Soc., Series 2* 16:273–293.

[31] Zaremba S (1968) Some applications of multidimensional integration by parts. *Ann. Poln. Math.* 21:85–96.