Statistical Science

Living on the Edge: An Unified Approach to Antithetic Sampling

Roberto Casarin, Radu V. Craiu, Lorenzo Frattarolo and Christian P. Robert

Abstract. We identify recurrent ingredients in the antithetic sampling literature leading to a unified sampling framework. We introduce a new class of antithetic schemes that includes the most used antithetic proposals. This perspective enables the derivation of new properties of the sampling schemes: i) optimality in the Kullback–Leibler sense; ii) closed-form multivariate Kendall's τ and Spearman's ρ; iii) ranking in concordance order and iv) a central limit theorem that characterizes stochastic behaviour of Monte Carlo estimators when the sample size tends to infinity. The proposed simulation framework inherits the simplicity of the standard antithetic sampling method, requiring the definition of a set of reference points in the sampling space and the generation of uniform numbers on the segments joining the points. We provide applications to Monte Carlo integration and Markov Chain Monte Carlo Bayesian estimation.

Key words and phrases: Antithetic variables, Countermonotonicity, Monte Carlo, Negative dependence, Variance reduction.

1. INTRODUCTION

The Monte Carlo method is at the core of model-based scientific exploration. In its simplest form, it relies on approximating an integral $\mathfrak{I} = \int f(\mathbf{x})\pi(d\mathbf{x})$ with $\hat{\mathfrak{I}}_d = \frac{1}{d}\sum_{i=1}^{d} f(\mathbf{X}_i)$ when π is a probability measure, F_π is the corresponding cumulative distribution function (CDF), $f : \mathbb{R}^p \mapsto \mathbb{R}$ is a integrable function with respect to π, d is the Monte Carlo sample size and $\mathbf{X}_1, \ldots, \mathbf{X}_d$ are independent, identically distributed (henceforth, iid) samples from π.

In modern computational problems, sampling from the distribution π may be expensive, in terms of either computational effort or time, so techniques needed to reduce the Monte Carlo sample size d, while maintaining the desired precision in estimation, are essential. A relevant class is represented by the *variance reduction techniques* that use statistical properties induced by the sampling design to reduce the variance $\mathbb{V}ar(\hat{\mathfrak{I}}_d)$. For instance, in the case $p = 1$, if the independence condition between samples $X_1, \ldots, X_d$ is dropped then, $\mathbb{V}ar(\hat{\mathfrak{I}}_d)$ becomes

$$\frac{1}{d^2}\sum_{i=1}^{d}\mathbb{V}ar\left(f\left(X_i\right)\right) + \frac{1}{d^2}\sum_{i \neq j}\mathbb{C}ov\left(f\left(X_i\right), f\left(X_j\right)\right), \quad (1)$$

which is reduced, compared to independent sampling, if the average covariance is negative.

Antithetic sampling designs aim at minimizing the covariances between samples while preserving their marginal distribution. A historical perspective on the strategies for antithetic sampling (e.g., Hammersley and Mauldon, 1956; Hammersley and Morton, 1956) allows us to better understand the rationale behind various constructions and to establish useful relationships with the results available from related fields, such as stochastic orders (e.g., Barlow and Proschan, 1975), optimal transport (e.g., Gaffke and Rüschendorf, 1981), Fréchet classes (e.g., Whitt, 1976), and group transformation (e.g., Andréasson, 1972). Our historical review identifies some key recurrent ingredients used to propose a unified framework for antithetic sampling. We introduce a new class of antithetic constructions that also includes some of the, to our knowledge, most used antithetic proposals, which are reviewed later in this section. The sampling schemes in the class are simple and consist in choosing once for all a deterministic set of points in a given dimension and the way those points are joined with segments. Then random vectors from the schemes are obtained by sampling on the segments.

Roberto Casarin is Professor at the Ca' Foscari University of Venice (e-mail: r.casarin@unive.it). Radu V. Craiu is Professor at the University of Toronto, Toronto, Ontario M5G 1Z5 Canada (e-mail: radu.craiu@utoronto.ca) Lorenzo Frattarolo is Researcher at the European Commission Joint Research Centre (e-mail: lorenzo.frattarolo@ec.europa.eu). Christian P. Robert is Professor at CEREMADE, Université Paris-Dauphine PSL and at the University of Warwick (e-mail: xian@ceremade.dauphine.fr).

Moreover, this new perspective enables the derivation of new properties of the sampling schemes for p stochastically independent replications ($p \geq 1$): *i)* optimality in the Kullback–Leibler sense; *ii)* closed-form multivariate Kendall's τ and Spearman's ρ; *iii)* ranking in concordance order, and *iv)* a central limit theorem that characterizes stochastic behavior when d tends to infinity.

The pairwise antithetic coupling introduced by Hammersley and Morton (1956) achieves variance reduction by generating $d/2$ (we assume d is even in (1)) iid pairs of negatively correlated random variables (X_{1i}, X_{2i}), $i = 1, \dots, d/2$. The joint bivariate distribution of (X_{1i}, X_{2i}) for each $i = 1, \dots, d/2$ achieves the lower Fréchet bound,

$$W\left(F_\pi\left(X_{1i}\right), F_\pi\left(X_{2i}\right)\right) = \max\left(F_\pi\left(X_{1i}\right) + F_\pi\left(X_{2i}\right) - 1, 0\right),$$

that represents the point-wise minimal joint cumulative distribution among the class of distributions having F_π as marginal (see Fréchet, 1935). This is achieved by sampling using the quantile coupling:

$$X_{1i} \sim \pi, \quad X_{2i} = F_\pi^{-1}\left(1 - F_\pi\left(X_{1i}\right)\right). \tag{2}$$

This procedure minimizes the correlation for any monotonic f in the case $d = 2$ and $p = 1$. The explanation of this reduction, as correctly pointed out by Whitt (1976), can be found in Hoeffding (1940) and Fréchet (1951), and it is due to the rearrangement inequality (see chapter X of Hardy et al., 1934). Interestingly, the construction cannot be unambiguously extended beyond pairs because the lower Fréchet bound of all d-variate distributions is a distribution only when $d = 2$. In particular, in dimension $d = 2$, for a given F_π, the Fréchet lower bound is the unique element in the set of bivariate distributions that is minimal for most dependence orders, i.e., there is no other element ranking lower than Fréchet bound in those orders. Beyond dimension $d = 2$, minimal elements are not unique. In the following, we focus on the concordance order, establishing a relationship with the variance of functions in one variable, i.e. $p = 1$.

DEFINITION 1.1 (Concordance Order (Joe, 1990)). Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors with CDFs F and G and survival function $\bar{F}$ and $\bar{G}$, respectively. Then $\mathbf{Y}$ is more concordant than $\mathbf{X}$ (written $\mathbf{X} \prec_C \mathbf{Y}$) if

$$F \leq G \text{ and } \bar{F} \leq \bar{G}. \tag{3}$$

As remarked by Joe (1990), for d dimensional vectors, $\mathbf{X} \prec_C \mathbf{Y}$ implies

$$\mathbb{V}ar\left(\sum_{l=1}^{d} b_l f\left(X_l\right)\right) \leq \mathbb{V}ar\left(\sum_{l=1}^{d} b_l f\left(Y_l\right)\right), \tag{4}$$

with f monotonic and any $b_l \geq 0$, $l = 1, \dots, d$. Equation (4) implies that concordance order is an efficiency order for Monte Carlo estimators when considering monotonic functions of dimension $p = 1$.[1] It follows that the best candidates for variance reduction, in the monotone case with $p = 1$ and $d > 2$, are the elements in the set of d-variate distribution with given marginals that are minimal in concordance order. We remark that given other notions of dependence order that imply concordance order, such as supermodular order (Müller and Scarsini, 2000) and correlation order (Lu and Yi, 2004), the set of minimal elements with respect to the other order is contained in the set of minimal elements with respect to concordance order (c.f. Remark 3.1 in Ahn and Fuchs, 2020) .

If we drop the monotonicity assumption about f in (1), the discussion and derivation of lower bounds for the variance are more complex and less general. If $p = 1$ and $d = 2$, for non-monotonic, bounded f, Hammersley and Mauldon (1956) prove that the lower bound of the variance can be attained only by a multivariate transformation of a single standard uniform random variable which, almost surely, is coordinate-wise monotonic. The proof relies on two main ingredients. First, the monotonic transformation introduces an approximate representation of the class of bivariate distributions with uniform marginals. The candidate member of the class is approximated by partitioning the unit square in sub-squares of side $1/n$. The approximation is a doubly stochastic matrix in which each element corresponds to a sub-square and has a value equal to the mass assigned to the corresponding sub-square. This construction relies on the bijective rearrangement (Puccetti and Wang, 2015) also known as measure-preserving transformation (e.g. Brown, 1966; Vitale, 1990). In the interpretation of Vitale (1990), for every random vector (U_1, U_2) on the unit square, with standard uniform marginals, there is a sequence of bijective maps f_n such that $(U_1, f_n(U_1))$ weakly converges to (U_1, U_2). Bijective rearrangements and the induced stochastic dependence (Durante and Sanchez, 2012) are relevant to our discussion of the antithetic constructions for $d > 2$.

The second ingredient is the Birkhoff-Von Neumann's decomposition (Birkhoff, 1946; Von Neumann, 1953) of doubly-stochastic matrices in which they are represented as convex combinations of permutation matrices. Handscomb (1958) extends the latter result to characterize the extremal points of multi-stochastic arrays as higher-dimension permutation arrays and to provide a generalization for $d > 2$ of the results in Hammersley and Mauldon (1956).

Those early results based on discretization give sufficient conditions to characterize the transformations

[1] The statement remains valid for importance sampling estimators. In the Markov Chain Monte Carlo theory, a related result is in Daduna and Szekli (2006) showing the equivalence of concordance order and the South West order (Mira, 2002) of asymptotic variances of Markov chains.

needed to obtain a minimum variance. Unfortunately, the characterizations are not constructive and do not provide feasible random sampling algorithms. Moreover, the existence of the optimal transformation minimizing the variance is not guaranteed.

This led earlier researchers to propose feasible yet sub-optimal sampling solutions, including Hammersley and Morton's (Hammersley and Morton, 1956) proposal for $d > 2$. Andréasson (1972), Andréasson and Dahlquist (1972), and Roach and Wright (1977) follow a group theoretic approach to span the set of antithetic vectors. In particular, Roach and Wright (1977), build on Andréasson (1972), Andréasson and Dahlquist (1972) and Tukey (1957), and draws a parallel with systematic sampling. Roach and Wright (1977) sampling solutions for $d = 2$, in the case of non-monotonic fs, rely on discretization, optimal transport, and use a branch and bound algorithm to explore the group of transformations leading to antithetic vectors. The group theoretic approach of Roach and Wright (1977) was also used in Fishman and Huang (1983) to obtain a reinterpretation of the original Hammersley and Morton (1956) proposal for $d > 2$. Their construction, named rotation sampling, is described next.

EXAMPLE 1 (Rotation sampling).

$$U_1 = U \sim \mathcal{U}[0,1],$$
$$U_l = \left(\frac{l-1}{d} + U\right) \bmod 1, \quad l \in \{2,\dots,d\}, \tag{5}$$

where $\mathcal{U}[0,1]$ denotes the standard uniform distribution.

This proposal is a particular case of our stochastic representation.

The extension to unbounded functions of the theorems in Hammersley and Mauldon (1956) and Handscomb (1958) can be found in Wilson (1979) for $d \geq 2$ and $p = 1$ and in Wilson (1983) for $d \geq 2$, $p \geq 1$. The latter paper combines discretization and bijective rearrangement with the optimal transport assignment problem to prove the results. Bijective rearrangement and Monge-Kantorowitch transportation problem are used in Gaffke and Rüschendorf (1981) to obtain minimum variance constructions for f equal to the identity function. The authors are the first to realize that the Hammersley and Morton (1956) bivariate antithetic vector has an almost sure constant sum, which is one of the main ingredients of our unified approach. Thus they minimize the variance in the case $p = 1$ and f equal to the identity function.

The relationship between constant sum and variance reduction is trivial. Random vectors of dimension $d \geq 2$ with constant sum achieve the smallest variance for the sum of their components. Beyond that, a recent stream of papers (see Ahn and Fuchs, 2020, and references therein) proves that the constant sum vectors are among the minimal vectors with respect to the concordance order. In particular, one possible generalization of the constant sum constraint is the following one.

DEFINITION 1.2 (l-countermonotonic). A d-dimensional random vector $\mathbf{U}$ with uniform marginals, is said to be l-countermonotonic (l-CTM), if there exist some index set $\mathcal{L} \subseteq \mathcal{D}$ with $\mathcal{D} = \{1,\dots,d\}$ and $|\mathcal{L}| = l$, a family $\{g_l\}_{l\in\mathcal{L}}$ of strictly increasing continuous functions $[0,1] \mapsto \mathbb{R}$ and some $k \in \mathbb{R}$ such that:

$$\sum_{l\in\mathcal{L}} g_l(U_l) = k \text{ a.s.} \tag{6}$$

Theorem 2 and Proposition 1 in Lee and Ahn (2014) show that the antithetic vector is the only element of the 2-CTM class and it is minimal in the concordance order.

In addition, conditions for achieving minimality in the concordance order were linked to l-CTM random vectors. Lee, Cheung and Ahn (2017) show that the set of d-CTM vectors is contained in the subset of elements minimal in concordance order. It follows then that d-CTM proposals represent valid candidates for variance reduction, in the monotone case with $p = 1$.

In this paper, we study some of the existing sampling methods and propose new constructions for d-CTM vectors of Uniform(0,1) random variables with a.s. constant sum, that is $g_l(U_l) = U_l$, $l \in \mathcal{D}$ in Eq (6). This subclass is known in the literature as strict d-CTM (Lee and Ahn, 2014). Gaffke and Rüschendorf (1981) recognize that Hammersley and Morton (1956) is strict 2-CTM and provide the first strict 3-CTM construction given below.

EXAMPLE 2 (Gaffke and Rüschendorf (1981) strict 3-CTM).

$$\begin{aligned} U_1 &= U, \qquad U \sim \mathcal{U}[0,1], \\ U_2 &= U + \frac{1}{2}\mathbb{I}_{[0,\frac{1}{2}]}(U) - \frac{1}{2}\mathbb{I}_{[\frac{1}{2},1]}(U), \\ U_3 &= -2U + \mathbb{I}_{[0,\frac{1}{2}]}(U) + 2\mathbb{I}_{[\frac{1}{2},1]}(U). \end{aligned} \tag{7}$$

For the case $d > 3$, the authors propose to generate a sequence of independent random vectors using their representation in (7) and the bivariate antithetic vector of Hammersley and Morton (1956).

EXAMPLE 3 (Gaffke and Rüschendorf (1981) strict d-CTM). Let $\tilde{d} = \lfloor (d-2)/2 \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer smaller than x, and let V_i, $i = 1,\dots,\tilde{d}$ be a sequence of independent random variables. Define

$$U_{2i-1} = V_i, \quad U_{2i} = 1 - V_i, \quad i = 1,\dots,\tilde{d}$$

with

$$U_{d-1} = V_{\tilde{d}+1}, \quad U_d = 1 - V_{\tilde{d}+1}, \tag{8}$$

if d even, and

$$U_{d-2} = V_{\tilde{d}+1},$$

$$U_{d-1} = V_{\tilde{d}+1} + \frac{1}{2}\mathbb{I}_{[0,\frac{1}{2}]}\left(V_{\tilde{d}+1}\right) - \frac{1}{2}\mathbb{I}_{[\frac{1}{2},1]}\left(V_{\tilde{d}+1}\right), \tag{9}$$

$$U_d = -2U + \mathbb{I}_{[0,\frac{1}{2}]}\left(V_{\tilde{d}+1}\right) + 2\mathbb{I}_{[\frac{1}{2},1]}\left(V_{\tilde{d}+1}\right),$$

if d is odd.

Almost contemporaneously, Arvidsen and Johnsson (1982) put forward the apparently different proposal given in the following.

EXAMPLE 4 (Arvidsen and Johnsson (1982) strict d-CTM).

$$\begin{aligned} &U_1 = U, \qquad U \sim \mathcal{U}[0,1], \\ &U_i = \left(2^{i-2}U_1 + 1/2\right) \bmod 1, \quad 2 \le i \le d-1, \\ &U_d = 1 - \left(2^{d-2}U_1\right) \bmod 1. \end{aligned} \tag{10}$$

We will show that for $d = 3$, the two proposals in Examples 2 and 4 coincide and are special cases of our general stochastic representation. Both constructions yield vectors with a constant sum, but Gaffke and Rüschendorf (1981) proposal's use of independent random variates for $d > 3$ made us wonder about its efficiency, especially since the results in Hammersley and Mauldon (1956) and Handscomb (1958) suggest that combinations of independent vectors could be sub-optimal. We compare different strict d-CTM constructions using the concordance order. According to Ahn and Fuchs (2020), all strict d-CTM have minimal multivariate Kendall's τ, but they can have different multivariate Spearman's ρ values. For example, Gaffke and Rüschendorf (1981), and Arvidsen and Johnsson (1982) proposals have the same values for multivariate Kendall's τ, but different ones for Spearman's ρ.

Other examples of strict d-CTM vectors, partially covered by our representation, can be found in Knott and Smith (2006), Lee and Ahn (2014), and after a linear transformation also the construction in Bubenik and Holbrook (2007) and references therein, can be seen as strict d-CTM vectors.

The range of application of CTM constructions has been extended to other marginal distributions. For instance, Rüschendorf and Uckelmann (2002), expands the work in Gaffke and Rüschendorf (1981) to random variables $\{Y_1, \ldots, Y_d\}$ with unimodal distributions using the Levy-Shepp form of the Khinchine representation theorem (Lévy, 1962; Shepp, 1962), as $Y_i = XV_i$ where $V_i \sim U(-1,1)$ for all $1 \le i \le d$. Hence, a CTM construction for $\{V_1, \ldots, V_d\}$ implies constant sum for $\{Y_1, \ldots, Y_d\}$. Trivially, such a vector will achieve the smallest variance for the sum of its components. In those cases, the literature refers to these vectors as complete or joint mix, differentiating between having identical or different marginals (see Puccetti and Wang, 2015, and references therein). Our general construction can be extended to non-uniform marginals following Rüschendorf and Uckelmann (2002).

In Rubinstein and Samorodnitsky (1987), a different extension of Handscomb (1958) theorem was proposed by dropping the bijective condition for the rearrangement. They prove the existence of antithetic solutions that minimize the variance. According to the authors, optimal antithetic solutions should be a function of only one uniform random variable, without restriction on the functional dependence. Unfortunately, dropping the bijective condition results in a tautological statement because, as shown in Brown (1966), Whitt (1976), Vitale (1990) and recently reformulated in theorem 1 of Puccetti and Wang (2015), every random vector can be expressed as a function of only one uniform random variable.

This difficulty of narrowing down conditions for the existence of optimal antithetic variables is linked to the challenge of extending the Birkoff-von Neumann representation to the continuous case. It is, in fact, well known that bijective rearrangements are only a sub-class of the extremal transformations. For example, the $d = 2$ case is known as Birkhoff's problem 111 (Isbell, 1955), and even if there exists a characterization (Lindenstrauss, 1965), the necessary and sufficient conditions in their most recognizable form (Moameni, 2016) are of limited practical relevance. For a discussion and an example of an extremal non-bijective class in the multivariate case, refer to Durante, Fernández Sánchez and Trutschnig (2014). Since a general characterization is out of reach, we solve the optimal transport problem for transformations in the extremal class and produce a stochastic representation that depends on a single standard uniform.

Historically, given the impossibility of obtaining optimal and feasible antithetic plans, by mid 80's the literature shifted the focus to negative dependence. In particular, a procedure considered close to antithetic sampling, but applicable to the general $d \ge 2$, $p \ge 1$ is the Latin Hypercube sampling introduced in McKay, Beckman and Conover (1979).

EXAMPLE 5 (McKay, Beckman and Conover (1979) Latin Hypercube). Given a standard uniform d-dimensional random vector $\mathbf{V}$ and $\mathcal{D}^\sigma = (\sigma(0), \ldots, \sigma(d-1))^T$, a permutation of $\{0, 1, \ldots, d-1\}$ independent of $\mathbf{U}$, set

$$\mathbf{U} = \frac{1}{d}\left(\mathcal{D}^\sigma + \mathbf{V}\right). \tag{11}$$

The simplicity of the method, the guarantee of asymptotic variance reduction (Stein, 1987) and the availability of a central limit theorem (Owen, 1992) made it one of the most common variance reduction strategies. The relationship with antithetic variates was studied in Craiu and Meng (2005), where through the introduction of an iterative version of the method, the Iterated Latin Hypercube (ILH), it is shown that, in the iteration limit, the resulting random vector has an almost-sure constant sum. Our new representation allows comparing ILH and its combination with other antithetic proposals. Finally, we extend the central limit theorem in Owen (1992), showing the irrelevance of the starting distribution when d goes to infinity, and the number of iterations is fixed.

The paper is structured as follows: Section 2 contains the description of the unified representation and its interpretation. Section 3 discusses distributional properties and concordance measures. New and old illustrations of the unified representation and their ranking are presented in Section 4 followed by the derivation of the general central limit theorem for Latin Hypercube in Section 5. Numerical illustrations are presented in Section 6 and the paper ends with a discussion of future directions for research in Section 7.

2. SAMPLING ON LINE SEGMENTS

We introduce a general method for constructing antithetic vectors whose components have a standard uniform, Uniform(0,1), as marginal distribution. Non-uniform variables can be obtained using various transformations, e.g. the inverse CDF method or the Levy-Shepp form of the Khinchine representation theorem (Lévy, 1962; Shepp, 1962). We study conditions for achieving d-CTM and show that several known countermonotonic random vectors used in variance reduction can be obtained as special cases of our general construction.

2.1 Standard Antithetic Construction

Our method relies on sampling with equal probability on a collection $\mathcal{S}$ of line segments in the d-dimensional Euclidean space. Since each segment is uniquely characterized by its endpoints or vertexes, the collection $\mathcal{S}$ can be equivalently represented by the set of vertex pairs that define the segments and their coordinates. This representation is efficient in large dimensions even when the segments share some of their vertexes.

More formally, let us define a vertex set $\mathcal{V} = \{1,\ldots,n\}$ as a set of points in the d-dimensional hypercube, the coordinates of the k-th vertex as the column vector $\mathbf{x}_k \equiv (x_{1k},\ldots,x_{dk})^T \in [0,1]^d$ and the coordinate matrix $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$ as the collection of vertex coordinates. We assume there is an edge $e = (i,j)$ between i and j, with $i < j$, if there is a segment joining the two vertices i and j, and define the collection of segments by the edge set $\mathcal{E} = \{(i,j) \in \mathcal{V}\times\mathcal{V}\}$. Then $\mathcal{G} = \{\mathcal{V},\mathcal{E}\}$ is an undirected graph and $\mathcal{S} = \{\mathcal{G},\mathbf{X}\}$ is the collection of segments. The lexicographic order on vertex indexing induces an order

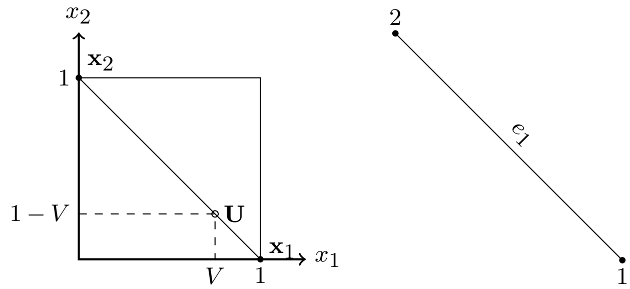


FIG 1. *Left: support set, i.e. the segment joining vertices 1 and 2 of coordinates x_1 and x_2, for the antithetic sampling. Right: dependence graph $\mathcal{G}$ with $\mathcal{V} = \{1,2\}$ and $\mathcal{E} = \{e_1\}$.*

on the edge set, such that the k-th element $e_k \in \mathcal{E}$ is uniquely associated to its couple of vertices, defining the map $\varphi_{\mathcal{E}} : \{1,\ldots,|\mathcal{E}|\} \mapsto \mathcal{E}$, $k \to (i(k), j(k))$.

Our stochastic construction relies on the graph representation $\mathcal{G}$ and requires a properly chosen vertex matrix $\mathbf{X}$ and two independent standard uniform random numbers

1. draw $V \sim \mathcal{U}[0,1]$ and $W \sim \mathcal{U}[0,1]$ independently;
2. choose with uniform probability on the edge set $\mathcal{E}$ the edge e_K by computing $K = \lfloor|\mathcal{E}|W\rfloor + 1$; obtain the random pair of vertices $(I,J) = (i(K), j(K))$ with $(i(K), j(K)) = \varphi_{\mathcal{E}}(K)$;
3. obtain a random point on the segment joining vertices I and J with uniform probability

$$\begin{aligned} U_1 &= x_{1I}V + (1-V)x_{1J}, \\ &\vdots \\ U_d &= x_{dI}V + (1-V)x_{dJ}. \end{aligned} \tag{12}$$

As in antithetic coupling, it is possible to use only one standard uniform number W by setting $V = \{|\mathcal{E}|W\}$. One can show that $\mathbb{P}(\{|\mathcal{E}|W\} \le u|\ \lfloor|\mathcal{E}|W\rfloor + 1 = k) = \mathbb{P}(\{|\mathcal{E}|W\} \le u)$. The following example shows that the standard antithetic method is a special case of our general sampling construction.

EXAMPLE 6. Let us consider $d = 2$ and sampling in the unit square on the diagonal joining the vertex 1 of coordinates $\mathbf{x}_1 \equiv (x_{11} = 1, x_{21} = 0)^T$ and the vertex 2 of coordinates $\mathbf{x}_2 \equiv (x_{12} = 0, x_{22} = 1)^T$ (see Figure 1). Let $V \sim \mathcal{U}[0,1]$ and compute

$$\begin{aligned} U_1 &= x_{11}V + (1-V)x_{12} = V, \\ U_2 &= x_{21}V + (1-V)x_{22} = 1 - V. \end{aligned}$$

Thus, sampling one uniform antithetic couple V and $1-V$ is equivalent to sampling on a segment (see the left plot in Figure 1), and the support of the samples can be summarized by the vertex coordinate matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \tag{13}$$

and the couple of vertices $e_1 = (1,2)$ of the segment we are sampling on (right plot).

The marginal uniformity of the samples on the segment follows from the convex combination representation and the standard uniform assumption for V

$$U_l \sim \mathcal{U}\left[\min\left(x_{l1}, x_{l2}\right), \max\left(x_{l1}, x_{l2}\right)\right], \ l = 1, 2. \tag{14}$$

In addition, the standard uniformity of U_l follows from the assumption $\max\left(x_{l1}, x_{l2}\right) = 1$ and $\min\left(x_{l1}, x_{l2}\right) = 0$ for all $l = 1, 2$, and the d-CTM property from conditions on the vertex coordinates: $x_{11} + x_{21} = x_{12} + x_{22} = 1$.

Example 6 illustrates the fact, which is easy to prove in full generality, that our construction generates samples with uniform probability on $\mathcal{S}$. However, it does not guarantee that all the marginal distributions of the components U_l, for $l \in \mathcal{D}$, are Uniform$(0,1)$. In the following, we study the conditions on $\mathcal{G}$ and $\mathbf{X}$, such that the variables U_l, $l \in \mathcal{D}$ are conditionally uniform, given the choice of the segment, and marginally Uniform$(0,1)$.

2.2 Uniformity

We provide conditions on the collection of segments $\mathcal{S}$ in the d-dimensional hypercube to achieve standard marginal uniformity and d-CTM when using our construction. Before presenting the general result we introduce some notation and discuss the main assumptions.

For the general case, the following condition rules out atomic and mixed probability measures.

ASSUMPTION 1 (Admissibility). The set $\mathcal{S} = \{\mathcal{G}, \mathbf{X}\}$ is admissible if all segments in $\mathcal{S}$ are not contained in any of the $(d-1)$-hyperplanes that are parallel to a $(d-1)$-dimensional hyperface of the unit hypercube $[0,1]^d$.

We provide some intuition for Assumption 1 through the following 2-dimensional example.

EXAMPLE 7. Consider the collection of segments in the left plot of Figure 2 with coordinate matrix

$$\mathbf{X} = \begin{pmatrix} \alpha\ \beta\ \gamma\ \alpha \\ \beta\ \alpha\ \beta\ \alpha \end{pmatrix}, \tag{15}$$

where $\alpha < \beta \leq \gamma \in \mathbb{R}$. According to the lexicographic order map $k \to (i(k), j(k))$ the edge set is: $\mathcal{E}^* = \{e_1^* = (1,2),\ e_2^* = (1,3),\ e_3^* = (1,4),\ e_4^* = (2,3),\ e_5^* = (2,4),\ e_6^* = (3,4)\}$ (see right plot). Since the sampling method can concentrate the probability mass at some points or along some directions of the hypercube we need to impose some admissibility conditions to have non-degenerate distributions. The first set of conditions excludes degenerate segments, which are segments with equal end-points. Thus, we rule out self-loops from the graph, i.e. edges from one vertex to itself. The second set

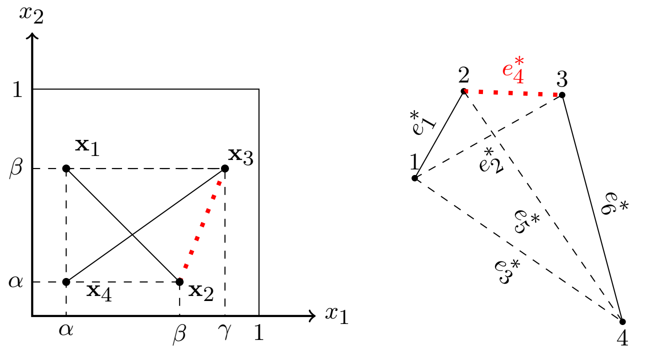


FIG 2. *Left: support set of the sampling scheme. Right: admissible (solid) and not-admissible (dashed) edges. Edge labeling follows the vertex lexicographic order.*

of admissibility conditions excludes edges where the distribution concentrates along some coordinates. The edge e_2^* joining vertex 1 to 3 is not admissible since:

$$U_1 = x_{11}V + x_{13}(1-V) = \gamma - (\gamma - \alpha)V,$$
$$U_2 = x_{21}V + x_{23}(1-V) = \beta,$$

and U_2 is almost surely constant conditionally on being on that edge. A similar remark applies to the edges e_3^* and e_5^*, whereas e_4^* (red dashed in Figure 2) is only admissible if $\beta < \gamma$.

In summary, the admissible edge set is:

$$\mathcal{E} = \begin{cases} \{e_1 = e_1^* = (1,2), e_2 = e_4^* = (2,3), \\ \qquad e_3 = e_6^* = (3,4)\}, & \text{if } \beta < \gamma, \\ \\ \{e_1 = e_1^* = (1,2), e_2 = e_6^* = (3,4)\}, & \text{if } \beta = \gamma. \end{cases}$$

In our construction, conditional uniformity is a necessary condition for standard marginal uniformity and the admissibility assumption implies conditional uniformity. To enhance the paper's readability, all proofs are deferred to the Appendix.

LEMMA 1 (Conditional Uniformity). Let $\mathcal{S}$ satisfy Assumption 1. Conditionally on being on the k-th segment of edge $e_k = (i(k), j(k))$, for each $l \in \mathcal{D}$, the random variable U_l in (12) is uniform on $[\alpha_{l,k}, \beta_{l,k}]$ with $\alpha_{l,k} = \min\left(x_{l,i(k)}, x_{l,j(k)}\right)$ and $\beta_{l,k} = \max\left(x_{l,i(k)}, x_{l,j(k)}\right)$.

Another requirement for our construction is that sampling points are in the unit hypercube which is guaranteed by the following range condition.

ASSUMPTION 2 (Range). The range requirement is satisfied if $\max\{x_{lk}, k = 1, \ldots, n\} = 1$ and $\min\{x_{lk}, k = 1, \ldots, n\} = 0$.

Since U_l is a convex combination of $x_{l,i}$, $i \in \{1,\dots,n\}$, Assumption 2 is needed in light of the requirement that $U_l \in [0,1]$ for each $l \in \mathcal{D}$.

In order to introduce the third assumption we need some notation. The set of admissible edges depends on the partitions induced by the distinct elements, sorted in ascending order, in the rows of $\mathbf{X}$. For each row $(x_{l,1},\dots,x_{l,n})$, of $\mathbf{X}$ with $l = 1,\dots,d$ we define $\mathbf{a}_l = (a_{l,1}, a_{l,2},\dots,a_{l,n_l-1}, a_{l,n_l})$ the sequence of $n_l \le n$ sorted distinct elements. The unique values define a partition, say $\{A_{l,m}\}_{m=2}^{n_l}$ of the unit interval with elements:

$$A_{l,m} = [a_{l,m-1}, a_{l,m}), \quad m \in \{2,\dots,n_l\}. \tag{16}$$

For each unique value a_{lk}, the position set $\mathcal{M}_{l,k} = \{i \in \{1,\dots,n\} : x_{l,i} = a_{l,k}\}$, $k = 1,\dots,n_l$ denotes the set of points that provides the unique projected values of $\mathbf{X}$'s columns into the l-th dimension. For each row, $\mathcal{M}_{l,k}$ must therefore satisfy:

$$\mathcal{M}_{l,m} \cap \mathcal{M}_{l,m'} = \emptyset, \quad \bigcup_{m=1}^{n_l} \mathcal{M}_{l,m} = \{1,\dots,n\},$$

and one can represent the coordinates of the vertex $\mathbf{x}_k$ by using the sets of positions of the unique values:

$$x_{l,k} = \sum_{m=1}^{n_l} a_{l,m} \mathbb{I}_{\mathcal{M}_{l,m}}(k), \, l \in \mathcal{D},$$

which will be used in the following to write the CM constraint.

For illustration purposes, let us focus on the first coordinate of Example 7 in the case $\beta < \gamma$. The sequence of $n_l = 3$ sorted distinct elements of the first row is $\mathbf{a}_1 = (a_{1,1} = \alpha, a_{1,2} = \beta, a_{1,3} = \gamma)$ and the sets of positions of the unique values are $\mathcal{M}_{1,1} = \{1,4\}$, $\mathcal{M}_{1,2} = \{2\}$, $\mathcal{M}_{1,3} = \{3\}$ which is a partition of $\{1,2,3,4\}$.

We denote with $\mathcal{G}_l \equiv \{\mathcal{V}_l, \mathcal{E}_l\}$ the projection of $\mathcal{G}$ on the l-coordinate induced by the unique values $\mathbf{a}_l$. $\mathcal{G}_l$ is the graph obtained by assigning a node to each of the n_l components of $\mathbf{a}_l$ and defining the edge set $\mathcal{E}_l = \{e_{l,1},\dots,e_{l,|\mathcal{E}|}\}$ through the map $\{1,\dots,|\mathcal{E}|\} \mapsto \mathcal{E}_l$, $k \in \{1,\dots,|\mathcal{E}|\} \to e_{l,k} = (m(k), m'(k)) \in \mathcal{E}_l$ where

$$m(k) = \{m \in \{1,\dots,n_l\} : i(k) \in \mathcal{M}_{l,m}\},$$
$$m'(k) = \left\{m' \in \{1,\dots,n_l\} : j(k) \in \mathcal{M}_{l,m'}\right\}.$$

We call $n^l_{(m,m')}$ the edges multiplicity between the projected nodes m and m', by convention if (m,m') is not in $\mathcal{E}_l$ then $n^l_{(m,m')} = 0$.

REMARK 1. In the proposed notation, Assumption 1 corresponds to $m(k) \ne m'(k)$ when $e_{l,k} = (m(k), m'(k))$ for all $k = 1,\dots,|\mathcal{E}|$, $l = 1,\dots,d$.

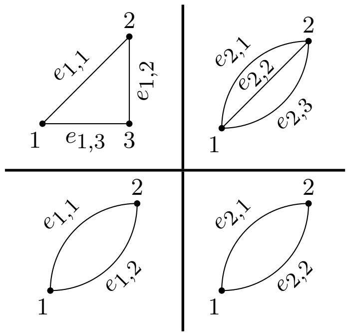


FIG 3. *Graph projection for the graph in Example 7 in the case $\beta < \gamma$ (top) and $\beta = \gamma$ (bottom), considering the first (left) or the second (right) coordinate.*

In Figure 3 we report the projection of the graph of Figure 2 onto the l-th coordinates given in Equation (15) of Example 7.

Set $\alpha_{l,k} = a_{l,m_\alpha(k)}$ and $\beta_{l,k} = a_{l,m_\beta(k)}$, where $m_\alpha(k)$ and $m_\beta(k)$ are the minimum and the maximum value between $m(k)$ and $m'(k)$, and define

$$\mathcal{K}_{l,m} \equiv \{k \in \{1,\dots,|\mathcal{E}|\} : m_\alpha(k) + 1 \le m \le m_\beta(k)\}.$$

We state now the condition on the coordinates of the sampling construction for marginal standard uniformity.

ASSUMPTION 3 (Coordinate). The following set of $n_l - 2$ equations in the variables $a_{l,m}$, $m = 2,\dots,n_l-1$ are satisfied

$$F_{l,m}(\mathbf{a}_l) = \frac{1}{|\mathcal{E}|} \sum_{k \in \mathcal{K}_{l,m}} \frac{1}{a_{l,m_\beta(k)} - a_{l,m_\alpha(k)}} - 1 = 0 \tag{17}$$

$m = 2,\dots,n_l-1$ with $a_{l,1} = 0$ and $a_{l,n_l} = 1$.

The following example clarifies the relationship between Assumptions 2-3 and standard uniformity.

EXAMPLE 8. We discuss separately the two cases: $\beta = \gamma$ and $\beta < \gamma$ since they correspond to two different admissible edge sets. If $\beta = \gamma$ the first component of $\mathbf{U}$ in the stochastic construction of Equation (12) is:

$$U_1 = \begin{cases} \alpha + (\beta - \alpha) V, & \text{if } K = 1 \\ \beta - (\beta - \alpha) V, & \text{if } K = 2. \end{cases}$$

Since we are sampling uniformly on the edge set $\mathcal{E}$ it follows that:

$$\mathbb{P}(K=1) = \mathbb{P}(K=2) = \frac{1}{|\mathcal{E}|} = \frac{1}{2}$$

and U_1 has the following marginal probability density function (PDF)

$$\begin{aligned} f(u_1) &= \frac{1}{2}\frac{1}{\beta-\alpha}\mathbb{I}_{[\alpha,\beta]}(u_1) + \frac{1}{2}\frac{1}{\beta-\alpha}\mathbb{I}_{[\alpha,\beta]}(u_1) \\ &= \frac{1}{\beta-\alpha}\mathbb{I}_{[\alpha,\beta]}(u_1). \end{aligned} \tag{18}$$

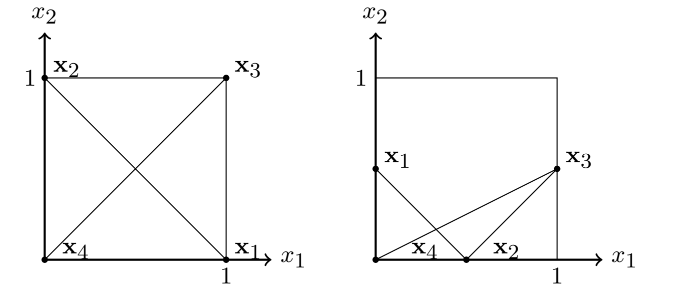


FIG 4. *Sampling on line segments. The support set of the 2-dimensional example with 4 vertexes. Cases: $\beta = \gamma$ (left) and $\beta < \gamma$ (right).*

We get a standard uniform random variable if $\alpha = 0$ and $\beta = 1$. By the same argument, U_2 is a standard uniform random variable. In conclusion, by sampling on the two diagonals of the unit square with a mixture of an antithetic couple $(V, 1 - V)$ and a comonotonic couple (V, V), the method can attain marginal standard uniformity (left plot in Figure 4).

If $\beta < \gamma$ the PDF of U_1 is:

$$f(u_1) = \frac{1}{3(\beta - \alpha)} \mathbb{I}_{[\alpha,\beta]}(u_1) + \frac{1}{3(\gamma - \beta)} \mathbb{I}_{[\beta,\gamma]}(u_1) + \frac{1}{3(\gamma - \alpha)} \mathbb{I}_{[\alpha,\gamma]}(u_1). \tag{19}$$

A necessary and sufficient condition for $u_1 \in [0,1]$ is $\alpha = \min\{x_{1k}, k = 1, \ldots, 4\} = 0$ and $\gamma = \max\{x_{1k}, k = 1, \ldots, 4\} = 1$ which implies the PDF is piece-wise constant on the partition $[0, \beta) \subset [0,1]$ and $[\beta, 1] \subset [0,1]$, induced by the unique values in the first row of $\mathbf{X}$, and is null on $[0, \beta) \cap [\beta, 1]$. For $\beta = 1/2$ the PDF is constant over the elements of the partition, i.e.:

$$\begin{cases} \frac{1}{\beta} + 1 = 3, & \text{if } u_1 \in [0, \beta), \\ \frac{1}{1-\beta} + 1 = 3, & \text{if } u_1 \in [\beta, 1] \end{cases} \tag{20}$$

which implies $U_1 \sim \mathcal{U}[0,1]$ and $U_2 \sim \mathcal{U}[0,1/2]$. Thus, for $\beta < \gamma$ our construction is not standard uniform along all coordinates of the vector. The right plot of Figure 4 shows that the range of U_1 is $[0,1]$ (horizontal axis) and the range of U_2 is $[0,1/2]$ (vertical axis). Similar arguments can be applied to show that imposing standard uniformity for U_2 requires $\min\{x_{1k}, k = 1, \ldots, 4\} = \alpha = 0$ and $\max\{x_{1k}, k = 1, \ldots, 4\} = \beta = 1$ which is not satisfied since $\beta < \gamma$.

We are ready to state the main result of this section which guarantees the marginal standard uniformity of our construction. In addition, we provide a method to find $\mathbf{a}_l$ satisfying the condition in Assumption 3.

THEOREM 1 (Marginal Standard Uniformity). Under Assumptions 1-3, each coordinate of the random vector $\mathbf{U} = (U_1, \ldots, U_d)$ in the stochastic representation (12) has a Uniform$(0,1)$ marginal distribution.

THEOREM 2. The set of equalities in (17) are satisfied iff $\mathbf{a}_l$ is a solution of the following convex minimization problem

$$\min_{\mathbf{a}_l \in [0,1]^{n_l}} \Psi_l(\mathbf{a}_l) \tag{21}$$

with constraints $a_{l,1} = 0$ and $a_{l,n_l} = 1$, where

$$\Psi_l(\mathbf{a}_l) = -\frac{1}{2|\mathcal{E}|} \sum_{m,m'=1}^{n_l} n^l_{(m,m')} \log |a_{l,m'} - a_{l,m}|.$$

Since the optimization problem is convex, if a solution exists it is a global minimum. In addition, since $\Psi_l(\mathbf{a}_l)$ is a sum of lower semi-continuous functions, it is lower semi-continuous. This, together with the compactness of the unit hypercube, guarantees the existence of a solution, by the lower version of the Weierstrass theorem (see for example Theorem 2.43 in Aliprantis and Border (2007)).

In the next theorem, we show that our construction satisfies an optimality criterion involving the Kullback–Leibler (KL) divergence from the uniform distribution. We remind the reader that the KL divergence of the probability measure $\mathbb{P}$ with respect to the probability measure $\mathbb{Q}$ is

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\right].$$

In our case, $\mathbb{P}$ is the joint measure of U_l and K given by the stochastic representation (12):

$$d\mathbb{P}(u_l, k) = \frac{1}{|\mathcal{E}|} f(u_l|\, K = k)\, du_l,$$

and $\mathbb{Q}$ is the joint uniform independent measure in U_l and K:

$$d\mathbb{Q}(u_l, k) = \frac{1}{|\mathcal{E}|} \mathbb{I}_{[0,1]}(u_l)\, du_l. \tag{22}$$

Let us denote with $\mathcal{P}_{d,n,\mathcal{E}}$ the class of measures with stochastic representation (12) and n nodes in the unit hypercube of dimension d connected by the edges in the set $\mathcal{E}$.

THEOREM 3. The minimization problem in (21) is equivalent to:

$$\min_{\mathbb{P} \in \mathcal{P}_{d,n,\mathcal{E}}} D_{KL}(\mathbb{P}||\mathbb{Q}) \tag{23}$$

with $\mathbb{Q}$ the joint uniform in Eq. 22.

Finally, we note that Assumptions 1-3 refer only to properties of the marginal distribution. Consequently, our marginal uniformity result holds also for the following generalization of our construction.

Given a d-dimensional $\mathbf{V} \sim F$ with $V_l \sim \mathcal{U}[0,1]$ for $l = 1,\ldots,d$ and the stochastic representation

$$U_1 = x_{1I}V_1 + x_{1J}(1 - V_1),$$
$$\vdots \tag{24}$$
$$U_d = x_{dI}V_d + x_{dJ}(1 - V_d),$$

we have the following corollary to Theorem 1.

COROLLARY 1. Under Assumptions 1-3 the l-th component of the random vector $\mathbf{U} = (U_1,\ldots,U_d)$ in the stochastic representation (24) is a standard uniform random variable.

The results in Theorems 1-3 allow us to find the l-th row $\mathbf{a}_l$, of the coordinates matrix $\mathbf{X}$ such that U_l in representation (12) is standard uniform. Thus, we have d minimization problems each with a different number of variables, n_l, where the dependence structure in $\mathbf{U}$ is given by the graph $\mathcal{G}$ of the segments in $\mathcal{S}$ and marginally encoded by $\mathbf{a}_l$ and projected graphs $\mathcal{G}_l$.

DEFINITION 2.1. Let $\mathcal{S} = (\mathcal{G}, \mathbf{X})$ be the segment set with $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The convex minimization problem

$$\min_{\{\mathbf{a}_l, l=1,\ldots,d\} \in [0,1]^{n_1+\cdots+n_d}} \sum_{l=1}^{d} \Psi_l(\mathbf{a}_l)$$

is called standard uniform on $\mathcal{S}$ problem.

In the previous optimization problem, as well as in those introduced later on in this section, the vertex set $\mathcal{V}$ and the edge set $\mathcal{E}$ are held fixed. We are optimizing on the position of the vertexes, i.e. in the coordinate matrix $\mathbf{X}$ (or equivalently on the collection of vectors $\mathbf{a}_l$, $l = 1,\ldots,d$). In the following sections, we will show how to rank d-dimensional construction based on a different number of vertexes n and different $\mathcal{E}$ by their amount of negative dependence, using concordance measures.

2.3 Strict Countermonotonicity on Segments

Limiting the study to strict d-CTM leads to an unique value for the constant k in (6):

$$k = \mathbb{E}\left[\sum_{j=1}^{d} U_j\right] = \sum_{j=1}^{d} \mathbb{E}[U_j] = \frac{d}{2}.$$

The constant sum condition can be written as a linear restriction on the coordinates of the vertices $\mathbf{x}_k$, that is

$$\sum_{l=1}^{d} U_l = \sum_{l=1}^{d} x_{lJ} + V\left[\sum_{l=1}^{d} x_{lI} - \sum_{l=1}^{d} x_{lJ}\right] = \frac{d}{2},$$

and since the previous relationship should be valid for all V and (I,J) (i.e. for all W in our setting) we obtain the condition that all vertices should be in the hyperplane of constant sum, i.e.

$$\sum_{l=1}^{d} x_{lk} = \sum_{l=1}^{d}\sum_{m=1}^{n_l} a_{l,m}\mathbb{I}_{\mathcal{M}_{l,m}}(k) = \frac{d}{2} \quad k = 1,\ldots,n. \tag{25}$$

The convexity of the minimization problem in Equation (2.1) is not altered by the inclusion of a linear constraint, nevertheless the constraint couples the coordinates and yields the following non-separable optimization problem in $n_d = \sum_{l=1}^{d} n_l$ variables.

DEFINITION 2.2. Let $\mathcal{S} = (\mathcal{G}, \mathbf{X})$ be the segment set with $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The convex minimization problem

$$\min_{\{\mathbf{a}_l, l=1,\ldots,d\} \in [0,1]^{n_1+\cdots+n_d}} \sum_{l=1}^{d} \Psi_l(\mathbf{a}_l)$$

subject to

$$\sum_{l=1}^{d}\sum_{m=1}^{n_l} a_{l,m}\mathbb{I}_{\mathcal{M}_{l,m}}(k) = \frac{d}{2}, \quad k = 1,\ldots,n$$

will be referred as the strict d-CTM on $\mathcal{S}$ problem.

Since all the constraints are affine the problem in the above definition represents an ordinary convex problem in the terminology of Rockafellar (1970) and can be solved using the method of Lagrange multipliers. Local minima are also global if they exist. Finally, the existence of a solution is guaranteed by lower semi-continuity of $\sum_{l=1}^{d} \Psi_l(\mathbf{a}_l)$ and the fact that the intersection of the unit hypercube with the hyperplane of constant sum is compact.

3. DISTRIBUTIONAL PROPERTIES

3.1 Distribution

The joint CDF of the stochastic representations of $\mathbf{U}$ in Equations (12) and (24) can be written using the distributions of the reflections of $\mathbf{V}$. Let I_d denote the identity matrix and $\mathbf{e}_l$ its l-th column. Reflections are defined in the following.

DEFINITION 3.1 (Reflections). Let $\mathbf{U} \in [0,1]^d$ be a random vector. The transformation $\mathbf{U} \mapsto \mathbf{W} = R_{l,\frac{1}{2}}(\mathbf{U})$ with $R_{l,\frac{1}{2}}(\mathbf{U}) = \left[I_d - 2\left(\mathbf{e}_l\mathbf{e}_l^T\right)\right]\mathbf{U} + \mathbf{e}_l$, defines a reflection of $\mathbf{U}$ with respect to the hyperplane defined by the l-th coordinate equal to 1/2. Given a index subset $\mathcal{L} \subseteq \mathcal{D}$, the sequential reflection transformation $R_{\mathcal{L},\frac{1}{2}}(\mathbf{U}) = [I_d - 2(\sum_{l\in\mathcal{L}} \mathbf{e}_l\mathbf{e}_l^T)]\mathbf{U} + \sum_{l\in\mathcal{L}} \mathbf{e}_l$ is the transformation obtained by reflecting $\mathbf{U}$ sequentially using $R_{l,\frac{1}{2}}$ for $l \in \mathcal{L}$ Then $R_{\mathcal{D},\frac{1}{2}}(\mathbf{U}) = (I_d - 2I_d)\mathbf{U} + \mathbf{1}_d = (1 - U_1,\ldots,1 - U_l,\ldots,1 - U_d)$ is the central inversion through the center of the unit hypercube.

Given a index subset $\mathcal{L}$, we denote the distribution of the reflection $R_{\mathcal{L},\frac{1}{2}}(\mathbf{V})$ with:

$$F_{\mathbf{V},\mathcal{L}}(\mathbf{u}) = \mathbb{P}\left(R_{\mathcal{L},\frac{1}{2}}(\mathbf{V}) \leq \mathbf{u}\right).$$

The marginal distribution of U_l conditionally on living on an edge e_k is $\mathcal{U}[\alpha_{l,k}, \beta_{l,k}]$ with CDF

$$F_{U_l|K}(u_l;k) = \frac{\max\{\alpha_{l,k}, \min\{\beta_{l,k}, u_l\}\} - \alpha_{l,k}}{\beta_{l,k} - \alpha_{l,k}}, \tag{26}$$

where $\alpha_{l,k}$, $\beta_{l,k}$ are defined in Lemma 1. For each $k = 1,\ldots,n$ define the sets:

$$\mathcal{L}_k^+ = \left\{l \in \{1,\ldots,d\} : x_{li(k)} - x_{lj(k)} \geq 0\right\},$$
$$\mathcal{L}_k^- = \left\{l \in \{1,\ldots,d\} : x_{li(k)} - x_{lj(k)} < 0\right\}.$$

THEOREM 4. The random vector $\mathbf{U} = (U_1,\ldots,U_d)$ in representation (24), conditional on $K = k$, has cumulative distribution function $F_{\mathbf{U}|K}(u_1,\ldots,u_d;k)$ given by

$$\mathbb{P}(U_1 \leq u_1,\ldots,U_d \leq u_d | K = k) = F_{\mathbf{V},\mathcal{L}_k^-}(\mathbf{v}_k)$$

where $\mathbf{v}_k = (v_{1,k},\ldots,v_{d,k})$ with $v_{l,k} = F_{U_l|K}(u_l;k)$.

The following result comes from summing over all possible values of K.

COROLLARY 2. The random vector $(U_1,\ldots,U_d)$ in representation (24), has distribution $F_{\mathbf{U}}(u_1,\ldots,u_d)$ given by

$$\mathbb{P}(U_1 \leq u_1,\ldots,U_d \leq u_d) = \frac{1}{|\mathcal{E}|}\sum_{k=1}^{|\mathcal{E}|} F_{\mathbf{V},\mathcal{L}_k^-}(\mathbf{v}_k).$$

For random vectors in representation (12) we are able to derive a closed-form expression of the conditional distribution of $\mathbf{U}$.

COROLLARY 3. $\mathbf{U} = (U_1,\ldots,U_d)$ in representation (12) when conditioning on $K = k$ has cumulative distribution function

$$F_{\mathbf{U}|K}(u_1,\ldots,u_d;k) = \max\left(v_k^+ + v_k^- - 1, 0\right),$$

where $v_k^+ = \min\{v_{l,k}, l \in \mathcal{L}_k^+\}$, $v_k^- = \min\{v_{l,k}, l \in \mathcal{L}_k^-\}$ and $v_{l,k} = F_{U_l|K}(u_l;k)$. The pairs of variables U_l and $U_{l'}$ $l \neq l'$ have cumulative distribution function $F_{U_l,U_{l'}|K}(u_l, u_{l'};k)$ whose form depends on l, l' as follows:

$$\begin{cases} \min(v_{l,k}, v_{l',k}), & \text{if } l,l' \in \mathcal{L}_k^+, \\ \min(v_{l,k}, v_{l',k}), & \text{if } l,l' \in \mathcal{L}_k^-, \\ \max(v_{l,k} + v_{l',k} - 1, 0), & \text{if } l \in \mathcal{L}_k^-, l' \in \mathcal{L}_k^+, \\ \max(v_{l,k} + v_{l',k} - 1, 0), & \text{if } l \in \mathcal{L}_k^+, l' \in \mathcal{L}_k^-. \end{cases} \tag{27}$$

From Corollary 3 and following the definition of Fréchet bound (Fréchet, 1951), the elements U_l and $U_{l'}$ of the stochastic representation (12) are monotonic in the same direction if $l, l' \in \mathcal{L}_k^+$ or $l, l' \in \mathcal{L}_k^-$ and antithetic if $l \in \mathcal{L}_k^-, l' \in \mathcal{L}_k^+$ or $l \in \mathcal{L}_k^+, l' \in \mathcal{L}_k^-$, conditionally on living on the k-th segment.

3.2 Multivariate Kendall's τ and Spearman's ρ

Multivariate Kendall's τ and Spearman's ρ are multivariate measures of concordance introduced in Joe (1990) as a generalization of the well-known bivariate measures. As in the bivariate case, they are invariant to monotonic transformations and increase with concordance order. They attain the maximal value of 1 in the extreme positive dependence case of the comonotonic coupling. They are zero in the independence case and negative in the case of negative dependence. Contrary to the bivariate case in which the minimum value of -1 is attained by the Fréchet lower bound, in the multivariate case, the extreme negative value generally depends on d and does not reach the value -1.

In particular, Fuchs, McCord and Schmidt (2018); Ahn and Fuchs (2020) show that d-CTM vectors have the same minimal multivariate Kendall's τ but different Spearman's ρ. Then, Spearman's ρ can be used to rank d-CTM vectors in concordance order.

Let $\mathbf{U}$ and $\mathbf{W}$ be independent random vectors with the same distribution $F_{\mathbf{U}} = F_{\mathbf{W}} = G$. The multivariate Kendall's τ is defined as:

$$\begin{aligned}\tau(F_{\mathbf{U}}) &= \frac{2^d}{2^{d-1}-1}\left[\int_{[0,1]^d} F_{\mathbf{U}}(\mathbf{u})\, dF_{\mathbf{U}}(\mathbf{u}) - \frac{1}{2^d}\right] \\ &= \frac{2^d}{2^{d-1}-1}\left[\mathbb{E}\left[G(\mathbf{W})\right] - \frac{1}{2^d}\right] \\ &= \frac{2^d}{2^{d-1}-1}\left[\mathbb{P}(\mathbf{U} \leq \mathbf{W}) - \frac{1}{2^d}\right].\end{aligned}$$

Fuchs, McCord and Schmidt (2018) show that if $\mathbf{U}$ and $\mathbf{W}$ are d-CTM then $\tau(F_{\mathbf{U}})$ attains its minimal value

$$\tau(F_{\mathbf{U}}) = \tau_{\min} = -\frac{1}{2^{d-1}-1}. \tag{28}$$

We provide an analytical expression of the Kendall's τ for vectors with the stochastic representation of Equation 24.

PROPOSITION 1. Let $\mathcal{S} = (\mathcal{G}, \mathbf{X})$ be the segment set with $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. Let $\mathbf{U}$ and $\mathbf{W}$ be two independent copies of a random vector with the generalized line segments representation (24) on $\mathcal{S}$, then:

$$\tau(F_{\mathbf{U}}) = \frac{2^d}{2^{d-1}-1}\left[\frac{1}{|\mathcal{E}|}\sum_{k_{\mathbf{U}}=1}^{|\mathcal{E}|}\mathbb{P}\left(R_{\mathbf{V},\mathcal{L}_{k_{\mathbf{U}}}^-}(\mathbf{V}) \leq \mathbf{Y}_{k_{\mathbf{U}}}\right) - \frac{1}{2^d}\right], \tag{29}$$

with $\mathbf{Y}_{k_{\mathbf{U}}} = \left(F_{U_1|K_{\mathbf{U}}}(W_1;k_{\mathbf{U}}),\ldots,F_{U_d|K_{\mathbf{U}}}(W_d;k_{\mathbf{U}})\right)$.

We use the definition of Spearman's ρ in Joe (1990)[2] for distributions with standard uniform marginals. Let $F(\mathbf{U})$ be the cumulative distribution function of the random vector $\mathbf{U}$; then the multivariate Spearman's ρ is

$$\begin{aligned}\rho(F_{\mathbf{U}}) &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\int_{[0,1]^d} F_{\mathbf{U}}(\mathbf{u})\,d\mathbf{u} - \frac{1}{2^d}\right)\\ &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\int_{[0,1]^d} \prod_{l=1}^{d} u_L dF_{\mathbf{U}}(\mathbf{u}) - \frac{1}{2^d}\right)\\ &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\mathbb{E}\left[\prod_{l=1}^{d} U_l\right] - \frac{1}{2^d}\right).\end{aligned}$$

The attainable lower bound $\rho_{\min}$ can be computed using the lower bound for $\mathbb{E}\left[\prod_{l=1}^{d} U_l\right]$ given in Corollary 4.1 of Wang and Wang (2011). We report the values in Table 1.

TABLE 1
Minimum values of multivariate Spearman's ρ, using the lower bound in Corollary 4.1 of Wang and Wang (2011).

d	2	3	4	5	10	20	50	100
$\rho_{\min}$	-1	-0.56	-0.32	-0.18	-0.01	$-1.99\cdot10^{-5}$	$-4.53\cdot10^{-14}$	$-7.97\cdot10^{-29}$

The following proposition provides the Spearman's ρ of the segment construction $\mathcal{S}$ and some sufficient conditions for the reduction of the Spearman's ρ.

PROPOSITION 2. Let the random vector $\mathbf{U}$ satisfy the generalized representation (24). The Spearman's ρ is:

$$\rho(F_{\mathbf{U}}) = \sum_{m=0}^{d} \sum_{\substack{\mathcal{L}_m \subseteq \mathcal{D}\\ |\mathcal{L}_m|=m}} \xi_{\mathcal{L}_m}\rho\left(F_{\mathbf{V},\mathcal{D}\setminus\mathcal{L}_m}\right) + \frac{2^d(d+1)}{2^d-(d+1)}\frac{1}{2^d}(\xi^*-1)$$

with

$$\xi_{\mathcal{L}_m} = \frac{1}{|\mathcal{E}|}\sum_{k=1}^{|\mathcal{E}|}\left(\prod_{l\in\mathcal{L}_m} x_{l,i(k)} \prod_{l\in\mathcal{D}\setminus\mathcal{L}_m} x_{l,j(k)}\right),$$

$$\xi^* = \sum_{m=0}^{d} \sum_{\substack{\mathcal{L}_m \subseteq \mathcal{D}\\ |\mathcal{L}_m|=m}} \xi_{\mathcal{L}_m} = \frac{1}{|\mathcal{E}|}\sum_{k=1}^{|\mathcal{E}|}\prod_{l=1}^{d}\left(x_{l,i(k)}+x_{l,j(k)}\right).$$

If $\mathbf{V}$ is reflection invariant, and $\xi^* \le 1$ the Spearman's ρ satisfies:

$$(30)\quad \rho(F_{\mathbf{U}}) = \xi^*\rho(F_{\mathbf{V}}) + (1-\xi^*)\left(-\frac{(d+1)}{2^d-(d+1)}\right),$$

$$(31)\quad \rho(F_{\mathbf{U}}) \le \rho(F_{\mathbf{V}}).$$

[2]The Spearman's ρ following the definition in Ahn and Fuchs (2020) is obtained as: $(\rho(F_U)+\rho(F_W))/2$, with $\mathbf{W} = R_{\mathcal{D},\frac{1}{2}}(\mathbf{U})$.

We provide a simplified formula for $\rho\left(F_{\mathbf{V},\mathcal{D}\setminus\mathcal{L}_m}\right)$ in the case of two line segment constructions which will be studied later on in this paper. For the line segment representation (12) we obtain

$$\begin{aligned}\rho\left(F_{\mathbf{V},\mathcal{D}\setminus\mathcal{L}_m}\right) &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\mathbb{E}\left[V^m(1-V)^{d-m}\right] - \frac{1}{2^d}\right)\\ &= \frac{2^d(d+1)}{2^d-(d+1)}\left(B(m+1,d-m+1) - \frac{1}{2^d}\right),\end{aligned}$$

where $B(x,y)$ is the Euler's beta function.

For the representation (24), under the independence assumption V_l, $l=1,\ldots,d$ iid, the Spearman's ρ is

$$\begin{aligned}\rho\left(F_{\mathbf{V},\mathcal{D}\setminus\mathcal{L}_m}\right) &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\prod_{l\in\mathcal{L}_m}\mathbb{E}[V_l]\prod_{l\in\mathcal{D}\setminus\mathcal{L}_m}\mathbb{E}[(1-V_l)] - \frac{1}{2^d}\right)\\ (32)\qquad &= \frac{2^d(d+1)}{2^d-(d+1)}\left(\frac{1}{2^d}-\frac{1}{2^d}\right) = 0.\end{aligned}$$

Constructions obtained as a random permutation of a line segment construction inherit its concordance order rank since the multivariate Kendall's τ and Spearman's ρ are permutation invariant. In addition, they satisfy the d-CTM property as stated in the following.

COROLLARY 4. Let $\mathbf{W}$ be the exchangeable version of a strict d-CTM line segment vector $\mathbf{U}$ obtained as random permutations of its components, then $\mathbf{W}$ is strict d-CTM and exchangeable.

4. SPECIAL CASES

In this section, we discuss several examples starting from the new constructions proposed in this paper and then reviewing the constructions proposed in the literature which are special cases of our stochastic representations (12) or (24). We use multivariate Spearman's ρ to rank the proposals in concordance order[3].

4.1 Circulant Variates

Obtaining the coordinate matrix $\mathbf{X}$ used in Equation 12 can be costly in high dimensions especially when numerical procedures are used to solve the optimization problem stated in Section 2.2. We propose suitable constraints on the segment set $\mathcal{S} = (\mathbf{X},\mathcal{G})$ to reduce the computational cost of our procedure. The proposed conditions on the coordinate matrix $\mathbf{X}$ allows for decoupling the CTM constraint.

[3]To keep the paper to a reasonable length, we report here the values of the multivariate Spearman's ρ for dimensions d from 2 to 5. We have evaluated the ordering of the constructions up to dimension $d=20$, and no changes were observed.

First, we assume that the number of vertices n is equal to the dimension d of the random vector and $\mathbf{X} = d/2\tilde{\mathbf{X}}$ where $\tilde{\mathbf{X}}$ is doubly stochastic, and obtain:

$$\sum_{i=1}^{d} x_{i,k} = \sum_{k=1}^{n} x_{i,k} = \frac{d}{2}.$$

This assumption allows us to simplify the optimization problem and to search for independent solutions for each row of $\mathbf{X}$.

We assume further constraints on the matrix $\mathbf{X}$ and on the graph $\mathcal{G}$ such that the same optimization problem is solved for all rows of $\mathbf{X}$. We assume the first coordinates of the d vertices are arranged in increasing order $x_{11} \leq \ldots \leq x_{1d}$ and compute the k-th coordinates as the k-th circular permutation of the first ones:

$$\begin{aligned} x_{k1} &= x_{1\,(k-1)(\mathrm{mod}\,d)+1}, \\ &\vdots \\ x_{ki} &= x_{1\,(i-1+(k-1))(\mathrm{mod}\,d)+1}, \\ &\vdots \\ x_{kd} &= x_{1\,(d-1+(k-1))(\mathrm{mod}\,d)+1}. \end{aligned} \tag{33}$$

The resulting coordinate matrix $\mathbf{X}$ is a circulant matrix with i-th row sum equal to the i-th column sum for all rows. Imposing $x_{11} \leq \ldots \leq x_{1d}$ implies the same set of $\mathbf{a}_l$ is used for all $l \in \mathcal{D}$, with the same multiplicities $|\mathcal{M}_{l,m}| = |\mathcal{M}_{1,m}|$ but with different positions $m \in \mathcal{M}_{l,m}$ as effect of the circular permutation.

Furthermore, we choose the edge set in such a way as to have the same projected graph for each set of coordinates and assume a circulant graph that is invariant by circular shifts of the vertexes.

DEFINITION 4.1 (Circulant Graph). Given a subset $\mathcal{L} \subseteq \left\{1, \ldots, \left\lfloor \frac{d}{2} \right\rfloor\right\}$ then the d-vertex circulant graph $\mathcal{C}_d(\mathcal{L})$ is a graph with vertices $1, \ldots, d$ and edge set $\mathcal{E}_{d,\mathcal{L}}$ is such that $(i,j) \in \mathcal{E}_{d,\mathcal{L}}$ if either $|i-j| \in \mathcal{L}$ or $(d - |i-j|) \in \mathcal{L}$.

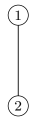
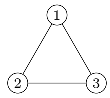
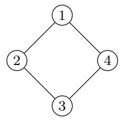
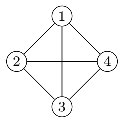
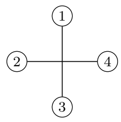
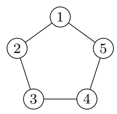
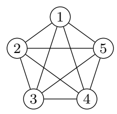
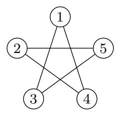
The circular symmetry imposed on the vertex coordinates and on the graph simplifies the optimization problem. Whatever the multiplicities $|\mathcal{M}_{1,m}|$, under Assumptions 1 and 2, we obtain the following results:

$$\Psi_l(\mathbf{a}_l) = \Psi_1(\mathbf{a}_1) = \Phi_1(\mathbf{x}_1)$$
$$= -\frac{1}{|\mathcal{E}_{d,\mathcal{L}}|} \sum_{(i,j)\in\mathcal{E}_{d,\mathcal{L}}} \log|x_{1,i} - x_{1,j}|,$$

$$\sum_{m=1}^{n_1} |\mathcal{M}_{1,k}|\, a_{1,m} = \sum_{i=1}^{d} x_{1,i}.$$

This rewriting of constraint and objective function allows one to minimize on $\mathbf{x}_1$ instead of $\mathbf{a}_l$. Minimization automatically excludes vectors that violate Assumption 1 given $\mathcal{E}_{d,\mathcal{L}}$, because for those cases the objective function is infinite. The following definitions introduce formally our new proposal called Circulant Variates (CCV).

TABLE 2
Circulant graphs $\mathcal{C}_d(\mathcal{L})$ of the Circulant Countermonotonic on Segments up to dimension 5.

Graph Label	$\mathcal{G}$	$\mathbf{x}_1$	Spearman's ρ
$\mathcal{C}_2(\{1\})$		$(0,1)$	-1
$\mathcal{C}_3(\{1\})$		$\left(0, \frac{1}{2}, 1\right)$	-0.5
$\mathcal{C}_4(\{1\})$		$\left(0, \frac{1}{3}, \frac{2}{3}, 1\right)$	-0.2840
$\mathcal{C}_4(\{1,2\})$		$\left(0, \frac{1}{2} - \frac{1}{2\sqrt{5}}, \frac{1}{2} + \frac{1}{2\sqrt{5}}, 1\right)$	-0.2763
$\mathcal{C}_4(\{2\})$		$(0,0,1,1)$	-0.2121
$\mathcal{C}_5(\{1\})$		$(0, 1/4, 2/4, 3/4, 1)$	-0.1659
$\mathcal{C}_5(\{1,2\})$		$\left(0, \frac{1}{2} - \frac{\sqrt{3}}{2\sqrt{7}}, \frac{1}{2}, \frac{1}{2} + \frac{\sqrt{3}}{2\sqrt{7}}), 1\right)$	-0.1577
$\mathcal{C}_5(\{2\})$		$\left(0, 0, \frac{1}{2}, 1, 1\right)$	-0.1385

DEFINITION 4.2 (Circulant Countermonotonic). We call the circulant matrix $\mathbf{X} \in \mathbb{R}^d \times \mathbb{R}^n$ whose rows are obtained by the d circular shifts of the first row, a solution of the Circulant Countermonotonic on Segments problem, on the circulant graph $\mathcal{C}_d(\mathcal{L})$, if $x_{1,1} = 0$, $x_{1,d} = 1$ and $\{x_{1,2}, \ldots, x_{1,d-1}\}$ solve the convex minimization problem:

$$\min_{\{x_{l,m}\}_{m=2}^{d-1} \in [0,1]^{d-2}} \Phi_1(\mathbf{x}_1)$$

with

$$\Phi_1(\mathbf{x}_1) = -\frac{1}{2|\mathcal{E}_{d,\mathcal{L}}|} \sum_{(i,j)\in\mathcal{E}_{d,\mathcal{L}}} \log|x_{1,i} - x_{1,j}|,$$

subject to:

$$\sum_{i=1}^{d} x_{1,i} = \frac{d}{2}.$$

DEFINITION 4.3 (Circulant Variates). Let $\mathcal{S}_{d,\mathcal{L}} = \{\mathbf{X}, C_d(\mathcal{L})\}$ a collection of segments such that $\mathbf{X}$ is a solution Circulant Countermonotonic on Segments problem on $\mathcal{C}_d(\mathcal{L})$. Variates obtained from the components of the d-dimensional random vector uniformly distributed on $\mathcal{S}_{d,\mathcal{L}}$ are Circulant Variates (CCV).

The following corollary is an application of Theorem 1 to CCV.

COROLLARY 5. CCV are marginally standard uniform and constant in sum.

For $C_d(\{1\})$ a solution can be derived as follows. The sum in each of the first $d-1$ equations of (12) has only two terms, and considering the m-th equation we have

$$\frac{1}{(x_{1m+1} - x_{1m})} + \frac{1}{(x_{1d} - x_{11})} = d.$$

Substituting the constraints $x_{11} = 0$ and $x_{1d} = 1$, we find that the x's are uniformly spaced on the unit interval and satisfy all the $d-2$ equations in (17). The case of $C_3(\{1\})$ was already studied in Nelsen and Úbeda-Flores (2012) . In their example, the probability mass has a distribution uniform on the edges of the triangle with vertices $\mathbf{x}_1 = (0, 1/2, 1)$, $\mathbf{x}_2 = (1/2, 1, 0)$, $\mathbf{x}_3 = (1, 0, 1/2)$. The other two vertices are different 3-cycles of the first one. This implies that the row sum of $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, is equal to its column sum and both are equal to $3/2$. Lee and Ahn (2014) show that the construction in Nelsen and Úbeda-Flores (2012) is 3-CTM.

Additionally, the exchangeable version of the $C_d(\{1\})$ construction is distribution-wise equivalent to the degenerate random balanced sampling introduced in Equation (8) of Gerow and Holbrook (1996). They propose to generate Z_1 as an uniform random variable on $[-1, 1]$ and obtain the remaining variables according to

$$Z_l = c_l - \frac{Z_1}{d-1}, \quad c_l = -1 + \frac{2l-3}{d-1},$$

$l \in \mathcal{D}$, and then randomly permute the Z_l. They show that $\sum_{l=1}^{d} Z_l = 0$ and that once permuted, the Z_l's are uniformly distributed on $[-1, 1]$. If we set $U_l = (Z_l + 1)/2$ then the permuted version can be written in terms of permuted $C_d(\{1\})$ construction:

PROPOSITION 3. The exchangeable version of $U_l = (Z_l + 1)/2$ $l \in \mathcal{D}$ has the same distribution of the exchangeable version of CCV with dependence graph $C_d(\{1\})$.

4.2 Rotation Sampling

Fishman and Huang (1983) rephrase the original Hammersley and Morton (1956) proposal for $d > 2$ obtaining an equivalent construction with the standard marginal uniformity that was missing in Hammersley and Morton (1956). Their construction was named rotation sampling because the modulo one arithmetic on which it is based is often associated with circular motion.

PROPOSITION 4. The line segment stochastic representation (12) of the rotation sampling in the Example 1 has $2d$ vertices with coordinate matrix of elements:

$$x_{l,m} = \begin{cases} \dfrac{l+m-1}{d}, & \text{if } m < d+2-l, \\ \dfrac{l+m-1-d}{d}, & \text{if } m \geq d+2-l, \end{cases}$$

$$x_{l,d+m} = \begin{cases} \dfrac{l+m-2}{d}, & \text{if } m < d+2-l, \\ \dfrac{l+m-2-d}{d}, & \text{if } m \geq d+2-l \end{cases}$$

and edge set:

$$\mathcal{E}^{RS} = \{(i, d+i) \, | i \in \{1, 2, \ldots, d\}\}.$$

COROLLARY 6. A rotation sampling random vector has standard uniform marginals and is not d-CTM.

We report in Table 3 the analytic values of the Multivariate Spearman's ρ for rotation sampling vectors up to dimension $d = 5$.

TABLE 3
Values of the multivariate Spearman's ρ for rotation sampling.

d	2	3	4	5
$\rho_{\min}$	-0.5	-0.33	-0.2168	-0.1372

For $d = 2$ this construction does not reduce to the usual antithetic variates. This is suggested by a value of multivariate Spearman's ρ different from the value of -1 attained by the Fréchet lower bound. Lacking the constant sum property, the proposal has multivariate Spearman's ρ larger than one of the d-CTM proposals considered in this paper.

4.3 Arvidsen and Johnson: A Fresh Look

In the pioneering paper of Arvidsen and Johnsson (1982), the objective of variance reduction is obtained by designing the first standard uniform d-CTM construction (10). Craiu and Meng (2005) show that this construction is displacing the binary digits of U_1 and give the name permuted displacement to its exchangeable version. We show in the next proposition the relationship between Arvidsen and Johnson sampling scheme and the scheme from Example 3.

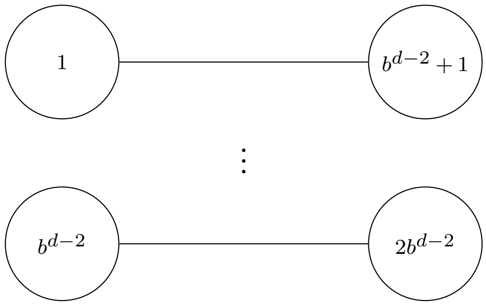


FIG 5. $\mathcal{G}$ *for b-based Arvidsen and Johnson' construction.*

PROPOSITION 5. For $d = 3$, the construction of Arvidsen and Johnsson (1982) given in Example 4 is equivalent to the antithetic proposal of Gaffke and Rüschendorf (1981) given in (7).

The Ardvisen and Johnson construction (AJ) and the following family of general constructions admit a line segment representation. The generalization is useful to show that Ardvisen and Johnson is the only d-CTM within this family and attains consequently the minimal Spearman's ρ within this family.

DEFINITION 4.4. Given $U_1 \sim \mathcal{U}[0,1]$ and $b \in \mathbb{N}$ the base-b Ardvisen and Johnson construction is:

$$U_i = \left(b^{i-2}U_1 + 1/b\right) \bmod 1,\ i = \{2,\dots,d-1\},$$

$$U_d = 1 - \left(b^{d-2}U_1\right) \bmod 1. \tag{34}$$

PROPOSITION 6. The line segment stochastic representation (12) of the construction in (34) has $2b^{d-2}$ verticeswith coordinate matrix $\mathbf{X} = (\mathbf{z}^T, \mathbf{y}^T)$ where:

$$\begin{aligned}
\mathbf{y}_1 &= \left(0, \frac{1}{b^{d-2}}, \dots, \frac{b^{d-2}-1}{b^{d-2}}\right)^T, & \mathbf{z}_1 &= \mathbf{y}_1 + \frac{1}{b^{d-2}}\mathbf{1}_{b^{d-1}},\\
\mathbf{y}_k &= \left(b^{k-2}\mathbf{y}_1 + \frac{1}{b}\right) \bmod 1, & \mathbf{z}_k &= \mathbf{y}_k + \frac{1}{b^{d-k}}\mathbf{1}_{b^{d-1}},\\
\mathbf{y}_d &= \mathbf{1}_{b^{d-1}}, & \mathbf{z}_d &= \mathbf{y}_d - \mathbf{1}_{b^{d-1}},\\
\mathbf{y} &= (\mathbf{y}_1, \dots, \mathbf{y}_d), & \mathbf{z} &= (\mathbf{z}_1, \dots, \mathbf{z}_d)
\end{aligned}$$

and edge set:

$$\mathcal{E}^{AJ} = \left\{\left(i, b^{d-2}+i\right) | i \in \{1,2,\dots,b^{d-2}\}\right\}.$$

COROLLARY 7. The base-b Ardvisen and Johnson random vector has standard uniform marginals. Only the case $b = 2$ is d-CTM.

Table 4 shows that the constructions in Definition 4.4 attain the lowest Spearman's ρ for $b = 2$.
Table 4 reports also the multivariate proposal of Gaffke and Rüschendorf (1981) (GR) described in Example 3. The table shows that GR performs better than the non d-CTM proposal, but worse than the AJ proposal because it

TABLE 4
Multivariate Spearman's ρ for base-b Ardvisen and Johnson random vectors and the multivariate proposal of Gaffke and Rüschendorf (1981) in Example 3.

Spearman's ρ		b = 1	2	3	4	5	GR
d	2	-1	-1	-1	-1	-1	-1
	3	-0.3333	-0.5000	-0.3333	-0.2083	-0.1200	-0.5000
	4	-0.0909	-0.2822	-0.1662	-0.0869	-0.0367	-0.2525
	5	0.0154	-0.1637	-0.0933	-0.0455	-0.0165	-0.1538

has independence as one of the main ingredients as discussed in the introduction. This also shows the effectiveness of multivariate ρ as a ranking measure for d-CTM vectors.

4.4 Latin Hypercube Iterations

In this section, we will reconsider the Iterated Latin Hypercube construction introduced in Craiu and Meng (2005) and establish a relationship with a d-dimensional generalization of the superstar introduced in Gerow and Holbrook (1996), and with our strict countermonotonic on segments construction.
The Latin Hypercube sampling, introduced by McKay, Beckman and Conover (1979), and further developed by Craiu and Meng (2005) with the goal of obtaining variance reduction in MCMC sampling, consists of the following steps: for $t = 0,\dots,T$ take an iid standard uniform d-dimensional random vector $\mathbf{U}_0$ and let $\mathcal{D}_t^\sigma = (\sigma_t(0), \dots, \sigma_t(d-1))^T$ be a permutation of $\{0,1,\dots,d-1\}$ independent of $\mathbf{U}_0,\dots,\mathbf{U}_{t-1}$ and

$$\mathbf{U}_t = \frac{1}{d}\left(\mathcal{D}_t^\sigma + \mathbf{U}_{t-1}\right). \tag{35}$$

If $t = 1$ (35) corresponds to the original Latin Hypercube Sampling, and $t > 1$ to the Iterated Latin Hypercube procedure introduced in Craiu and Meng (2005). It was shown in Craiu and Meng (2006) that ILH iterations represent an Iterated Function System with probabilities (IFSP) $\left([0,1]^d, (w_\sigma), p_\sigma\right)$ with w_σ similitudes with contraction ratio d^{-1} associated to each permutation σ of $\{1,\dots,d\}$

$$w_\sigma(\mathbf{u}) = \left(\frac{\sigma(1)-1}{d} + \frac{u_1}{d}, \dots, \frac{\sigma(d)-1}{d} + \frac{u_d}{d}\right). \tag{36}$$

We can show that the 3-dimensional superstar considered in Gerow and Holbrook (1996) can be obtained by using the same IFSP:

PROPOSITION 7. The 3-dimensional superstar proposed in Gerow and Holbrook (1996):

$$\mathbf{X}_t = f_k(\mathbf{X}_{t-1}) = \frac{1}{3}\mathbf{X}_{t-1} + \frac{2}{3}\mathbf{V}_k \tag{37}$$

with $\mathbf{V}_k$ a random permutation of $\{-1,0,1\}$ and an initial $\mathbf{X}_0$ a 3-dimensional vector such that $-1 \le \mathbf{X}_{i0} \le 1$ and $\sum_{i=1}^{3} \mathbf{X}_{i0} = 0$, up to a change of support, is generated by the same IFSP of the 3-dimensional version of the Iterated Latin Hypercube construction introduced in Craiu and Meng (2005).

PROPOSITION 8. The line segment stochastic representation (24) of the construction in (35) has $2d!$ vertices with coordinates $x_{l,k} = (\sigma_k(l) + 1)d^{-1}$ and $x_{l,d!+k} = (\sigma_k(l))d^{-1}$, $k = 1, \ldots, d!$ and edge set:

$$\mathcal{E}^{LH} = \{(i, d! + i) \,|\, i \in \{1, 2, \ldots, d!\}\}.$$

Craiu and Meng (2005) obtained standard marginal uniformity for the special case of an initial vector with iid components. In the superstar case, the original distribution is concentrated on points but converges to standard marginal uniforms as shown in Gerow and Holbrook (1996). In Craiu and Meng (2005), it is also shown that in the limit $t \to \infty$ the ILH is d-CTM. We show that the ILH iterations preserve marginal uniformity and constant sum in the general case.

COROLLARY 8. Let $\mathbf{U}_{t-1}$ be a dependent random vector of dimension d, whose coordinates add up to d/2 (a.s.) and each coordinate has a $\mathcal{U}[0,1]$ distribution. Then the random vector $\mathbf{U}_t$ in (35) has all its coordinates marginally $\mathcal{U}[0,1]$ distributed and adding up to $d/2$.

The preservation of strict d-CTM property raises the possibility of using ILH iterations on Arvidsen and Johnson's and CCV constructions. In addition, using results in section 3.2, we can give a closed-form expression for the multivariate Kendall's τ and Spearman's ρ for ILH iteration applied to $\mathbf{V}$ with i.i.d components or in representation (12) and rank the obtained random vector in the concordance order. In particular, ILH iterations on the Arvisen and Johnson construction and CCV have a constant sum and minimal multivariate Kendall's τ (equation (28)). Spearman's ρ for those constructions can be easily obtained using the deterministic composition and equation (32). Using the segment representation in Proposition 8, the ILH(T) proposal applied to $\mathbf{V}$ with iid components, can be expressed as a T-fold deterministic composition. Each composition expands the cardinality of vertex and edge sets by $d!$, resulting in a vertex set of cardinality $2(d!)^T$ and an edge set of cardinality $(d!)^T$. To maintain a feasible notation we substitute the index k of the different edges with the multi-index $\{k_t\}_{t=1}^{T}$ where each $k_t = 1, \ldots, d!$:

$$\alpha_{l,k_1,\ldots,k_T} = \sum_{r=1}^{T} \frac{\sigma_{k_r}(l)}{d^{T-r+1}}, \tag{38}$$

$$\beta_{l,k_1,\ldots,k_T} = \sum_{r=1}^{T} \frac{\sigma_{k_r}(l)}{d^{T-r+1}} + \frac{1}{d^T}. \tag{39}$$

The following proposition allows for computing multivariate Kendall's τ and Spearman's ρ for the ILH(T) case.

PROPOSITION 9. Let $\mathbf{U}_0$ be a d dimensional random vector with iid $\mathcal{U}[0,1]$ components and let $\mathbf{U}_T$ be the vector obtained by applying to it the ILH transformation in (35). Then the Kendall's τ is

$$\tau = \frac{1}{2^{d-1} - 1}\left(\frac{1}{(d!)^T} - 1\right),$$

and the Spearmans' ρ has coefficient ξ^* in Equation (30) equal to

$$\xi^* = \sum_{m_0+m_1+\ldots+m_T=d} \binom{d}{m_0, m_1, \ldots, m_T} \xi^*_{m_0,\ldots,m_T},$$

where

$$\xi^*_{m_0,\ldots,m_T} = \prod_{t=1}^{T} \binom{d}{m_t}^{-1} d^{-m_0} \sum_{\{i_1,\ldots,i_{m_t}\} \in \mathcal{P}_{\mathcal{D},m_t}} \prod_{l=1}^{m_t} \frac{2(i_l - 1)}{d^{T-t+1}}$$

and $\mathcal{C}_{\mathcal{D},m_2}$ denotes the set of combinations of the elements of $\mathcal{D}$ with m_t elements.

In the one iteration case, i.e., $T = 1$, we obtain the following expression:

$$\begin{aligned}\xi^* &= \prod_{l=0}^{d-1} \frac{(2l+1)}{d} = \prod_{l=1}^{d} \frac{(2l-1)}{d} \\ &= \frac{(2d-1)!}{d^d\left(2^{d-1}(d-1)!\right)} = \frac{\prod_{l=d}^{2d-1} l}{d^d 2^{d-1}}\end{aligned}$$

and using the arithmetic and geometric means inequality we obtain the bound:

$$\xi^*_1 \le \frac{\left(\frac{1}{d}\sum_{l=d}^{2d-1} l\right)^d}{d^d 2^{d-1}} \le 2\left(\frac{3d-1}{4d}\right)^d \le 1.$$

For the sake of comparison, in Figure 6 we summarize Kendall's τ and Spearman's ρ for all the constructions discussed in this section. Since Kendall's τ and Spearman's ρ are permutation invariant, the same ranking applies to the exchangeable versions of the constructions.

5. A CENTRAL LIMIT THEOREM

In this section, we study the Central Limit Theorem for our best-performing classes of variates. While the derivation does not yield a minimality result, it complements the one in Equation 4 because it guarantees asymptotic variance reduction for all square-integrable functions.

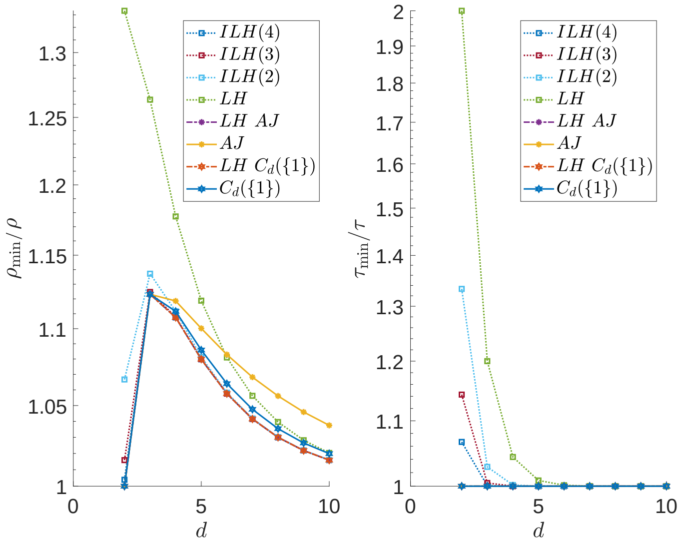


FIG 6. *Ratio of minimum value attainable by multivariate association measures and the values attained by different antithetic random vectors (vertical axis) as a function of dimension d (horizontal axis). Left: the Spearman's ρ measure. Right: the Kendall's τ measure. Note: in each plot, larger values indicate constructions farther from the minimum.*

DEFINITION 5.1 (Generalized Latin Hypercube Sample). Let σ_i, $i = 1, \dots, p$ be independent random permutations of $0, \dots, d-1$ and $\mathbf{V}^i = \left(V_1^i, \dots, V_d^i\right)$, $i = 1, \dots, p$ random vectors, independent from the σ_i and from each other, identically distributed with probability measure μ. A $p \times d$ matrix $\mathbf{U}$ is a Generalized Latin Hypercube Sample if it has the stochastic representation

$$U_l^i = \frac{\sigma_i(l) + 1}{d} V_l^i + \frac{\sigma_i(l)}{d}\left(1 - V_l^i\right) \tag{40}$$

with $i = 1, \dots, p$ and $l = 1, \dots, d$.

We remark that the constructions with the lowest multivariate Spearman's rho ILH(T), LH-$C_d(\{1\})$, and LH-AJ generate Generalized Latin Hypercube Samples.

The following lemma introduces the irrelevance of the distribution on the $\mathbf{V}^i$, $i = 1, \dots, p$.

LEMMA 2. Consider a a-LH sample and a b-LH sample where a and b are two different Radon measures. The following relationship holds for every function f locally integrable with respect to both measures:

$$\mathbb{E}_{a-LH}\left[\left(\frac{1}{d}\sum_{l=1}^{d} f(\mathbf{U}_l)\right)^r\right] - \mathbb{E}_{b-LH}\left[\left(\frac{1}{d}\sum_{l=1}^{d} f(\mathbf{U}_l)\right)^r\right] = o(1).$$

Given the previous lemma, we are able to show that the asymptotic distribution is the same as the ordinary Latin Hypercube. In particular, Stein (1987) express the variance of the Latin Hypercube using the ANOVA decomposition of the function f:

$$f(\mathbf{u}) = \mathbb{E}_{IID}[f(\mathbf{U})] + \sum_{i=1}^{p} f_i(u_i) + r(\mathbf{u}) \tag{41}$$
$$f_i(u_i) = \mathbb{E}_{IID}[f(\mathbf{u}) - \mathbb{E}_{IID}[f(\mathbf{U})] \,|U_i = u_i],$$

where $r(\mathbf{u})$ denotes the residual from the additive decomposition (Owen, 1992).

THEOREM 5. Let $\bar{X} = \frac{1}{d}\sum_{l=1}^{d} f(\mathbf{U}_l)$ with $\mathbf{U}_l$ $l \in \mathcal{D}$ from a μ-LH sample with μ being Radon and f being bounded and locally integrable with respect to μ. Then $\sqrt{d}\left(\bar{X} - E_{IID}(X)\right)$ converges in distribution to $\mathcal{N}\left(0, \int_{[0,1]^p} r(\mathbf{u})^2 d\mathbf{u}\right)$, where $r(\mathbf{u})$ is introduced in (41).

REMARK 2. The hypothesis of $\mathbf{U}^j$ being independent of $\mathbf{U}^k$ for all $j \neq k$ $j, k = 1, \dots, p$ in the definition of Generalized Latin Hypercube Sample is not restrictive as it seems. In practice, one can use the inverse Rosenblatt transform (see for example Rüschendorf (2013), Theorem 1.12) to obtain samples from a generic distribution. Then, if the composition of f with the inverse Rosenblatt transform is bounded, we are still under the incidence of Theorem 5. A multivariate version of the central limit theorem can be obtained as in Corollary 1 of Owen (1992). Concerning the boundedness assumption on f, it is probably too restrictive because in Loh (1996) a multivariate Berry-Essen type bound for the standardized multivariate version of $\bar{X}$, in the case of Latin Hypercube, is obtained under the assumption that the multivariate function $\mathbf{f}$ involved is Lebesgue measurable and $\mathbb{E}\|\mathbf{f}\|^3 < \infty$.

6. NUMERICAL ILLUSTRATIONS

We illustrate the performance of the methods presented in this paper using several simulation exercises involving standard Monte Carlo, Markov chain Monte Carlo and Sequential Monte Carlo algorithms.[4]

One of the critical dimensions used to rank our countermonotonic vectors is sampling time, as shown in Figure 7. It is obtained by averaging over 1000 independent replications. In each experiment, we sampled 5000 antithetic vectors of dimension $2 \leq d \leq 20$. Sampling schemes have been implemented in Matlab on a Windows 10 laptop with an Intel i7-6500U CPU and 8 GB of RAM.

Permuted $C_d(\{1\})$ outperforms the other competitors. We report the times for the segment version that randomly

[4]Replication codes for the examples in the paper can be found at https://github.com/Frattalol/Livingontheedge

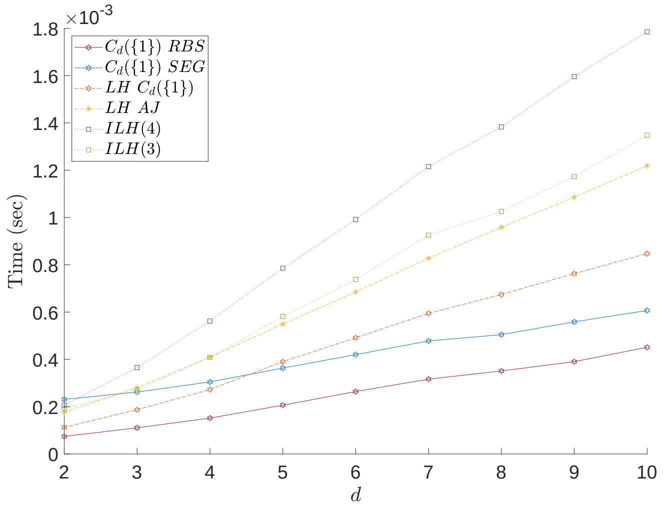


FIG 7. *Sampling times (vertical axis) as a function of the sampling dimension d (horizontal axis) for different methods (different symbols and colors). All estimates are averages over 1000 experiments. In each experiment, 5,000 antithetic vectors of dimensions d are sampled following a given method.*

permutes stochastic representation (12) and the RBS version that uses the equivalence with random balanced sampling described in Proposition 3. The latter choice is faster, and it is also used as the base for the Latin Hypercube iteration in $LH\text{-}C_d(\{1\})$.

Consequently, all the experiments in the section use the exchangeable version of LH-$C_d(\{1\})$. In fact, LH-$C_d(\{1\})$ vector reaches the lower value of multivariate Kendall's τ and Spearman's ρ (Figure 6) but is also a faster sampling scheme.

6.1 Monte Carlo Integration

The first evaluation of our methodology is for the Monte Carlo integration on the unit hypercube, with integration points chosen according to the three competing schemes. In standard Monte Carlo (MC), the sampling points are iid, and each sampling point corresponds to a random vector of dimension equal to the number of variables in the function. In our antithetic method (LH-$C_d(\{1\})$) the variables used to populate the same coordinate for different sampling points are from an antithetic vector. Across coordinates, these antithetic vectors are independent of one another. For the Quasi-Monte Carlo's Sobol scheme (QMC Sobol), a deterministic sequence of points uniformly covers the hypercube of dimension equal to the number of variables. In the evaluation of the integration problem complexity, the effective dimension plays an important role. We refer to Owen (2003), Wang and Fang (2003) and Wang and Sloan (2005) for the formal definition of truncation p_t and superposition p_s dimensions. Owen (2003) shows that a low p_s is necessary for QMC to surpass the computational efficiency of MC when the sample sizes are at practical levels. Wang and Fang (2003) and Wang and Sloan (2005) show that the integrands commonly used in option pricing have $p_t \simeq p$ and $p_s \leq 2$ and explain, using those results, QMC's good performance in this domain. According to our Theorem 5 for integrands with $p_s = 1$, i.e. for functions that are well approximated by sums of one-dimensional functions, LH-$C_d(\{1\})$ should be efficient in reducing the variance. Theorem 5 also guarantees that our method cannot perform worse than MC, asymptotically in the number of points. To investigate the role of effective dimension in the relative performance of the three competing methods, we use the two-parameter function introduced in Wang and Sloan (2005):

$$f(\mathbf{x}) = \prod_{i=1}^{p} \left(1 + a\tau^i (x_i - 1/2)\right). \tag{42}$$

Varying the parameter a has more effect on p_t than on p_s and varying the parameter τ has the opposite effect. We consider a high dimensional function ($p = 100$) and different specifications of effective dimensions, according to the parameters reported in Table 5.

TABLE 5
Wang and Fang (2003) effective dimensions (truncation and superposition dimensions, p_t and p_s, respectively) of the integrand function in Equation (42) for dimension $p = 100$ and different parameter settings (columns).

a	0.1	0.1	0.1	0.1	0.1	1	1	1	1	1	10	10	10	10	10
τ	0.1	0.5	0.8	0.9	1	0.1	0.5	0.8	0.9	1	0.1	0.5	0.8	0.9	1
p_t	2	4	11	22	100	2	4	11	23	100	2	5	17	39	100
p_s	1	1	1	1	2	1	1	2	3	14	1	3	8	15	96

Effective dimensions are computed using the methods for multiplicative functions introduced in Wang and Fang (2003).

Figure 8 shows the mean square error (MSE) (vertical axis) and the computing time (horizontal axis) of Monte Carlo (red), QMC (black) and circulant variates LH-$C_d(\{1\})$ (blue) sampling, for the different effective dimensions p_t and p_s (different plots) given in Table 5. Our LH-$C_d(\{1\})$ method has the best performance when the superposition dimension is equal to 1. The performance is decreasing in the truncation dimension. For $p_s = 2, 3$, QMC performs better than the method proposed here when the number of points used in the integration is high. QMC advantage increases in the truncation dimension. At moderate p_s QMC dominates. In those cases, our proposal is slower than MC but reaches the same MSE. In the extreme case of f being almost full dimensional (right lower corner), $C_d(\{1\})$ is performing as badly as MC, but QMC is doing orders of magnitudes worse. These numerical results are in line with the result in Theorem 5 and

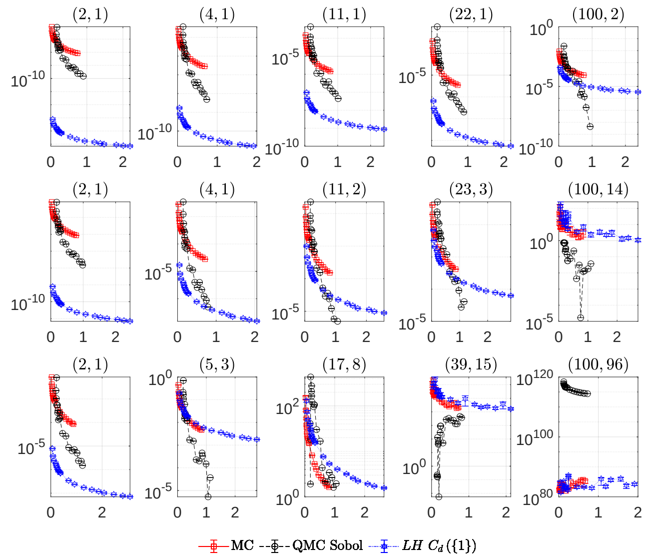


FIG 8. *Monte Carlo integration of the Wang and Sloan functions with $p = 100$ and different effective dimensions. Mean square error (vertical axis) and computing time in thousands of seconds (horizontal axis). Different plots use different effective dimensions (p_t, p_s). In each plot: Monte Carlo (red), QMC (black), and circulant variates LH-$C_d(\{1\})$ (blue) sampling. For each line: different number of samples from 10 to 1,000 (circles). For each setting, all statistics are averages over 10,000 experiments.*

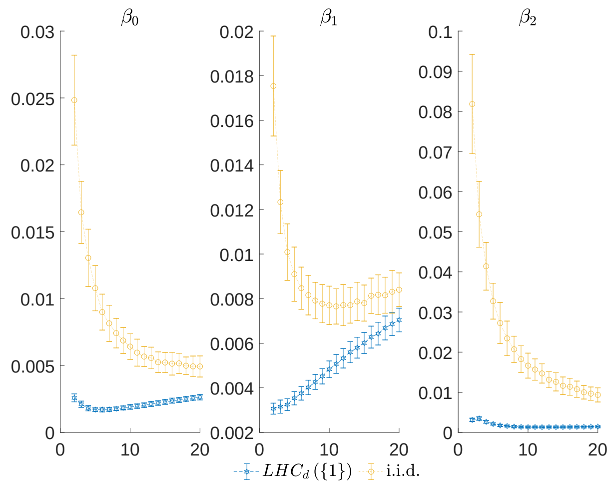


FIG 9. *Monte Carlo variance of the posterior mean estimator (vertical axis) corresponding to a different number of antithetic variates d (horizontal axis) for the parameters β_0 (left panel), β_1 (center panel) and β_2 (right panel). In each plot: the average variance of antithetic Gibbs (blue dots) and of iid Gibbs (yellow dots) with their range (vertical segments). Note: all estimates are based on 100 independent experiments. In each experiment, the Gibbs sampler runs for 10 seconds.*

indicate that our method should be used when the superposition dimension is low and when there is no information about effective dimensions of the integrand since in the worst case, it reproduces the precision of standard MC estimates.

6.2 Markov Chain Monte Carlo

6.2.1 Bayesian inference on Probit (van Dyk and Meng, 2001) The data used are taken from van Dyk and Meng (2001) and represent the clinical characteristics summarized by two covariates of 55 patients, of which 19 were diagnosed with lupus. The disease indicator is modelled as independent Bernoulli variables $Y_i \sim \mathcal{B}er\left(\Phi\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)\right)$ where Φ is the standard normal CDF and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ is a the vector of parameters. The objective is to sample from the posterior distribution corresponding to the flat prior for $\boldsymbol{\beta}$. We adopt the standard Gibbs sampler with latent variables $\psi_i \sim \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, 1\right)$ of which we consider only the sign. We repeat the following alternating two steps to obtain draws from the posterior. First, we sample from $\boldsymbol{\beta}|\boldsymbol{\psi} \sim \mathcal{N}\left(\tilde{\boldsymbol{\beta}}, \left(X^TX\right)^{-1}\right)$ with $\tilde{\boldsymbol{\beta}} = \left(X^TX\right)^{-1}X^T\boldsymbol{\psi}$ with X the data matrix whose i-th rows is $\mathbf{x}_i$. Then from $\psi_i\,|\beta, Y_i \sim \mathcal{TN}\left(\mathbf{x}_i^T\boldsymbol{\beta}, 1, Y_i\right)$ where $\mathcal{TN}\left(\mu, \sigma^2, Y\right)$ is a the normal distribution with mean μ and variance σ^2, truncated to be positive if $Y > 0$ or negative otherwise. Further details of the algorithms can be found in Craiu and Meng (2005).

6.2.2 Metropolis within Gibbs Hierarchical Poisson (Gelfand and Smith, 1990) This second example concerns the counts of failures $\mathbf{s} = (s_1, \ldots, s_n)$ for $n = 10$ pumps in a nuclear power plant. The time $\mathbf{t} = (t_1, \ldots, t_n)$ of operation are known. The model assumes $s_k \sim \mathcal{P}oi\left(\lambda_k t_k\right)$ and $\lambda_k \sim \mathcal{G}a\left(\alpha, \beta\right)$ with parameters $\alpha, \beta > 0$. The objective of the inference is the posterior distribution of α and β, to which we assign an exponential prior with mean 1 and a Gamma prior $\mathcal{G}a\left(0.1, 1\right)$, respectively. Using the conjugate priors, it is easy to obtain a Gamma distribution for λ_1 given λ_2, ..., λ_n and β, and those variables can be easily sampled using a Gibbs step. Sampling α is slightly more difficult. In fact, we have:

$$\mathbb{P}\left(\alpha\,|\lambda_1, \ldots, \lambda_n, \beta\right) \propto \exp\left[\alpha\left(n\log\beta + \sum_{k=1}^{n}\log\lambda_k - 1\right) - n\log\Gamma\left(\alpha\right)\right].$$

We then sample α with a random walk Metropolis-Hastings (MH) step with a deterministic scan. We show results for the case when we antithetically couple the uniform draws for acceptance rejection choice and when we are not doing it. The former case is the one for which Frigessi, Gåsemyr and Rue (2000) reports the worst performance of the usual two-variates antithetic coupling of chains, in agreement with our results.

6.2.3 Pseudo Marginal Metropolis-Hastings Stochastic Volatility (Gerber and Chopin, 2015) The last application targets a state-of-the-art methodology, the Pseudo Marginal MH (PCMH) proposed by Andrieu and Roberts (2009), which is able to estimate models with intractable likelihoods that are approximated using a particle filter. In

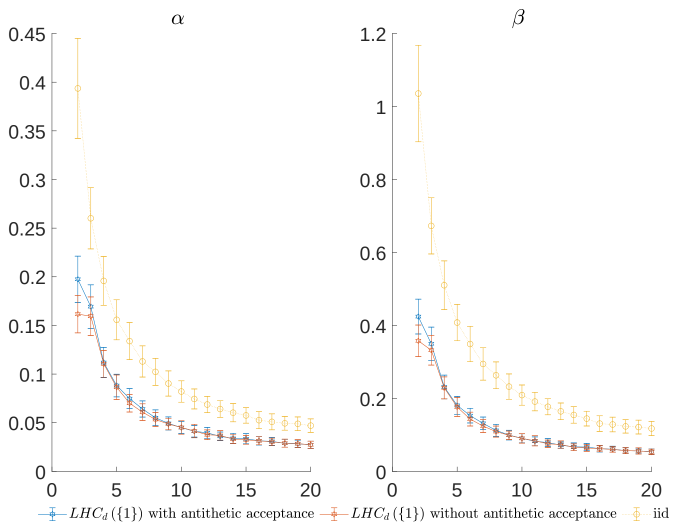


FIG 10. *Monte Carlo variance of the posterior mean estimator (vertical axis) corresponding to a different number of antithetic variates d (horizontal axis) for the parameters α (left panel) and β (right panel). In each plot: the average variance for the MH with (blue dots), without (red dots) antithetic acceptance rule and standard iid MH (yellow dots) and their ranges (vertical segments). Note: all estimates are based on 100 independent experiments. In each experiment, the Gibbs sampler runs for 10 seconds.*

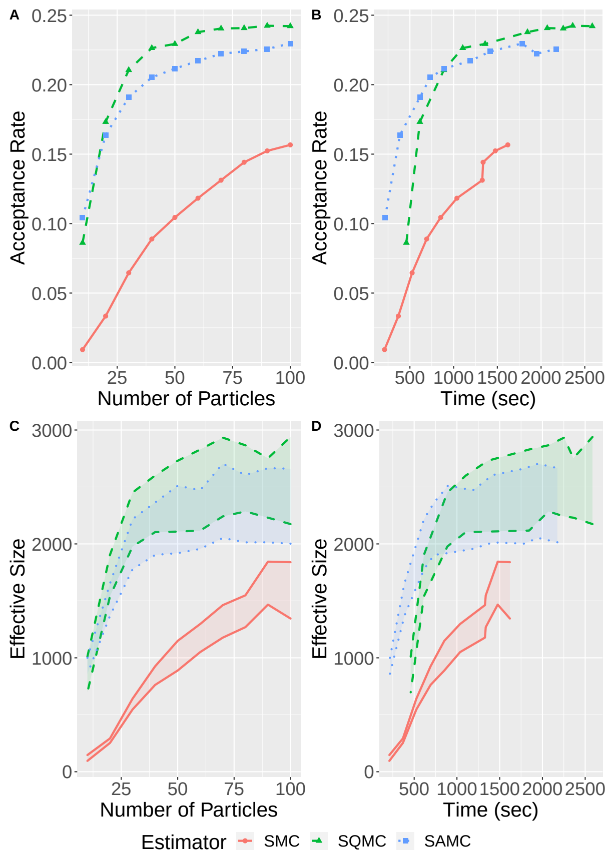


FIG 11. *Acceptance rate (top) and effective sample size (bottom) of the PCMH using Sequential Monte Carlo (SMC), Sequential Quasi-Monte Carlo (SQMC), and Sequential Antithetic Monte Carlo (SAMC) (different colors). Acceptance rate of the Metropolis step (vertical axis) versus the number of particles (horizontal axis, Panel A) and computing time (horizontal axis, Panel B). Maximum and minimum effective sample size (vertical axis) versus number of particles (horizontal axis, Panel C) and computing time (horizontal axis, Panel D).*

particular, following Gerber and Chopin (2015), we consider the bivariate stochastic volatility model introduced in Chan, Kohn and Kirby (2006):

$$\begin{aligned} \mathbf{y}_t &= S_t^{1/2}\boldsymbol{\epsilon}_t, \\ \mathbf{x}_t &= \boldsymbol{\mu} + \Phi\left(\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \Psi^{1/2}\nu_t, \\ S_t &= \operatorname{diag}\left(\exp\left(x_{1t}, x_{2t}\right)\right), \\ (\boldsymbol{\epsilon}_t, \boldsymbol{\nu}_t) &\sim \mathcal{N}\left(\mathbf{0}_4, C\right) \end{aligned}$$

with $\mathbf{y}_t = (y_{1t}, y_{2t})^T$ and $\mathbf{x}_t = (x_{1t}, x_{2t})^T$ observable and latent log-volatility vectors, Φ and Ψ diagonal matrices and C a correlation matrix. Following those authors we take independent uniform and gamma prior distributions:

$$\phi_{ii} \sim \mathcal{U}[0,1], \tag{43}$$

$$\frac{1}{\psi_{ii}} \sim \mathcal{G}a\left(10\exp\left(-10\right), 10\exp\left(-3\right)\right) \tag{44}$$

and a flat prior for μ, where ϕ_{ii} and ψ_{ii} denote the diagonal elements of Φ and Ψ, respectively. In addition, we assume that C is uniformly distributed on the space of correlation matrices. To sample from the posterior distribution of the parameters, we use a Gaussian random-walk MH algorithm with covariance matrix calibrated by Gerber and Chopin (2015) so that the acceptance probability of the algorithm becomes, as N tends to infinity, close to 25%. We consider the mean-corrected daily returns on the Nasdaq and Standard and Poor's 500 indices for the period ranging from January 3rd, 2012, to October 21st, 2013, so that the data set contains 452 observations. Figure 11 PCMH algorithms using sequential quasi-Monte Carlo and antithetic Monte Carlo are equivalent in acceptance rate and effective sample size (Panels A and C) when a low number of particles (up to 20) is used. Nevertheless, antithetic Monte Carlo achieves larger acceptance rates (AR) and effective sample size (ESS) with a lower computing time (Panels B and D). When a larger number of particles is used (above 20), the performances are equivalent in terms of ESS, whereas SQMC is better in terms of AR.

7. DISCUSSION

The development of antithetic constructions has generated a rich class of methods to accompany the evolution of Monte Carlo sampling algorithms. We enrich this class with a new antithetic method, the circulant variates (CCV), that satisfies the countermonotonicity, exchangeability, and marginal uniformity conditions. In particular, the marginal uniformity condition is linked to the Kullback–Leibler optimality.

The principle behind the proposal, relying on sampling on segments, leads to a unification of several classical antithetic constructions: rotation sampling, Latin hypercube, permuted displacement, and random balanced sampling.

Within this common framework, we provide a convenient representation of the antithetic vectors in terms of graphs, i.e. vertices, and edges, and evaluate theoretically their distributions and concordance measures. The latter allows us to rank the methods within the class of sampling on segments. The constructions based on circulant graphs (CCV) with the smallest number of edges rank best. Also, the best CCV outperforms the existing constructions reviewed in the paper. We also demonstrate a central limit theorem in the case of asymptotically increasing vector size.

Leveraging on Iterated Latin Hypercube (ILH) properties, we combine the two methods by using the CCV construction to initialize the ILH construction and reduce the number of iterations. This reduces the simulation cost and improves the concordance lower bound. The numerical experiments include MCMC, Sequential MC, Quasi-MC, and classical MC integration. The proposed methods outperform standard implementations and are competitive with Quasi-Monte Carlo methods in scenarios with low effective dimensions. The asymptotic variance reduction with respect to standard MC, implied by Theorem 5, is confirmed by our numerical experiments. The results hold for all square-integrable functions. Moreover, the variance reduction is larger for functions that are well approximated by sums of one-dimensional functions.

Future work includes possible extensions of the theory for KL optimality beyond the marginal univariate case. An investigation of the relationship between the line segment representation and orthogonal array-based Latin Hypercubes (Tang, 1993) could lead to an improvement of the performance for superposition dimensions bigger than 1. We notice that different methods that satisfy countermonotonicity, exchangeability, and marginal uniformity yield different variance reductions in practice. The results for the CCV suggest a non-trivial relationship between graph topology, concordance order, and variance reduction. A more general and ambitious goal is to propose a mathematical framework directed at identifying the additional features that produce these differences.

Our sampling method has been successfully applied within the Bayesian estimation framework of the European Commission's multi-country model (Albonico et al., 2019). We expect that the proposed simulation technique will find direct application in other fields of computational mathematics and statistics.

ACKNOWLEDGMENTS

The views expressed by Lorenzo Frattarolo are the author's alone and do not necessarily correspond to those of the European Commission. We thank for their useful comments the participants at: the *1st Italian-French Statistics Workshop*, October 2017, Italy; the *32nd European Statisticians Meeting*, Italy, July 2019; the *Workshop "Monte Carlo methods, and approximate dynamic programming"*, France, October 2019; the *XXII Workshop on Quantitative Finance*, Italy, April 2022; the *49th Annual Meeting of the Statistical Society of Canada*, May 2022; the *XXXII Conference of the Association of European Operational Research Societies*, Finland, July 2022; *the 24th International Conference on Computational Statistics*, Bologna, Italy, August 2022; and the *12th European Bayesian Econometrics Seminar*, Salzburg, Austria, September 2022.

FUNDING

Radu V. Craiu's research was supported by NSERC of Canada grant RGPIN-249547. Roberto Casarin acknowledges the support from: the Venice Centre for Risk Analytics (VERA) at the Ca' Foscari University of Venice; the European Union - Next Generation EU - Project 'GRINS - Growing Resilient, INclusive and Sustainable'; the National Recovery and Resilience Plan (NRRP) Spoke 4. Christian P. Robert's research has been supported by the French Government under both a senior chair (2016-2021) from l'Institut Universitaire de France and under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR19-P3IA-0001 (PRAIRIE 3IA Institute), as well as the ERC Synergy grant 101071601 (OCEAN).

SUPPLEMENTARY MATERIAL

Supplement to "Living on the Edge: An Unified Approach to Antithetic Sampling"
(doi: 10.1214). The Supplement (Casarin, Craiu, Frattarolo, Robert, 2023) contains some background on antithetic vector transformation theory and proof of the results in the paper.

REFERENCES

AHN, J. Y. and FUCHS, S. (2020). On Minimal Copulas under the Concordance Order. *Journal of Optimization Theory and Applications* **184** 762–780.

ALBONICO, A., CALÉS, L., CARDANI, R., CROITOROV, O., FERRONI, F., GIOVANNINI, M., HOHBERGER, S., PATARACCHIA, B., PERICOLI, F. M., RACIBORSKI, R., RATTO, M., ROEGER, W. and VOGEL, L. (2019). Comparing post-crisis dynamics across Euro Area countries with the Global Multi-country model. *Economic Modelling* **81** 242-273.

ALIPRANTIS, C. D. and BORDER, K. C. (2007). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

ANDRÉASSON, I. J. (1972). Combinations of antithetic methods in simulation Technical Report No. 72.49, Kungliga Tekniska Högskolan. Department of Information Processing. Computer Science.

ANDRÉASSON, I. J. and DAHLQUIST, G. (1972). Groups of antithetic transformations in simulation Technical Report No. 72.57, Kungliga Tekniska Högskolan. Department of Information Processing. Computer Science.

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725.

ARVIDSEN, N. I. and JOHNSSON, T. (1982). Variance reduction through negative correlation, a simulation study. *Journal of Statistical Computation and Simulation* **15** 119–127.

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. *International series in decision processes*. Holt, Rinehart and Winston.

BIRKHOFF, G. (1946). Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A* **5** 147–151.

BROWN, J. (1966). Approximation theorems for Markov operators. *Pacific Journal of Mathematics* **16** 13–23.

BUBENIK, O. and HOLBROOK, J. (2007). Densities for random balanced sampling. *Journal of Multivariate Analysis* **98** 350 - 369.

CHAN, D., KOHN, R. and KIRBY, C. (2006). Multivariate Stochastic Volatility Models with Correlated Errors. *Econometric Reviews* **25** 245-274.

CRAIU, R. V. and MENG, X. L. (2005). Multiprocess Parallel Antithetic Coupling for Backward and Forward Markov Chain Monte Carlo. *The Annals of Statistics* **33** 661-697.

CRAIU, R. V. and MENG, X.-L. (2006). Meeting Hausdorff in Monte Carlo: A surprising tour with antihype fractals. *Statistica Sinica* **16** 77–91.

DADUNA, H. and SZEKLI, R. (2006). Dependence ordering for Markov processes on partially ordered spaces. *Journal of Applied Probability* **43** 793–814.

DURANTE, F., FERNÁNDEZ SÁNCHEZ, J. and TRUTSCHNIG, W. (2014). Multivariate copulas with hairpin support. *Journal of Multivariate Analysis* **130** 323-334.

DURANTE, F. and SANCHEZ, J. F. (2012). On the approximation of copulas via shuffles of Min. *Statistics & Probability Letters* **82** 1761 - 1767.

FISHMAN, G. S. and HUANG, B. D. (1983). Antithetic Variates Revisited. *Commun. ACM* **26** 964–971.

FRÉCHET, M. (1935). Généralisation du théoreme des probabilités totales. *Fundamenta Mathematicae* **25** 379-387.

FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3ê serie, Sciences, Sect. A* **14** 53–77.

FRIGESSI, A., GÅSEMYR, J. and RUE, H. (2000). Antithetic Coupling of Two Gibbs Sampler Chains. *The Annals of Statistics* **28** 1128-1149.

FUCHS, S., MCCORD, Y. and SCHMIDT, K. D. (2018). Characterizations of copulas attaining the bounds of multivariate Kendall's tau. *Journal of Optimization Theory and Applications* **178** 424–438.

GAFFKE, N. and RÜSCHENDORF, L. (1981). On a class of extremal problems in statistics. *Mathematische Operationsforschung und Statistik. Series Optimization* **12** 123-135.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85** 398-409.

GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 509-579.

GEROW, K. and HOLBROOK, J. (1996). Statistical sampling and fractal distributions. *Math. Intelligencer* **18** 12–22.

HAMMERSLEY, J. M. and MAULDON, J. G. (1956). General principles of antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society* **52** 476–481.

HAMMERSLEY, D. C. and MORTON, K. V. (1956). A new Monte Carlo technique: antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society* **52** 449–475.

HANDSCOMB, D. C. (1958). Proof of the antithetic variates theorem for n > 2. *Mathematical Proceedings of the Cambridge Philosophical Society* **54** 300–301.

HARDY, G. H., COLLECTION, K. M. R., LITTLEWOOD, J. E., PÓLYA, G., PÓLYA, G. and LITTLEWOOD, D. E. (1934). *Inequalities*. Cambridge University Press.

HOEFFDING, W. (1940). Masstabinvariante Korrelationtheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universitat Berlin* **5** 179-233.

ISBELL, J. R. (1955). Birkhoff's problem 111. *Proceedings of the American Mathematical Society* **6** 217–218.

JOE, H. (1990). Multivariate concordance. *Journal of Multivariate Analysis* **35** 12 - 30.

KNOTT, M. and SMITH, C. (2006). Choosing joint distributions so that the variance of the sum is small. *Journal of Multivariate Analysis* **97** 1757 - 1765.

LEE, W. and AHN, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance: Mathematics and Economics* **56** 68 - 79.

LEE, W., CHEUNG, K. C. and AHN, J. Y. (2017). Multivariate countermonotonicity and the minimal copulas. *Journal of Computational and Applied Mathematics* **317** 589–602.

LÉVY, P. (1962). Extensions d'un théorème de D. Dugué et M. Girault. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **1** 159–173.

LINDENSTRAUSS, J. (1965). A Remark on Extreme Doubly Stochastic Measures. *The American Mathematical Monthly* **72** 379–382.

LOH, W. L. (1996). On Latin hypercube sampling. *The Annals of Statistics* **24** 2058–2080.

LU, T.-Y. and YI, Z. (2004). Generalized correlation order and stop-loss order. *Insurance: Mathematics and Economics* **35** 69–76.

MCKAY, M. D., BECKMAN, R. J. and CONOVER, W. J. (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.

MIRA, A. (2002). Ordering and improving the performance of Monte Carlo Markov Chains. *Statistical Science* **16** 340-350.

MOAMENI, A. (2016). Supports of Extremal Doubly Stochastic Measures. *Canadian Mathematical Bulletin* **59** 381–391.

MÜLLER, A. and SCARSINI, M. (2000). Some Remarks on the Supermodular Order. *Journal of Multivariate Analysis* **73** 107-119.

NELSEN, R. B. and ÚBEDA-FLORES, M. (2012). Directional dependence in multivariate distributions. *Annals of the Institute of Statistical Mathematics* **64** 677–685.

OWEN, A. B. (1992). A Central Limit Theorem for Latin Hypercube Sampling. *Journal of the Royal Statistical Society: Series B (Methodological)* **54** 541-551.

OWEN, A. B. (2003). The Dimension Distribution, and Quadrature Test Functions. *Statistica Sinica* **13** 1–17.

PUCCETTI, G. and WANG, R. (2015). Extremal Dependence Concepts. *Statistical Science* **30** 485 – 517.

ROACH, W. and WRIGHT, R. (1977). Optimal antithetic sampling plans. *Journal of Statistical Computation and Simulation* **5** 99-114.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Landmarks in Mathematics and Physics*. Princeton University Press.

RUBINSTEIN, Y. R. and SAMORODNITSKY, G. (1987). A Modified Version of Handscomb's Antithetic Variates Theorem. *SIAM Journal on Scientific and Statistical Computing* **8** 82-98.

RÜSCHENDORF, L. (2013). *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*. *Springer Series in Operations Research and Financial Engineering*. Springer Berlin Heidelberg.

RÜSCHENDORF, L. and UCKELMANN, L. (2002). *Variance Minimization and Random Variables with Constant Sum* In *Distributions With Given Marginals and Statistical Modelling* 211–222. Springer Netherlands, Dordrecht.

SHEPP, L. (1962). Symmetric random walk. *Transactions of the American Mathematical Society* **104** 144–153.

STEIN, M. (1987). Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics* **29** 143–151.

TANG, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association* **88** 1392-1397.

TUKEY, J. W. (1957). Antithesis or regression? *Mathematical Proceedings of the Cambridge Philosophical Society* **53** 923–924.

VAN DYK, D. A. and MENG, X. L. (2001). The Art of Data Augmentation. *Journal of Computational and Graphical Statistics* **10** 1-50.

VITALE, R. A. (1990). On Stochastic Dependence and a Class of Degenerate Distributions. *Topics in Statistical Dependence* **16** 459.

VON NEUMANN, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games* **2** 5–12.

WANG, X. and FANG, K.-T. (2003). The effective dimension and quasi-Monte Carlo integration. *Journal of Complexity* **19** 101-124.

WANG, X. and SLOAN, I. H. (2005). Why are high-dimensional finance problems often of low effective dimension? *SIAM Journal on Scientific Computing* **27** 159–183.

WANG, B. and WANG, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis* **102** 1344 - 1360.

WHITT, W. (1976). Bivariate Distributions with Given Marginals. *The Annals of Statistics* **4** 1280 – 1289.

WILSON, J. R. (1979). Proof of the antithetic-variates theorem for unbounded functions. *Mathematical Proceedings of the Cambridge Philosophical Society* **86** 477–479.

WILSON, J. R. (1983). Antithetic Sampling with Multivariate Inputs. *American Journal of Mathematical and Management Sciences* **3** 121-144.