

# Quasi-random numbers for copula models

Mathieu Cambou · Marius Hofert · Christiane Lemieux

the date of receipt and acceptance should be inserted later

**Abstract** The present work addresses the question how sampling algorithms for commonly applied copula models can be adapted to account for quasi-random numbers. Besides sampling methods such as the conditional distribution method (based on a one-to-one transformation), it is also shown that typically faster sampling methods (based on stochastic representations) can be used to improve upon classical Monte Carlo methods when pseudo-random number generators are replaced by quasi-random number generators. This opens the door to quasi-random numbers for models well beyond independent margins or the multivariate normal distribution. Detailed examples (in the context of finance and insurance), illustrations and simulations are given and software has been developed and provided in the R packages `copula` and `qrng`.

**Keywords** Quasi-random numbers · copulas · conditional distribution method · Marshall–Olkin algorithm · tail events · risk measures

**Mathematics Subject Classification (2000)** 62H99 · 65C60

## 1 Introduction

In many applications, in particular in finance and insurance, the quantities of interest can be written as  $\mathbb{E}[\Psi_0(\mathbf{X})]$ , where  $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$  is a random vector with distribution function  $H$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and

---

Mathieu Cambou  
Institute of Mathematics, Station 8, EPFL, 1015 Lausanne, Switzerland,  
E-mail: mathieu.cambou@epfl.ch

Marius Hofert  
Department of Statistics and Actuarial Science, University of Waterloo,  
Canada, E-mail: marius.hofert@uwaterloo.ca

Christiane Lemieux  
Department of Statistics and Actuarial Science, University of Waterloo,  
Canada, E-mail: clemieux@uwaterloo.ca

$\Psi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$  is a measurable function. The components of  $\mathbf{X}$  are typically dependent. To account for this dependence, the distribution of  $\mathbf{X}$  can be modeled by

$$H(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d, \quad (1)$$

where  $F_j(x) = \mathbb{P}(X_j \leq x)$ ,  $j \in \{1, \dots, d\}$ , are the marginal distribution functions of  $H$  and  $C : [0, 1]^d \rightarrow [0, 1]$  is a *copula*, i.e., a distribution function with standard uniform univariate margins; for an introduction to copulas, see (McNeil et al 2005, Chapter 5), Nelsen (2006) or Joe (2014). A copula model such as (1) allows one to separate the dependence structure from the marginal distributions. This is especially useful in the context of model building and sampling in the case where  $\mathbb{E}[\Psi_0(\mathbf{X})]$  mainly depends on the dependence between the components of  $\mathbf{X}$ , so on  $C$ ; for examples of this type, see Section 5.

In terms of copula model (1), we may then write

$$\mathbb{E}[\Psi_0(\mathbf{X})] = \mathbb{E}[\Psi(\mathbf{U})]$$

where  $\mathbf{U} = (U_1, \dots, U_d) : \Omega \rightarrow \mathbb{R}^d$  is a random vector with distribution function  $C$ ,  $\Psi : [0, 1]^d \rightarrow \mathbb{R}$  is defined as

$$\Psi(u_1, \dots, u_d) = \Psi_0(F_1^-(u_1), \dots, F_d^-(u_d)),$$

and  $F_j^-(p) = \inf\{x \in \mathbb{R} : F_j(x) \geq p\}$ ,  $j \in \{1, \dots, d\}$ , are the marginal quantile functions. If  $C$  and the margins  $F_j$ ,  $j \in \{1, \dots, d\}$ , are known, we can use Monte Carlo simulation to estimate  $\mathbb{E}[\Psi(\mathbf{U})]$ . For a (pseudo-)random sample  $\{\mathbf{U}_i : i = 1, \dots, n\}$  from  $C$ , the (classical) Monte Carlo estimator of  $\mathbb{E}[\Psi(\mathbf{U})]$  is given by

$$\frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i) \approx \mathbb{E}[\Psi(\mathbf{U})].$$

The main challenge of a Monte Carlo simulation is thus the sampling of the copula. This challenge also holds for quasi-Monte Carlo (QMC) methods, and is actually amplified by

the fact that these methods are more sensitive to certain properties of the function  $\Psi$ . Thus the choice of the construction method of a stochastic representation for  $C$  can have complex effects on the performance of QMC methods, a feature that does not show up when using Monte Carlo methods. The present work includes a careful analysis of these effects, as they must be thoroughly understood in order to successfully replace pseudorandom numbers by quasi-random numbers into different copula sampling algorithms.

Let us briefly summarize the idea behind QMC methods and how they can be used for copula models; more precise definitions on some of the concepts used here will be given later. The idea is to start with a so-called *low-discrepancy point set*  $P_n = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq [0, 1]^k$ , with  $k \geq d$ , which is designed so that its empirical distribution over  $[0, 1]^k$  is closer (in a sense to be defined later) to the uniform distribution  $U[0, 1]^k$  than a set of independent and identically (i.i.d.) random points is. Assuming that for  $\mathbf{U}' \sim U[0, 1]^k$  we have a transformation  $\phi_C$  such that  $\phi_C(\mathbf{U}') \sim C$ , we can then construct the approximation

$$\frac{1}{n} \sum_{i=1}^n \Psi(\phi_C(\mathbf{v}_i)) \approx \mathbb{E}[\Psi(\mathbf{U})]. \quad (2)$$

Figure 1 shows the points  $\phi_C(\mathbf{v}_i)$  obtained using either pseudo-random or quasi-random numbers, for a transformation  $\phi_C$  designed for the *Clayton copula*.

QMC methods are typically used to approximate integrals of functions over the unit cube via

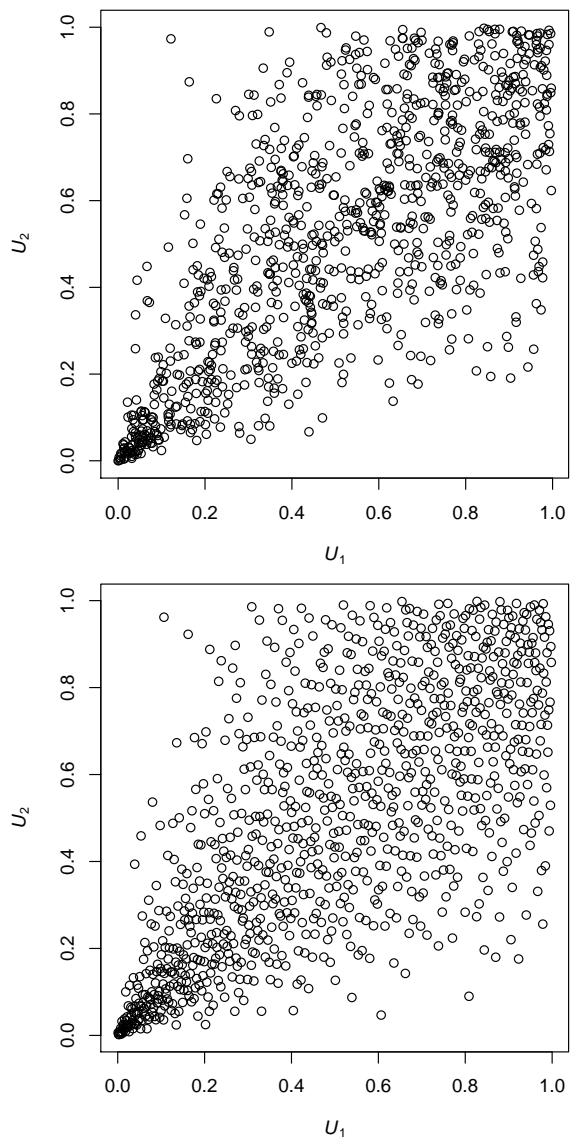
$$Q_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{v}_i) \approx \int_{[0,1]^k} f(\mathbf{v}) d\mathbf{v} = I(f). \quad (3)$$

A widely used upper bound for the integration error  $|I(f) - Q_n|$  is the Koksma–Hlawka inequality (see, e.g., Niederreiter (1992)), which is of the form  $V(f)D^*(P_n)$ , where  $V(f)$  measures the variation of  $f$  in the sense of Hardy and Krause, while  $D^*(P_n)$  measures the discrepancy of  $P_n$ , i.e., how far  $P_n$  is from  $U[0, 1]^k$ .

To analyze the properties of the QMC approximation (2) for  $\mathbb{E}[\Psi(\mathbf{U})]$ , there are two possible approaches. The first one is to define  $f = \Psi \circ \phi_C$  and work within the traditional framework given by (3), the Koksma–Hlawka inequality with this composed function and the low-discrepancy point set  $P_n$ . The second one is to think of (2) as approximating

$$\mathbb{E}[\Psi(\mathbf{U})] = \int_{[0,1]^d} \Psi(\mathbf{u}) dC(\mathbf{u})$$

and work with generalizations of the Koksma–Hlawka inequality that apply to measures other than the Lebesgue measure; see Hlawka and Mück (1972) or Aistleitner and Dick (2015). In the latter case, we work with the function  $\Psi$  and view the transformation  $\phi_C$  as one that is applied to the low-discrepancy point set  $P_n$  rather than to  $\Psi$ . That is, here we work with the transformed point set  $\tilde{P}_n = \{\phi_C(\mathbf{v}_1), \dots, \phi_C(\mathbf{v}_n)\}$



**Figure 1** 1000 realizations of a bivariate Clayton copula with  $\theta = 2$  (Kendall's tau equals 0.5), generated by a pseudo-random number generator (top) and by a quasi-random number generator (bottom).

and analyze its quality via measures of discrepancy that quantify its distance to  $C$  rather than comparing  $P_n$  to  $U[0, 1]^k$ .

QMC methods have been used in a variety of applications, but so far most of the problems considered have been such that the stochastic models used can be formulated using independent random variables (e.g., a vector of dependent normal variates can be written as a linear transformation of independent normal variates). In such cases, the transformation of the uniform vector  $\mathbf{v}$  into observations from the desired stochastic model can be easily obtained by transforming each component  $v_j$  of  $\mathbf{v}$  using the inverse transform method, which is deemed to work well with QMC in part because of its monotonicity, and also because it corresponds to an overall one-to-one transformation from  $[0, 1]^d$  to  $\mathbb{R}^d$ .

In the more general copula setting considered in this paper, at first sight the so-called conditional distribution method (CDM) (which is the inverse of the copula-based version of the Rosenblatt transform) appears to be a good choice to use with quasi-random numbers, as it is the direct multivariate extension of the inverse transform mentioned in the previous paragraph. Namely, it is a one-to-one transformation that maps  $[0, 1]^d$  to  $[0, 1]^d$  and it is monotone in each variable. A transformation with  $k = d$  is certainly desirable (and preferable to a many-to-one transformation with  $k > d$ ) when used in conjunction with QMC methods, since these methods do better on problems of lower dimension. Also, intuitively the monotonicity should be helpful to preserve the smoothness of  $\Psi$  (for the first approach) and the low-discrepancy of  $P_n$  (for the second approach). An additional advantage of the CDM is that it is applicable to any copula  $C$  (and the only known algorithm such general). However, the involved (inverses of the) conditional copulas are often challenging to evaluate which has led to other sampling algorithms being more frequently used. An example is the Marshall–Olkin algorithm for sampling Archimedean copulas, which we also address in this work.

The paper is organized as follows. Section 2 provides a short introduction to quasi-random numbers. Section 3 shows how quasi-random samples from various copulas (and thus multivariate models with these dependence structures) can be obtained using different sampling algorithms. Detailed examples are given. In Section 4 we discuss the theoretical background supporting each of the two approaches mentioned earlier to analyze the use of low-discrepancy sequences for copula sampling. Numerical results are provided in Section 5. Finally, Section 6 includes concluding remarks and a discussion of future work. Note that most results and figures presented in this paper (as well as additional experiments conducted) can be found in the R packages `copula` (see the vignette `qrng`) and `qrng` (see `demo(basket_options)` and `demo(test_functions)`).

## 2 Quasi-random numbers

Here we assume that a random sample  $\{\mathbf{U}_i : i = 1, \dots, n\}$  from a copula  $C$  can be generated by transforming a random sample  $\{\mathbf{U}'_i : i = 1, \dots, n\}$  from  $U[0, 1]^k$  with  $k \geq d$ ; several algorithms for copula models fall under this setup. Due to the independence of the vectors  $\mathbf{U}'_i$ , realizations of the sample  $\{\mathbf{U}'_i : i = 1, \dots, n\}$  (obtained by so-called pseudo-random number generators (PRNGs)) will inevitably show regions of  $[0, 1]^k$  which are lacking points and other areas of  $[0, 1]^k$  which contain more samples than expected by the uniform distribution. To reduce this problem of an inhomogeneous concentration of samples, quasi-random number generators (QRNGs) do not aim at mimicking i.i.d. samples but instead at producing a homogeneous coverage of  $[0, 1]^k$ .

The homogeneity of a sequence of points over the unit hypercube can be measured by its discrepancy, which relates to the error incurred by representing the (Lebesgue-)measure of subsets of the unit hypercube by the fraction of points in these subsets. Quasi-random sequences aim at achieving smaller discrepancy than pseudo-random number sequences and are thus also called *low-discrepancy sequences*. The rest of this section reviews concepts related to QRNG that are used in this paper. The reader is referred to Niederreiter (1992), Morokoff (1998), (Glasserman 2004, Chapter 5), and Dick and Pillichshammer (2010) for more details.

### 2.1 Discrepancy

The notion of discrepancy applies to sequences of points  $X = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$  in the unit hypercube  $[0, 1]^k$ . Denote by  $P_n = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq [0, 1]^k$  the first  $n$  points of the sequence. Let  $\mathcal{J}^*$  be the set of intervals of  $[0, 1]^k$  of the form  $[\mathbf{0}, \mathbf{z}] = \prod_{j=1}^k [0, z_j]$ , where  $0 < z_j \leq 1, j = 1, \dots, k$ . Then the *discrepancy function* of  $P_n$  on an interval  $[\mathbf{0}, \mathbf{z}]$  is the difference

$$E([\mathbf{0}, \mathbf{z}]; P_n) = \frac{A([\mathbf{0}, \mathbf{z}]; P_n)}{n} - \lambda([\mathbf{0}, \mathbf{z}]),$$

where  $A([\mathbf{0}, \mathbf{z}]; P_n) = \#\{i; 1 \leq i \leq n, \mathbf{v}_i \in [\mathbf{0}, \mathbf{z}]\}$  is the number of points from  $P_n$  that fall in  $[\mathbf{0}, \mathbf{z}]$  and  $\lambda([\mathbf{0}, \mathbf{z}]) = \prod_{j=1}^k z_j$  is the Lebesgue measure of  $[\mathbf{0}, \mathbf{z}]$ .

The *star discrepancy*  $D^*$  of  $P_n$  is defined by

$$D^*(P_n) = \sup_{[\mathbf{0}, \mathbf{z}] \in \mathcal{J}^*} |E([\mathbf{0}, \mathbf{z}]; P_n)|.$$

An infinite sequence  $X$  satisfying  $D^*(P_n) \in O(n^{-1} \log^k n)$  is said to be a *low-discrepancy sequence*.

For a function  $\Psi : [0, 1]^k \rightarrow \mathbb{R}$ , we have the well-known Koksma–Hlawka error bound given by

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{v}_i) - \mathbb{E}[\Psi(\mathbf{U}')] \right| \leq V(\Psi) D^*(P_n), \quad (4)$$

where  $\mathbf{U}' \sim U[0, 1]^k$  and  $V(\Psi)$  denotes the variation of the function  $\Psi$  in the sense of Hardy and Krause. See Owen (2005) for a detailed account of the notion of variation and its applicability in practice. We also refer the interested reader to Hartinger et al (2004) and Sobol' (1973) for results handling unbounded functions (and thus of unbounded variation).

### 2.2 Low-discrepancy sequences

There are two main approaches for constructing low-discrepancy sequences: integration lattices and digital sequences. Only the latter are used in this paper, so our discussion will focus on those.

Digital sequences contain the well-known constructions of Sobol' (1967), Faure (1982), and Niederreiter (1987), and

are also closely related to the sequence proposed in Halton (1960). The basic building block for this construction is the *van der Corput sequence in base  $b \geq 2$* , defined as

$$S_b(i) = \sum_{r=0}^{\infty} a_r(i) b^{-r-1}, \quad i \in \mathbb{N}, \quad (5)$$

where  $a_r(i)$  is the  $r$ th digit of the  $b$ -adic expansion of  $i - 1 = \sum_{r=0}^{\infty} a_r(i) b^r$ . To construct a sequence of points in  $[0, 1]^k$  from this one-dimensional sequence, one approach is the one proposed by Halton (1960), which consists of choosing  $k$  pairwise relatively prime integers  $b_1, \dots, b_k$  and defining the  $i$ th point of the sequence as

$$\mathbf{v}_i = (S_{b_1}(i), \dots, S_{b_k}(i)), \quad i \in \mathbb{N}.$$

Another possibility is to fix the base  $b$ , and choose  $k$  linear transformations that are then applied to the digits  $a_r(i)$  from the expansion of  $i - 1$  before being used in (5) to define a real number between 0 and 1. More precisely, let  $M_1, \dots, M_k$  be (unbounded) “ $\infty \times \infty$ ” matrices with entries in  $\mathbb{Z}_b$  and let

$$S_b^{M_j}(i) = \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} m_{r+1, l+1} a_l(i) b^{-r-1}, \quad (6)$$

where  $m_{r,l}$  is the element in the  $r$ th row and  $l$ th column of  $M_j$ . Here we assume for simplicity that  $b$  is prime and all operations in (6) are performed in the finite field  $\mathbb{F}_b$ . One can then define a sequence of points in  $[0, 1]^k$  by taking

$$\mathbf{v}_i = (S_b^{M_1}(i), \dots, S_b^{M_k}(i)) \quad (7)$$

as its  $i$ th point. Sobol’ was the first to propose such a construction, working in base 2 and defining the matrices  $M_1, \dots, M_k$  so that he was able to prove that the obtained sequence has  $D^*(P_n) \in O(n^{-1} \log^k n)$ ; see Sobol’ (1967).

We also point out that Halton sequences can be generalized using the same idea as in (7). That is, one can choose matrices  $M_1, \dots, M_k$  with the elements of  $M_j$  in  $\mathbb{Z}_{b_j}$ , and “scramble” the digits of the expansion of  $i - 1$  before reverting them via (5) to produce a number between 0 and 1. A very simple way to achieve this is via diagonal matrices  $M_j$ , each containing a well-chosen element (or factor) of  $\mathbb{Z}_{b_j}$ . In our numerical experiments, we use such an approach, with the factors provided in Faure and Lemieux (2009); see the R package `qrng` for an implementation.

### 2.3 Randomized quasi-Monte Carlo

In contrast to the error rate  $O(n^{-1/2})$  for Monte Carlo methods based on PRNGs, approximations based on QRNGs have the advantage of having an error in  $O(n^{-1} \log^k n)$  when the function of interest  $\Psi$  is of bounded variation. However, in practice it is also important to be able to estimate the corresponding error. While bounds such as the Koksma–Hlawka

inequality are useful to understand the behaviour of approximations based on quasi-random sequences, they do not provide a practical way to estimate the error. To circumvent this problem, an approach that is often used is to *randomize* a low-discrepancy point set in such a way that its high uniformity (or low discrepancy) is preserved, but at the same time unbiased estimators can be constructed (and sampled) from it. Another advantage of this approach is that variance expressions can be derived and compared with Monte Carlo sampling for wider classes of functions, i.e., not necessarily of bounded variation (see Owen (1997a), Lemieux (2009) and the references therein). This approach gives rise to *randomized quasi-Monte Carlo (RQMC)* methods.

To apply this approach, we need a randomization function  $r: [0, 1]^s \times [0, 1]^k \rightarrow [0, 1]^k$  with  $s \geq k$  such that for any fixed  $\mathbf{v} \in [0, 1]^k$ , we have that if  $\mathbf{U}' \sim U[0, 1]^s$ , then  $r(\mathbf{U}', \mathbf{v}) \sim U[0, 1]^k$ . Hence the individual RQMC samples have the same properties as those from a random sample; the difference lies in the fact that the RQMC samples are dependent.

An early randomization scheme originally proposed by Cranley and Patterson (1976) is to take

$$r(\mathbf{U}', \mathbf{v}) = (\mathbf{v} + \mathbf{U}') \bmod 1. \quad (8)$$

A randomized point set is then obtained by generating a uniform vector  $\mathbf{U}'$  and letting  $\tilde{P}_n(\mathbf{U}') = \{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n\}$ , where  $\tilde{\mathbf{v}}_i = r(\mathbf{U}', \mathbf{v}_i)$ ,  $i \in \{1, \dots, n\}$ . Hence the same shift  $\mathbf{U}'$  is applied to all points in  $P_n$ . If we let  $\mathbf{U}'_1, \dots, \mathbf{U}'_B$  be independent  $U[0, 1]^s$  vectors, we can construct  $B$  i.i.d. unbiased estimators

$$\hat{\mu}_n^l = \frac{1}{n} \sum_{\tilde{\mathbf{v}}_i \in \tilde{P}_n(\mathbf{U}'_i)} \Psi(\phi_C(\tilde{\mathbf{v}}_i)), \quad l \in \{1, \dots, B\}$$

for  $\mathbb{E}[\Psi(\mathbf{U})]$ , whose variances can be estimated via the sample variance of  $\hat{\mu}_n^1, \dots, \hat{\mu}_n^B$ .

In addition to the simple random shift described in (8), several other randomization schemes have been proposed and studied. A popular randomization method for digital nets is to “scramble” them, an idea originally proposed by Owen (1995) and subsequently studied by Owen (1997a), Owen (1997b), Owen (2003), Matousěk (1998) and Hong and Hickernell (2003), among others.

A simpler randomization for digital nets is to use the digital counterpart of (8), where instead of adding two real numbers modulo 1, we add (in  $\mathbb{Z}_b$ ) the digits of their base  $b$  expansion. That is, for  $u = \sum_{r=0}^{\infty} u_r b^{-r-1}$  and  $v = \sum_{r=0}^{\infty} v_r b^{-r-1}$ , we let

$$u \oplus_b v = \sum_{r=0}^{\infty} ((u_r + v_r) \bmod b) b^{-r-1}$$

and define  $r(\mathbf{u}, \mathbf{v}) = \mathbf{u} \oplus_b \mathbf{v}$ , where the  $\oplus_b$  operation is applied component-wise to the  $k$  coordinates of  $\mathbf{u}$  and  $\mathbf{v}$ . The same idea can be applied to randomize Halton sequences (as shown, e.g., in Lemieux (2009)), but where a different base  $b$  is used

in each of the  $k$  coordinates. Digital shifts for the Sobol' and generalized Halton sequences are available in our R package `qrng`.

### 3 Quasi-random copula samples

Sampling procedures for a  $d$ -dimensional copula  $C$  can be viewed as transformations  $\phi_C : [0, 1]^k \rightarrow [0, 1]^d$  for some  $k \geq d$ , such that, for  $\mathbf{U}' \sim U[0, 1]^k$ ,  $\mathbf{U} := \phi_C(\mathbf{U}') \sim C$ ; that is,  $\phi_C$  transforms independent  $U[0, 1]$  random variables to  $d$  dependent random variables with distribution function  $C$ .

The case  $k = d$  is mostly known and applied as *conditional distribution method (CDM)* and involves the inversion method for sampling univariate conditional copulas (although, for example, for Archimedean copulas another transformations  $\phi_C$  with  $k = d$  is known; see Wu et al (2006)). This approach thus naturally uses  $d$  independent  $U[0, 1]$  random variables as input. The case  $k \geq d$  (often:  $k > d$ ) is typically known as *stochastic representation* and is usually based on sampling  $k$  univariate random variables from elementary probability distributions, as we will see in Section 3.2.

In what follows we consider the above two approaches and show how they can be adapted to quasi-random number generation.

#### 3.1 Conditional distribution method and other one-to-one transformations ( $k = d$ )

##### 3.1.1 Theoretical background

The only known general sampling approach which works for any copula is the CDM. For  $j \in \{2, \dots, d\}$ , let

$$\begin{aligned} C_{j|1\dots j-1} &= C(u_j | u_1, \dots, u_{j-1}) \\ &= \mathbb{P}(U_j \leq u_j | U_1 = u_1, \dots, U_{j-1} = u_{j-1}) \end{aligned}$$

denote the *conditional copula of  $U_j$  at  $u_j$  given  $U_1 = u_1, \dots, U_{j-1} = u_{j-1}$* . If

$$C_{j|1\dots j-1}^- = C^-(u_j | u_1, \dots, u_{j-1})$$

denotes the corresponding quantile function, the CDM is given as follows; see Embrechts et al (2003) or (Hofert 2010, p. 45).

**Theorem 1 (Conditional distribution method)** *Let  $C$  be a  $d$ -dimensional copula,  $\mathbf{U}' \sim U[0, 1]^d$ , and  $\phi_C^{CDM}$  be given by*

$$\begin{aligned} U_1 &= U'_1, \\ U_2 &= C_{2|1}^-(U'_2 | U_1), \\ &\vdots \\ U_d &= C_{d|1\dots d-1}^-(U'_d | U_1, \dots, U_{d-1}). \end{aligned}$$

*Then  $\mathbf{U} = (U_1, \dots, U_d) = \phi_C^{CDM}(\mathbf{U}') \sim C$ .*

To find the conditional copulas  $C_{j|1\dots j-1}$ , for  $j \in \{2, \dots, d\}$ , for a specific copula  $C$ , the following result (which holds under mild assumptions) is often applied. A rigorous proof can be found in (Schmitz 2003, p. 20), an implementation is provided by the function `cCopula()` in the R package `copula`. The corollary that follows is an immediate consequence of Sklar's theorem, for example.

**Theorem 2 (Computing conditional copulas)** *Let  $C$  be a  $d$ -dimensional copula, which, for  $d \geq 3$ , admits continuous partial derivatives with respect to the first  $d - 1$  arguments. For  $j \in \{2, \dots, d\}$  and  $u_l \in [0, 1]$ ,  $l \in \{1, \dots, j\}$ ,*

$$\begin{aligned} C_{j|1\dots j-1} &= \frac{D_{j-1\dots 1} C_{1\dots j}}{D_{j-1\dots 1} C_{1\dots j-1}} \\ &= \frac{D_{j-1\dots 1} C_{1\dots j}}{c_{1\dots j-1}}, \end{aligned} \quad (9)$$

*where  $D_{j-1\dots 1}$  denotes the derivative with respect to the first  $j - 1$  arguments,  $C_{1\dots j}$  denotes the marginal copula corresponding to the first  $j$  components and  $c_{1\dots j-1}$  denotes the density of  $C_{1\dots j-1}$ . If  $C$  admits a density, then (9) equals*

$$C_{j|1\dots j-1} = \frac{\int_0^{u_j} c_{1\dots j}(u_1, \dots, u_{j-1}, z_j) dz_j}{c_{1\dots j-1}}. \quad (10)$$

**Corollary 1 (Conditional copulas for general multivariate distributions)** *Let  $H$  be a  $d$ -dimensional absolutely continuous distribution function with density  $h$ , margins  $F_1, \dots, F_d$  and copula  $C$ . For  $j \in \{2, \dots, d\}$  and  $u_l \in [0, 1]$ ,  $l \in \{1, \dots, j\}$ ,*

$$\begin{aligned} C_{j|1\dots j-1} &= H_{j|1\dots j-1}(F_j^-(u_j) | F_1^-(u_1), \dots, F_{j-1}^-(u_{j-1})) \end{aligned} \quad (11)$$

$$= \frac{\int_{-\infty}^{F_j^-(u_j)} h_{1\dots j}(F_1^-(u_1), \dots, F_{j-1}^-(u_{j-1}), z_j) dz_j}{h_{1\dots j-1}(F_1^-(u_1), \dots, F_{j-1}^-(u_{j-1}))}. \quad (12)$$

##### 3.1.2 Examples

We now present several copula families and show how the corresponding conditional copulas and their inverses can be computed.

##### Elliptical copulas

An *elliptical copula* describes the dependence structure of an elliptical distribution; for the latter, see Cambanis et al (1981), Embrechts et al (2002), Embrechts et al (2003), or (McNeil et al 2005, Sections 3.3, 5). The most prominent two families in the class of elliptical copulas are the Gauss and the  $t$  copulas.

*Gauss copulas.* Gauss copulas are given by

$$C_P^{\text{Ga}}(\mathbf{u}) = \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

where  $\Phi_P$  denotes the  $d$ -variate normal distribution function with location vector  $\mathbf{0}$  and scale matrix  $P$  (a correlation matrix) and  $\Phi^{-1}$  is the standard normal quantile function. Consider the dimension to be  $j$  and let  $\mathbf{X} \sim \Phi_P$  with  $\mathbf{X} = (\mathbf{X}_{1:(j-1)}, X_j)$ . Furthermore, assume

$$P = \begin{pmatrix} P_{1:(j-1),1:(j-1)} & P_{1:(j-1),j} \\ P_{j,1:(j-1)} & P_{j,j} \end{pmatrix}$$

to be positive definite. It follows from (Fang et al 1990, p. 45 and 78) that

$$X_j | \mathbf{X}_{1:(j-1)} = \mathbf{x}_{1:(j-1)} \sim \mathcal{N}(\mu_{j|1:(j-1)}(\mathbf{x}_{1:(j-1)}), P_{j|1:(j-1)}),$$

where

$$\begin{aligned} \mu_{j|1:(j-1)}(\mathbf{x}_{1:(j-1)}) &= P_{j,1:(j-1)}(P_{1:(j-1),1:(j-1)})^{-1} \mathbf{x}_{1:(j-1)}, \\ P_{j|1:(j-1)} &= P_{j,j} - P_{j,1:(j-1)}(P_{1:(j-1),1:(j-1)})^{-1} P_{1:(j-1),j}; \end{aligned} \quad (13)$$

so  $H_{j|1\dots j-1}$  is again normal. With  $\Phi^{-1}(\mathbf{u}_{1:(j-1)}) = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_{j-1}))$ , it follows from (11) that

$$\begin{aligned} C_{j|1\dots j-1} &= H_{j|1\dots j-1}(\Phi^{-1}(u_j) | \Phi^{-1}(\mathbf{u}_{1:(j-1)})) \\ &= \Phi_{\mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})), P_{j|1:(j-1)}}(\Phi^{-1}(u_j)) \\ &= \Phi\left(\frac{\Phi^{-1}(u_j) - \mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)}))}{\sqrt{P_{j|1:(j-1)}}}\right) \end{aligned}$$

and thus that

$$\begin{aligned} C_{j|1\dots j-1}^- &= \Phi\left(\Phi^{-1}\left(\mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})), P_{j|1:(j-1)}\right)(u_j)\right) \\ &= \Phi\left(\mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})) + \sqrt{P_{j|1:(j-1)}}\Phi^{-1}(u_j)\right). \end{aligned}$$

An implementation of this inverse is provided via `cCopula` (, `inverse=TRUE`) in the R package `copula`.

*t copulas.*  $t$  copulas are given by

$$C_{v,P}^t(\mathbf{u}) = t_{v,P}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d)),$$

where  $t_{v,P}$  denotes the  $d$ -variate  $t_v$  distribution function with location vector  $\mathbf{0}$  and scale matrix  $P$  (a correlation matrix) and  $t_v^{-1}$  is the standard  $t_v$  quantile function. The following proposition provides formulas for the conditional copulas and their inverses; see (Fang et al 1990, Section 2.5) and (Joe 2014, Section 2.7) for more details.

**Proposition 1 (Conditional  $t$  copulas and inverses)** *With the notation as in the Gauss case and*

$$P^{-1} = \begin{pmatrix} P_{1:(j-1),1:(j-1)}^{-1} & P_{1:(j-1),j}^{-1} \\ P_{j,1:(j-1)}^{-1} & P_{j,j}^{-1} \end{pmatrix},$$

*the conditional  $t$  copula at  $u_j$ , given  $u_1, \dots, u_{j-1}$ , and its inverse are given by*

$$C_{j|1\dots j-1} = t_{v+j-1}\left(s_1\left(\sqrt{P_{j,j}^{-1}}t_v^{-1}(u_j) + s_2\right)\right),$$

$$C_{j|1\dots j-1}^- = t_v\left(\frac{t_{v+j-1}^{-1}(u_j)/s_1 - s_2}{\sqrt{P_{j,j}^{-1}}}\right),$$

*where  $s_1 = s_1(\mathbf{x}_{1:(j-1)})$  and  $s_2 = s_2(\mathbf{x}_{1:(j-1)})$  for  $\mathbf{x}_{1:(j-1)} = (t_v^{-1}(u_1), \dots, t_v^{-1}(u_{j-1}))$  and*

$$\begin{aligned} s_1(\mathbf{x}_{1:(j-1)}) &= \sqrt{\frac{v+j-1}{\mathbf{v} + \mathbf{x}_{1:(j-1)}^\top (P_{1:(j-1),1:(j-1)})^{-1} \mathbf{x}_{1:(j-1)}}}, \\ s_2(\mathbf{x}_{1:(j-1)}) &= \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),j}^{-1} / \sqrt{P_{j,j}^{-1}}. \end{aligned}$$

Figure 2 displays 1000 samples from a  $t$  copula with three degrees of freedom and correlation parameter  $\rho = P_{1,2} = 1/\sqrt{2}$  (Kendall's tau equals 0.5), once drawn with a PRNG (top) and once with a QRNG (bottom). We can visually confirm in this case that the low discrepancy of the latter is preserved. How this seemingly good feature translates into better estimators of the form (2) will be studied further through the theoretical results of Section 4 and the numerical experiments of Section 5.

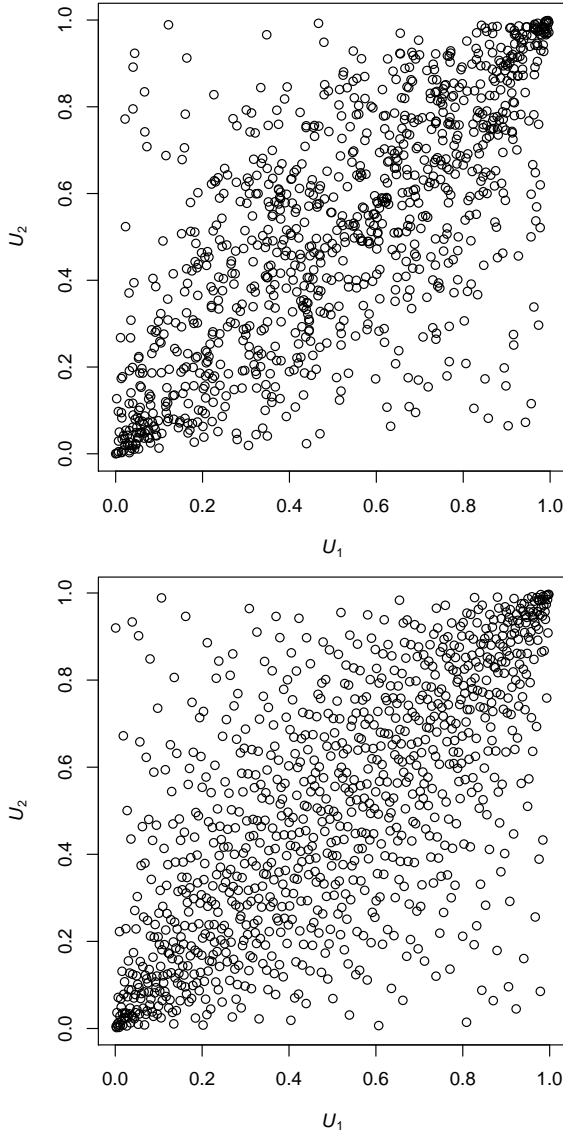
*Archimedean copulas*

An (*Archimedean*) *generator* is a continuous, decreasing function  $\psi: [0, \infty] \rightarrow [0, 1]$  which satisfies  $\psi(0) = 1$ ,  $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$ , and which is strictly decreasing on  $[0, \inf\{t : \psi(t) = 0\}]$ . A  $d$ -dimensional copula  $C$  is called *Archimedean* if it permits the representation

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)),$$

where  $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ , and for some generator  $\psi$  with inverse  $\psi^{-1}: [0, 1] \rightarrow [0, \infty]$ , where  $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$ . For applications and the importance of Archimedean copulas in the realm of finance and insurance, see, e.g., Hofert et al (2013).

McNeil and Nešlehová (2009) show that a generator defines an Archimedean copula if and only if  $\psi$  is  $d$ -monotone, meaning that  $\psi$  is continuous on  $[0, \infty]$ , admits derivatives  $\psi^{(l)}$  up to the order  $l = d - 2$  satisfying  $(-1)^l \psi^{(l)}(t) \geq 0$  for all  $l \in \{0, \dots, d - 2\}$ ,  $t \in (0, \infty)$ , and  $(-1)^{d-2} \psi^{(d-2)}(t)$  is decreasing and convex on  $(0, \infty)$ .



**Figure 2** 1000 realizations of a  $t$  copula with three degrees of freedom and correlation parameter  $\rho = 1/\sqrt{2}$  (Kendall's tau equals 0.5), generated by a PRNG (top) and by a QRNG (bottom).

Assuming  $\psi$  to be sufficiently often differentiable, conditional Archimedean copulas follow from Theorem 2 and are given by

$$C_{j|1\dots j-1} = \frac{\psi^{(j-1)}\left(\sum_{l=1}^j \psi^{-1}(u_l)\right)}{\psi^{(j-1)}\left(\sum_{l=1}^{j-1} \psi^{-1}(u_l)\right)}, \quad (14)$$

where  $u_l \in [0, 1]$ ,  $l \in \{1, \dots, j\}$ , and thus

$$C_{j|1\dots j-1}^- = \psi\left(\psi^{(j-1)-1}\left(u_j \psi^{(j-1)}\left(\sum_{l=1}^{j-1} \psi^{-1}(u_l)\right)\right) - \sum_{l=1}^{j-1} \psi^{-1}(u_l)\right). \quad (15)$$

The generator derivatives  $\psi^{(j-1)}$  and their inverses  $\psi^{(j-1)-1}$  can be challenging to compute. The former are known explicitly for several Archimedean families and certain generator transformations; see Hofert et al (2012) for more details. To compute the inverses, one can use numerical root-finding on  $[0, 1]$ ; see `cCopula(..., inverse=TRUE)` in the R package `copula`. This can be applied, e.g., in the case of Gumbel copulas.

The following example shows the case of a Clayton copula family, for which (15) can be given explicitly and thus where the CDM is tractable; this explicit formula is also utilized by `cCopula(..., inverse=TRUE)`.

*Example 1 (Clayton copulas)* If  $\psi(t) = (1+t)^{-1/\theta}$ ,  $t \geq 0$ ,  $\theta > 0$ , denotes a generator of the Archimedean Clayton copula, then  $\psi^{(j)}(t) = (-1)^j (1+t)^{-(j+1/\theta)} \prod_{l=0}^{j-1} (l+1/\theta)$ . Therefore, (14) equals

$$C_{j|1\dots j-1} = \left( \frac{1-j + \sum_{l=1}^j u_l^{-\theta}}{2-j + \sum_{l=1}^{j-1} u_l^{-\theta}} \right)^{j-1/\theta}$$

and (15) equals

$$C_{j|1\dots j-1}^- = \left( 1 + \left( 1 - (j-1) + \sum_{l=1}^{j-1} u_l^{-\theta} \right) \left( u_j^{-\frac{1}{j-1+1/\theta}} - 1 \right) \right)^{-\frac{1}{\theta}}.$$

Figure 1 displays 1000 samples from a Clayton copula with  $\theta = 2$  (Kendall's tau equals 0.5), once drawn with a PRNG (top) and once with a QRNG (bottom).

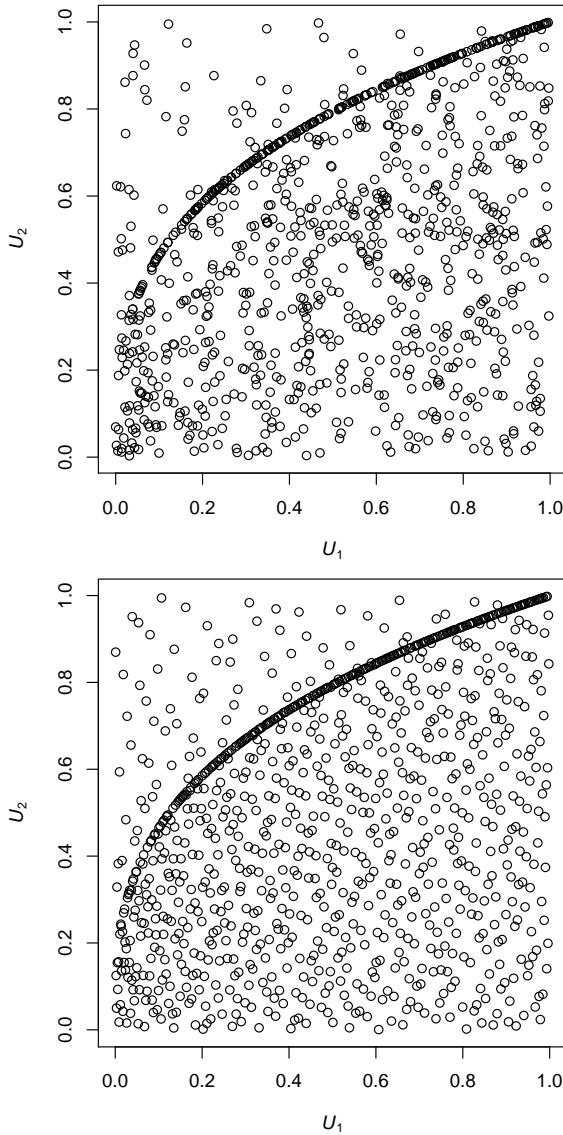
### Marshall–Olkin copulas

A class of bivariate copulas for which  $C_{2|1}^-$  is explicit is the class of Marshall–Olkin copulas  $C(u_1, u_2) = \min\{u_1^{1-\alpha_1} u_2, u_1 u_2^{1-\alpha_2}\}$ ,  $\alpha_1, \alpha_2 \in (0, 1)$ , where one can show that

$$C_{2|1}^- = \begin{cases} \frac{u_1^{\alpha_1}}{1-\alpha_1} u_2, & \text{if } u_2 \in [0, (1-\alpha_1) u_1^{\alpha_1(\frac{1}{\alpha_2}-1)}], \\ u_1^{\alpha_1/\alpha_2}, & \text{if } u_2 \in ((1-\alpha_1) u_1^{\alpha_1(\frac{1}{\alpha_2}-1)}, u_1^{\alpha_1(\frac{1}{\alpha_2}-1)}), \\ u_2^{\frac{1}{1-\alpha_2}}, & \text{if } u_2 \in [u_1^{\alpha_1(\frac{1}{\alpha_2}-1)}, 1]. \end{cases}$$

Figure 3 shows 1000 samples, once drawn from a PRNG (top) and once from a QRNG (bottom). Here again we can visually confirm the low discrepancy.

Another class of copulas not discussed in this section which is naturally sampled with the CDM and thus can easily be adapted to construct corresponding quasi-random numbers is the class of pair copula constructions; see, e.g., Kurowicka and Cooke (2007). For this purpose, we modified the function `RVineSim()` in the R package `VineCopula` (version  $\geq 1.3$ ). It now allows to pass a matrix of quasi-random numbers to be transformed to the corresponding samples from a pair copula



**Figure 3** 1000 realizations of a Marshall–Olkin copula with  $\alpha_1 = 0.25$  and  $\alpha_2 = 0.75$  (Kendall’s tau equals roughly 0.23), generated by a PRNG (top) and by a QRNG (bottom).

construction; see the vignette `qrng` in the R package `copula` for examples. Note that if sampling of the R-vine involves numerical root-finding (required for certain copula families), the corresponding numerical inaccuracy may have an effect on the low discrepancy of the generated samples.

### 3.2 Stochastic representations ( $k \geq d$ , typically $k > d$ )

#### 3.2.1 Theoretical background

As mentioned above, pair-copula constructions are one of the rare copula classes for which the CDM is applied in practice. For most other copula classes and families, faster sampling algorithms derived from stochastic representations of  $\mathbf{U} \sim C$

are known, especially for  $d \gg 2$ . They are mostly class- and family-specific, as can be seen in the examples below.

#### 3.2.2 Examples

##### Elliptical copulas

Gauss and  $t$  copulas are typically sampled via their stochastic representations.

*Gauss copulas.* A random vector  $\mathbf{X} \sim \Phi_P$  admits the stochastic representation  $\mathbf{X} = \mathbf{A}\mathbf{Z}$  where  $\mathbf{A}$  denotes the lower triangular matrix from the Cholesky decomposition  $P = \mathbf{A}\mathbf{A}^\top$  and  $\mathbf{Z}$  is a vector of  $d$  independent standard normal random variables. A random vector  $\mathbf{U} \sim C_P^{\text{Ga}}$  thus admits the stochastic representation  $\Phi(\mathbf{X}) = \Phi(\mathbf{A}\mathbf{Z})$  for  $\mathbf{Z} = (\Phi^{-1}(U'_1), \dots, \Phi^{-1}(U'_d))$  and  $\mathbf{U}' \sim U[0, 1]^d$ ; here  $\Phi$  is assumed to act on  $\mathbf{A}\mathbf{Z}$  component-wise. Note that for Gauss copulas, this sampling approach is equivalent to the CDM.

*t copulas.* A random vector  $\mathbf{X} \sim t_{v,P}$  admits the stochastic representation  $\mathbf{X} = \sqrt{\mathbf{W}}\mathbf{A}\mathbf{Z}$  where  $\mathbf{A}$  and  $\mathbf{Z}$  are as above and  $\mathbf{W} = 1/\Gamma$  for  $\Gamma$  following a Gamma distribution with shape and rate equal to  $v/2$ . A random vector  $\mathbf{U} \sim C_{v,P}^t$  thus admits the stochastic representation  $t_v(\mathbf{X}) = t_v(\sqrt{\mathbf{W}}\mathbf{A}\mathbf{Z})$ ; as before,  $t_v$  is assumed to act on  $\sqrt{\mathbf{W}}\mathbf{A}\mathbf{Z}$  component-wise. Note that for  $t$  copulas with finite  $v$ , this sampling approach is different from the CDM.

##### Archimedean copulas

The conditional independence approach behind the Marshall–Olkin algorithm for sampling Archimedean copulas is one example for transformations  $\phi_C$  for  $k > d$ ; see Marshall and Olkin (1988). For this algorithm,  $k = d + 1$  and one uses the fact that for an Archimedean copula  $C$  with completely monotone generator  $\psi$ ,

$$\mathbf{U} = (\psi(E_1/V), \dots, \psi(E_d/V)) \sim C, \quad (16)$$

where  $V \sim F = \mathcal{L}\mathcal{S}^{-1}[\psi]$ , independent of  $E_1, \dots, E_d \sim \text{Exp}(1)$ ; here,  $F = \mathcal{L}\mathcal{S}^{-1}[\psi]$  denotes the distribution function corresponding to  $\psi$  by Bernstein’s Theorem ( $\mathcal{L}\mathcal{S}^{-1}[\cdot]$  denotes the inverse Laplace–Stieltjes transform). To give an explicit expression for the transformation  $\phi_C = \phi_C^{\text{MO}}$  in this case, we assume that  $v_1$  is used to generate  $V$  via the inversion method, and  $v_{j+1}$  is used to generate  $E_j$ , for  $j \in \{1, \dots, d\}$ . Then we have that  $\phi_C^{\text{MO}} = (\phi_{C,1}^{\text{MO}}, \dots, \phi_{C,d}^{\text{MO}})$ , where

$$\phi_{C,j}^{\text{MO}} = \phi_{C,j}^{\text{MO}}(v_1, v_{j+1}) = \psi\left(\frac{-\log v_{j+1}}{F^{-1}(v_1)}\right), \quad j \in \{1, \dots, d\}. \quad (17)$$

We can use a low-discrepancy sequence in  $k = d + 1$  dimensions to produce a sample based on this method. Having



$k = d + 1$  instead of  $k = d$  is a slight disadvantage, since it is well known that the performance of QMC methods tends to deteriorate with increasing dimensions.

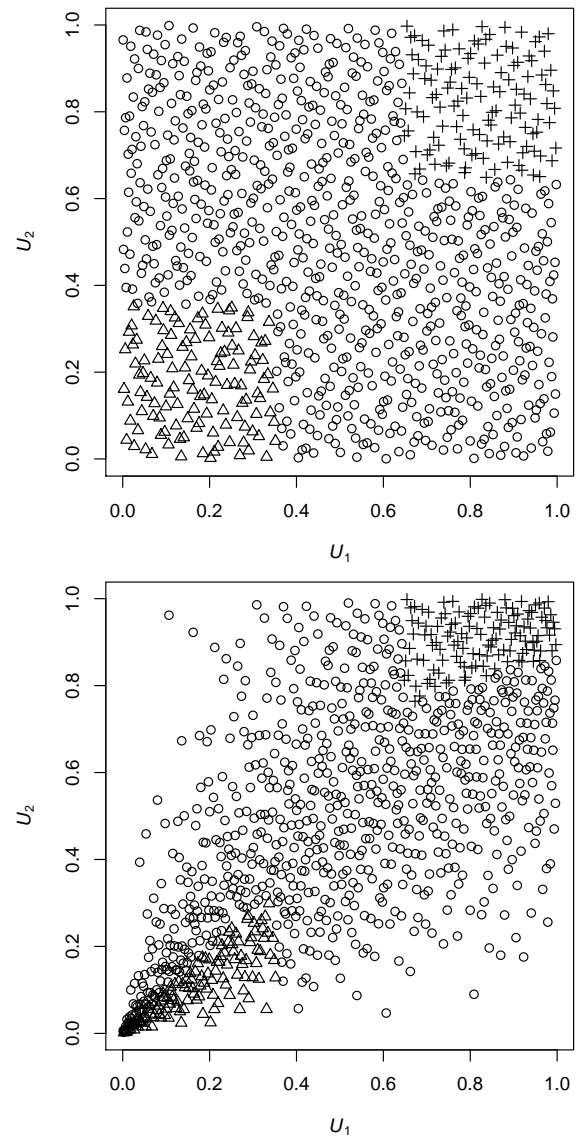
To explore the effect of the transformation  $\phi_C$  on  $P_n$ , we generated 1000 realizations of a three-dimensional Halton sequence; see the top of Figure 4 where we used different markers for points falling in two non-overlapping regions in  $[0, 1]^2$ . The first two of the three dimensions are then mapped via  $\phi_C^{\text{CDM}}$  (see the bottom of Figure 4) to a Clayton copula with parameter  $\theta = 2$  (such that Kendall's tau equals 0.5). As we can see, the non-overlapping marked regions remain non-overlapping after the one-to-one transformations have been applied. To study the effect of the Marshall-Olkin algorithm, we look at when the first dimension of the Halton sequence is mapped to a Gamma  $\Gamma(1/\theta, 1)$  distribution by inversion of  $v_1$  (the distribution of  $V$  in (16) for a Clayton copula) and the last two to unit exponential distributions (by inversion of  $1 - v_j$  for  $j = 2, 3$ , so that the obtained  $u_j$  is increasing in each of  $v_1$  and  $v_{j+1}$  for  $j = 1, 2$ ). The top of Figure 5 shows the second and third coordinates of the Halton sequence, and marks the points belonging to two different three-dimensional intervals (this is why not all two-dimensional points are marked in the two-dimensional projected regions). We see on the bottom of Figure 5 that here again, the marked regions remain non-overlapping. However, it should also be clear that two points in a given interval defined over the second and third dimension could end up in very different locations after this transform, if the corresponding first coordinates are far apart. Hence, the fact that the Marshall-Olkin transform uses  $k = d + 1$  uniforms (and thus is not one-to-one) makes it more challenging to understand its effect when used with quasi-random numbers. On the other hand, because it is designed so that the first uniform  $v_1$  is very important, it may work quite well with QMC since these methods are known to perform better when a small number of variables are important (i.e., see Dick and Pillichshammer (2010); Lemieux (2009)). This combination (QRNG with the Marshall-Olkin approach) is studied further in Section 4, with numerical results provided in Section 5.

### Marshall-Olkin copulas

Bivariate ( $d = 2$ ) Marshall-Olkin copulas  $C$  also allow for a stochastic representation in our framework  $\phi_C$  for  $k > d$ . For example, it is easy to check that for  $(U'_1, U'_2, U'_3) \sim U[0, 1]^3$ ,

$$(\max\{U_1'^{\frac{1}{1-\alpha_1}}, U_3'^{\frac{1}{\alpha_1}}\}, \max\{U_2'^{\frac{1}{1-\alpha_2}}, U_3'^{\frac{1}{\alpha_2}}\}) \sim C.$$

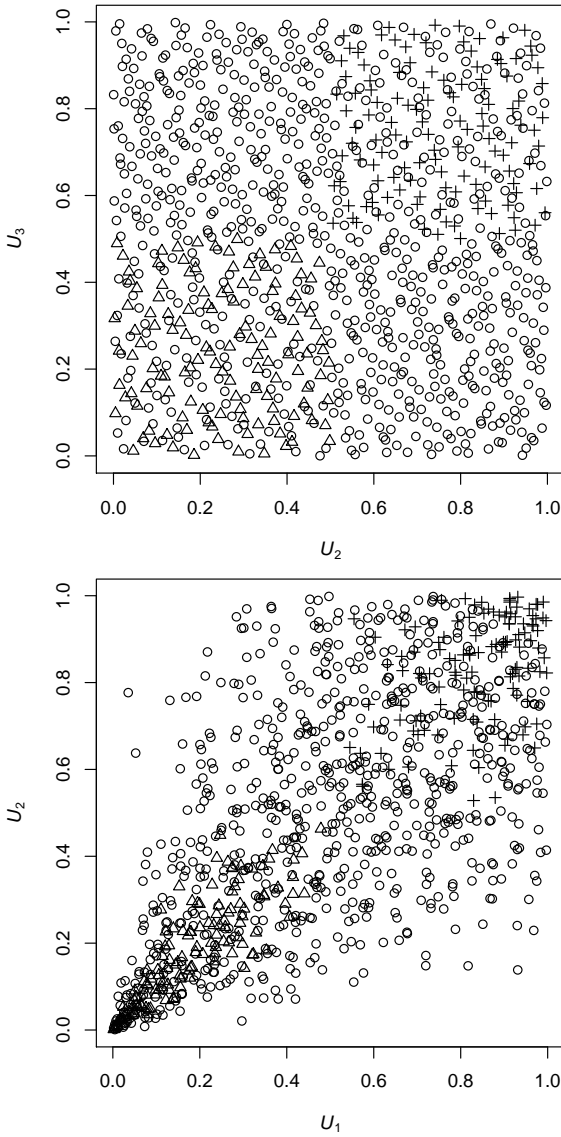
This construction can be generalized to  $d > 2$  (but we omit further details about Marshall-Olkin copulas in the remaining part of this paper).



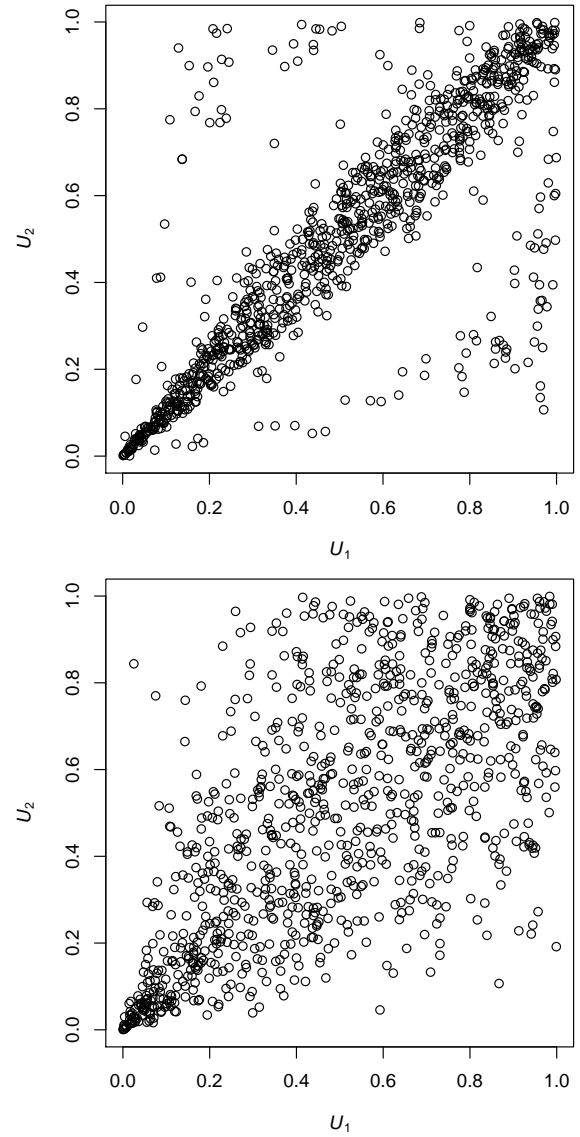
**Figure 4** 1000 realizations of the first two components of a three-dimensional Halton sequence with marked points ( $\triangle$  and  $+$ ) in the respective regions  $[0, \sqrt{1/8}]^2$  and  $[1 - \sqrt{1/8}, 1]^2$  (top): corresponding  $\phi_C^{\text{CDM}}$ -transformed points (bottom) to a Clayton copula with  $\theta = 2$  (Kendall's tau equals 0.5).

### 3.3 Words of caution

The plots showing copula samples obtained from QRNGs that we have seen so far have been promising, in that the additional uniformity (or low discrepancy) compared to pseudo-sampling was visible. Here we want to add a word of caution to the effect that it is crucial to work with high quality quasi-random numbers, as defects that exist with respect to their uniformity on the unit cube will translate into poor copula samples. Figure 6 illustrates this by showing two-dimensional copula samples obtained from quasi-random numbers of poor quality, corresponding to the projection on coordinates (20,21) of the first 1000 points of the Halton



**Figure 5** 1000 realizations of the second and third components of a three-dimensional Halton sequence with marked points ( $\Delta$  and  $+$ ) corresponding to the respective regions  $[0, 0.5]^3$  and  $[0.5, 1]^3$  (top): corresponding  $\phi_C^{\text{MO}}$ -transformed points (bottom) to a Clayton copula with  $\theta = 2$  (Kendall's tau equals 0.5).



**Figure 6** Samples obtained from Clayton copula with  $\theta = 2$  with the CDM method based on coordinates 20 and 21 of the Halton sequence (top) and the generalized Halton sequence (bottom).

#### 4 Analyzing the performance of copula sampling with quasi-random numbers

In this section, we discuss the two approaches outlined in the introduction to analyze the validity of sampling algorithms for copulas that are based on low-discrepancy sequences.

##### 4.1 Composing the sampling method with the function of interest $\Psi$

Our goal here is to assess the quality of a quasi-random sampling method for copula models by viewing the transformation  $\phi_C$  as being composed with the function  $\Psi$  of interest,

sequence (top) and a similar sample obtained from a generalized Halton sequence (bottom), which was designed to address defects of this type in the Halton sequence. More precisely, here the problem is that this particular projection is based on the twin prime numbers 71 and 73 for the base. Defects of this type are discussed further, e.g., in Morokoff and Caflisch (1994).

so that we can work in the usual Koksma–Hlawka setting based on uniform discrepancy measures.

Given that a copula transform  $\phi_C = (\phi_{C,1}, \dots, \phi_{C,d})$  is regular enough, denote its Jacobian by

$$J_{\phi_C} = \frac{\partial (\phi_{C,1}, \dots, \phi_{C,d})}{\partial (u_1, \dots, u_d)},$$

and write

$$\begin{aligned} \mathbb{E}[\Psi(\mathbf{U})] &= \int_{[0,1]^d} \Psi(\mathbf{u})c(\mathbf{u})d\mathbf{u} \\ &= \int_{[0,1]^d} \Psi(\phi_C(\mathbf{v}))c(\phi_C(\mathbf{v}))|J_{\phi_C}(\phi_C(\mathbf{v}))|d\mathbf{v}. \end{aligned}$$

In the case of  $\phi_C = \phi_C^{\text{CDM}}$ , one can easily show that  $|J_{\phi_C}(\phi_C(\mathbf{v}))| = c(\phi_C(\mathbf{v}))^{-1}$ , and thus

$$\mathbb{E}[\Psi(\mathbf{U})] = \int_{[0,1]^d} \Psi(\phi_C(\mathbf{v}))d\mathbf{v}. \quad (18)$$

While the properties of the CDM approach allow one to directly show (18) in its integral form as done above, this equality holds more generally for any transformation  $\phi_C : [0, 1]^k \rightarrow [0, 1]^d$  such that  $\phi_C(\mathbf{U}) \sim C$  whenever  $\mathbf{U} \sim U[0, 1]^k$ ; see also Cafilisch (1998); Pillards and Cools (2006).

In the case where (18) holds, one can apply the Koksma–Hlawka error bound (4) to transformed samples.

**Proposition 2 (Koksma–Hlawka bound for a change of variables)** *Let  $\mathbf{U} \sim C$ ,  $\phi_C$  such that (18) holds, and  $\mathbf{u}_i = \phi_C(\mathbf{v}_i)$  for  $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$  in  $[0, 1]^d$ . Then*

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(P_n)V(\Psi \circ \phi_C).$$

Note that  $V(\Psi) < \infty$  does not imply  $V(\Psi \circ \phi_C) < \infty$  in general. To get further insight into the conditions required to have a finite bound on the integration error, we work with a slight variation of the above bound that is given in (Niederreiter 1992, pp. 19–20) (see also (Hlawka and Mück 1972, (4')) and (Hlawka 1961, (4))), where the term  $V(\Psi \circ \phi_C)$  is replaced by an expression given in terms of the partial derivatives of  $\Psi \circ \phi_C$  assuming the latter exist and are continuous. It is given by

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(P_n)\|\Psi \circ \phi_C\|_{d,1},$$

where

$$\|\Psi \circ \phi_C\|_{d,1} = \sum_{l=1}^s \sum_{\alpha} \int_{(0,1)^l} \left| \frac{\partial^l \Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}} \right| dv_{\alpha_1} \dots dv_{\alpha_l}$$

and the second sum is taken over all nonempty subsets  $\alpha = \{\alpha_1, \dots, \alpha_l\} \subseteq \{1, \dots, d\}$ . Furthermore, the notation  $\mathbf{1}$  in  $\Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})$  means that each variable  $v_j$  with

$j \notin \{\alpha_1, \dots, \alpha_l\}$  is set to 1.

The following proposition provides sufficient conditions on the functional  $\Psi$  and on the copula  $C$  to ensure that  $\|\Psi \circ \phi_C\|_{d,1} < \infty$  when  $\phi_C = \phi_C^{\text{CDM}}$ .

**Proposition 3 (Conditions to have bounded variation with variable change in the CDM)** *Assume that  $\Psi$  has continuous mixed partial derivatives up to total order  $d$  and there exist  $m, M, K > 0$  such that for all  $\mathbf{u} \in (0, 1)^d$ ,  $c(\mathbf{u}) \geq m > 0$ ,  $C_{i|1\dots i-1} = C(u_i | u_1, \dots, u_{i-1})$  and*

$$\left| \frac{\partial^k C_{i|1\dots i-1}}{\partial u_{\alpha_1} \dots \partial u_{\alpha_k}} \right| \leq M, \quad \alpha_1, \dots, \alpha_k \in \{1, \dots, i\}, \quad (19)$$

for each  $1 \leq k \leq i \leq d$ . Furthermore, assume that for all  $1 \leq k \leq l \leq d$  and  $\{\alpha_1, \dots, \alpha_l\} \subseteq \{1, \dots, d\}$ , we have

$$\left| \frac{\partial^k \Psi(u_1, \dots, u_d)}{\partial u_{\beta_1} \dots \partial u_{\beta_k}} \right| \leq K, \quad \beta_j \in \{\alpha_1, \dots, \alpha_l\}, \quad 1 \leq j \leq k. \quad (20)$$

Then there exists a constant  $C^{(d)}$  (independent of  $n$  but dependent on  $\Psi$ ) such that for the choice  $\phi_C = \phi_C^{\text{CDM}}$ , we have

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(\mathbf{v}_1, \dots, \mathbf{v}_n)KC^{(d)}(M^d/m)^{2d-1},$$

where  $\mathbf{u}_i = \phi_C^{\text{CDM}}(\mathbf{v}_i)$ ,  $i = 1, \dots, n$ .

*Proof* See (Hlawka and Mück 1972, (11) and the remark thereafter).

*Remark 1* 1) As we will see in the discussion preceding the next proposition, in general, to ensure that  $\|\Psi \circ \phi_C\|_{d,1} < \infty$  holds, a possible approach is to bound the mixed partial derivatives involving  $\Psi$  and then to verify that the mixed partial derivatives involving  $\phi_C$  are integrable. As explained in Hlawka and Mück (1972), Condition (19) ensures that the latter condition is verified in the case of the CDM (or Rosenblatt) transform, and avoids having to deal with the function  $\phi_C$  and its partial derivatives. Unfortunately (and while it may seem easier to work with the conditional copulas  $C_{j|1\dots j-1}$  than with  $\phi_C$ ), in many cases the copulas involved do not have bounded mixed partial derivatives everywhere, with singularities appearing near the boundaries when one or more arguments are 0 or 1. A non-trivial case where we were able to verify (19) is for the Eyrard–Farlie–Gumbel–Morgenstern copula (see Jaworski et al (2010)), assuming the parameters are chosen so that the density  $c(\mathbf{u})$  and thus the denominator of  $C_{j|1\dots j-1}$  is bounded away from 0 for all  $\mathbf{u} \in [0, 1]^d$ .

- 2) We note that the conditions given in (20) are not the same as those required to prove that

$$\|\Psi\|_{d,1} = \sum_{l=1}^d \sum_{\alpha} \int_{[0,1]^l} \left| \frac{\partial^l \Psi(u_{\alpha_1}, \dots, u_{\alpha_l}, \mathbf{1})}{\partial u_{\alpha_1} \dots \partial u_{\alpha_l}} \right| du_{\alpha_1} \dots du_{\alpha_l}$$

is bounded; in the latter case, we only need to consider mixed partial derivatives of order at most one in each variable (since the  $\alpha_j$ 's are distinct). However, in (20), the  $\beta_j$ 's are not necessarily distinct. In particular, this means that we need to consider the partial derivative of  $\Psi$  of order  $d$  with respect to each variable and make sure it is bounded.

Let us now move away from the CDM method and consider a general transformation  $\phi_C$ . In order to study  $\|\Psi \circ \phi_C\|_{d,1}$ , we first need to decompose mixed partial derivatives of the form

$$\frac{\partial^l (\Psi \circ \phi_C)(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}}$$

in terms of  $\Psi$  and  $\phi_C$  separately. To do so, we follow Hlawka and Mück (1972), as well as (Constantine and Savits 1996, Theorem 2.1), and obtain an expression for the mixed partial derivative of a composition of functions via the representation

$$\begin{aligned} & \frac{\partial^l \Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}} \\ &= \sum_{1 \leq |\beta| \leq l} \frac{\partial^{|\beta|} \Psi}{\partial \beta_1 u_1 \dots \partial \beta_d u_d} \sum_{s=1}^l \sum_{\gamma, k} c_{\gamma} \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C,k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1}^{\gamma_{j,1}} \dots \partial v_{\alpha_l}^{\gamma_{j,l}}} \end{aligned} \quad (21)$$

where  $\beta \in \mathbb{N}_0^d$  and  $|\beta| = \sum_{j=1}^d \beta_j$ . Here we do not specify over which values of  $\gamma_j$  and  $k_j$  the inner sum in the above expression is taken: details can be found in the proof of Proposition 4. But let us point out that in the product over  $j$ , each index  $\alpha_1, \dots, \alpha_l$  appears exactly once. On the other hand – and as noted in item 2) of Remark 1 above – in the mixed partial derivative of  $\Psi$ , a given variable can appear with order larger than 1.

From (21), we see that a sufficient condition to show that  $\|\Psi \circ \phi_C\|_{d,1} < \infty$  is to establish that all products of the form

$$\frac{\partial^{|\beta|} \Psi}{\partial \beta_1 u_1 \dots \partial \beta_d u_d} \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C,k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1}^{\gamma_{j,1}} \dots \partial v_{\alpha_l}^{\gamma_{j,l}}}, \quad s \in \{1, \dots, l\}, \quad (22)$$

are in  $L_1$ .

We note that for the MO algorithm (assuming as we did in (17) that  $v_1$  is used to generate  $V$  and  $v_{j+1}$  is used to generate  $E_j$ ),  $\phi_{C,j}$  is a function of  $v_1$  and  $v_{j+1}$  only, for  $j \in \{1, \dots, d\}$ . Hence the only partial derivatives of  $\phi_{C,j}$  that are nonzero are those with respect to variables in  $\{v_1, v_{j+1}\}$ . This observation

is helpful to prove the following result, which shows that the error bound obtained when using the MO algorithm has the desired behavior induced by the low-discrepancy point set used to generate the copula samples; note that it does not show that  $\Psi \circ \phi_C$  has bounded variation. Its proof can be found in the appendix.

**Proposition 4 (Error behaviour for MO for continuous  $V$ )** Let  $\phi_C^{MO}$  be the transformation associated with the Marshall–Olkin algorithm, as given in (17), and that  $V \sim F$  is continuously distributed. Let  $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$  be the point set in  $[0, 1]^{d+1}$  used to produce copula samples via the transformation  $\phi_C^{MO}$  and let  $\mathbf{u}_i = \phi_C^{MO}(\mathbf{v}_i)$ . If

- 1) the point set  $P_n$  excludes the origin and there exists some  $p \geq 1$  such that  $\min_{1 \leq i \leq n} v_{i,1} \geq 1/pn$ ;
- 2) the function  $\Psi$  satisfies  $|\Psi(\mathbf{u})| < \infty$  for all  $\mathbf{u} \in [0, 1]^{d+1}$  and

$$\frac{\partial^{|\beta|} \Psi}{\partial \beta_1 u_1 \dots \partial \beta_d u_d} < \infty \text{ for all } \beta = (\beta_1, \dots, \beta_d), \quad (23)$$

with  $\beta_l \in \{0, \dots, d\}$  and  $|\beta| \leq d$ ;

- 3) and the generator  $\psi(\cdot)$  of the Archimedean copula  $C$  is such that

- a)  $\psi'(t) + t\psi''(t)$  has at most one zero  $t^*$  in  $(0, \infty)$  and it satisfies  $-t^* \psi'(t^*) < \infty$ ; and
- b)  $F^{-1}(1 - 1/pn)$  is in  $O(n^a)$  for some constant  $a > 0$ ;

then there exists a constant  $C^{(d)}$  (independent of  $n$  but dependent on  $\Psi$  and  $\phi_C^{MO}$ ) such that

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq C^{(d)} (\log n) D^*(P_n).$$

*Remark 2* We note that if  $\mathbb{E}[V] < \infty$ , as is the case for Clayton's copula family, Condition 3) b) can be easily checked via Markov's inequality. In the case of the Gumbel copula,  $V$  has an  $\alpha$ -stable distribution and it can be shown that  $P(V > x) \leq cx^{-\alpha}$  for  $x \geq x_0$  and for some constant  $c$ , where  $c$  and  $x_0$  depend both on the parameters of the distribution; see (Nolan 2014, Theorem 1.12). Therefore  $F^{-1}(1 - 1/pn)$  can be bounded by a constant time  $n^a$  in this case (namely by  $\max\{x_0, (c/pn)^{1/\alpha}\}$ ). As for Condition 3) a), one can show that  $t^* = \theta$  and  $t^* = 1$  are the only zeros for the Clayton and Gumbel copulas, respectively.

When  $F$  is discrete, we can split the problem into sub-problems based on the value taken by  $V$ . Then, in each case, the bounded variation condition is much easier to verify, because the transformation  $\phi_C$  given  $V$  is essentially one-dimensional as it is mapping each  $v_j$  to an exponential  $E_{j-1}$  for  $j \in \{2, \dots, d+1\}$ . Its proof can be found in the appendix.

**Proposition 5 (Error behaviour for the Marshall–Olkin algorithm for discrete  $V$ )** Let  $\phi_C^{MO}$  be the transformation associated with the Marshall–Olkin algorithm, as given in (17)

and assume  $C$  is an Archimedean copula whose distribution function  $F$  of  $V$  is discrete. Let  $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$  be the point set in  $[0, 1]^{d+1}$  used to produce copula samples via the transformation  $\phi_C^{MO}$  and let  $\mathbf{u}_i = \phi_C^{MO}(\mathbf{v}_i)$ . If (23) holds and

- 1) there exists some  $p \geq 1$  such that the point set  $P_n$  satisfies  $\max_{1 \leq i \leq n} v_{i,1} \leq 1 - 1/pn$ ;
- 2) there exist constants  $c > 0$  and  $q \in (0, 1)$  such that  $1 - F(l) \leq cq^l$  for  $l \geq 1$ ;

then there exists a constant  $C^{(d)}$  (independent of  $n$  but dependent on  $\Psi$  and  $\phi_C^{MO}$ ) such that

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq C^{(d)} (\log n) D^*(P_n).$$

*Remark 3* We note that the Frank, Joe and Ali-Mikhail-Haq copulas are such that  $F$  is discrete. The condition on the tail of  $F$  stated in the proposition can be shown to hold for the Frank and Ali-Mikhail-Haq copulas, but not for the Joe copula (the distribution of  $V$  in this case has a Sibuya distribution, for which no moments exists, i.e., it has a very fat tail).

Let us now move on to RQMC methods. We already mentioned that an advantage they have over their deterministic counterparts is that expressions for their variance can be obtained under weaker conditions than what is required to get error expressions and bounds for deterministic sampling schemes. The following result shows that similar variance expressions also hold after composing  $\Psi$  with  $\phi_C$ .

**Proposition 6 (Variance expression with a change of variables)** *If  $\tilde{P}_n = \{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n\}$  is a randomly digitally shifted net with corresponding RQMC estimator  $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \Psi(\phi_C(\tilde{\mathbf{v}}_i))$  and if  $\text{Var}(\Psi(\mathbf{U})) < \infty$  with  $\mathbf{U} \sim C$ , then we have that*

$$\text{Var}(\hat{\mu}_n) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_d^*} |\widehat{\Psi \circ \phi_C}(\mathbf{h})|^2, \quad (24)$$

where  $\mathcal{L}_d^*$  is the dual net of the deterministic net that has been shifted to get  $\tilde{P}_n$ , and  $\hat{f}(\mathbf{h})$  is the Walsh coefficient of  $f$  at  $\mathbf{h}$ , while

$$\begin{aligned} (\widehat{\Psi \circ \phi_C})(\mathbf{h}) &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) P(\mathbf{h}, \mathbf{k}), \\ P(\mathbf{h}, \mathbf{k}) &= \int_{[0,1]^d} e^{2\pi i(\mathbf{k} \cdot \phi_C(\mathbf{w}) - \mathbf{h} \cdot \mathbf{w})} d\mathbf{w} \\ &= \mathbb{E} \left[ e^{2\pi i(\mathbf{k} \cdot \phi_C(\mathbf{W}) - \mathbf{h} \cdot \mathbf{W})} \right], \quad \mathbf{W} \sim \mathbf{U}[0, 1]^d. \end{aligned}$$

*Proof* It is clear from Theorem 3 in the appendix and using Representation (18) that (24) holds and the condition  $\text{Var}(\Psi(\mathbf{U})) < \infty$  with  $\mathbf{U} \sim C$  ensures it is finite. So what remains to be shown is the expression for the Walsh coefficient

of the composed function  $\Psi \circ \phi_C$ . It is obtained as follows:

$$\begin{aligned} (\widehat{\Psi \circ \phi_C})(\mathbf{h}) &= \int_{[0,1]^d} \Psi(\phi_C(\mathbf{w})) e^{-2\pi i \langle \mathbf{h}, \mathbf{w} \rangle_b} d\mathbf{w} \\ &= \int_{[0,1]^d} \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) e^{2\pi i \langle \mathbf{k}, \phi_C(\mathbf{w}) \rangle_b} e^{-2\pi i \langle \mathbf{h}, \mathbf{w} \rangle_b} d\mathbf{w} \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) \int_{[0,1]^d} e^{2\pi i (\langle \mathbf{k}, \phi_C(\mathbf{w}) \rangle_b - \langle \mathbf{h}, \mathbf{w} \rangle_b)} d\mathbf{w} \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) P(\mathbf{h}, \mathbf{k}), \end{aligned}$$

where the third equality holds thanks to Fubini's theorem.

By adding assumptions on the smoothness of  $\Psi$  and thus on the behavior of its Walsh coefficients, one could obtain improved convergence rates for the variance given in (24) compared to the  $O(1/n)$  we get with MC, something we plan to study in future work.

## 4.2 Transforming the low-discrepancy samples

As mentioned in the introduction, we can think of  $\phi_C$  as transforming the point set  $P_n$  instead of being composed with  $\Psi$ . The integration error can then be analyzed via a generalized version of the Koksma–Hlawka inequality such as the one studied in Aistleitner and Dick (2015), which we now explain.

Similarly to the Lebesgue case we define the *copula-discrepancy function* with respect to a copula-induced measure  $P_C$  on an interval  $B$  (i.e.,  $P_C(B) = \mathbb{P}(\mathbf{U} \in B)$  for  $\mathbf{U} \sim C$ ) as

$$E_C(B; P_n) = \frac{A(B; P_n)}{n} - P_C(B).$$

Let  $\mathcal{J}$  be the set of intervals of  $[0, 1]^d$  of the form  $[a, \mathbf{b}) = \prod_{j=1}^d [a_j, b_j)$ , where  $0 \leq a_j \leq b_j \leq 1$ . The *copula-discrepancy*  $D_C$  of  $P_n$  is then defined as

$$D_C(P_n) = \sup_{B \in \mathcal{J}} |E_C(B; P_n)|, \quad (25)$$

and similarly for  $D_C^*(P_n)$ , the *star-copula-discrepancy function* when the sup in (25) is taken over  $\mathcal{J}^*$  instead.

The generalization of the Koksma–Hlawka inequality studied in (Aistleitner and Dick 2015, Theorem 1) then provides

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq V(\Psi) D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n),$$

where we assume  $\mathbf{u}_i = \phi_C(\mathbf{v}_i)$ ,  $i \in \{1, \dots, n\}$ . To get some insight on this upper bound, we need to know how  $D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n)$  behaves as a function of  $n$ . Unfortunately, in general we cannot prove that  $D^*(\mathbf{v}_1, \dots, \mathbf{v}_n) \in O(n^{-1} \log^d n)$  implies that

$D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) \in O(n^{-1} \log^d n)$ . Here are a few things we can say, though.

First, an obvious case for which discrepancy is preserved is when  $\phi_C$  maps rectangles to rectangles, because then  $\phi_C(B) \in \mathcal{J}$  for all  $B \in \mathcal{J}$ , and thus

$$D_C(\tilde{P}_n) \leq D(P_n),$$

$$D_C^*(\tilde{P}_n) \leq D^*(P_n),$$

where  $\tilde{P}_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ . However, this only happens when  $C$  is the independence copula, and in this case the equality holds. This is not a very interesting case since our focus here is on dependence modelling.

For the more realistic setting where  $\phi_C$  does not map rectangles to rectangles, the following result from Hlawka and Mück (1972) holds and gives a much slower convergence rate for  $D_C^*(\tilde{P}_n)$ .

**Proposition 7** *Let  $C$  be such that the Rosenblatt transform  $\phi_C^{-1}$  is Lipschitz continuous on  $[0, 1]^d$  w.r.t. the sup-norm  $\|\cdot\|_\infty$ , and  $\{\mathbf{u}_i = \phi_C(\mathbf{v}_i)\}$  for some sequence of points  $\{\mathbf{v}_i\}$  in  $[0, 1]^d$ . Then*

$$D_C(\{\mathbf{u}_1, \dots, \mathbf{u}_n\}) \leq c(d)D(\{\mathbf{v}_1, \dots, \mathbf{v}_n\})^{1/d},$$

for some function  $c(d)$ , constant in  $n$ .

Note that the above results fully depend on the properties of  $\phi_C$ . The aim would then be to choose  $\phi_C$  such that a low-discrepancy sequence  $\{\phi_C(\mathbf{v}_i)\}$  w.r.t. the copula measure  $P_C$  results whenever applied to a low-(Lebesgue-)discrepancy sequence  $\{\mathbf{v}_i\}$ . A more fundamental approach would be to directly produce a low-discrepancy sequence  $\{\mathbf{u}_i\}$  where the discrepancy is measured w.r.t. the copula measure  $C$ . This is something we intend to study in future work.

Now, computing  $D_C$  or  $D_C^*$  is usually not feasible in practice. If we replace the sup-norm by the  $L_2$ -norm, we obtain  $L_2$ -discrepancies which are usually more practical to compute. Let  $L_2$ -discrepancies  $T_C(\mathbf{u}_1, \dots, \mathbf{u}_n)$  and  $T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n)$  be defined by

$$T_C(\mathbf{u}_1, \dots, \mathbf{u}_n) = \left( \int_{\{(\mathbf{y}, \mathbf{z}) \in [0, 1]^{2d}; y_i < z_i\}} \left( \frac{A([\mathbf{y}, \mathbf{z}]; P_n)}{n} - P_C([\mathbf{y}, \mathbf{z}]) \right)^2 d\mathbf{y}d\mathbf{z} \right)^{1/2},$$

and

$$T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) = \left( \int_{[0, 1]^d} \left( \frac{A([\mathbf{0}, \mathbf{z}]; P_n)}{n} - C(\mathbf{z}) \right)^2 d\mathbf{z} \right)^{1/2},$$

respectively. Proceeding similarly to Morokoff and Caflisch (1994),  $T_C^*$  can be computed as

$$\begin{aligned} T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{i=1}^d (1 - \max(u_{k,i}, u_{l,i})) + \int_{[0, 1]^d} C(\mathbf{z})^2 d\mathbf{z} \\ &\quad - \frac{2}{n} \sum_{k=1}^n \int_{u_{k,1}}^1 \dots \int_{u_{k,d}}^1 C(\mathbf{z}) d\mathbf{z}. \end{aligned}$$

If we consider a convex combination  $C(u_1, \dots, u_d) = \lambda \prod_{i=1}^d u_i + (1 - \lambda) \min(u_1, \dots, u_d)$ ,  $\lambda \in (0, 1)$ , of the independence copula and the upper Fréchet–Hoeffding bound, then one can compute  $T_C^*$  explicitly via

$$\begin{aligned} T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{i=1}^d (1 - \max(u_{k,i}, u_{l,i})) + \frac{\lambda^2}{3^d} \\ &\quad + \frac{2(1 - \lambda)^2}{(d+1)(d+2)} + \frac{2\lambda(1 - \lambda)d!}{\prod_{i=1}^d (2i+1)} - \frac{\lambda}{n2^{d-1}} \sum_{k=1}^n \prod_{i=1}^d (1 - u_{k,i}^2) \\ &\quad - \frac{2(1 - \lambda)}{n} \sum_{k=1}^n \left( \sum_{i_1=1}^d \sum_{i_2 \neq i_1}^d \dots \sum_{i_d \neq i_1, \dots, i_{d-1}}^d \frac{1 - u_{k,i_d}^{d+1}}{(d+1)!} \right. \\ &\quad \left. - \sum_{l=1}^{d-1} \sum_{i_1=1}^d \dots \sum_{i_l \neq i_1, \dots, i_{l-1}}^d \frac{u_{k,i_l}^{l+1} (1 - u_{k,i_{l+1}})}{(l+1)!} \right). \end{aligned}$$

## 5 Numerical results

Through typical examples from the realm of finance and insurance and a few test functions, we now illustrate in this section the efficiency of QRNG in comparison to standard (P)RNG for copula sampling. More precisely, we compare Monte Carlo sampling approaches with two types of QRNGs based on randomized low-discrepancy sequences: The Sobol' sequence and the generalized Halton sequence, both randomized with a digital shift. Variance/error estimates are obtained by using  $B = 25$  i.i.d. copies of the randomized sequence and comparisons are made with MC sampling based on the same total number of replications. Each plot includes lines showing  $n^{-0.5}$ ,  $n^{-1}$  and/or  $n^{-1.5}$  convergence rates. In addition, on top of each plot and for each QRNG method, we provide the regression estimate of  $\alpha$  such that the variance/error is in  $O(n^{-\alpha})$ . For PRNG, we only show the results with the CDM sampling algorithm, since the choice of method does not affect the error or variance very much. On the other hand, for QRNG we show the results both with CDM and MO (when applicable), as this seems to sometimes make a difference. Understanding better why it is so and under what circumstances a sampling algorithm perform better when used in conjunction with QRNG will be a subject of further research.

While the examples given in the next section illustrates the use of our proposed method in typical contexts where they might be used, the test functions results in the section that follows are meant to focus on assessing the performance of QRNG compared to PRNG on the sole basis of generating copula samples  $\mathbf{U}$  – without including the effect of the marginal distributions – and also to see if the sampling algorithm (CDM or MO) has an effect on the performance of QRNG.

Finally, we note that QRNG based on Sobol' point sets is typically slightly faster than PRNG, while the generalized Halton sequence runs slower than PRNG.

## 5.1 Examples from the realm of finance and insurance

Consider a random vector  $\mathbf{X} = (X_1, \dots, X_d)$  modeling  $d$  risks in a portfolio of stocks or insurance losses. We assume that the  $j$ th marginal distribution is either log-normal with  $X_j \sim \text{LN}(\log(100) + \mu - \sigma^2/2, \sigma^2)$ ,  $j \in \{1, \dots, d\}$ , where  $\mu = 0.0001$  and  $\sigma = 0.2$ , or Pareto distributed with the same mean and variance as in the log-normal case. The copula  $C$  of  $\mathbf{X}$  throughout this numerical study is either a Clayton or an exchangeable  $t$  copula with three degrees of freedom. To allow a comparable degree of dependence, we will use the same Kendall's tau for both models. This easily translates to the parameter  $\theta$  of a Clayton copula via the relationship  $\theta = 2\tau(1 - \tau)^{-1}$  and to the correlation parameter  $\rho$  of an exchangeable  $t$  copula via  $\rho = \sin(\pi\tau/2)$ . We denote  $S = \sum_{j=1}^d X_j$  and consider the estimation of the following functionals  $\Psi(\mathbf{X})$ :

- the Best-Of Call option payoff  $(\max X_i - K)^+$ ;
- the Basket Call option payoff  $(d^{-1}S - K)^+$ ;
- the Value-at-Risk at level 0.99 on the aggregated risks

$$\text{VaR}_{0.99}(S) = F_S^{-1}(0.99) = \inf \{x \in \mathbb{R} : F_S(x) \geq 0.99\},$$

- the expected shortfall at level 0.99 on the aggregated risks

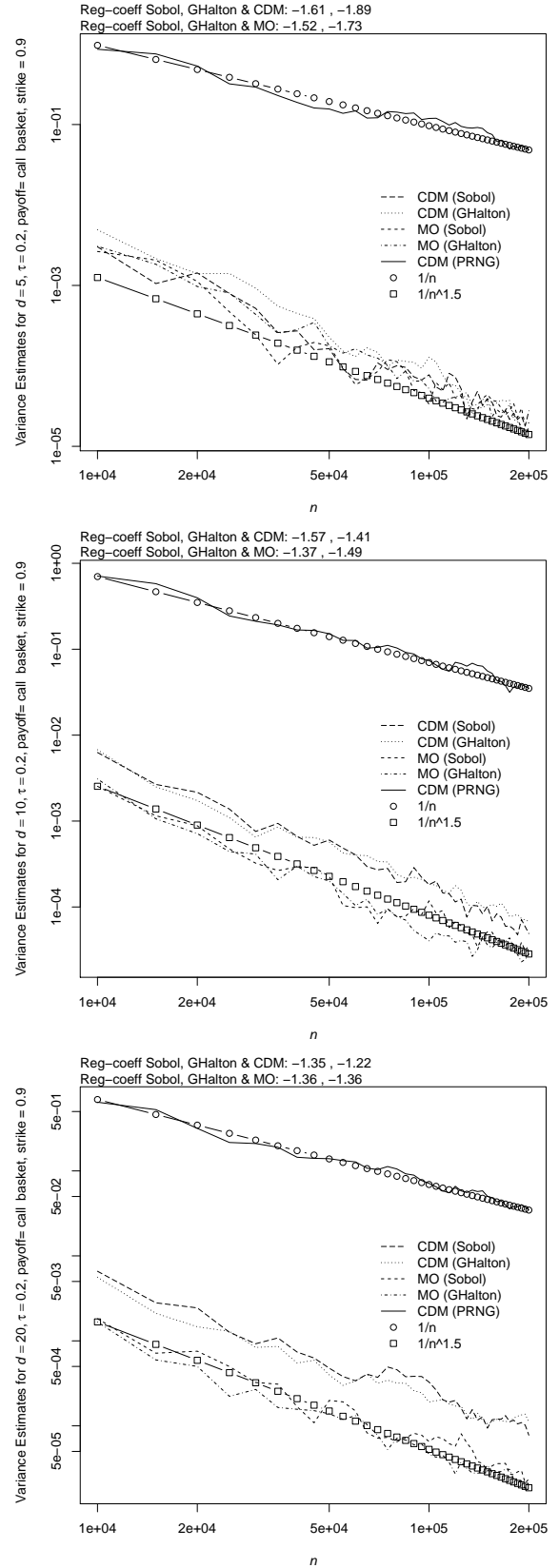
$$\text{ES}_{0.99}(S) = \frac{1}{1 - 0.99} \int_{0.99}^1 F_S^{-1}(u) du;$$

- the contribution of a given margin (without loss of generality, the first margin was chosen) to  $\text{ES}_{0.99}$  of the sum under the Euler principle, see Tasche (2008)

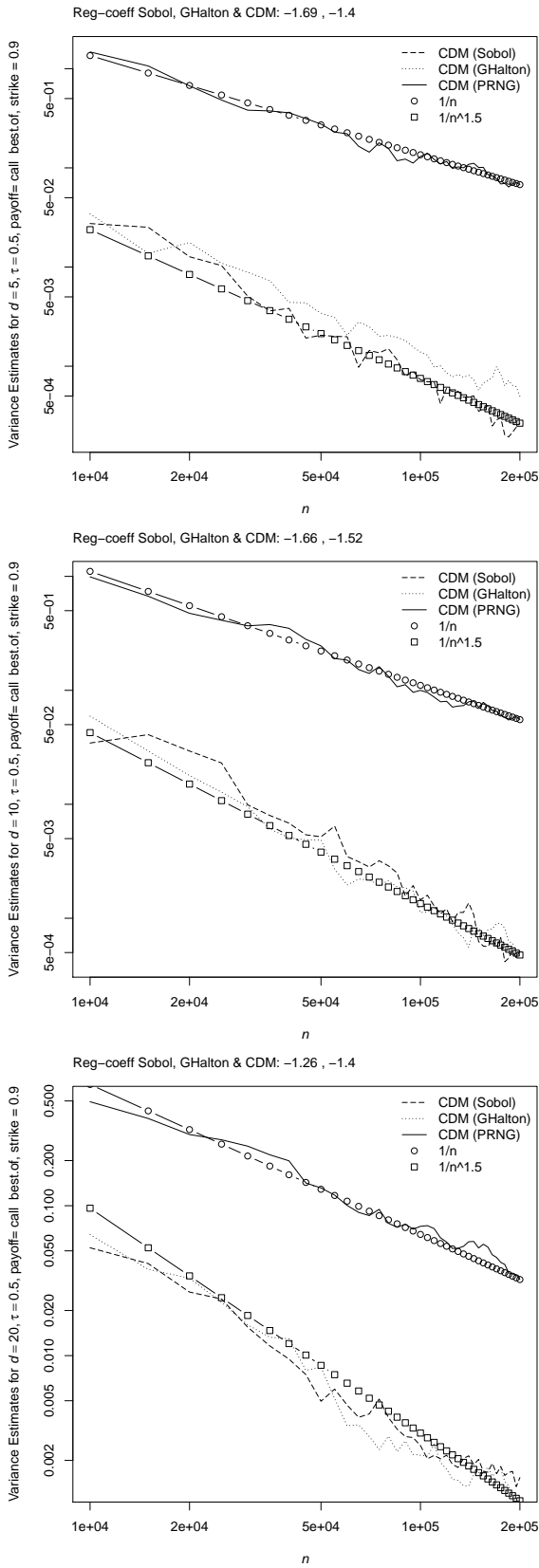
$$\mathbb{E}[X_1 | S > F_S^{-1}(\alpha)];$$

we refer to this functional as Allocation First.

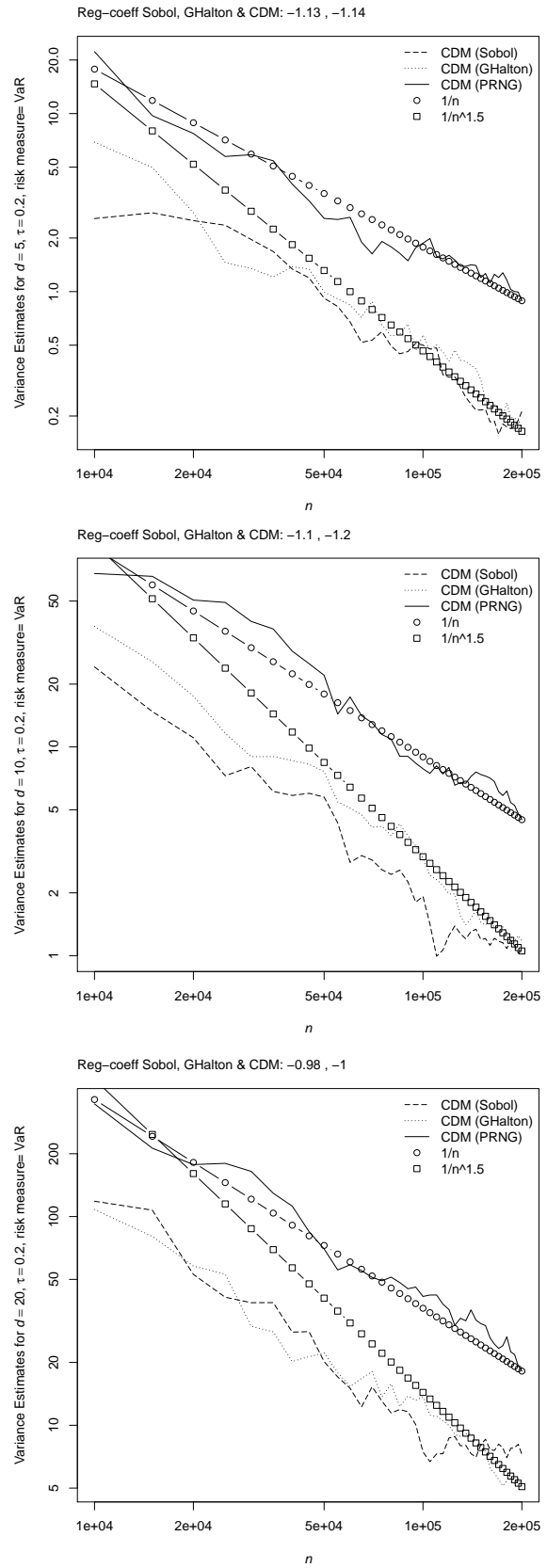
Figures 7, 8, 9, and 10 (as well as Figures 13 and 14 in the online supplement) display selected variance estimates for Clayton and  $t$  copulas with Kendall's tau parameter equal to 0.2 and 0.5, using either lognormal or Pareto margins, in dimensions  $d = 5, 10, 20$  (displayed in different rows) and sample sizes  $n \in \{10\,000, 15\,000, \dots, 200\,000\}$ . In the Clayton case, the experiment uses both the MO and CDM sampling methods. For the  $t$  copulas, while there is a sampling approach based on a stochastic representation (as seen in Section 3.2.2), there is no version of the MO algorithm available, so we only use the CDM method. In addition, both the Sobol' and generalized Halton QRNGs are used. In all cases, we see that the variances associated with the Sobol' and generalized Halton quasi-random sequences are smaller and converge faster than the Monte Carlo variance. It is not clearly determined whether one sampling method is performing considerably better than the other. But we note that in some cases, such as the estimate of the Basket Call with  $\tau = 0.2$  in  $d = 20$  dimensions (Figure 7, bottom) the MO sampling seems to perform better than CDM.



**Figure 7** Variance estimates for the functional Basket Call with log-normal margins based on  $B = 25$  repetitions for a Clayton copula with parameter such that Kendall's tau equals 0.2 for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).

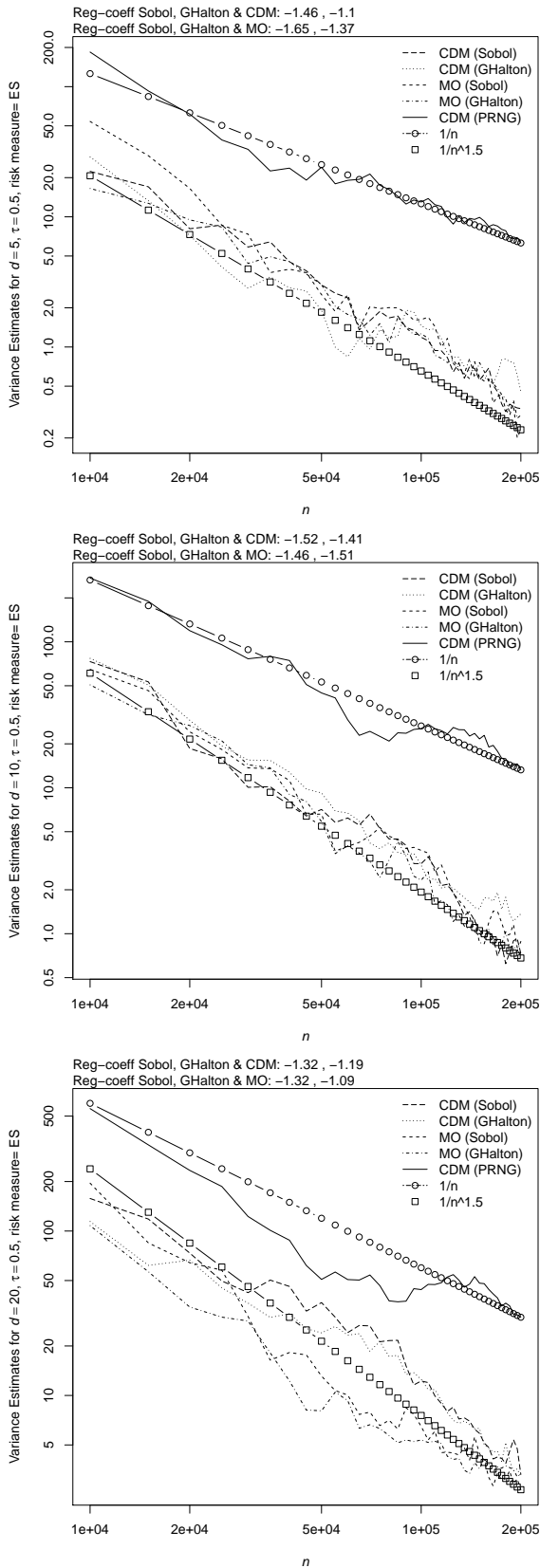


**Figure 8** Variance estimates for the functional Best-Of Call with Pareto margins based on  $B = 25$  repetitions for an exchangeable  $t$  copula with three degrees of freedom such that Kendall's tau equals 0.5 for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).



**Figure 9** Variance estimates for the functional  $VaR_{0.99}$  with lognormal margins for an exchangeable  $t$  copula with three degrees of freedom such that Kendall's tau equals 0.2 based on  $B = 25$  repetitions for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).





**Figure 10** Variance estimates for the functional  $ES_{0.99}$  with Pareto margins for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on  $B = 25$  repetitions for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).

## 5.2 Test functions

We now consider integration results on two different test functions. The results are shown in Figures 11, 12, 15, and 16 (the latter is in the online supplement), which are based on a Clayton (or  $t$ ) copula with  $\tau = 0.2$  and  $\tau = 0.5$ , respectively. The first test function is given by

$$\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d,$$

where the vector  $\mathbf{u}$  is obtained after transforming the uniform points  $\mathbf{v}$  using either the CDM transform or the MO algorithm. Recall that the former requires  $d$ -dimensional points (using either a PRNG or a QRNG), whereas the latter requires  $(d + 1)$ -dimensional points. Note that  $\Psi_1$  integrates exactly to 1 with respect to the copula-induced measure, since  $U_j \sim U[0, 1]$ ,  $j \in \{1, \dots, d\}$ . While we know the exact value of the integral in this case, we still compare estimators based on  $B$  i.i.d. copies of either MC or RQMC, but we plot the average absolute error rather than the estimated variance.

The second test function is given by

$$\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u})),$$

where

$$g_1(\mathbf{v}) = \prod_{j=1}^d \frac{|4v_j - 1| + \alpha_j}{1 + \alpha_j}, \quad \alpha_j = j,$$

which is often used as a test function in the QMC literature; see, e.g., Faure and Lemieux (2009) and the references therein. So here we first apply the inverse of the CDM transform to the copula-transformed points obtained either using the CDM approach or MO, and then apply the  $d$ -dimensional function  $g_1$ . While this has the effect of simply applying the standard test function  $g_1$  to the original sample points  $\mathbf{v}_i$  in the case of the CDM, in the case of the MO algorithm, we are not falling back on the original points  $\mathbf{v}_i$ . The hope is that if MO does not preserve so well the low discrepancy of the original points, this function would be able to detect this problem.

While the second test function is mostly interesting for Archimedean copulas, the first one can be used more generally. This is why in the results reported in Figures 11 and 12, we also consider an exchangeable  $t$  copula with three degrees of freedom and Kendall's  $\tau$  equal to 0.2.

For both test functions, we see that the Sobol' and generalized Halton sequences always clearly outperform Monte Carlo, with an error often converging in  $O(n^{-1})$  rather than the  $O(n^{-0.5})$  corresponding to Monte Carlo. For the first function  $\Psi_1$ , both the CDM and MO methods perform about the same. We believe this might be due to the simplicity of  $\Psi_1$ —a sum of univariate powers of the  $u_j$ 's—and the fact that both methods perform equally well in the univariate case when combined with RQMC. Looking at the results for  $\Psi_2$ ,

we see that with RQMC the CDM method performs better than MO, as there is no copula transformation performed in the case of CDM. However, RQMC with MO is still better than Monte Carlo, which suggests that the MO algorithm manages to preserve the low discrepancy of the original point set.

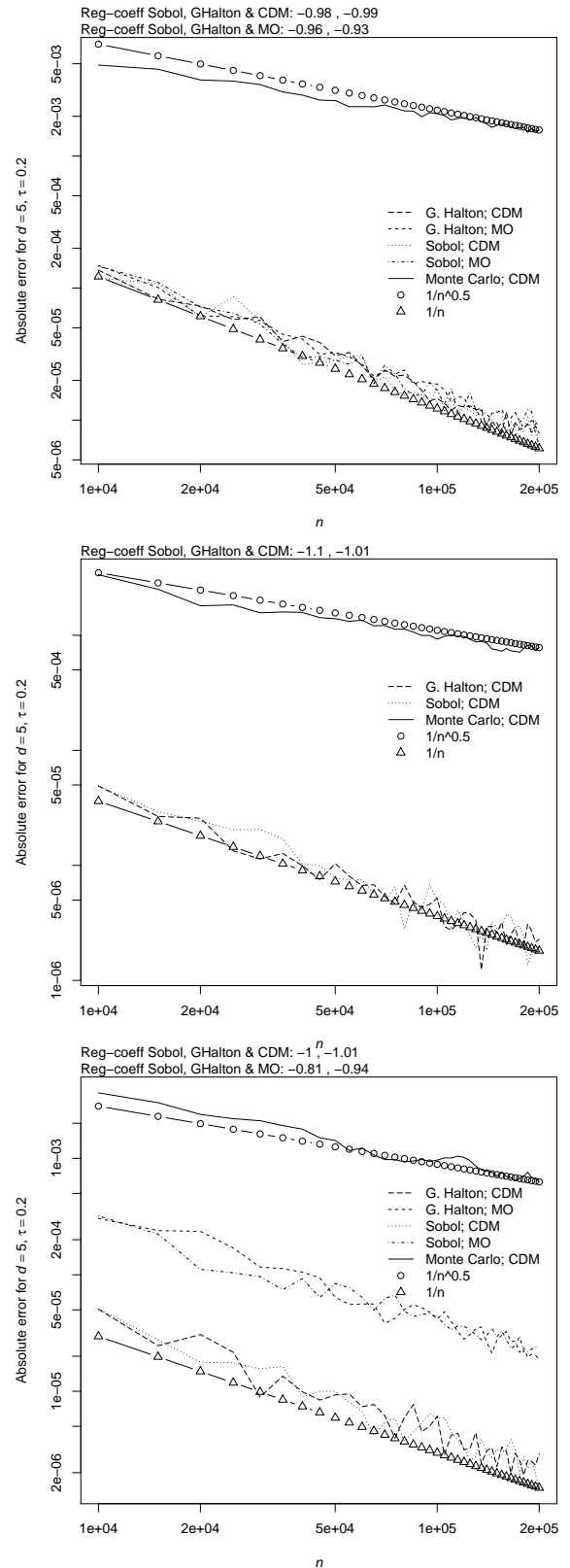
Overall, the numerical results presented in this section suggest that replacing PRNG by QRNG within algorithms that sample from copulas reduces the variance of the obtained estimators and typically also improve the convergence rate of the variance. The improvement was seen across the range of dimensions from 5 to 20 that were tested, and the different functions considered. In small dimensions, different copula sampling algorithms (CDM and MO) tend to perform similarly, but when the dimension increases we sometimes see differences in their performance, with so far neither of the two algorithms clearly dominating the other.

## 6 Conclusion and discussion

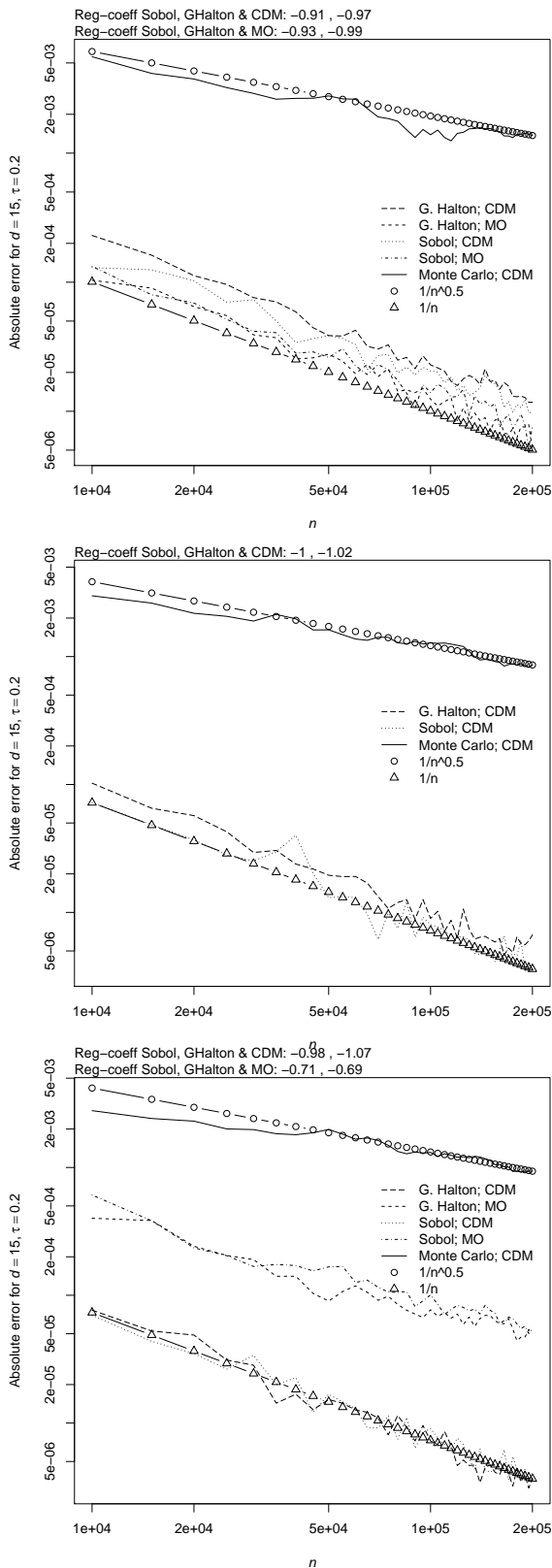
The main goal of this paper was to show how copula samples can be generated using quasi-random numbers. This is of interest when replacing PRNGs by QRNGs in applications involving dependent samples, possibly in higher dimensions. We have seen that different sampling approaches can be used, with a focus on the CDM approach and, additionally for Archimedean copulas, on the Marshall–Olkin algorithm. We have studied the error behaviour for both methods and have seen that in order to prove that the composed function  $\Psi \circ \phi_C$  is smooth enough to satisfy the Koksma–Hlawka bound for the error, sufficient conditions on the function  $\Psi$  are that it must have smooth higher order mixed partial derivatives. For the Marshall–Olkin algorithm, we have shown that for several Archimedean copula families, the corresponding transformation  $\phi_C$  is smooth enough to guarantee the good behaviour of the error bound. The superiority of QRNG over PRNG for copula sampling was illustrated on several examples, including a simulation addressing an application in the realm of finance and insurance. Most of the results in this paper are reproducible using the R packages `copula` and `qrng`.

Some ideas for future work would be to follow-up on Proposition 6 and to analyze the speed of decay of the Walsh coefficients of the composed function  $\Psi \circ \phi_C$ , based on assumptions on the speed of decay of the Walsh coefficients of  $\Psi$  and the properties of the sampling method  $\phi_C$ .

Concerning the copula-induced discrepancy studied in Section 4.2, a possible avenue for future research would be to construct point sets that directly minimize this discrepancy, without first transforming a (uniform-based) low-discrepancy sample. In addition, proving error bounds based on the  $L_2$ -discrepancy would be useful, as this discrepancy measure is easier to compute. Finally, numerically computing the copula-induced discrepancy for samples transformed either using



**Figure 11** Average absolute errors for the test functions  $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$  (top) and  $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$  (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.2 based on  $B = 25$  repetitions for  $d = 5$ ; the middle plot shows results for  $\Psi_1(\mathbf{u})$  and an exchangeable  $t$  copula with 3 degrees of freedom and Kendall's tau of 0.2.



**Figure 12** Average absolute errors for the test functions  $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$  (top),  $\Psi_2(\mathbf{u}) = K_C(\mathbf{C}(\mathbf{u})) + 1/2$  (middle), and  $\Psi_3(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$  (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.2 based on  $B = 25$  repetitions for  $d = 15$ ; the middle plot shows results for  $\Psi_1(\mathbf{u})$  and an exchangeable  $t$  copula with 3 degrees of freedom and Kendall's tau of 0.2.

the CDM or the MO algorithm would give us some insight as to how conservative the bound given in Proposition 7 is.

**Acknowledgements** We thank the Associate Editor and the two anonymous reviewers for their helpful comments, which helped us improve this paper. The first author wishes to thank SCOR for their financial support. The second and third authors acknowledge the support of NSERC through grants #5010 and #238959, respectively.

**References**

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

Caffisch R (1998) Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* 7:1–49

Cambanis S, Huang S, Simons G (1981) On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* 11(3):368–385

Constantine G, Savits T (1996) A multivariate Faa Di Bruno formula with applications. *Transactions of the American Mathematical Society* 348(2):503–520

Cranley R, Patterson TNL (1976) Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis* 13(6):904–914

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

Embrechts P, McNeil AJ, Straumann D (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls. *Risk Management: Value at Risk and Beyond* pp 176–223

Embrechts P, Lindskog F, McNeil AJ (2003) Modelling dependence with copulas and applications to risk management. In: Rachev S (ed) *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, pp 329–384

Fang KT, Kotz S, Ng KW (1990) *Symmetric Multivariate and Related Distributions*. Chapman & Hall/CRC

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

Faure H, Lemieux C (2009) Generalized Halton sequence in 2008: A comparative study. *ACM Transactions on Modeling and Computer Simulation* 19:Article 15

Glasserman P (2004) *Monte Carlo Methods in Financial Engineering*. Springer, New York

Halton JH (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik* 2:84–90

Hartering J, Kainhofer R, Tichy R (2004) Quasi-Monte Carlo algorithms for unbounded, weighted integration problems. *Journal of Complexity* 20:654–668

Hlawka E (1961) Über die diskrepanz mehrdimensionaler folgen mod 1. *Mathematische Zeitschrift* 77:273–284

Hlawka E, Mück R (1972) über eine transformation von gleichverteilten folgen ii. *Computing* 9(2):127–138

Hofert M (2010) *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. PhD thesis, University of Ulm

Hofert M, Mächler M, McNeil AJ (2012) Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis* 110:133–150

Hofert M, Mächler M, McNeil AJ (2013) Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique* 154(1):25–63

- Hong HS, Hickernell FJ (2003) Algorithm 823: Implementing scrambled digital sequences. *ACM Transactions on Mathematical Software* 29:95–109
- Horn R, Johnson C (2012) *Matrix Analysis*. Cambridge University Press, Cambridge
- Jaworski P, Durante F, Härdle WK, Rychlik T (eds) (2010) *Copula Theory and Its Applications*, Lecture Notes in Statistics – Proceedings, vol 198. Springer
- Joe H (2014) *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton
- Kurowicka D, Cooke RM (2007) Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics & Data Analysis* 51:2889–2906
- Lemieux C (2009) *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics, Springer, New York
- Marshall AW, Olkin I (1988) Families of multivariate distributions. *Journal of the American Statistical Association* 83(403):834–841
- Matoušek J (1998) On the  $L_2$ -discrepancy for anchored boxes. *Journal of Complexity* 14:527–556
- McNeil A, Frey R, Embrechts P (2005) *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton
- McNeil AJ, Nešlehová J (2009) Multivariate Archimedean copulas,  $d$ -monotone functions and  $l_1$ -norm symmetric distributions. *The Annals of Statistics* 37(5b):3059–3097
- Morokoff W (1998) Generating quasi-random paths for stochastic processes. *SIAM Review* 40(4):765–788
- Morokoff W, Cafilisch R (1994) Quasi-random sequences and their discrepancies. *SIAM J Sci Comput* 15(6):1251–1279
- Nelsen R (2006) *An Introduction to Copulas*, 2nd edn. Springer, New York
- Niederreiter H (1987) Point sets and sequences with small discrepancy. *Monatshefte für Mathematik* 104:273–337
- Niederreiter H (1992) *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA
- Nolan J (2014) *Stable distributions – models for heavy tailed data*. <http://academic2.american.edu/~jpnolan/stable/chap1.pdf>
- Owen AB (1995) Randomly permuted  $(t, m, s)$ -nets and  $(t, s)$ -sequences. In: Niederreiter H, Shiue PJS (eds) *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Lecture Notes in Statistics, vol 106, Springer-Verlag, New York, pp 299–317
- Owen AB (1997a) Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis* 34(5):1884–1910
- Owen AB (1997b) Scrambled net variance for integrals of smooth functions. *Annals of Statistics* 25(4):1541–1562
- Owen AB (2003) Variance and discrepancy with alternative scramblings. *ACM Transactions on Modeling and Computer Simulation* 13:363–378
- Owen AB (2005) Multidimensional variation for quasi-Monte Carlo. In: Fan J, Li G (eds) *International Conference on Statistics in honour of Professor Kai-Tai Fang's 65th birthday*, World Scientific Publications, Hackensack, NJ, pp 49–74
- Pillards T, Cools R (2006) Using box-muller with low discrepancy points. In: et al MG (ed) *ICCSA 2006, Lecture Notes in Computer Science*, vol 3984, Springer-Verlag, Berlin Heidelberg, pp 780–788
- Schmitz V (2003) *Copulas and Stochastic Processes*. PhD thesis, Institute of Statistics, Aachen University
- Sobol' IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* 7:86–112
- Sobol' IM (1973) Calculation of improper integrals using uniformly distributed sequences. *Soviet Mathematics Doklady* 14:734–738
- Tasche D (2008) *Capital Allocation to Business Units and Sub-Portfolios: the Euler Principle*. In: Resti A (ed) *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, Risk Books,

London, pp 423–453

- Wu F, Valdez EA, Sheris M (2006) Simulating Exchangeable Multivariate Archimedean Copulas and its Applications. *Communications in Statistics – Simulation and Computation* 36(5):1019–1034

## Appendix

For all the randomization schemes mentioned in Section 2, in addition to having a simple method to estimate the variance of the corresponding RQMC estimator, results giving exact expressions for this variance are also known and typically rely on using a well-chosen series expansion of the function  $\Psi$  of interest. The following result recalls this variance expression in the case of randomly digitally shifted net; see Lemieux (2009) for a detailed proof. This result is used in the proof of Proposition 6 in Section 4.1.

**Theorem 3 (Variance for randomly digitally shifted nets)** Let  $\tilde{P}_n = \{\tilde{v}_1, \dots, \tilde{v}_n\}$  be a randomly digitally shifted net in base  $b$  with corresponding RQMC estimator  $\hat{\mu}_n$  given by

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{v}_i)$$

and assume  $\text{Var}(\Psi(\mathbf{U})) < \infty$  for  $\mathbf{U} \sim U[0, 1]^d$ . Then we have that

$$\text{Var}(\hat{\mu}_n) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_d^*} |\hat{\Psi}(\mathbf{h})|^2,$$

where  $\hat{\Psi}(\mathbf{h})$  is the Walsh coefficient of  $\Psi$  at  $\mathbf{h}$ , given by

$$\hat{\Psi}(\mathbf{h}) = \int_{[0,1]^d} f(\mathbf{u}) e^{-2\pi i(\mathbf{h}, \mathbf{u})_b} d\mathbf{u}$$

where  $(\mathbf{h}, \mathbf{u})_b = \frac{1}{b} \sum_{j=1}^d \sum_{l=0}^{\infty} h_{j,l} u_{j,l+1}$  with  $h_{j,l}$  and  $u_{j,l}$  obtained from the base  $b$  expansion of  $h_j$  and  $u_j$ , respectively, and  $\mathcal{L}_d^* = \{\mathbf{h} \in \mathbb{Z}^d : \langle \mathbf{h}, \mathbf{v}_i \rangle_b \in \mathbb{Z}, \forall i = 1, \dots, n\}$  is the dual net of the deterministic net  $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$  that has been shifted to get  $\tilde{P}_n$ .

## Proofs

*Proof (Proof of Proposition 4)* We start by providing more details on the expression (21), which is given by:

$$\frac{\partial^l \Psi \circ \Phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}} = \sum_{1 \leq |\boldsymbol{\beta}| \leq l} \frac{\partial^{|\boldsymbol{\beta}|} \Psi}{\partial \beta_1 u_1 \dots \partial \beta_d u_d} \sum_{s=1}^l \sum_{(\mathbf{k}, \boldsymbol{\gamma}) \in p_s(\boldsymbol{\beta}, \boldsymbol{\alpha})} c_{\boldsymbol{\gamma}} \prod_{j=1}^s \frac{\partial^{|\boldsymbol{\gamma}_j|} \Phi_{C, k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial \gamma_{j,1} v_{\alpha_1} \dots \partial \gamma_{j,l} v_{\alpha_l}}$$

where  $\boldsymbol{\beta} \in \mathbb{N}_0^d$ ,  $|\boldsymbol{\beta}| = \sum_{j=1}^d \beta_j$ , and the set  $p_s(\boldsymbol{\beta}, \boldsymbol{\alpha})$  includes pairs  $(\mathbf{k}, \boldsymbol{\gamma})$  such that  $\mathbf{k}$  is an  $s$ -dimensional vector  $\mathbf{k} = (k_1, \dots, k_s)$  where each  $k_j \in \{1, \dots, d\}$ , and  $\boldsymbol{\gamma}$  is an  $sl$ -dimensional vector  $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s)$  where each  $\boldsymbol{\gamma}_j$  is an  $l$ -dimensional vector whose entries are either 0 or 1, and  $\sum_{j=1}^s \boldsymbol{\gamma}_{j,i} = 1$  for  $i \in \{1, \dots, l\}$ . Finally, the  $c_{\boldsymbol{\gamma}}$  are constants, which are defined in detail in Constantine and Savits (1996), along with further information on the precise definition of  $p_s(\mathbf{k}, \boldsymbol{\gamma})$ .

As mentioned in Section 4.1, a sufficient condition to show that  $\|\Psi \circ \Phi_C\|_{d,1} < \infty$  is to establish that all products of the form (22) are in  $L_1$ , which we recall is given by

$$\frac{\partial^{|\boldsymbol{\beta}|} \Psi}{\partial \beta_1 u_1 \dots \partial \beta_d u_d} \prod_{j=1}^s \frac{\partial^{|\boldsymbol{\gamma}_j|} \Phi_{C, k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial \gamma_{j,1} v_{\alpha_1} \dots \partial \gamma_{j,l} v_{\alpha_l}},$$

for  $s \in \{1, \dots, l\}$  and  $(\mathbf{k}, \boldsymbol{\gamma}) \in p_s(\boldsymbol{\beta}, \boldsymbol{\alpha})$ .

Recall also that for the MO algorithm,  $\phi_{C,l}$  is a function of  $v_1$  and  $v_{l+1}$  only, for  $l = 1, \dots, d$ . Hence the only partial derivatives of  $\phi_{C,l}$  that are nonzero are those with respect to variables in  $\{v_1, v_{l+1}\}$ .

Now, since we assume that (23) holds, then it means we just need to show that the product found in (22) is in  $L_1$ , under the conditions stated in the proposition. In turn, we first show that this holds if the following bounds hold for the mixed partial derivatives of  $\phi_C$ :

$$\int_0^1 \left| \frac{\partial \phi_{C,l}(v_1 = 1, v_{l+1})}{\partial v_{l+1}} \right| dv_{l+1} < \infty, \quad (26)$$

$$\int_{[0,1]^l} \left| \frac{\partial^2 \phi_{C,1}(v_1, v_2)}{\partial v_1 \partial v_2} \prod_{j=2}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right| dv_1 dv_2 \dots dv_l < \infty, \text{ and} \quad (27)$$

$$\int_{[0,1]^{l-1}} \left| \frac{\partial \phi_{C,r}(v_1, v_{r+1} = 1)}{\partial v_1} \left( \prod_{j=1, j \neq r}^{l-1} \left| \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right| dv_j \right) \right| dv_j < \infty \quad (28)$$

for all  $l \leq d+1$ ,

We have three cases to consider.

*Case 1:*  $1 \notin I$ . Then the product in (22) is given by

$$\prod_{j=1}^l \left| \frac{\partial \phi_{C,j}(v_1 = 1, v_{j+1})}{\partial v_{j+1}} \right|,$$

where we assumed w.l.o.g. that  $I = \{2, \dots, l+1\}$ ,  $s = l$  and  $k_j = j+1$  for  $j \in \{1, \dots, s\}$ . Since each term in the product depends on a distinct variable, the product is in  $L_1$  if (26) holds.

*Case 2:*  $1 \in I$  and  $j$  such that  $\gamma_{j,1} = 1$  has  $k_j + 1 \notin I$ . This case can be analyzed w.l.o.g. by assuming  $I$  is of the form  $I = \{1, \dots, r, r+2, \dots, l+1\}$  for some  $r \geq 1$ . In that case, the products in (22) are of the form

$$\left| \frac{\partial \phi_{C,r}(v_1, v_{r+1} = 1)}{\partial v_1} \prod_{j=1, j \neq r}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right|$$

and is thus in  $L_1$  as long as (28) holds.

*Case 3:*  $1 \in I$  and  $j$  such that  $\gamma_{j,1} = 1$  has  $k_j + 1 \in I$ . In this case, we can assume w.l.o.g. that  $I = \{1, \dots, l\}$  and therefore the products in (22) are of the form

$$\left| \frac{\partial^2 \phi_{C,r}(v_1, v_{r+1})}{\partial v_1 \partial v_{r+1}} \prod_{j=1, j \neq r}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right|$$

and is thus in  $L_1$  as long as (27) holds.

The last part of the proof is to show that (26), (27), and (28) hold. First we study the partial derivatives involved in these expressions and find they are given by:

$$\begin{aligned} \frac{\partial \phi_{C,1}(v_1, v_2)}{\partial v_1} &= \psi' \left( -\frac{\log(v_2)}{x_1} \right) \frac{\log v_2}{x_1^2} \frac{\partial x_1}{\partial v_1}, \\ \frac{\partial \phi_{C,1}(v_1, v_2)}{\partial v_2} &= -\psi' \left( -\frac{\log(v_2)}{x_1} \right) \frac{1}{x_1 v_2}, \\ \frac{\partial^2 \phi_{C,1}(v_1, v_2)}{\partial v_1 \partial v_2} &= \frac{\partial x_1}{\partial v_1} \frac{1}{v_2 x_1^2} \left[ \psi' \left( -\frac{\log v_2}{x_1} \right) - \frac{\log v_2}{x_1} \psi'' \left( -\frac{\log v_2}{x_1} \right) \right], \end{aligned}$$

where  $x_1 = F^{-1}(v_1)$  and  $\frac{\partial x_1}{\partial v_1} = 1/f(x_1)$ , where  $f$  is the pdf corresponding to  $F$ , which exists since we assumed  $F$  was continuous. Now, the partial derivatives with respect to either  $v_1$  or  $v_2$  are clearly non-negative for all  $v_1$  and  $v_2$ . Hence it is easy to see that (26) and (28) hold, because we can remove the absolute value inside the integrals and therefore, these integrals amount to take differences/sums of  $\phi_{C,r}(\cdot, \cdot)$  at different values over its domain, which obviously yields a finite value since  $\phi_{C,r}(\cdot, \cdot)$  always takes values in  $[0, 1]$ .

As for the mixed partial derivative with respect to  $v_1$  and  $v_2$ , our assumption on  $\psi'(t) + t\psi''(t)$  implies we have at most one sign change over the domain of the integral. If there is no sign change, the argument used in the previous paragraph to handle (26) and (28) can be used to show (27) is bounded. If there is one sign change, then we let  $t^*$  be such that

$$\psi'(t) + t\psi''(t) \leq 0 \quad \text{for } 0 \leq t \leq t^* \text{ and } \psi'(t) + t\psi''(t) \geq 0 \quad \text{for } t^* \geq t.$$

Then let  $q(v)$  be a function such that  $-\log q(v)/F^{-1}(v) = t^*$ . For instance, one can verify that for the Clayton copula,  $q(v) = e^{-\theta F^{-1}(v)}$ . When integrating the absolute value of the mixed partial derivative  $\partial^2 \phi_{C,1}(v_1, v_2)/\partial v_1 \partial v_2$ , we get

$$\begin{aligned} & \int_0^1 \frac{\partial x_1}{\partial v_1} \frac{1}{x_1^2} \left[ \int_0^{q(v_1)} \frac{1}{v_2} \left( \psi' \left( -\frac{\log v_2}{x_1} \right) - \frac{\log v_2}{x_1} \psi'' \left( -\frac{\log v_2}{x_1} \right) \right) dv_2 \right. \\ & \quad \left. + \int_{q(v_1)}^1 \frac{1}{v_2} \left( -\psi' \left( -\frac{\log v_2}{x_1} \right) + \frac{\log v_2}{x_1} \psi'' \left( -\frac{\log v_2}{x_1} \right) \right) dv_2 \right] dv_1 \\ &= 2 \int_0^1 \frac{\partial x_1}{\partial v_1} \frac{1}{x_1^2} \left[ \psi'(-\log q(v_1)/x_1) \log q(v_1) \right] dv_1 \\ &= -2t^* \psi'(t^*) \int_0^1 \frac{1}{F^{-1}(v_1)} \frac{\partial F^{-1}(v_1)'}{\partial v_1} dv_1 = -2t^* \psi'(t^*) \log F^{-1}(v_1) \Big|_0^1. \end{aligned} \quad (29)$$

Now, in most cases  $F^{-1}(1)$  is not bounded, and thus we cannot prove that  $\Psi \circ \phi_C$  has bounded variation. However, from there we can still get the upper bound on the error given in the result, by using a technique initially developed by Sobol' (1973) to handle improper integrals, and later by Hartinger et al (2004) to deal with unbounded integration problems taken w.r.t. to a measure that is not necessarily uniform (as studied in Section 4.2). Note that to apply their approach more easily, we need to make a small change and assume that rather than generating  $V$  as  $F^{-1}(v_1)$ , we use  $F^{-1}(1-v_1)$ , so that in our study of the variation above (via the integral of the absolute value of the mixed partial derivatives), the boundedness condition fails at  $v_1 = 0$  instead of  $v_1 = 1$ . Following the approach in Hartinger et al (2004) (see their Equation (24)) and taking  $\mathbf{c} = (1/pn, 0, \dots, 0)$ , the integration error satisfies

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \\ & \leq \frac{1}{pn} \Psi(1, \dots, 1) + D^*(P_n) V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C) + I_{rest} \end{aligned}$$

where  $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C)$  denotes the variation of  $\Psi \circ \phi_C$  over  $[\mathbf{c}, \mathbf{1}]$  and

$$I_{rest} = \left| \int_0^1 \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} - \int_{\mathbf{c}} \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} \right| \leq \frac{M}{pn} \text{ for some } M > 0,$$

since we assumed  $|\Psi(\mathbf{u})|$  was bounded. As for  $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C)$ , we can infer from the steps that led to (29) that it is bounded by a constant times  $\log F^{-1}(1-1/pn) \leq a \log n + \log c$  by assumption. Therefore there exists a constant  $K^{(d)}$  such that  $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C) \leq K^{(d)} \log n$ .

*Proof (Proof of Proposition 5)* Let  $p_l$  be such that  $P(V = l) = p_l$ , for  $l \geq 1$ . Let  $P_l = \sum_{k=1}^l p_k$  for  $l \geq 1$  and  $P_0 = 0$ . We also let  $\phi_C^l(v_2, \dots, v_{d+1}) = \phi_C(P_{l-1}, v_2, \dots, v_{d+1})$  for  $l \geq 1$  (transformation  $\phi_C$  when  $v_1$  generates the value  $l$  for  $V$ ). Consider a given value of  $n$  and low-discrepancy point set  $P_n$ . If we use inversion to generate  $V$ , then we have that the subset  $P_n^l = \{v_i : P_{l-1} < v_{i,1} \leq P_l\}$  will be used to produce copula samples with  $V = l$ . Let  $\tilde{n}_l = |P_n^l|$  and  $n_l = np_l$ . It is clear that if  $l$  becomes too large, then  $\tilde{n}_l$  will eventually be 0. Let  $L(n)$  be the largest value of  $l$  such that  $\tilde{n}_l > 0$ , and let  $\bar{p}_l = \tilde{n}_l/n$ . Then we can write

$$\left| \int_{[0,1]^{d+1}} \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} - \frac{1}{n} \sum_{i=1}^n \Psi \circ \phi_C(\mathbf{v}_i) \right|$$

$$\begin{aligned}
&\leq \left| \sum_{l=1}^{L(n)} p_l \left( \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{n_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right) \right| \\
&\quad + \sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\
&\leq \sum_{l=1}^{L(n)} \tilde{p}_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{\tilde{n}_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right| \\
&\quad + \sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\
&\quad + \sum_{l=1}^{L(n)} |(p_l - \tilde{p}_l)| \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\
&\leq \sum_{l=1}^{L(n)} (\tilde{p}_l A(n, d)) + B(n, d) + C(n, d),
\end{aligned}$$

where  $A(n, d)$ ,  $B(n, d)$ , and  $C(n, d)$  are bounds such that

$$\begin{aligned}
\left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{n_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right| &\leq A(n, d) \\
\sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| &\leq B(n, d) \\
\sum_{l=1}^{L(n)} |(p_l - \tilde{p}_l)| \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| &\leq C(n, d).
\end{aligned}$$

First, by definition of  $D^*(P_n)$  we have  $|\tilde{n}_l - n_l| \leq 2nD^*(P_n)$  and thus  $|\tilde{p}_l - p_l| \leq 2D^*(P_n)$ . Hence we can take  $C(n, d) = 2E(|\Psi(\mathbf{U})|)D^*(P_n)$ . Similarly, we can show that  $\sum_{l=L(n)+1}^{\infty} p_l \leq D^*(P_n)$  and can therefore take  $B(n, d) = E(|\Psi(\mathbf{U})|)D^*(P_n)$ . The analysis of the expression to be bounded by  $A(n, d)$  is more complicated. First, we note that under the assumption we have on  $\Psi$  and its partial derivatives, we need to show that the product in (22) is in  $L_1$ , but where each  $\phi_{C,k_j}$  is replaced by  $\phi_{C,k_j}^l$  for a given  $l$ . Since  $\phi_{C,k_j}^l$  is solely a function of  $v_{k_j+1}$ , then it means that the only relevant products to consider are of the form

$$\prod_{j=1}^r \frac{\partial \phi_{C,k_j}^l(v_{k_j+1})}{\partial v_{k_j+1}} \quad (30)$$

in which each term is of the form  $-\psi' \left( \frac{-\log v_{k_j+1}}{l} \right) \frac{1}{lv_{k_j+1}}$  which is non-negative for any  $v_{k_j+1}$ . Using a similar reasoning to the one used in the proof of Proposition 4 (to conclude that (26) and (28) hold), it is easy to see that (30) is in  $L_1$ .

What remains to be done is to analyze the discrepancy of  $P_n^l$ . That is, here we are looking for a bound on  $\sup_{\mathbf{z} \in \mathcal{I}^*} |E(\mathbf{z}; P_n^l)|$ , where we recall that  $\mathcal{I}^*$  is the set of intervals of  $[0, 1]^d$  of the form  $\mathbf{z} = \prod_{j=1}^d [0, z_j)$ , where  $0 < z_j \leq 1$ . So consider a given  $\mathbf{z} \in [0, 1]^d$ . Then  $E(\mathbf{z}; P_n^l) = A(\mathbf{z}; P_n^l)/\tilde{n}_l - \lambda(\mathbf{z})$ . Let  $\mathbf{z}_1 = (P_l, \mathbf{z})$  and  $\mathbf{z}_2 = (P_{l-1}, \mathbf{z})$ , which are both in  $[0, 1]^{d+1}$ . Note that  $A(\mathbf{z}_1; P_n) - A(\mathbf{z}_2; P_n) = A(\mathbf{z}; P_n^l)$ . By definition of  $D^*(P_n)$ , it is not hard to see that

$$\left| \frac{A(\mathbf{z}_1; P_n) - A(\mathbf{z}_2; P_n)}{n} - p_l \lambda(\mathbf{z}) \right| \leq 2D^*(P_n)$$

and therefore

$$\left| \frac{A(\mathbf{z}; P_n^l)}{\tilde{n}_l} - \frac{n_l}{\tilde{n}_l} \lambda(\mathbf{z}) \right| \leq 2D^*(P_n) \frac{n}{\tilde{n}_l}.$$

Using the fact that  $|\tilde{n}_l - n_l| \leq 2nD^*(P_n)$ , after some further simplifications we get that

$$\left| \frac{A(\mathbf{z}; P_n^l)}{\tilde{n}_l} - \lambda(\mathbf{z}) \right| \leq 4D^*(P_n) \frac{n}{\tilde{n}_l}.$$

Hence we can take  $A(n, d) = 4D^*(P_n) \frac{n}{\tilde{n}_l}$  and then get  $\sum_{l=1}^{L(n)} \tilde{p}_l A(n, d) \leq 4L(n)D^*(P_n)$ . To show that the overall bound for the integration error is of the form  $(\log n)D^*(P_n)$  times a constant, we simply need to show that  $L(n) \in O(\log n)$ . But this follows from our assumptions on  $P_n$  and  $F$ , since by definition,  $L(n)$  is the largest integer such that  $1 - F(L(n)) > 1/pn$  but we also have  $1 - F(L(n)) \leq cq^{L(n)}$ , hence

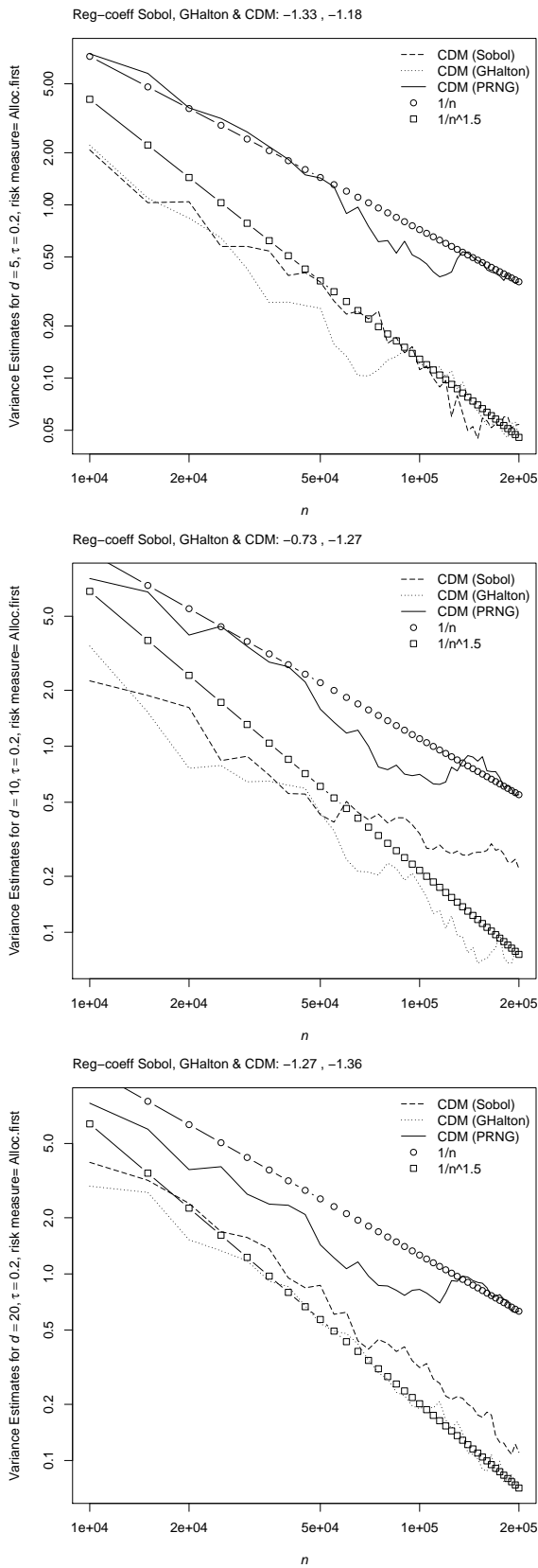
$$1/pn < cq^{L(n)} \Leftrightarrow L(n) \log(1/q) - \log c < \log p + \log n$$

and thus  $L(n) \leq (\log n + \log p + \log c) / \log(1/q)$ , as required.

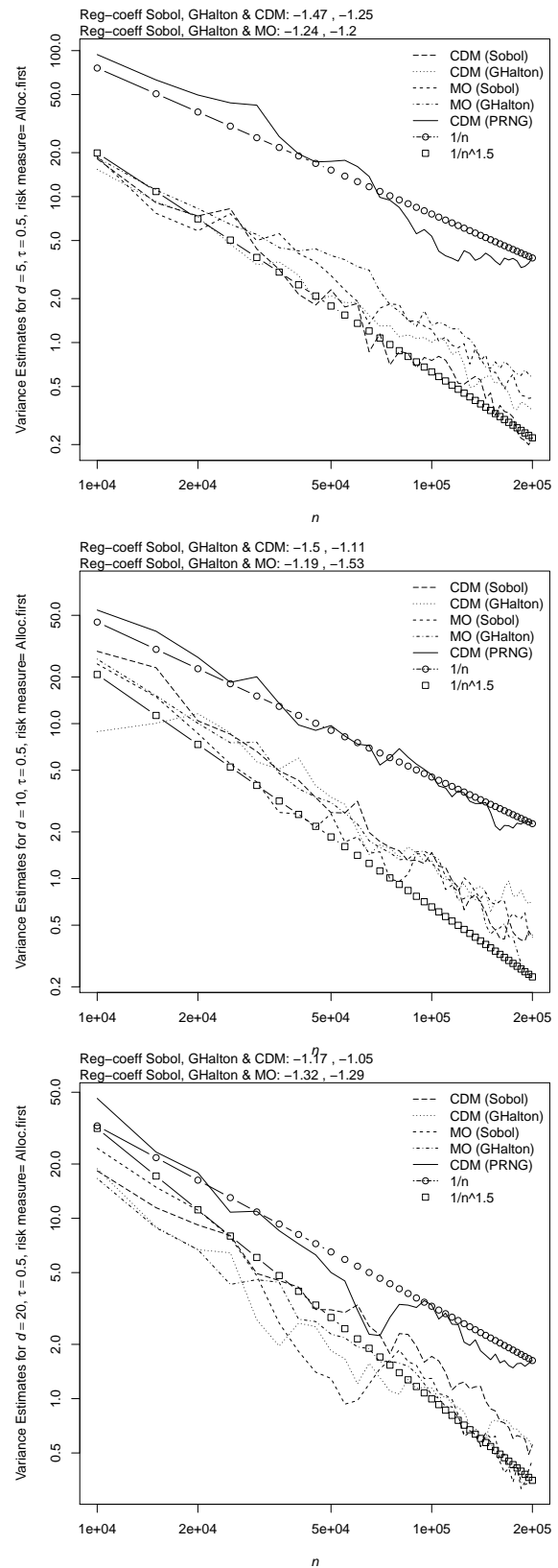
## Online supplement

### Additional numerical results

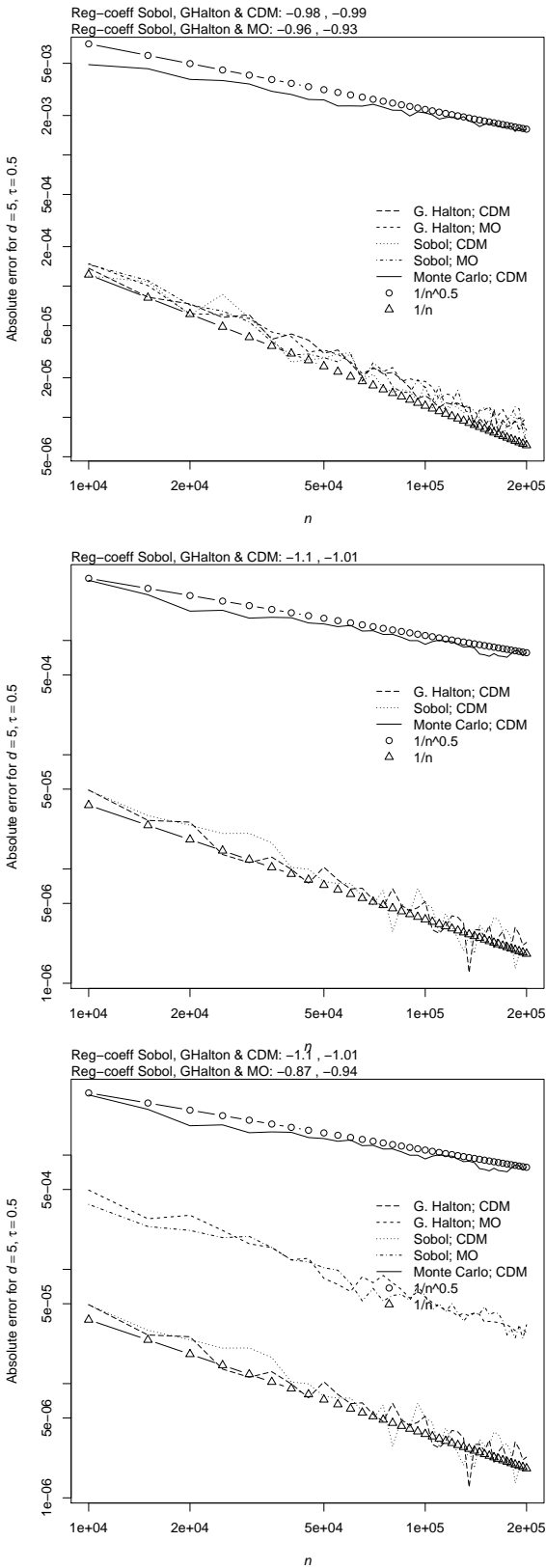
Here we provide a few additional results for the experimental setup described in Section 5, namely for the finance and insurance examples (see Figures 13 and 14) and then for the test functions (see Figures 15 and 16).



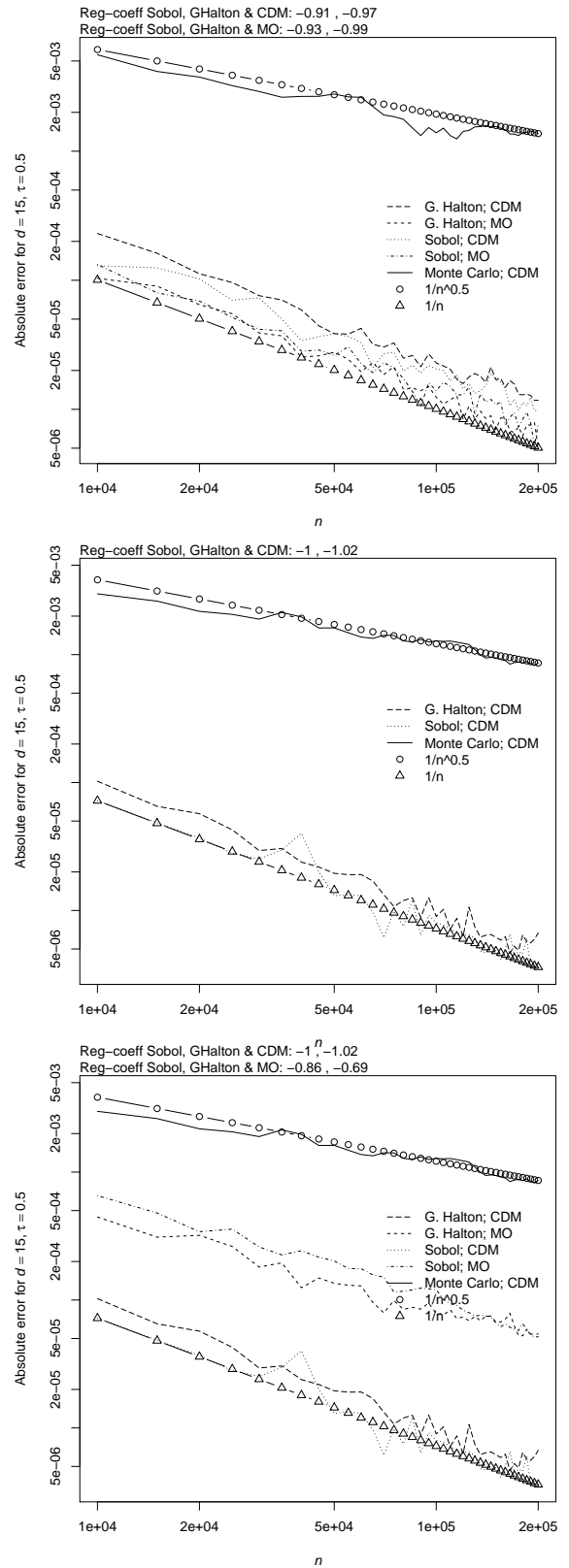
**Figure 13** Variance estimates for the functional Allocation First for lognormal margins and an exchangeable  $t$  copula with three degrees of freedom such that Kendall's tau equals 0.2 based on  $B = 25$  repetitions for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).



**Figure 14** Variance estimates for the functional Allocation First with Pareto margins and for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on  $B = 25$  repetitions for  $d = 5$  (top),  $d = 10$  (middle) and  $d = 20$  (bottom).



**Figure 15** Average absolute errors for the test functions  $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$  (top) and  $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$  (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on  $B = 25$  repetitions for  $d = 5$ : the middle plot shows results for  $\Psi_1(\mathbf{u})$  and an exchangeable  $t$  copula with three degrees of freedom and Kendall's tau of 0.2.



**Figure 16** Average absolute errors for the test functions  $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$  (top),  $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$  (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on  $B = 25$  repetitions for  $d = 15$ : the middle plot shows results for  $\Psi_1(\mathbf{u})$  and an exchangeable  $t$  copula with three degrees of freedom and Kendall's tau of 0.2.