

Importance Sampling and Stratification for Copula Models

Philipp Arbenz, Mathieu Cambou, Marius Hofert, Christiane Lemieux, and Yoshihiro Taniguchi

Dedicated to Ian H. Sloan on the occasion of his 80th birthday.

Abstract An importance sampling approach for sampling from copula models is introduced. The proposed algorithm improves Monte Carlo estimators when the functional of interest depends mainly on the behaviour of the underlying random vector when at least one of its components is large. Such problems often arise from dependence models in finance and insurance. The importance sampling framework we propose is particularly easy to implement for Archimedean copulas. We also show how the proposal distribution of our algorithm can be optimized by making a connection with stratified sampling. In a case study inspired by a typical insurance application, we obtain variance reduction factors sometimes larger than 1000 in comparison to standard Monte Carlo estimators when both importance sampling and quasi-Monte Carlo methods are used.

1 Introduction

Many applications in finance and insurance lead to the problem of calculating a functional of the form $\mu = \mathbb{E}(\Psi_0(\mathbf{X}))$, where $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Psi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function. If the components of \mathbf{X} cannot be assumed to be independent, it is popular

Philipp Arbenz
SCOR
e-mail: philipp.arbenz@gmail.com

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

Marius Hofert · Christiane Lemieux · Yoshihiro Taniguchi (✉)
University of Waterloo, Waterloo, Ontario, Canada
e-mail: marius.hofert@uwaterloo.ca; clemieux@uwaterloo.ca; ytaniguc@uwaterloo.ca

to model the distribution function H of \mathbf{X} with a copula C , such that $H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, $\mathbf{x} \in \mathbb{R}^d$, where $F_j(x) = \mathbb{P}(X_j \leq x)$, $j \in \{1, \dots, d\}$, are the univariate margins of H and $C : [0, 1]^d \rightarrow [0, 1]$ is a copula. A copula allows one to separate the dependence structure from the marginal distributions, which is useful for constructing multivariate stochastic models. We assume the reader to be familiar with copulas and refer to [15] or [17] for an introduction; see also Section 2 for important background information.

A drawback of using such flexible models is that an analytical form for the quantity of interest $\mathbb{E}(\Psi_0(\mathbf{X}))$ rarely exists, and thus numerical methods must be applied to evaluate it. Preferably, the employed techniques should be applicable to high-dimensional problems, which are common in finance. An advantage of Monte Carlo (MC) simulation is that the rate of convergence of its error is independent of the dimensionality of a given problem. Nevertheless, the convergence rate of plain MC is generally slow so that MC is often combined with some variance reduction technique (VRT) to improve the precision of estimators.

Importance sampling (IS) is a VRT often used for rare event simulations. IS attempts to reduce the variance of the MC estimator of $\mathbb{E}(\Psi_0(\mathbf{X}))$ by sampling \mathbf{X} more frequently from the important region where $|\Psi_0(\mathbf{X})|$ is large. While there are many publications that design IS for Gaussian and t -copula models, [2, 8, 12, 19] for instance, not much attention has been given to IS for other types of copulas, including Archimedean copulas. To our knowledge, [3] is the only work which develops IS for Archimedean copulas.

The main contribution of this paper is the study of IS techniques that do not rely on a specific copula structure. We consider the case where the functional Ψ_0 of interest depends mainly on the behaviour of the random vector \mathbf{X} when at least one of the components is large. Such problems often arise from dependence models in the realm of finance and insurance. We propose a new IS framework for this setup which can be implemented for all classes of copula models from which sampling is feasible. The main idea of our proposed IS approach is to oversample sets of the form $[0, 1]^s \setminus [0, \lambda_k]^s$ for $0 \leq \lambda_1 \leq \dots \leq \lambda_M \leq 1$. Explicit algorithms are given in the case of Archimedean copulas. We also examine how to optimally choose the proposal distribution by making a connection with stratified sampling (SS), which is then used to propose yet another estimator based on our general IS setup.

While the plain MC method generates samples based on pseudo-random numbers, quasi-Monte Carlo (QMC) methods use a low-discrepancy sequence (LDS) to draw samples. An LDS has the property of covering the unit cube $[0, 1]^d$ more uniformly than pseudo-random numbers generally do. This usually leads to approximations whose error converges to 0 faster than with MC. Furthermore, these methods can be randomized in a way that preserves their low discrepancy but allows for error estimation. Such randomized QMC methods can thus be seen as a VRT. QMC has been primarily used for multinormal models and has shown substantial improvements over plain MC. Recently, its effectiveness for sampling copula models was studied and demonstrated theoretically and empirically in [1]. Building up on that work, it is natural to try to combine QMC with our proposed IS approach.

The rest of this work is organized as follows. In Section 2, we motivate our proposed IS method and give the necessary background on Archimedean copulas and QMC methods for copulas. In Section 3 we introduce a general IS setup for copula models, and then show in Section 4 how to exploit the Marshall–Olkin stochastic representation of Archimedean copulas to design an efficient sampling algorithm for IS. We show that the proposed IS scheme is very similar to stratified sampling and then develop sampling methods for SS estimators. In Section 5, we derive variance expressions for the IS and SS estimators. By minimizing such variance expressions, we derive the optimal calibration for our proposal distribution, for both IS and SS estimators. In Section 6, we investigate the effectiveness of the proposed IS and SS schemes using numerical experiments. All proofs are deferred to the appendix.

2 Motivation and Background

In a copula model, we may write $\mu = \mathbb{E}(\Psi_0(\mathbf{X})) = \mathbb{E}(\Psi(\mathbf{U}))$, 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 given by

$$\Psi(u_1, \dots, u_d) = \Psi_0(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \quad (1)$$

and $F_j^{-1}(p) = \inf\{x \in \mathbb{R} : F_j(x) \geq p\}$ for $j \in \{1, \dots, d\}$.

If C and F_1, \dots, F_d are known, we can use MC simulation to estimate $\mathbb{E}(\Psi(\mathbf{U}))$. For a random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ of \mathbf{U} , the MC estimator of $\mathbb{E}(\Psi(\mathbf{U}))$ is

$$\hat{\mu}_{\text{MC},n} = \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i). \quad (2)$$

In this paper, we consider the case where Ψ is large only when at least one of its arguments is close to 1, or equivalently, if at least one of the components of \mathbf{X} is large. This assumption is inspired by several applications in insurance:

- The fair premium of a stop loss cover with deductible D is $\mathbb{E}(\max\{\sum_{j=1}^d X_j - D, 0\})$. The corresponding functional is $\Psi(\mathbf{u}) = \max\{\sum_{j=1}^d F_j^{-1}(u_j) - D, 0\}$; see the left-hand side of Figure 1 for a contour plot of Ψ for two Pareto margins.
- Risk measures for an aggregate sum $S = \sum_{j=1}^d X_j$, such as value-at-risk, $\text{VaR}_\alpha(S)$, or expected shortfall, $\text{ES}_\alpha(S)$, $\alpha \in (0, 1)$, cannot in general be written as an expectation of type $\mathbb{E}(\Psi_0(\mathbf{X}))$. However, they are functionals of the aggregate distribution function $F_S(x) = \mathbb{P}(S \leq x) = \mathbb{E}(\Psi(\mathbf{U}; x))$, where $\Psi(\mathbf{u}; x) = I_{\{F_1^{-1}(u_1) + \dots + F_d^{-1}(u_d) \leq x\}}$. We can therefore write

$$\text{VaR}_\alpha(S) = \inf\{x \in \mathbb{R} : \mathbb{E}(\Psi(\mathbf{U}; x)) \geq \alpha\}, \quad \text{ES}_\alpha(S) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(S) du, \quad (3)$$

which depend only on those x for which $\mathbb{E}(\Psi(\mathbf{U}; x)) \geq \alpha$ holds. This is determined by the tail behaviour of S , which is strongly influenced by the properties of the copula C when at least one component is close to 1. Note that capital

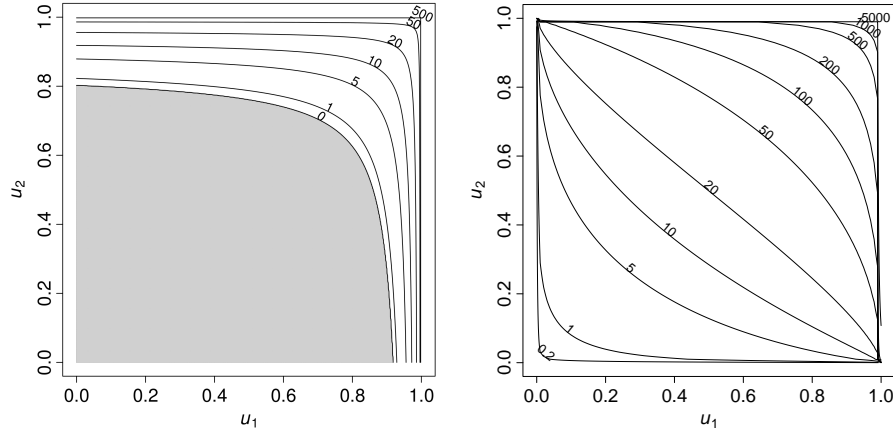


Fig. 1 *Left:* Contour lines for the excess function $\Psi(u_1, u_2) = \max\{F_1^{-1}(u_1) + F_2^{-1}(u_2) - 10, 0\}$, where the margins are Pareto distributed with $F_1(x) = 1 - (1+x/4)^{-2}$ and $F_2(x) = 1 - (1+x/8)^{-2}$. The grey area indicates where Ψ is zero. *Right:* Contour lines for the product function $\Psi(u_1, u_2) = F_1^{-1}(u_1)F_2^{-1}(u_2)$, where $X_1 \sim \text{LN}(2, 1)$ and $X_2 \sim \text{LN}(1, 1.5)$.

allocation methods such as the Euler principle for expected shortfall behave similarly, see [21] and [15], page 260.

Note that in this framework we follow the convention of [15, Remark 2.1] that \mathbf{X} refers to a loss and $-\mathbf{X}$ to a profit, which is more common in an actuarial context. One could have equally well worked with the profit-and-loss random variable $-\mathbf{X}$ by changing the area of interest to where components of \mathbf{X} are small.

2.1 Archimedean Copulas and Sampling Methods

Archimedean copulas form a popular class of copulas in actuarial science and risk management, as they can capture various types of tail dependence. An Archimedean copula admits the representation

$$C(u_1, \dots, u_d) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \quad (4)$$

where ψ is a univariate function called *generator* and is such that $\psi : [0, \infty) \rightarrow [0, 1]$ with $\psi(0) = 1$ and $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$; also $\psi(t)$ is continuous and strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$. We review two sampling techniques applicable to Archimedean copulas.

Conditional distribution method

The conditional distribution method (CDM) is a sampling technique that in principle works for any copula. For $j \in \{2, \dots, d\}$, let

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

be the conditional distribution of the j th component given the first $j - 1$ components. As a function of u_j , $C_{j|1\dots j-1}(u_j | u_1, \dots, u_{j-1})$ is a univariate distribution function on $[0, 1]$ and thus can be sampled via inversion. Doing this iteratively in j based on the previously computed component samples leads to a sample from C according to the CDM. The efficiency of this sampling method depends on the computational cost required to evaluate the conditional quantile functions $C_{j|1\dots j-1}^{-1}(u_j | u_1, \dots, u_{j-1})$, which in many cases are not available in closed form. For Archimedean copulas, there exists a more efficient sampling method, which we now describe.

Marshall–Olkin algorithm

It is well established that ψ induces an Archimedean copula for any dimension $d \geq 2$ if and only if ψ is the Laplace–Stieltjes transform of a distribution function of some positive random variable V , the so-called *frailty*. Based on V , one can derive the stochastic representation

$$\left(\psi\left(\frac{E_1}{V}\right), \dots, \psi\left(\frac{E_d}{V}\right) \right) \sim C, \quad (5)$$

where $E_1, \dots, E_d \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ are independent of the positive frailty random variable V whose Laplace–Stieltjes transform is ψ . This sampling method is known as *Marshall–Olkin (MO) algorithm*; see [14].

For many popular Archimedean copulas, the frailty random variable V from the MO algorithm has a known distribution, for instance V is Gamma distributed for Clayton copulas; see Table 1 for information about five popular Archimedean copulas and the corresponding frailty random variables V , and see [11, Table 1] for the details concerning Table 1. In Section 4 we develop an IS algorithm that exploits the MO representation of Archimedean copulas.

Family	Parameter	$\psi(t)$	V
Ali-Mikhail-Haq	$\theta \in [0, 1)$	$(1 - \theta) / (\exp(t) - \theta)$	$\text{Geo}(1 - \theta)$
Clayton	$\theta \in (0, \infty)$	$(1 + t)^{-1/\theta}$	$\Gamma(1/\theta, 1)$
Frank	$\theta \in (0, \infty)$	$-\log(1 - (1 - e^{-\theta}) \exp(-t)) / \theta$	$\text{Log}(1 - e^{-\theta})$
Gumbel	$\theta \in [1, \infty)$	$\exp(-t^{1/\theta})$	$\text{Stable}(1/\theta, 1, \cos^\theta(\pi/(2\theta)), I_{\{\theta=1\}}; 1)$
Joe	$\theta \in [1, \infty)$	$1 - (1 - \exp(-t))^{1/\theta}$	$\text{Sibuya}(1/\theta)$

Table 1 Popular Archimedean generators and corresponding frailty distributions.

2.2 Quasi-Monte Carlo and Copula Models

The combination of QMC and copulas is studied in depth in [1]. To describe how it works, let $\eta : [0, 1)^{d+k} \rightarrow [0, 1)^d$ for $k \geq 0$ be some transformation function such that $\eta(\mathbf{U}') \sim C$ for $\mathbf{U}' \sim \text{U}[0, 1)^{d+k}$. The choice of η corresponds to the choice of sampling methods for C , such as CDM or MO. The plain MC estimator (2) thus becomes

$$\hat{\mu}_{\text{MC},n} = \frac{1}{n} \sum_{i=1}^n \Psi(\eta(\mathbf{U}'_i)), \quad \mathbf{U}'_i \stackrel{\text{ind.}}{\sim} \text{U}[0, 1)^{d+k}. \quad (6)$$

To use QMC, we replace the point set $\{\mathbf{U}'_i, i = 1, \dots, n\}$ by a low-discrepancy point set. The choice of sampling algorithm η is not very important to control the MC error, but it is for QMC, as explained in [1]. The sampling algorithms we propose in this work are applicable to both MC and QMC, and numerical results for both methods are reported in Section 6. For QMC we use a Sobol' sequence [20] and apply to it a randomization based on a digital-shift (see [13, Section 6.2.2]) so that we can construct unbiased estimators and compute confidence intervals for the quantity of interest by using replication.

3 Importance Sampling for Copula Models

IS is a popular variance reduction technique for rare event simulations. Suppose we want to estimate $\mathbb{E}(\Psi(\mathbf{U}))$ where $\mathbf{U} \sim C$, for a d -dimensional copula C . In IS, we draw samples from some proposal distribution $\tilde{\mathbf{U}} \sim G$ and construct the estimator

$$\hat{\mu}_{\text{IS},n} = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{\mathbf{U}}_i) w(\tilde{\mathbf{U}}_i), \quad \tilde{\mathbf{U}}_i \stackrel{\text{ind.}}{\sim} G, \quad (7)$$

where $w(\mathbf{u}) = \frac{dC(\mathbf{u})}{dG(\mathbf{u})}$ is the Radon-Nikodym derivative of C with respect to G . The function w works as a weight function so that the estimator remains unbiased after changing the distribution. Intuitively, the variance of the IS estimator is smaller than the variance of the plain MC estimator if the proposal distribution is concentrated around the important region, which we characterized in Section 2 as the region where the maximal component of a sample point is close to 1.

In order to define the proposal distribution G , we suggest a mixing approach by taking a weighted average of a multivariate distribution function $C_\lambda : [0, 1]^d \rightarrow [0, 1]$ over different values of λ . Let F_Λ denote a discrete distribution function of a random variable $\Lambda : \Omega \mapsto [0, 1]$, defined by $q_k := \mathbb{P}(\Lambda = \lambda_k)$, $k = 1, \dots, M$. We then define the distribution function G of $\tilde{\mathbf{U}}$ as a mixture of C_λ with respect to F_Λ :

$$G(\mathbf{u}) = \sum_{k=1}^M q_k C_{\lambda_k}(\mathbf{u}),$$

where C_λ is a distorted version of the copula C itself that concentrates samples in a region of the form $[0, 1]^d \setminus [0, \lambda]^d$. Note that the C_λ we will construct (see (8)) is a copula only if $C(\lambda \mathbf{1}) = 0$, but C_λ does not need to be copula for our approach to work.

We will see that this mixture approach is natural in order to allow C to be absolutely continuous with respect to G . In particular, the absolute continuity is guaranteed for any copula C if the following assumption is satisfied.

Assumption 1. *The random variable Λ satisfies $\mathbb{P}(\Lambda = 0) > 0$.*

In order to obtain a well defined weight function w and an unbiased estimator $\hat{\mu}_{\text{IS},n}$, Assumption 1 must be fulfilled. Note that this assumption does not require

particular conditions on C . Although it seems restrictive, we will see that it is also needed to have a consistent estimator $\hat{\mu}_{\text{IS},n}$. Moreover, ensuring $\mathbb{P}(\Lambda = 0) > 0$ can be seen as a form of defensive mixture sampling, where a fraction of samples are drawn from the original distribution [9]. Defensive sampling bounds the IS weights away from infinity (as will be seen in Lemma 2) so that the resulting estimator has a finite variance. To that end, we assume Assumption 1 to be satisfied in what follows.

The construction of the proposal distribution G as a C_λ -mixture directly yields a sampling method, as one can draw a realization of G by first drawing $\Lambda \sim F_\Lambda$ and then $\tilde{\mathbf{U}} \sim C_\Lambda$. Therefore, the following algorithm can be used to construct $\hat{\mu}_{\text{IS},n}$:

Algorithm 1 General IS algorithm for copulas

- 1: Fix $n \in \mathbb{N}$.
 - 2: Draw $\Lambda_i \sim F_\Lambda, i \in \{1, \dots, n\}$.
 - 3: Draw $\tilde{\mathbf{U}}_i \sim C_{\Lambda_i}, i \in \{1, \dots, n\}$.
 - 4: Calculate $w(\tilde{\mathbf{U}}_i) = dC(\tilde{\mathbf{U}}_i)/dG(\tilde{\mathbf{U}}_i), i \in \{1, \dots, n\}$.
 - 5: Return $\hat{\mu}_{\text{IS},n} = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{\mathbf{U}}_i)w(\tilde{\mathbf{U}}_i)$.
-

The following lemma establishes consistency and asymptotic normality of the estimator $\hat{\mu}_{\text{IS},n}$.

Lemma 1. *Suppose that $\text{Var}(\Psi(\mathbf{U})) < \infty$ and that $w(\cdot) \leq B$ for a constant $B < \infty$. Then*

1. $\hat{\mu}_{\text{IS},n}$ converges \mathbb{P} -almost surely to μ ;
2. $\sigma^2 = \text{Var}(\Psi(\tilde{\mathbf{U}})w(\tilde{\mathbf{U}})) < \infty$ and $\sqrt{n}(\hat{\mu}_{\text{IS},n} - \mu) \rightarrow \mathcal{N}(0, \sigma^2)$ in distribution.

We will later show that under some mild assumptions on F_Λ , the weight function will indeed be bounded on $[0, 1]$.

The form of C_λ we propose to work with is the distribution of \mathbf{U} conditioned on the event that at least one of its components exceeds λ :

$$\begin{aligned} C_\lambda(\mathbf{u}) &= \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d \mid \max\{U_1, \dots, U_d\} > \lambda) \\ &= \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d \mid \mathbf{U} \notin [0, \lambda]^d) \\ &= \frac{C(\mathbf{u}) - C(\min\{u_1, \lambda\}, \dots, \min\{u_d, \lambda\})}{1 - C(\lambda \mathbf{1})}, \end{aligned} \quad (8)$$

where $\lambda \mathbf{1} = \lambda(1, \dots, 1) = (\lambda, \dots, \lambda) \in [0, 1]^d$. By putting mass of Λ on $(0, 1)$, we can put more weight on the region of the copula where at least one component is large. For instance, if F_Λ is discrete and $\mathbb{P}(\Lambda = 0) = \mathbb{P}(\Lambda = 0.9) = 0.5$, then 50% of the samples of $\tilde{\mathbf{U}}$ are constrained to lie only in $[0, 1]^d \setminus [0, 0.9]^d$ while the other 50% of the samples will lie on $[0, 1]^d$. Note that the mass on $[0, 1]^d \setminus [0, 0.9]^d$ would then be higher than 50% since we can still sample from $[0, 1]^d \setminus [0, 0.9]^d$ when $\Lambda = 0$. On the other hand, the case $\mathbb{P}(\Lambda = 0) = 1$ yields $G = C$ since $C_\lambda = C$ for $\lambda = 0$.

We now describe how the weight function w based on the above choice for C_λ can be calculated.

Theorem 1. *The Radon–Nikodym derivative $w(\mathbf{u}) = dC(\mathbf{u})/dG(\mathbf{u})$ is given by*

$$w(\mathbf{u}) = \left(\sum_{k=1}^M \frac{I_{\{\lambda_k \leq \max\{u_1, \dots, u_d\}\}}}{1 - C(\lambda_k \mathbf{1})} q_k \right)^{-1}.$$

In order to simplify the notation, let $\tilde{w} : [0, 1] \rightarrow [0, \infty)$ be defined as

$$\tilde{w}(u) = \left(\sum_{k=1}^M \frac{I_{\{\lambda_k \leq u\}}}{1 - C(\lambda_k \mathbf{1})} q_k \right)^{-1}. \quad (9)$$

Therefore we have that $w(\mathbf{u}) = \tilde{w}(\max\{u_1, \dots, u_d\})$. In order to evaluate \tilde{w} , it is sufficient to calculate (or approximate) $C(\lambda_k \mathbf{1})$ for $k \in \{1, \dots, M\}$. These values must be calculated only once and thus this approach is fast and can be easily implemented. In particular, the density of C does not have to be evaluated to calculate w (or \tilde{w}). This is an advantage in comparison to most other IS algorithms, for which the existence of the density of C is required.

Lemma 2. *Under Assumption 1, \tilde{w} is bounded from above by $\mathbb{P}(\Lambda = 0)^{-1}$ on $[0, 1]$.*

As a consequence of Lemma 2, Assumption 1 is not only sufficient to obtain existence of the weights, but it also guarantees that they are bounded. In virtue of Lemma 1, this guarantees consistency and asymptotic normality of the IS estimator.

Note that our approach could be generalized to other forms of C_λ and F_Λ (e.g., not necessarily discrete). In such cases the evaluation of the weight function \tilde{w} might be more demanding and require the use of numerical integration schemes.

4 Importance Sampling Algorithm for Archimedean Copulas

While the IS method from the previous section can be applied to any copula, sampling from C_λ is in general difficult. A possible solution could be to use rejection sampling, but we do not pursue this approach here as we expect it would not work very well with QMC sampling. In this section, we instead focus on developing sampling algorithms for $\mathbf{U} | U_{(d)} := \max\{U_1, \dots, U_d\} > \lambda$ when \mathbf{U} follows an Archimedean copula with generator ψ . This corresponds to Step 3 in Algorithm 1. In light of (5), we have $(U_1, \dots, U_d) \stackrel{d}{=} (\psi(\frac{E_1}{V}), \dots, \psi(\frac{E_d}{V}))$ where $E_j \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ and V is the corresponding frailty random variable. The condition $U_{(d)} > \lambda$ can then be written as $E_{(1)} < \psi^{-1}(\lambda)V$, where $E_{(1)} \sim \text{Exp}(d)$ is the first order statistic of $\{E_1, \dots, E_d\}$. In summary, sampling from $\mathbf{U} | U_{(d)} > \lambda$ is equivalent to sampling from $(E_1, \dots, E_d, V) | E_{(1)} < \psi^{-1}(\lambda)V$. Algorithm 2 summarizes the sampling method for this conditional distribution where we let $\gamma = \psi^{-1}(\lambda)$. Proposition 1 asserts that samples from Algorithm 2 have the right distribution.

Proposition 1. *Let E_1, \dots, E_d be iid positive random variables and V be a positive random variable independent of the E_j 's. Then a sample (E_1, \dots, E_d, V) constructed as in Steps 1–3 of Algorithm 2 has the distribution $(E_1, \dots, E_d, V) | (E_{(1)} < \gamma V)$.*

Algorithm 2 Sampling Step of the IS algorithm for Archimedean copulas**Require:** $0 < \gamma = \psi^{-1}(\lambda) < \infty$.1: Draw $(E_{(1)}, V) | (E_{(1)} < \gamma\mathcal{V})$.2: Draw $(E_1, \dots, E_d) | E_{(1)}$.3: Let $U_j = \psi(E_j/V)$ for $j \in \{1, \dots, d\}$.4: **return** (U_1, \dots, U_d) .

While Proposition 1 holds for general (positive) E_j 's and V , we now give detailed explanations of how to do the sampling for Steps 1 and 2 of Algorithm 2, i.e., when $E_j \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$ and V is the frailty random variable. We assume that V is continuous for the derivations below. We need only minor modifications for the discrete case.

Step 1: Sample $(E_{(1)}, V) | (E_{(1)} < \gamma\mathcal{V})$

We want to sample from the joint distribution of $(E_{(1)}, V)$ conditioned on the event $(E_{(1)} < \gamma\mathcal{V})$. Let $f_{E_{(1)}}(x)$ and $f_V(v)$ be the density of $E_{(1)}$ and V , respectively. Further, let $f_{(E_{(1)}, V) | (E_{(1)} < \gamma\mathcal{V})}(x, v)$ be the conditional joint density of $(E_{(1)}, V)$ given $E_{(1)} < \gamma\mathcal{V}$. Then by independence of $E_{(1)}$ and V

$$f_{(E_{(1)}, V) | (E_{(1)} < \gamma\mathcal{V})}(x, v) = \beta f_{E_{(1)}}(x) f_V(v) I(x < \gamma v), \quad (10)$$

where $1/\beta = \mathbb{P}(E_{(1)} < \gamma\mathcal{V}) = \mathbb{P}(U_{(d)} > \lambda) = (1 - C(\lambda \mathbf{1})) = (1 - \psi(d\psi^{-1}(\lambda)))$. We use conditional sampling to sample from this density, that is, we first sample V from the marginal conditional density $f_{V | (E_{(1)} < \gamma\mathcal{V})}$ of (10) then draw $E_{(1)}$ from (10) given V . Note that

$$f_{V | (E_{(1)} < \gamma\mathcal{V})}(v) = \beta f_V(v) \int_0^{\gamma v} f_{E_{(1)}}(x) dx = \beta f_V(v) (1 - \exp(-d\gamma v)). \quad (11)$$

Unfortunately, the density (11) does not belong to a known parametric family for most Archimedean copulas. Nonetheless, there exist efficient numerical algorithms that allow one to sample from a univariate distribution given its probability density function. For instance, the NINIGL Algorithm in [4] achieves this through numerical inversion techniques. Such algorithms could become costly if they had to be applied for several values of Λ . However in our numerical experiments, the threshold random variable Λ only takes a small number of distinct values, such as 10, which is much less than the number of simulations, which is of order 10,000. Furthermore for each value of $\Lambda = \lambda$, we sample from (11) thousands of times, which makes the overhead required to initialize the sampling algorithms negligible.

After sampling V from (11), we want to draw $E_{(1)}$ given V . Let the conditional density of $E_{(1)}$ be denoted by $f_{E_{(1)} | (E_{(1)} < \gamma\mathcal{V}, V)}(x | V)$. Then

$$f_{E_{(1)} | (E_{(1)} < \gamma\mathcal{V}, V)}(x | V) = \frac{d \exp(-dx) I(x < \gamma V)}{1 - \exp(-d\gamma V)}$$

and we can draw a sample from this density using the inversion technique. In particular, we generate $U \sim \mathcal{U}[0, 1)$ and then let $E_{(1)} = -\frac{1}{d} \log(1 - U(1 - e^{-\gamma d^V}))$.

Step 2: Sampling $(E_1, \dots, E_d) | E_{(1)}$

Suppose we have drawn $E_{(1)} = x_{(1)}$ from Step 1. Let $f(x_1, \dots, x_d) = \exp(-\sum_{i=1}^d x_i)$ be the joint density of (E_1, \dots, E_d) . Note that each E_j is as likely to be the minimum. Consider the case where E_1 is the minimum. The conditional distribution is

$$f(x_1, \dots, x_d | E_1 = E_{(1)}, E_{(1)} = x_{(1)}) = \frac{e^{-x_{(1)} - \sum_{j=2}^d x_j}}{(1/d) d e^{-d x_{(1)}}} = e^{-\sum_{j=2}^d (x_j - x_{(1)})} \cdot I_{\{E_1 = x_{(1)}\}}. \quad (12)$$

We can sample from (12) by letting $E_j = \text{Exp}(1) + x_{(1)}$ independently for $j \in \{2, \dots, d\}$.

Since any of the E_j 's can be the minimum, we pick the index for the minimum component randomly from 1 to d and sample the rest of the components accordingly. This sampling method works for MC, but may not work very well for QMC. When randomly choosing the index for the minimum component, we potentially destroy the structure of the LDS. So, if we are working with an LDS, the CDM based on Proposition 2 below is preferred.

Proposition 2. *Suppose E_1, \dots, E_d are iid $\text{Exp}(1)$. Then*

$$\begin{aligned} \mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = x) \\ = \begin{cases} 1 - \exp\{-(x_k - x)\}, & \text{if } x_j = x \text{ for some } j \in \{1, \dots, k-1\}, \\ \frac{1}{d-k+1} I_{\{x_k < x\}} + \frac{d-k}{d-k+1} (1 - \exp\{-(x_k - x)\}), & \text{otherwise.} \end{cases} \end{aligned} \quad (13)$$

To sample E_1, \dots, E_d , we let k take the successive values $k \in \{1, \dots, d\}$ in (13) and proceed by inversion.

4.1 Stratified Sampling Alternative to Importance Sampling

Recall from Algorithm 1 and the form of C_λ given in (8) that our IS scheme starts with sampling a threshold random variable $\Lambda = \lambda_k$ and then proceeds by sampling $\tilde{\mathbf{U}} \sim \mathbf{U} | (T = \max\{U_1, \dots, U_d\} > \lambda_k)$. Instead, we can construct a stratified sampling (SS) estimator based on the samples from $\mathbf{U} | (\lambda_{k+1} > T \geq \lambda_k)$. That is, we stratify the domain of \mathbf{U} along with T . Suppose Λ takes M distinct values as $0 = \lambda_1 < \dots < \lambda_M < 1$. Let $\lambda_{M+1} = 1$ for convenience. Then we can define M strata as

$$\Omega_k = \{\mathbf{u} \in [0, \lambda_{k+1})^d | \mathbf{u} \notin [0, \lambda_k)^d\}, \quad k = 1, \dots, M. \quad (14)$$

By construction, $\lambda_k \leq T < \lambda_{k+1}$ if and only if $\mathbf{u} \in \Omega_k$. We can then construct the SS estimator

$$\hat{\mu}_{\text{SS},n} = \sum_{k=1}^M \frac{p_k}{n_k} \sum_{i=1}^{n_k} \Psi(\tilde{\mathbf{U}}_i^{(k)}), \quad (15)$$

where p_k is the stratum probability, n_k is the number of samples allocated to the stratum Ω_k , and $\tilde{\mathbf{U}}_i^{(k)} \stackrel{\text{ind.}}{\sim} \mathbf{U} | \Omega_k$. For Archimedean copulas, $p_k = \psi(d\psi^{-1}(\lambda_{k+1})) - \psi(d\psi^{-1}(\lambda_k))$. In Section 5, we show that the SS estimator has a smaller variance than the IS estimator. It is easy to show that sampling from Ω_k is equivalent to sampling from $(E_1, \dots, E_d, V) | \psi^{-1}(\lambda_{k+1})V < E_{(1)} \leq \psi^{-1}(\lambda_k)V$. Let $\gamma_k = \psi^{-1}(\lambda_k)$ for all $k \in \{1, \dots, M+1\}$. Algorithm 3 summarizes the procedure to sample from each stratum.

Algorithm 3 Sampling $\mathbf{U}_{k,j}$ in SS algorithm for Archimedean copulas

Require: $0 < \gamma_{k+1} < \gamma_k < \infty$.

- 1: Draw $(E_{(1)}, V) | (\gamma_{k+1}V < E_{(1)} \leq \gamma_k V)$.
 - 2: Draw $(E_1, \dots, E_d) | E_{(1)}$.
 - 3: Let $U_j = \psi(E_j/V)$ for $j \in \{1, \dots, d\}$.
 - 4: **return** (U_1, \dots, U_d) .
-

In this algorithm, Step 2 is exactly the same as for the IS case (Algorithm 2). For Step 1, we use conditional sampling to draw samples from the joint conditional density of $(E_{(1)}, V) | (\gamma_{k+1}V < E_{(1)} \leq \gamma_k V)$. By using an argument similar to the one used for Step 1 of Algorithm 2, we can show that the marginal conditional density of V is

$$f_{V|(E_{(1)} < \gamma V)}(v) = \beta f_V(v) (\exp(-d\gamma_{k+1}v) - \exp(-d\gamma_k v)), \quad (16)$$

where $f_V(v)$ is the density of V and $\beta = 1/p_k = 1/\psi[d\psi^{-1}(\lambda_{k+1})] - \psi[d\psi^{-1}(\lambda_k)]$. Conditional on V drawn from (16), generate $U \sim \text{U}[0, 1)$ and then let $E_{(1)} = -\frac{1}{d} \log [e^{-\gamma_{k+1}dy} - U(e^{-\gamma_{k+1}dy} - e^{-\gamma_k dy})]$. Then $(E_{(1)}, V)$ follows the desired distribution.

Remark 1. We can follow Algorithm 3 to sample from the SS distribution under QMC, if the number of samples to be drawn is fixed. In some cases, however, we want to keep running simulations until some error criterion is met. Since SS requires to have a subset of points allocated to each stratum, combining it with QMC for n not fixed is challenging. This is because when the total sample size is increased by successive increments, it means possibly disjoint subsets of a QMC point set will be used in a given stratum, which is undesirable. Whether or not this allocation over successive increments can be done in a clever way that exploits the uniformity of low-discrepancy sequences is a question we leave for future research.

5 Variance Analysis and Calibration Method

In this section, we analyze the variance of the IS and SS estimators and then propose calibration methods for choosing the q_k 's designed to minimize the variance of the respective estimators. We also show that the SS scheme is more flexible to calibrate and gives an estimate with a smaller variance than the IS estimator.

We define the strata $\Omega_1, \dots, \Omega_M$ as in (14) and $C_k = C(\lambda_k \mathbf{1})$ for $k \in \{1, \dots, M\}$. The following proposition gives the variance of the IS estimator.

Proposition 3. *Let $\hat{\mu}_{\text{IS},n}$ be the IS estimator described in Algorithm 1 with C_λ given in (8). Then its variance is given by*

$$\text{Var}(\hat{\mu}_{\text{IS},n}) = \frac{1}{n} \left(\sum_{k=1}^M p_k \left(\sum_{l=1}^k \frac{q_l}{1-C_l} \right)^{-1} \mu_k^{(2)} - \mu^2 \right), \quad (17)$$

where $p_k = \mathbb{P}(\mathbf{U} \in \Omega_k)$, $q_k = \mathbb{P}(\Lambda = \lambda_k)$ and $\mu_k^{(2)} = \mathbb{E}(\Psi^2(\mathbf{U}) | \Omega_k)$.

For the optimal calibration, we want to choose the q_k 's so that (17) is minimized. The following proposition gives an analytical expression for the optimal calibration. For convenience, we define $\mu_0^{(2)} = 0$.

Proposition 4. *The set of q_k 's that minimize (17) under the condition $\mu_1^{(2)} \leq \dots \leq \mu_M^{(2)}$ with $\mu_0^{(2)} = 0$ for convenience, is*

$$q_k^{\text{opt}} = \frac{(1-C_k) \left(\sqrt{\mu_k^{(2)}} - \sqrt{\mu_{k-1}^{(2)}} \right)}{\sum_{k=1}^M (1-C_k) \left(\sqrt{\mu_k^{(2)}} - \sqrt{\mu_{k-1}^{(2)}} \right)}, \quad k \in \{1, \dots, M\}. \quad (18)$$

Remark 2. If the condition $\mu_1^{(2)} \leq \dots < \mu_M^{(2)}$ is not met, some of the q_k^{opt} 's given by (18) will be negative, which makes the IS scheme infeasible. Note that $q_k^{\text{opt}} < 0$ means that ever having the event $[\Lambda = \lambda_k]$ makes the overall variance greater than when $q_k^{\text{opt}} = 0$. We propose to then remove λ_k from the support of Λ if $q_k^{\text{opt}} < 0$. Accordingly, the strata Ω_k 's will change so the stratum second moments need to be recomputed for the optimal calibration.

Of course, we do not know the true values of the $\mu_k^{(2)}$'s in practice, so we have to replace them with estimates. As often done for Neyman allocation, we can first run a pilot study with a small number of simulations and estimate the $\mu_k^{(2)}$'s. The condition $\mu_1^{(2)} \leq \dots < \mu_M^{(2)}$ means that the outer strata must have greater stratum second moments than the inner strata. We refer to this condition as *increasing second moment (ISM) condition*. Whether this ISM condition is met depends on the problem at hand. In this paper, we specifically work with $\Psi(\mathbf{U})$ which is large when at least one component of \mathbf{U} is large. This assumption on Ψ and the ISM condition are not

incompatible, although there is no guarantee that the former implies the latter. If the ISM is satisfied, then we can substitute (18) into (17) and obtain

$$\text{Var}(\hat{\mu}_{\text{IS},n}^{\text{opt}}) = \frac{1}{n} \left(\left(\sum_{k=1}^M p_k \sqrt{\mu_k^{(2)}} \right)^2 - \mu^2 \right). \quad (19)$$

Using the Cauchy-Schwarz inequality, we can show that

$$\text{Var}_{\mathbb{Q}}(\hat{\mu}_{\text{IS},n}^{\text{opt}}) = \frac{1}{n} \left(\left(\sum_{k=1}^M p_k \sqrt{\mu_k^{(2)}} \right)^2 - \mu^2 \right) \leq \frac{1}{n} \left(\sum_{k=1}^M p_k \mu_k^{(2)} - \mu^2 \right) = \text{Var}(\hat{\mu}_{\text{MC},n}).$$

Equality holds only when $\mu_k^{(2)}$ is the same for all k . Except for this restrictive case, the IS estimator with the optimal choice of q_k 's always has a smaller variance than the plain MC counterpart. If the ISM is not met, there is no analytical form for the optimal q_k 's. We can still find the optimal values using widely available convex optimization solvers in this case. If we let $q_1 = 1$ and $q_k = 0$ for $k = 2, \dots, M$, the proposal distribution is the same as the original distribution. That is, IS becomes plain MC. Hence, if we choose the q_k 's appropriately, the IS estimator cannot do worse than plain MC. In this sense, the IS estimator is similar to an SS estimator.

Now that we have derived the variance expression and the optimal choice of q_k 's, we move on to the stratified sampling estimator (15). Using simple algebra, one can show that $\text{Var}(\hat{\mu}_{\text{SS},n}) = \sum_{k=1}^M p_k^2 \sigma_k^2 / n_k$, where $\sigma_k^2 = \text{Var}(\Psi(\mathbf{U}) | \Omega_k)$, $k \in \{1, \dots, M\}$ are the stratum variances. The optimal n_k 's are given by Neyman allocation

$$n_k = \frac{n p_k \sigma_k}{\sum_{k=1}^M p_k \sigma_k}. \quad (20)$$

Unlike the IS estimator, there is no restriction on this optimal allocation, i.e., we do not need σ_k to be increasing with k . In this sense, the SS estimator is more flexible.

Since the true strata variances are unknown, we have to replace them with estimates. Investigating the optimal calibration formula for IS (18) and SS (20), it appears that the estimation error of the strata moments (the $\mu_k^{(2)}$'s for IS and the σ_k^2 's for SS) has greater impact on the estimated calibration for IS than for SS. Since q_k for IS depends on $\sqrt{\mu_k^{(2)}} - \sqrt{\mu_{k-1}^{(2)}}$, the estimation error comes from both estimating $\mu_{k-1}^{(2)}$ and $\mu_k^{(2)}$. On the other hand, for SS, n_k depends on σ_k , so the estimation error comes from estimating σ_k^2 alone. Consequently, the approximation is likely to deviate more from the actual optimal calibration for IS than for SS.

Going back to IS and as discussed in [9], instead of choosing $\Lambda = \lambda_k$ with probability q_k it is more efficient to stratify Λ . That is, take $n_k = n q_k$ observations with $\Lambda = \lambda_k$. Let $\hat{\mu}_{\text{IS},n}^{\text{det}}$ denote such a stratified IS estimator. Generally $n q_k$ is not an integer and needs to be rounded. If each n_k is large enough, this rounding effect is negligible. The following proposition compares the variance of the three estimators.

Proposition 5. *Suppose we have an IS estimator with $\mathbb{P}(\Lambda = \lambda_k) = q_k, 1 \leq k \leq M$. If the $\mu_k = \mathbb{E}(\Psi(\mathbf{U})|\Omega_k)$ are not all equal and n is large enough, then there exists some strata sample allocation (n_1, \dots, n_M) for the SS estimator such that $\text{Var}(\hat{\mu}_{\text{SS},n}) \leq \text{Var}(\hat{\mu}_{\text{IS},n}^{\text{det}}) \leq \text{Var}(\hat{\mu}_{\text{IS},n})$.*

This result trivially holds when we use the optimal q_k 's (18) for stratified and unstratified IS and use the optimal allocation (20) for SS. Since the SS estimator is more flexible for calibration and it has a smaller variance than both IS estimators, the SS approach is preferred if the sampling efforts for (11) and (16) are not significantly different. Nonetheless, depending on the type of the underlying copula, sampling from the IS distribution could be much easier than sampling from the SS distribution.

6 Numerical Examples

In this section, we investigate the efficiency of the IS and SS estimators introduced in this paper. We consider the valuation of tail-related quantities of a portfolio consisting of stocks from companies in the financial industry listed on the S&P 100. The five stocks in the portfolio are AIG, Allstate Corp., American Express Inc., Bank of New York and Citigroup Inc. Their stock symbols are AIG, ALL, AXP, BK and C, respectively. We assume that the value of the portfolio is 100 and that all the portfolio weights are equal. The data are daily negative log-returns of these five companies from 2010-01-01 to 2016-04-01 (1571 data points). We fit GARCH(1,1)-models with t -innovations to each return series to filter out the volatility clustering effect using the R package “rugarch” [6]. The fitted standardized residuals do not exactly follow a t -distribution, so we fit a semi-parametric distribution to the residuals using the R package “spd” [7]. The fitted model uses a kernel density estimate for the centre of the distribution and fits a generalized Pareto distribution to the tails. The use of generalized Pareto distribution to model the GARCH filtered residuals to estimate tail-related risk measures in a univariate setting is studied by McNeil and Frey [16]. We let $S = \sum_{j=1}^d X_j$ denote the portfolio loss over a one day period with

$$X_j = 100\omega_j \left(1 - \sum_{j=1}^d \exp(a_j - b_j \tilde{F}_j^{-1}(U_j)) \right),$$

where d is the number of assets, ω_j 's are the portfolio weights, a_j 's are the means of the log-returns, b_j 's are the fitted conditional standard deviations from the GARCH(1,1) model, \tilde{F}_j 's are the fitted semi-parametric distributions from the R package “spd”, and (U_1, \dots, U_d) follows the fitted copula. We use the R package “distr” [18] to sample from (11) and (16).

Using the R package “copula” [10], we fit the Gumbel, Frank, Clayton and Joe copulas to the standardized residuals based on MLE. Among the four Archimedean copulas, the Gumbel copula with $\theta = 1.604$ gives the best fit in terms of log-likelihood, followed by a Frank copula with $\theta = 4.06$. Hence we proceed assuming that the model we consider is well approximated by a Gumbel or a Frank copula.

The three functionals we estimate are stop loss $\mathbb{E}(\max\{S - D, 0\})$ with $D = 3$ for Gumbel and $D = 2$ for Frank, $\text{VaR}_{0.99}$ and $\text{ES}_{0.99}$ of S . To define C_Λ , we use $\lambda_k = 1 - (\frac{1}{2})^{k-1}$ for $k = 1, \dots, M$, with $M = 10$. When constructing an IS estimator, we stratify Λ regardless of whether we use MC or QMC. When we calibrate the q_k 's for IS according to (18) and SS according to (20), we use ES as our objective function.

Objective function	d	Estimate	Gumbel						Frank					
			MC			QMC			MC			QMC		
			IS	SS	Plain	IS	SS	IS	SS	Plain	IS	SS		
$\mathbb{E}(\max\{S - D, 0\})$	5	0.012	67	168	33	1730	8085	0.011	6.4	11	14	85	161	
	20	0.010	49	40	51	1128	3488	0.0034	4.6	4.1	5.7	46	46	
$\text{VaR}_{0.99}(S)$	5	3.2	10	26	8.4	39	98	2.4	9.7	9.0	2.6	32	26	
	20	3.04	7.9	7.2	5.8	19	28	2.1	4.3	4.8	3.6	16	19	
$\text{ES}_{0.99}(S)$	5	4.2	89	175	29	6019	16989	2.8	17	21	7.1	250	373	
	20	4.03	49	39	48	1296	4205	2.3	4.6	3.8	4.0	38	36	
Run time	5		3.6	3.7	1.8	3.7	3.8		3.6	3.7	1.1	3.6	3.7	
	20		2.0	1.9	1.2	1.9	2.0		1.7	1.8	1.1	1.9	2.0	

Table 2 Estimates and variance reduction factors for the Gumbel and Frank copulas based on $n = 30000$, $d = 5$.

Table 2 shows the estimates, variance reduction factors and computational times for the three functionals for the five different estimators based on Gumbel and Frank copulas, respectively. We used 30 randomizations to estimate the variance of each estimator (MC and QMC). The estimates shown are based on SS estimators with QMC. Variance reduction factors are defined to be the ratios of the variance of the plain MC estimators over the variance of the estimators with the respective VRTs. The last row of Table 2 shows the increase in computation time compared to plain MC. We see that both IS and SS reduce the variance by large amounts and this is amplified when combined with QMC. Note that SS estimators generally give smaller variances than the IS estimators, as suggested by Proposition 5. For IS and SS estimators with and without QMC, we see that the largest variance reduction factors are for ES. This makes sense as we calibrate the q_k 's to minimize the variance of the ES estimator.

We also repeat the same experiment but with a portfolio of 20 stocks from large companies in the financial industry traded on NYSE (the full list is available from the authors); the results are displayed under $d = 20$ in Table 2. Figures 2 and 3 show the log-variance of the three different MC-based estimators for different n .

Appendix

Proof of Lemma 1. Since $\mathbb{E}(\Psi(\tilde{\mathbf{U}})_w(\tilde{\mathbf{U}})) = \mathbb{E}(\Psi(\mathbf{U}))$, consistency follows directly from the Strong Law of Large Numbers. Note that $\mathbb{E}(\Psi(\tilde{\mathbf{U}})^2_w(\tilde{\mathbf{U}})^2) = \mathbb{E}(\Psi(\mathbf{U})^2_w(\mathbf{U})) \leq \mathbb{E}(\Psi(\mathbf{U})^2)B < \infty$, where the first equality is justified by a

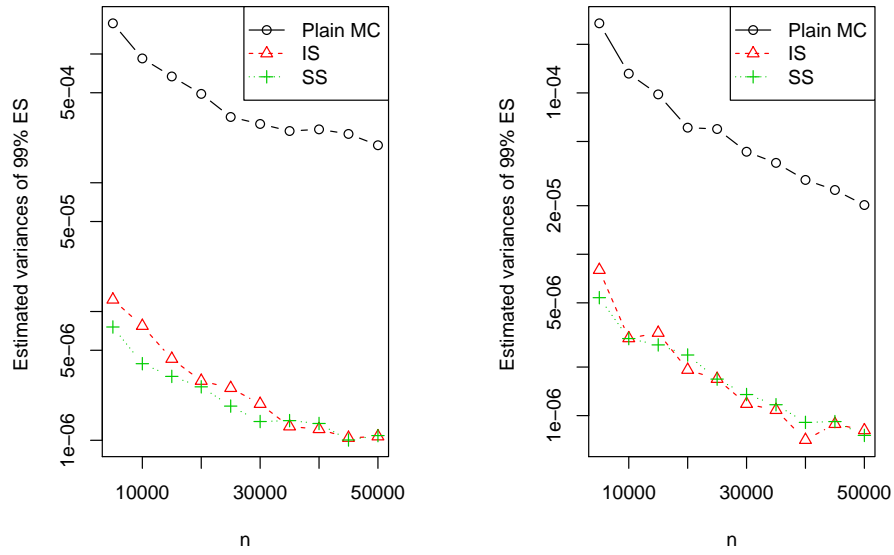


Fig. 2 Estimated variances of plain MC, IS and SIS estimators of $ES_{0.99}$ for a Gumbel copula (left-hand side) and a Frank copula (right-hand side) for different n and for $d = 5$.

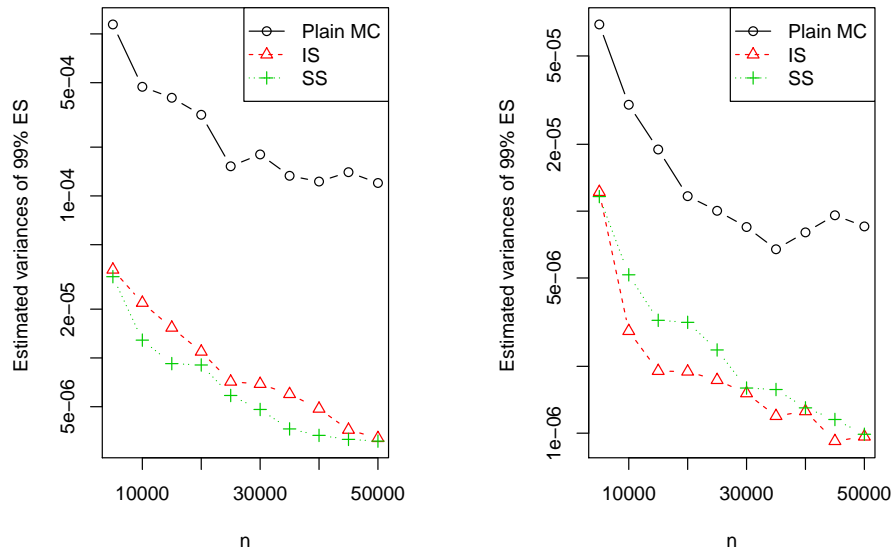


Fig. 3 Estimated variances of plain MC, IS and SIS estimators of 99% ES for a Gumbel copula (left-hand side) and a Frank copula (right-hand side) for different n and for $d = 5$

change of measure. We can immediately deduce asymptotic normality of $\hat{\mu}_{\text{IS},n}$ by the Central Limit Theorem, see, for example, Section 2.4 in [5], page 110. \square

Proof of Theorem 1. Due to Leibniz' integral rule, $dG(\mathbf{u}) = \int_0^1 dC_\lambda(\mathbf{u})dF_\Lambda(\lambda)$. From the definition of C_λ , we can derive the differential

$$dC_\lambda(\mathbf{u}) = \begin{cases} 0, & \mathbf{u} \in [0, \lambda]^d, \\ \frac{dC(\mathbf{u})}{1-C(\lambda\mathbf{1})}, & \text{otherwise.} \end{cases}$$

Using both identities, we obtain

$$dG(\mathbf{u}) = dC(\mathbf{u}) \int_0^1 \frac{I_{\{\lambda \leq \max\{u_1, \dots, u_d\}\}}}{1-C(\lambda\mathbf{1})} dF_\Lambda(\lambda),$$

leading to the desired result. \square

Proof of Lemma 2. Since $C(\lambda\mathbf{1}), \lambda \in [0, 1]$, the diagonal section of the copula C and the distribution function F_Λ are both increasing functions. The weight function \tilde{w} is thus decreasing on $[0, 1]$, and is bounded above by $\tilde{w}(0) = \mathbb{P}(\Lambda = 0)^{-1} < \infty$. \square

Proof of Proposition 1. We sample $(E_1, \dots, E_d, V) | (E_{(1)} < \gamma V)$ using conditional distribution sampling. That is, we first sample $(E_{(1)}, V) | (E_{(1)} < \gamma V)$, which is the Step 1 of Algorithm 2. Given the $(E_{(1)}, V)$ drawn, we then want to sample $(E_1, \dots, E_d) | (E_{(1)} < \gamma V, E_{(1)}, V)$ which is equivalent to sampling $(E_1, \dots, E_d) | E_{(1)}$ and this is the Step 2 of the algorithm. \square

Proof of Proposition 2. First, consider the case where $x_j = x$ for some $j = 1, \dots, k-1$. Without loss of generality assume that $x_1 = x$, i.e., $E_1 = E_{(1)}$. So we want to find $\mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = E_1 = x)$. From (12), the conditional distribution of E_k is $x + \text{Exp}(1)$. So the above probability equals

$$\mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = x) = 1 - e^{-(x_k - x)}. \quad (21)$$

Next, we consider the case $x_j \neq x$ for all $j = 1, \dots, k-1$. This means that $E_j = E_{(1)}$ for some $j = k, \dots, d$. Since all E_j are iid, there is a $\frac{1}{d-k+1}$ probability that $E_k = E_{(1)}$. In such a case $E_k = x$ with probability 1 as we are given $E_{(1)} = x$. Suppose $E_k \neq E_{(1)}$, which occurs with probability of $\frac{d-k}{d-k+1}$. Then we need to find the probability

$$\begin{aligned} & \mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = x, E_j \neq E_{(1)}, j = 1, \dots, k) \\ &= \sum_{j=k+1}^d \frac{1}{d-k} \mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = x, E_j = E_{(1)}) \\ &= \mathbb{P}(E_k \leq x_k | E_1 = x_1, \dots, E_{k-1} = x_{k-1}, E_{(1)} = x, E_d = E_{(1)}) = 1 - e^{-(x_k - x)}. \end{aligned}$$

The last equality again holds by (12) and the result follows. \square

Proof of Proposition 3. Recall that the IS estimator (7) is

$$\hat{\mu}_{\text{IS},n} = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{U}_i) w(\tilde{U}_i) = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{U}_i) \tilde{w}(t_i), \quad (22)$$

where $t_i = \max(\tilde{\mathbf{U}}_{i,1}, \dots, \tilde{\mathbf{U}}_{i,d})$, and where the weight function (9) is

$$\tilde{w}(u) = \left(\sum_{k=1}^M \frac{I_{\{\lambda_k \leq u\}}}{1 - C_k} q_k \right)^{-1}. \quad (23)$$

Hence \tilde{w} is constant over each stratum Ω_k . Thus, for $\mathbf{u} \in \Omega_k$, we can define the stratum weight as

$$w_k = \left(\sum_{l=1}^k \frac{q_l}{1 - C_l} \right)^{-1}, \quad k \in \{1, \dots, M\}. \quad (24)$$

The second moment of $w(\tilde{\mathbf{U}})\Psi(\tilde{\mathbf{U}})$ is

$$\begin{aligned} \mathbb{E}(w^2(\tilde{\mathbf{U}})\Psi^2(\tilde{\mathbf{U}})) &= \mathbb{E}(w(\mathbf{U})\Psi^2(\mathbf{U})) = \sum_{k=1}^M p_k \mathbb{E}(w(\mathbf{U})\Psi^2(\mathbf{U}) | \mathbf{U} \in \Omega_k) \\ &= \sum_{k=1}^M p_k w_k \mathbb{E}(\Psi^2(\mathbf{U}) | \mathbf{U} \in \Omega_k) = \sum_{k=1}^M p_k w_k \mu_k^{(2)} = \sum_{k=1}^M p_k \left(\sum_{l=1}^k \frac{1}{1 - C_l} q_l \right)^{-1} \mu_k^{(2)}. \end{aligned}$$

The third equality holds because the weight function $\tilde{w}(t)$ is constant over each stratum. The last equality follows from (24). Then the variance of the IS estimator based on n samples is $\text{Var}(\hat{\mu}_{\text{IS},n}) = \frac{1}{n} \left(\sum_{k=1}^M p_k \left(\sum_{l=1}^k \frac{1}{1 - C_l} q_l \right)^{-1} \mu_k^{(2)} - \mu^2 \right)$. \square

Proof of Proposition 4. Since the variance expression (17) is convex in q_k 's, we can solve the minimization problem using Lagrange multipliers. First, we simplify (17) so that the minimization problem becomes easier. Let $\tilde{p}_k = \mathbb{P}(\tilde{\mathbf{U}} \in \Omega_k)$, the stratum probability under the proposal distribution. Observe that

$$\begin{aligned} \tilde{p}_k &= \sum_{l=1}^M q_l \cdot \mathbb{P}(\mathbf{U} \in \Omega_k | \Lambda = \lambda_k) = \sum_{l=1}^k q_l \cdot \mathbb{P}(\mathbf{U} \in \Omega_k | \Lambda = \lambda_k) \\ &= \sum_{l=1}^k q_l \cdot \mathbb{P}(\mathbf{U} \in \Omega_k | \Lambda = \lambda_k) = \sum_{l=1}^k q_l \frac{p_k}{1 - C_l} = p_k \sum_{l=1}^k \frac{q_l}{1 - C_l}. \end{aligned} \quad (25)$$

By (23) and (25), we can write $w_k = p_k / \tilde{p}_k$. The weight w_k is the ratio of probabilities of a sample falling onto stratum Ω_k under the original distribution and the proposal distribution. Plugging this expression into (17), we have

$$\text{Var}(\hat{\mu}_{\text{IS},n}) = \frac{1}{n} \left(\sum_{k=1}^M \frac{p_k^2}{\tilde{p}_k} \mu_k^{(2)} - \mu^2 \right). \quad (26)$$

Using the Lagrange multiplier method, we can show that the optimal \tilde{p}_k is

$$\tilde{p}_k^{\text{opt}} = p_k \sqrt{\mu_k^{(2)}} / \sum_{k=1}^M p_k \sqrt{\mu_k^{(2)}}. \quad (27)$$

Note that this optimal choice of \tilde{p}_k 's resembles the Neyman allocation, the optimal allocation under stratified sampling.

Using the relation $q_k = (1 - C_k) \left(\frac{\tilde{p}_k}{p_k} - \frac{\tilde{p}_{k-1}}{p_{k-1}} \right)$, (with $\tilde{p}_0/p_0 = 0$) the optimal q_k is

$$q_k^{\text{opt}} \propto (1 - C_k) \left(\sqrt{\mu_k^{(2)}} - \sqrt{\mu_{k-1}^{(2)}} \right), \quad (\text{with } \mu_0^{(2)} = 0). \quad (28)$$

The assumption that $\mu_1^{(2)} \leq \dots \leq \mu_M^{(2)}$ ensures that $q_k^{\text{opt}} \geq 0$ for $k = 1, \dots, M$. \square

Proof of Proposition 5. We have $\hat{\mu}_{\text{IS},n}^{\text{det}} = \frac{1}{n} \sum_{k=1}^M \sum_{j=1}^{nq_k} \Psi(\tilde{\mathbf{U}}_{ki}) w(\tilde{\mathbf{U}}_{ki})$, $\tilde{\mathbf{U}}_{ki} \stackrel{iid}{\sim} \mathbf{U} | \Lambda = \lambda_k$. Thus $\text{Var}(\hat{\mu}_{\text{IS},n}^{\text{det}}) = \mathbb{E} [\text{Var}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}) | \Lambda)] / n + O(1/n^2)$ (term due to rounding nq_k). Since $\text{Var}(\hat{\mu}_{\text{IS},n}) = \frac{1}{n} \text{Var}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}))$, we have $\text{Var}(\hat{\mu}_{\text{IS},n}^{\text{det}}) \leq \text{Var}(\hat{\mu}_{\text{IS},n})$ as long as n is large enough for the $O(1/n^2)$ term due to rounding to be smaller than $\text{Var}(\mathbb{E}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}) | \Lambda)) / n > 0$. As shown before, $\tilde{p}_k = \mathbb{P}(\tilde{\mathbf{U}} \in \Omega_k) = p_k \sum_{l=1}^k \frac{q_l}{1 - C_l}$.

Consider an SS estimator with $n_k = n\tilde{p}_k$. Then $\text{Var}(\hat{\mu}_{\text{SS},n}) = \frac{1}{n} \sum_{k=1}^M \frac{p_k^2}{\tilde{p}_k} \sigma_k^2$. Also $\text{Var}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}) | \Lambda = \lambda_k) = \text{Var}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}) | T > \lambda_k) \geq \mathbb{E}[\text{Var}(\Psi(\tilde{\mathbf{U}}) w(\tilde{\mathbf{U}}) | T > \lambda_k, T \in \Omega_j)] = \sum_{j=k}^M \frac{p_j}{1 - C_k} w_j^2 \sigma_j^2$. Then, using (24) and $w_k = p_k / \tilde{p}_k$ we get

$$\begin{aligned} \text{Var}(\hat{\mu}_{\text{IS},n}^{\text{det}}) &\geq \frac{1}{n} \sum_{k=1}^M q_k \sum_{j=k}^M \frac{p_j}{1 - C_k} w_j^2 \sigma_j^2 = \frac{1}{n} \sum_{k=1}^M p_k w_k^2 \sigma_k^2 \sum_{j=1}^k \frac{q_j}{1 - C_j} \\ &= \frac{1}{n} \sum_{k=1}^M p_k w_k \sigma_k^2 = \frac{1}{n} \sum_{k=1}^M \frac{p_k^2}{\tilde{p}_k} \sigma_k^2 = \text{Var}(\hat{\mu}_{\text{SS},n}). \end{aligned}$$

\square

Acknowledgements We thank the anonymous referees for their helpful comments. The third and fourth author gratefully acknowledge the financial support of NSERC Canada through grant numbers #5010 and #238959, respectively.

References

1. Cambou, M., Hofert, M., Lemieux, C.: Quasi-random numbers for copula models. *Statistics and Computing* **27**(5), 1307–1329 (2017)
2. Chan, J., Kroese, D.: Efficient estimation of large portfolio loss probabilities in t-copula models. *European Journal of Operational Research* **205**(2), 361–367 (2010)
3. Choe, G., Jang, H.: Efficient algorithms for basket default swap pricing with multivariate Archimedean copulas. *Insurance: Mathematics and Economics* **48**(2), 205–213 (2011)
4. Derflinger, G., Hörmann, W., Leydold, J.: Random variate generation by numerical inversion when only the density is known. *ACM Trans. Model. Comput. Simul.* **20**(4), 1–25 (2010)
5. Durrett, R.: *Probability: theory and examples*, fourth edn. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2010)
6. Ghalanos, A.: rugarch: Univariate GARCH models. (2015). R package version 1.3-6.
7. Ghalanos, A.: spd: Semi-Parametric Distribution. (2015). R package version 2.0-1.

8. Glasserman, P., Li, J.: Importance sampling for portfolio credit risk. *Management science* **51**(11), 1643–1656 (2005)
9. Hesterberg, T.: Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37**(2), 185–194 (1995)
10. Hofert, M., Kojadinovic, I., Maechler, M., Yan, J.: *copula: Multivariate Dependence with Copulas* (2016). R package version 0.999-15
11. Hofert, M., Mächler, M., et al.: Nested archimedean copulas meet R: The nacopula package. *Journal of Statistical Software* **39**(9), 1–20 (2011)
12. Huang, P., Subramanian, D., Xu, J.: An importance sampling method for portfolio cvar estimation with gaussian copula models. In: *Proceedings of the 2010 Winter Simulation Conference (WSC)*, pp. 2790 – 2800 (2010)
13. Lemieux, C.: *Monte Carlo and Quasi-Monte-Carlo Sampling*. Springer (2009)
14. Marshall, A., Olkin, I.: Families of multivariate distributions. *Journal of the American Statistical Association* **83**(403), 834–841 (1988)
15. McNeil, A., Frey, R., Embrechts, P.: *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton (2005)
16. McNeil, A.J., Frey, R.: Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**(3), 271–300 (2000)
17. Nelsen, R.: *An Introduction to Copulas*, second edn. Springer, New York (2006)
18. Ruckdeschel, P., Kohl, M., Stabla, T., Camphausen, F.: S4 classes for distributions. *R News* **6**(2), 2–6 (2006)
19. Sak, H., Hörmann, W., Leydold, J.: Efficient risk simulations for linear asset portfolios in the t-copula model. *European Journal of Operational Research* **202**(3), 802–809 (2010)
20. Sobol, I.: On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.* **7**(4), 86–112 (1967)
21. Tasche, D.: Capital Allocation to Business Units and Sub-Portfolios: the Euler Principle. In: A. Resti (ed.) *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, pp. 423–453. Risk Books, London (2008)