Efficient simulation for dependent rare events

with applications to extremes

Lars Nørvang Andersen · **Patrick J. Laub** ·
Leonardo Rojas-Nandayapa

Abstract We consider the general problem of estimating probabilities which arise as a union of dependent events. We propose a flexible series of estimators for such probabilities, and describe variance reduction schemes applied to the proposed estimators. We derive efficiency results of the estimators in rare-event settings, in particular those associated with extremes. Finally, we examine the performance of our estimators in a numerical example.

1 Introduction

The estimators in this paper apply to quite general problems, so we will first introduce them in the framework of our main example, namely, as estimators relating to rare maxima of dependent random vectors. For a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with maximum $M = \max_i X_i$, the first problem we consider is estimating

$$\alpha(\gamma) = \mathbb{P}(M > \gamma).$$

This problem has many applications in many areas, for example in actuarial science (e.g. default probabilities [6]), finance (e.g. probability of 'knock-out' in a

Lars Nørvang Andersen
Aarhus University
E-mail: larsa@math.au.dk

Patrick J. Laub
University of Queensland and Aarhus University
E-mail: p.laub@[uq.edu.au|math.au.dk]

Leonardo Rojas-Nandayapa
University of Liverpool
E-mail: leorojas@liverpool.ac.uk

barrier option [14]), survival analysis, reliability [32] and engineering (e.g. failure probability of a series circuit).

We construct estimators for this probability, which are in terms of

$$E(\gamma) = \sum_{i=1}^{d} \mathbb{1}\{X_i > \gamma\},$$

the random variable which counts the number of X_i which exceed γ.[1] Our two main estimators in this setting are

$$\widehat{\alpha}_1 = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) + \frac{1}{R} \sum_{r=1}^{R} (1 - E_r(\gamma)) \mathbb{1}\{E_r(\gamma) \geq 2\}, \text{ and} \qquad (1)$$

$$\widehat{\alpha}_2 = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) - \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \mathbb{P}(X_i > \gamma, X_j > \gamma) \qquad (2)$$

$$+ \frac{1}{R} \sum_{r=1}^{R} \left[1 - E_r(\gamma) + \frac{E_r(\gamma)(E_r(\gamma) - 1)}{2} \right] \mathbb{1}\{E_r(\gamma) \geq 3\}.$$

where $R \in \mathbb{N}$ and the $E_r(\gamma)$s are derived from i.i.d. samples of $\boldsymbol{X}$. The fact that these unbiased estimators of $\alpha(\gamma)$ follows from Proposition 1 below. Estimation of $\mathbb{P}(M > \gamma)$ is a difficult problem and treatments in the literature of make distributional assumptions on $\boldsymbol{X}$. One such example is Adler et al. [2] where $\boldsymbol{X}$ is assumed to be multivariate normal. In this case, our estimator $\widehat{\alpha}_1$, with appropriate importance sampling, is the same as one of the estimators from [2].

The next problem we consider is estimating

$$\beta_n(\gamma) := \mathbb{E}[Y \mathbb{1}\{E_\gamma \geq n\}]$$

for $n = 1, \ldots, d$ and some random variable Y. We do not make any assumptions of independence between the $\{X_i > \gamma\}$ events themselves or between the events and Y.

The subcase of $Y = 1$ a.s. has some interesting examples:

$$\beta_1(\gamma) = \mathbb{P}(M > \gamma) = \alpha(\gamma), \quad \text{and} \quad \beta_n(\gamma) = \mathbb{P}(X_{(n)} > \gamma)$$

where $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(d)}$ are the order statistics of $\boldsymbol{X}$. The probability of a parallel circuit failing is a simple application for $\mathbb{P}(X_{(n)} > \gamma)$.

Our main β_1 estimator uses the fact that

$$\{M > \gamma\} := \bigcup_{i=1}^{d} \{X_i > \gamma\} = \bigcup_{i=1}^{d} \{X_1 \leq \gamma, \ldots, X_{i-1} \leq \gamma, X_i > \gamma\} \qquad (3)$$

where the events in the union on the right are disjoint. This supplies a form of β_1 which is amenable to efficient Monte Carlo estimation:

$$\beta_1 = \sum_{i=1}^{d} \mathbb{E}[Y \mathbb{1}\{X_1 \leq \gamma, \ldots, X_{i-1} \leq \gamma\} \mid X_i > \gamma] \mathbb{P}(X_i > \gamma). \qquad (4)$$

---

[1] We use $\mathbb{1}\{\cdot\}$ to denote the indicator function, and $\mathbb{1}\{\emptyset\} = 1$.

As previously mentioned, while they are main example and motivation, the extremes considered so far are a very specific instance of estimators. We now turn our attention to the general set-up treated in the paper.

Let $A(\gamma) = \cup_{i=1}^{d} A_i(\gamma)$ be the union of events $A_1(\gamma), \ldots, A_d(\gamma)$ for an index parameter $\gamma \in \mathbb{R}$. We consider the problem of estimating $\mathbb{P}(A(\gamma))$ when the events are rare, that is, $\mathbb{P}(A(\gamma)) \to 0$ as $\gamma \to \infty$. Define

$$\alpha(\gamma) := \mathbb{P}(A(\gamma)) \quad \text{and} \quad E(\gamma) := \sum_{i=1}^{d} \mathbb{1}\{A_i(\gamma)\}.$$

Note that we recover our introductory example by having $A_i(\gamma) = \{X_i > \gamma\}$. Aside from this example, $A(\gamma)$ is quite general (a union of arbitrary events) and many interesting events arising in applied probability and statistics can be formulated as a union. The quantity $\beta_n(\gamma)$ is reminiscent of *expected shortfall* from risk management [29].

Traditional Monte Carlo methods are unreliable in the rare-event setting. We will use standard techniques from the *rare-event simulation methodology*, such as importance sampling for variance reduction and applicable measures of efficiency: *bounded relative error* and *logarithmic efficiency*, cf. [7, 17, 36]. The resulting estimators are among the most efficient possible under the most general assumptions.

The paper is structured as follows. In Sections 2 and 3 we formally introduce our estimators for $\alpha(\gamma)$ and $\beta_n(\gamma)$ respectively, we prove their validity, and show how to combine them with some existing variance reduction techniques; the efficiency properties for the general estimators are analysed in Section 4, in addition we further investigate the efficiency for certain important dependence structures. Finally, we evaluate the numerical performance of the estimators in Section 5.

2 Estimators of α

In the following, we first explain the construction of our estimators of α, then discuss possible variance reduction schemes. As the γ notation can be cumbersome, we simply write $A = A(\gamma)$, $A_i = A_i(\gamma)$, $E = E_\gamma$, $\alpha = \alpha(\gamma)$ and $\beta_n = \beta_n(\gamma)$. Similarly, we write $\sum_i$, $\sum_{i<j}$, $\cup_i$, $\cap_i$ for $\sum_{i=1}^{d}, \sum_{1=i<j}^{d}, \cup_{i=1}^{d}$ and $\cap_{i=1}^{d}$.

2.1 Proposed estimators of α

The inclusion–exclusion formula (IEF) provides a representation of α as a summation whose terms are decreasing in size. The formula is

$$\alpha = \mathbb{P}(A) = \sum_{i=1}^{d} (-1)^{i+1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i \right). \tag{5}$$

The IEF can rarely be used as its summands are increasingly difficult to calculate numerically. The $\mathbb{P}(A_i)$ terms are typically known, and the $\mathbb{P}(A_i, A_j)$ terms can frequently be calculated, however the remaining higher-dimensional terms are

normally intractable for numerical integration algorithms (cf. the *curse of dimensionality* [7, Chapter IX]). Truncating the summation leads to bias, and indeed by the Bonferroni inequalities we have:

$$\alpha \le \sum_{i=1}^{k}(-1)^{i-1}\sum_{|I|=i}\mathbb{P}\left(\bigcap_{i\in I}A_i\right) \quad \text{if } 1 \le k < d \text{ and } k \text{ is odd,} \tag{6}$$

$$\alpha \ge \sum_{i=1}^{k}(-1)^{i-1}\sum_{|I|=i}\mathbb{P}\left(\bigcap_{i\in I}A_i\right) \quad \text{if } 1 < k < d \text{ and } k \text{ is even.} \tag{7}$$

This higher-order intractability motivates our estimators which use the IEF rewritten in terms of $E = \sum_i \mathbb{1}\{A_i\}$.

Proposition 1 *For $i = 1, \ldots, d$,*

$$\sum_{|I|=i}\mathbb{1}\{\cap_{i\in I}A_i\} = \binom{E}{i}\mathbb{1}\{E \ge i\}. \tag{8}$$

Proof

$$\sum_{|I|=i}\mathbb{1}\{\cap_{i\in I}A_i\} = \sum_{k=i}^{d}\sum_{|I|=i}\mathbb{1}\{\cap_{i\in I}A_i, E = k\} = \sum_{k=i}^{d}\binom{k}{i}\mathbb{1}\{E = k\} = \binom{E}{i}\mathbb{1}\{E \ge i\}.$$

$\square$

Taking the expectation of (8) gives

$$\sum_{|I|=i}\mathbb{P}\left(\bigcap_{i\in I}A_i\right) = \mathbb{E}\left[\binom{E}{i}\mathbb{1}\{E \ge i\}\right] \quad \text{for } i = 1, \ldots, d.$$

So the following has mean α, and forms the nucleus of our $\widehat{\alpha}_i$ estimators:

$$\sum_{i=1}^{d}(-1)^{i-1}\binom{E}{i}\mathbb{1}\{E \ge i\}. \tag{9}$$

We present estimators which deterministically *calculate* the first larger terms of the IEF (5) and Monte Carlo (MC) *estimate* the remaining smaller terms using sample means of (8). We begin by constructing the single-replicate estimator $\widehat{\alpha}_1$ where the first summand is calculated and the remaining terms are estimated:

$$\widehat{\alpha}_1 := \sum_i \mathbb{P}(A_i) + \sum_{i=2}^{d}\left[(-1)^{i-1}\binom{E}{i}\mathbb{1}\{E \ge i\}\right]$$

$$= \sum_i \mathbb{P}(A_i) + (1 - E)\mathbb{1}\{E \ge 2\}, \quad \text{using} \quad \sum_{k=0}^{n}(-1)^{k-1}\binom{n}{k} = 0.$$

In identical fashion, the single-replicate estimator calculating the first two terms from the IEF is

$$\widehat{\alpha}_2 := \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) + \sum_{i=3}^{d} \left[(-1)^{i-1} \binom{E}{i} \mathbb{1}\{E \geq i\} \right]$$

$$= \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) + \left[1 - E + \frac{E(E-1)}{2} \right] \mathbb{1}\{E \geq 3\}.$$

Thus, for $n \in \{1, \ldots, d-1\}$,[2]

$$\widehat{\alpha}_n := \sum_{i=1}^{n} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i \right) + \left[\sum_{i=0}^{n} (-1)^i \binom{E}{i} \right] \mathbb{1}\{E \geq n+1\}. \qquad (10)$$

Thus, $\{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{d-1}\}$ is a collection of estimators which allows the user to control the computational division of labour between numerical integration and Monte Carlo estimation. We will furthermore let $\widehat{\alpha}_0$ be the crude Monte Carlo estimator $\mathbb{1}\{E \geq 1\}$, and note that this falls under the definition in (10) if we interpret the empty sum as zero.

The $\widehat{\alpha}_n$ estimators are of decreasing variance in n, however each estimator carries the assumption that one can perform accurate numerical integration for 1 up to n dimensions. As numerical integration can be slow and unreliable in high dimensions we focus on $\widehat{\alpha}_1$, and also show the numerical performance of $\widehat{\alpha}_2$.

In practice, theses estimators will exhibit very modest improvements when compared against their truncated IEF counterparts (i.e., the right side of (6) and (7)). When combined with importance sampling, as in Section 2.4, the improvement is marked. Furthermore, we will show that these estimators possess desirable efficiency properties which are preserved after combining with importance sampling.

2.2 Discussion of $\widehat{\alpha}_1$ estimator

The estimator $\widehat{\alpha}_1$ has some nice interpretations. Recall the Boole–Fréchet inequalities

$$\max_i \mathbb{P}(A_i) \leq \alpha = \mathbb{P}(A) \leq \sum_i \mathbb{P}(A_i) =: \overline{\alpha}. \qquad (11)$$

The stochastic part of $\widehat{\alpha}_1$ is an unbiased estimate of $\overline{\alpha} - \alpha \leq 0$. That is to say, $\widehat{\alpha}_1$ MC estimates the difference between the target quantity α and its upper bound given by the Boole–Fréchet inequalities, $\overline{\alpha}$. Similarly, we often have

$$\alpha(\gamma) \sim \sum_i \mathbb{P}(A_i(\gamma)),\,^{3}$$

for example when the A_i exhibit a weak dependence structure. In this case, we can say that $\widehat{\alpha}_1$ MC estimates the difference between α and its (first-order) asymptotic expansion.

---

[2] Note that by the IEF, we have $\widehat{\alpha}_d := \alpha$, so this possibility is ignored.

[3] Using the standard notation that $f(x) \sim g(x)$ means $\lim_{x \to \infty} f(x)/g(x) = 1$.

2.3 Relation of $\widehat{\alpha}_n$ estimators to control variates

An alternative construction of $\{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{d-1}\}$ is to add *control variates* to the crude Monte Carlo estimator $\widehat{\alpha}_0$. We begin by adding the control variate E to $\widehat{\alpha}_0$ with weight $\tau \in \mathbb{R}$:

$$\widehat{\alpha}_1^\tau := \mathbb{1}\{E \geq 1\} - \tau \Big[E - \sum_i \mathbb{P}(A_i) \Big].$$

Setting $\tau = 1$ means this estimator simplifies to $\widehat{\alpha}_1$. Next, we add the control variates E and $-\frac{1}{2}E(E-1)$ to $\widehat{\alpha}_0$, and setting the corresponding weights to 1 gives $\widehat{\alpha}_2$. This pattern goes on.

2.4 Combining $\widehat{\alpha}_1$ with importance sampling

The family of estimators $\widehat{\alpha}_n$ can be combined with the variance reduction technique called *importance sampling* (IS), cf. [7,17]. Standard IS theory suggests that we should focus on IS distributions where the event of interest $A = \cup_i A_i = \{E \geq 1\}$ occurs almost surely. A convenient way of constructing such a distribution is as a *mixture distribution*. Say that we condition on A_i with probability

$$p_i := \frac{\mathbb{P}(A_i)}{\sum_j \mathbb{P}(A_j)} = \frac{\mathbb{P}(A_i)}{\overline{\alpha}}, \qquad \text{for } i = 1, \ldots, d.$$

A heuristic motivation for this selection comes from a rare-event setting where the asymptotic relationship $\mathbb{P}(A_i(\gamma), A_j(\gamma)) = \mathrm{o}(\mathbb{P}(A_i(\gamma)))$ often occurs for all $i \neq j$. In such a case

$$\mathbb{P}\left(A_i(\gamma) \mid A(\gamma)\right) = \frac{\mathbb{P}(A_i(\gamma))}{\sum_j \mathbb{P}(A_j(\gamma))(1 + \mathrm{o}(1))} \sim p_i(\gamma), \quad \text{as } \gamma \to \infty.$$

Now consider the measure

$$\mathbb{Q}^{[1]}(\mathscr{A}) = \sum_i p_i \, \mathbb{P}(\mathscr{A} \mid A_i) \qquad \forall \mathscr{A} \in \mathcal{F},$$

which induces the likelihood ratio of $L^{[1]} := \mathrm{d}\,\mathbb{Q}^{[1]} / \mathrm{d}\,\mathbb{P} = \overline{\alpha}/E$. As

$$\overline{\alpha} + (1 - E)\mathbb{1}\{E \geq 2\}L^{[1]} = \overline{\alpha}\Big(1 + \frac{1 - E}{E}\Big) = \frac{\overline{\alpha}}{E} \quad \text{under } \mathbb{Q}^{[1]},$$

we can see that $\widehat{\alpha}_1$ under this change of measure, with $R \in \mathbb{N}$ replicates, is

$$\widehat{\alpha}_1^{[1]} := \frac{1}{R} \sum_{r=1}^R \frac{\overline{\alpha}}{E_r^{[1]}}, \tag{12}$$

where the superscript "[1]" indicates that the $E_r^{[1]}$ are (independently) sampled under $\mathbb{Q}^{[1]}$. This estimator corresponds to one from the paper of Adler et al. [1], though applied in a more general way (they consider rare maxima of normally distributed vectors).

Continuing in the same pattern, consider the *second-order* IS distributions where $\{E \geq 2\}$ occurs almost surely, to be applied to $\widehat{\alpha}_2$. Say that we choose to condition on $A_i \cap A_j$ with probability

$$p_{ij} := \frac{\mathbb{P}(A_i, A_j)}{\sum_{m<n} \mathbb{P}(A_m, A_n)} = \frac{\mathbb{P}(A_i, A_j)}{q}, \qquad \text{for } 1 \leq i < j \leq d \,,$$

defining $q := \sum_{i<j} \mathbb{P}(A_i, A_j)$. Now consider the measure

$$\mathbb{Q}^{[2]}(\mathscr{A}) = \sum_{i<j} p_{ij} \, \mathbb{P}(\mathscr{A} \mid A_i, A_j) \qquad \forall \mathscr{A} \in \mathcal{F},$$

which induces a likelihood ratio of

$$L^{[2]} := \frac{\mathrm{d}\,\mathbb{Q}^{[2]}}{\mathrm{d}\,\mathbb{P}} = \frac{q}{\sum_{i<j} \mathbb{1}\{A_i A_j\}} = \frac{q}{\binom{E}{2}} = \frac{2q}{E(E-1)} \,.$$

Thus, after simplifying, the estimator $\widehat{\alpha}_2$ under $\mathbb{Q}^{[2]}$ is

$$\widehat{\alpha}_2^{[2]} := \overline{\alpha} - \frac{2q}{R} \sum_{r=1}^{R} \frac{1}{E_r^{[2]}}. \tag{13}$$

Remark 1 *As the $\mathbb{Q}^{[2]}$-mean of $\frac{2}{E}$ is less than 1, this fraction can be seen as a correction term for the two-term truncation of (5). We know from (7) that $\alpha \geq \overline{\alpha} - q$.*

$$\diamond$$

Both of these IS algorithms have some extra requirements for their use. The first-order estimators require that we can simulate from $\mathbb{P}(\cdot \mid A_i)$ and can calculate the $\mathbb{P}(A_i)$. The second-order estimator requires that we can simulate from $\mathbb{P}(\cdot \mid A_i, A_j)$ and that we can calculate the $\mathbb{P}(A_i)$ and $\mathbb{P}(A_i, A_j)$. In the rare maxima case, integration routines in MATHEMATICA or MATLAB can usually calculate these probabilities; it is simulating from the conditional distributions which can be the prohibitive requirement, particularly for $\widehat{\alpha}_2^{[2]}$.

3 Estimators of β_n

Now, we turn our attention to the estimation of

$$\beta_n := \mathbb{E}[Y \mathbb{1}\{E \geq n\}] \,.$$

We start with β_1, and rewrite the partition (3) in terms of the general A_i:

$$A := \bigcup_{i=1}^{d} A_i = A_1 \cup (A_1^{\mathsf{c}} A_2) \cup \cdots \cup (A_1^{\mathsf{c}} \ldots A_{d-1}^{\mathsf{c}} A_d) \,. \tag{14}$$

This gives us (the generalised version of (4))

$$\begin{aligned}
\beta_1 = \, &\mathbb{E}[Y \mid A_1] \, \mathbb{P}(A_1) + \mathbb{E}[Y \mathbb{1}\{A_1\} \mid A_2] \, \mathbb{P}(A_2) \\
&+ \cdots + \mathbb{E}[Y \mathbb{1}\{A_1^{\mathsf{c}} \ldots A_{d-1}^{\mathsf{c}}\} \mid A_d] \, \mathbb{P}(A_d) \,.
\end{aligned}$$

If we assume it is possible to sample from the $\mathbb{P}(\,\cdot\mid A_i)$ conditional distributions—the same assumption required to use the first-order IS estimator $\widehat{\alpha}_1^{[1]}$ from Section 2.4—then each of these conditional expectations can be estimated by sample means:

$$\widehat{\beta}_1 := \sum_{i=1}^{d} \frac{\mathbb{P}(A_i)}{\lceil R/d \rceil} \sum_{r=1}^{\lceil R/d \rceil} Y_{i,r} \mathbb{1}\{A_1^{\mathsf{c}} \ldots A_{i-1}^{\mathsf{c}}\}_{i,r} \tag{15}$$

Where $Y_{i,r}$ and $\mathbb{1}\{\cdot\}_{i,r}$, $r = 1, \ldots, R$ are sampled independently and conditional on A_i. The following proposition gives the partition of the event $\{E \geq i\}$:

Proposition 2 *Consider a finite collection of events $\{A_1, \ldots, A_d\}$ and for each subset $I \subset \{1, 2, \ldots, d\}$ define* [4]

$$B_I := \bigcap_{j \in I} A_j, \qquad C_I := \bigcap_{\substack{k \notin I, \\ k < \max I}} A_k^c.$$

Then

$$\{E \geq m\} = \bigcup_{|I|=m} B_I = \bigcup_{|I|=m} B_I C_I. \tag{16}$$

Moreover, the collection of sets $\{B_I C_I : |I| = m\}$ is disjoint.

Proof The first equality in (16) is straightforward from the definition of the random variable E. For the second equality we note that the relation $\supseteq$ follows trivially; to prove the opposite relation $\subseteq$ it remains to show that if ω is such that $\omega \in B_I$ and $\omega \notin C_I$, then there exists I' such that $|I'| = m$ and $\omega \in B_{I'} C_{I'}$. Notice that if $\omega \notin C_I$, then there exists a nonempty set J satisfying $\max J < \max I$, with $j \in J$ if and only if $\omega \notin A_j^c$. Select I' as the set formed by the smaller m elements of $I \cup J$. In consequence,

$$\omega \in \Big(\bigcap_{j \in I \cup J} A_j \Big) \Big(\bigcap_{\substack{k \notin I \cup J, \\ k \leq \max I}} A_k^c \Big) \subseteq \Big(\bigcap_{j \in I'} A_j \Big) \Big(\bigcap_{\substack{k \notin I', \\ k \leq \max I'}} A_k^c \Big) = B_{I'} C_{I'}.$$

This completes the proof of the second equivalence in (16).

Next we show that the collection of sets $\{B_I C_I : |I| = m\}$ is disjoint. Consider two sets of indexes I_1 and I_2 such that $|I_1| = |I_2| = m$ and $I_1 \neq I_2$. Take i such that $i \in I_1$, $i \notin I_2$ and w.l.o.g. further assume that $i < \max I_2$. Then $B_{I_1} \subseteq A_i$ while $C_{I_2} \subseteq A_i^c$. $\qquad\square$

This proposition implies that

$$\beta_n = \mathbb{E}\Big[Y \mathbb{1}\Big\{ \bigcup_{|I|=n} B_I \Big\} \Big] = \mathbb{E}\Big[Y \mathbb{1}\Big\{ \bigcup_{|I|=n} B_I C_I \Big\} \Big] = \sum_{|I|=n} \mathbb{E}\left[Y \mathbb{1}\{C_I\} \,\big|\, B_I \right] \mathbb{P}(B_I).$$

---

[4] Using the convention that $\cap_\emptyset = \Omega$.

Therefore, if (i) reliable estimates of $\mathbb{P}\left(B_I\right)$ are available, and (ii) it is possible to simulate from the conditional measures $\mathbb{P}\left(\cdot \mid B_I\right)$, then the following is an unbiased estimator of $\mathbb{E}[Y\mathbb{1}\{E \geq n\}]$:

$$\widehat{\beta}_n := \sum_{|I|=n} \frac{\mathbb{P}(B_I)}{\lceil R/\binom{d}{n}\rceil} \sum_{r=1}^{\lceil R/\binom{d}{n}\rceil} Y_{I,r}\mathbb{1}\{C_I\}_{I,r} \tag{17}$$

Where, similar to before, $Y_{I,r},\mathbb{1}\{\cdot\}_{I,r}$ indicates independent sampling conditioned on B_I.

Notice that a permutation of the sets $A_1,\ldots,A_d$ will result in a different collection of events C_I, and also a slightly different estimator.

3.1 Applying $\widehat{\beta}_i$ to estimate α

The $\widehat{\beta}_i$ estimators can be used in various ways to estimate the probability $\alpha = \mathbb{P}(A)$. The simplest way is to set $Y = 1$ a.s. in $\widehat{\beta}_1$ (17), leading to the estimator

$$\widehat{(\beta_1 \ddagger \alpha)} := \mathbb{P}(A_1) + \sum_{i=2}^{d} \frac{\mathbb{P}(A_i)}{\lceil R/(d-1)\rceil} \sum_{r=1}^{\lceil R/(d-1)\rceil} \mathbb{1}\{A_1^{\mathsf{c}}\ldots A_{i-1}^{\mathsf{c}}\}_{i,r}, \tag{18}$$

using the notation from (15). Note, we achieve minor improvement in (18) over (17) when $Y = 1$ a.s. as $\mathbb{E}[1 \mid A_1] = 1$ does not require estimation.

More effective estimators can be constructed if we use $\widehat{\beta}_n$ to estimate terms from $\widehat{\alpha}_{n-1}$ (10). We label the random terms in $\widehat{\alpha}_n$ as

$$R_n := \Big[\sum_{i=0}^{n}(-1)^i\binom{E}{i}\Big]\mathbb{1}\{E \geq n+1\}, \quad \text{and say} \quad \mathcal{R}_n := \mathbb{E}[R_n]. \tag{19}$$

Now, if we choose $Y := \sum_{i=0}^{n-1}(-1)^i\binom{E}{i}$ then it is obvious that

$$\beta_n := \mathbb{E}\Big\{\Big[\sum_{i=0}^{n-1}(-1)^i\binom{E}{i}\Big]\mathbb{1}\{E \geq n\}\Big\} = \mathcal{R}_{n-1}.$$

This leads to the set of estimators

$$\widehat{(\beta_n \ddagger \alpha)} := \sum_{i=1}^{n-1}(-1)^{i-1}\sum_{|I|=i}\mathbb{P}\Big(\bigcap_{i \in I}A_i\Big)$$

$$+ \sum_{|I|=n}\frac{\mathbb{P}(B_I)}{\lceil R/\binom{d}{n}\rceil}\sum_{r=1}^{\lceil R/\binom{d}{n}\rceil}\Big[\sum_{i=0}^{n-1}(-1)^i\binom{E_{I,r}}{i}\Big]\mathbb{1}\{E \geq n\}_{I,r},$$

for $n = 2,\ldots d-1$. In particular, for $n = 2$

$$\widehat{(\beta_2 \ddagger \alpha)} := \sum_{i}\mathbb{P}(A_i) + \sum_{i<j}\frac{\mathbb{P}(A_i,A_j)}{\lceil R/\binom{d}{2}\rceil}\sum_{r=1}^{\lceil R/\binom{d}{2}\rceil}(1 - E_{ij,r})\mathbb{1}\{E \geq 2\}_{ij,r}, \tag{20}$$

where the ij subscript indicates sampling conditional on A_iA_j, similar to before.

4 Efficiency results

In this section we analyse the performance of the estimators in a rare-event setting. Recall that in such a setting, $\{A_1(\gamma), \ldots, A_d(\gamma)\}$ denotes an indexed collection of not necessarily independent rare events and our objective is to calculate $\alpha(\gamma) = \mathbb{P}(\bigcup_i^d A_i(\gamma))$ as $\gamma \to \infty$. For such a *rare-event* estimation problem there are specialised concepts of efficiency. In Section 4.2 these definitions of efficiency are introduced. In addition, we provide efficiency criteria for the proposed estimators under very general assumptions.

In Sections 4.3 and 4.4 we specialise in rare events associated with extremes. In such a framework, we show when the estimator $\widehat{\alpha}_1$ is efficient for: i) a vast array of multivariate distributions with identical marginals in Section 4.3, and ii) the specific cases of normal and elliptical distributions in Section 4.4. For this section we take the number of replicates R to be 1.

4.1 Variance Reduction

First we compare the efficiency of our proposed estimator $\widehat{\alpha}_1$ against that of the crude Monte Carlo (CMC) estimator $\widehat{\alpha}_0(\gamma)$ of $\alpha(\gamma) := \mathbb{P}(A(\gamma))$. An upper bound for $\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_0(\gamma)$ is

$$\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_0(\gamma) = \mathbb{P}(A(\gamma))[1 - \mathbb{P}(A(\gamma)] < \mathbb{P}(A(\gamma)) \leq \sum_i \mathbb{P}(A_i(\gamma)).$$

This implies that the variance of the CMC estimator is of order $\mathcal{O}(\max_i \mathbb{P}(A_i(\gamma)))$, which is the best possible without making any further assumptions. In contrast an upper bound of $\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma) = \mathbb{V}\mathrm{ar}\,R_1$, where $R_1 = (1-E)\mathbb{1}\{E \geq 2\}$ from (19), is

$$\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma) \leq \mathbb{E}[R_1^2] < 2\,\mathbb{E}\left[\binom{E}{2}\mathbb{1}\{E \geq 2\}\right] \underset{(8)}{=} 2\sum_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma)). \tag{21}$$

Thus the variance of our estimator $\widehat{\alpha}_1(\gamma)$ is of order $\mathcal{O}(\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma)))$, so we can conclude that $\widehat{\alpha}_1(\gamma)$ is asymptotically superior to CMC.

Next we turn our attention to the estimator $\widehat{\beta}_n$. The following proposition shows that the reduction of variance of the estimator $\widehat{\beta}_n$ is of at least of a factor $\max_{|I|=n} \mathbb{P}(B_I)$ with respect to the non-conditional (crude) version estimator $\widehat{\beta}_n^{[0]}$ defined as

$$\widehat{\beta}_n^{[0]} := \sum_{|I|=n} \frac{1}{\lceil R/\binom{d}{n}\rceil} \sum_{r=1}^{\lceil R/\binom{d}{n}\rceil} Y_{Ir}\mathbb{1}\{B_I C_I\} \tag{22}$$

Proposition 3

$$\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) \leq \max_{|I|=n} \mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}).$$

Proof Let $W_I := Y\mathbb{1}\{C_I\}$. By independence of the W_I we can write the variance of $\widehat{\beta}_n$ as

$$\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) = \mathbb{V}\mathrm{ar}\left(\sum_{|I|=n} W_I\,\mathbb{P}(B_I)\,\Big|\,B_I\right) = \sum_{|I|=n} \mathbb{P}(B_I)^2\,\mathbb{V}\mathrm{ar}(W_I \mid B_I)$$

$$\leq \max_{|I|=n} \mathbb{P}(B_I) \sum_{|I|=n} \mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(W_I \mid B_I).$$

Now, observe that

$$\mathbb{P}(B_I) \, \mathbb{V}\mathrm{ar}(W_I \mid B_I) \leq \mathbb{P}(B_I) \, \mathbb{E}[W_I^2 \mid B_I] - \mathbb{P}(B_I)^2 \, \mathbb{E}[W_I \mid B_I]^2$$
$$= \mathbb{E}[W_I^2 \mathbb{1}\{B_I\}] - \mathbb{E}[W_I \mathbb{1}\{B_I\}]^2 = \mathbb{V}\mathrm{ar}\left[W_I \mathbb{1}\{B_I\}\right].$$

Thus we have proven that

$$\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) \leq \max_{|I|=n} \mathbb{P}(B_I) \sum_{|I|=n} \mathbb{V}\mathrm{ar}(W_I \mathbb{1}\{B_I\}) = \max_{|I|=n} \mathbb{P}(B_I) \sum_{|I|=n} \mathbb{V}\mathrm{ar}(\widehat{\beta}_0).$$

$\square$

4.2 Efficiency criteria

We now ask if and when $\widehat{\alpha}_1$ and $\widehat{\beta}_n$ are efficient in the rare-event sense. We must first define efficiency, as there are several common benchmarks for the efficiency of a rare-event estimator.

Definition 1 *An estimator $\widehat{p}_\gamma$ of some rare probability p_γ which satisfies $\forall \varepsilon > 0$*

$$\limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar} \, \widehat{p}_\gamma}{p_\gamma^{2-\varepsilon}} = 0 \qquad (23\mathrm{a})$$

$$\limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar} \, \widehat{p}_\gamma}{p_\gamma^2} < \infty \qquad (23\mathrm{b})$$

$$\limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar} \, \widehat{p}_\gamma}{p_\gamma^2} = 0 \qquad (23\mathrm{c})$$

has logarithmic efficiency *(LE)* (23a)*, bounded relative error *(BRE)* (23b)*, or vanishing relative error *(VRE)* (23c) *respectively.*

The levels of efficiency in Definition 1 are given in increasing order of strength, that is, VRE $\Rightarrow$ BRE $\Rightarrow$ LE. As VRE is often too difficult a goal, we focus on BRE and LE. The following proposition gives an alternative form of the conditions in (23) for the specific case of our estimator $\widehat{\alpha}_1$.

Proposition 4 *The estimator $\widehat{\alpha}_1$ has LE iff it holds that $\forall \varepsilon > 0$*

$$\limsup_{\gamma \to \infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^{2-\varepsilon}} = 0, \qquad (24)$$

and has BRE iff

$$\limsup_{\gamma \to \infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^2} < \infty. \qquad (25)$$

Proof We prove the LE claim (24). Proof of the BRE claim follows the same arguments.
($\Rightarrow$) We can see that

$$\mathbb{V}\mathrm{ar} \, \widehat{\alpha}_1(\gamma) \geq \mathbb{V}\mathrm{ar} \, \mathbb{1}\{E \geq 2\} = \mathbb{P}(E \geq 2) \, \mathbb{P}(E \leq 1), \quad \mathbb{P}(E \leq 1) \to 1, \qquad (26)$$

and

$$\mathbb{P}(E \geq 2) \geq \binom{d}{2}^{-1} \sum_{n=2}^{d} \binom{n}{2} \mathbb{P}(E = n) \underset{(8)}{=} \binom{d}{2}^{-1} \sum_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma)). \qquad (27)$$

So, $\forall \varepsilon > 0$,

$$
0 \underset{(23a)}{=} \limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma)}{\mathbb{P}(A)^{2-\varepsilon}} \underset{(11)\,\&\,(26)}{>} \limsup_{\gamma \to \infty} \frac{\mathbb{P}(E \geq 2)}{\left(\sum_k \mathbb{P}(A_k(\gamma))\right)^{2-\varepsilon}}
$$

$$
\underset{(27)}{\geq} \left[d^{2-\varepsilon}\binom{d}{2}\right]^{-1} \limsup_{\gamma \to \infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^{2-\varepsilon}}
$$

which implies (24).

($\Leftarrow$) We can see that, $\forall \varepsilon > 0$,

$$
\limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma)}{\mathbb{P}(A)^{2-\varepsilon}} \underset{(11)\,\&\,(21)}{<} \limsup_{\gamma \to \infty} \frac{2\sum_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{(\max_k \mathbb{P}(A_k(\gamma)))^{2-\varepsilon}}
$$

$$
\leq 2\binom{d}{2} \limsup_{\gamma \to \infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^{2-\varepsilon}} \underset{(24)}{=} 0,
$$

which implies (23a). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 1 *If the A_i events are independent then the estimator $\widehat{\alpha}_1$ has BRE.*

For the efficiency of our $\widehat{\beta}_n$ estimators, the following proposition provides a very simple yet non-trivial condition for BRE.

Proposition 5 *The estimator $\widehat{\beta}_n(\gamma)$ has BRE if*

$$
\limsup_{\gamma \to \infty} \frac{\max_{|I|=n} \mathbb{P}(B_I)}{\beta_n(\gamma)} < \infty.
$$

Proof By Proposition 3 and the hypothesis we have

$$
\limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar}(\widehat{\beta}_n(\gamma))}{\beta_n^2(\gamma)} \leq \limsup_{\gamma \to \infty} \frac{\max_{|I|=n} \mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma))}{\beta_n^2(\gamma)}
$$

$$
\leq c \limsup_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma))}{\beta_n(\gamma)}\,.
$$

Since $\widehat{\beta}_n^{[0]}$ is an estimator in crude form then $\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma)) = \mathcal{O}(\beta_n(\gamma))$ as $\gamma \to \infty$, so the proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 1 *The estimator $\widehat{(\beta_1 \ddagger \alpha)}$ from (18) has BRE.*

4.3 Efficiency for identical marginals and dependence

In this and the following subsections, we concentrate on rare events associated to extremes. More precisely, we let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an arbitrary random vector and define $M = \max_i X_i$. Therefore, we define $A_i(\gamma) = \{X_i > \gamma\}$ implying that the event of interest A is equivalent to $\{M > \gamma\}$.

In this subsection, we assume the X_i have identical marginal distributions. This simplifies the condition for BRE of $\widehat{\alpha}_1$, (25), so that it is now solely determined by the *copula* of $\boldsymbol{X}$. We investigate some common tail dependence measures of copulas (tail dependence parameter and residual tail index) and also some common families of copulas (Archimedean copulas) to see when the estimator $\widehat{\alpha}_1$ exhibits efficiency.

4.3.1 Asymptotic dependence

The most basic measurement of tail dependence between a pair (X_i, X_j) with common marginal distribution F and copula C_{ij} (cf. [24,30]) is

$$\lambda_{ij} = \lim_{v \to 1} \mathbb{P}(X_i > v \mid X_j > v) = \lim_{v \to 1} \frac{1 - C_{ij}(v, v)}{1 - v}$$

where $\lambda_{ij} \in [0, 1]$ is called the *(upper) tail dependence parameter (or coefficient)* [24, 29]. We say the (X_i, X_j) pair exhibit *asymptotic independence* (AI) when $\lambda_{ij} = 0$, or *asymptotic dependence* (AD) when $\lambda_{ij} > 0$. The canonical examples given for each case are the (non-degenerate) bivariate normal distribution for AI, and the bivariate Student t distribution for AD [38].

For $\widehat{\alpha}_1$ to have BRE, all pairs in $\boldsymbol{X}$ must exhibit AI. This is a necessary but not sufficient condition, therefore we will employ a more refined tail dependence measurement.

4.3.2 Residual tail index

We must first define two classes of functions:

- $L(x)$ is *slowly-varying* (at ∞) if $L(cx)/L(x) \to 1$ as $x \to \infty$ for all $c > 0$,
- $f(x)$ is *regularly-varying* (at ∞) with index $\tau > 0$ if it takes the form $f(x) = L(x)x^{-\tau}$ for some $L(x)$ which is slowly-varying (cf. [9,34]).

We will assume, w.l.o.g., the marginals of $\boldsymbol{X}$ to be unit Fréchet distributed (i.e., $F_1(x) = \exp(-x^{-1}) \sim 1 - x^{-1}$). Ledford and Tawn [26–28] first noted that the joint survivor functions for a wide array of bivariate distributions satisfy

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) \sim L(\gamma)\gamma^{-1/\eta} \qquad \text{as } \gamma \to \infty \qquad (28)$$

for a slowly-varying $L(\gamma)$ and an $\eta \in (0, 1]$. In other words, (28) says that $\mathbb{P}(X_i > \gamma, X_j > \gamma)$ is regularly-varying with index $1/\eta$.

The index is called the *residual tail index* [15,31].[5] When (X_i, X_j) exhibit AD (AI) then we typically have $\eta = 1$ $(\eta < 1)$.[6] For independent components we have $\eta = 1/2$, so Ledford and Tawn [26] describe bivariate distributions with $\eta = 1/2$ as having *near independence*. When $\eta < 1/2$ the random pair take large values together less frequently than they would if independent.

Returning to our original problem of estimating $\alpha(\gamma)$, let us label the residual tail index for every (X_i, X_j) pair of $\boldsymbol{X}$ as η_{ij}. Also, let $\eta = \max_{ij} \eta_{ij}$ and L be the associated slowly varying function. The following proposition outlines how these values relate to efficiency of $\widehat{\alpha}_1$:

Proposition 6 *If* (28) *is satisfied for the maximal pair of $\boldsymbol{X}$, that is,*

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) \sim L(\gamma)\gamma^{-1/\eta} \qquad \text{as } \gamma \to \infty \,,$$

then the estimator $\widehat{\alpha}_1$ has: i) BRE if $\eta < 1/2$ or if $\eta = 1/2$ and $L(\gamma) \not\to \infty$ as $\gamma \to \infty$, ii) LE if $\eta = 1/2$.

---

[5] The older (and less insightful) name for η is the *coefficient of tail dependence* [26,33].

[6] Hashorva [20] has found a case where an elliptically distributed (X_i, X_j) has $\eta = 1$ and AI.

Table 1: Residual tail dependence index η and $L(x)$ for various copulas. This is a subset of Table 1 of [21] (their row numbers are preserved).

(a) Copulas with BRE.

#	Name	η	$L(x)$
1	Ali-Mikhail-Haq	0.5	$1 + \tau$
2	BB10 in Joe	0.5	$1 + \theta/\tau$
3	Frank	0.5	$\delta/(1 - e^{-\delta})$
4	Morgenstern	0.5	$1 + \tau$
5	Plackett	0.5	δ
6	Crowder	0.5	$1 + (\theta - 1)/\tau$
7	BB2 in Joe	0.5	$\theta(\delta + 1) + 1$
8	Pareto	0.5	$1 + \delta$
9	Raftery	0.5	$\delta/(1 - \delta)$

(b) Copulas without BRE.

#	Name	η	$L(x)$
11	Joe	1	$2 - 2^{1/\delta}$
12	BB8 in Joe	1	$2 - 2(1 - \delta)^{\theta - 1}$
13	BB6 in Joe	1	$2 - 2^{1/(\delta\theta)}$
14	Extreme value	1	$2 - V(1, 1)$
15	B11 in Joe	1	δ
16	BB1 in Joe	1	$2 - 2^{1/\delta}$
17	BB3 in Joe	1	$2 - 2^{1/\theta}$
18	BB4 in Joe	1	$2^{-1/\delta}$
19	BB7 in Joe	1	$2 - 2^{1/\theta}$

Proof Label the components of $\boldsymbol{X}$ such that

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) = \mathbb{P}(X_1 > \gamma, X_2 > \gamma)$$

then the condition for LE becomes, $\forall \varepsilon > 0$

$$\limsup_{\gamma \to \infty} \frac{\max_{i<j} \mathbb{P}(X_i \geq \gamma, X_j \geq \gamma)}{\max_k \mathbb{P}(X_k \geq \gamma)^{2-\varepsilon}} = \limsup_{\gamma \to \infty} \frac{L(\gamma)\gamma^{-1/\eta}}{(\gamma^{-1})^{2-\varepsilon}} = \limsup_{\gamma \to \infty} L(\gamma)\gamma^{2 - \frac{1}{\eta} - \varepsilon} = 0$$

which is equivalent to $\eta \in (0, 1/2]$; the $\eta = 1/2$ case has LE as $\gamma^{-\varepsilon} L(\gamma) \to 0$ for all $\varepsilon > 0$ (see Proposition 1.3.6 part (v) of [9]). Similarly we have BRE for $\eta \in (0, 1/2)$, but for the $\eta = 1/2$ case we also require that $L(\gamma) \not\to \infty$. □

Heffernan [21] has conveniently compiled a directory of η and $L(x)$ for many copulas which satisfy (28). A summary of these results is given in Table 1. In reading Heffernan's directory, one can spot two trends: normally $\eta \in \{1/2, 1\}$ and L is a constant. The oft-cited Gaussian copula is the only exception for both of these trends in Heffernan's directory, having $\eta = (1 + \rho)/2$ and $L(x) \propto (\log x)^{-\rho/(1+\rho)}$; Section 4.4 deals with the Gaussian case in detail.

4.3.3 Archimedean Copulas

Some of the most frequently used copulas are in the family of *Archimedean copulas*. These are very general models and are widely used in applications due to their flexibility. A copula is Archimedean if there exists a function ψ such that the copula C can be written as

$$C(u_1, \ldots, u_n) = \psi^{\leftarrow}(\psi(u_1) + \cdots + \psi(u_n)).$$

The function ψ, called the *generator* of the copula, defines a copula if its functional inverse is the Laplace transform of a non-negative random variable. For Archimedean copulas we can restate the BRE condition (25) in terms of the generator ψ.

Theorem 1 (Thm. 3.4 of [13]) *Let $(U_1, \ldots, U_n) \sim C$ where C is an Archimedean copula with generator ψ. If $\psi^{\leftarrow}$ is twice continuously differentiable and its second derivative is bounded at 0 then $\forall\, i \neq j$*

$$\lim_{u \to 0} \frac{\mathbb{P}(U_i \geq 1 - ux_1, U_j \geq 1 - ux_2)}{u^2} < \infty$$

for any $0 < x_1, x_2 < \infty$.

Corollary 2 *Consider using $\widehat{\alpha}_1$ for a distribution with common marginal distributions and a copula C. If C satisfies the conditions of Theorem 1 then $\widehat{\alpha}_1$ has BRE.*

Charpentier and Segers [13] have helpfully created a directory of Archimedean copulas from which we can see if the BRE conditions from Corollary 2 are satisfied. Using this information, we provide a summary of the efficiency status of many Archimedean copulas in Table 2.

The efficiency of $\widehat{\alpha}_1$ can be proved without the assumption of identical marginal distributions, but the efficiency must be shown case-by-case for each family of distributions. The next section does this for the multivariate normal distribution and for some elliptical distributions.

4.4 Efficiency for the case of normal and elliptical distributions

The efficiency characteristics of normally and elliptically distributed random vectors are very similar. This section defines these distributions, outlines their asymptotic properties, then shows the conditions in which $\widehat{\alpha}_1$ exhibits levels of asymptotic efficiency.

4.4.1 Definitions and categories of elliptical distributions

Let $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and positive-definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. Denote the corresponding density $\varphi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\cdot, \cdot)$, and write $\sigma_i^2 := \boldsymbol{\Sigma}_{ii}$, $\rho_{ij} := \boldsymbol{\Sigma}_{ij}/(\sigma_i \sigma_j)$. The normal distribution belong to the class of *elliptical distributions*, which we denote $\mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$, where F is the c.d.f. of a positive r.v. We define $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ as

$$\boldsymbol{X} \overset{\mathcal{D}}{=} \boldsymbol{\mu} + R\boldsymbol{C}U \tag{29}$$

where $R \sim F$ is called the *radial component*, U is (independent of R and) distributed uniformly on the d-dimensional unit hypersphere, and $\boldsymbol{C} \in \mathbb{R}^{d \times d}$ satisfies $\boldsymbol{C}\boldsymbol{C}^{\top} = \boldsymbol{\Sigma}$. For background on elliptical distributions, see [5]. The efficiency of $\widehat{\alpha}_1$ turns out to be related with max-domain of attraction (MDA) of the radial component. The MDA is known from standard extreme value theory, see [15].

We consider some subclasses of elliptical distributions depending on the MDA of the radial distribution:

- $F \in \mathrm{MDA}(\text{Fréchet})$, then Theorem 4.3 of [23] implies that $\boldsymbol{X}$ has asymptotic dependence and $\widehat{\alpha}_1$ is never efficient (see Section 4.3.1).
- $F \in \mathrm{MDA}(\text{Weibull})$, then components of $\boldsymbol{X}$ are light-tailed and uninteresting (in a rare-event context).
- $F \in \mathrm{MDA}(\text{Gumbel})$, this is the interesting case which includes the normal distribution. Hashorva [19] label these the *type I elliptical random vectors*.

Table 2: Examples of Archimedean copula families. Names (if they are named) and generator functions are listed, as are the ranges for which θ is valid and the subset of θ which ensures that $\widehat{\alpha}_1$ has BRE. A Θ in the final column means that all valid θ ensure BRE. The families listed appear in Table 4.1 of [30] and Table 1 of [13].

#	Name	Generator $\psi(t)$	Valid θ	Efficient θ
1	Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,\infty)$	Θ
2		$(1-t)^\theta$	$[1,\infty)$	$\{1\}$
3	Ali–Mikhail–Haq	$\log\frac{1-\theta(1-t)}{t}$	$[-1,1)$	Θ
4	Gumbel–Hougaard	$(-\log t)^\theta$	$[1,\infty)$	$\{1\}$
5	Frank	$-\log\frac{\mathrm{e}^{-\theta t}-1}{\mathrm{e}^{-\theta}-1}$	$\mathbb{R}$	$\Theta\setminus\{0\}$
6		$-\log[1-(1-t)^\theta]$	$[1,\infty)$	$\{1\}$
7		$-\log[\theta t+(1-\theta)]$	$(0,1]$	Θ
8		$\frac{1-t}{1+(\theta-1)t}$	$[1,\infty)$	Θ
9		$\log(1-\theta\log t)$	$(0,1]$	Θ
10		$\log(2t^{-\theta}-1)$	$(0,1]$	Θ
11		$\log(2-t^\theta)$	$(0,1/2]$	Θ
12		$(\frac{1}{t}-1)^\theta$	$[1,\infty)$	$\{1\}$
13		$(1-\log t)^\theta-1$	$(0,\infty)$	Θ
14		$(t^{-1/\theta}-1)^\theta$	$[1,\infty)$	$\{1\}$
15		$(1-t^{1/\theta})^\theta$	$[1,\infty)$	$\{1\}$
16		$(\frac{\theta}{t}+1)(1-t)$	$[0,\infty)$	Θ
17		$-\log\frac{(1+t)^{-\theta}-1}{2^{-\theta}-1}$	$\mathbb{R}$	$\Theta\setminus\{0\}$
18		$\mathrm{e}^{\theta/(t-1)}$	$[2,\infty)$	$\emptyset$
19		$\mathrm{e}^{\theta/t}-\mathrm{e}^\theta$	$(0,\infty)$	Θ
20		$\mathrm{e}^{t^{-\theta}}-\mathrm{e}$	$(0,\infty)$	Θ
21		$1-[1-(1-t)^\theta]^{1/\theta}$	$[1,\infty)$	$\{1\}$
22		$\arcsin(1-t^\theta)$	$(0,1]$	Θ

4.4.2 Efficiency for type I elliptical distributions

Take $\boldsymbol{X}\sim\mathcal{ELL}(\boldsymbol{\mu},\boldsymbol{\Sigma},F)$ where the radial distribution $F\in\mathrm{MDA}(\mathrm{Gumbel})$ has support $(0,x_F)$, for some $x_F\in\overline{\mathbb{R}}$, and where $\{\sigma_1,\dots,\sigma_d\}$ are in decreasing order. By definition of the Gumbel MDA, one can find a scaling function $w(x)$ satisfying

$$\lim_{x\to x_F}\frac{\overline{F}(x+t/w(x))}{\overline{F}(x)}=\mathrm{e}^{-t}.$$

One frequently takes $w(x):=\overline{F}(x)/\int_x^{x_F}\overline{F}(s)\,\mathrm{d}s$. Also, define $v_i(\gamma):=(\gamma-\mu_i)/\sigma_i$ and $a_{ij}:=\sigma_j/\sigma_i$. If $\rho_{ij}\geq a_{ij}$ then set

$$\mu_{ij}:=\mu_j\quad\text{and}\quad\kappa_{ij}:=\sigma_j$$

otherwise for $\rho_{ij} < a_{ij}$

$$\mu_{ij} := \frac{\mu_i - a_{ij}\rho_{ij}(\mu_1 + \mu_2) + a^2\mu_j}{\alpha_{ij}(1 - \rho_{ij}^2)} \quad \text{and} \quad \kappa_{ij} := \frac{\sigma_i^2\sigma_j^2(1 - \rho_{ij}^2)}{\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j + \sigma_j^2} \, .$$

We now apply the asymptotic properties outlined in the Appendix to assess the efficiency of $\widehat{\alpha}_1$ for type I elliptical distributions.

Theorem 2 *Consider* $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ *where* $F \in MDA(Gumbel)$, *and let*

$$\kappa := \max_{i<j} \kappa_{ij}, \quad \mu := \max_{i<j \, : \, \kappa=\kappa_{ij}} \mu_{ij}, \quad and \quad \upsilon(\gamma) := (\gamma - \mu)/\kappa + \mathrm{o}(1) \, .$$

If $\kappa > \sigma_1$,[7] *then* $\widehat{\alpha}_1$ *has LE if*

$$\forall \varepsilon > 0 \quad \limsup_{\gamma \to x_F} \frac{w(\upsilon(\gamma))\overline{F}(\upsilon(\gamma))}{w(\upsilon_1(\gamma))\overline{F}(\upsilon_1(\gamma))^{2-\varepsilon}} < \infty \, . \tag{30}$$

Moreover, if (30) *holds for* $\varepsilon = 0$ *then* $\widehat{\alpha}_1$ *has BRE.*

Proof It follows from (33) and Theorem 3 in the Appendix. □

Example 2 (Kotz Type III) *One family of type I elliptical distributions, is the* Kotz Type III *distributions, defined by*

$$\overline{F}(\gamma) = (K + \mathrm{o}(1))\gamma^N \exp(-ru^\delta), \qquad w(\gamma) = r\delta\gamma^{\delta-1}, \gamma > 0,$$

with $K, \delta, N > 0$. *In this case it is clear that*

$$\lim_{\gamma \to \infty} \frac{w(\upsilon(\gamma))}{w(\upsilon_1(\gamma))} = \left(\frac{\sigma_1}{\kappa}\right)^{\delta-1} < \infty,$$

while

$$\limsup_{\gamma \to \infty} \frac{\overline{F}(\upsilon(\gamma))}{\overline{F}(\upsilon_1(\gamma))^2}$$

$$= \limsup_{\gamma \to \infty} \left(\frac{\sigma_1^2}{\kappa\gamma}\right)^N \exp\left\{-r\left(\left(\frac{\gamma-\mu}{\kappa}\right)^\delta - 2\left(\frac{\gamma-\mu_1}{\sigma_1}\right)^\delta\right)\right\},$$

$$= \limsup_{\gamma \to \infty} \left(\frac{\sigma_1^2}{\kappa\gamma}\right)^N \exp\left\{-r\left(\frac{\gamma^\delta - \delta\mu\gamma^{\delta-1} + \mathrm{o}(\gamma^{\delta-1})}{\kappa^\delta} - \frac{\gamma^\delta - \delta\mu_1\gamma^{\delta-1} + \mathrm{o}(\gamma^{\delta-1})}{\sigma_1^\delta/2}\right)\right\} \, .$$

Hence, $\widehat{\alpha}_1$ *has BRE in the following cases*

- $\sigma_1^\delta > 2\kappa^\delta$, *or*
- $\sigma_1^\delta = 2\kappa^\delta$, $\delta > 1$ *and* $\mu_1 > \mu$.

The estimator $\widehat{\alpha}_1$ *has LE if* $\sigma_1^\delta = 2\kappa^\delta$, *and inefficient when* $\sigma_1^\delta < 2\kappa^\delta$.

Example 3 (Normal distributions) *The normal distribution is a Kotz III type distribution with* $\delta = 2$. *Hence,* $\widehat{\alpha}_1$ *has BRE if* $\sigma_1^2 > 2\kappa^2$, *or* $\sigma_1^2 = 2\kappa^2$ *and* $\mu_1 > \mu$. *The estimator* $\widehat{\alpha}_1$ *has LE if* $\sigma_1^2 = 2\kappa^2$, *and is inefficient when* $\sigma_1^2 < 2\kappa^2$.

---

[7] This implies that the Savage condition (see Appendix) is fulfilled at least for one pair.

Frequently, a set of random variables represents as a stochastic process $\{X_n\}_{n\geq 0}$. The value of $\mathbb{P}(M > \gamma)$, with $M := \max_{1\leq n\leq d} X_n$, in such cases usually valuable. The simplest case to take is when all X_n have identical marginals such as in stationary processes; one such example is the autoregressive (AR) process.

Example 4 (AR(1) processes) *Say $X_t = \varphi X_{t-1} + \varepsilon_t$, where $|\varphi| < 1$ and ε_t are i.i.d. $\mathcal{N}_1(0,\sigma_\varepsilon^2)$, and we start the process in stationarity. We have that each X_i has the same marginal distribution, $X_i \sim \mathcal{N}_1(0, \sigma_\varepsilon^2/(1-\varphi^2))$, and*

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) = \begin{cases} \mathbb{P}(X_\bullet > \gamma, X_{\bullet+1} > \gamma) & \text{if } \varphi > 0 \\ \mathbb{P}(X_\bullet > \gamma, X_{\bullet+2} > \gamma) & \text{if } \varphi < 0 \\ \mathbb{P}(X_\bullet > \gamma)^2 & \text{if } \varphi = 0 \end{cases} \cdot$$

For $\varphi \neq 0$ we know that

$$(X_{\bullet+1} \mid X_\bullet = \gamma) \sim \mathcal{N}_1(\varphi\gamma, \sigma_\varepsilon^2), \text{ and } (X_{\bullet+2} \mid X_\bullet = \gamma) \sim \mathcal{N}_1(\varphi^2\gamma, \sigma_\varepsilon^2(1-\varphi^4)/(1-\varphi^2)).$$

When $\varphi = 0$ the X_i are independent and $\widehat{\alpha}_1$ is trivially efficient, and when $\varphi \in (-1,1) \setminus \{0\}$ we have (noting that $\{X_\bullet > \gamma\} \to \{X_\bullet = \gamma\}$) that

$$\lim_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)}{\max_i \mathbb{P}(X_i > \gamma)^2} = \lim_{\gamma\to\infty} \frac{\mathbb{P}(X_\bullet > \gamma, X_{\bullet+(1\,\text{or}\,2)} > \gamma)}{\mathbb{P}(X_\bullet > \gamma)^2}$$

$$= \lim_{\gamma\to\infty} \frac{\mathbb{P}(X_{\bullet+(1\,\text{or}\,2)} > \gamma \mid X_\bullet = \gamma)}{\mathbb{P}(X_\bullet > \gamma)}$$

$$= 0$$

as $\sigma_\varepsilon^2 < \sigma_\varepsilon^2(1-\varphi^4)/(1-\varphi^2)) < \sigma_\varepsilon^2/(1-\varphi^2)$. Therefore, we have BRE of $\widehat{\alpha}_1$ for all stationary AR(1) processes.

5 Numerical experiments

We explore the performance of the estimators for the problem of $\mathbb{P}(M > \gamma)$ for $M = \max_i X_i$, where $\boldsymbol{X}$ is multivariate normal and multivariate Laplace distributed. The following notation is used: $\boldsymbol{X}_{-i}$ ($\boldsymbol{X}_{-i,-j}$) is the random vector $\boldsymbol{X}$ with X_i (X_i and X_j) removed, $\boldsymbol{0}$ is the vector of zeros, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{x}^\top$ is the transpose of $\boldsymbol{x}$, and $X \perp\!\!\!\perp Y$ means X and Y are independent. We use some standard distributions: $\mathcal{E}(\lambda)$ for exponential ($f(x) \propto e^{-\lambda x}$), $\mathcal{IG}(\mu, \lambda)$ for inverse Gaussian ($f(x) \propto x^{-3/2}e^{-\lambda(x-\mu)^2/(2\mu^2 x)}$), $\mathcal{L}$ for Laplace (defined in Case 2 below). The Matlab and Mathematica code used to generate them are available online [3].

Case 1: Multivariate Normal distributions

Let $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I} + \rho$; that is, each $X_i \sim \mathcal{N}_1(0,1)$ and $\mathbb{C}\text{orr}(X_i, X_j) = \rho$. We implement the first- and second-order IS regimes. The necessary conditional distributions are well-known and simple; both $\boldsymbol{X}_{-i} \mid X_i$ and $\boldsymbol{X}_{-i,-j} \mid (X_i, X_j)$ are normally distributed [4]. Sampling from $X_i \mid X_i > \gamma$ can be

easily done by acceptance–rejection with shifted exponential proposals [35] (or by inverse transform sampling [7, Remark 2.4], though this can be problematic using only double precision arithmetic). To simulate $(X_i, X_j) \mid \min\{X_i, X_j\} > \gamma$ we use Botev's MATLAB library [10], but also remark that a Gibb's sampler is a commonly used alternative [11,35].

Case 2: Multivariate Laplace distributions

Let $\boldsymbol{X} \sim \mathcal{L}$. We can define this distribution by

$$\boldsymbol{X} \overset{\mathcal{D}}{=} \sqrt{R}\boldsymbol{Y}\,, \quad \text{where } \boldsymbol{Y} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{I}), R \sim \mathcal{E}(1), \boldsymbol{Y} \perp\!\!\!\perp R\,.$$

The distribution has been applied in a financial context [22], and is examined in [16,25]. From the former we have that the density of $\mathcal{L}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = 2(2\pi)^{-d/2} K_{(d/2)-1}\left(\sqrt{2\boldsymbol{x}^\top \boldsymbol{x}}\right)\left(\sqrt{\tfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{x}}\right)^{1-(d/2)}$$

where $K_n(\cdot)$ denotes the modified Bessel function of the second kind of order n.

To implement the first-order IS algorithm we need the conditional distributions $X_i \mid X_i > \gamma$ and $\boldsymbol{X}_{-i} \mid X_i$. Assuming $\gamma > 0$ we can derive that $(X_i \mid X_i > \gamma) \sim \mathcal{E}(\sqrt{2})$. Further calculation gives

$$\boldsymbol{X}_{-1} \mid X_1 \overset{\mathcal{D}}{=} \frac{X_1}{Y_1}\boldsymbol{Y}_{-1} \mid (\sqrt{R}Y_1 = X_1) \overset{\mathcal{D}}{=} \frac{X_1}{Y_{1,X_1}}\boldsymbol{Y}_{-1}\,,$$

where $Y_{1,X_1} \sim (Y_1 \mid \sqrt{R}Y_1 = X_1)$, noting that $Y_{1,X_1} \perp\!\!\!\perp \boldsymbol{Y}_{-1}$ because of the independence between the entries of $\boldsymbol{Y}$. Direct calculation gives

$$f_{Y_i \mid \sqrt{R}Y_i}(y_i \mid x_i) = 2\,|y_i|\exp\left\{-x_i^2/y_i^2 - x_i^2/2 + \sqrt{2}\,|x_i|\right\}/(\sqrt{\pi}y_i^2)$$

which is the density of $\sqrt{X}$ where $X \sim \mathcal{IG}(\sqrt{2}|x_i|, 2x_i^2)$. This is summarised in the following algorithm.

Algorithm 1 Sampling $\boldsymbol{X}_{-i} \mid X_i > \gamma$ for the Laplace distribution

---

1: $X_i \leftarrow \mathcal{E}(\sqrt{2})$
2: $Y_{i,X_i} \leftarrow \mathcal{IG}(\sqrt{2}|X_i|, 2X_i^2)$.
3: $\boldsymbol{Y}_{-i} \leftarrow \mathcal{N}_{d-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$.
4: **return** $X_i\boldsymbol{Y}_{-i}/Y_{i,X_i}$.

5.1 Test setup

The estimators tested are $\widehat{\alpha}_0$ (crude Monte Carlo) and $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_1^{[1]}$, $\widehat{\alpha}_2^{[2]}$, $\widehat{(\beta_1 \ddagger \alpha)}$, $\widehat{(\beta_2 \ddagger \alpha)}$, defined in (1), (2), (12), (13), (18) and (20) respectively. As a reference, we show the true value α (calculated by numerical integration using MATHEMATICA), and the first two truncations of the IEF: $\overline{\alpha}(\gamma) := \sum_i \mathbb{P}(X_i > \gamma)$ and $\overline{\alpha}(\gamma) - q(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)$. Each estimator is given $R = 10^6$, and an asterisk is placed in table entries where the corresponding estimate had 0 variance (i.e., the estimator had degenerated).

5.2 Results

Estimators	γ			
	2	4	6	8
α	5.633e-02	1.095e-04	3.838e-09	2.481e-15
$\widehat{\alpha}_0$	5.651e-02	1.140e-04	0*	0*
$\overline{\alpha}$	9.100e-02	1.267e-04	3.946e-09	2.488e-15
$\overline{\alpha}-q$	4.000e-02	1.055e-04	3.827e-09	2.480e-15
$\widehat{\alpha}_1$	5.650e-02	1.047e-04	3.946e-09*	2.488e-15*
$\widehat{\alpha}_2$	5.605e-02	1.075e-04	3.827e-09*	2.480e-15*
$\widehat{\alpha}_1^{[1]}$	5.637e-02	1.096e-04	3.837e-09	2.481e-15
$\widehat{\alpha}_2^{[2]}$	5.633e-02	1.095e-04	3.838e-09	2.481e-15
$\widehat{(\beta_1 \ddagger \alpha)}$	5.634e-02	1.095e-04	3.838e-09	2.480e-15
$\widehat{(\beta_2 \ddagger \alpha)}$	5.631e-02	1.095e-04	3.838e-09	2.481e-15

Table 3: Estimates of $\mathbb{P}(M > \gamma)$ where $M = \max_i X_i$ and $\boldsymbol{X} \sim \mathcal{N}_4(\boldsymbol{0}_4, \boldsymbol{\Sigma})$, $\rho = 0.75$.

Estimators	γ			
	2	4	6	8
$\widehat{\alpha}_0$	3.109e-03	4.075e-02	1*	1*
$\overline{\alpha}$	6.154e-01	1.566e-01	2.822e-02	3.142e-03
$\overline{\alpha}-q$	2.899e-01	3.665e-02	2.827e-03	1.147e-04
$\widehat{\alpha}_1$	2.977e-03	4.429e-02	2.822e-02*	3.142e-03*
$\widehat{\alpha}_2$	5.077e-03	1.839e-02	2.827e-03*	1.147e-04*
$\widehat{\alpha}_1^{[1]}$	6.918e-04	4.639e-04	1.747e-04	2.192e-05
$\widehat{\alpha}_2^{[2]}$	7.838e-08	8.647e-05	1.237e-05	4.010e-08
$\widehat{(\beta_1 \ddagger \alpha)}$	6.564e-05	7.046e-05	6.227e-05	4.362e-05
$\widehat{(\beta_2 \ddagger \alpha)}$	3.493e-04	1.593e-05	6.883e-06	3.340e-07

Table 4: Absolute relative errors of the estimates in Table 3.

Estimators	γ			
	2	4	6	8
$\widehat{\alpha}_0$	2.309e-01	1.068e-02	0	0
$\widehat{\alpha}_1$	2.557e-01	5.099e-03	0	0
$\widehat{\alpha}_2$	1.885e-01	1.414e-03	0	0
$\widehat{\alpha}_1^{[1]}$	2.817e-02	3.071e-05	4.650e-10	9.972e-17
$\widehat{\alpha}_2^{[2]}$	9.901e-03	4.244e-06	1.908e-11	8.575e-19
$\widehat{(\beta_1 \ddagger \alpha)}$	1.929e-02	2.089e-05	3.197e-10	6.994e-17
$\widehat{(\beta_2 \ddagger \alpha)}$	1.306e-02	5.265e-06	2.310e-11	1.035e-18

Table 5: Standard deviations of the estimates in Table 3.

Estimators	γ			
	6	8	10	12
α	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{\alpha}_0$	3.910e-04	2.000e-05	2.000e-06	0*
$\overline{\alpha}$	4.130e-04	2.441e-05	1.443e-06	8.527e-08
$\overline{\alpha}-q$	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{\alpha}_1$	4.120e-04	2.441e-05*	1.443e-06*	8.527e-08*
$\widehat{\alpha}_2$	4.093e-04*	2.435e-05*	1.442e-06*	8.526e-08*
$\widehat{\alpha}_1^{[1]}$	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{(\beta_1 \ddagger \alpha)}$	4.093e-04	2.435e-05	1.442e-06	8.526e-08

Table 6: Estimates of $\mathbb{P}(M > \gamma)$ where $M = \max_i X_i$ and $\boldsymbol{X} \sim \mathcal{L}$, $d = 4$.

Estimators	γ			
	6	8	10	12
$\widehat{\alpha}_0$	4.472e-02	1.786e-01	3.873e-01	1*
$\overline{\alpha}$	8.959e-03	2.473e-03	6.987e-04	2.003e-04
$\overline{\alpha}-q$	8.067e-05	8.266e-06	8.757e-07	9.506e-08
$\widehat{\alpha}_1$	6.516e-03	2.473e-03*	6.987e-04*	2.003e-04*
$\widehat{\alpha}_2$	8.067e-05*	8.266e-06*	8.757e-07*	9.506e-08*
$\widehat{\alpha}_1^{[1]}$	8.470e-06	1.023e-05	3.019e-05	1.577e-05
$\widehat{(\beta_1 \ddagger \alpha)}$	4.515e-05	2.948e-05	2.151e-06	2.833e-06

Table 7: Absolute relative errors of the estimates in Table 6.

Estimators	γ			
	6	8	10	12
$\widehat{\alpha}_0$	1.977e-02	4.472e-03	1.414e-03	0
$\widehat{\alpha}_1$	1.000e-03	0	0	0
$\widehat{\alpha}_2$	0	0	0	0
$\widehat{\alpha}_1^{[1]}$	2.735e-05	8.581e-07	2.752e-08	8.189e-10
$\widehat{(\beta_1 \ddagger \alpha)}$	1.937e-05	6.086e-07	1.908e-08	5.990e-10

Table 8: Standard deviations of the estimates in Table 6.

Test cases		γ				Test cases		γ			
d	ρ	2	4	6	8	d	ρ	2	4	6	8
	-0.25	0.00957	1*	1*	1*		-0.25	1*	1*	1*	1*
3	0	0.00255	1*	1*	1*	3	0	0.151*	1*	1*	1*
	0.5	0.00166	1*	1*	1*		0.5	0.0764	1*	1*	1*
	0.75	0.005	0.165	1*	1*		0.75	0.0172	0.754	1*	1*
	-0.25	0.00955	1*	1*	1*		-0.25	1*	1*	1*	1*
4	0	0.0185	1*	1*	1*	4	0	0.189	1*	1*	1*
	0.5	0.00139	1*	1*	1*		0.5	0.0153	1*	1*	1*
	0.75	0.00484	0.283	1*	1*		0.75	0.0175	0.502	1*	1*
Average		0.00663	0.806	1	1	Average		0.308	0.907	1	1
(a) $\widehat{\alpha}_1$ to $\overline{\alpha}$						(b) $\widehat{\alpha}_2$ to $\overline{\alpha}-q$					

Table 9: Ratios of absolute relative errors for pairs of estimators. Each row corresponds to a separate distribution for $\boldsymbol{X}$, each being $\mathcal{N}_d$ distributed with standard normal marginals and constant correlation ρ.

5.3 Discussion

We begin with some trends which we expected to find in the results:

- all estimators outperform crude Monte Carlo $\widehat{\alpha}_0$,
- the estimators which calculate $\mathbb{P}(X_i > \gamma)$ outperform those which don't,
- the estimators which calculate $\mathbb{P}(X_i > \gamma, X_j > \gamma)$ outperform those which only use the univariate $\mathbb{P}(X_i > \gamma)$,
- the importance sampling estimators improve upon their original counterparts,
- the second-order IS improves upon the first-order IS.

Also noticed in the performance of the $\widehat{\alpha}$ estimators:

- the $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ estimators often degenerated (i.e. had zero variance) to $\overline{\alpha}$ and $\overline{\alpha}-q$ respectively,
- the degeneration begin for smaller γ when the $\boldsymbol{X}$ had a weaker dependence structure.

Table 9 shows the degeneration of the estimators in various examples involving multivariate normal distributions.

The fact that the estimators degenerate is not wholly undesirable, as they degenerate to the deterministic functions $\overline{\alpha}$ and $\overline{\alpha}-q$ which are highly accurate when degeneration occurs. Obviously, for very large γ one would not resort to Monte Carlo methods as the asymptote $\overline{\alpha}$ would be accurate enough for most purposes; one could use the $\widehat{\alpha}$ estimators until the sample variance is below some threshold, then switch to the faster deterministic estimators $\overline{\alpha}$ and $\overline{\alpha}-q$.

Regarding the $\widehat{(\beta_1 \ddagger \alpha)}$ and $\widehat{(\beta_2 \ddagger \alpha)}$ estimators:

- their performance is roughly the same as than their $\widehat{\alpha}_1^{[1]}$ and $\widehat{\alpha}_2^{[2]}$ counterparts,
- they perform better when the dependence between the variables is weak.

One must remember that the $\widehat{\beta}_i$ estimators are valid for a much larger class of problems (estimating expectations, not just probabilities). Also, we would expect that the $\widehat{\beta}_i$-based estimators compare favorably to the $\widehat{\alpha}_i^{[i]}$ IS-based estimators when d is large, as the method involves no likelihood term which can degenerate.

6 Conclusion

In this paper we presented new estimators for the tail probability of a union of dependent rare events. The key idea in both estimators is that the tail probability of the such a rare event can be well approximated by the Bonferroni approximations:

$$\alpha = \mathbb{P}(A) \approx \sum_{i=1}^{k} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \text{ for } k = 1, 2 \,.$$

We provided conditions which ensure $\widehat{\alpha}_1$ and $\widehat{\beta}_i$ have logarithmic efficiency and bounded relative error. The estimators were tested on the classical example of rare maxima of random vectors. Furthermore, we note the fact that our $\widehat{\beta}_i$ estimators can be applied to a more general setting which could make useful for a larger variety of estimation problems.

6.1 Future work

In this paper we did not discuss stratification strategies for $\widehat{\beta}_i$ that could result in further reductions in variance. Nor did we investigate which permutations of the A_i minimise the variance of $\widehat{\beta}_i$. Further investigation into the use of $\widehat{\beta}_i$ to estimate tail probabilities of order statistics would be of value.

Acknowledgements

References

1. Adler, R.J.: An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Lecture Notes-Monograph Series **12**, i–155 (1990)
2. Adler, R.J., Blanchet, J.H., Liu, J., et al.: Efficient Monte Carlo for high excursions of Gaussian random fields. The Annals of Applied Probability **22**(3), 1167–1214 (2012)
3. Andersen, L.N., Laub, P.J., Rojas-Nandayapa, L.: Online accompaniment for "Efficient simulation for dependent rare events with applications to extremes" (2016). Available at `https://github.com/Pat-Laub/RareMaxima`
4. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, second edn. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1984)
5. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, third edn. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ (2003)

6. Asmussen, S., Albrecher, H.: Ruin Probabilities, second edn. Advanced Series on Statistical Science & Applied Probability, 14. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2010). DOI 10.1142/9789814282536. URL `http://dx.doi.org/10.1142/9789814282536`

7. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis, *Stochastic Modelling and Applied Probability series*, vol. 57. Springer (2007)

8. Berman, S.: Sojourns and Extremes of Stochastic Processes. CRC Press (1992)

9. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation, vol. 27. Cambridge university press (1989)

10. Botev, Z.I.: The normal law under linear restrictions: Simulation and estimation via minimax tilting. Journal of the Royal Statistical Society, Series B **79**, pp. 1–24

11. Breslaw, J.: Random sampling from a truncated multivariate normal distribution. Applied Mathematics Letters **7**(1), 1–6 (1994)

12. Bryc, W.: The Normal Distribution: Characterizations with Applications, vol. 100. Springer Science & Business Media (2012)

13. Charpentier, A., Segers, J.: Tails of multivariate Archimedean copulas. Journal of Multivariate Analysis **100**(7), 1521–1537 (2009)

14. Cont, R.: Encyclopedia of Quantitative Finance (2010)

15. De Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer Science & Business Media (2007)

16. Eltoft, T., Kim, T., Lee, T.W.: On the multivariate Laplace distribution. IEEE Signal Processing Letters **13**(5), 300–303 (2006)

17. Glasserman, P.: Monte Carlo Methods in Financial Engineering, *Stochastic Modelling and Applied Probability series*, vol. 53. Springer (2003)

18. Hashorva, E.: Asymptotics and bounds for multivariate Gaussian tails. Journal of Theoretical Probability **18**(1), 79–97 (2005)

19. Hashorva, E.: Asymptotic properties of type I elliptical random vectors. Extremes **10**(4), 175–206 (2007)

20. Hashorva, E.: On the residual dependence index of elliptical distributions. Statistics & Probability Letters **80**(13), 1070–1078 (2010)

21. Heffernan, J.E.: A directory of coefficients of tail dependence. Extremes **3**(3), 279–290 (2000)

22. Huang, Z., Shahabuddin, P.: Rare-event, heavy-tailed simulations using hazard function transformations, with applications to value-at-risk. In: Proceedings of the 2003 Winter Simulation Conference, 2003. (2003)

23. Hult, H., Lindskog, F., et al.: Multivariate extremes, aggregation and dependence in elliptical distributions. Advances in Applied Probability **34**(3), 587–608 (2002)

24. Joe, H.: Multivariate Models and Multivariate Dependence Concepts. CRC Press (1997)

25. Kotz, S., Kozubowski, T.J., Podgórski, K.: Asymmetric multivariate Laplace distribution. In: The Laplace Distribution and Generalizations, pp. 239–272. Springer (2001)

26. Ledford, A.W., Tawn, J.A.: Statistics for near independence in multivariate extreme values. Biometrika **83**(1), 169–187 (1996)

27. Ledford, A.W., Tawn, J.A.: Modelling dependence within joint tail regions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **59**(2), 475–499 (1997)

28. Ledford, A.W., Tawn, J.A., et al.: Concomitant tail behaviour for extremes. Advances in Applied Probability **30**(1), 197–215 (1998)

29. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools, 2nd edn. Princeton University Press (2015)

30. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer Science & Business Media (2007)

31. Nolde, N.: Geometric interpretation of the residual dependence coefficient. Journal of Multivariate Analysis **123**, 85–95 (2014)

32. Rausand, M., Høyland, A.: System Reliability Theory: Models, Statistical Methods, and Applications, vol. 396. John Wiley & Sons (2004)

33. Resnick, S.: Hidden regular variation, second order regular variation and asymptotic independence. Extremes **5**(4), 303–336 (2002)

34. Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer (1987)

35. Robert, C.P.: Simulation of truncated normal variables. Statistics and computing **5**(2), 121–125 (1995)

36. Rubinstein, R.Y., Kroese, D.P.: Simulation and the Monte Carlo method, vol. 707. John Wiley & Sons (2011)

37. Savage, I.R.: Mills ratio for multivariate normal distributions. J. Res. Nat. Bur. Standards Sect. B **66**, 93–96 (1962)
38. Sibuya, M.: Bivariate extreme statistics. Annals of the Institute of Statistical Mathematics **11**(2), 195–210 (1960)

A Elliptical distribution asymptotics

A.1 Asymptotic properties of normal distributions

In general, for an $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$, Theorem 2.6.1 of Bryc [12] states that for all measurable $A \subset \mathbb{R}^d$ the

$$\lim_{n \to \infty} \frac{1}{n^2} \log \mathbb{P}(\boldsymbol{X} \geq nA) = -\inf_{\boldsymbol{x} \in A} \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}. \tag{31}$$

The asymptotic properties of elliptical distributions also relate to this quadratic programming problem, which Hashorva [18,19] denotes as

$$\mathcal{P}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{t}) := \text{minimise } \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} \text{ under the linear constraint } \boldsymbol{x} \geq \boldsymbol{t}. \tag{32}$$

The program $\mathcal{P}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{t})$ is usually minimised at the boundary $\boldsymbol{t}$, and hence the asymptotic form (31) is very simple. This occurs when $\boldsymbol{\Sigma}^{-1}\boldsymbol{t} > \boldsymbol{0}$ (componentwise), a condition often called the *Savage condition* after Richard Savage [37]. For the cases when the Savage condition fails, the asymptotics change as some components of $\boldsymbol{X}$ become irrelevant in the limit. Figure 1 graphically shows some contours of $\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$ for some $\boldsymbol{\Sigma}$ which do and do not satisfy the Savage condition.
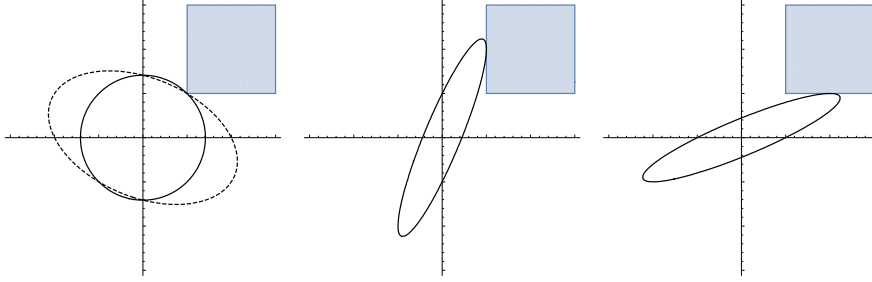

Fig. 1: Contours of $\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$ for example $\boldsymbol{\Sigma}$ which: (a) satisfy the Savage condition (i.e., $\boldsymbol{\Sigma}^{-1}\boldsymbol{1} > \boldsymbol{0}$), and (b)–(c) do not satisfy the condition. The covariance matrices, in MATLAB notation, are: (a) $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{\Sigma} = [2, -1/2; -1/2, 1]$, (b) $\boldsymbol{\Sigma} = [1, 2; 2, 5]$, and (c) $\boldsymbol{\Sigma} = [5, 2; 2, 1]$.

A.2 Asymptotic properties of type I elliptical distributions

Take $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ where the radial distribution $F \in \text{MDA}(\text{Gumbel})$ has support $(0, x_F)$, for some $x_F \in \overline{\mathbb{R}}$, and where $\{\sigma_1, \ldots, \sigma_d\}$ are in decreasing order. The univariate and bivariate asymptotics, $\mathbb{P}(X_i > \gamma)$ and $\mathbb{P}(X_i > \gamma, X_j > \gamma)$, can be written in terms of the scaling function $w(\gamma)$ and of $\overline{F}((\gamma - \mu)/\kappa)$ for some particular μ and κ. Theorem 12.3.1 of Berman [8] gives the univariate case,

$$\mathbb{P}(X_i > \gamma) = (1 + \text{o}(1)) \frac{\overline{F}(v_i(\gamma))}{\sqrt{2\pi v_i(\gamma) w(v_i(\gamma))}} \quad \text{as } \gamma \to x_F \tag{33}$$

where $v_i(\gamma) = (\gamma - \mu_i)/\sigma_i$. The bivariate case, i.e. $\mathbb{P}(X_i > \gamma, X_j > \gamma)$, relies on the following constants. Define $a_{ij} := \sigma_j/\sigma_i$. If $\rho_{ij} \geq a_{ij}$ then define

$$\mu_{ij} := \mu_j \quad \text{and} \quad \kappa_{ij} := \sigma_j$$

otherwise for $\rho_{ij} < a_{ij}$

$$\mu_{ij} := \frac{\mu_i - a_{ij}\rho_{ij}(\mu_1 + \mu_2) + a^2\mu_j}{\alpha_{ij}(1 - \rho_{ij}^2)} \quad \text{and} \quad \kappa_{ij} := \frac{\sigma_i^2\sigma_j^2(1 - \rho_{ij}^2)}{\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j + \sigma_j^2}.$$

Theorem 3 *Let (X_i, X_j) be a pair from a type I elliptical random vector $\boldsymbol{X} \sim E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ and consider $\gamma \nearrow x_F$. Then with $v_{ij}(\gamma) = (\gamma - \mu_{ij})/\kappa_{ij} + c_{ij}(\gamma)$ for some $c_{ij}(\gamma) \in o(1)$,*

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) = \overline{F}(v_{ij}(\gamma)) \times \begin{cases} \left(2\pi v_{ij}(\gamma)w(v_{ij}(\gamma))\right)^{-1/2}(1 + o(1)), & \text{if } \rho_{ij} > a_{ij}, \\ \left(2\pi v_{ij}(\gamma)w(v_{ij}(\gamma))\right)^{-1}(C_{a,\rho} + o(1)), & \text{if } \rho_{ij} < a_{ij}, \end{cases}$$

for a $C_{a,\rho} \in \mathbb{R}_+$. Furthermore, if either $\mu_i \geq \mu_j$ or $\lim_{\gamma \to x_F} w(\gamma)/\gamma < \infty$, then there exists a $C_\rho \in \mathbb{R}_+$ such that

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) = \overline{F}(v_{ij}(\gamma))\left(2\pi v_{ij}(\gamma)w(v_{ij}(\gamma))\right)^{-1/2}(C_\rho + o(1)), \quad \text{if } \rho_{ij} = a_{ij}.$$

Proof Use Theorem 2 of Hashorva [19]. First we consider the case $a_{ij} < \rho_{ij}$. In such a case it holds that

$$\lim_{\gamma \to x_F} \sqrt{\frac{w(v_j(\gamma))}{v_j(\gamma)}} \, (v_i(\gamma) - \rho_{ij}v_j(\gamma)) = \lim_{\gamma \to x_F} \sqrt{w(v_j(\gamma))v_j(\gamma)} \left(\frac{v_i(\gamma)}{v_j(\gamma)} - \rho_{ij}\right)$$

$$= \lim_{\gamma \to x_F} \sqrt{w(v_j(\gamma))v_j(\gamma)} \, (a_{ij} - \rho_{ij}) = -\infty.$$

Hence, the hypotheses of Case i) of Theorem 2 of Hashorva [19] hold and the first result follows. In the case where $a_{ij} = \rho_{ij}$ then

$$\lim_{\gamma \to x_F} \sqrt{\frac{w(v_j(\gamma))}{v_j(\gamma)}} \, (v_i(\gamma) - \rho_{ij}v_j(\gamma)) = \lim_{\gamma \to x_F} \sqrt{\frac{w(v_j(\gamma))}{v_j(\gamma)}} \frac{(\mu_j - \mu_i)}{\sigma_i}.$$

The last limit remains bounded from above if either $\mu_i > \mu_j$ or $\lim_{\gamma \to \infty} w(\gamma)/\gamma < \infty$. For the case $a_{ij} > \rho_{ij}$ we define $a_{ij}(\gamma) := v_i(\gamma)/v_j(\gamma)$ so $\lim_{\gamma \to \infty} a_{ij}(\gamma) = a_{ij}$.

We let

$$\tau_{ij}(\gamma) = \sqrt{\frac{1 - 2a_{ij}(\gamma)\rho_{ij} + a_{ij}^2(\gamma)}{1 - \rho_{ij}^2}}, \quad \tau_{ij} := \lim_{\gamma \to \infty} \tau_{ij}(\gamma) = \sqrt{\frac{1 - 2a_{ij}\rho_{ij} + a_{ij}^2}{1 - \rho_{ij}^2}}.$$

The results follows by noting that

$$v_j(\gamma)\tau_{ij}(\gamma) = v_{ij}(\gamma), \quad v_{ij}(\gamma) = \frac{\gamma - \mu_{ij}}{\tau_{ij}} + o(1).$$

$\square$