Efficient simulation for dependent rare events

with applications to extremes

Patrick J. Laub

The University of Queensland & Aarhus University

UNSW Statistics Seminar

19/05/17

Quick Bio

SE $\rightarrow$ Math. Now based at University of Queensland & Aarhus University, Denmark.

Supervisors: Phil Pollett, Søren Asmussen

Coauthors: Leonardo Rojas-Nandayapa, Lars Nørvang Andersen

Inviter: Peter Straka

Outline

- Introduction of problems & estimators
- Discussion of estimators & improvements
- Efficiency results
- Limitations

First problem

For a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with maximum $M = \max_i X_i$, the first problem we consider is estimating

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

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.

First glance at estimators

Our two main estimators in this setting are

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

$$\widehat{\alpha}_2 = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) - \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \Big[1 - E(\gamma) + \frac{E(\gamma)(E(\gamma)-1)}{2}\Big]\mathbb{1}\{E_r(\gamma) \geq 3\}\,.$$

ACEMS

Second problem

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 $\mathbf{X}$. The probability of a parallel circuit failing is a simple application for $\mathbb{P}(X_{(n)} > \gamma)$.

General setup

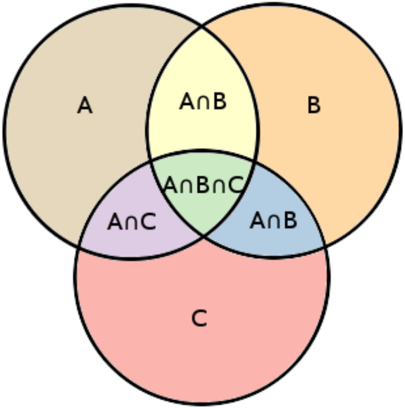
Let $A(\gamma) = \cup_{i=1}^{d} A_i(\gamma)$ be the union of events $A_1(\gamma), \ldots, A_d(\gamma)$ for an index parameter $\gamma \in \mathbb{R}$. We consider the problem of estimating $\mathbb{P}(A(\gamma))$ when the events are rare, that is, $\mathbb{P}(A(\gamma)) \to 0$ as $\gamma \to \infty$.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

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

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.

ACEMS

Inclusion–exclusion



$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - [\mathbb{P}(A, B) + \mathbb{P}(A, C) + \mathbb{P}(B, C)] + \mathbb{P}(A, B, C)$$

The inclusion–exclusion formula (IEF) provides a representation of α as a summation whose terms are decreasing in size. The formula is, for $A = \cup_i A_i$,

$$\alpha = \mathbb{P}(A) = \sum_{i=1}^{d} \mathbb{P}(A_i) - \sum_{1=i<j}^{d} \mathbb{P}(A_i, A_j) + \cdots + (-1)^{d+1} \mathbb{P}(A_1, \ldots, A_d)$$

$$= \sum_{i=1}^{d} (-1)^{i+1} \sum_{|I|=i} \mathbb{P}\Big(\bigcap_{j \in I} A_j \Big) .$$

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* [**asmussen2007stochastic**]).

Bonferonni inequalities

Truncating the summation can lead to bias, and indeed by the Bonferroni inequalities we have:

$$
\begin{aligned}
\mathbb{P}(A) = \mathbb{P}(\cup_i A_i) = \alpha &\le \sum_i \mathbb{P}(A_i) \quad \text{(Boole–Fréchet)} \\
\alpha &\ge \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) \\
\alpha &\le \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) + \sum_{i<j<k} \mathbb{P}(A_i, A_j, A_k)
\end{aligned}
$$

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

Constructing IEF estimators

Remember IEF:

$$\alpha = \sum_{i=1}^{d} (-1)^{i+1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{j \in I} A_j \right) = \sum_{i=1}^{d} (-1)^{i+1} \mathbb{E}\left[\sum_{|I|=i} \mathbb{1}\left(\bigcap_{j \in I} A_j \right) \right]$$

Proposition

For $i = 1, \ldots, d$, $\sum_{|I|=i} \mathbb{1}\left\{ \cap_{j \in I} A_j \right\} = \binom{E}{i} \mathbb{1}\{E \geq i\}$

Proof.

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

□

Estimators

$$\mathbb{E}\Big[\sum_{i=1}^{d}(-1)^{i-1}\binom{E}{i}\mathbb{1}\{E \geq i\}\Big] = \sum_{i=1}^{d}(-1)^{i-1}\,\mathbb{E}\Big[\binom{E}{i}\mathbb{1}\{E \geq i\}\Big]$$
$$= \mathrm{IEF}_1 + \mathrm{IEF}_2 + \cdots + \mathrm{IEF}_d$$
$$= \alpha\,.$$

We present estimators which deterministically *calculate* the first larger terms of the IEF and Monte Carlo (MC) *estimate* the remaining smaller terms using sample means of the above.

First estimator

We begin by constructing the single-replicate estimator $\widehat{\alpha}_1$ where the first summand is calculated and the remaining terms are estimated:

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

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

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

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

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

General form of the estimators

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

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

Properties of these estimators

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*. N.B. If we look at $\widehat{\alpha}_0$ we get the CMC estimator $\mathbb{1}\{E \geq 1\}$.

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. When combined with importance sampling the improvement is marked.

We do assume knowledge of marginal distributions.

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

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

$$\max_i \mathbb{P}(A_i) \le \alpha = \mathbb{P}(A) \le \sum_i \mathbb{P}(A_i) =: \overline{\alpha}\,. \tag{2}$$

The stochastic part of $\widehat{\alpha}_1$ is an unbiased estimate of $\overline{\alpha} - \alpha \le 0$. 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

$$\alpha(\gamma) \sim \sum_i \mathbb{P}(A_i(\gamma))\,,^{1}$$

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.

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

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

An alternative construction of $\{\widehat{\alpha}_1, \dots, \widehat{\alpha}_{d-1}\}$ 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.

ACEMS

Importance sampling (first-order)

Standard IS theory says condition on $A = \cup_i A_i = \{E \geq 1\}$ occuring. We use a *mixture distribution* as a proposal. Say that we condition on A_i with probability

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

Why? If $\mathbb{P}(A_i(\gamma), A_j(\gamma)) = \mathrm{o}(\mathbb{P}(A_i(\gamma)))$ often occurs for all $i \neq j$, then

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

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

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

$$\Rightarrow \widehat{\alpha}_1^{[1]} := \frac{1}{R} \sum_{r=1}^{R} \frac{\overline{\alpha}}{E_r^{[1]}}\,, \tag{3}$$

where the $E_r^{[1]}$ are iid from $\mathbb{Q}^{[1]}$. Same as Adler et al. **[adler1990introduction]**.

Importance sampling (second-order)

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

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

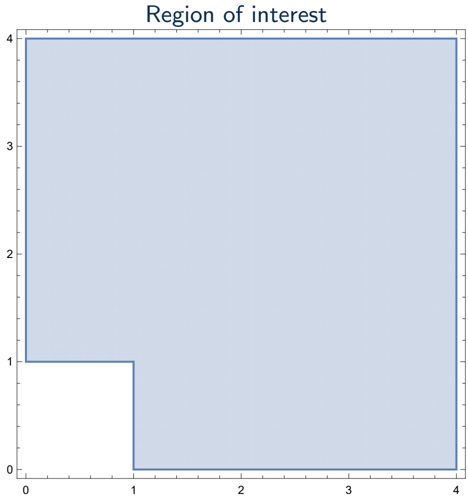
which induces a likelihood ratio of

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

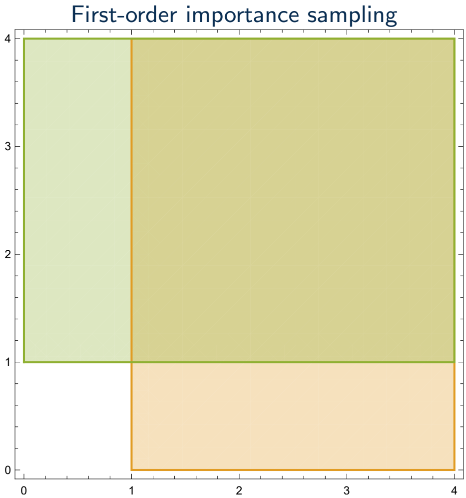
Thus, after simplifying, the estimator $\widehat{\alpha}_2$ under $\mathbb{Q}^{[2]}$ is

$$\widehat{\alpha}_2^{[2]} := \overline{\alpha} - \frac{2q}{R} \sum_{r=1}^{R} \frac{1}{E_r^{[2]}}\,. \tag{4}$$

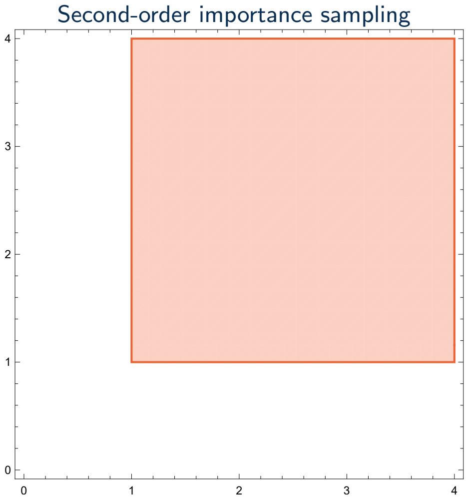
Example: $\alpha(1) = \mathbb{P}(\max\{X_1, X_2\} > 1)$



Example: $\alpha(1) = \mathbb{P}(\max\{X_1, X_2\} > 1)$



Example: $\alpha(1) = \mathbb{P}(\max\{X_1, X_2\} > 1)$



Importance sampling (extra requirements)

First-order IS:

- can simulate from $\mathbb{P}(\cdot \mid A_i)$,
- can calculate the $\mathbb{P}(A_i)$.

Second-order IS:

- can simulate from $\mathbb{P}(\cdot \mid A_i, A_j)$,
- can calculate the $\mathbb{P}(A_i)$ and $\mathbb{P}(A_i, A_j)$.

Normally (at least for extremes) can calculate $\mathbb{P}(A_i)$ and $\mathbb{P}(A_i, A_j)$ with MATHEMATICA or MATLAB. The prohibitive part is being able to simulate from conditionals.

Second problem – 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 the partition

$$A := \bigcup_{i=1}^{d} A_i = A_1 \cup (A_1^c A_2) \cup \cdots \cup (A_1^c \ldots A_{d-1}^c A_d) . \tag{5}$$

This gives us

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

If we assume it is possible to sample from the $\mathbb{P}(\,\cdot \mid A_i)$ conditional distributions (same as for $\widehat{\alpha}_1^{[1]}$) then each of these conditional expectations can be estimated by sample means:

$$\widehat{\beta}_1 := \sum_{i=1}^{d} \frac{\mathbb{P}(A_i)}{\lceil R/d \rceil} \sum_{r=1}^{\lceil R/d \rceil} Y_{i,r} \mathbb{1}\{A_1^c \ldots A_{i-1}^c\}_{i,r} . \tag{6}$$

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

Partition

Proposition

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

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

Then

$$\{E \geq m\} = \bigcup_{|I|=m} B_I = \bigcup_{|I|=m} B_I C_I. \tag{7}$$

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

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

General estimators of β

This proposition implies that

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

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 $\mathbb{P}(\,\cdot \mid B_I)$, then the following is an unbiased estimator of $\mathbb{E}[Y\mathbb{1}\{E \geq n\}]$:

$$\widehat{\beta}_n := \sum_{|I|=n} \frac{\mathbb{P}(B_I)}{\lceil R/\binom{d}{n}\rceil} \sum_{r=1}^{\lceil R/\binom{d}{n}\rceil} Y_{I,r}\mathbb{1}\{C_I\}_{I,r}\,. \tag{8}$$

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

Efficiency (definition)

Definition

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

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

has *logarithmic efficiency*, *bounded relative error*, or *vanishing relative error* respectively.

Efficiency (for our estimators)

Proposition

If for the estimator $\widehat{\alpha}_1$ *(*$\forall \varepsilon > 0$*)*

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

then the estimator has LE, BRE respectively.

Proposition

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

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

Efficiency results

- If the A_i are independent events then the estimator $\widehat{\alpha}_1$ has BRE.
- More generally? Again consider rare maxima, and to simplify, consider $X_i \overset{\mathcal{D}}{=} X_j$.
    - If $\exists$ asymptotic dependence ($\lambda > 0$), then $\widehat{\alpha}_1$ doesn't have BRE.
    - If asymptotic independence ($\lambda = 0$), need to look at *residual tail index* η:
        - BRE if $\eta < 1/2$.
        - LE if $\eta = 1/2$.
    - For exchangable Archimedean copulas with generator ψ, we have BRE if $\psi^{\leftarrow} \in C^2$ and $(\psi^{\leftarrow})''$ is bounded at 0.
    - For $\mathbf{X} \sim \mathcal{ELL}(\mu, \mathbf{\Sigma}, F)$ where $F \in \text{MDA(Gumbel)}$, we have conditions for when $\widehat{\alpha}_1$ has LE and when BRE. (This gives normal case.)
- The estimator $(\widehat{\beta_1 \ddagger \alpha})$ from has BRE.

Look at

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

where $\lambda_{ij} \in [0, 1]$ is called the *(upper) tail dependence parameter (or coefficient)*.

The canonical examples are the (non-degenerate) bivariate normal distribution for AI, and the bivariate Student t distribution for AD.

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

Residual tail index

Ledford and Tawn first noted that the joint survivor functions for a wide array of bivariate distributions satisfy

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

for a slowly-varying $L(\gamma)$ and an $\eta \in (0, 1]$.

In other words, this 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* (or, confusingly, the *coefficient of tail dependence*).

Efficiency (using residual tail index)

Proposition

If the Ledford & Tawn form is satisfied for the maximal pair of $\mathbf{X}$*, 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:*

1. *BRE if* $\eta < 1/2$ *or if* $\eta = 1/2$ *and* $L(\gamma) \not\to \infty$ *as* $\gamma \to \infty$*,*
2. *LE if* $\eta = 1/2$*.*

Proof.

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

□

Copulas and their residual tail indices

Table: Residual tail dependence index η and $L(x)$ for various copulas. This is a subset of Table 1 of [**heffernan2000directory**] (their row numbers are preserved).

#	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	$\delta/(1-\mathrm{e}^{-\delta})$
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	$\delta/(1-\delta)$

(a) Copulas with 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	$2-2^{1/\theta}$

(b) Copulas without BRE.

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

Theorem (Thm. 3.4 of [charpentier2009tails])

Let $(U_1, \dots, U_n) \sim C$ 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

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

Efficiency (cases)

- If the A_i are independent events then the estimator $\widehat{\alpha}_1$ has BRE.
- More generally? Again consider rare maxima, and to simplify, consider $X_i \overset{\mathcal{D}}{=} X_j$.
  - If $\exists$ asymptotic dependence ($\lambda > 0$), then $\widehat{\alpha}_1$ doesn't have BRE.
  - If asymptotic independence ($\lambda = 0$), need to look at *residual tail index* η:
    - BRE if $\eta < 1/2$.
    - LE if $\eta = 1/2$.
  - For exchangable Archimedean copulas with generator ψ, we have BRE if $\psi^{\leftarrow} \in C^2$ and $(\psi^{\leftarrow})''$ is bounded at 0.
  - For $\mathbf{X} \sim \mathcal{ELL}(\mu, \mathbf{\Sigma}, F)$ where $F \in \text{MDA(Gumbel)}$, we have conditions for when $\widehat{\alpha}_1$ has LE and when BRE. (This gives normal case.)
- The estimator $(\widehat{\beta_1 \ddagger \alpha})$ from has BRE.

Numerical example: multivariate normal ($R = 10^6$)

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

Numerical example: multivariate normal ($R = 10^6$)

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: Relative errors of the estimates of $\mathbb{P}(M > \gamma)$ where $\mathbf{X} \sim \mathcal{N}_4(\mathbf{0}_4, \mathbf{\Sigma})$, $\rho = 0.75$.

Numerical example: multivariate Laplace ($R = 10^6$)

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

Numerical example: multivariate Laplace ($R = 10^6$)

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: Relative errors of the estimates of $\mathbb{P}(M > \gamma)$ where $\mathbf{X} \sim \mathcal{L}$, $d = 4$.

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

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

The distribution has been applied in a financial context [**huang2003rare**], and is examined in [**eltoft2006multivariate**, **kotz2001asymmetric**]. From the former we have that the density of $\mathcal{L}$ is

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

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

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

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

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 do not,
- 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 $\mathbf{X}$ had a weaker dependence structure.

Limitations

We do assume knowledge of marginal distributions. If we just have joint pdf...

Asymptotic properties $\not\Rightarrow$ finite-term accuracy

Who actually wants to estimate probabilities of events under 10^{-10}?

Who actually believes probability estimates of events under 10^{-10}?

Limitations

We do assume knowledge of marginal distributions. If we just have joint pdf...

Asymptotic properties $\not\Rightarrow$ finite-term accuracy

Who actually wants to estimate probabilities of events under 10^{-10}?

Who actually believes probability estimates of events under 10^{-10}?

Thanks for listening!