Simple and efficient estimators for rare-event maxima and sums

Patrick J. Laub

University of Queensland, Brisbane Australia and
Aarhus University, Aarhus Denmark

Greenland

August 1, 2016

Outline

Part 1: $\mathbb{P}(\max\{X_1, \ldots, X_d\} > \gamma)$ for large γ

- present a collection of (raw) estimators,
- tradeoff between numerical integration and MC estimation,
- bounded relative error.

Joint with Leonardo Rojas-Nandayapa (UL) & Lars Nørvang Andersen (AU).

Outline

Part 1: $\mathbb{P}(\max\{X_1, \dots, X_d\} > \gamma)$ for large γ

- present a collection of (raw) estimators,
- tradeoff between numerical integration and MC estimation,
- bounded relative error.

Joint with Leonardo Rojas-Nandayapa (UL) & Lars Nørvang Andersen (AU).

Part 2: $\mathbb{P}(X_1 + \cdots + X_d > \gamma)$ for large γ

- present a framework for IS estimation,
- uses asymptotic information about the sum,
- incorporates an MCMC step.

Joint with Thomas Taimre (UQ).

Outline

Part 1: $\mathbb{P}(\max\{X_1, \ldots, X_d\} > \gamma)$ for large γ

- present a collection of (raw) estimators,
- tradeoff between numerical integration and MC estimation,
- bounded relative error.

Joint with Leonardo Rojas-Nandayapa (UL) & Lars Nørvang Andersen (AU).

Part 2: $\mathbb{P}(X_1 + \cdots + X_d > \gamma)$ for large γ

- present a framework for IS estimation,
- uses asymptotic information about the sum,
- incorporates an MCMC step.

Joint with Thomas Taimre (UQ).

My priorities: explain the intuition, mention theoretical and numerical results, get feedback.

Rare maxima

Say $\boldsymbol{X} = (X_1, \dots, X_d) \sim F(\cdot)$, and $M := \max\{X_1, \dots, X_d\}$. Want to know

$$\ell(\gamma) := \mathbb{P}(M \geq \gamma)\,.$$

Assume that γ becomes large, so $\ell(\gamma)$ is a rare event.

Rare maxima

Say $\boldsymbol{X} = (X_1, \ldots, X_d) \sim F(\cdot)$, and $M := \max\{X_1, \ldots, X_d\}$. Want to know

$$\ell(\gamma) := \mathbb{P}(M \geq \gamma) \,.$$

Assume that γ becomes large, so $\ell(\gamma)$ is a rare event.

Original Idea: If $\{X_i \geq \gamma\}$ is **rare**, then the chance of *two* or more components $\geq \gamma$ must be **'second-order rare'**! I.e.,

$$\mathbb{P}(M \geq \gamma) \approx \sum_{i=1}^{d} \mathbb{P}(X_i \geq \gamma) \,. \qquad \diamond$$

Rare maxima

Say $\boldsymbol{X} = (X_1, \ldots, X_d) \sim F(\cdot)$, and $M := \max\{X_1, \ldots, X_d\}$. Want to know

$$\ell(\gamma) := \mathbb{P}(M \geq \gamma)\,.$$

Assume that γ becomes large, so $\ell(\gamma)$ is a rare event.

Original Idea: If $\{X_i \geq \gamma\}$ is **rare**, then the chance of *two* or more components $\geq \gamma$ must be **'second-order rare'**! I.e.,

$$\mathbb{P}(M \geq \gamma) \approx \sum_{i=1}^{d} \mathbb{P}(X_i \geq \gamma)\,. \qquad \diamond$$

Indeed, the *Boole–Fréchet inequalities* tell us that

$$\max_i\{\mathbb{P}(X_i \geq \gamma)\} \leq \mathbb{P}(M \geq \gamma) \leq \sum_{i=1}^{d} \mathbb{P}(X_i \geq \gamma)\,.$$

Inclusion–exclusion

As

$$\ell(\gamma) = \mathbb{P}(M \geq \gamma) \; = \; \mathbb{P}(X_1 \geq \gamma \cup \cdots \cup X_d \geq \gamma)$$

Inclusion–exclusion

As

$$\ell(\gamma) = \mathbb{P}(M \geq \gamma) \;=\; \mathbb{P}(X_1 \geq \gamma \cup \cdots \cup X_d \geq \gamma)$$

we can use the inclusion–exclusion principle

$$\ell(\gamma) = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \cdots + (-1)^{d-1}\,\mathbb{P}(X_1 > \gamma, \ldots, X_d > \gamma)\,,$$

Inclusion–exclusion

As

$$\ell(\gamma) = \mathbb{P}(M \geq \gamma) \;=\; \mathbb{P}(X_1 \geq \gamma \cup \cdots \cup X_d \geq \gamma)$$

we can use the inclusion–exclusion principle

$$\ell(\gamma) = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \cdots + (-1)^{d-1}\, \mathbb{P}(X_1 > \gamma, \dots, X_d > \gamma)\,,$$

which can be rewritten as

$$\ell(\gamma) = \sum_{i=1}^{d} \Big[(-1)^{i-1} \sum_{\substack{I \subset \{1,\dots,d\} \\ |I| = i}} \mathbb{P}(\{X_j > \gamma; j \in I\}) \Big]\,.$$

Rare maxima

Define $E_\gamma(\omega) := \sum_{i=1}^d \mathbf{1}\{X_i(\omega) > \gamma\}$, the number of *exceedences* over level γ for the vector $\boldsymbol{X}(\omega)$. Note,

$$\mathbb{E}\Big[\sum_{i=1}^{d}(-1)^{i-1}\binom{E_\gamma}{i}\mathbf{1}\{E_\gamma \geq i\}\Big] = \ell(\gamma)\,. \tag{1}$$

Rare maxima

Define $E_\gamma(\omega) := \sum_{i=1}^d \mathbf{1}\{X_i(\omega) > \gamma\}$, the number of *exceedences* over level γ for the vector $\boldsymbol{X}(\omega)$. Note,

$$\mathbb{E}\left[\sum_{i=1}^{d}(-1)^{i-1}\binom{E_\gamma}{i}\mathbf{1}\{E_\gamma \geq i\}\right] = \ell(\gamma)\,. \tag{1}$$

In particular,

$$\mathbb{E}\left[\binom{E_\gamma}{1}\mathbf{1}\{E_\gamma \geq 1\}\right] = \mathbb{E}[E_\gamma] = \sum_i \mathbb{P}(X_i > \gamma)\,,$$

$$\mathbb{E}\left[-\binom{E_\gamma}{2}\mathbf{1}\{E_\gamma \geq 2\}\right] = \mathbb{E}\left[-\frac{1}{2}E_\gamma(E_\gamma - 1)\right] = -\sum_{i<j}\mathbb{P}(X_i > \gamma, X_j > \gamma)\,,$$

and so on.

Building an estimator 1

Firstly, we make the estimator where the first summand of (1) is calculated, and the remaining terms are MC estimated:

$$\widehat{\ell}_1(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) + \sum_{i=2}^{d} \left[(-1)^{i-1} \binom{E_\gamma}{i} \mathbf{1}\{E_\gamma \geq i\} \right].$$

Building an estimator 1

Firstly, we make the estimator where the first summand of (1) is calculated, and the remaining terms are MC estimated:

$$\widehat{\ell}_1(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) + \sum_{i=2}^{d} \left[(-1)^{i-1} \binom{E_\gamma}{i} \mathbf{1}\{E_\gamma \geq i\} \right].$$

As $\sum_{k=0}^{n} (-1)^{k-1} \binom{n}{k} = 0$, we can simplify this as

$$\widehat{\ell}_1(\gamma) = \sum_i \mathbb{P}(X_i > \gamma) + (1 - E_\gamma) \mathbf{1}\{E_\gamma \geq 2\}.$$

Building an estimator 2

The estimator where the first two summands of incl-excl. are calculated is

$$\widehat{\ell}_2(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \sum_{i=3}^{d} \left[(-1)^{i-1} \binom{E_\gamma}{i} \mathbf{1}\{E_\gamma \geq i\} \right]$$

Building an estimator 2

The estimator where the first two summands of incl-excl. are calculated is

$$\widehat{\ell}_2(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \sum_{i=3}^{d} \left[(-1)^{i-1} \binom{E_\gamma}{i} \mathbf{1}\{E_\gamma \geq i\} \right]$$

which simplifies to

$$\widehat{\ell}_2(\gamma) = \sum_i \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) + \left[1 - E_\gamma + \frac{1}{2} E_\gamma (E_\gamma - 1) \right] \mathbf{1}\{E_\gamma \geq 3\}\,.$$

Building the general estimator

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

$$\widehat{\ell}_n(\gamma) := \sum_{i=1}^{n} \Big[(-1)^{i-1} \sum_{\substack{I \subset \{1,\ldots,d\} \\ |I|=i}} \mathbb{P}(\{X_j > \gamma; j \in I\})\Big] + \Big[\sum_{i=0}^{n}(-1)^i \binom{E_\gamma}{i}\Big] \mathbf{1}\{E_\gamma \geq n+1\}\,.$$

Nice interpretations

Overall: set of estimators $\{\widehat{\ell}_1, \ldots, \widehat{\ell}_{d-1}\}$ that control

Numerical Integration	Monte Carlo Simulation

$\widehat{\ell}_1(\gamma)$ specifically: Uses MC to estimate the difference between $\mathbb{P}(M > \gamma)$ and its Boole–Fréchet upper bound, $\sum_i \mathbb{P}(X_i > \gamma)$.

Also we often have

$$\mathbb{P}(M > \gamma) \sim \sum_i \mathbb{P}(X_i > \gamma) \quad \text{as } \gamma \to \infty\,.$$

Here, $\widehat{\ell}_1$ uses MC to estimate the difference between $\mathbb{P}(M > \gamma)$ and its (first-order) asymptotic expansion.

Stuff I don't have time for

- Another derivation of estimators using *control variates*.
- We prove *bounded relative error* for $\widehat{\ell_1}$ in many cases.
- Associated importance sampling regimes are nice.
- Some numerical results (not many competitor algorithms).

Stuff I don't have time for

- Another derivation of estimators using *control variates*.
- We prove *bounded relative error* for $\widehat{\ell_1}$ in many cases.
- Associated importance sampling regimes are nice.
- Some numerical results (not many competitor algorithms).

Now time for something (not so) completely different!

Rare Sums

Estimate:

$$\ell(\gamma) := \mathbb{P}(S > \gamma)\,,$$

for large γ, where

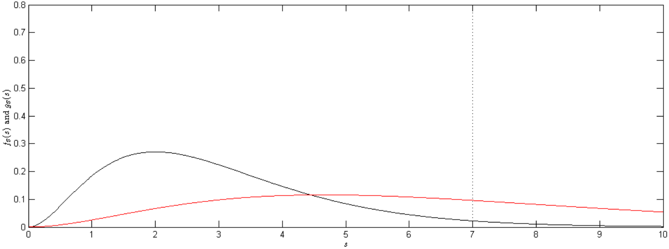
$$S := X_1 + \cdots + X_d\,, \qquad \boldsymbol{X} := (X_1, \ldots, X_d) \sim f_{\boldsymbol{X}}\,.$$

We assume each $X_i \geq 0$ is absolutely continuous, with marginal distributions F_i and densities f_i.

Importance Sampling

Importance sampling (IS):

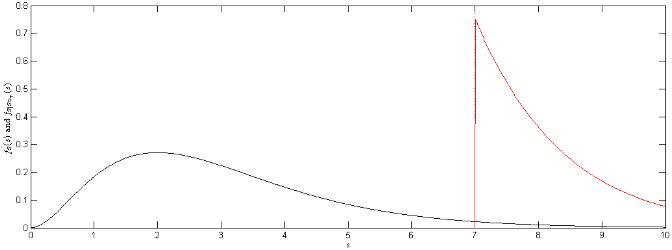
$$\ell(\gamma) = \mathbb{E}_{f_{\boldsymbol{X}}}\left[\mathbf{1}_{\{S>\gamma\}}\right] = \mathbb{E}_{g_{\boldsymbol{X}}}\left[\mathbf{1}_{\{S>\gamma\}}\,\frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{g_{\boldsymbol{X}}(\boldsymbol{X})}\right].$$



Importance Sampling

Importance sampling (IS):

$$\ell(\gamma) = \mathbb{E}_{f_{\boldsymbol{X}}}\left[\mathbf{1}_{\{S>\gamma\}}\right] = \mathbb{E}_{g_{\boldsymbol{X}}}\left[\mathbf{1}_{\{S>\gamma\}}\,\frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{g_{\boldsymbol{X}}(\boldsymbol{X})}\right].$$



Zero variance IS density:

$$g_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}\,|\,S>\gamma}(\boldsymbol{x}) := \frac{f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{S>\gamma\}}(\boldsymbol{x})}{\ell(\gamma)}.$$

Good importance sampling

When estimating $\ell(\gamma) = \mathbb{P}(S > \gamma)$, we need to choose $g_{\boldsymbol{X}}$ so that

1. $\{S > \gamma\}$ happens more often (preferably every time), and
2. $g_{\boldsymbol{X}}$ is similar to $f_{\boldsymbol{X}}$ when $\{S > \gamma\}$.

Transform the problem

Change of variables (to $\approx L^1$ polar form):

$$\boldsymbol{X} \longrightarrow \left(X_1 + \cdots + X_d, \frac{\boldsymbol{X}}{X_1 + \cdots + X_d} \right) =: (S, \boldsymbol{\Theta}) \,.$$

The density of $(S, \boldsymbol{\Theta})$ is

$$f_{(S,\boldsymbol{\Theta})}(s, \boldsymbol{\theta}) = f_{\boldsymbol{X}}(s\boldsymbol{\theta}) \times |s|^{d-1} \,.$$

The Main Idea

Conceptually

$$f_{(S,\mathbf{\Theta})} = f_S \times f_{\mathbf{\Theta}|S} \,,$$

however we cannot access the RHS here. Consider IS densities

$$g_{(S,\mathbf{\Theta})} = g_S \times g_{\mathbf{\Theta}|S}$$

where we *can* access g_S and $g_{\mathbf{\Theta}|S}$.

Try to find a $g_{(S,\mathbf{\Theta})}$ where

$$g_{(S,\mathbf{\Theta})} \approx f_{(S,\mathbf{\Theta})} \quad \text{(for large } S\text{)}\,,$$

by making sure

$$g_S \approx f_S \quad \text{and} \quad g_{\mathbf{\Theta}|S} \approx f_{\mathbf{\Theta}|S} \quad \text{(for large } S\text{)}\,.$$

Resulting IS Estimator

$$\ell(\gamma) = \mathbb{E}_{g_{(S,\mathbf{\Theta})}} \left[\mathbf{1}_{\{S>\gamma\}} \frac{f_{(S,\mathbf{\Theta})}(S,\mathbf{\Theta})}{g_S(S) g_{\mathbf{\Theta}|S}(\mathbf{\Theta}|S)} \right] .$$

Resulting IS Estimator

$$\ell(\gamma) = \mathbb{E}_{g_{(S,\boldsymbol{\Theta})}} \left[\mathbf{1}_{\{S>\gamma\}} \, \frac{f_{(S,\boldsymbol{\Theta})}(S,\boldsymbol{\Theta})}{g_S(S) g_{\boldsymbol{\Theta}|S}(\boldsymbol{\Theta}|S)} \right] .$$

Truncate $g_{(S,\boldsymbol{\Theta})}$ to $\{S > \gamma\}$ and use this as the IS density:

$$\ell(\gamma) = \overline{G}_S(\gamma) \, \mathbb{E}_{g_{(S,\boldsymbol{\Theta})|S>\gamma}} \left[\frac{f_{(S,\boldsymbol{\Theta})}(S,\boldsymbol{\Theta})}{g_S(S) g_{\boldsymbol{\Theta}|S}(\boldsymbol{\Theta}|S)} \right] ,$$

where $\overline{G}_S(\gamma) := \int_\gamma^\infty g_S(s) \, \mathrm{d}s$.

Resulting IS Estimator

$$\ell(\gamma) = \mathbb{E}_{g_{(S,\boldsymbol{\Theta})}} \left[\mathbf{1}_{\{S>\gamma\}} \frac{f_{(S,\boldsymbol{\Theta})}(S,\boldsymbol{\Theta})}{g_S(S) g_{\boldsymbol{\Theta}|S}(\boldsymbol{\Theta}|S)} \right] .$$

Truncate $g_{(S,\boldsymbol{\Theta})}$ to $\{S > \gamma\}$ and use this as the IS density:

$$\ell(\gamma) = \overline{G}_S(\gamma) \, \mathbb{E}_{g_{(S,\boldsymbol{\Theta})|S>\gamma}} \left[\frac{f_{(S,\boldsymbol{\Theta})}(S,\boldsymbol{\Theta})}{g_S(S) g_{\boldsymbol{\Theta}|S}(\boldsymbol{\Theta}|S)} \right] ,$$

where $\overline{G}_S(\gamma) := \int_\gamma^\infty g_S(s) \, \mathrm{d}s$. Thus, the general form of the estimator is

$$\widehat{\ell}_{\mathrm{Gen}}(\gamma) := \overline{G}_S(\gamma) \frac{f_{(S,\boldsymbol{\Theta})}(S,\boldsymbol{\Theta})}{g_S(S) g_{\boldsymbol{\Theta}|S}(\boldsymbol{\Theta} \mid S)}$$

for $S \sim g_{S|S>\gamma} := g_S \mathbf{1}_{\{S>\gamma\}} / \overline{G}_S(\gamma)$ and $\boldsymbol{\Theta} \sim g_{\boldsymbol{\Theta}|S}(\,\cdot\,|S)$.

Choosing Radial Approximation g_S

Often know asymptotics of $\overline{F}_S(\gamma) = \mathbb{P}(S > \gamma)$,

$$\overline{F}_S(\gamma) = c_S \overline{\mathfrak{F}}_S(\gamma) \times [1 + \mathrm{o}(1)]$$

$$\Rightarrow f_S(s) = c_S \mathfrak{f}_S(s) \times [1 + \mathrm{o}(1)] \,.$$

In $\widehat{\ell}_{\mathrm{Gen}}(\gamma)$ choose $g_S = \mathfrak{f}_S$. Thus, draw $S \sim \mathfrak{f}_{S \mid S > \gamma}$, $\boldsymbol{\Theta} \sim g_{\boldsymbol{\Theta} \mid S}(\,\cdot \mid S)$, then

$$\widehat{\ell}_{\mathrm{TP}}(\gamma) := c_S \overline{\mathfrak{F}}_S(\gamma) \times \frac{f_{(S, \boldsymbol{\Theta})}(S, \boldsymbol{\Theta})}{c_S \mathfrak{f}_S(S) g_{\boldsymbol{\Theta} \mid S}(\boldsymbol{\Theta} \mid S)} \,.$$

Choosing Radial Approximation g_S

Often know asymptotics of $\overline{F}_S(\gamma) = \mathbb{P}(S > \gamma)$,

$$\overline{F}_S(\gamma) = c_S \overline{\mathfrak{F}}_S(\gamma) \times [1 + \mathrm{o}(1)]$$

$$\Rightarrow f_S(s) = c_S \mathfrak{f}_S(s) \times [1 + \mathrm{o}(1)] \,.$$

In $\widehat{\ell}_{\mathrm{Gen}}(\gamma)$ choose $g_S = \mathfrak{f}_S$. Thus, draw $S \sim \mathfrak{f}_{S \mid S > \gamma}$, $\boldsymbol{\Theta} \sim g_{\boldsymbol{\Theta} \mid S}(\,\cdot \mid S)$, then

$$\widehat{\ell}_{\mathrm{TP}}(\gamma) := \underbrace{c_S \overline{\mathfrak{F}}_S(\gamma)}_{\text{Asymptotic}} \times \underbrace{\frac{f_{(S,\boldsymbol{\Theta})}(S, \boldsymbol{\Theta})}{c_S \mathfrak{f}_S(S) g_{\boldsymbol{\Theta} \mid S}(\boldsymbol{\Theta} \mid S)}}_{1+\mathrm{o}(1)} \,.$$

Choosing Angular Approximation $g_{\mathbf{\Theta}|S}$ in Theory

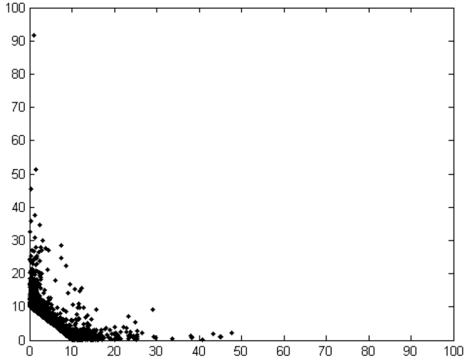
For the estimator to be very accurate, we must understand *how* the summands behave when producing a large sum. We consider three cases:

1. Subexponential tail decay: $\lim_{\gamma \to \infty} \overline{F^{*n}}(\gamma)/\overline{F}(\gamma) = n$.
2. Superexponential tail decay: $\lim_{\gamma \to \infty} \gamma^{-1} \log(\overline{F}(\gamma)) = -\infty$
3. "Exponential" tail decay: $\lim_{\gamma \to \infty} \gamma^{-1} \log(\overline{F}(\gamma)) = -\lambda$, $\lambda \in (0, \infty)$

Subexponential Tail Decay

Example

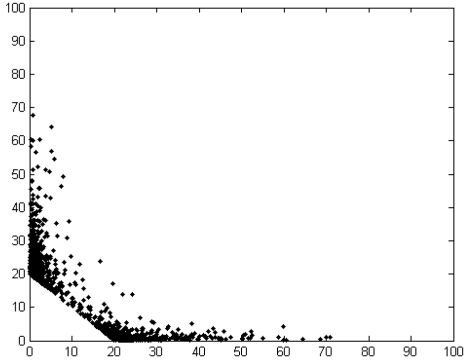
$X_1, X_2 \sim \mathrm{LogN}(0,1)$: $f(x) = (x\sqrt{2\pi})^{-1}\,\mathrm{e}^{-\frac{\log(x)^2}{2}}$.



Subexponential Tail Decay

Example

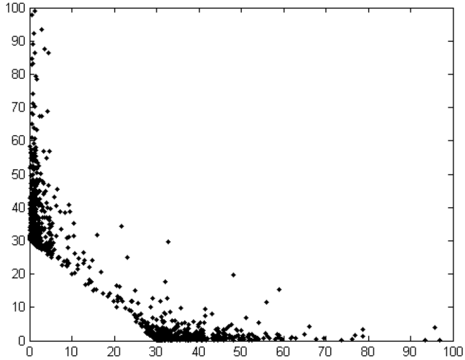
$X_1, X_2 \sim \mathrm{LogN}(0,1)$: $f(x) = (x\sqrt{2\pi})^{-1}\,\mathrm{e}^{-\frac{\log(x)^2}{2}}$.



Subexponential Tail Decay

Example

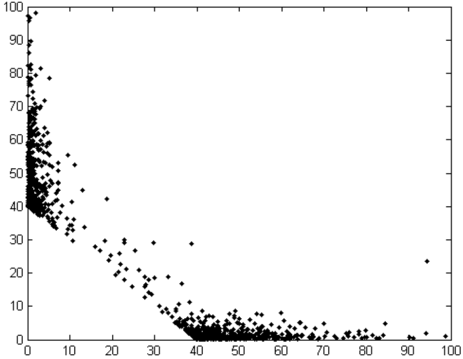
$X_1, X_2 \sim \mathrm{LogN}(0,1)$: $f(x) = (x\sqrt{2\pi})^{-1}\,\mathrm{e}^{-\frac{\log(x)^2}{2}}$.



Subexponential Tail Decay

Example

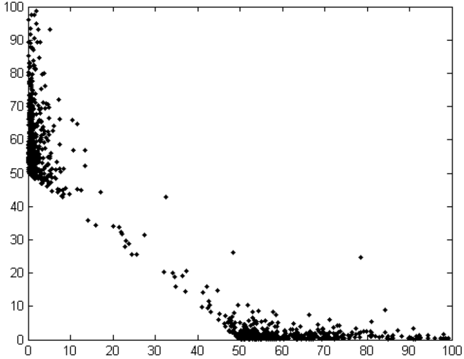
$X_1, X_2 \sim \mathrm{LogN}(0,1)$: $f(x) = (x\sqrt{2\pi})^{-1}\,\mathrm{e}^{-\frac{\log(x)^2}{2}}$.



Subexponential Tail Decay

Example

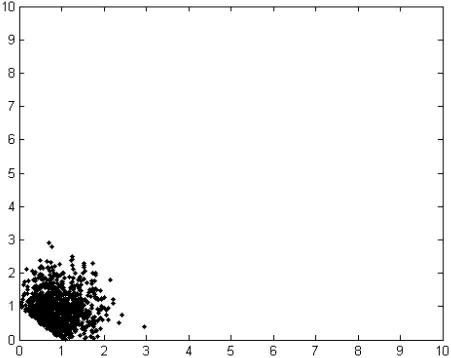
$X_1, X_2 \sim \mathrm{LogN}(0,1)$: $f(x) = (x\sqrt{2\pi})^{-1}\, \mathrm{e}^{-\frac{\log(x)^2}{2}}$.



Superexponential Tail Decay

Example

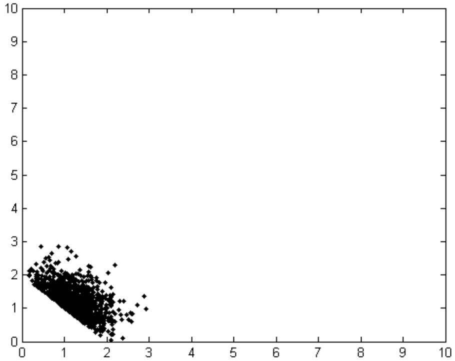
$X_1, X_2 \sim \mathrm{Weib}(2,1)$: $f(x) = 2\,x\,\mathrm{e}^{-x^2}$.



Superexponential Tail Decay

Example

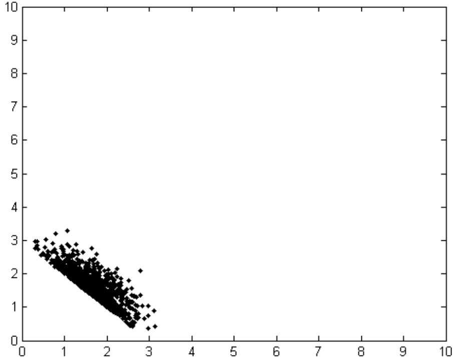
$X_1, X_2 \sim \mathrm{Weib}(2,1)$: $f(x) = 2\,x\,\mathrm{e}^{-x^2}$.



Superexponential Tail Decay

Example

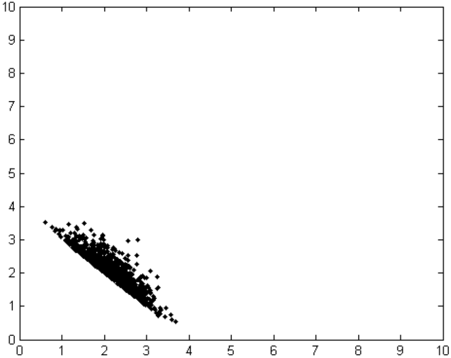
$X_1, X_2 \sim \mathrm{Weib}(2,1)$: $f(x) = 2\,x\,\mathrm{e}^{-x^2}$.



Superexponential Tail Decay

Example

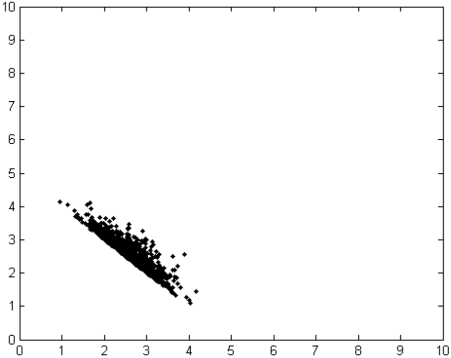
$X_1, X_2 \sim \mathrm{Weib}(2, 1)$: $f(x) = 2\,x\,\mathrm{e}^{-x^2}$.



Superexponential Tail Decay

Example

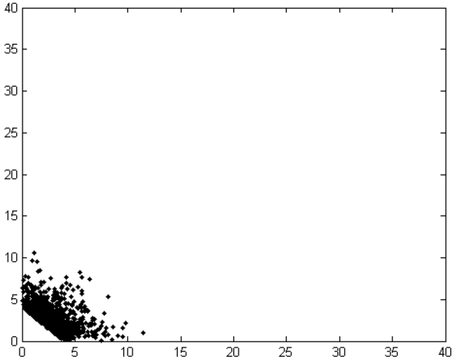
$X_1, X_2 \sim \mathrm{Weib}(2, 1)$: $f(x) = 2\,x\,\mathrm{e}^{-x^2}$.



Exponential Tail Decay

Example

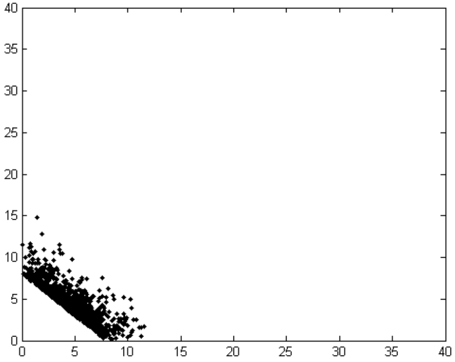
$X_1, X_2 \sim \mathrm{Gamma}(2, 1)$: $f(x) = x\,\mathrm{e}^{-x}$.



Exponential Tail Decay

Example

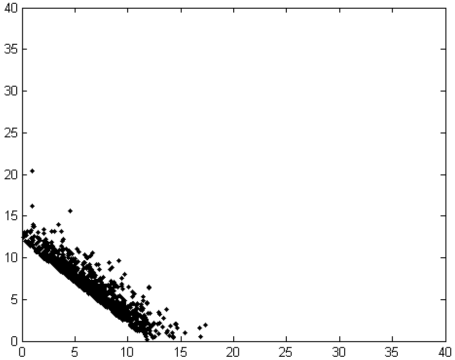
$X_1, X_2 \sim \mathrm{Gamma}(2,1)$: $f(x) = x\,\mathrm{e}^{-x}$.



Exponential Tail Decay

Example

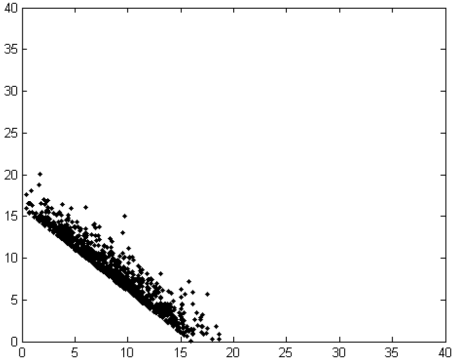
$X_1, X_2 \sim \mathrm{Gamma}(2,1)$: $f(x) = x\,\mathrm{e}^{-x}$.



Exponential Tail Decay

Example

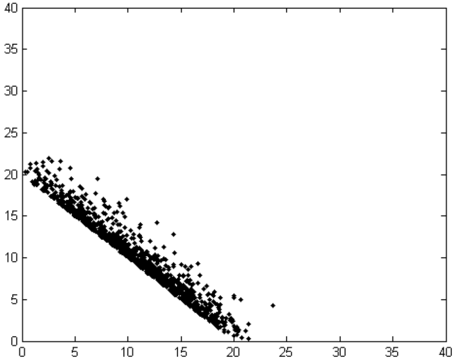
$X_1, X_2 \sim \mathrm{Gamma}(2, 1)$: $f(x) = x\,\mathrm{e}^{-x}$.



Exponential Tail Decay

Example

$X_1, X_2 \sim \mathrm{Gamma}(2, 1)$: $f(x) = x\,\mathrm{e}^{-x}$.



Choosing Angular Approximation $g_{\mathbf{\Theta}|S}$ in Theory

These considerations indicate behaviour of $g_{\mathbf{\Theta}|S}$ when γ is large.

Unfortunately, does not immediately suggest the appropriate functional form for $g_{\mathbf{\Theta}|S}$.

In the remainder, we present a preliminary approach to addressing this.

Select $g_{\mathbf{\Theta}|S}$ from some family of distributions which has the appropriate support.

Choosing Angular Approximation $g_{\Theta|S}$ in Practice

Definition (Dirichlet distribution)

The $\mathrm{Dirichlet}(\boldsymbol{\alpha})$ distribution has

$$f_{\mathrm{Dir}}(\boldsymbol{\theta}) = \frac{\Gamma\left(\sum_{i=1}^{d} \alpha_i\right)}{\Gamma\left(\prod_{i=1}^{d} \alpha_i\right)} \prod_{i=1}^{d} \theta_i^{\alpha_i - 1}, \qquad \boldsymbol{\theta} \in \mathbb{S}^{d-1},$$

as its density, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ is a vector of positive constants. $\diamond$

Choosing Angular Approximation $g_{\boldsymbol{\Theta}|S}$ in Practice

Sample many $(\boldsymbol{X} \mid S > \gamma)$ using MCMC, and find which Dirichlet distribution is the closest fit (VM or CE), by

$$\min_{\boldsymbol{\alpha} \,:\, \boldsymbol{\alpha} \in \mathbb{R}_+^d} \frac{1}{T} \sum_{t=1}^{T} \frac{f_{(S,\boldsymbol{\Theta})}(\boldsymbol{\Theta}^{[t]})}{f_{\mathrm{Dir}}(\boldsymbol{\Theta}^{[t]}; \boldsymbol{\alpha})} \quad \text{or} \quad \max_{\boldsymbol{\alpha} \,:\, \boldsymbol{\alpha} \in \mathbb{R}_+^d} \frac{1}{T} \sum_{t=1}^{T} \log f_{\mathrm{Dir}}(\boldsymbol{\Theta}^{[t]}; \boldsymbol{\alpha})$$

where $\boldsymbol{\Theta}^{[t]} \sim f_{(S,\boldsymbol{\Theta}) | S > \gamma}$.

Note: RHS is just maximum likelihood.

The Algorithm

Algorithm

- $\widehat{\alpha} \leftarrow$ *fitted Dirichlet parameter (from MCMC & optim).*
- *For* $r = 1, \ldots, R$*:*
  1. *Generate* $S^{[r]} \sim \mathfrak{f}_{S \mid S > \gamma}$ *and* $\boldsymbol{\Theta}^{[r]} \sim f_{\mathrm{Dir}}(\cdot\,; \widehat{\alpha})$.
  2. *Compute*

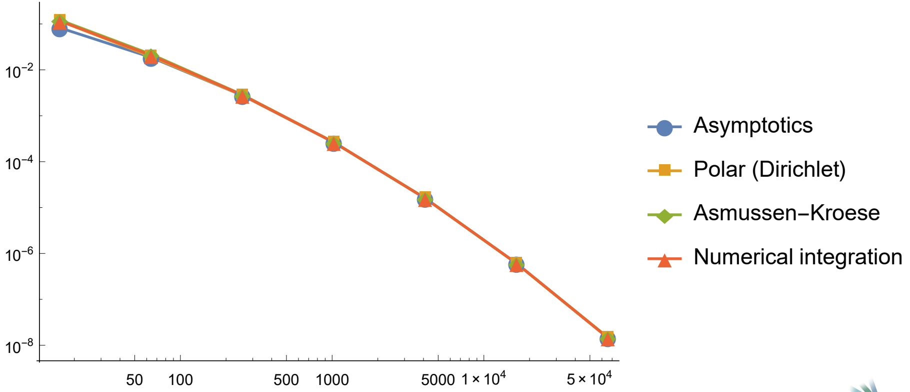
$$\mathcal{R}^{[r]} = \frac{f_{(S,\boldsymbol{\Theta})}(S^{[r]}, \boldsymbol{\Theta}^{[r]})}{c_S \mathfrak{f}_S(S^{[r]})\, f_{\mathrm{Dir}}(\boldsymbol{\Theta}^{[r]}; \widehat{\alpha})}$$

Return

$$\widehat{\ell}_{\mathrm{TP}}(\gamma) := c_S \overline{\mathfrak{F}}_S(\gamma) \times \frac{1}{R} \sum_{r=1}^{R} \mathcal{R}^{[r]} .$$

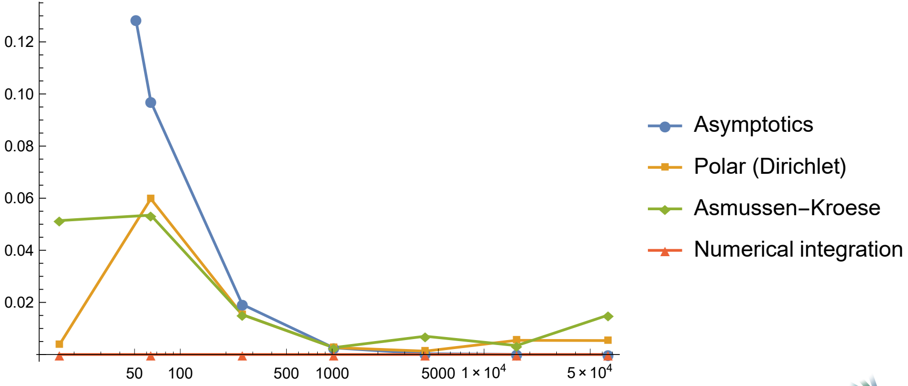
Example: Dependent Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$



Example: Dependent Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$



More stuff I don't have time for

- In some regimes, we show *vanishing relative error* for $\widehat{\ell}_{\mathrm{TP}}$.
- Many numerical experiments, conclusion is $\widehat{\ell}_{\mathrm{TP}} \sim$ Asmussen–Kroese.
- MCMC is a terribly unreliable friend.

Questions for the audience!

- (Part 1) Have you seen this before?
- (Part 2) Any suggestions here?
- (Both) Do these seem useful?

Tak for at lytte!

Supported by: University of Queensland, Aarhus University, ACEMS.

Efficiency conditions for $\widehat{\ell_1}$

An estimator $\widehat{p}(\gamma)$ of $p(\gamma)$ has *bounded relative error (BRE)* if

$$\lim_{\gamma \to \infty} \frac{\mathbb{V}\mathrm{ar}\ \widehat{p}(\gamma)}{p(\gamma)^2} \overset{?}{<} \infty \,.$$

Efficiency conditions for $\widehat{\ell}_1$

An estimator $\widehat{p}(\gamma)$ of $p(\gamma)$ has *bounded relative error (BRE)* if

$$\lim_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\ \widehat{p}(\gamma)}{p(\gamma)^2} \overset{?}{<} \infty\,.$$

We can show that $\widehat{\ell}_1(\gamma)$ has BRE iff

$$\lim_{\gamma\to\infty} \frac{\max_{i\neq j} \mathbb{P}(X_i \geq \gamma, X_j \geq \gamma)}{\max_i \{\mathbb{P}(X_i \geq \gamma)\}^2} < \infty\,,$$

Efficiency conditions for $\widehat{\ell}_1$

An estimator $\widehat{p}(\gamma)$ of $p(\gamma)$ has *bounded relative error (BRE)* if

$$\lim_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\ \widehat{p}(\gamma)}{p(\gamma)^2} \overset{?}{<} \infty \,.$$

We can show that $\widehat{\ell}_1(\gamma)$ has BRE iff

$$\lim_{\gamma\to\infty} \frac{\max_{i\neq j} \mathbb{P}(X_i \geq \gamma, X_j \geq \gamma)}{\max_i \{\mathbb{P}(X_i \geq \gamma)\}^2} < \infty \,,$$

that is,

$$\max_{i\neq j} \mathbb{P}(X_i \geq \gamma, X_j \geq \gamma) = \mathcal{O}(\max_i \{\mathbb{P}(X_i \geq \gamma)\}^2) \quad \text{as } \gamma \to \infty \,.$$

Correlation in the limit

Assume common marginals. Most common measure is the *coefficient of asymptotic upper tail dependence*

$$\lambda(X_i \,|\, X_j) := \lim_{\gamma \to \infty} \mathbb{P}(X_i \geq \gamma \,|\, X_j \geq \gamma)\,.$$

A finer measure is *residual tail dependence*. It states that[1] as $\gamma \to \infty$

$$\mathbb{P}(X_i \geq \gamma, X_j \geq \gamma) \;\sim\; \mathscr{L}(\gamma)\,\gamma^{-\frac{1}{\eta}}$$

for some slowly-varying function $\mathscr{L}(\cdot)$ and $\eta \in [0, 2]$.

[1] Given we transform to unit Fréchet marginal distributions.

Efficiency conditions for $\widehat{\ell}_1$

Must have $\forall\ i \neq j$ that $\lambda(X_i \mid X_j) = 0$, but this isn't strong enough. Say

$$\frac{\mathbb{P}(X_i \geq \gamma, X_j \geq \gamma)}{\mathbb{P}(X_k \geq \gamma)^2} \sim \frac{\mathscr{L}(\gamma)\, \gamma^{-\frac{1}{\eta}}}{\gamma^{-2}} = \mathscr{L}(\gamma)\, \gamma^{2-\frac{1}{\eta}}\,,$$

then the condition for $\widehat{\ell}_1$ having BRE becomes

$$\limsup_{\gamma\to\infty} \mathscr{L}(\gamma)\, \gamma^{2-\frac{1}{\eta}} < \infty \Leftrightarrow \eta \in [0, \frac{1}{2}) \text{ or } (\eta = \frac{1}{2} \text{ and } \mathscr{L}(\gamma) \to 0)\,.$$

Copulas for which $\widehat{\ell}_1$ has BRE

This is a subset of Table 1 of Heffernan (2000):

#	**Name**	η	$L(x)$
1	Ali-Mikhail-Haq	0.5	$1+\alpha$
2	BB10 in Joe	0.5	$1+\theta/\alpha$
3	Frank	0.5	$\delta/(1-\mathrm{e}^{-\delta})$
4	Morgenstern	0.5	$1+\alpha$
5	Plackett	0.5	δ
6	Crowder	0.5	$1+(\theta-1)/\alpha$
7	BB2 in Joe	0.5	$\theta(\delta+1)+1$
8	Pareto	0.5	$1+\delta$
9	Raftery	0.5	$\delta/(1-\delta)$
10	Gaussian ($\rho \le 0$)	$\frac{1+\rho}{2}$	$C_\rho(\log t)^{-\frac{\rho}{1+\rho}}$

Table: Copulas with BRE.

Choosing Radial Approximation g_S

In many cases, know asymptotics

$$\ell(\gamma) = (1 + \mathrm{o}(1))\, c_S \overline{\mathfrak{F}}_S(\gamma)\,,$$

where $\overline{\mathfrak{F}}_S$ is the complementary cdf of a random variable with density $\mathfrak{f}_S$.

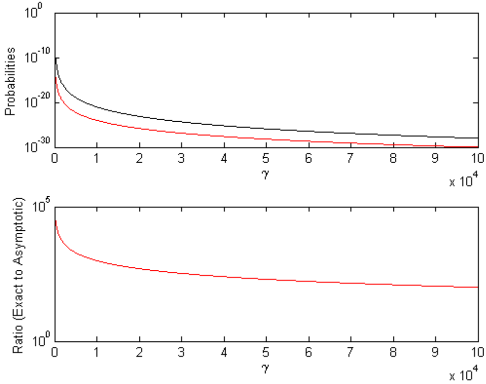
Tempting to use (deterministic) estimator $\widehat{\ell}_{\mathrm{Asym}}(\gamma) := c_S \overline{\mathfrak{F}}_S(\gamma)$.

Can be wildly inaccurate for moderate γ.

Choosing Radial Approximation g_S

Example ($f_i(x) = \lambda_i \alpha_i (1 + \lambda_i x)^{-(\alpha_i+1)}$, $i = 1, 2.$)

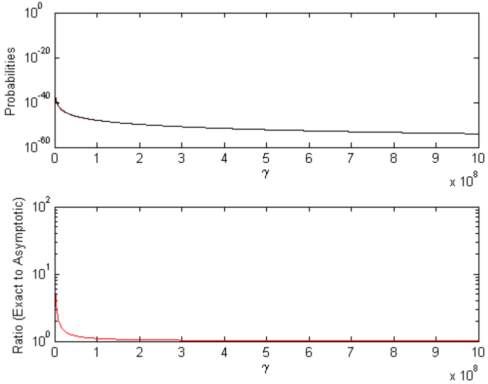
E.g. $\alpha_1 = 6$, $\lambda_1 = 1$, $\alpha_2 = 7$, and $\lambda_2 = 0.1$; $\widehat{\ell}_{\text{Asym}}(\gamma) = (1+\gamma)^{-6}$.



Choosing Radial Approximation g_S

Example ($f_i(x) = \lambda_i \alpha_i (1 + \lambda_i x)^{-(\alpha_i+1)}$, $i = 1, 2$.)

E.g. $\alpha_1 = 6$, $\lambda_1 = 1$, $\alpha_2 = 7$, and $\lambda_2 = 0.1$; $\widehat{\ell}_{\mathrm{Asym}}(\gamma) = (1+\gamma)^{-6}$.



Efficiency Criteria

Definition

An estimator $\widehat{p}(\gamma)$ of some real probability $p(\gamma)$ which satisfies $\forall \varepsilon > 0$

$$\limsup_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{p}(\gamma)}{p(\gamma)^{2-\varepsilon}} = 0 \quad (2a) \qquad \limsup_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{p}(\gamma)}{p(\gamma)^{2}} < \infty \quad (2b) \qquad \limsup_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{p}(\gamma)}{p(\gamma)^{2}} = 0 \quad (2c)$$

has *logarithmic efficiency* (2a), *bounded relative error/strong efficiency* (2b), or *vanishing relative error* (2c) respectively.

Efficiency of the Estimator $\widehat{\ell}_{\mathrm{TP}}(\gamma)$

Define $\mathbb{S}^{d-1} := \{\boldsymbol{\theta} \in \mathbb{R}_+^d \ : \ \boldsymbol{\theta}^\top \mathbf{1} = 1\}$.

Assumption (Assumption 1)

$$\limsup_{\gamma\to\infty} \sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \frac{f_{\boldsymbol{\Theta}|S}(\boldsymbol{\theta}|\gamma)}{g_{\boldsymbol{\Theta}|S}(\boldsymbol{\theta}|\gamma)} \le K \,,$$

for some $1 \le K < \infty$.

Proposition

Under Assumption 1, the estimator $\widehat{\ell}_{\mathrm{TP}}(\gamma)$ *has bounded relative error as* $\gamma \to \infty$*. If* $K = 1$ *in Assumption 1, the estimator* $\widehat{\ell}_{\mathrm{TP}}(\gamma)$ *has vanishing relative error as* $\gamma \to \infty$.

Some Copulas and Their Impact on $c_S \overline{\mathfrak{F}}_S(\gamma)$

Proposition (Asymptotic right-tail independence)

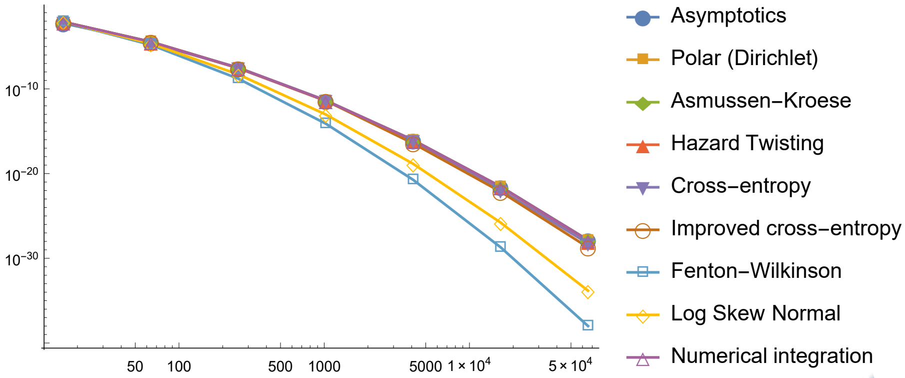
Clayton, Ali–Mikhail–Haq, Frank, Farlie–Gumbel–Morgenstern, Gaussian (all pairwise correlations $\notin \{-1, 1\}$).

Proposition (Asymptotic right-tail dependence)

Joe, Gumbel–Hougaard, t, Gaussian (all pairwise correlations $\in \{-1, 1\}$).

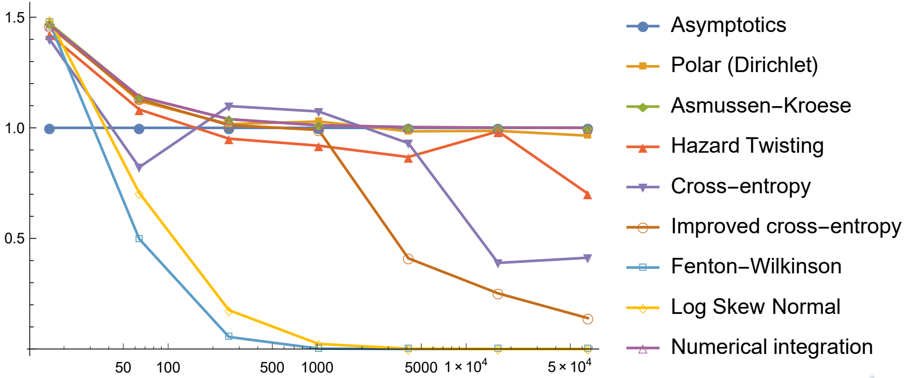
Example: Iid Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



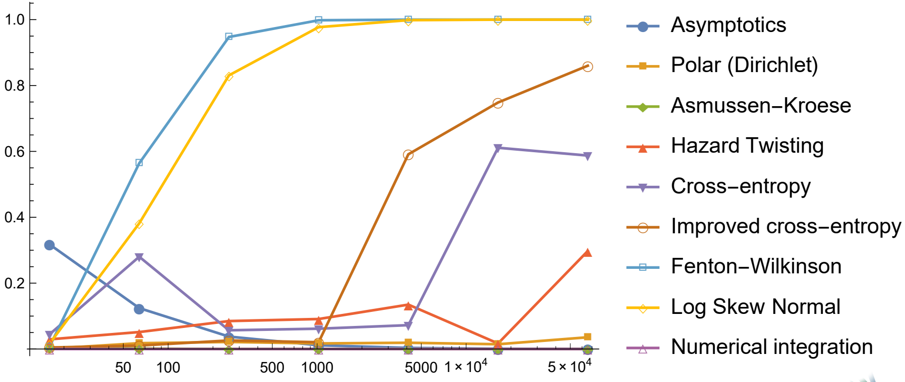
Example: Iid Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



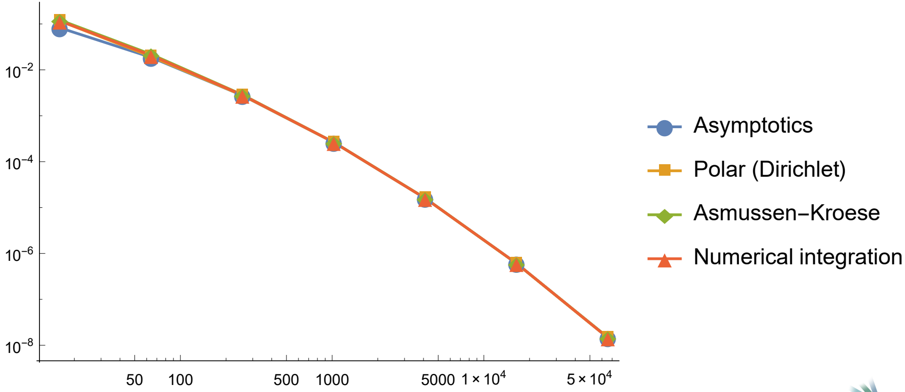
Example: Iid Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



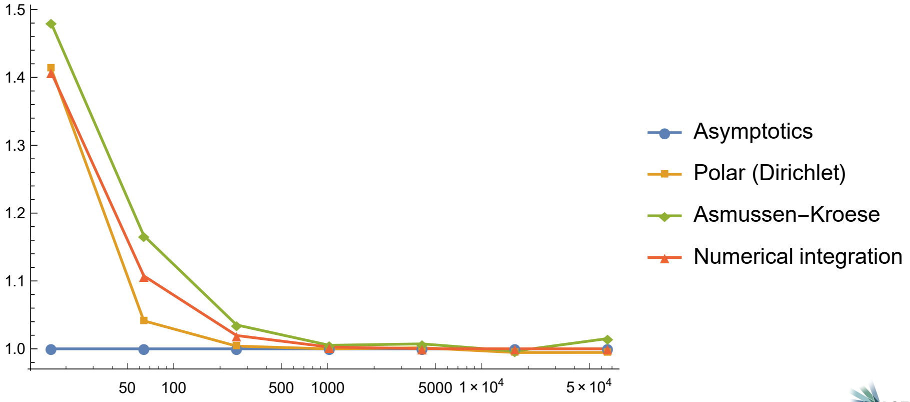
Example: Dependent Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$



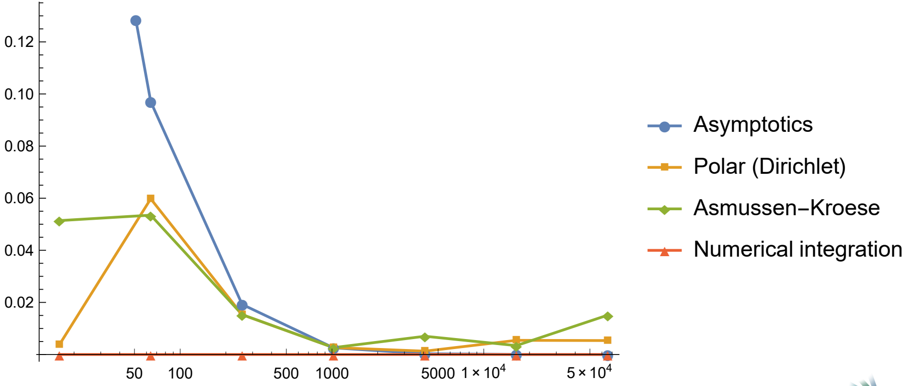
Example: Dependent Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$



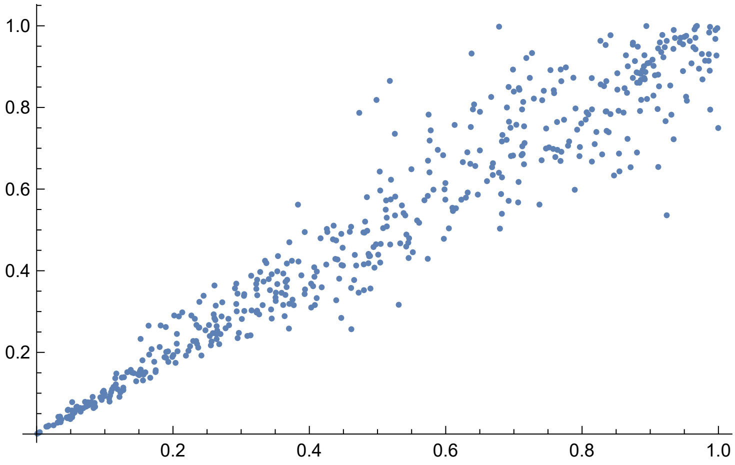
Example: Dependent Lognormals ($d = 2$)

$$\mu = \mathbf{0},\ \Sigma = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$



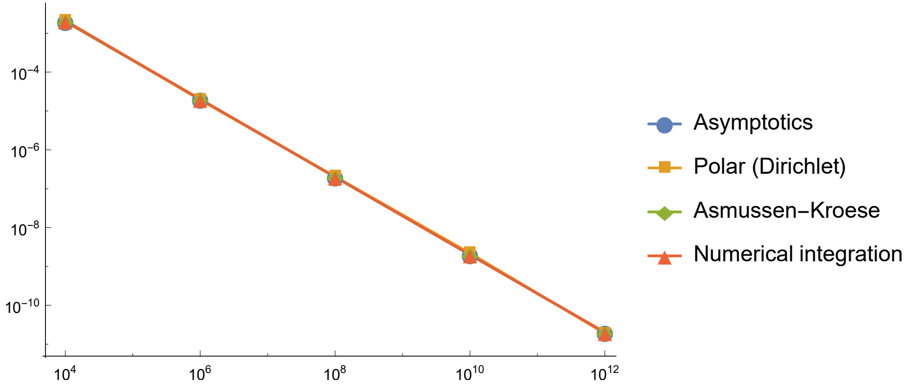
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Clayton copula.



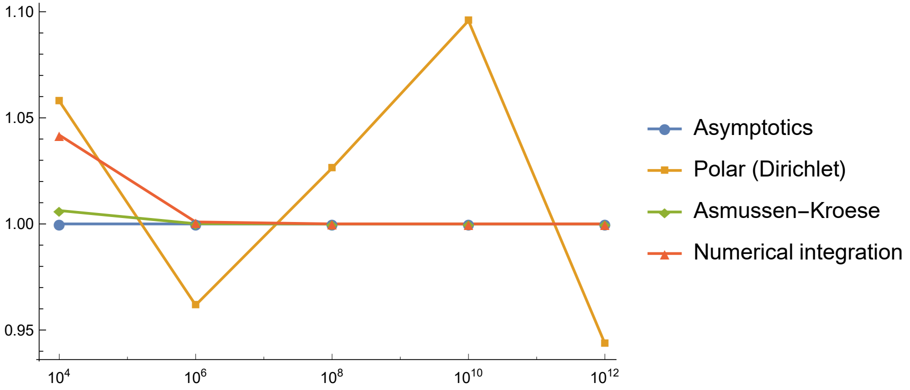
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Clayton copula.



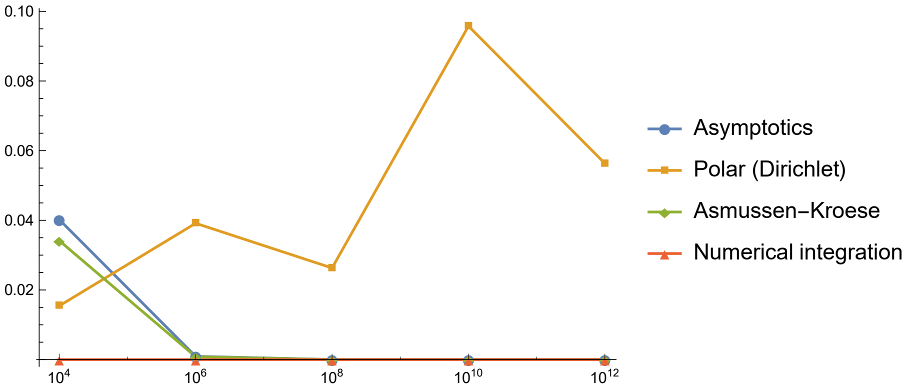
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Clayton copula.



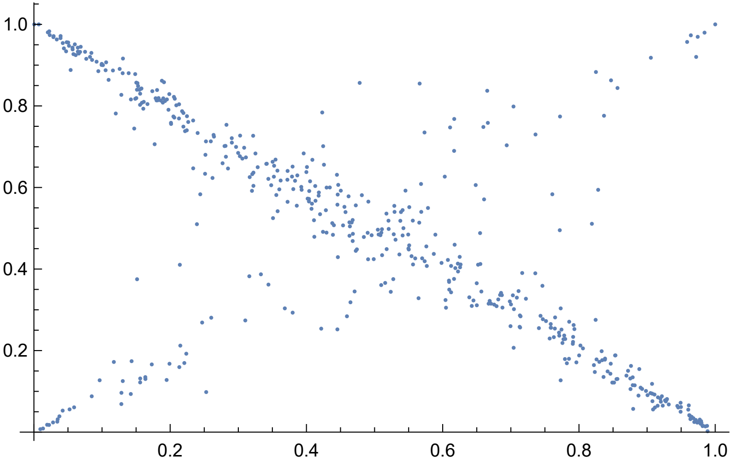
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Clayton copula.



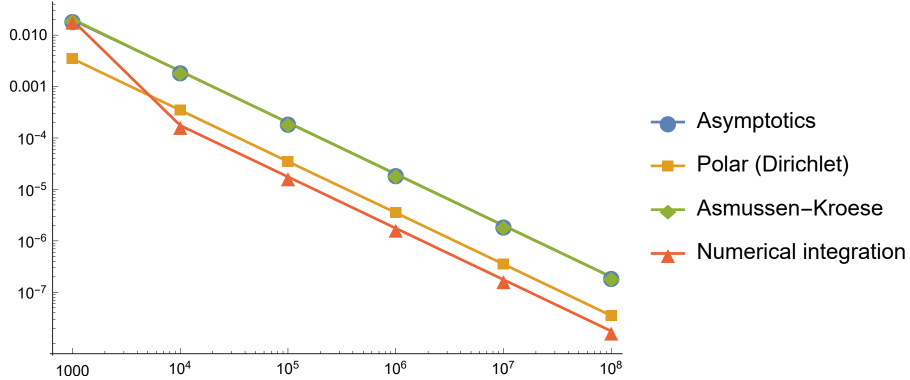
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Multivariate t copula.



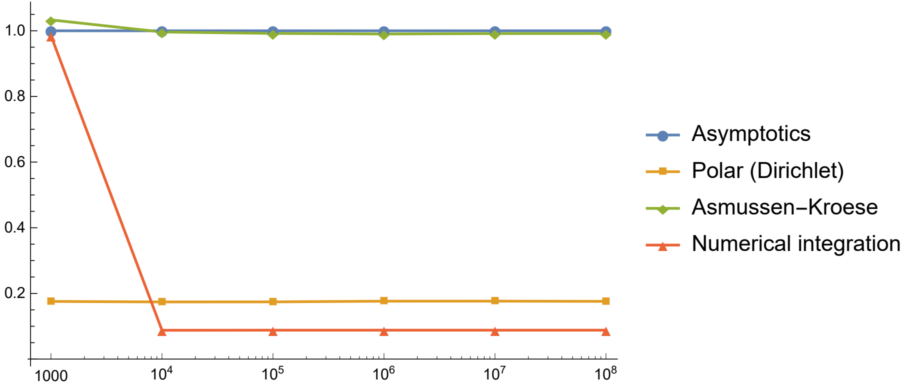
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Multivariate t copula.



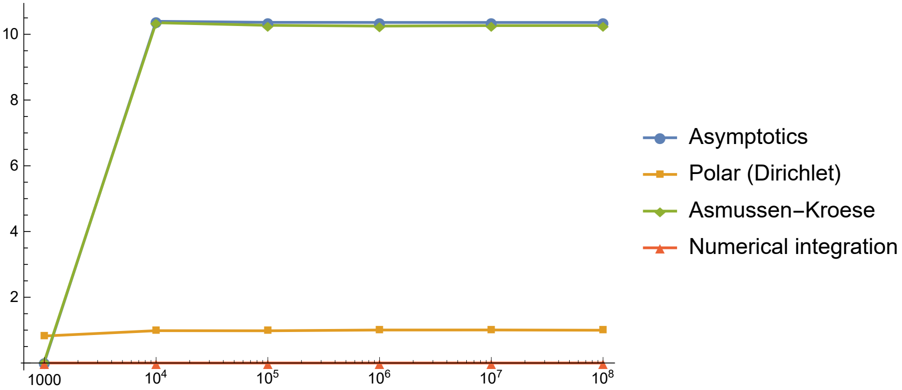
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Multivariate t copula.



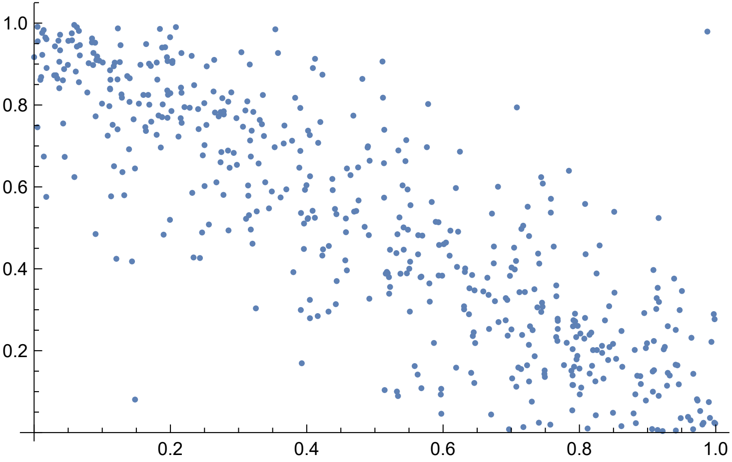
Example: Dependent Paretos ($d = 2$)

$\alpha = 10$. Multivariate t copula.



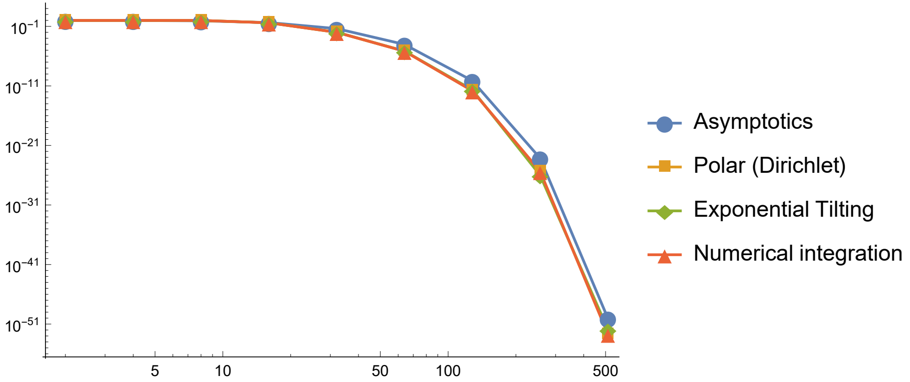
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Frank copula.



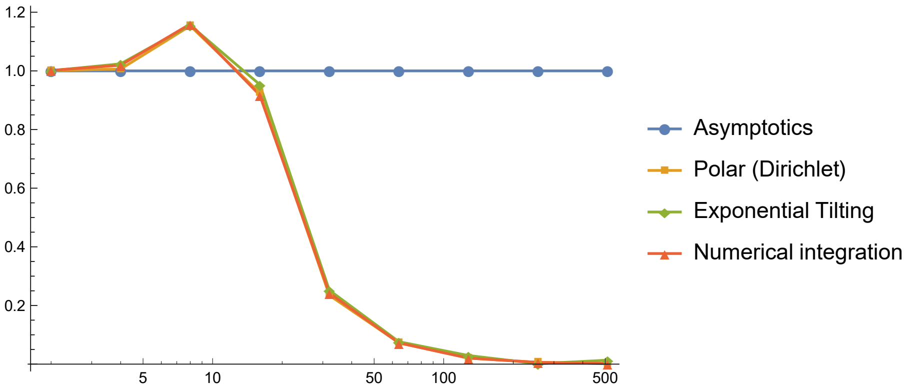
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Frank copula.



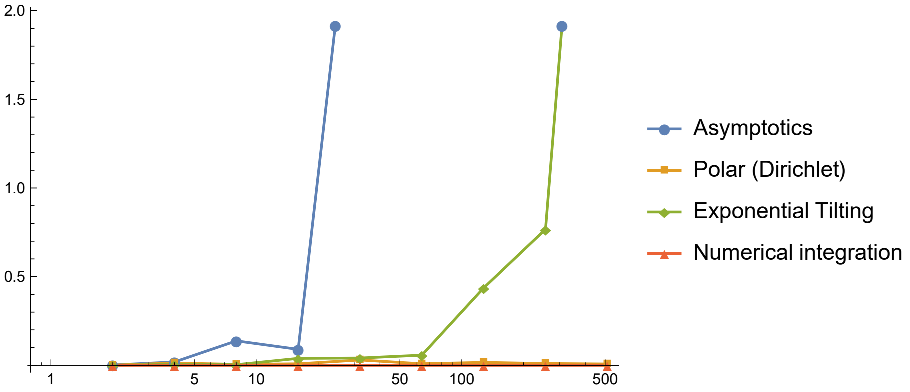
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Frank copula.



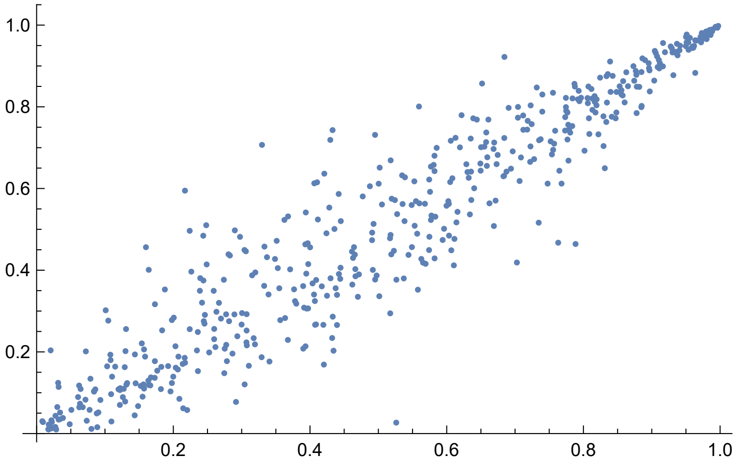
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4.$ Frank copula.



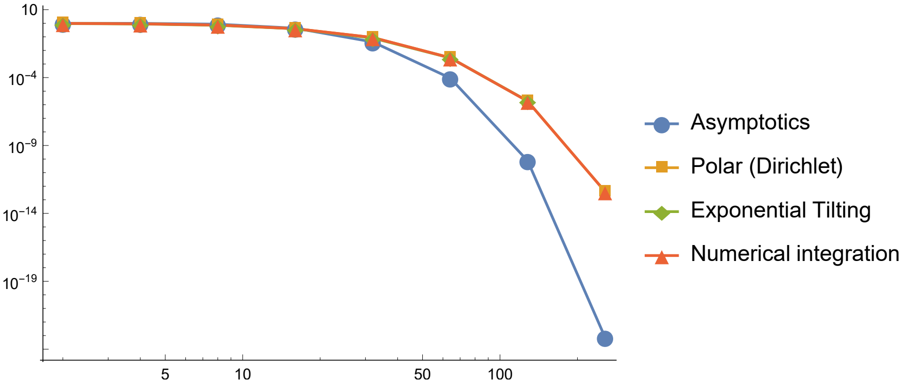
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Gumbel–Hougaard copula.



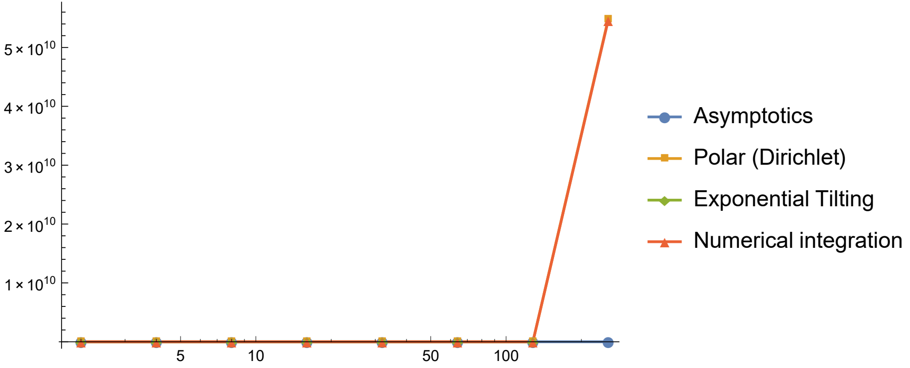
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Gumbel–Hougaard copula.



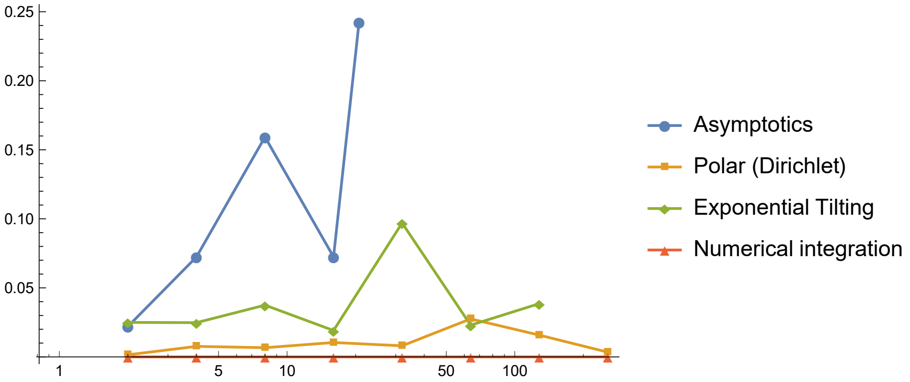
Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Gumbel–Hougaard copula.



Example: Dependent Gammas ($d = 2$)

$\alpha = 2, \lambda = 4$. Gumbel–Hougaard copula.



Discussion

Asymptotic form is assumed for $\widehat{\ell}_{\mathrm{TP}}(\gamma)$; however, reasonable practical performance should be achievable with "educated guess" in $\widehat{\ell}_{\mathrm{Gen}}(\gamma)$.

More precise asymptotics for $\mathbf{\Theta} | S > \gamma$?

Extensions: e.g. $S = \max\{X_1, \ldots, X_d\}$, $X_i \in \mathbb{R}$.