Approximating failure probabilities for multivariate extremes

Anna Kiriliouk [1]   Chen Zhou [2]

[1]University of Namur, Belgium
[2]Erasmus University Rotterdam, the Netherlands

Workshop: combining causal inference and extreme value theory in the study of climate extremes and their causes

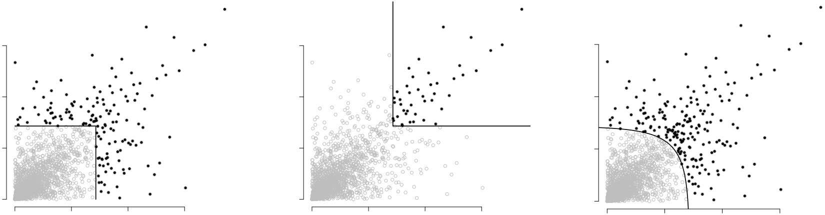
June 29th, 2022

Outline

Failure probabilities for multivariate extremes

Goal: estimating the failure probability $\mathbb{P}[\boldsymbol{X} \in C]$

- for $\boldsymbol{X} \in [0, \infty)^d$ with d "large";
- C a so-called failure region;
- based on n iid observations;
- using the framework of multivariate regular variation.



Example: weighted sums of components

- For example,
$$C_{\text{sum}} = \{\boldsymbol{y} \in [0,\infty)^d : \boldsymbol{v}^T \boldsymbol{y} > x\}$$
for $x > 0$ large and $\boldsymbol{v} = (v_1, \ldots, v_d) > \mathbf{0}$ a vector of weights such that $v_1 + \ldots + v_d = 1$.

- Failure probability $\mathbb{P}[\boldsymbol{X} \in C] = \mathbb{P}[\boldsymbol{v}^T \boldsymbol{X} > x]$.
  - Finance: Value-at-Risk or Expected Shortfall of a portfolio loss representing aggregated stock returns.
  - Flood risk management: aggregated precipitation (spatially and/or temporally) to estimate the risk of a flood occurring because of prolonged extreme rain.

- Here, it is not possible to separate marginal and dependence structure modeling.

Multivariate regular variation framework (Resnick, 2007)

- Suppose that $\boldsymbol{X}$ is multivariate regularly varying; there exists a sequence $b_n \to \infty$ and a limit measure $\nu_{\boldsymbol{X}}$ such that

$$n\,\mathbb{P}\left[b_n^{-1}\boldsymbol{X} \in \cdot\,\right] \xrightarrow{v} \nu_{\boldsymbol{X}}(\,\cdot\,), \qquad \text{as } n \to \infty,$$

  where $\xrightarrow{v}$ denotes vague convergence on $\mathbb{E}_0 = [0,\infty]^d \setminus \{\mathbf{0}\}$.
- The limit measure $\nu_{\boldsymbol{X}}$ is the exponent measure.
- Homogeneity: there exists an $\alpha > 0$, the tail index of $\boldsymbol{X}$, such that

$$\nu_{\boldsymbol{X}}(tC) = t^{-\alpha}\nu_{\boldsymbol{X}}(C), \qquad t > 0, C \subset \mathbb{E}_0.$$

- For large n,

$$\mathbb{P}\left[\boldsymbol{X} \in C\right] = \frac{1}{n}\left\{n\,\mathbb{P}\left[b_n^{-1}\boldsymbol{X} \in b_n^{-1}C\right]\right\} \approx \frac{1}{n}\nu_{\boldsymbol{X}}\left(b_n^{-1}C\right).$$

Multivariate regular variation framework (Resnick, 2007)

- Let $\|\cdot\|$ denote a norm on $\mathbb{R}^d$ and consider

$$(R, \boldsymbol{W}) = \left(\|\boldsymbol{X}\|, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \right).$$

- Multivariate regular variation is equivalent to

$$n\mathbb{P}\left[\left(b_n^{-1} R, \boldsymbol{W} \right) \in \cdot \right] \xrightarrow{v} \mu_\alpha \times H_{\boldsymbol{X}}(\cdot), \qquad \text{as } n \to \infty,$$

- The measure $H_{\boldsymbol{X}}$ on $\mathbb{S}_{d-1} := \{\boldsymbol{w} \in \mathbb{E}_0 : \|\boldsymbol{w}\| = 1\}$ is called the angular measure and μ_α is given by $\mu_\alpha((x, \infty]) = x^{-\alpha}$ for $x > 0$.
- We have, for $B \subset \mathbb{S}_{d-1}$,

$$\nu_{\boldsymbol{X}}(\{\boldsymbol{x} \in \mathbb{E}_0 : \|\boldsymbol{x}\| > s, \boldsymbol{x}/\|\boldsymbol{x}\| \in B\}) = s^{-\alpha} H_{\boldsymbol{X}}(B)$$

and

$$\nu_{\boldsymbol{X}}(\mathrm{d}r \times \mathrm{d}\boldsymbol{w}) = \alpha r^{-\alpha-1} \, \mathrm{d}r \, \mathrm{d}H_{\boldsymbol{X}}(\boldsymbol{w})$$

Pairwise tail dependence measures

- Parametric models for $\nu_{\boldsymbol{X}}$ or $H_{\boldsymbol{X}}$ are numerous.
- Model fit is often assessed by checking if the model-implied pairwise dependence matches the non-parametrically estimated pairwise dependence.
- Goal: build a parametric model by directly matching a pairwise dependence measure.
- The marginal variable X_j satisfies $(j = 1, \ldots, d)$

$$\lim_{n\to\infty} n\,\mathbb{P}\left[X_j > b_n x\right] = x^{-\alpha} \int_{\mathbb{S}_{d-1}} w_j^{\alpha}\,\mathrm{d}H_{\boldsymbol{X}}(\boldsymbol{w}) =: x^{-\alpha}\sigma_{jj}.$$

- If we consider $\sqrt{X_j X_k}$ (for $j, k = 1, \ldots, d$),

$$\lim_{n\to\infty} n\,\mathbb{P}[\sqrt{X_j X_k} > b_n x] = x^{-\alpha} \int_{\mathbb{S}_{d-1}} w_j^{\alpha/2} w_k^{\alpha/2}\,\mathrm{d}H_{\boldsymbol{X}}(\boldsymbol{w}) =: x^{-\alpha}\sigma_{jk}.$$

Tail pairwise dependence matrix

- Larsson and Resnick (2012) introduced the tail pairwise dependence matrix (TPDM) $\Sigma_{\boldsymbol{X}}$ of $\boldsymbol{X}$,

$$\Sigma_{\boldsymbol{X}} = (\sigma_{jk})_{j,k=1,\ldots,d}, \quad \text{with} \quad \sigma_{jk} = \int_{\mathbb{S}_{d-1}} w_j^{\alpha/2} w_k^{\alpha/2} \, \mathrm{d}H_{\boldsymbol{X}}(\boldsymbol{w}).$$

- The TPDM has positive entries only and is positive semi-definite (Cooley and Thibaud, 2019).
- Two variables X_j, X_k are tail dependent if and only if $\sigma_{jk} > 0$.
- Let $\|\cdot\|$ denote the L_α norm. Then the total mass of the spectral measure equals

$$H_{\boldsymbol{X}}(\mathbb{S}_{d-1}) = \sum_{j=1}^{d} \sigma_{jj} = \mathrm{tr}(\Sigma_{\boldsymbol{X}}).$$

The max-linear model

- In a max-linear model, each component of a d-dimensional vector $\boldsymbol{Y}$ can be interpreted as the maximum shock among a set of q independent heavy-tailed factors.
  - Let $A = (a_{il})$ denote a $d \times q$ matrix with non-negative entries
  - Let $Z_1, \ldots, Z_q$ be independent Fréchet(α) random variables,

$$\boldsymbol{Y} := A \times_{\max} \boldsymbol{Z} := \left(\max_{l=1,\ldots,q} a_{1l} Z_l, \ldots, \max_{l=1,\ldots,q} a_{dl} Z_l \right)^T.$$

- Fougeres et al. (2013) showed that the max-linear model is dense in the class of d-dimensional multivariate extreme-value distributions.
- The max-linear model is used in Gissibl and Klüppelberg (2018), Cui and Zhang (2018), Einmahl et al. (2018) and Janßen and Wan (2020), among others.

The TPDM of a max-linear model

- The spectral measure of $\boldsymbol{Y}$ is (under the L_α norm)

$$H_{\boldsymbol{Y}}(\,\cdot\,) = \sum_{l=1}^{q} \|\boldsymbol{a}_l\|^\alpha \, \delta_{\boldsymbol{a}_l/\|\boldsymbol{a}_l\|}(\,\cdot\,),$$

  where $\boldsymbol{a}_l$ is the l-th column of A.
- As noticed in Cooley and Thibaud (2019), the TPDM of $\boldsymbol{Y}$ has elements

$$\sigma_{jk} = \sum_{l=1}^{q} a_{jl}^{\alpha/2} a_{kl}^{\alpha/2}$$

- In other words,

$$\Sigma_{\boldsymbol{Y}} = A_* A_*^T, \qquad \text{where } A_* := \left(a_{jk}^{\alpha/2}\right)_{j,k=1,\ldots,d}$$

Max-linear model and failure probabilities

- The failure region

$$C_{\max}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{E}_0 : y_1 > x_1 \text{ or } \ \dots \ \text{ or } y_d > x_d\}$$

has exponent measure

$$\nu(C_{\max}(\boldsymbol{x})) = \sum_{l=1}^{q} \max_{j=1,\dots,d} \left(\frac{a_{jl}}{x_j}\right)^{\alpha}.$$

- The failure region

$$C_{\text{sum}}(\boldsymbol{v}, x) = \{\boldsymbol{y} \in \mathbb{E}_0 : v_1 y_1 + \ldots + v_d y_d > x\}.$$

has exponent measure

$$\nu(C_{\text{sum}}(\boldsymbol{v}, x)) = x^{-\alpha} \sum_{l=1}^{q} \left(\boldsymbol{v}^T \boldsymbol{a}_l\right)^{\alpha}.$$

The max-linear model and the TPDM

- Can we construct a max-linear model $\boldsymbol{Y}$ such that $\Sigma_{\boldsymbol{Y}}$ matches the (estimated) $\Sigma_{\boldsymbol{X}}$?
- Cooley and Thibaud (2019) show that
  - As $q \to \infty$, the class of max-linear angular measures is dense in the class of possible angular measures.
  - If attention is restricted to the TPDM, a max-linear model with finite q is sufficient to exactly match $\Sigma_{\boldsymbol{X}}$.
- For any (estimated) TPDM, we could construct a max-linear model with the same TPDM.
- The coefficient matrix of the max-linear model is obtained through a completely positive decomposition of $\Sigma_{\boldsymbol{X}}$.
  - A matrix Σ is completely positive if it can be decomposed as $\Sigma = AA^T$, where the matrix A has non-negative entries.
- The algorithm to find completely positive decompositions is complicated; theoretically, $q \leq d(d+1)/2 + 4$.

An alternative goal: finding an approximate decomposition

- Any symmetric positive semi-definite matrix can be decomposed as $\Sigma = AA^T$ through the Cholesky decomposition:
  - the matrix A may contain negative elements;
  - A is a $d \times d$ lower-triangular matrix;
  - the decomposition is not unique.
- We search for an approximate completely positive decomposition
  - with $q = d$;
  - not necessarily matching all elements in Σ;
  - target: a lower-triangular matrix A.
- For this presentation, consider $\alpha = 2$; then $A = A_*$.
- Write

$$A = \begin{pmatrix} a_{11} & \mathbf{0}^T \\ \boldsymbol{a}_{-1} & A^{(-1,-1)} \end{pmatrix},$$

where $A^{(-1,-1)}$ is a $(d-1) \times (d-1)$ lower-triangular matrix with non-negative elements.

Matching AA^T and $\Sigma_{\boldsymbol{X}}$

- We can calculate the TPDM $\Sigma_{\boldsymbol{X}}$ as

$$AA^T = \begin{pmatrix} a_{11}^2 & a_{11}(\boldsymbol{a}_{-1})^T \\ a_{11}(\boldsymbol{a}_{-1}) & \boldsymbol{a}_{-1}\boldsymbol{a}_{-1}^T + A^{(-1,-1)}(A^{(-1,-1)})^T \end{pmatrix} = (\sigma_{jk})$$

- Hence, we obtain
  - a_{11} as $\sqrt{\sigma_{11}}$;
  - $\boldsymbol{a}_{-1}$ from $\sigma_{j1}/\sqrt{\sigma_{11}}$ for $j = 2, \cdots, d$;
  - the TPDM of the other $(d-1)$ dimensions by taking $\Sigma_{\boldsymbol{X}}^{(-1,-1)}$ and subtracting $\boldsymbol{a}_{-1}(\boldsymbol{a}_{-1})^T$
- Can we do that?
  - $A^{(-1,-1)}(A^{(-1,-1)})^T$ is also a completely positive matrix;
  - It implies that for $j, k \neq 1$

$$\sigma_{jk} \geq a_{j1}a_{k1} \quad \Rightarrow \quad \sigma_{jk}\sigma_{11} \geq \sigma_{j1}\sigma_{k1}.$$

  - This is a necessary condition to perform the above "algorithm".

A criterion for the reverse algorithm

- The reverse algorithm works only if for all $j, k \in \{2, \ldots, d\}$,

$$\frac{\sigma_{j1}\sigma_{k1}}{\sigma_{jk}\sigma_{11}} \leq 1$$

- What if this does not hold for some TPDM $\Sigma_{\boldsymbol{X}}$?
- Define

$$D_1(\Sigma_{\boldsymbol{X}}) := \max\left\{j, k \in \{2, \ldots, d\} : \frac{\sigma_{j1}\sigma_{k1}}{\sigma_{jk}\sigma_{11}}\right\}.$$

  - If $D_1(\Sigma_{\boldsymbol{X}}) \leq 1$ the algorithm works;
  - If $D_1(\Sigma_{\boldsymbol{X}}) > 1$ we still have

$$\sigma_{jk}\sigma_{11} \geq \frac{\sigma_{j1}\sigma_{k1}}{D_1(\Sigma_{\boldsymbol{X}})}.$$

- We can take $a_{11} = \sqrt{D_1(\Sigma_{\boldsymbol{X}})\sigma_{11}}$ and do the algorithm, at the cost that $a_{11}^2 > \sigma_{11}$.

An approximate decomposition algorithm

- Define for any $i \in \{1, \dots, d\}$,

$$D_i(\Sigma_{\boldsymbol{X}}) := \max\left\{ j, k \in \{1, \dots, d\} \setminus \{i\} : \frac{\sigma_{ji}\sigma_{ki}}{\sigma_{jk}\sigma_{ii}} \right\}.$$

- Let $\boldsymbol{\tau}_i = (\tau_{1,i}, \dots, \tau_{d,i})^T$ with

$$\tau_{j,i} = \begin{cases} \sigma_{ji} \left(\sigma_{ii} \max(D_i, 1)\right)^{-1/2} & \text{if } j \neq i, \\ \left(\sigma_{ii} \max(D_i, 1)\right)^{1/2} & \text{if } j = i. \end{cases}$$

- Let $\boldsymbol{\tau}_{-i} := (\tau_{1,i}, \dots, \tau_{i-1,i}, \tau_{i+1,i}, \dots, \tau_{d,i})^T \in \mathbb{R}^{d-1}$ and

$$\Sigma_{\boldsymbol{X}}^{(i)} := \Sigma_{\boldsymbol{X}}^{(-i,-i)} - \boldsymbol{\tau}_{-i}\boldsymbol{\tau}_{-i}^T \quad \in \mathbb{R}^{(d-1)\times(d-1)}.$$

Proposition

For all $i \in \{1, \dots, d\}$, the matrix $\Sigma_{\boldsymbol{X}}^{(i)}$ is a TPDM since

1. $\Sigma_{\boldsymbol{X}}^{(i)} \geq 0$ *component-wise;*
2. $\Sigma_{\boldsymbol{X}}^{(i)}$ *is positive semi-definite.*

An approximate decomposition algorithm

- The proposition holds for all TPDMs, not only those obtained through a triangular matrix.
- Let $\Sigma_{\boldsymbol{X}}$ be an (estimated) TPDM and let $i_1 \mapsto i_2 \mapsto \ldots \mapsto i_d$ denote a path, where $(i_1, \ldots, i_d)$ is a permutation of $(1, \ldots, d)$.
  - determines which column will be treated first.
- The iterative algorithm:
  - Obtain the vector $\boldsymbol{\tau}_{i_1}$ by taking $i = i_1$ and fill in the first column of the matrix A with $\boldsymbol{\tau}_{i_1}$.
  - The targeted TPDM is then reduced to a $(d-1) \times (d-1)$ matrix.
  - In step j, fill the j-th column of A by applying the algorithm to the targeted $(d-j+1) \times (d-j+1)$ TPDM matrix.
    - Set the elements in the $i_1, i_2, \cdots, i_{j-1}$-th row to zero.
    - Fill in the other $(d-j+1)$ elements by the $(d-j+1)$-dimensional $\boldsymbol{\tau}$ vector obtained in this step.
    - The dimension of the targeted TPDM is reduced by one.
- Then A satisfies $\sigma_{\boldsymbol{X}_{jk}} = [AA^T]_{jk}$ for $j \neq k$ and $\sigma_{\boldsymbol{X}_{jj}} \leq [AA^T]_{jj}$.

Choosing an optimal path

- If $\boldsymbol{X}$ follows a max-linear model constructed from a lower-triangular parameter matrix A, by choosing the path $1 \mapsto 2 \mapsto \ldots \mapsto d$, the exact max-linear model is recovered.
- In general, smart path choices can lead to exact decompositions!
  - A simple approach: pick the lowest value of D_i in each step;
  - An exhaustive approach: build a "tree" of possibilities;
  - A pragmatic approach: in each step, pick a random "branch" in the tree, until an end "leaf". If the procedure stops with less than d steps, restart the entire procedure from the beginning.
- In practice:
  - For small d ($d \leq 20$), we can obtain thousands of exact decompositions in half an hour.
  - For moderate d ($d \leq 40$), we still find a moderate to large number of exact decompositions.
  - For high d (e.g. $d = 150$), it is rare to find exact decompositions, but approximate ones may be quite satisfactory.

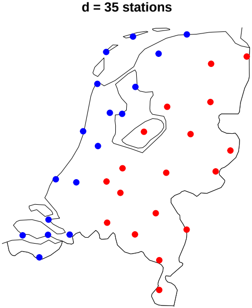
Estimation of the TPDM

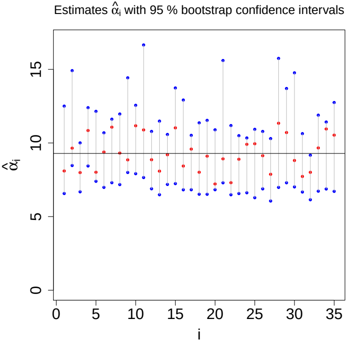
- Suppose (i.i.d.) data $\boldsymbol{X}_i$ are available, for $i = 1, 2, \ldots, n$.
- We first need to estimate:
  - the tail index α;
  - the mass of the angular measure $m = H_{\boldsymbol{X}}(\mathbb{S}_{d-1})$.
- For $i = 1, \ldots, n$, let $R_i = \|\boldsymbol{X}_i\|$ and $\boldsymbol{W}_i = \boldsymbol{X}_i / R_i$.
- Let r_0 be a high quantile of the empirical distribution of $R_1, \ldots, R_n$.
- Write $n_{\text{exc}} = \sum_{i=1}^{n} \mathbb{1}\{R_i > r_0\}$.
- Then σ_{jk} can be estimated by

$$\widehat{\sigma}_{jk} = \frac{\widehat{m}}{n_{\text{exc}}} \sum_{i=1}^{n} W_{ij}^{\widehat{\alpha}/2} W_{ik}^{\widehat{\alpha}/2} \mathbb{1}\{R_i > r_0\}$$

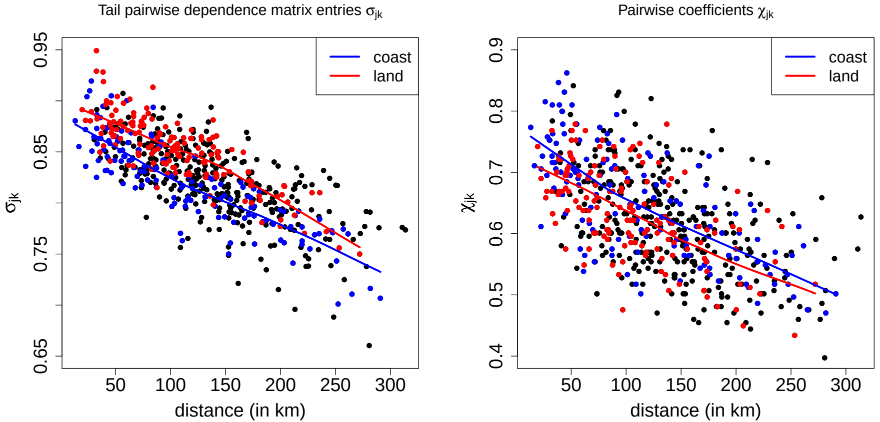
Example: daily maximum speeds of wind gusts

- Daily maximal speeds of wind gusts, measured in km/h, observed at 35 weather stations in the Netherlands during extended winter (October–March), $n = 3827$.
  - here, focus on inland stations only ($d = 18$)
- Marginal analysis: $\widehat{\alpha} \approx 9.3$ falls in the 95% confidence intervals of all marginals





Wind gusts: pairwise dependence coefficients



Failure region $C_{\max}$

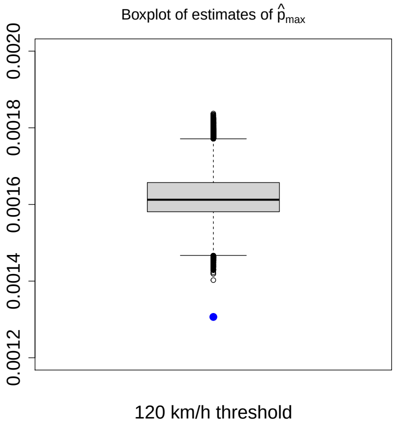
- Let $\boldsymbol{X} = (X_1, \ldots, X_{18})$ represent the maximum wind gusts at the 18 inland stations.
- As an example, let's calculate the probability that the maximum wind gust exceeds x at at least one station,

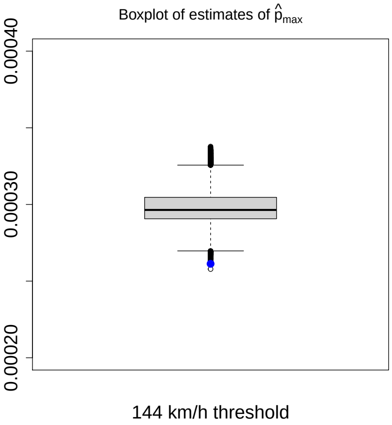
$$p_{\max}(x) = \mathbb{P}[\max(\boldsymbol{X}) > x],$$

- The KNMI issues an alarm for wind gusts exceeding 120 km/h.
- In February 2022, storm Eunice caused massive damage in Europe. The maximum wind gust measured in the Netherlands was 144 km/h (setting the record for harshest inland wind ever measured).
- We calculate 10 000 exact decompositions of $\widehat{\Sigma}_{\boldsymbol{X}}$ (computing time: $\sim$ 15 minutes).

Results: estimations of $p_{\max}$

Empirical estimates in blue; 5 exceedances (left) and 1 exceedance (right).





References

Cooley, D. and E. Thibaud (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika 106*(3), 587–604.

Cui, Q. and Z. Zhang (2018). Max-linear competing factor models. *Journal of Business & Economic Statistics 36*(1), 62–74.

Einmahl, J. H., A. Kiriliouk, and J. Segers (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes 21*, 205–233.

Fougeres, A.-L., C. Mercadier, and J. P. Nolan (2013). Dense classes of multivariate extreme value distributions. *Journal of Multivariate Analysis 116*, 109–129.

Gissibl, N. and C. Klüppelberg (2018). Max-linear models on directed acyclic graphs. *Bernoulli 24*(4A), 2693–2720.

Janßen, A. and P. Wan (2020). *k*-means clustering of extremes. *Electronic Journal of Statistics 14*(1), 1211–1233.

Larsson, M. and S. I. Resnick (2012). Extremal dependence measure and extremogram: the regularly varying case. *Extremes 15*(2), 231–256.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.