Metric embeddings of tail correlation matrices

Anja Janßen, Otto-von-Guericke University Magdeburg

joint work with

Sebastian Neblung, University of Hamburg,

and

Stilian Stoev, University of Michigan, Ann Arbor

ERCE Seminar, March 15th 2023

1. Introduction: Tail correlation matrices

2. Max-stable random vectors and L^1-embedding of the spectral distance

3. Tawn-Molchanov models and ℓ_1-embedding of the spectral distance

4. Conclusion

Disclaimer

Some parts of this talk are well-known for decades! Yet, we hope that a new perspective can provide new opportunities for applications.

1. Introduction: Tail correlation matrices

2. Max-stable random vectors and L^1-embedding of the spectral distance

3. Tawn-Molchanov models and ℓ_1-embedding of the spectral distance

4. Conclusion

One often looks for a handy, bivariate measure of **tail dependence** of random vectors $\mathbf{X} = (X_1, \dots, X_n)$, similar to the covariance or correlation matrix but only for extremes.

Tail correlation matrix

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector with identical marginal distributions and let $x_F := \inf\{x \in \mathbb{R} : P(X_1 \le x) = 1\}$. If the limit

$$\lambda_{i,j} := \lim_{x \to x_F} P(X_i > x | X_j > x) = \lim_{x \to x_F} \frac{P(X_i > x, X_j > x)}{P(X_j > x)}, \quad 1 \le i, j \le n$$

exists, then we call $\lambda_{i,j} \in [0,1]$ the **tail correlation coefficient** of X_i and X_j.
We call the matrix

$$\mathbf{\Lambda} := (\lambda_{i,j})_{1 \le i,j \le n}$$

the **tail correlation matrix** (TCM) of $\mathbf{X}$.

Advantages of the TCM:

- Easy to interpret: Fixed range of $[0, 1]$ and the larger, the more tail dependence
- (Comparably) easy to estimate: Transform observations from random vector to identical margins (via rank transform) and use empirical joint exceedance probabilities over high threshold (=quantile).
- (Partial knowledge of) bivariate tail correlation coefficients allow for (crude) upper and lower bounds on other probabilities of the form

$$\lim_{x \to x_F} \frac{P(\max_{i \in I} X_i > x)}{P(X_1 > x)}, \qquad \lim_{x \to x_F} \frac{P(\min_{i \in I} X_i > x)}{P(X_1 > x)}$$

for sets $I \subset \{1, \ldots, n\}$.

Disadvantages of the TCM:

- Provides too little information for relevant applications (for example exceedance probabilities of linear combinations of $\mathbf{X}$)
- Provides no information in case of weak tail dependence (for example Gaussian copula, for which $\mathbf{\Lambda} = \mathbf{I}_n$)
- Hard to identify whether a matrix is a TCM or not. **Hard to identify whether a matrix is a TCM or not.**

How can we check whether a given matrix is a TCM? And why do we want to do it in the first place?

- For estimators of the TCM, we would like them to have the same properties as a TCM.
- In order to find proper bounds (best or worse case scenarios) on possible values of the TCM (and derived higher-order tail exceedance probabilities)

Necessary properties:

- Entries of $\boldsymbol{\Lambda}$ need to be in $[0, 1]$ and entries on diagonal have to be 1's.

Characterization via Bernoulli-compatibility, Embrechts et al. (2016)

An $n \times n$-matrix $\boldsymbol{\Lambda}$ is a TCM if and only if

- It has 1's on its diagonal
- There exists a random vector $(\xi_i)_{1 \le i \le n} \in \{0, 1\}^n$ and some $c \ge 0$ such that
$$\boldsymbol{\Lambda}_{i,j} = c \cdot E(\xi_i \xi_j), \quad 1 \le i, j \le n.$$

This characterization can be derived from writing

$$\lim_{x \to x_F} \frac{P(X_i > x, X_j > x)}{P(X_j > x)} = \lim_{x \to x_F} \frac{E(\mathbb{1}_{\{X_i > x\}} \mathbb{1}_{\{X_j > x\}})}{P(X_j > x)}.$$

Characterization via Bernoulli-compatibility, Embrechts et al. (2016)

An $n \times n$-matrix $\mathbf{\Lambda}$ is a TCM if and only if

- It has 1's on its diagonal
- There exists a random vector $(\xi_i)_{1 \le i \le n} \in \{0,1\}^n$ and some $c \ge 0$ such that
$$\mathbf{\Lambda}_{i,j} = c \cdot E(\xi_i \xi_j), \quad 1 \le i,j \le n.$$

- The above Theorem allows to derive further necessary properties of a TCM, for example that it is positive-semidefinite.
- It does, however, help little in order to check whether a given matrix is a TCM or not, because knowledge of the matrix
$$E(\xi_i \xi_j)_{1 \le i,j \le n}$$
does (usually) not allow to re-construct the distribution of $(\xi_i)_{1 \le i \le n} \in \{0,1\}^n$.

Our aim is thus to learn more about the properties of a TCM.

Our definition of the tail correlation coefficients imply the following properties of $1-\lambda_{i,j}$:

- $1-\lambda_{i,j} = 1-\lim_{x\to x_F} P(X_i > x | X_j > x) \in [0,1]$, so $(i,j)\mapsto 1-\lambda_{i,j}$ is non-negative.
- $1-\lambda_{i,i} = 1-\lim_{x\to x_F} P(X_i > x | X_i > x) = 0$.
- $1-\lambda_{i,j} = 1-\lim_{x\to x_F} \frac{P(X_i>x,X_j>x)}{P(X_j>x)} = 1-\lim_{x\to x_F} \frac{P(X_j>x,X_i>x)}{P(X_i>x)} = 1-\lambda_{j,i}$, so the map $(i,j)\mapsto 1-\lambda_{i,j}$ is symmetric.
- For i,j,k

$$\begin{aligned}
(1-\lambda_{i,k}) &= \lim_{x\to x_F} P(X_i \le x | X_k > x) = \lim_{x\to x_F} \frac{P(X_i \le x, X_k > x)}{P(X_k > x)} \\
&= \lim_{x\to x_F} \frac{P(X_i \le x, X_j \le x, X_k > x) + P(X_i \le x, X_j > x, X_k > x)}{P(X_k > x)} \\
&\le \lim_{x\to x_F} \frac{P(X_j \le x, X_k > x) + P(X_i \le x, X_j > x)}{P(X_k > x)} \\
&= 1-\lambda_{j,k} + 1-\lambda_{i,j}
\end{aligned}$$

If Λ is a TCM, then the map

$$(i,j)\mapsto d(i,j) := 1-\lambda_{i,j}$$

defines a **semi-metric** on $\{1,\dots,n\}$. We call it **spectral distance**.

This distance between random variables is not new, but previously appeared in a specific framework of max-stable random vectors.

(Multiplicative) Max-stable random vectors

Let $\mathbf{X} = (X_1, \ldots, X_n)$ with $P(\mathbf{X} \neq \mathbf{0}) > 0$ be a random vector such that, for i.i.d. copies $\mathbf{X}^{(i)} = (X_1^{(i)}, \ldots, X_n^{(i)}), i \in \mathbb{N}$ of $\mathbf{X}$ there exist $a_m > 0$ such that, for all $m \in \mathbb{N}$,

$$a_m^{-1} \left(\max_{i=1,\ldots,m} X_1^{(i)}, \ldots, \max_{i=1,\ldots,m} X_n^{(i)} \right) \stackrel{d}{=} (X_1, \ldots, X_n).$$

Then, $\mathbf{X}$ is a (special case of a) **max-stable** random vector.
If we can choose $a_m = m$, then $\mathbf{X}$ is called **simple max-stable**.

One can easily show: A simple max-stable vector has Fréchet(1)-margins, i.e. $P(X_i \leq x) = \exp(-c_i/x), x > 0$ for some scale parameter $c_i > 0$.

Of course, not all random vectors are max-stable. But typically random vectors transformed to Fréchet(1)-margins satisfy the following property:

Max-stable random vectors evolve as limits

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector with Fréchet(1)-margins such that, for i.i.d. copies $\mathbf{X}^{(i)} = (X_1^{(i)}, \ldots, X_n^{(i)}), i \in \mathbb{N}$ of $\mathbf{X}$ there exists a non-degenerate random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ such that

$$m^{-1}\left(\max_{i=1,\ldots,m} X_1^{(i)}, \ldots, \max_{i=1,\ldots,m} X_n^{(i)}\right) \stackrel{d}{\Rightarrow} (Y_1, \ldots, Y_n).$$

Then, $\mathbf{Y}$ is simple max-stable and for all sets $A \subset \mathbb{R}^n$ bounded away from 0 we have

$$\lim_{x\to\infty} \frac{P(\mathbf{X} \in xA)}{P(X_1 > x)} = \lim_{x\to\infty} \frac{P(\mathbf{Y} \in xA)}{P(Y_1 > x)}.$$

Thus, the tail correlation coefficients of $\mathbf{X}$ and $\mathbf{Y}$ coincide and we will in the following focus on simple max-stable random vectors.

How can we alternatively characterize such max-stable random vectors?
We can use Poisson point processes for this!

Point process representation of simple max-stable random vectors

Equivalent are:

- There exist $(U_i)_{i\in\mathbb{N}} \overset{i.i.d.}{\sim} \text{Unif}([0,1])$, independent of $(\Gamma_i)_{i\in\mathbb{N}}$, which are an enumeration of points from a Poisson point process on $(0,\infty)$ with intensity $\nu((x,\infty)) = x^{-1}, x > 0$. Furthermore, for $1 \leq l \leq n$ there exist $f_l : [0,1] \to [0,\infty), f_l \in L^1([0,1])$ such that

$$X_l = \max_{i\in\mathbb{N}} \Gamma_i f_l(U_i), \quad l = 1,\ldots,n.$$

- The random vector $\mathbf{X} = (X_1,\ldots,X_n)$ is simple max-stable and its marginal distributions are of the form

$$P(X_l \leq x) = \exp(-\int_0^1 f_l(u)du/x), \quad x > 0.$$

We call the $f_l, l = 1,\ldots,n$ the **spectral functions** of the max-stable random vector.

The spectral functions are not uniquely determined, but the value of

$$\int_0^1 |f_i(u) - f_j(u)| du = \|f_i - f_j\|_{L^1}$$

is, and this was introduced in Davis & Resnick (1989) as a distance (semi-metric) between components of a max-stable random vector.
One can show that if $\int_0^1 f_l(u)du = 1, 1 \leq l \leq n$, then

$$1 - \lambda_{i,j} = 1 - \lim_{x\to\infty} P(X_i > x | X_j > x) = 0.5\|f_i - f_j\|_{L^1},$$

which implies

L^1-embeddability of the spectral distance

If $\mathbf{X} = (X_1, \ldots, X_n)$ is a simple max-stable random vector with standardized margins $P(X_i \leq x) = \exp(-1/x), 1 \leq i \leq n$, then its spectral distance, given by $\mathbf{D} := \mathbf{1} - \mathbf{\Lambda}$, is L^1-embeddable, i.e. there exist functions $g_1, \ldots, g_n \in L^1([0,1])$ such that

$$d_{i,j} = \|g_i - g_j\|_{L^1}, \quad 1 \leq i, j \leq n.$$

Note, however, that L^1-embeddability is difficult to check, and the spectral functions cannot be reconstructed from observations.

1 Introduction: Tail correlation matrices

2 Max-stable random vectors and L^1-embedding of the spectral distance

3 Tawn-Molchanov models and ℓ_1-embedding of the spectral distance

4 Conclusion

It is well-known that the L^1-embeddability of a semi-metric is equivalent to the ℓ_1-embeddability of a semi-metric:

ℓ_1-embeddability of a semi-metric

A semi-metric $d : \{1, \ldots, n\}^2 \to [0, \infty), (i, j) \mapsto d_{i,j}$ is called ℓ_1-**embeddable** if there exists a suitable dimension m and n points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ such that

$$d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|_1, \quad 1 \le i, j \le n.$$

If $m = 1$, then we call d a **line metric**.

From the previous, we thus know that our spectral distance is ℓ_1-embeddable and in the following we will find out more about this embedding. Indeed, it will lead us to a special class of simple max-stable vectors.

Let us look at the following simple class of random vectors

Tawn-Molchanov random vectors

For all $\emptyset \neq J \subset \{1, \ldots, n\}$ let $\beta_J \geq 0$ and let Y_J be i.i.d. standard Fréchet(1). Set

$$\mathbf{1}_J = (\mathbb{1}_{i \in J})_{1 \leq i \leq n}, \quad \emptyset \neq J \subset \{1, \ldots, n\}.$$

Then,

$$(X_1, \ldots, X_n) = \max_{\emptyset \neq J \subset \{1, \ldots, n\}} \beta_J Y_J \mathbf{1}_J$$

is a max-stable random vector, known as a **Tawn-Molchanov** model (TM model), see Strokorb & Schlather (2015).

- One can easily show that for such a random vector

$$P(\min_{i \in S} X_i > x) \sim x^{-1} \sum_{J \supset S} \beta_J, \quad x \to \infty. \tag{1}$$

- Thus, the coefficients β_J can be determined from joint exceedance probabilities and Tawn-Molchanov models can be used as simple approximations for more general max-stable random vectors.
- For all max-stable random vectors $\tilde{\mathbf{X}}$ sharing the asymptotics (1) with the TM-model $\mathbf{X}$, we have

$$P(\mathbf{X} \leq \mathbf{x}) \leq P(\tilde{\mathbf{X}} \leq \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n.$$

It is easy to derive the spectral distance for a Tawn-Molchanov model with standard Fréchet(1)-margins.
Remember / Verify:

$$
\begin{aligned}
d_{i,j} &= 1 - \lim_{x\to\infty} P(X_i > x | X_j > x) \\
&= \lim_{x\to\infty} P(X_i < x | X_j > x) \\
&= \frac{1}{2}\lim_{x\to\infty} \frac{P(X_i < x, X_j > x) + P(X_i > x, X_j < x)}{P(X_1 > x)} \\
&= \frac{1}{2} \sum_{\substack{\emptyset \neq J \subset \{1,\ldots,n\} \\ |J \cap \{i,j\}| = 1}} \beta_J \\
&= \sum_{\emptyset \neq J \subset \{1,\ldots,n\}} \left| \frac{\beta_J}{2}(\mathbf{1}_J)_i - \frac{\beta_J}{2}(\mathbf{1}_J)_j \right|
\end{aligned}
$$

Thus, for Tawn-Molchanov models, the coefficients β_J directly provide the ℓ_1-embedding of the spectral distance in dimension $m = 2^n - 1$.

1 Introduction: Tail correlation matrices

2 Max-stable random vectors and L^1-embedding of the spectral distance

3 Tawn-Molchanov models and ℓ_1-embedding of the spectral distance

4 Conclusion

- If $\mathbf{\Lambda}$ is a valid TCM of a simple max-stable random vector, then $\mathbf{D} := \mathbf{1} - \mathbf{\Lambda}$ defines a semi-metric.
- This semi-metric is both L^1- and ℓ_1-embeddable, with the embedding in direct correspondence to the spectral functions and Tawn-Molchanov coefficients.
- In general, none of the two embeddings is unique.
- However, problems about embeddability of a metric usually suffer from the high algorithmic complexity. In particular, it is known that the decision problem whether a semi-metric is ℓ_1-embeddable or not is **NP-hard** (see Deza & Laurent (1997)).

Here comes one more detail, swept under the rug so far

Necessary and sufficient conditions for TCMs

We have seen that L^1- and ℓ_1 embeddability of $\mathbf{D} := \mathbf{1} - \mathbf{\Lambda}$ are **necessary** properties for $\mathbf{\Lambda}$ being a TCM. They are, however, not sufficient. In particular, for $\mathbf{D}$ being an ℓ_1-embeddable distance, $\mathbf{\Lambda} := 1 - \mathbf{D}$ does not even need to have non-negative entries.

But with a few (technical) intermediate steps, we could confirm a conjecture of Shyamalkumar & Tao (2020)

Algorithmic complexity of TCM realization problem, J. et al (2022)

The decision problem whether a semi-metric is ℓ_1-embeddable can be transformed (in polynomial time) into the decision problem whether a matrix $\mathbf{\Lambda}$ is a TCM. Thus, the decision problem whether a matrix is a valid TCM or not is NP-hard.

- In general, the ℓ_1-embedding of a spectral distance, and thus, the corresponding Tawn-Molchanov-model to a TCM is not unique.

Uniqueness of ℓ_1-embedding, J. et al (2022)

If the spectral distance is a line metric (i.e. it can be ℓ_1-embedded in dimension 1), then this embedding is unique. In this case, the corresponding Tawn-Molchanov model is uniquely determined and $\beta_J > 0$ only for

$$J = \{1, \dots, i\} \text{ or } J = \{i, \dots, n\}, \quad 1 \leq i \leq n$$

for a suitable reordering of margins.

- Open question: If the representation is not unique, does there exist an "optimal" representation?

Some references

Deza, M. M. und Laurent, M.:
Geometry of Cuts and Metrics
Springer (1997)

Bandelt, H. J., & Dress, A. W.:
A canonical decomposition theory for metrics on a finite set.
Advances in mathematics, **92** (1992)

Davis, R. A., & Resnick, S. I.:
Basic properties and prediction of max-ARMA processes.
Adv. Appl. Probab. **21** (1989)

Embrechts, P., Hofert, M., & Wang, R.:
Bernoulli and tail-dependence compatibility
Ann. Appl. Probab. **26** (2016)

Shyamalkumar, N. D. und Tao, S.:
On tail dependence matrices
Extremes **23** (2020)

Janßen, A, Neblung, S. und Stoev, S.:
Tail dependence, exceedance sets and metric embeddings
Arxiv 2212.01044 (2022).

Strokorb, K., & Schlather, M.:
An exceptional max-stable process fully parameterized by its extremal coefficients.
Bernoulli **21** (2015)