Learning extremal graphical structures in high dimensions

Sebastian Engelke[1] **Michaël Lalancette**[2] Stanislav Volgushev[3]

[1]Research Center for Statistics, University of Geneva

[2]Research Chair on Mathematical Statistics, Technical University of Munich

[3]Department of Statistical Sciences, University of Toronto

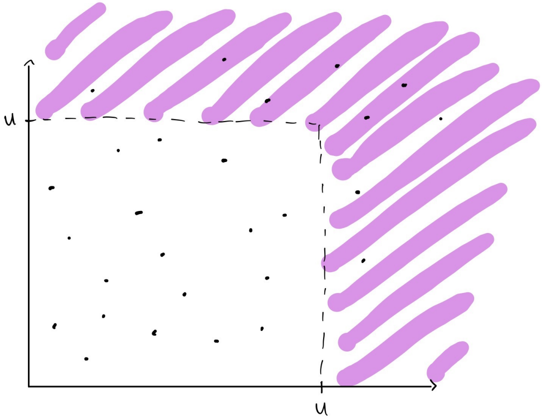
ETH-UCPH-TUM Workshop on Graphical Models, 2022-10-14

Tail (or extremal) dependence

- Random vector $\boldsymbol{X} \in \mathbb{R}^d$
- Tail dependence can be defined as the dependence structure of $\boldsymbol{X}$ in extreme regions/conditional on an extreme event
- Extreme events:

$$\{X_1 > u\} \quad \text{or} \quad \{\max X_i > u\} \quad \text{or} \quad \{\min X_i > u\}$$

Tail dependence: illustration



Multivariate Pareto distributions

- Suppose that

$$\mathbb{P}\big(F(\boldsymbol{X}) \leq 1 - q/\boldsymbol{x} \mid \max_i F_i(X_i) > 1 - q\big) \longrightarrow \mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}), \quad q \downarrow 0,$$

  where $F(\boldsymbol{X}) := (F_1(X_1), \ldots, F_d(X_d))$
- "Given that at least one component of $\boldsymbol{X}$ exceeds it's $(1-q)$th quantile, $q/(1-F(\boldsymbol{X})) \approx \boldsymbol{Y}$ in distribution"
- $\boldsymbol{Y}$ is *multivariate Pareto* (MP)
- $\boldsymbol{X}$ is in the *domain of attraction* of $\boldsymbol{Y}$ (tail dependence of $\boldsymbol{X} \approx \boldsymbol{Y}$)
- Goal: observe $\boldsymbol{X}$, but model and estimate $\boldsymbol{Y}$ directly because extrapolation!

Multivariate Pareto distributions: 3 defining properties

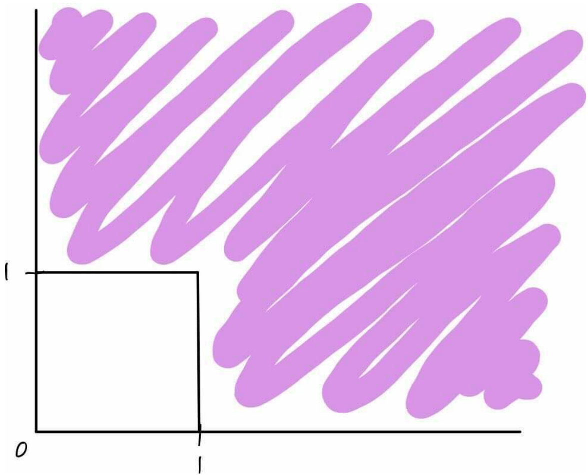
- $\boldsymbol{Y} \in \mathcal{L} := \{\boldsymbol{y} \geq 0 : \|\boldsymbol{y}\|_\infty > 1\}$
- $\mathbb{P}(Y_1 > 1) = \cdots = \mathbb{P}(Y_d > 1)$
- For $A \subset \mathcal{L}$ and $t \geq 1$,

$$\mathbb{P}(\boldsymbol{Y} \in tA) = t^{-1}\mathbb{P}(\boldsymbol{Y} \in A)$$

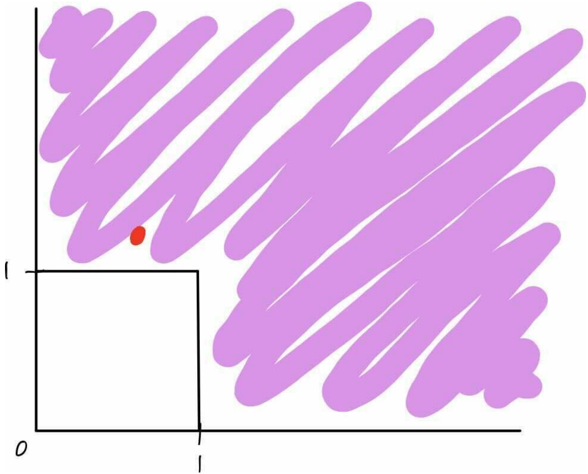
- If $\boldsymbol{Y}$ has density f, for $\boldsymbol{y} \in \mathcal{L}$, $t \geq 1$,

$$f(t\boldsymbol{y}) = t^{-(d+1)} f(\boldsymbol{y})$$

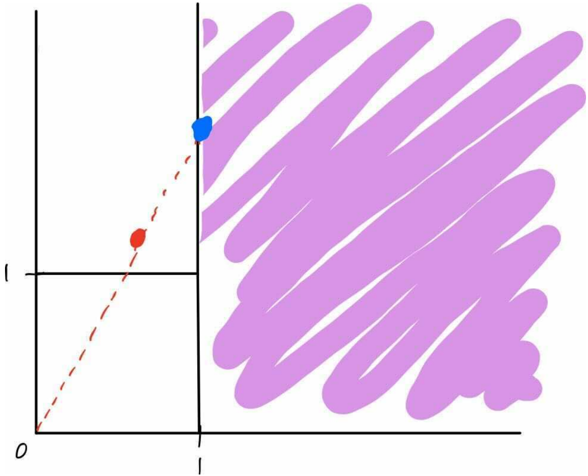
Multivariate Pareto distributions: homogeneity



Multivariate Pareto distributions: homogeneity



Multivariate Pareto distributions: homogeneity



Multivariate Pareto distributions: conditional independence

- Suppose $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ has a *positive* density f
- Support $\mathcal{L} \neq$ product space
- We say that $Y_i \perp_e Y_j \mid \boldsymbol{Y}_{\setminus\{i,j\}}$ if for some (all) $m \notin \{i,j\}$,

$$Y_i \perp Y_j \mid \{\boldsymbol{Y}_{\setminus\{i,j\}}, Y_m > 1\}$$

- Heuristic:

$$\begin{aligned} f \text{ factorizes on } \mathcal{L} &\iff f \text{ factorizes on } \{\boldsymbol{y} \in \mathcal{L} : y_m > 1\} \\ &\iff \text{density of } \boldsymbol{Y} \mid Y_m > 1 \text{ factorizes on its domain} \\ &\iff Y_i \perp Y_j \mid \{\boldsymbol{Y}_{\setminus\{i,j\}}, Y_m > 1\} \end{aligned}$$

Extremal graphical models

- $G := (V := [d], E)$ an undirected, connected graph
- $\boldsymbol{Y}$ is an *extremal graphical model* on G if for each pair (i, j),

$$Y_i \perp_e Y_j \mid \boldsymbol{Y}_{\setminus\{i,j\}} \iff (i,j) \notin E$$

- Faithfulness
- Engelke & Hitz (2020, JRSSB)

Hüsler–Reiss distributions

- A family of MP distributions, parametrized by an *extremal variogram* $\Gamma \in \mathbb{R}^{d \times d}$
- If $\boldsymbol{Y} \sim \mathrm{HR}(\Gamma)$,

$$\Gamma_{ij} = \mathbb{V}\mathrm{ar}(\log Y_i - \log Y_j \mid Y_m > 1)$$

Estimating Hüsler–Reiss distributions: the empirical variogram

- $\boldsymbol{X}$ in the domain of attraction of $\boldsymbol{Y} \sim \mathrm{HR}(\Gamma)$, iid data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \boldsymbol{X}$
- For $m \in V$, estimate Γ_{ij} by

$$\widehat{\Gamma}_{ij}^{(m)} := \widehat{\mathbb{V}\mathrm{ar}}\Big(\log(1 - \widetilde{F}_i(X_{ti})) - \log(1 - \widetilde{F}_j(X_{tj})) \mid \widetilde{F}_m(X_{tm}) > 1 - k/n\Big),$$

where k large, k/n small, $\widetilde{F}_i$ are empirical df

- $\widehat{\Gamma} := d^{-1} \sum_{m=1}^{d} \widehat{\Gamma}^{(m)}$

Theorem (Engelke, L. & Volgushev, 2022)

Under (mild) assumptions, with probability at least $1 - \delta$,

$$\|\widehat{\Gamma} - \Gamma\|_\infty \lesssim \left(\frac{k}{n}\right)^\xi (\log(n/k))^2 + \sqrt{\frac{\log d + \log \frac{1}{\delta}}{k}}.$$

HR graphical models

- Why this model?
- If $\boldsymbol{Y} \sim \mathrm{HR}(\Gamma)$, $\boldsymbol{Y} / Y_m \mid Y_m > 1$ is log-Gaussian with "covariance matrix"

$$\Sigma^{(m)} := \frac{1}{2}(\Gamma_{im} + \Gamma_{jm} - \Gamma_{ij})_{i,j \in V}$$

- So for $\{i,j\} \not\ni m$,

$$Y_i \perp_e Y_j \mid \boldsymbol{Y}_{\setminus \{i,j\}} \iff \Theta^{(m)}_{ij} = 0,$$

  where $\Theta^{(m)}$ is the (pseudo)inverse of $\Sigma^{(m)}$
- Extremal graph structure is encoded into the zero pattern of the matrices $\Theta^{(m)}$
- Estimate those and combine them through majority voting

`EGlearn`: learning HR graphical models

- For $m \in V$,
  1. Compute
  $$\widehat{\Sigma}^{(m)} := \frac{1}{2}(\widehat{\Gamma}_{im} + \widehat{\Gamma}_{jm} - \widehat{\Gamma}_{ij})_{i,j \in V}, \quad m \in V$$
  2. Throw $\widehat{\Sigma}^{(m)}$ into a base learner $\mathcal{A}$ to obtain a sparse estimate $\widehat{\mathbb{1}}\{\Theta^{(m)} \neq 0\}$
- For each pair (i, j), add an edge to $\widehat{E}$ if and only if
$$\frac{1}{d-2} \# \left\{ m \in V \setminus \{i, j\} : \widehat{\mathbb{1}}\{\Theta^{(m)}_{ij} \neq 0\} = 1 \right\} > \frac{1}{2}$$
- Graph estimate $\widehat{G} := (V, \widehat{E})$

EGlearn: illustration

$$
\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 \\ \cdot & 0 & \cdot & 1 \\ \cdot & 1 & 1 & \cdot \end{pmatrix}
\quad
\begin{pmatrix} \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & 1 & \cdot \end{pmatrix}
\quad
\begin{pmatrix} \cdot & 1 & \cdot & 0 \\ 1 & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & \cdot \end{pmatrix}
\quad
\begin{pmatrix} \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}
$$

Figure: Estimated sparsity pattern of $\Theta^{(m)}$, $m = 1, 2, 3, 4$

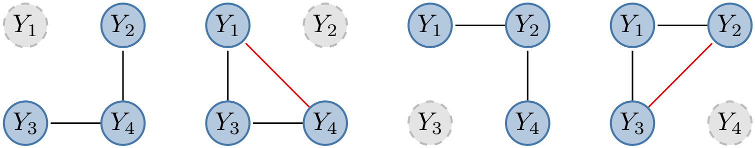


Figure: Corresponding votes

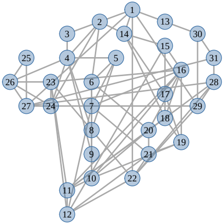
EGlearn: model selection consistency

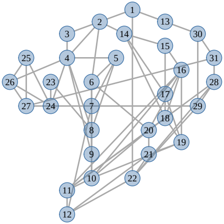
Theorem (Engelke, L. & Volgushev, 2022)

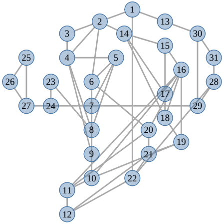
If $\mathcal{A}$ is neighborhood selection or graphical lasso, under assumptions,

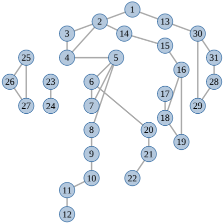
$$\mathbb{P}(\widehat{G} = G) \longrightarrow 1$$

as long as $\log d = o(k/(\log k)^8)$.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Implementation

Available in the package `graphicalExtremes` (Engelke, Hitz & Gnecco, 2019)

Selected references

Extremal graphical models

Engelke, S. and A. S. Hitz (2020). Graphical models for extremes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol. 82*, 871–932.

Engelke, S., A. S. Hitz and N. Gnecco (2019). `graphicalExtremes`. *R package*.

Engelke, S., M. Lalancette and S. Volgushev (2022+). Learning extremal graphical structures in high dimensions. *arXiv preprint arXiv:2111.00840*

Gaussian graphical models and sparse precision matrix estimation

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics 34(3)*, 1436–1462.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika 94(1)*, 19–35.

Summary

- Extremal graphical models allow sparse representation of extremal dependence structure
- In the HR parametric family, they can be learned from data even in exponentially high dimension
- We do so using majority voting combined with Gaussian graphical modeling tools
- Preprint available on arXiv
- Extensions:
  - Beyond HR
  - Independence and disconnected graphs
  - DAGs, SEMs and causality?

Thank you for your attention!

mic-lalancette.github.io