Asymptotics and Concentration of Empirical Variograms

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

[1]Research Center for Statistics, University of Geneva

[2]Department of Statistical Sciences, University of Toronto

Extreme Value Analysis, 02/07/2021

Multivariate peaks-over-threshold

- $\boldsymbol{X} \in [0, \infty)^d$ has unit Pareto marginals
- Tail dependence of $\boldsymbol{X}$ is often thought of as the dependence structure of $\boldsymbol{X}$, assuming $\|\boldsymbol{X}\|_\infty$ is large
- Formally, if

$$\left\{ q\boldsymbol{X} \;\middle|\; \|\boldsymbol{X}\|_\infty > q^{-1} \right\} \rightsquigarrow \boldsymbol{Y}, \quad q \to 0,$$

  the distribution of $\boldsymbol{Y}$ is called a *multivariate Pareto (MP) distribution*
- $\boldsymbol{X}$ is in the *domain of attraction* (DA) of $\boldsymbol{Y}$

Multivariate Pareto distributions

- Although MP are the main focus when doing multivariate peaks-over-threshold inference, they are related to the well known property of *max-stability*
- If $\boldsymbol{X}$ is in the MDA of a max-stable distribution $\boldsymbol{Z}$, then $\boldsymbol{X}$ is also in the DA of a MP $\boldsymbol{Y}$, and $\boldsymbol{Z}$ and $\boldsymbol{Y}$ define each other
- In particular, $\{$distribution of $\boldsymbol{Z}\} \Leftrightarrow \{$distribution of $\boldsymbol{Y}\} \Leftrightarrow R$, defined as

$$R(\boldsymbol{x}) := \lim_{q\to 0} q^{-1}\mathbb{P}\big(X_1 \geq (qx_1)^{-1}, \ldots, X_d \geq (qx_d)^{-1}\big), \quad \boldsymbol{x} \in [0,\infty]^d$$

- So every max-stable dependence model has a unique "associated" MP distribution (if $\boldsymbol{Z}$ is Hüsler–Reiss, $\boldsymbol{Y}$ is *Hüsler–Reiss Pareto* (HRP))

Extremal graphical models

- Engelke & Hitz (2020) construct graphical models for MP distributions
- Graphical model *selection* for MP distributions
- Two special cases of extremal graphical models
    1. G is a tree (but $\boldsymbol{Y}$ is "arbitrary" MP)
    2. $\boldsymbol{Y}$ is HRP (but G is "arbitrary" graph)

Extremal variograms

- In both cases, the graph structure is encoded into the *extremal variogram matrix* $\Gamma^{(m)}$ of $\boldsymbol{Y}$ rooted at variable $m \in V = \{1, \ldots, d\}$,

$$\Gamma_{ij}^{(m)} := \mathbb{V}\text{ar}(\log Y_i - \log Y_j \mid Y_m > 1), \quad i, j \in V$$

- When $\boldsymbol{Y}$ is HRP, all the $\Gamma^{(m)} = \Gamma$, the parameter matrix, i.e. $\Gamma^{(m)}$ fully characterize the HRP distributions

Empirical variograms

- Knowledge of an extremal variogram $\Gamma^{(m)} \Longrightarrow$ Identification of the full tree structure and of the full model of $\boldsymbol{Y}$ is HRP
- Estimation of $\Gamma^{(m)} \Longrightarrow$
  - Estimation of extremal tree models (Engelke & Volgushev, 2020)
  - Estimation of Hüsler–Reiss distributions
    - Without enforced sparsity on the graphical model (Engelke & al., 2015)
    - With enforced sparsity on the graphical model, through L^1 penalization (Sebastian Engelke's talk) or through a total positivity constraint (Frank Röttger's talk)
- Motivates study of the *empirical variogram*

Empirical variograms

- (iid) data $\boldsymbol{X}_t := (X_{t1}, \dots, X_{td})$, $1 \leq t \leq n$, in the DA of $\boldsymbol{Y}$
- Select $1 \leq k := k_n \leq n$
- Define approximate MP observations $\widehat{Y}_{ti} := \frac{k}{n}\left(1 - \widehat{F}_i(X_{it})\right)^{-1}$, so k of them exceed 1
- Calculate sample variances

$$\widehat{\Gamma}_{ij}^{(m)} := \frac{1}{k}\sum_{t=1}^{n}\left(\log \widehat{Y}_{ti} - \log \widehat{Y}_{tj}\right)^2 \mathbb{1}\left\{\widehat{Y}_{tm} \geq 1\right\} - \left(\frac{1}{k}\sum_{t=1}^{n}\left(\log \widehat{Y}_{ti} - \log \widehat{Y}_{tj}\right)\mathbb{1}\left\{\widehat{Y}_{tm} \geq 1\right\}\right)^2$$

- Question: how well does $\widehat{\Gamma}^{(m)}$ estimate $\Gamma^{(m)}$?
- (Engelke & Volgushev, 2020) prove that $\widehat{\Gamma}_{ij}^{(m)} \xrightarrow{P} \Gamma_{ij}^{(m)}$

Tail assumption

- Recall that

$$q^{-1}\mathbb{P}\big(X_1 \geq (qx_1)^{-1}, \ldots, X_d \geq (qx_d)^{-1}\big) \longrightarrow R(\boldsymbol{x})$$

- Tail assumption: There exists K, $\xi > 0$: as $q \to 0$,
  1. For $|J| \in \{2, 3\}$,

$$\sup_{\boldsymbol{x}_J \in [0,1]^{|J|}} \left| q^{-1}\mathbb{P}\big(X_j \geq (qx_j)^{-1},\ j \in J\big) - R_J(\boldsymbol{x}_J) \right| = Kq^{\xi}$$

  2. For $i \neq j$,

$$1 - R_{ij}(q^{-1}, 1) \leq Kq^{\xi}$$

  Only an assumption on the tail model (satisfied by any HR distribution, except perfect independence)

- "Choice of k" assumption: There exist $0 < \alpha \leq \beta < 2\xi/(2\xi + 1)$ such that

$$n^{\alpha} \lesssim k \lesssim n^{\beta}$$

Density assumption

- Density assumption: the functions R_{ij} have continuous partial derivatives and densities r_{ij} satisfying

$$r_{ij}(x, 1-x) \leq K(x(1-x))^{\varepsilon}, \quad x \in (0,1)$$

  for some $\varepsilon > 0$
- Satisfied if $\boldsymbol{Y}$ is HRP, unless perfect dependence or independence
- Not the weakest possible

Asymptotic distribution

Conjecture (Engelke, L. & Volgushev, 2021+)

Under the "tail", "choice of k" and "density" assumptions,

$$\sqrt{k}\big(\widehat{\Gamma}^{(m)} - \Gamma^{(m)}\big)_{m\in V} \rightsquigarrow \big(W^{(m)}\big)_{m\in V}$$

for a Gaussian $\big(W^{(m)}\big)_{m\in V}$.

Consequences:

- Asymptotic normality of the estimator of the parameters of HR distributions in (Engelke & al., 2015)
- Confidence sets and tests for graphical models (in fixed dimension)
- Not informative in growing dimension (e.g. $d > n$)

Concentration

Theorem (Engelke, L. & Volgushev, 2021+)

Let $\delta \geq d^3 e^{-\sqrt{k}}$. Under the "tail" and "choice of k" assumptions, with probability at least $1-\delta$

$$\max_{i,j,m \in V} \left|\widehat{\Gamma}_{ij}^{(m)} - \Gamma_{ij}^{(m)}\right| \leq C(\log n)^2 \sqrt{\frac{\log d + \log(1/\delta)}{k}}.$$

Further, under the "density" assumption, with probability at least $1-\delta$

$$\max_{i,j,m \in V} \left|\widehat{\Gamma}_{ij}^{(m)} - \Gamma_{ij}^{(m)}\right| \leq \bar{C} \sqrt{\frac{\log d + \log(1/\delta)}{k}}.$$

Corollaries: Extremal graph learning guarantees

- Extremal tree learning (Engelke & Volgushev, 2020): $\mathbb{P}(\widehat{G} = G) \to 1$ if

$$c_1 n^{\beta/2} - \log d \to \infty$$

- Extremal graphical lasso (Sebastian Engelke's talk): $\mathbb{P}(\widehat{G} = G) \to 1$ if

$$c_2 n^{\beta/2} - \log d \to \infty$$

- MTP_2 constrained graph estimation (Frank Röttger's talk): learning guarantees in high dimension?

Bonus: Discussion of the proofs

- Since everything is conditioned on $Y_m > 1$, we write $Y_i^{(m)} := Y_i \mid Y_m > 1$ and $\widehat{Y}_i^{(m)} := \widehat{Y}_i \mid \widehat{Y}_m > 1$
- Assume i, j, m distinct
- Then

$$\Gamma_{ij}^{(m)} = \mathbb{E}[(\log Y_i^{(m)})^2] + \mathbb{E}[(\log Y_j^{(m)})^2] - \mathbb{E}[(\log Y_i^{(m)})(\log Y_j^{(m)})] - \left(\mathbb{E}[\log Y_i^{(m)}] - \mathbb{E}[\log Y_j^{(m)}]\right)^2$$

- Similarly

$$\widehat{\Gamma}_{ij}^{(m)} = \widehat{\mathbb{E}}[(\log \widehat{Y}_i^{(m)})^2] + \widehat{\mathbb{E}}[(\log \widehat{Y}_j^{(m)})^2] - \widehat{\mathbb{E}}[(\log \widehat{Y}_i^{(m)})(\log \widehat{Y}_j^{(m)})] - \left(\widehat{\mathbb{E}}[\log \widehat{Y}_i^{(m)}] - \widehat{\mathbb{E}}[\log \widehat{Y}_j^{(m)}]\right)^2$$

Bonus: Discussion of the proofs

- $\mathbb{P}(Y_i^{(m)} \geq x, Y_j^{(m)} \geq y) = R_{ijm}(1/x, 1/y, 1)$
- Using that for $X_1, X_2 \geq 0$

$$\mathbb{E}[X_1 X_2] = \int_0^\infty \int_0^\infty \mathbb{P}(X_1 \geq x_1, X_2 \geq x_2) dx_1 dx_2$$

- Obtain expression for $\mathbb{E}[(\log Y_i^{(m)})(\log Y_j^{(m)})]$:

$$\int_0^1 \int_0^1 \frac{R_{ijm}(x, y, 1)}{xy} dxdy - \int_0^1 \int_1^\infty \frac{R_{ijm}([x, \infty), y, 1)}{xy} dxdy$$
$$- \int_1^\infty \int_0^1 \frac{R_{ijm}(x, [y, \infty), 1)}{xy} dxdy + \int_1^\infty \int_1^\infty \frac{R_{ijm}([x, \infty), [y, \infty), 1)}{xy} dxdy$$

- R_{ijm} can be seen as a measure, so e.g. $R_{ijm}([x, \infty), y, 1) := R_{jm}(y, 1) - R_{ijm}(x, y, 1)$

Bonus: Discussion of the proofs

- Similarly, $\widehat{\mathbb{E}}[(\log \widehat{Y}_i^{(m)})(\log \widehat{Y}_j^{(m)})]$ is equal to

$$\int_0^1 \int_0^1 \frac{\bar{R}_{ijm}(x,y,1)}{xy} dxdy - \int_0^1 \int_1^\infty \frac{\bar{R}_{ijm}([x,\infty),y,1)}{xy} dxdy - \int_1^\infty \int_0^1 \frac{\bar{R}_{ijm}(x,[y,\infty),1)}{xy} dxdy + \int_1^\infty \int_1^\infty \frac{\bar{R}_{ijm}([x,\infty),[y,\infty),1)}{xy} dxdy$$

- The *tail empirical copula*

$$\bar{R}_{ijm}(x,y,1) := \frac{1}{k} \sum_{t=1}^{n} \mathbb{1}\left\{ \widehat{F}_i(X_{ti}) \geq 1 - \frac{k}{n}x, \widehat{F}_j(X_{tj}) \geq 1 - \frac{k}{n}y, \widehat{F}_m(X_{tm}) \geq 1 - \frac{k}{n} \right\}$$

Bonus: Discussion of the proofs

- Good news: reduced the problem to studying the *tail empirical copula process* $\sqrt{k}(\bar{R}_{ijm} - R_{ijm})$
- Well known that $\sqrt{k}(\bar{R}_{ijm} - R_{ijm})$ converges to a GP in $\ell^\infty(\mathcal{K})$ for compact $\mathcal{K}$
- Similarly, can easily get concentration result for $\sup_{\mathcal{K}} |\bar{R}_{ijm} - R_{ijm}|$
- Bad news: None of those are sufficient, since we consider unbounded sets and an unbounded weighting
- Our fix: choose $\eta > 0$ and index the tail empirical copula process by the functions

$$f_{ijm,x,y,z} := \begin{cases} (xy)^{-\eta}\mathbb{1}_{[0,x]\times[0,y]\times[0,z]}, & 0 < x,y,z \leq 1 \\ x^\eta y^{-\eta}\mathbb{1}_{[x,\infty)\times[0,y]\times[0,z]}, & 1 < x < \infty, 0 < y,z \leq 1 \\ (xy)^\eta \mathbb{1}_{[x,\infty)\times[y,\infty)\times[0,z]}, & 1 < x,y < \infty, 0 < z \leq 1 \end{cases}$$

Thank you for your attention! Questions?

References

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

Engelke, S., A. Malinowski, Z. Kabluchko, and M. Schlather (2015). Estimation of hüsler–reiss distributions and brown–resnick processes. *J. R. Stat. Soc. B 77*, 239–265.

Engelke, S. and S. Volgushev (2020). Structure learning for extremal tree models. *arXiv preprint arXiv:2012.06179*.