MATHÉMATIQUES
VISION
APPRENTISSAGE

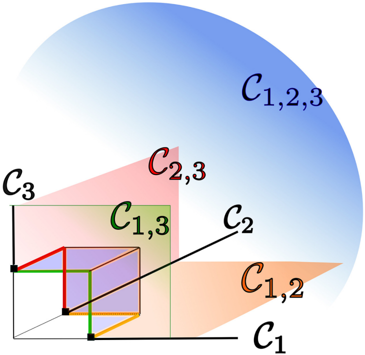
Sparse representation of multivariate extremes with applications to anomaly detection

Mathis Hardion

Problem of interest

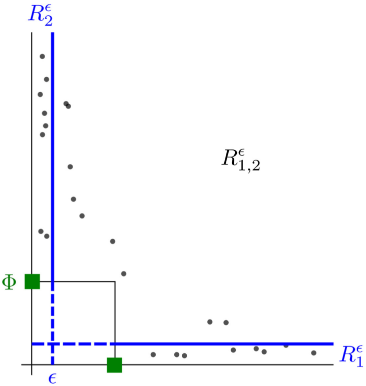
- Study the dependence structure of extremes to find the main directions among which they may happen to reduce problem dimensionality
  - Bridge the gap between existing low-dimensional methods and more complex problems
  - Leverage the multivariate regular variation hypothesis by estimating the angular measure
- As a byproduct, detect anomalies as extremes in improbable directions

Formulation



$$\mathcal{C}_\alpha := \left\{ \mathbf{v} \geq 0 \;\middle|\; \|\mathbf{v}\|_\infty \geq 1, \mathbf{v}^{|\alpha} > 0, \mathbf{v}^{|\alpha^c} = 0 \right\}$$

Figure 1: truncated cones in $\mathbb{R}^3$



$$R_\alpha^\varepsilon := \left\{ \mathbf{v} \geq 0 \;\middle|\; \|\mathbf{v}\|_\infty \geq 1, \mathbf{v}^{|\alpha} \geq \varepsilon, \mathbf{v}^{|\alpha^c} < \varepsilon \right\}$$

Figure 2: truncated ε-rectangles in $\mathbb{R}^2$

Figures from [4]

Formulation

- Usual rank transform $\mathbf{V} := \left(\frac{1}{1-F_j(X^j)}\right)_{1\leq j\leq d}$ and regular variation with exponent measure μ: $\forall \mathbf{v}, n\mathbb{P}\left(\frac{1}{n}\mathbf{V} \in [\mathbf{0}, \mathbf{v}]^c\right) \underset{n\to\infty}{\longrightarrow} \mu\left([\mathbf{0}, \mathbf{v}]^c\right)$.

Lemma

$$\mu\left(R_\alpha^\varepsilon\right) \underset{\varepsilon\to 0}{\longrightarrow} \mu\left(\mathcal{C}_\alpha\right).$$

Problem

Given $(\mathbf{X}_i)_i \overset{\text{i.i.d}}{\sim} \mathbf{X}$, estimate $\mathcal{M} := \left(\mu\left(\mathcal{C}_\alpha\right)\right)_{\alpha\subset\{1,\dots,d\}}$ and derive non-asymptotic bounds on the error.

MATHÉMATIQUES VISION APPRENTISSAGE

Hypotheses

1. The marginal cdfs $(F_j)_{1\leq j\leq d}$ are continuous.

2. For $\varnothing \neq \alpha \coloneqq \{i_1, \ldots, i_r\} \subset \{1, \ldots, d\}$, $\mu_\alpha(\cdot) \coloneqq \mu(\cdot \cap \mathcal{C}_\alpha)$ is absolutely continuous with respect to $dx_\alpha \coloneqq dx_{i_1} \ldots dx_{i_r}$.

3. The angular density is uniformly bounded, so that there exists $M > 0$ verifying

$$\sum_{\substack{\beta \subset \{1,\ldots,d\} \\ |\beta|\geq 2}} \sup_{i\in\beta} \sup_{\Omega_{\beta,i}} \frac{d\Phi_{\alpha,i_0}}{dx_{\alpha\setminus\{i_0\}}} < M.$$

MATHÉMATIQUES VISION APPRENTISSAGE

Nonparametric estimation of $\mathcal{M}$

- $\widehat{\mathbf{V}}_i$ computed with the ECDF
- $\widehat{\mathbb{P}}_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{\mathbf{V}}_i}$ their empirical distribution
- Following the RV, set $\widehat{\mu}_n(\cdot) := \frac{n}{k_n}\widehat{\mathbb{P}}_n\left(\frac{n}{k_n}\cdot\right)$, $\frac{n}{k_n} \to \infty$
- Build the estimator $\widehat{\mathcal{M}}(\alpha) := \widehat{\mu}_n(R_\alpha^\varepsilon) = \frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{1}\left\{\widehat{\mathbf{V}}_{\sigma(i)}^{|\alpha} \geq \frac{n}{k_n}\varepsilon, \widehat{\mathbf{V}}_{\sigma(i)}^{|\alpha^c} < \frac{n}{k_n}\varepsilon\right\}$

MATHÉMATIQUES VISION APPRENTISSAGE

Error bounds

- Triangle inequality:

$$\|\widehat{\mathcal{M}} - \mathcal{M}\|_\infty \leq \max_\alpha |\mu - \widehat{\mu}_n| (R_\alpha^\varepsilon) + \max_\alpha |\mu(\mathcal{C}_\alpha) - \mu(R_\alpha^\varepsilon)|$$

- Extend bounds in [3]:

Property

There exists $C > 0$ such that for $0 < \varepsilon < \frac{1}{4}$, $\delta \geq e^{-k_n}$, with probability at least $1 - \delta$,

$$\max_\alpha |\mu - \widehat{\mu}_n| (R_\alpha^\varepsilon) \leq Cd\sqrt{\frac{1}{\varepsilon k_n} \ln\left(\frac{d+3}{\delta}\right)} + 2 \max_{\alpha,\beta} \sup_{\mathbf{0} \leq \mathbf{x},\mathbf{z} \leq 2\varepsilon^{-1}} \left| \frac{n}{k_n} \tilde{F}_{\alpha,\beta}\left(\frac{k_n}{n}\mathbf{x}, \frac{k_n}{n}\mathbf{z}\right) - g_{\alpha,\beta}(\mathbf{x}, \mathbf{z}) \right|.$$

MATHÉMATIQUES VISION APPRENTISSAGE

Error bounds

Property

Under assumptions 2 and 3,

$$|\mu(R_\alpha^\varepsilon) - \mu(\mathcal{C}_\alpha)| \leq M d^2 \varepsilon.$$

Theorem

Under assumptions 2 and 3, there exists $C < 0$, such that for $0 < \varepsilon < \frac{1}{4}$, $\delta \geq e^{-k_n}$, with probability at least $1 - \delta$,

$$\|\widehat{\mathcal{M}} - \mathcal{M}\|_\infty \leq C d \left(\sqrt{\frac{1}{\varepsilon k_n} \ln\left(\frac{d+3}{\delta}\right)} + M d \varepsilon \right) + 2 \max_{\alpha,\beta} \sup_{\mathbf{0} \leq \mathbf{x},\mathbf{z} \leq 2\varepsilon^{-1}} \left| \frac{n}{k_n} \tilde{F}_{\alpha,\beta}\left(\frac{k_n}{n}\mathbf{x}, \frac{k_n}{n}\mathbf{z}\right) - g_{\alpha,\beta}(\mathbf{x},\mathbf{z}) \right|.$$

MATHÉMATIQUES VISION APPRENTISSAGE

Thresholding

- To deal with noise and gain extra sparsity: remove values of $\widehat{\mathcal{M}}(\alpha)$ under θ
- For instance, $\theta = p\left|\left\{\alpha \,\middle|\, \widehat{\mathcal{M}}(\alpha) > 0\right\}\right|^{-1} \sum_\alpha \widehat{\mathcal{M}}(\alpha)$ for $p > 0$
- $\widehat{\mathcal{M}}$ is an ERM, thresholding $\leftrightarrow$ L^1 regularization

DAMEX

- Introduce the score function $\widehat{s}(\mathbf{x}) := \frac{\widehat{\mathcal{M}}(\alpha(\mathbf{x}))}{\|\widehat{T}(\mathbf{x})\|_\infty}$ which plays a similar role to a p-value
- Evaluation on labeled dataset : train on normal region, test on extremes
- Compare with iForest [6], Local Outlier Factor [1]

	DAMEX			IsolationForest		LocalOutlierFactor	
k_n	AUC ROC	AUC PR	AFD	AUC ROC	AUC PR	AUC ROC	AUC PR
$n^{\frac{1}{4}}$	0.503	0.054	8.76	0.947	0.496	**0.996**	**0.961**
$\sqrt{n}$	0.895	0.678	24.6	0.884	0.614	**0.994**	**0.982**
$n^{\frac{3}{4}}$	0.817	0.773	54.0	0.715	0.498	**0.994**	**0.987**
$n^{\frac{1}{4}}\ln(n)$	0.939	0.806	26.2	0.911	0.638	**0.994**	**0.981**

TABLE 1: Results on extreme region with varying k_n, $\varepsilon = 0.01$, $p = 0.1$

References

[1] Markus M. Breunig et al. "LOF: Identifying Density-Based Local Outliers". In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. SIGMOD '00. Dallas, Texas, USA: Association for Computing Machinery, 2000, pp. 93–104. DOI: 10.1145/342009.335388. URL: https://doi.org/10.1145/342009.335388.

[2] The scikit-learn community. *Forest covertypes dataset*. scikit-learn 1.3.2 documentation. URL: https://scikit-learn.org/stable/datasets/real_world.html#covtype-dataset (visited on 03/12/2023).

[3] N. Goix, A. Sabourin, and S. Clémençon. "Learning the dependence structure of rare events: a non-asymptotic study". In: *Proc. COLT*. 2015.

[4] Nicolas Goix, Anne Sabourin, and Stephan Clémençon. "Sparse representation of multivariate extremes with applications to anomaly detection". In: *Journal of Multivariate Analysis* 161 (2017), pp. 12–31. DOI: https://doi.org/10.1016/j.jmva.2017.06.010. URL: https://www.sciencedirect.com/science/article/pii/S0047259X17304062.

[5] M. Hardion. *damex notebook*. URL: https://github.com/mhardion/damex.

[6] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation Forest". In: *2008 Eighth IEEE International Conference on Data Mining*. 2008, pp. 413–422. DOI: 10.1109/ICDM.2008.17.

[7] Yongcheng Qi. "Almost sure convergence of the stable tail empirical dependence function in multivariate extreme statistics". English (US). In: *Acta Mathematicae Applicatae Sinica* 13.2 (1997), pp. 167–175. DOI: 10.1007/BF02015138.

First bound construction

$$R(\mathbf{x}, \mathbf{z}, \alpha, \beta) := \left\{ \mathbf{y} \in [\mathbf{0}, \boldsymbol{\infty}]^d, \mathbf{y}^{|\alpha} \geq \mathbf{x}^{|\alpha}, \mathbf{y}^{|\beta} < \mathbf{z}^{|\beta} \right\}.$$

$$\mathbf{U} := \mathbf{V}^{-1}$$

$$\tilde{F}_{\alpha,\beta}(\mathbf{x}, \mathbf{z}) = \mathbb{P}\left(\mathbf{U} \in R(\mathbf{x}, \mathbf{z}, \alpha, \beta)\right),$$

$$g_{\alpha,\beta}(\mathbf{x}, \mathbf{z}) := \lim_{t\to\infty} \tilde{F}_{\alpha,\beta}\left(t^{-1}\mathbf{x}, t^{-1}\mathbf{z}\right)$$

$$g_{\alpha,\beta}(\mathbf{x}, \mathbf{z}) = \mu\left(R\left(\mathbf{x}^{-1}, \mathbf{z}^{-1}, \alpha, \beta\right)\right).$$

Natural empirical version $\widehat{g}_{n,\alpha,\beta}$ of $g_{\alpha,\beta}$ from (12): one recovers

$$\widehat{g}_{n,\alpha,\beta} = \widehat{\mu}_n\left(R\left(\mathbf{x}^{-1}, \mathbf{z}^{-1}, \alpha, \beta\right)\right).$$

$$\tilde{\boldsymbol{\varepsilon}}^{|\alpha} = \mathbf{1}^{|\alpha},\ \tilde{\boldsymbol{\varepsilon}}^{|\alpha^c} = \boldsymbol{\varepsilon}^{|\alpha^c}$$

$$R_\alpha^\varepsilon = R(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}, \alpha, \alpha^c) \setminus R(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\varepsilon}}, \alpha, \{1, \ldots, d\}),$$

$$\begin{aligned} |\mu - \widehat{\mu}_n|\left(R_\alpha^\varepsilon\right) &\leq |\mu - \widehat{\mu}_n|\left(R(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}, \alpha, \alpha^c)\right) + |\mu - \widehat{\mu}_n|\left(R(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\varepsilon}}, \alpha, \{1, \ldots, d\})\right) \\ &\leq 2 \max_\beta \sup_{\varepsilon \leq \mathbf{x}, \mathbf{z}} |\mu - \widehat{\mu}_n|\left(R(\mathbf{x}, \mathbf{z}, \alpha, \beta)\right). \end{aligned}$$

Scoring function motivation

$$A_{\mathbf{x}} := \left\{ \mathbf{y} \,\middle|\, T(\mathbf{y}) \in R^{\varepsilon}_{\alpha(\mathbf{x})}, \|T(\mathbf{y})\|_\infty \geq \|T(\mathbf{x})\| \right\}.$$

$$\begin{aligned}
\mathbb{P}(\mathbf{X} \in A_{\mathbf{x}}) &= \mathbb{P}(\mathbf{V} \in \|T(\mathbf{x})\|_\infty R^{\varepsilon}_{\alpha(\mathbf{x})}) \\
&= \mathbb{P}(\|\mathbf{V}\|_\infty \geq \|T(\mathbf{x})\|_\infty)\mathbb{P}(\mathbf{V} \in \|T(\mathbf{x})\|_\infty R^{\varepsilon}_{\alpha(\mathbf{x})} \mid \|\mathbf{V}\|_\infty \geq \|T(\mathbf{x})\|_\infty) \\
&= \underbrace{\mathbb{P}(\|\mathbf{U}\|_\infty \leq \|T(\mathbf{x})\|^{-1}_\infty)}_{=\|T(\mathbf{x})\|^{-1}_\infty \text{ (assumption 1)}} \underbrace{\mathbb{P}(\mathbf{V} \in \|T(\mathbf{x})\|_\infty R^{\varepsilon}_{\alpha(\mathbf{x})} \mid \|\mathbf{V}\|_\infty \geq \|T(\mathbf{x})\|_\infty)}_{\xrightarrow[\substack{\|T(\mathbf{x})\|_\infty \to \infty \\ \varepsilon \to 0}]{} \frac{\mathcal{M}(\alpha(\mathbf{x}))}{\mu([\mathbf{0},\mathbf{1}]^c)}},
\end{aligned}$$

Other numerics

- if the densities are constant, $M \leq d$
- Minimize $\frac{1}{\sqrt{\varepsilon k_n}} + d^2\varepsilon$: gives $\varepsilon = \frac{\sqrt{k_n}}{d^{\frac{4}{3}}}$

	DAMEX			IsolationForest		LocalOutlierFactor	
N	AUC ROC	AUC PR	AFD	AUC ROC	AUC PR	AUC ROC	AUC PR
80000	0.924	0.711	20.9	0.873	0.551	0.994	0.981
150000	0.906	0.639	20.7	0.890	0.600	0.994	0.981

Table 2: Results on extreme region with varying N, $k_n = n^{\frac{1}{4}} \ln(n)$, $\varepsilon = \frac{(k_n)^{\frac{1}{3}}}{d^{\frac{4}{3}}}$, $p = 0.1$

MATHÉMATIQUES VISION APPRENTISSAGE