Inference for Hidden Regular Variation in Multivariate Extremes

Grant Weller
Department of Statistics
Carnegie Mellon University

Department of Statistics Seminar
Columbia University
April 14, 2014

joint work with Dan Cooley, Colorado State University

Funding:
NSF-DMS-0905315, Weather and Climate Impacts Assessment Program (NCAR),
2011-2012 SAMSI UQ program

Multivariate extremes

The aim of a multivariate extreme value analysis is to describe the joint upper tail of a random vector.

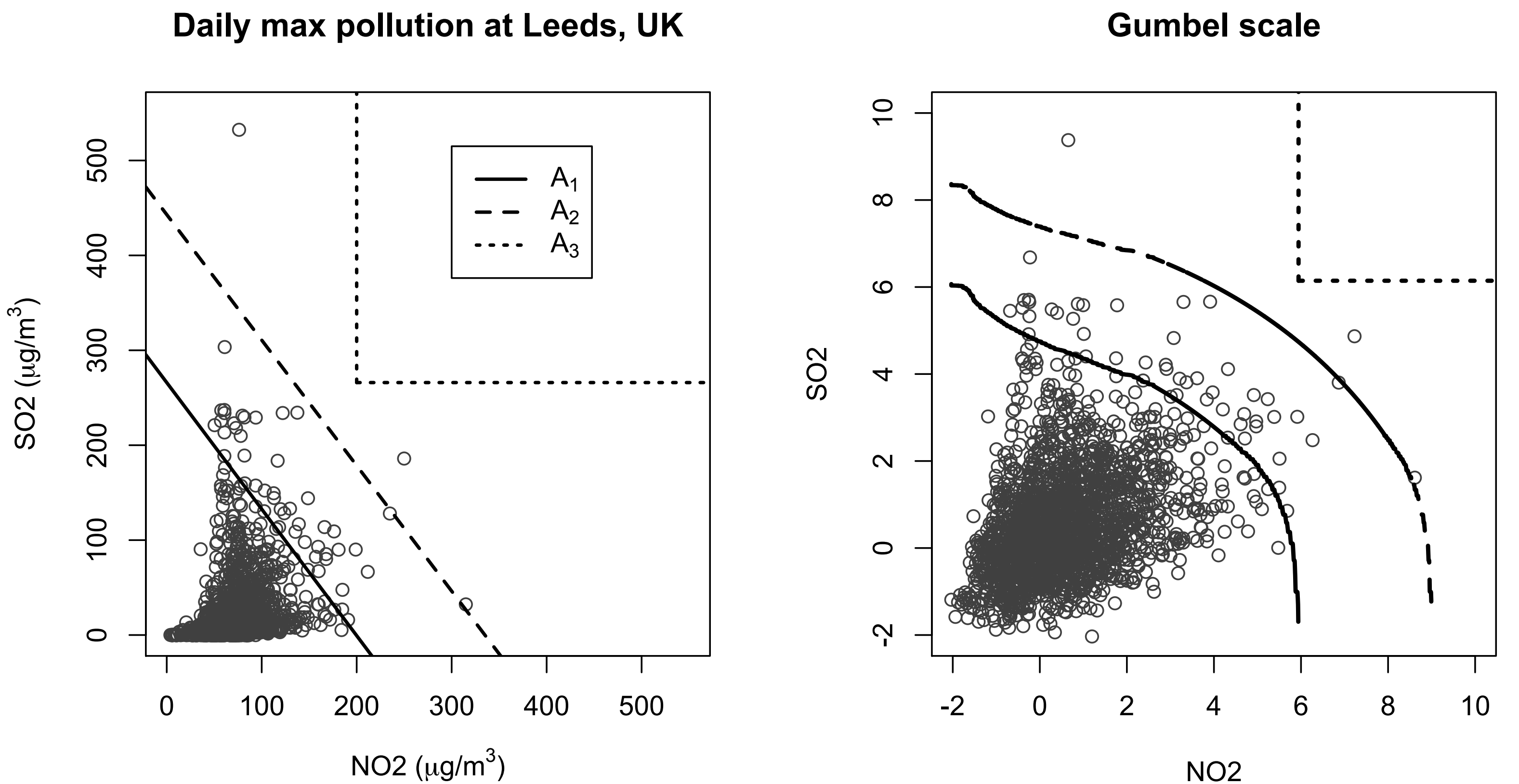
Examples:

- Estimating risk from combined effect of simultaneous extremes in multiple variables
- Modeling storms affecting several locations simultaneously
- Risk assessment in financial portfolios
- Long-range dependence in network traffic

Modeling is based on asymptotic results; *not necessary* to know the entire underlying distribution.

Motivating example

Daily air pollution measurements in Leeds, UK ($n = 2078$).



Aim: model-based estimates of risk set probabilities

Multivariate tail models

- Univariate extremes: described by 'three types'

$$F(x) = \exp\{-(1 + \xi z)^{-1/\xi}\}$$

- Multivariate: no finite parameterization
- Separation of marginal & dependence structures
- Most models assume *asymptotic dependence* of components

This talk: a model for tails when asymptotic dependence does not hold.

Tail models in practice

Marginal and dependence effects are typically handled separately.

Given iid realizations from a d-dimensional random vector $\mathbf{X}$:

1. Estimate (upper tails of) marginal distributions $X_1, ..., X_d$
2. Transform to something nice*
3. Fit dependence model

*'Nice' = regularly varying; often, unit Fréchet ($F(z) = \exp\{-z^{-1}\}$)

A bit 'copula-like'...but focuses specifically on the tail.

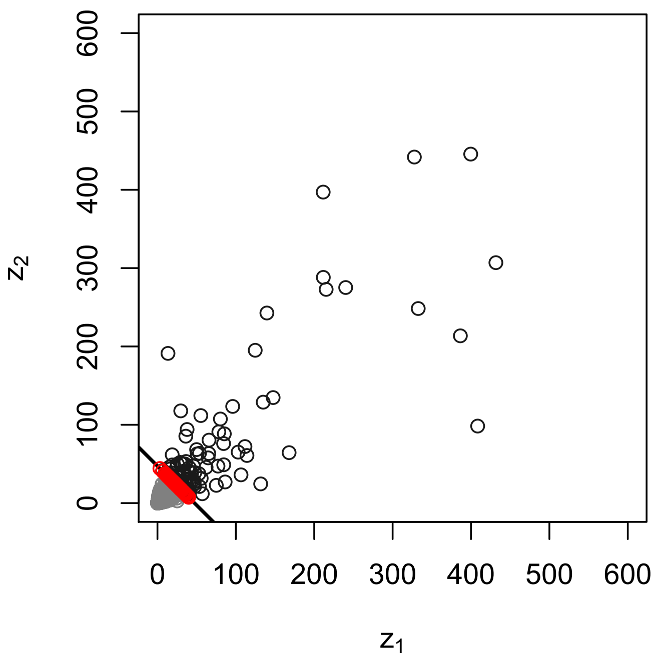
Outline

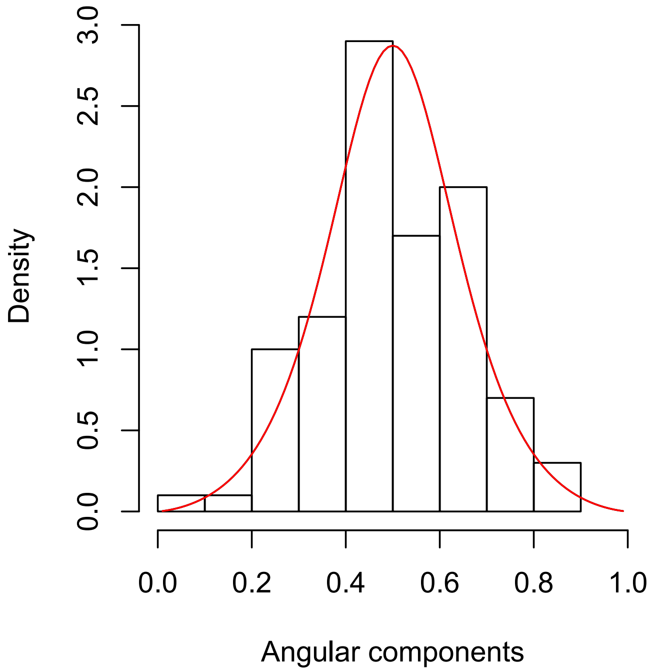
- Crash course on multivariate regular variation
- Hidden regular variation
- Sum characterization of HRV
- Inference for HRV via MCEM
- Application: air pollution data

Multivariate regular variation

Intuitive description: joint tail decay like a power function.

- Spectral decomposition
- Tail dependence described by an *angular measure*





MRV definition 1 (Resnick, 2007)

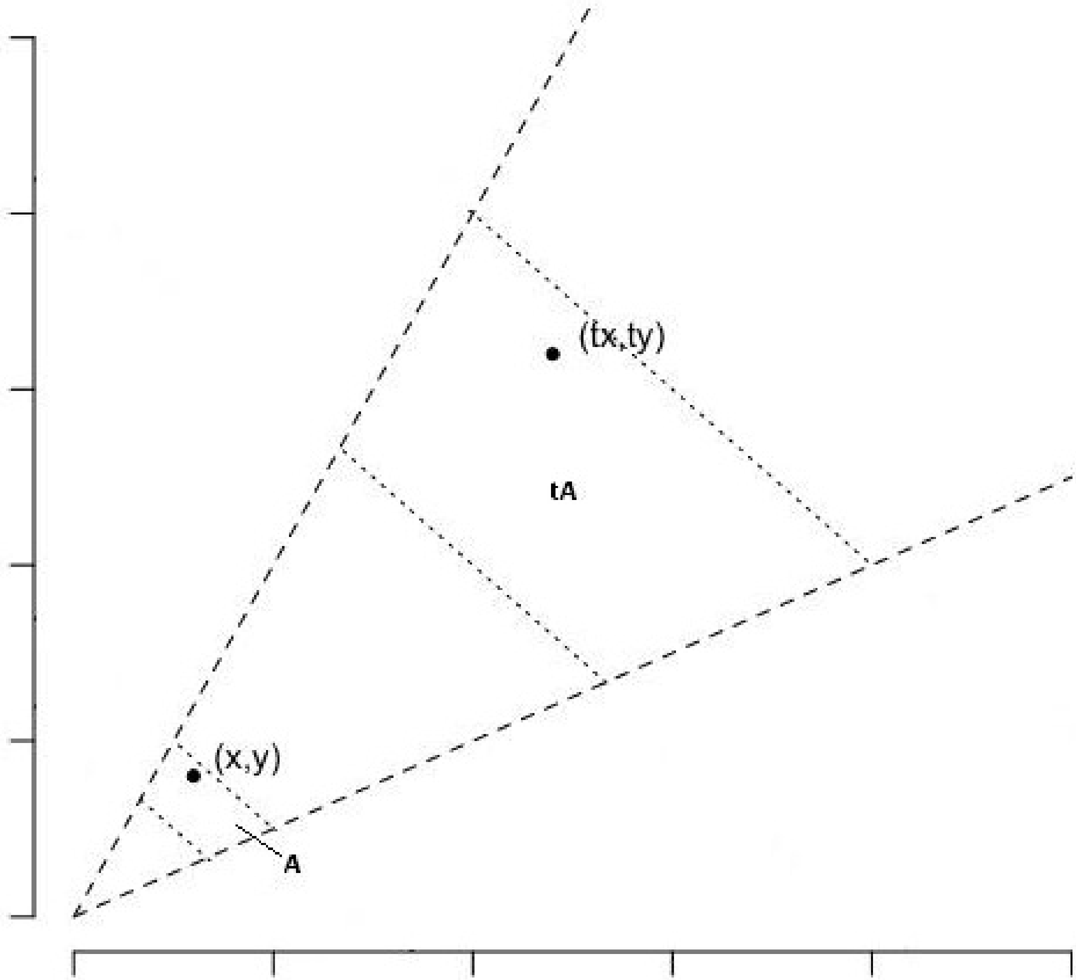
A random vector $\mathbf{Z}$ is regular varying on the cone $\mathfrak{C} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$ if there exists a normalizing sequence $\{a_n\}_{n=1}^{\infty}$ with $a_n \to \infty$ such that

$$n\mathbb{P}\left(\frac{\mathbf{Z}}{a_n} \in \cdot\right) \xrightarrow{v} \nu(\cdot)$$

as $n \to \infty$ in $M_+(\mathfrak{C})$, where $\xrightarrow{v}$ denotes vague convergence of measures.

- Scaling property: $\nu(tA) = t^{-\alpha}\nu(A)$ for $t > 0$, where α is called the *tail index*
- Extremes of the multivariate sample occur according to the limiting measure ν
- a_n is regular varying of index $1/\alpha$ (i.e. $a_n \sim n^{1/\alpha}$)

If marginals are unit Fréchet, $\alpha = 1$.



$$\nu(tA) = t^{-\alpha}\nu(A)$$

MRV definition 2: the angular measure

Define 'radial' and 'angular' components $R = \|\mathbf{Z}\|$, $\mathbf{W} = \mathbf{Z}\|\mathbf{Z}\|^{-1}$, where $\|\cdot\|$ is any norm on $\mathfrak{C}$.

The regular variation condition can then be written

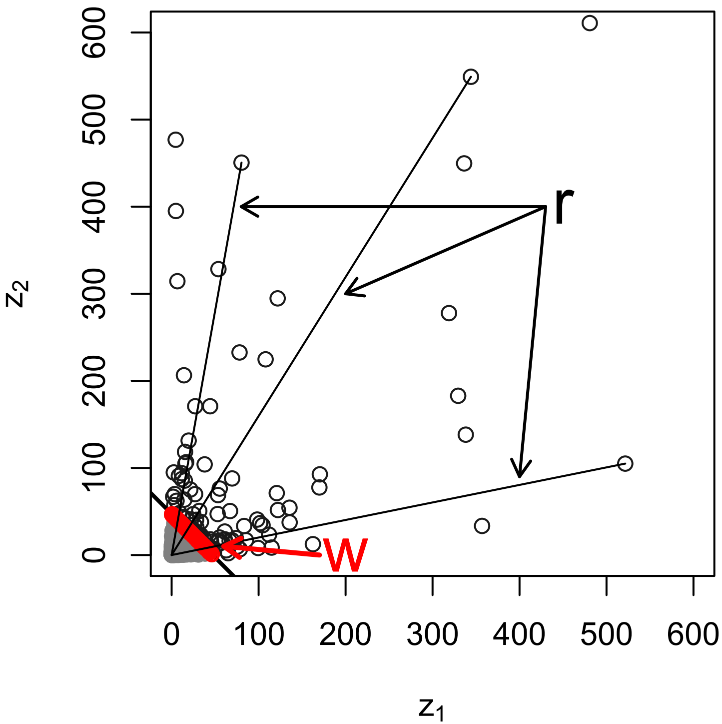
$$n\mathbb{P}\left(\frac{R}{a_n} > r, \mathbf{W} \in B\right) \xrightarrow{v} r^{-\alpha} H(B)$$

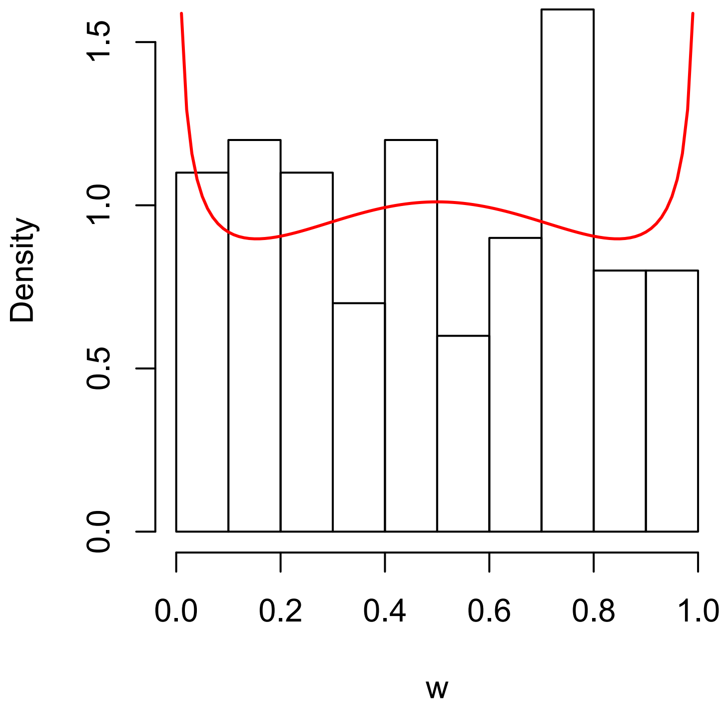
for any Borel set $B \in \mathcal{N} = \{\mathbf{z} \in \mathfrak{C} : \|\mathbf{z}\| = 1\}$.

- H is a measure on $\mathcal{N}$ which characterizes tail dependence
- By choice of a_n, H can be made to be a probability measure
- Equivalent convergence of Poisson point process

Loosely, $(R, \mathbf{W}) \sim r^{-1-\alpha} H(d\mathbf{w})$ for large r.

Radial and angular components





Likelihood inference

For a fixed sample of size n, assume

$$n\mathbb{P}\left(\frac{R}{a_n} > r, \mathbf{W} \in \cdot\right) \cong r^{-\alpha} H(\cdot)$$

for $r > r_0$ (large).

Likelihood:

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{z}_1, \ldots, \mathbf{z}_n) \propto \exp\{-r_0^{-\alpha}\} \prod_{i=1}^{N_0} r_i^{-(1+\alpha)} h(\mathbf{w}_i; \boldsymbol{\theta})$$

where $r_i = \|\mathbf{z}_i\|$ and $\mathbf{w}_i = \mathbf{z}_i \|\mathbf{z}_i\|^{-1}$, for the points $\mathbf{z}_1, \ldots, \mathbf{z}_{N_0}$ with $r_i > r_0$.

Parameters can be estimated via numerical optimization.

e.g. (*Coles and Tawn, 1991; Cooley et al., 2010; Ballani and Schlather, 2011.*)

Outline

- Crash course on Multivariate regular variation
- Hidden regular variation
- Sum characterization of HRV
- Estimation for HRV via MCEM
- Application: air pollution data

When multivariate regular variation fails

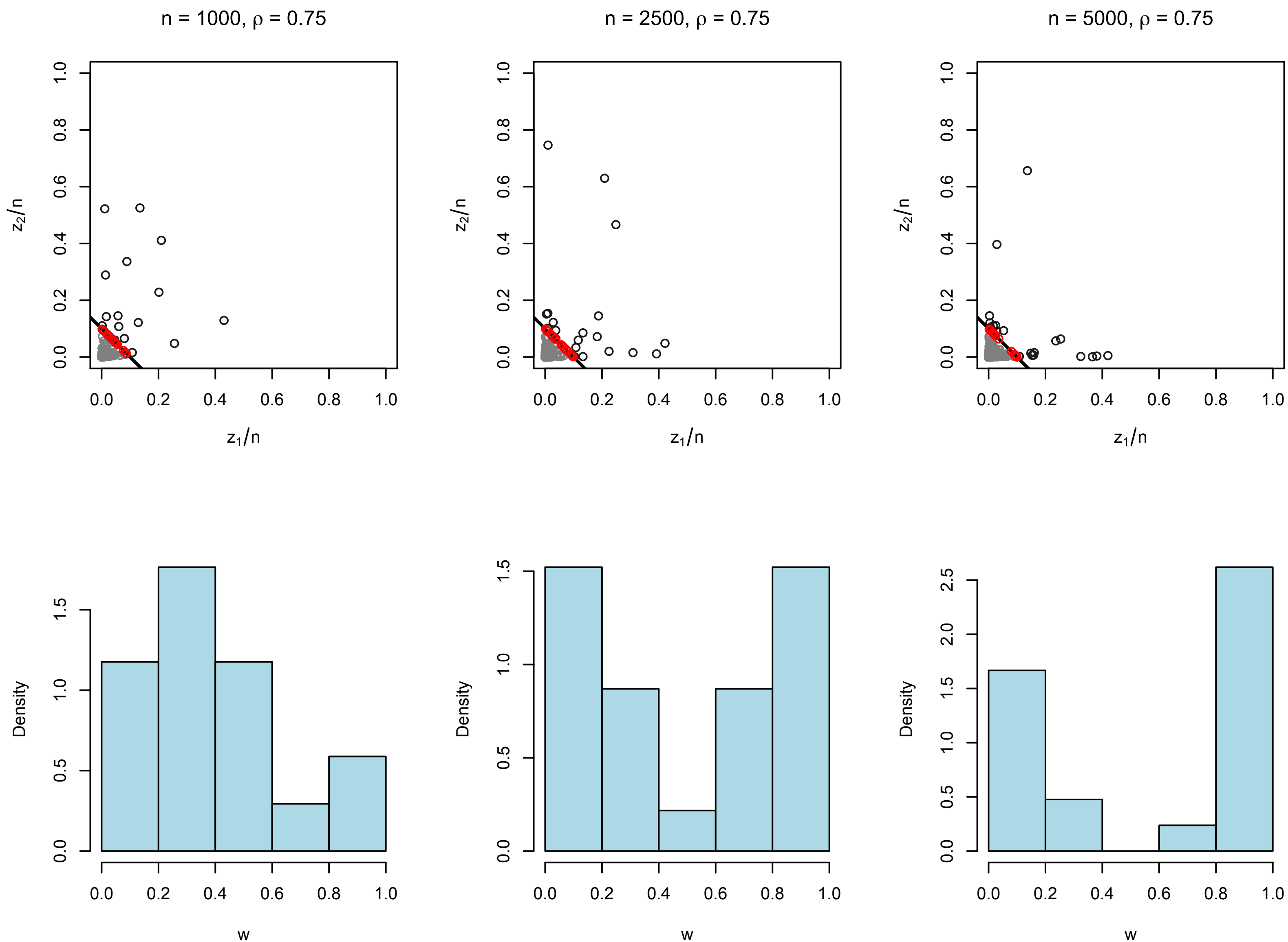
In some cases, H places zero mass on regions of $\mathcal{N}$.

Example: asymptotic independence in $d = 2$:

$$\lim_{z \to \infty} \mathbb{P}(Z_2 > z \mid Z_1 > z) = 0.$$

- H consists of point masses at $\{0\}$ and $\{1\}$ (under $\|\cdot\|_1$)
- e.g. bivariate Gaussian with correlation $\rho < 1$

Normalization by a_n kills off sub-asymptotic dependence structure.



n = 1000, ρ = 0.75
n = 2500, ρ = 0.75
n = 5000, ρ = 0.75
z2/n
z1/n
Density
w
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.5
2.0
2.5

Statistical issues

- Taking the limiting form as exact equality assumes

$$\mathbb{P}(a_n^{-1} Z_1 > z_1, a_n^{-1} Z_2 > z_2) = 0,$$

  a potentially dangerous assumption in practice.

- On the other hand, incorrectly assuming asymptotic dependence will result in overestimation of joint tail risks

Need to account for positive dependence in the presence of asymptotic independence.

Hidden regular variation (Resnick, 2002)

A regular varying random vector $\mathbf{Z}$ exhibits hidden regular variation on a subcone $\mathfrak{C}_0 \subset \mathfrak{C}$ if $\nu(\mathfrak{C}_0) = 0$ and there exists a sequence $b_n \to \infty$ with $b_n/a_n \to 0$ s.t.

$$n\mathbb{P}\left(\frac{\mathbf{Z}}{b_n} \in \cdot\right) \xrightarrow{v} \nu_0(\cdot)$$

as $n \to \infty$ in $M_+(\mathfrak{C}_0)$.

- Scaling: $\nu_0(tA) = t^{-\alpha_0}\nu_0(A)$ for measurable $A \in \mathfrak{C}_0$, $\alpha_0 \geq \alpha$
- ν_0 is Radon but *not necessarily finite*

Equivalently,

$$t\mathbb{P}\left(\frac{R}{b_n} > r, \mathbf{W} \in B\right) \xrightarrow{v} r^{-\alpha_0}H_0(B)$$

for B a Borel set of $\mathcal{N}_0 = \mathfrak{C}_0 \cap \mathcal{N}$ (e.g. $\mathcal{N}_0 = (0, 1)$).

H_0 is called the *hidden angular measure*.

Example: bivariate Gaussian

Consider $\mathbf{Z}$ with Fréchet margins and Gaussian dependence, $\rho \in (-1, 1)$. Recall ν places mass only on the axes of $\mathfrak{C}$.

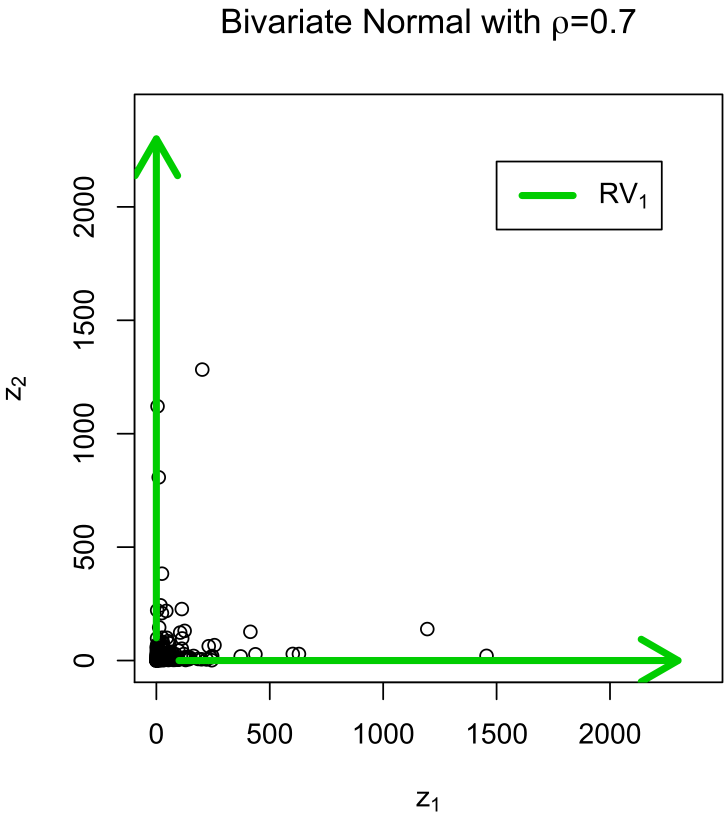
Define $\eta = (1+\rho)/2$, the *coefficient of tail dependence* (Ledford and Tawn, 1997).

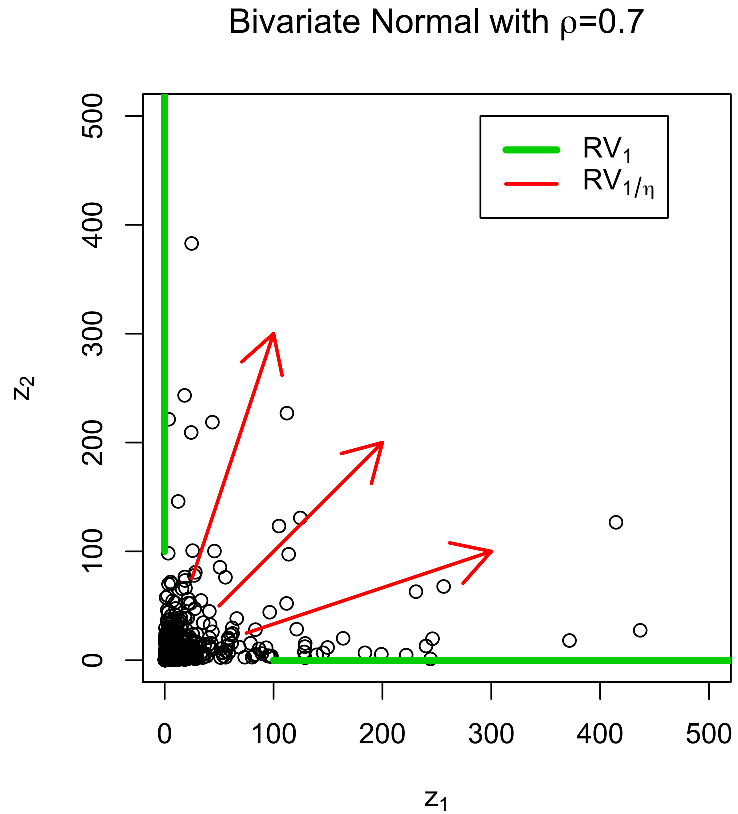
- $\mathbf{Z}$ exhibits hidden regular variation of order $\alpha_0 = 1/\eta$
- The density of the hidden measure ν_0 can be written

$$\nu_0(dr \times dw) = \frac{1}{\eta} r^{-(1+1/\eta)} dr \times \underbrace{\frac{1}{4\eta}\{w(1-w)\}^{-1/2\eta-1} dw}_{H_0(dw)}$$

H_0 is infinite on $(0, 1)$.

Example: bivariate Gaussian





Goal: a model for hidden regular variation.

Characterizations of HRV

How do we generate multivariate models with hidden regular variation?

- Mixture method (Maulik & Resnick, 2005)
- Maxima method (de Haan & Zhou, 2011)
- This work: additive method

Our focus: finite-sample inference

Tail equivalence (Maulik and Resnick, 2004)

Two random vectors $\mathbf{X}$ and $\mathbf{V}$ are *tail equivalent* on the cone $\mathfrak{C}^*$ if

$$n\mathbb{P}\left(\frac{\mathbf{X}}{b_n^*} \in \cdot\right) \xrightarrow{v} \nu_*(\cdot) \qquad \text{and} \qquad n\mathbb{P}\left(\frac{\mathbf{V}}{b_n^*} \in \cdot\right) \xrightarrow{v} c\nu_*(\cdot)$$

as $n \to \infty$ in $M_+(\mathfrak{C}^*)$ for $c > 0$.

'Extremes of $\mathbf{X}$ and $\mathbf{V}$ samples taken in $\mathfrak{C}^*$ will have the same asymptotic properties.'

Outline

- Crash course on multivariate regular variation
- Hidden regular variation
- Sum characterization of HRV
- Inference for HRV via MCEM
- Application: air pollution data

Sum characterization of HRV

Suppose $\mathbf{Z}$ is regular varying on $\mathfrak{C}$ with hidden regular variation on $\mathfrak{C}_0$:

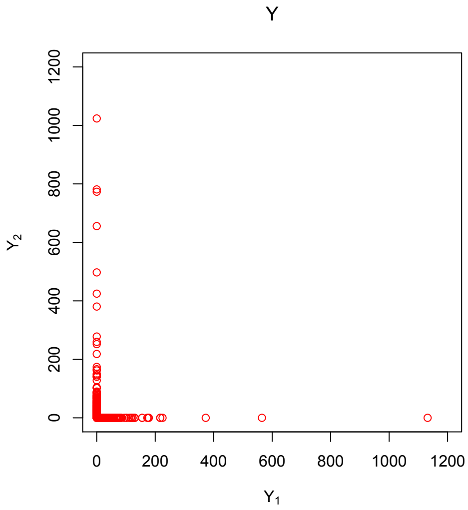
$$n\mathbb{P}\left(\frac{\mathbf{Z}}{a_n} \in \cdot\right) \xrightarrow{v} \nu(\cdot) \qquad \text{in } M_+(\mathfrak{C}) \qquad \text{and}$$

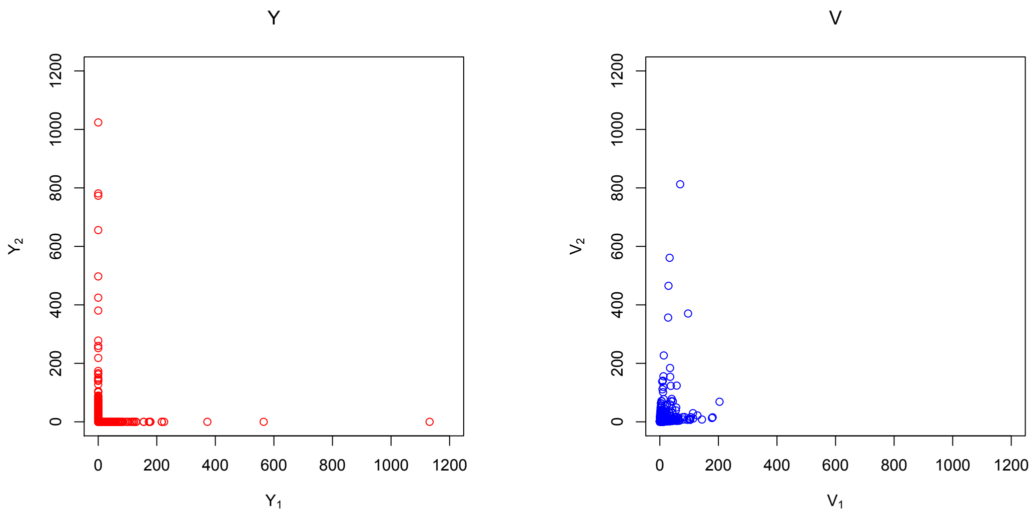
$$n\mathbb{P}\left(\frac{\mathbf{Z}}{b_n} \in \cdot\right) \xrightarrow{v} \nu_0(\cdot) \qquad \text{in } M_+(\mathfrak{C}_0)$$

with $\nu(\mathfrak{C}_0) = 0$ and $b_n/a_n \to 0$ as $t \to \infty$.

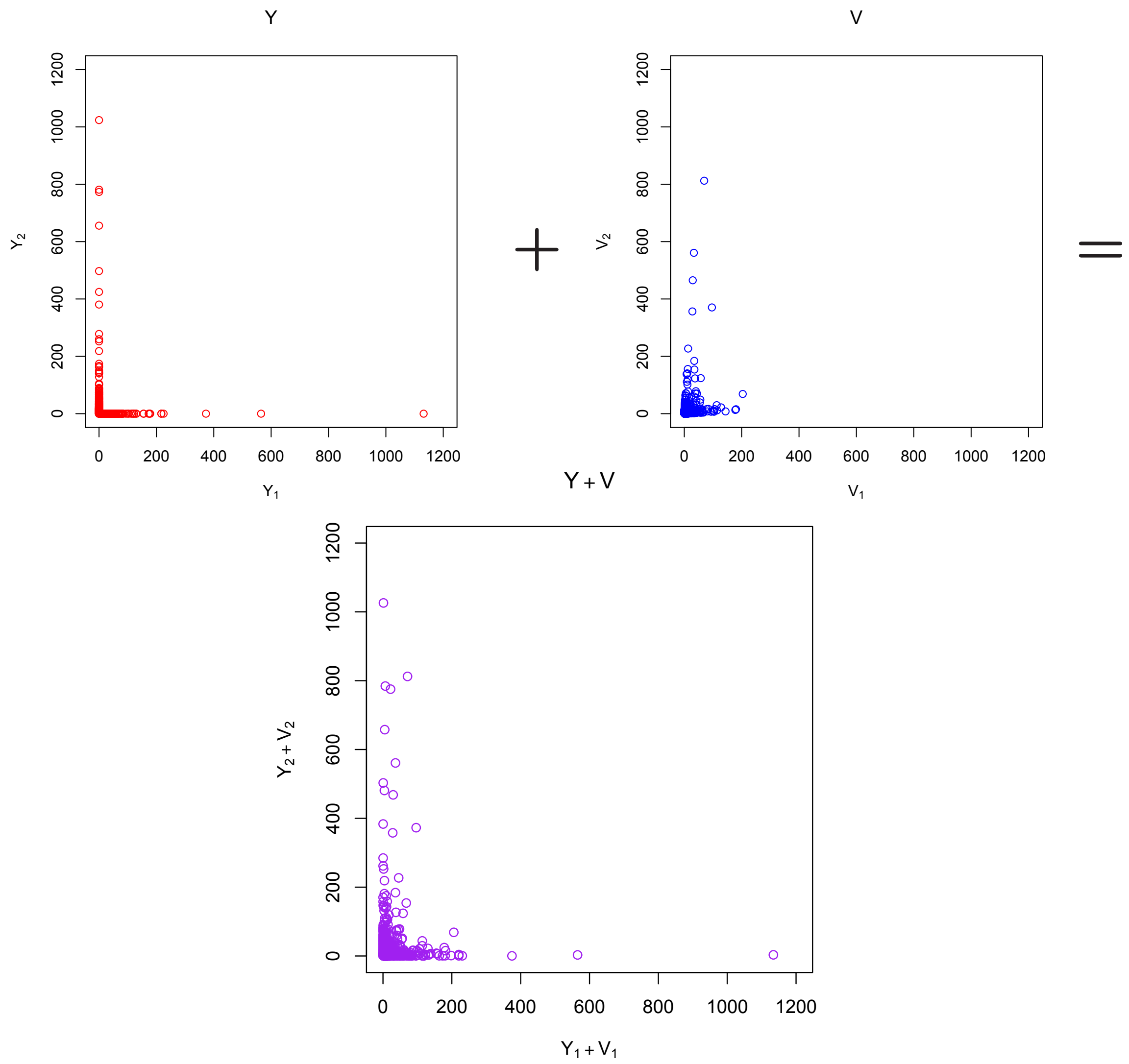
Aim: representation for $\mathbf{Z}$ which captures tail behavior on both $\mathfrak{C}$ and $\mathfrak{C}_0$ and from which an inference procedure may be developed.

We propose the sum $\mathbf{Y} + \mathbf{V}$.


Y
Y2
Y1
0
200
400
600
800
1000
1200



Y
V
Y_2
Y_1
V_2
V_1
0
200
400
600
800
1000
1200



No point falls exactly on an axis.

Construction of $\mathbf{Y} + \mathbf{V}$

Define $\mathbf{Y} = R\mathbf{W}$, with $\mathbb{P}(R > a_n) \sim 1/n$ and $\mathbf{W}$ drawn from limiting angular measure H. *Notice that* $\mathbf{Y}$ *has support only on* $\mathfrak{C} \setminus \mathfrak{C}_0$*.*

Let $\mathbf{V} \in [0, \infty)^d$ be regular varying on $\mathfrak{C}_0$ with limit measure ν_0:

$$n\mathbb{P}\left(\frac{\mathbf{V}}{b_n} \in \cdot\right) \xrightarrow{v} \nu_0(\cdot) \quad \text{in } M_+(\mathfrak{C}_0).$$

Further assume that on $\mathfrak{C}$,

$$\mathbb{P}(\|\mathbf{V}\| > r) \sim cr^{-\alpha^*}$$

as $r \to \infty$, with $c > 0$ and

$$\alpha^* > \alpha \vee (\alpha_0 - \alpha).$$

Assume R, $\mathbf{W}$, $\mathbf{V}$ are independent.

Tail equivalence result

Then

$$n\mathbb{P}\left(\frac{\mathbf{Y}+\mathbf{V}}{a_n}\in\cdot\right)\xrightarrow{v}\nu(\cdot)\text{ in }M_+(\mathfrak{C})$$

(Jessen and Mikosch, 2006).

Furthermore, tail equivalence also holds on $\mathfrak{C}_0$:

Theorem. *With* $\mathbf{Y}$ *and* $\mathbf{V}$ *as defined above,*

$$n\mathbb{P}\left(\frac{\mathbf{Y}+\mathbf{V}}{b_n}\in\cdot\right)\xrightarrow{v}\nu_0(\cdot)\textit{ in }M_+(\mathfrak{C}_0).$$

View $\mathbf{Z}$ as a sum of ‘first-order’ $\mathbf{Y}$ and ‘second-order’ $\mathbf{V}$.

The sum $\mathbf{Y}+\mathbf{V}$ is *tail equivalent* to $\mathbf{Z}$ on *both* $\mathfrak{C}$ *and* $\mathfrak{C}_0$.

Conditions on $\mathbf{V}$

We require two identifiability conditions on the tail index α^* of $\mathbf{V}$ on the full cone $\mathfrak{C}$:

- $\alpha^* > \alpha$ identifies the regular variation of $\mathbf{Z}$ as that of $\mathbf{Y}$
- $\alpha^* > \alpha_0 - \alpha$ ensures that the HRV of $\mathbf{Z}$ is the same as $\mathbf{V}$

Das and Resnick (2014): what happens for "$=$" or "$<$" in second item.

No known case of HRV with *infinite measure* ν_0 which satisfies both...

Infinite measure example: bivariate Gaussian

$\mathbf{Z}$ has Fréchet margins and Gaussian dependence ($\rho < 1$).
Recall: H_0 is *infinite* on $\mathcal{N}_0 = (0, 1)$.

Poses difficulty near the axes of $\mathfrak{C}$.

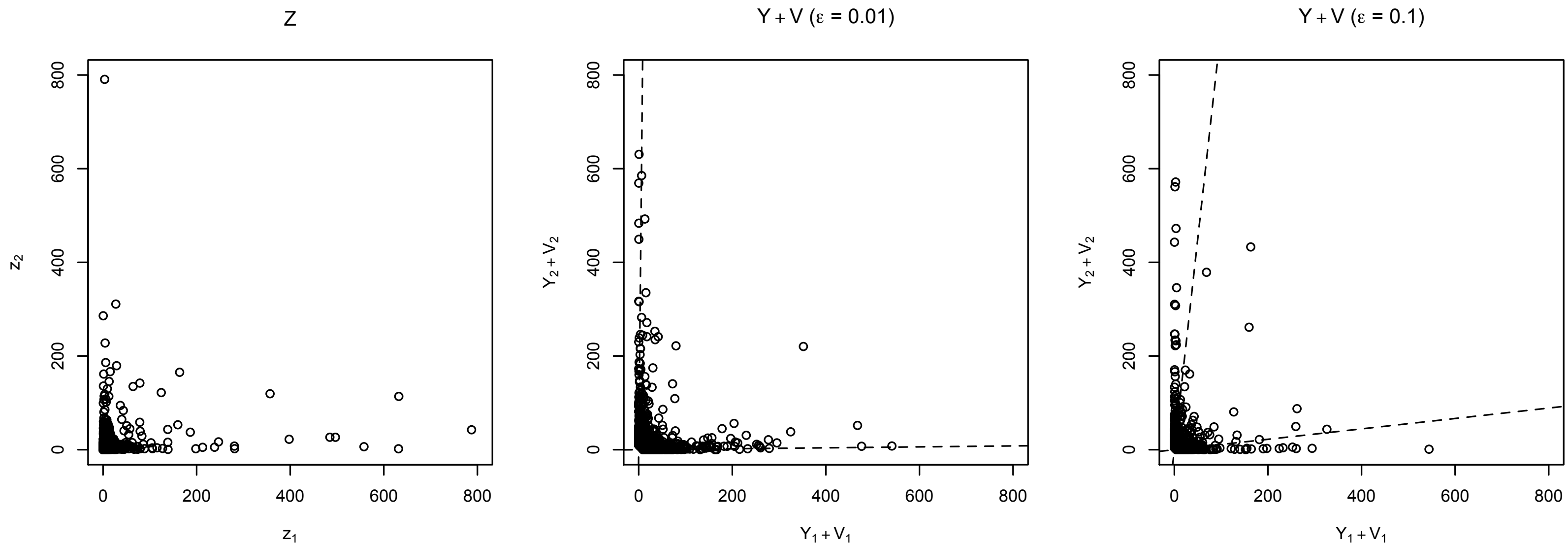
Proposed construction of $\mathbf{V}$:

- Restrict to $\mathfrak{C}_0^\epsilon = \mathfrak{C}_0 \cap \mathcal{N}_0^\epsilon$, where $\mathcal{N}_0^\epsilon = [\epsilon, 1 - \epsilon]$ for $\epsilon \in (0, 1/2)$.
- Simulate W_0 from probability density $H_0(dw)/H_0(\mathcal{N}_0^\epsilon)$
- Let R_0 follow a Pareto distribution with $\alpha = 1/\eta$
- $\mathbf{V} = [R_0 W_0, R_0(1 - W_0)]^T$

$\mathbf{Y} + \mathbf{V}$ is tail equivalent to $\mathbf{Z}$ on $\mathfrak{C}$ and $\mathfrak{C}_0^\epsilon$.

Sum representation of bivariate Gaussian

Example with $\rho = 0.5$ $(n = 2500)$:



For any set completely contained in $\mathfrak{C}_0^\epsilon$ we achieve the correct limit measure ν_0.

Choice of ϵ involves a trade-off between:

- Size of the subcone on which tail equivalence holds
- Threshold at which $\mathbf{Y} + \mathbf{V}$ is a useful approximation

(*Weller and Cooley, 2014*)

Outline

- Crash course on multivariate regular variation
- Hidden regular variation
- Sum characterization of HRV
- Inference for HRV via MCEM
- Application: air pollution data

EM algorithm

Observe realizations from $\mathbf{Z}$, tail equivalent to $\mathbf{Y}+\mathbf{V}$. Assume parametric forms and perform ML inference via EM.

If we assume $\mathbf{Z} = \mathbf{Y} + \mathbf{V}$,

$$\begin{aligned}\log f(\mathbf{z};\boldsymbol{\theta}) &= \int \log f(\mathbf{z},\mathbf{y},\mathbf{v};\boldsymbol{\theta}) f(\mathbf{y},\mathbf{v} \mid \mathbf{z};\boldsymbol{\theta}^{(k)}) d\mathbf{y} d\mathbf{v} \\ &\quad - \int \log f(\mathbf{y},\mathbf{v} \mid \mathbf{z};\boldsymbol{\theta}) f(\mathbf{y},\mathbf{v} \mid \mathbf{z};\boldsymbol{\theta}^{(k)}) d\mathbf{y} d\mathbf{v} \\ &:= Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) - H(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}).\end{aligned}$$

- EM constructs and maximizes Q
- MLE obtained as long as H is maximized at $\boldsymbol{\theta}^{(k)}$

Here: $\mathbf{Z}$ and $\mathbf{Y}+\mathbf{V}$ are only *tail equivalent*; $\boldsymbol{\theta}$ governs tail behavior of $\mathbf{Y}$ and $\mathbf{V}$. Requires a modification of the EM setup.

EM for extremes

Consider distributions with densities $g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ and $g_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta})$ which are tail equivalent to the true distributions; i.e.,

$$\begin{aligned} g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) &\cong f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \quad \text{for} \quad \|\mathbf{y}\| > r^*_{\mathbf{Y}} \\ g_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta}) &\cong f_{\mathbf{V}}(\mathbf{v};\boldsymbol{\theta}) \quad \text{for} \quad \|\mathbf{v}\| > r^*_{\mathbf{V}}, \end{aligned}$$

Complete likelihood is based on limiting Poisson point processes for $\mathbf{Y}$ and $\mathbf{V}$.

- E step: expectation is taken with respect to $g(\mathbf{y},\mathbf{v} \mid \mathbf{z};\boldsymbol{\theta}^{(k)})$.
- M step: maximization is taken over only 'large' $\mathbf{y}$ and $\mathbf{v}$.

We show

$$H(\boldsymbol{\theta}^{(k)} \mid \boldsymbol{\theta}^{(k)}) - H(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) \geq 0$$

using Jensen's inequality.

Implementation via MCEM

Q is not available in closed form.

At the E step of the $(k+1)^{th}$ iteration, simulate from

$$g_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}^{(k)})g_{\mathbf{V}}(\mathbf{z}-\mathbf{y};\boldsymbol{\theta}^{(k)}) \propto g(\mathbf{y},\mathbf{v} \mid \mathbf{z};\boldsymbol{\theta}^{(k)})$$

for all $\mathbf{z}$ and use the simulated realizations to compute

$$\widehat{Q}_m(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \frac{1}{m}\sum_{j=1}^{m} \ell(\boldsymbol{\theta};\mathbf{z},\mathbf{y}_j,\mathbf{v}_j).$$

employing Poisson point process likelihoods for *large* realizations of $\mathbf{Y}$ and $\mathbf{V}$.

Key idea: likelihood only depends on $\boldsymbol{\theta}$ for 'large' $\mathbf{y}$ and $\mathbf{v}$!

Uncertainty estimates obtained via Louis' method.

Simulation study: bivariate Gaussian

Simulate 10^4 realizations from a bivariate Gaussian distribution with correlation ρ, transform marginals to unit Fréchet.

Tail equivalent on $\mathfrak{C}$ and $\mathfrak{C}_0^\epsilon$ to $\mathbf{Y}+\mathbf{V}$, where $\mathbf{V}$ has angular measure

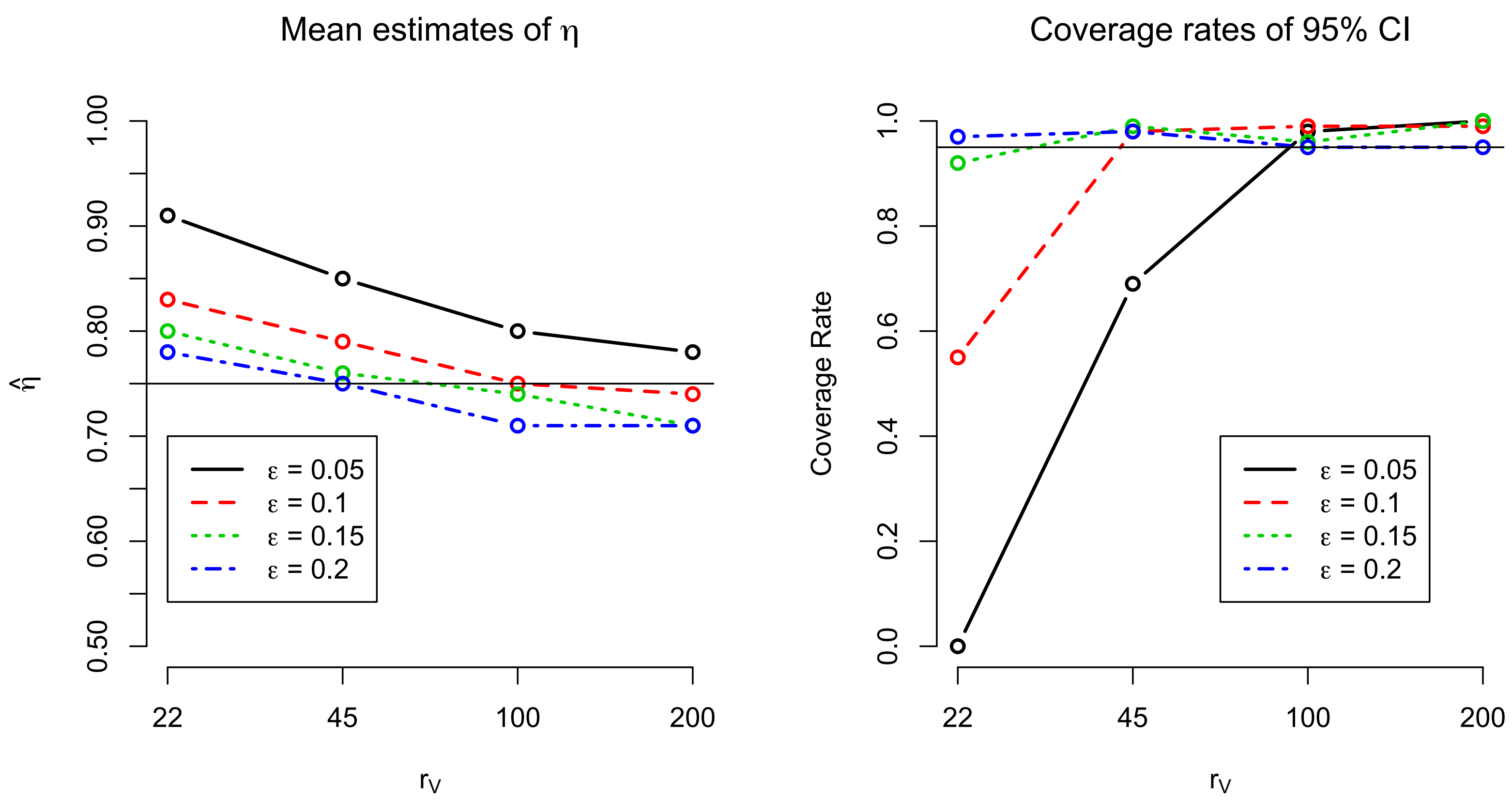
$$H_0(dw) = \frac{1}{4\eta}\{w(1-w)\}^{-1/2\eta-1}dw.$$

Aim: estimate $\eta = (1+\rho)/2$ from the ϵ-restricted model.

- Must select both ϵ and $r^*_{\mathbf{V}}$
- Trade-off in finite sample estimation problems
- b_n also unknown – incorporated as a scale parameter

Results

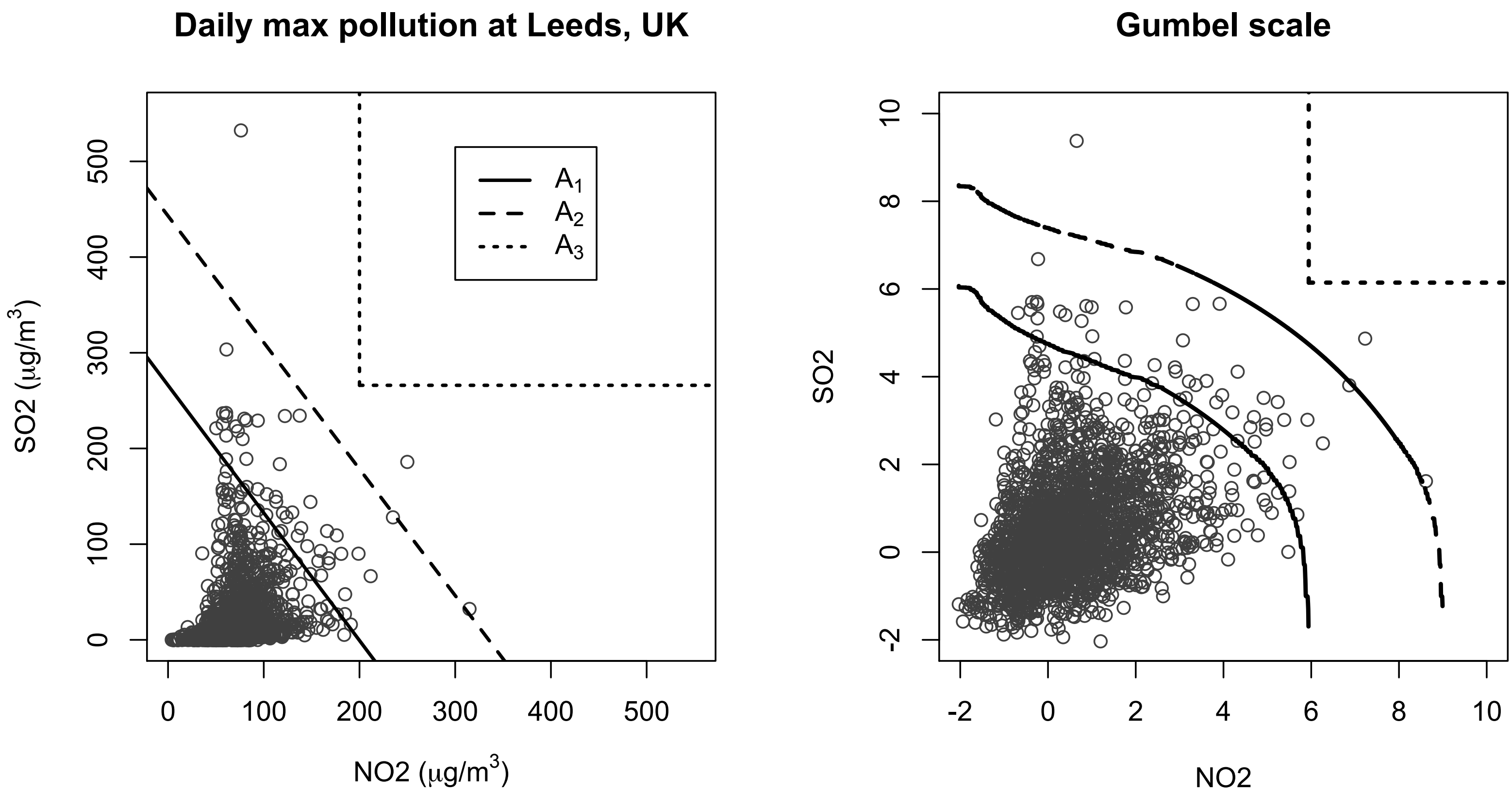
Shown for $\eta = 0.75$ ($\rho = 0.5$), based on 200 replicates



Outline

- Crash course on multivariate regular variation
- Hidden regular variation
- Sum characterization of HRV
- Inference for HRV via MCEM
- Application: air pollution data

Air pollution data



- Aim: estimate risk set probabilities
- Heffernan & Tawn (2004): asymptotic independence

Competing approaches

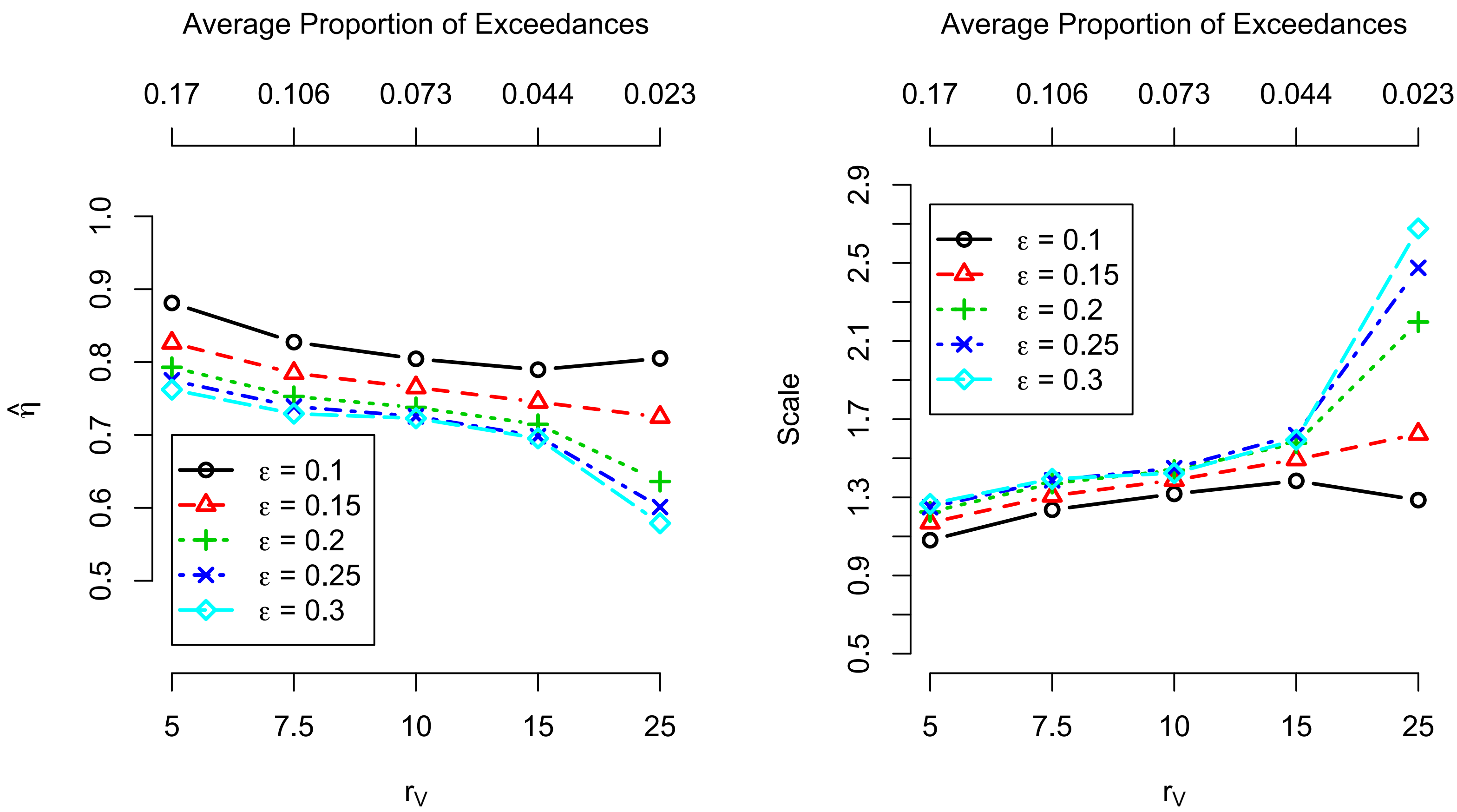
Examine three modeling approaches:

1. Fit the $\mathbf{Y}+\mathbf{V}$ model with ϵ-restricted infinite hidden measure model via proposed MCEM procedure.

2. Assume asymptotic dependence with bivariate logistic angular dependence. Fit to largest 10% of observations in terms of L_1 norm. Estimate $\widehat{\beta} = 0.72$.

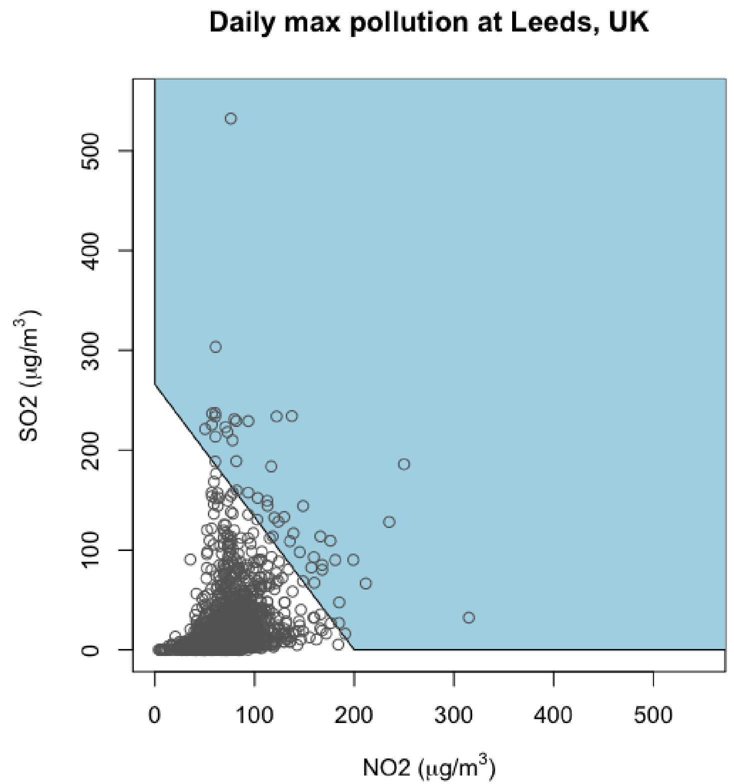
3. Conditional model of Heffernan & Tawn (2004).

Diagnostics



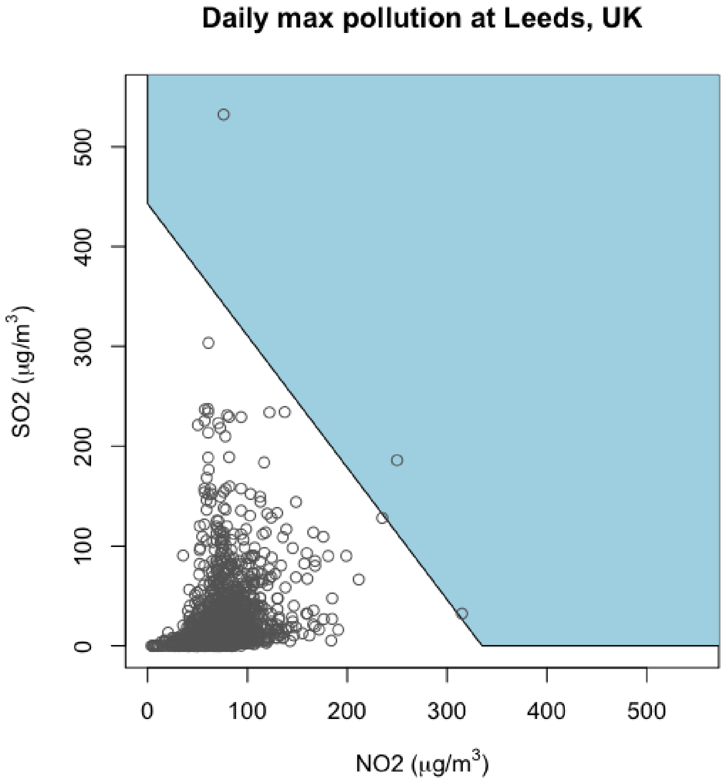
Select $\epsilon = 0.2$ and $r_V^* = 7.5$. Estimate $\hat{\eta} = 0.753$.

Results - risk set estimates



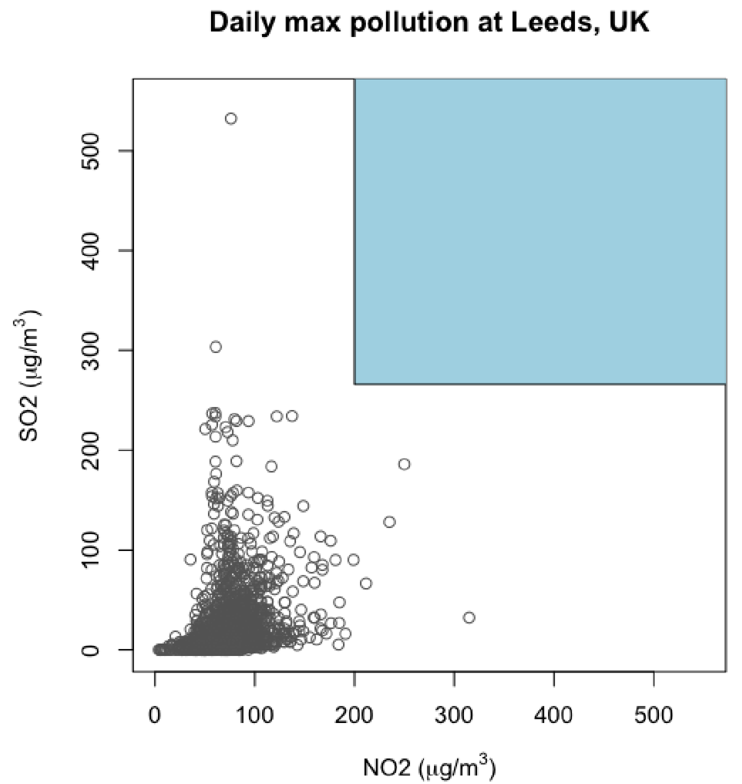
Model	$\widehat{\mathbb{P}} \times 100$	#expected	p-val
1 $(\mathbf{Y}+\mathbf{V})$	2.07	43.1	0.13
2 (asy. dep.)	2.52	52.3	0.41
3 (H & T)	2.26	47.0	0.35
Empirical	2.41	50	–

Results - risk set estimates



Model	$\widehat{\mathbb{P}} \times 100$	#expected	p-val
1 ($\mathbf{Y} + \mathbf{V}$)	0.14	2.71	0.51
2 (asy. dep.)	0.35	7.22	0.07
3 (H & T)	0.15	3.06	0.64
Empirical	0.14	3	–

Results - risk set estimates



Model	$\widehat{\mathbb{P}} \times 100$	#expected	p-val
1 $(\mathbf{Y}+\mathbf{V})$	0.012	0.25	0.78
2 (asy. dep.)	0.084	1.75	0.17
3 (H & T)	0.005	0.10	0.90
Empirical	0	0	–

Summary

This work introduces a sum representation for regular varying random vectors possessing hidden regular variation.

- Useful for finite sample simulation & estimation
- Asymptotically justified by tail equivalence result
- Difficulty arises when ν_0 is infinite
  - Our fix: restrict ν_0 to a compact cone
  - Others: de Haan & Zhou (2011), Mitra & Resnick (2010)
- Likelihood estimation via modified MCEM algorithm
- Captures tail dependence in the presence of asymptotic independence

Ongoing Work

Model selection procedures:

- Asymptotic dependence at weak level vs.
- Asymptotic independence + hidden regular variation

Extension of ideas to spatial settings - lattice data.

References

Das, B. and Resnick, S. I. (2014). Generation and detection of multivariate regular variation and hidden regular variation. *arXiv preprint arXiv:1403.5774*.

de Haan, L. and Zhou, C. (2011). Extreme residual dependence for random vectors and processes. *Adv. Appl. Prob.*, 43(1):217–242.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc.* , B 66(3):497–546.

Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math.*, 80(94):171–192.

Maulik, K. and Resnick, S. I. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Mitra, A. and Resnick, S. I. (2010). Hidden regular variation and detection of hidden risks. *Stoch. Models*, 27:591–614.

Ramos, A. and Ledford, A. W. (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc.*, B 71(1):219–241.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Weller, G. B. and Cooley, D. (2014). A sum characterization of hidden regular variation with likelihood inference via expectation–maximization. *Biometrika*, 101(1):17–36.