Going beyond the Hill: An introduction to Multivariate Extreme Value Theory

Philippe Naveau naveau@lsce.ipsl.fr

Laboratoire des Sciences du Climat et l'Environnement (LSCE)
Gif-sur-Yvette, France

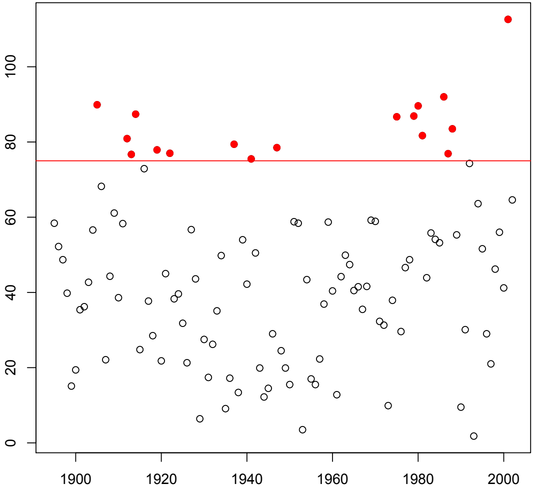
Books: Coles (1987), Embrechts et al. (1997), Resnick (2006)

FP7-ACQWA, GIS-PEPER, MIRACLE & ANR-McSim, MOPERA

25 juin 2015

EVT = Going beyond the data range

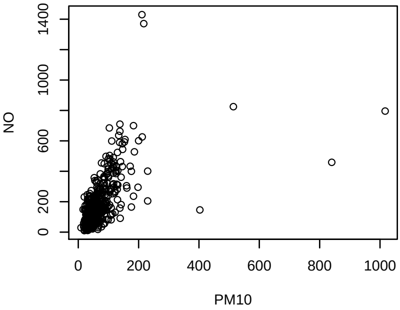
What is the probability of observing intensities above an high threshold ?

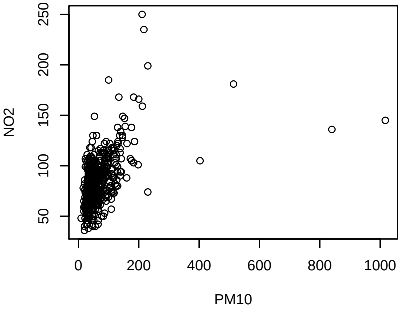


March precipitation amounts recorded at Lille (France) from 1895 to 2002. The 17 black dots corresponds to the number of excesses above the threshold $u_n = 75$ mm. This number can be conceptually viewed as a random sum of Bernoulli (binary) events.

An example in three dimensions

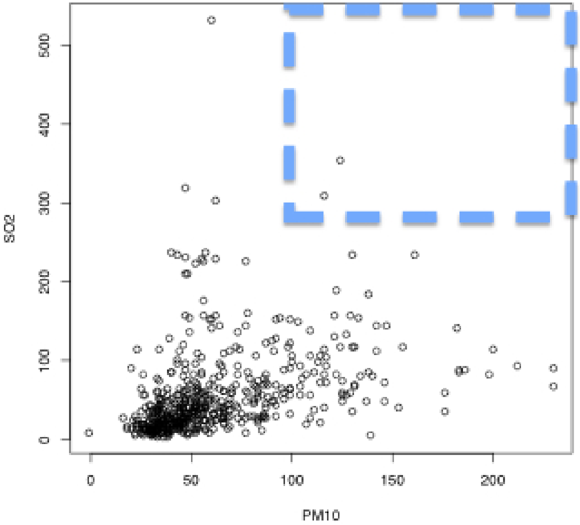
Air pollutants (Leeds, UK, winter 94-98, daily max) NO vs. PM10 (left), SO2 vs. PM10 (center), and SO2 vs. NO (right) (Heffernan& Tawn 2004, Boldi & Davison, 2007)



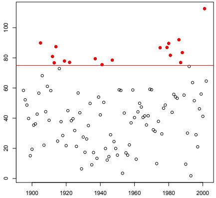


Typical question in multivariate EVT

What is the probability of observing data in the blue box ?



Siméon Denis Poisson (1781-1840)



Counting excesses

As a sum of random binary events, the variable N_n that counts the number of events above the threshold u_n has mean $n\,Pr(X > u_n)$

Poisson's theorem [1] in 1837

If u_n such that

$$\lim_{n \to \infty} n\,Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n follows approximately a **Poisson variable** N.

1. Give HW

Poisson and maxima

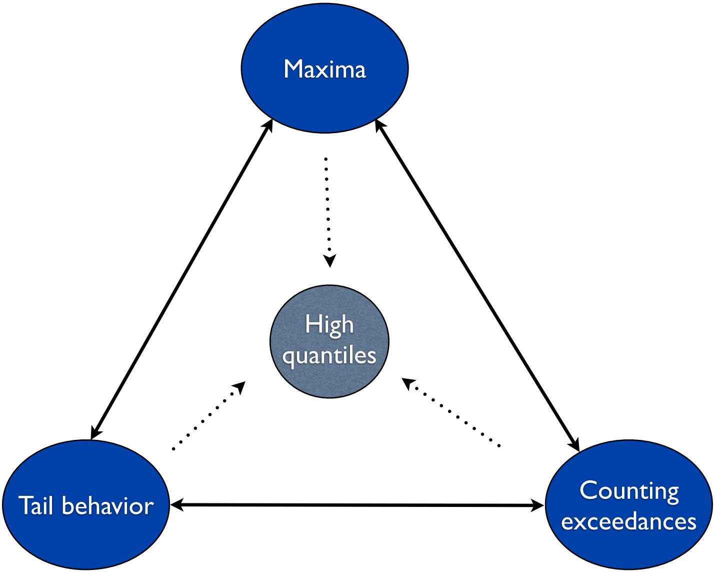
Counting = max

$$Pr(M_n \leq u_n) = Pr(N_n = 0) \text{ with } M_n = \max(X_1, \ldots, X_n)$$

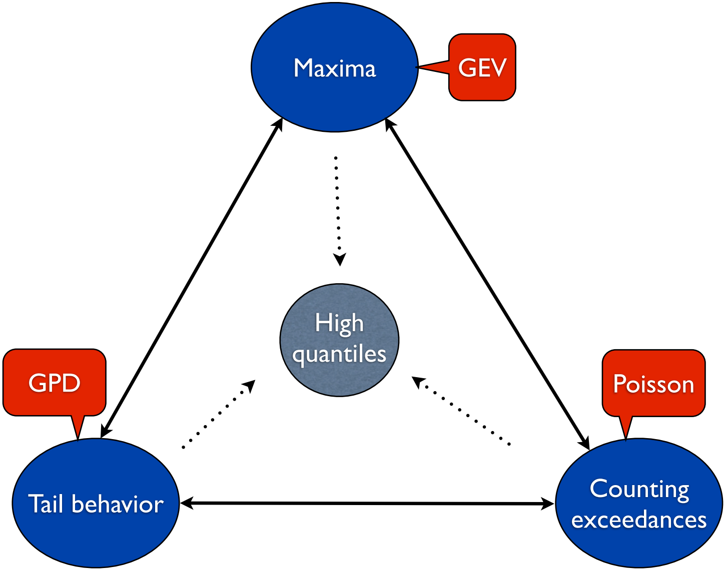
Poisson's at work

$$\lim_{n\to\infty} Pr(M_n \leq u_n) = \lim_{n\to\infty} Pr(N_n = 0) = Pr(N = 0) = \exp(-\lambda)$$

Equivalences

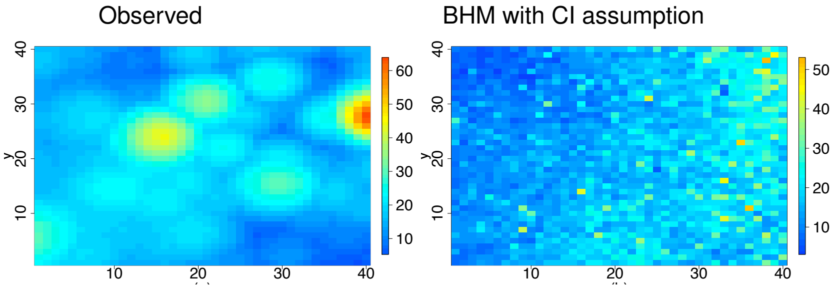


An univariate summary



Limits of the univariate approach

Independence or conditional independence assumptions



Ribatet, Cooley and Davison (2010)

Why is Multivariate EVT needed ?

- Compute confidence intervals
- Calculating probabilities of joint extreme events
- Clustering of regions
- Extrapolation of extremes
- Downscaling of extremes
- Trading time for space (for small data sets)
- etc

A fundamental question[2] for iid bivariate vector (X_i, Y_i)

Suppose that we have unit Fréchet margins at the limit

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x) = \lim_{n\to\infty} P(\max(Y_1,\ldots,Y_n)/a_n \leq x) = \exp(-x^{-1})$$

with a_n such that

$$P(X > a_n) = 1/n$$

2. L. de Hann, S. Resnick

A fundamental question[2] for iid bivariate vector (X_i, Y_i)

Suppose that we have unit Fréchet margins at the limit

$$\lim_{n\to\infty} P(\max(X_1,\dots,X_n)/a_n \le x) = \lim_{n\to\infty} P(\max(Y_1,\dots,Y_n)/a_n \le x) = \exp(-x^{-1})$$

with a_n such that

$$P(X > a_n) = 1/n$$

$$\lim_{n\to\infty} P(\max(X_1,\dots,X_n)/a_n \le x, \max(Y_1,\dots,Y_n)/a_n \le y) = ??$$

2. L. de Hann, S. Resnick

Why is the solution so ugly ?

If

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = G(x,y)$$

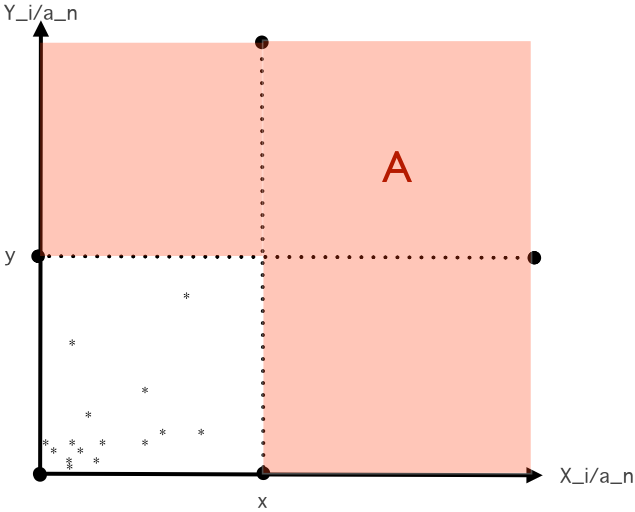
then

$$G(x,y) = \exp\left(-\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\, dH(w)\right)$$

where $H(.)$ such that $\int_0^1 w\, dH(w) = 1$

Still counting

$$P(\max(X_1, \ldots, X_n)/a_n \leq x, \max(Y_1, \ldots, Y_n)/a_n \leq y) = P(N_n(A) = 0)$$



Still counting

$$P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = P(N_n(A) = 0)$$

Poisson again

If

$$\lim_{n\to\infty} E(N_n(A)) = \Lambda(A),$$

then

$$\lim_{n\to\infty} P(N_n(A) = 0) = P(N(A) = 0) = \exp(-\Lambda(A))$$

Still counting

$$P(\max(X_1,\dots,X_n)/a_n \le x, \max(Y_1,\dots,Y_n)/a_n \le y) = P(N_n(A) = 0)$$

Poisson again

If

$$\lim_{n\to\infty} E(N_n(A)) = \Lambda(A),$$

then

$$\lim_{n\to\infty} P(N_n(A) = 0) = P(N(A) = 0) = \exp(-\Lambda(A))$$

One of the main question

- What are the properties of $\Lambda(A)$?

Back to univariate case : Fréchet margins

Poisson condition

$$\lim_{n\to\infty} nP(X/a_n \in A_x) = \Lambda_x(A_x)$$

with

$$\Lambda_x(A_x) = x^{-1}, \text{ for } A_x = [x, \infty)$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) =$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-x^{-1} - y^{-1})$$

Hence

$$x^{-1} + y^{-1} = \Lambda_x(A_x) + \Lambda_y(A_y) = \Lambda(A)$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-x^{-1} - y^{-1})$$

Hence

$$x^{-1} + y^{-1} = \Lambda_x(A_x) + \Lambda_y(A_y) = \Lambda(A)$$

The general case

$$\Lambda(A) \le \Lambda_x(A_x) + \Lambda_y(A_y)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) =$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\max(1/x, 1/y))$$

Hence,

$$\max(1/x, 1/y) = \max(\Lambda_x(A_x), \Lambda_x(A_y)) = \Lambda(A)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\max(1/x,1/y))$$

Hence,

$$\max(1/x,1/y) = \max(\Lambda_x(A_x),\ \Lambda_x(A_y)) = \ \Lambda(A)$$

The general case

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \le \ \Lambda(A)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\max(1/x,1/y))$$

Hence,

$$\max(1/x,1/y) = \max(\Lambda_x(A_x),\ \Lambda_x(A_y)) = \Lambda(A)$$

The general case

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \leq \Lambda(A)$$

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \leq \Lambda(A) \leq \Lambda_x(A_x) + \Lambda_y(A_y)$$

Scaling property

Univariate case with $\Lambda_x(A_x) = x^{-1}$

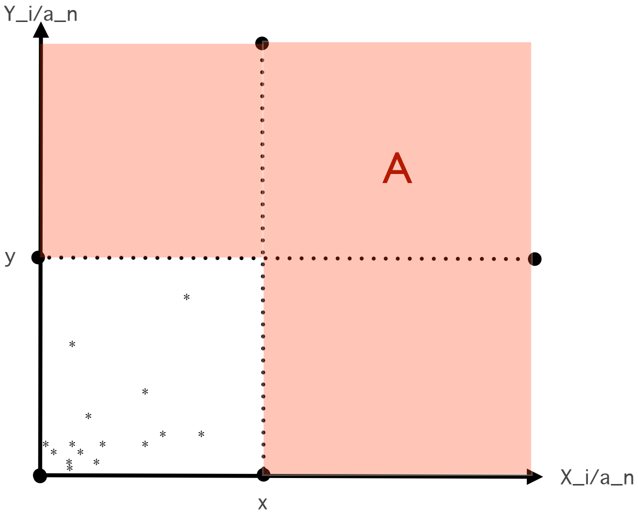
$$\Lambda_x(tA_x) = t^{-1}\Lambda_x(A_x)$$

Bivariate case

$$\Lambda(tA) = t^{-1}\Lambda(A)?$$

Going back to maxima

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\Lambda(A)) = P(M_X \le x, M_Y \le y)$$



Going back to maxima

$$P(M_X \leq x, M_Y \leq y) = \exp(-\Lambda(A))$$

Scaling

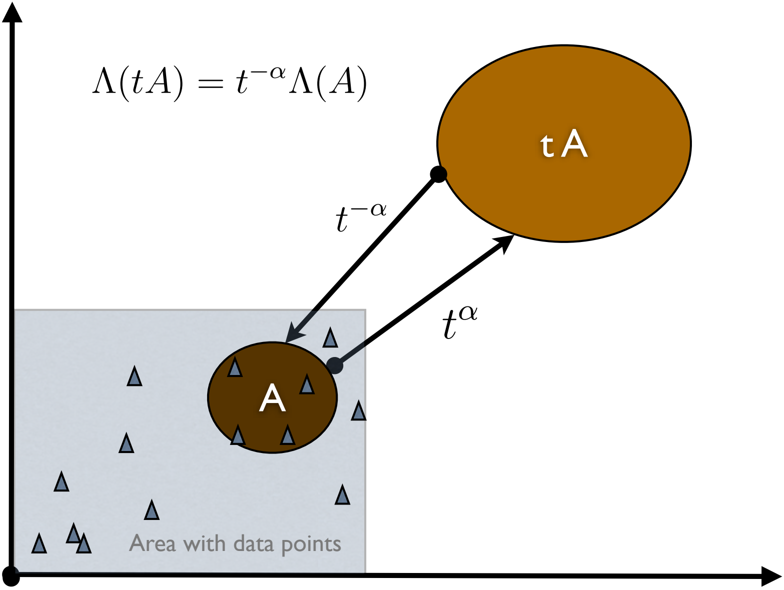
$$\Lambda(tA) = t^{-1}\Lambda(A)$$

is equivalent to

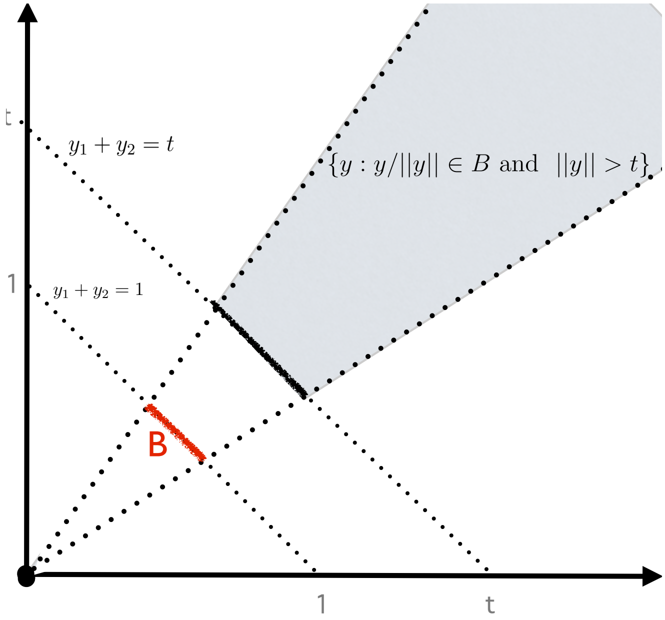
Max-stability

$$\begin{aligned} P^t(M_X \leq t\, x, M_Y \leq t\, y) &= (\exp(-\Lambda(tA)))^t = \exp(-t\Lambda(tA)) \\ &= \exp(-\Lambda(A)) \\ &= P(M_X \leq x, M_Y \leq y) \end{aligned}$$

Scaling property : an essential property of inference



Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$ with $||\mathbf{y}|| = y_1 + y_2$



Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$

A special case

$$A = \{\mathbf{z} = (x, y) : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}$$

where $||\mathbf{z}|| = x + y$ and B any set belonging to the unit sphere

A surprising property

$$\begin{aligned} tA &= \{t\mathbf{z} : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}, \\ &= \{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}, \text{ with } \mathbf{u} = t\mathbf{z}. \end{aligned}$$

This implies

$$\Lambda\left(\{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}\right) = t^{-1}H\left(B\right)$$

where $H(.)$ is the mean measure restricted to the unit sphere and often called the spectral measure.

Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$

A special case

$$A = \{\mathbf{z} = (x, y) : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}$$

where $||\mathbf{z}|| = x + y$ and B any set belonging to the unit sphere

A surprising property

$$\begin{aligned} tA &= \{t\mathbf{z} : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}, \\ &= \{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}, \text{ with } \mathbf{u} = t\mathbf{z}. \end{aligned}$$

This implies

$$\Lambda\left(\{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}\right) = t^{-1} H(B)$$

where $H(.)$ is the mean measure restricted to the unit sphere and often called the spectral measure.

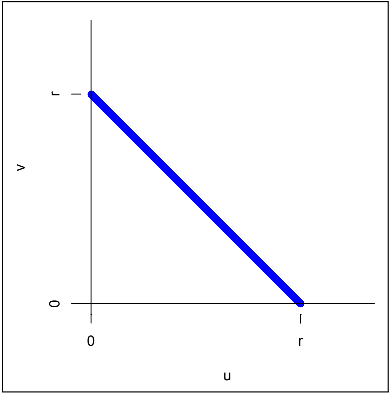
Independence between the strength of event $||\mathbf{z}|| = x + y$ and the location on the unit simplex

Polar coordinates

2D

$r = (u + v)$ and

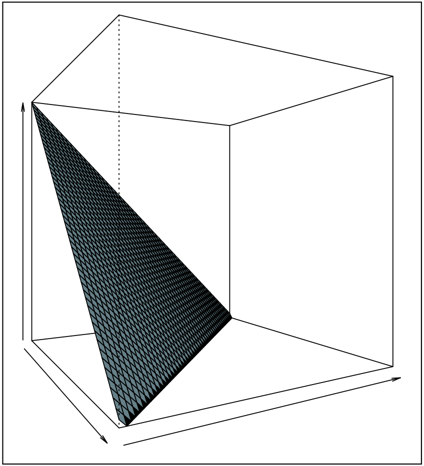
$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



3D

$r = (u + v + w),$

$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}, \theta_3 = \frac{w}{r}$

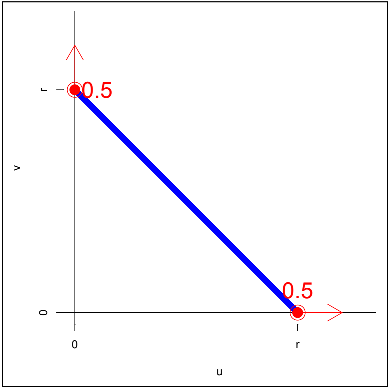


2D Polar coordinates

2D : INDEPENDENT CASE

$r = (u + v)$ and

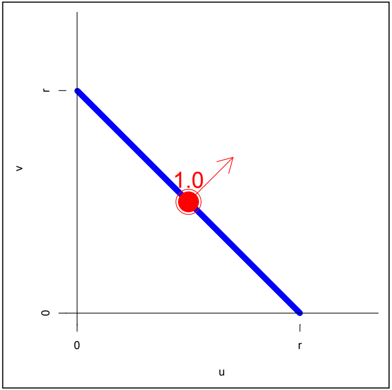
$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



2D : COMPLETE DEPENDENCE

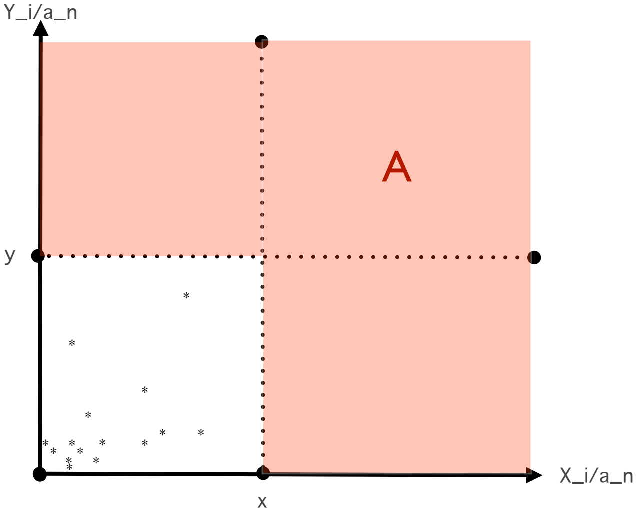
$r = (u + v)$ and

$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



Again, back to maxima

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\Lambda(A))$$



Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\Lambda(A))$$

Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\Lambda(A))$$

Changing coordinates : $r = u + v$ **and** $w = u/(u+v)$

$$\begin{aligned}
(u,v) \notin A \quad &\Leftrightarrow \quad u < x \text{ and } \; v < y, \\
&\Leftrightarrow \quad r < x/w \text{ and } \; r < y/(1-w), \\
&\Leftrightarrow \quad r < \min(x/w, y/(1-w))
\end{aligned}$$

Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\Lambda(A))$$

Changing coordinates : $r = u + v$ and $w = u/(u+v)$

$$\begin{aligned}(u,v) \notin A \quad &\Leftrightarrow \quad u < x \text{ and } \; v < y, \\ &\Leftrightarrow \quad r < x/w \text{ and } \; r < y/(1-w), \\ &\Leftrightarrow \quad r < \min(x/w, y/(1-w))\end{aligned}$$

Computing $\Lambda(A)$

$$\begin{aligned}\Lambda(A) \quad &= \quad \int_{w\in[0,1]} \int_{r>\min(x/w,y/(1-w))} r^{-2} dH(w) \\ &= \quad \int_{w\in[0,1]} \max(w/x, (1-w)/y) dH(w)\end{aligned}$$

Rewriting the counting rate in function of $H(dw)$

$$\Lambda(A) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) H(dw)$$

Scaling property checked

$$\Lambda(tA) = t^{-1}\Lambda(tA)$$

Max-stable vector

If

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = G(x,y)$$

then

$$-\log G(x,y) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\, dH(w)$$

where $H(.)$ such that $\int_0^1 w\, dH(w) = 1$

Max-stable vector properties

$$G(x, y) = \exp\left[-\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w)\right]$$

and $H(.)$ such that $\int_0^1 w\, dH(w) = 1$

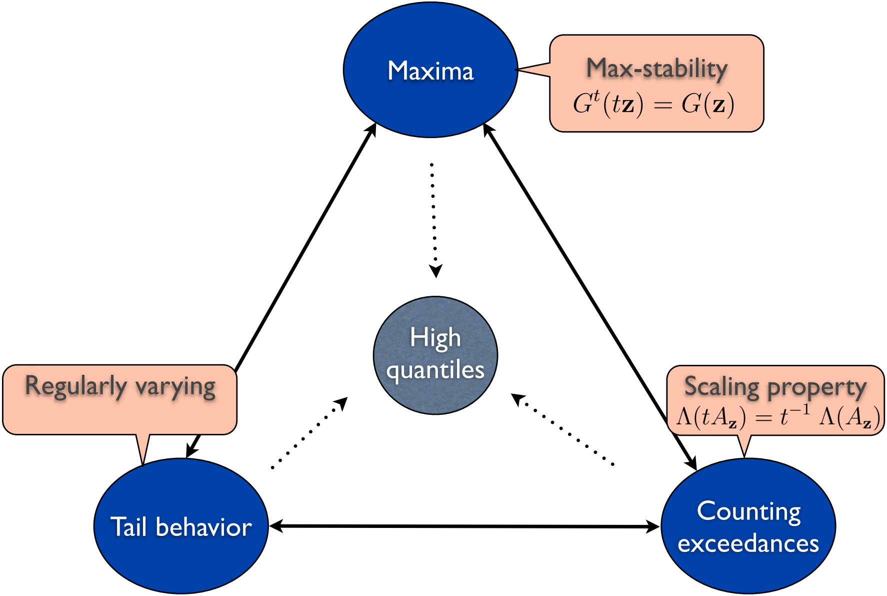
Max-stability

$$G^t(tx, ty) = G(x, y), \text{ for any } t > 0$$

Marginals : unit-Fréchet

$$G(x, \infty) = G(\infty, x) = \exp(-1/x)$$

A multivariate summary



Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

$$\Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of Wired.*

A quick summary about the basics

Learned lessons

- Multivariate maxima can be handled with Poisson counting processes
- "Polar coordinates" allows to see the independence between the strength of the event and the dependence structure that lives on the simplex
- The dependence structure has not explicit expressions (in contrast to the margins and to the Gaussian case)
- Max-stable property = scaling property for the Poisson intensity
- Conceptually easy to go from the bivariate to the multivariate case

Remaining questions

- How to make the inference of the dependence structure ?
- How can we use this dependence structure ?

Inference

Strategies for either the marginal behavior or the dependence

- Parametric : (+) Reduce dimensionality & easy to deal with covariates (-) impose a parametric form, model selection needed
- Non-parametric : (+) General without strong assumptions, (-) no practical for large dimension (curse of dimensionality), difficult to insert covariates

Inference

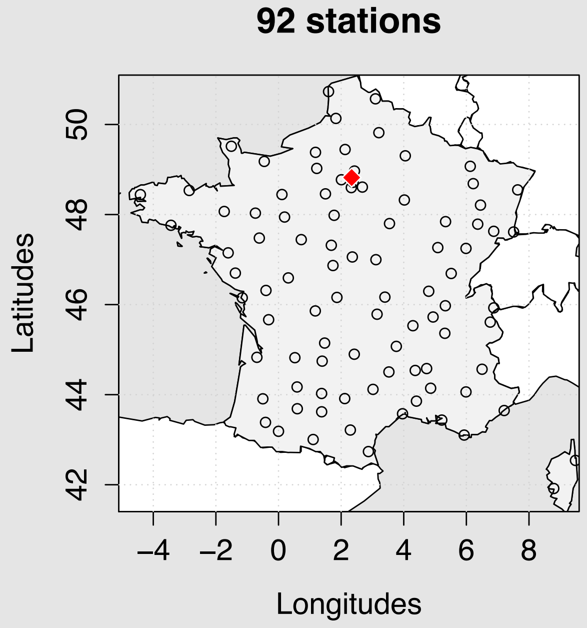
Strategies for either the marginal behavior or the dependence

- Parametric : (+) Reduce dimensionality & easy to deal with covariates (-) impose a parametric form, model selection needed
- Non-parametric : (+) General without strong assumptions, (-) no practical for large dimension (curse of dimensionality), difficult to insert covariates

Techniques

- Maximizing the likelihood : (+) easy to integrate covariates (-) impose a parametric form, no straightforward for large dimension
- Bayesian inference : (+) easy to insert expert knowledge, (-) no straightforward for large dimension (slow)
- Methods of moments : (+) fast and simple to understand, can be non-parametric (-) no straightforward to have covariates

Hourly precipitation in France, 1992-2011 (Olivier Mestre)



Our game plan to handle extremes from this big rainfall dataset

	Spatial scale	
	Large (country)	Local (region)
Problem	Dimension reduction	Spectral density in moderate dimension
Data	Weekly maxima of hourly precipitation	Heavy hourly rainfall excesses
Method	Clustering algorithms for maxima	Mixture of Dirichlet

Without imposing a given parametric structure

Clustering of maxima (joint work with E. Bernard, M. Vrac and O. Mestre)

Task 1

Clustering 92 grid points into around 10-20 climatologically homogeneous groups wrt spatial dependence

Clusterings

Challenges

- Comparing apples and oranges
- An average of maxima (centroid of a cluster) is not a maximum
- variances have to be finite
- Difficult interpretation of clusters

Questions

- How to find an appropriate metric for maxima ?
- How to create cluster centroids that are maxima ?

A central question (assuming that $\mathbb{P}[M(x) < v] = \mathbb{P}[M(y) < u] = \exp(-1/u)$)

$$\mathbb{P}[M(x) < u, M(y) < v] = \exp\left[-\int_0^1 \max\left(\frac{w}{u}, \frac{1-w}{v}\right)\, dH(w)\right]$$

θ = Extremal coefficient

$$\mathbb{P}\left[M(x) < u, M(y) < u\right] = \left(\mathbb{P}\left[M(x) < u\right]\right)^{\theta}$$

Interpretation

- Independence $\Rightarrow \theta = 2$
- $M(x) = M(y) \Rightarrow \theta = 1$
- Similar to correlation coefficients for Gaussian but ...
- No characterization of the **full** bivariate dependence

A L1 marginal free distance (Cooley, Poncet and N., 2005, N. and al., 2007)

$$d(x, y) = \frac{1}{2}\mathbb{E}\,|F_y(M(y)) - F_x(M(x))|$$

A L1 marginal free distance (Cooley, Poncet and N., 2005, N. and al., 2007)

$$d(x,y) = \frac{1}{2}\mathbb{E}\,|F_y(M(y)) - F_x(M(x))|$$

If $M(x)$ and $M(y)$ bivariate GEV, then

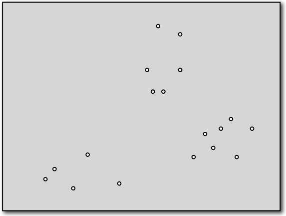
$$\text{extremal coefficient} = \frac{1+2d(x,y)}{1-2d(x,y)}$$

Clusterings

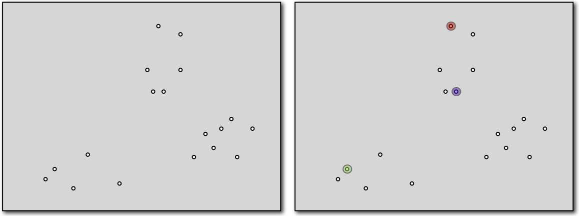
Questions

- How to find an appropriate metric for maxima ?
- **How to create cluster centroids that are maxima** ?

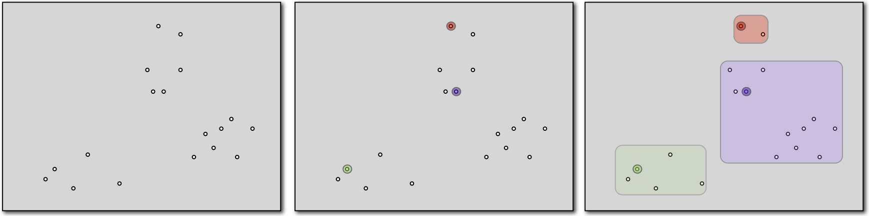
Partitioning Around Medoids (PAM) (Kaufman, L. and Rousseeuw, P.J. (1987))



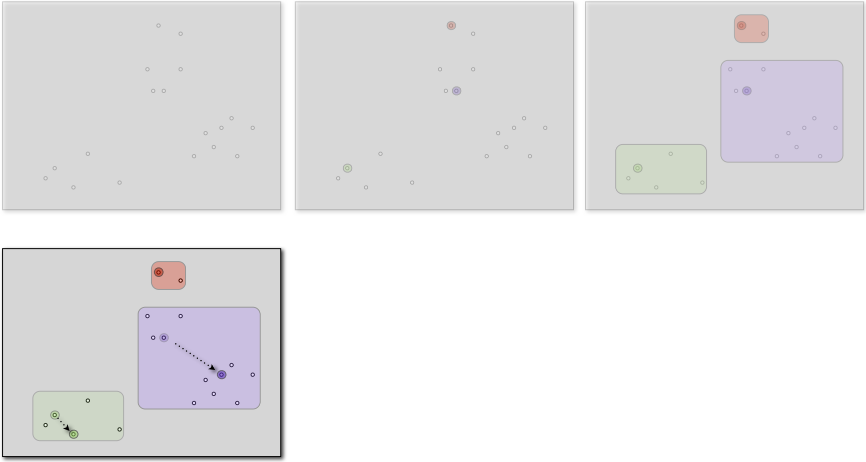
PAM : Choose K initial mediods



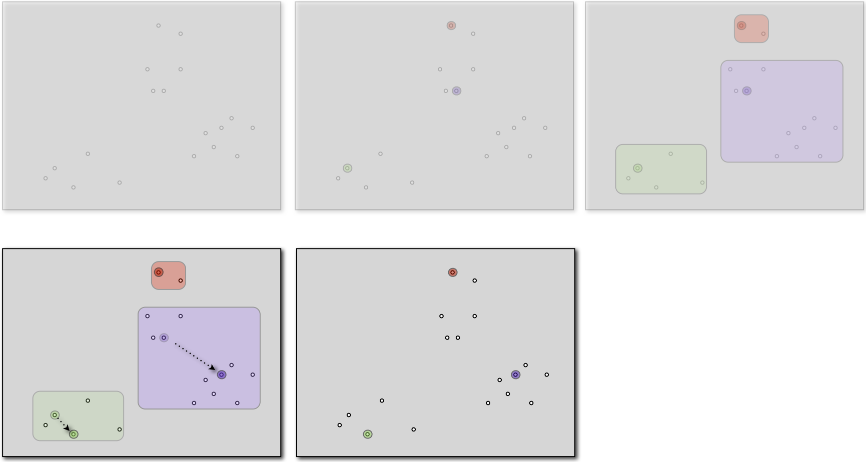
PAM : Assign each point to each closest mediod



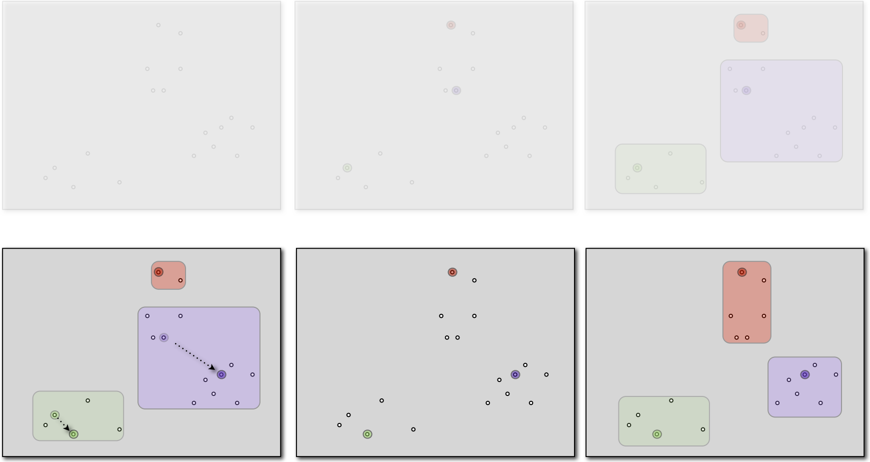
PAM : Recompute each mediod as the gravity center of each cluster



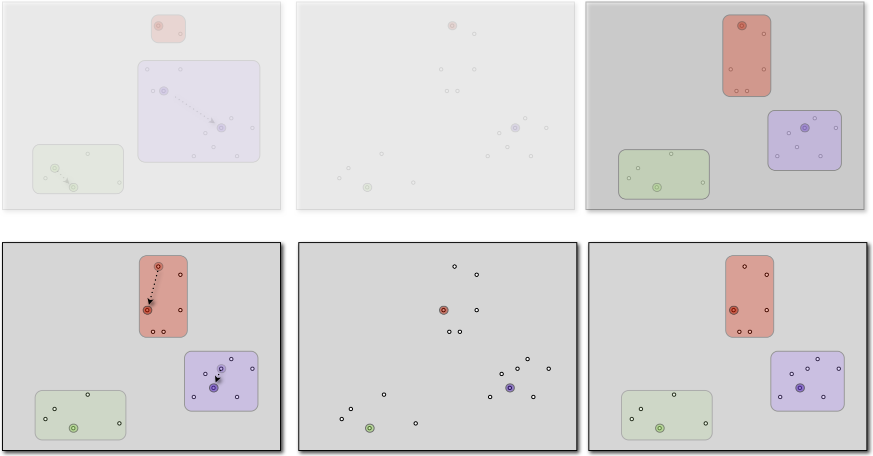
PAM : continue if a mediod has been moved



PAM : Assign each point to each closest mediod

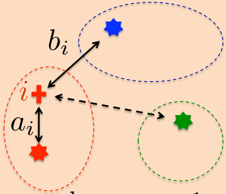


PAM : Recompute each mediod as the gravity center of each cluster



- **Clustering validation**

SILHOUETTE COEFFICIENT

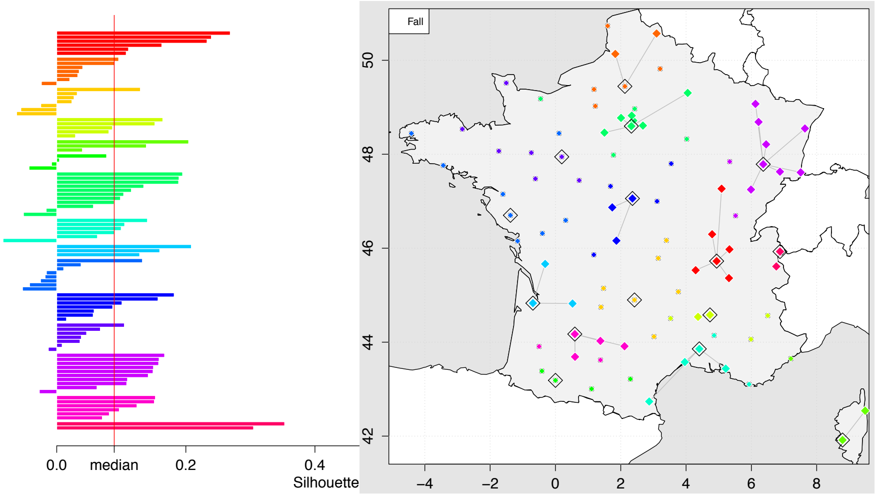


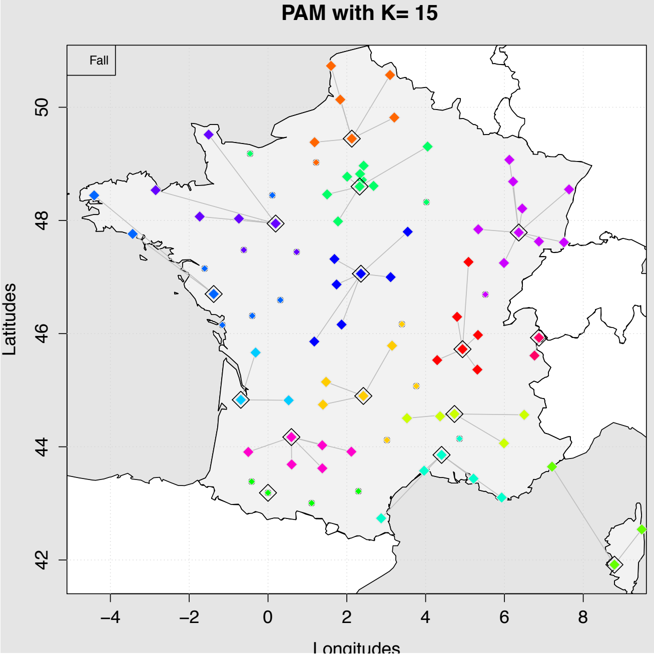
$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

$a_i \ll b_i, \quad s_i \approx 1 \quad \longrightarrow$ Well classified

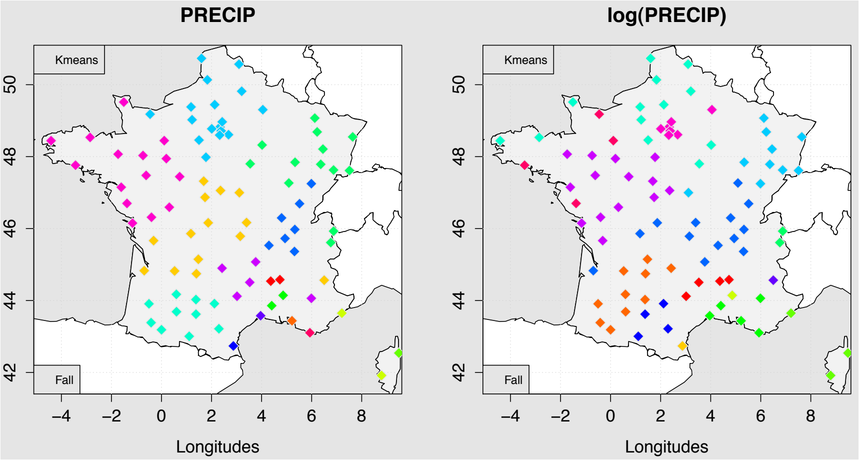
$a_i \sim b_i, \quad s_i \approx 0 \quad \longrightarrow$ Neutral

$a_i \gg b_i, \quad s_i \approx -1 \quad \longrightarrow$ Badly classified


0.0
median
0.2
0.4
Silhouette
Fall
50
48
46
44
42
−4
−2
0
2
4
6
8


PAM with K= 15
Fall
Latitudes
Longitudes
50
48
46
44
42
−4
−2
0
2
4
6
8

Applying the kmeans algorithm to maxima (15 clusters)



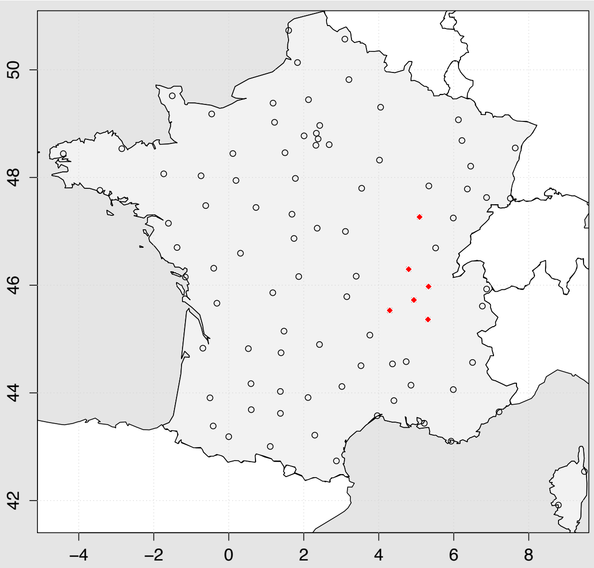
Summary on clustering of maxima

- Classical clustering algorithms (kmeans) are not in compliance with EVT
- Madogram provides a convenient distance that is marginal free and very fast to compute
- PAM applied with mado preserves maxima and gives interpretable results
- R package available on my web site

Our game plan to handle extremes from this rainfall dataset

	Spatial scale	
	Large (country)	Local (region)
Problem	Dimension reduction	Spectral density in moderate dimension
Data	Weekly maxima of hourly precipitation	Heavy hourly rainfall excesses
Method	Clustering algorithms for maxima	Mixture of Dirichlet

Back to the cluster



Bayesian Dirichlet mixture model for multivariate excesses (joint work with A. Sabourin)

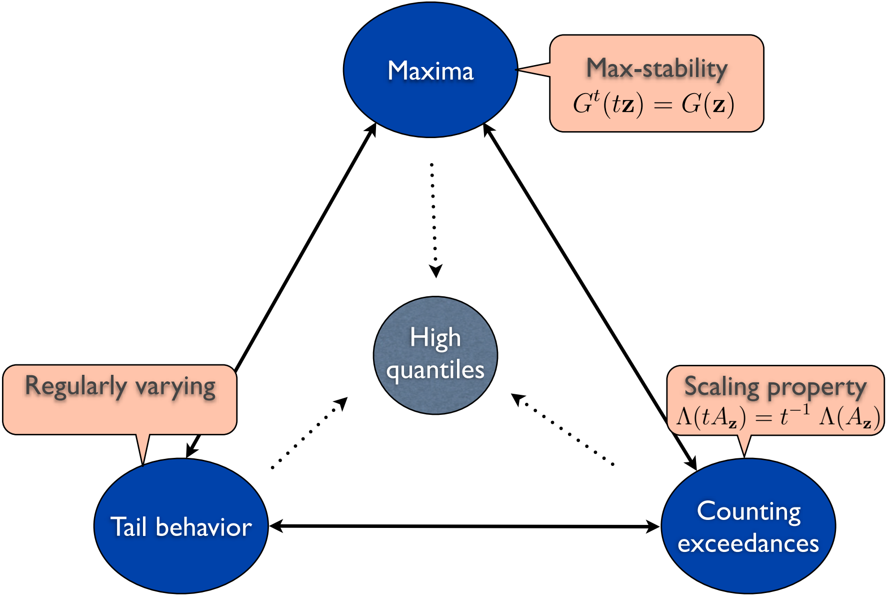
Meteo-France data

Wet hourly events at the regional scale (temporally declustered) of moderate dimensions (from 2 to 8)

Task 2

Assessing the dependence among rainfall excesses

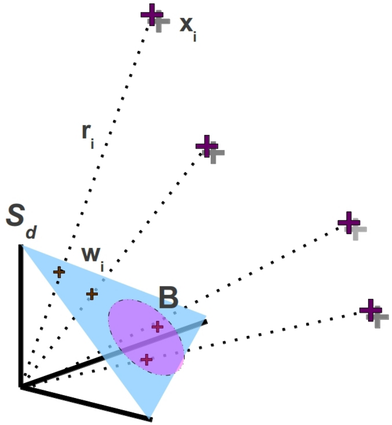
Multivariate Extreme Value Theory (de Haan, Resnick and others)



Defining radius and angular points

Example with $d = 3$ and $\mathbf{X} = (X_1, X_2, X_3)$ such that $\mathbf{P}(X_i < x) = e^{\frac{-1}{x}}$

$$\text{Simplex } \mathbf{S}_3 = \left\{ \mathbf{w} = (w_1, w_2, w_3) : \sum_{i=1}^{3} w_i = 1,\ w_i \geq 0 \right\}.$$



Mathematical constraints on the distribution of the angular points H

$$\mathbf{P}\left(\mathbf{W} \in B, R > r\right) \underset{r \to \infty}{\sim} \frac{1}{r} H(B)$$

Features of H

- H can be non-parametric
- The gravity center of H has to be centered on the simplex

$$\forall i \in \{1, \ldots, d\},\ \int_{\mathbf{S}_d} w_i \,\mathrm{d}H(\mathbf{w}) = \frac{1}{d}$$

A few references on Bayesian non-parametric and semi-parametric spectral inference

M.-O. Boldi and A. C. Davison.
A mixture model for multivariate extremes.
JRSS : Series B (Statistical Methodology), 69(2) :217–229, 2007.

S. Guillotte, F. Perron, and J. Segers.
Non-parametric bayesian inference on bivariate extremes.
JRSS : Series B (Statistical Methodology), 2011.

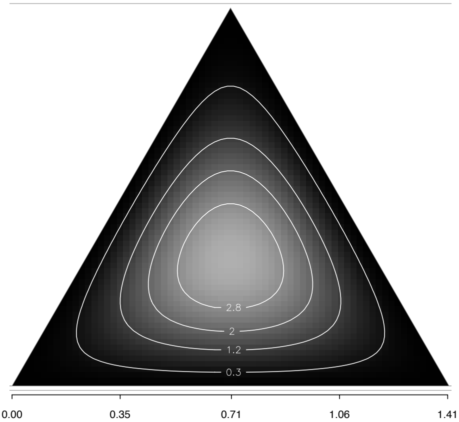
A. Sabourin and P. Naveau.
Bayesian Drichlet mixture model for multivariate extremes.
Computational Statistics & Data Analysis, 2014.

P.J. Green.
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.
Biometrika, 82(4) :711, 1995.

Roberts, G.O. and Rosenthal, J.S.
Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains
The Annals of Applied Probability,16,4,2123 :2139, 2006.

Dirichlet distribution

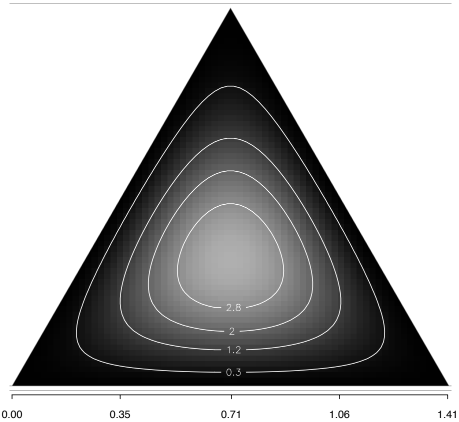
$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d,\ \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^{d} \Gamma(\nu\mu_i)} \prod_{i=1}^{d} w_i^{\nu\mu_i - 1}\,.$$



$\boldsymbol{\mu} = (1/3, 1/3, 1/3)$ and $\nu = 9$

Dirichlet distribution

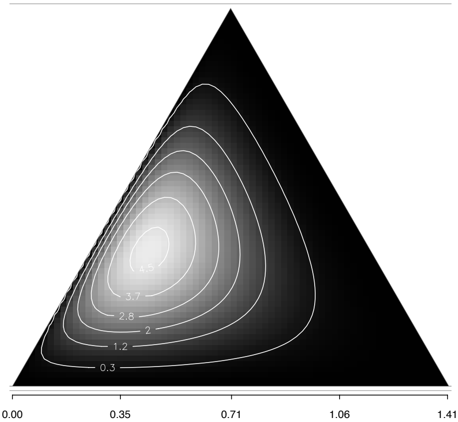
$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d,\ \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^{d} \Gamma(\nu\mu_i)} \prod_{i=1}^{d} w_i^{\nu\mu_i - 1}\,.$$



$\boldsymbol{\mu} = (1/3, 1/3, 1/3)$ and $\nu = 9$

Dirichlet distribution

$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d, \ \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^d \Gamma(\nu\mu_i)} \prod_{i=1}^d w_i^{\nu\mu_i - 1} .$$



$\boldsymbol{\mu} = (.15, .35, .05)$ and $\nu = 9$

But this one is not centered ! !

Mixture of Dirichlet distribution

Boldi and Davision, 2007

$$h_{(\boldsymbol{\mu},\mathbf{p},\boldsymbol{\nu})}(\mathbf{w}) = \sum_{m=1}^{k} p_m \, \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}_{\cdot\,,\,m}, \nu_m)$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot\,,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$

Mixture of Dirichlet distribution

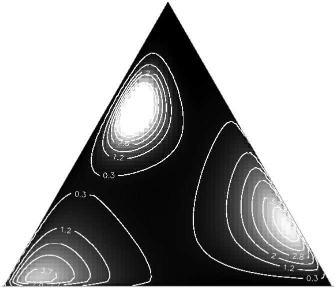
Boldi and Davision, 2007

$$h_{(\boldsymbol{\mu},\mathbf{p},\boldsymbol{\nu})}(\mathbf{w}) = \sum_{m=1}^{k} p_m \, \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}_{\cdot\,,\,m}, \nu_m)$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot\,,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$

Constraint on (μ, p)

$$p_1\, \boldsymbol{\mu}_{.,1} + \cdots + p_k\, \boldsymbol{\mu}_{.,k} = (\tfrac{1}{d}, \ldots, \tfrac{1}{d})$$



Inference of Dirichlet density mixtures

Boldi and Davison (2007)

Prior of $[\mu|\mathbf{p}]$ defined on the set

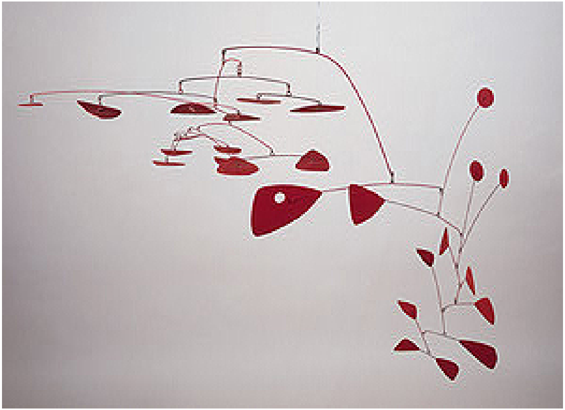
$$p_1\, \boldsymbol{\mu}_{.,1} + \cdots + p_k\, \boldsymbol{\mu}_{.,k} = (\tfrac{1}{d}, \ldots, \tfrac{1}{d})$$

- Sequential inference : first **p**, then $\boldsymbol{\mu}$ one coordinate after the other
- skewed, not interpretable, slow sampling
- Difficult inference in dimension > 3

Inference of Dirichlet density mixtures

How to build priors for $(\mathbf{p}, \mu)$ such that

$$p_1 \, \boldsymbol{\mu}_{\cdot,1} + \cdots + p_k \, \boldsymbol{\mu}_{\cdot,k} = \left(\tfrac{1}{d}, \ldots, \tfrac{1}{d}\right)$$



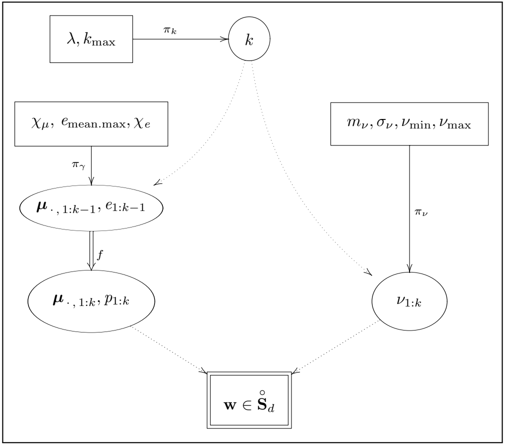
Unconstrained Bayesian modeling for $\Theta = \coprod_{k=1}^{\infty} \Theta_k; \quad \Theta_k = \left\{ (\mathbf{S}_d)^{k-1} \times [0,1)^{k-1} \times (0,\infty]^{k-1} \right\}$

Prior

$$
\begin{aligned}
& k \sim \text{Truncated geometric} \\
& \boldsymbol{\mu}_{\cdot,m} | (\boldsymbol{\mu}_{\cdot,1:m-1}, e_{1:m-1}) \sim \text{Dirichlet} \\
& e_m | (\boldsymbol{\mu}_{\cdot,1:m}, e_{1:m-1}) \sim \text{Beta} \\
& \nu_m \sim \text{logN}
\end{aligned}
$$

Posterior sampling : MCMC reversible jumps

Summary of the Bayesian scheme



Summary of the Bayesian schemes

Boldi and Davison (2012) Our approach

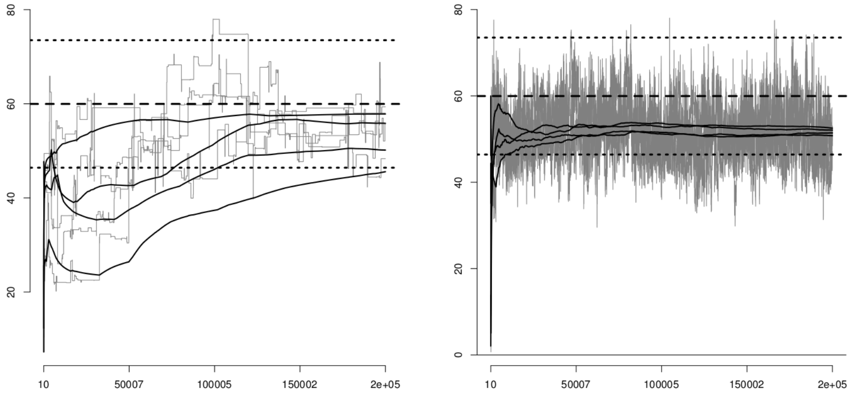
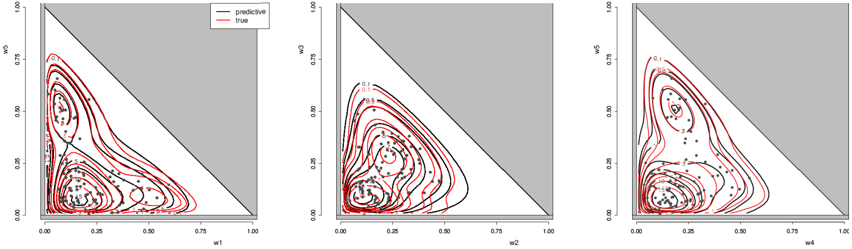


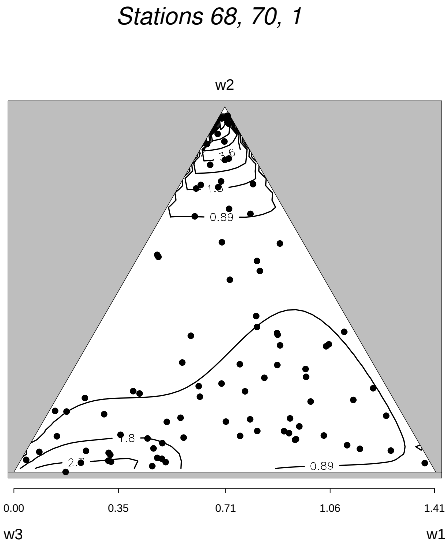
Figure 5: Convergence monitoring with five-dimensional data in the original DM model (left panel) and in the re-parametrized v
with four parallel chains in each model. Grey lines: Evolution of $\langle g, h_{\theta_t(j)} \rangle$. Black, solid lines: cumulative mean. Dashed line

Simulation example with $d = 5$ and $k = 3$



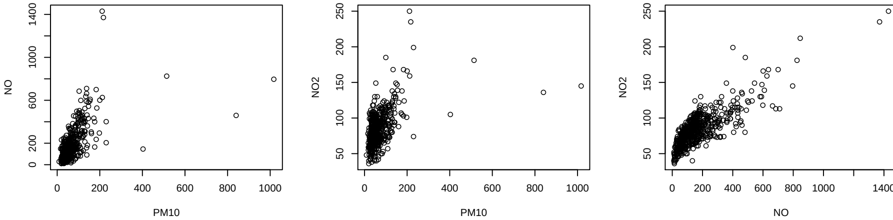
$$T_2 = 150\,10^3, \; T_1 = 50\,10^3.$$

Back to our excesses of the “Lyon” cluster



Coming back to Leeds

Air pollutants (Leeds, UK, winter 94-98, daily max) NO vs. PM10 (left), SO2 vs. PM10 (center), and SO2 vs. NO (right) (Heffernan& Tawn 2004, Boldi & Davison, 2007)



Coming back to Leeds

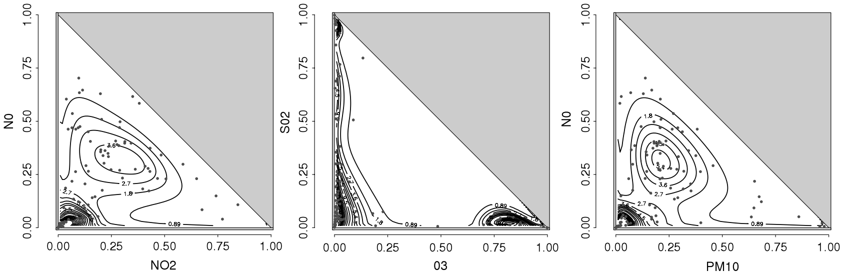


Fig. 6. Five dimensional Leeds data set: posterior predictive density. Black lines: projections of the predictive angular density defined on the four-dimensional simplex $\mathbf{S}_5$ onto the two-dimensional faces. Gray dots: projections of the 100 points with greatest L^1 norm.

Take home messages

Conclusions

- Clustering of weekly maxima with PAM is fast and gives spatially coherent structures
- Bayesian semi-parametric mixture can handle moderate dimensions and provide credibility intervals

Going further

- Anne Sabourin = a Bayesian semi-parametric mixture for censored data with an application to paleo-flood data

References

- Bernard, E., et al.. Clustering of maxima : Spatial dependencies among heavy rainfall in france. Journal of Climate, 2013, [**R package**].
- Sabourin, A. , Naveau, P. Dirichlet Mixture model for multivariate extremes. To appear in Computational Statistics and Data Analysis. [**R package**].
- Naveau P. et al., Modeling Pairwise Dependence of Maxima in Space. Biometrika, (2009)

Take home messages part II

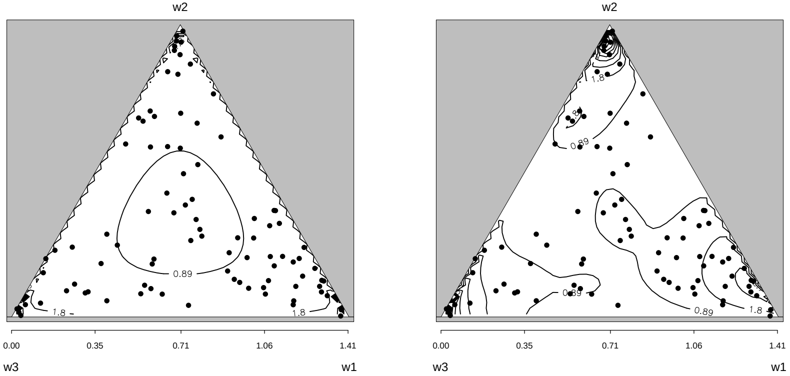
- Extremes here means very rare
- It is possible to estimate the dependence between bivariate extremes
- Multivariate EVT may help characterizing extremes dependencies in space or time
- Modeling trade off between parametric and non-parametric approaches

Advertisements

- Blog : Beyond the Hill
- International Meetings on Statistical Climatology, Canmore, Alberta, June 6-10, 2016.
- Summer School on Extreme value modeling and water resources, Lyon, June 13-24 2016.
- Workshop on environmental extremes, Aussois, France, June 26-30 2016.
- Banff International Research Station workshop on extremes, 13-17 June, 2016

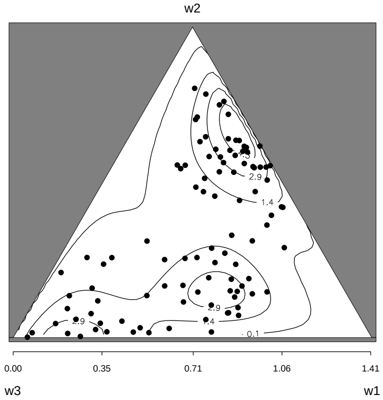
Different results from different Monte Carlo chains ?

Stations 68, 70, 42

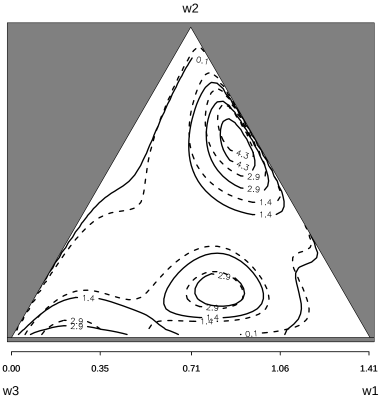


Simulation example with $d = 3$ and $k = 3$

Simulated points with true density

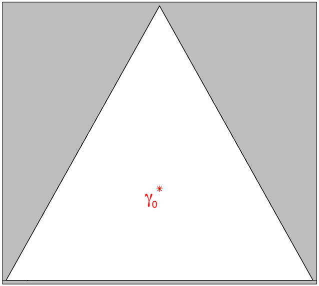


Predictive density



New parametrisation

Ex : $k = 4$ and $d = 3$

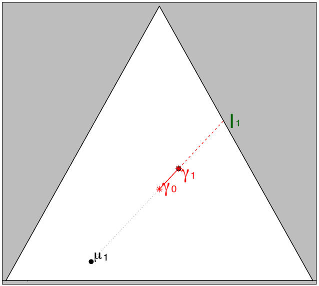


γ_m : "Equilibrium" centers built from $\boldsymbol{\mu}_{.,m+1}, \dots, \boldsymbol{\mu}_{.,k}$.

$$\gamma_m = \sum_{j=m+1}^{k} \frac{p_j}{p_{m+1} + \dots + p_k} \boldsymbol{\mu}_{.,j}$$

New parametrisation

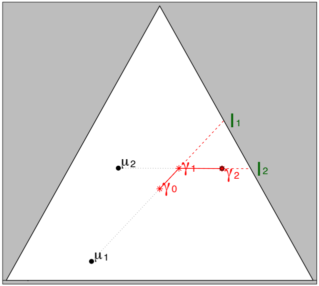
Ex : $k = 4$ and $d = 3$



$$\boldsymbol{\mu}_{\cdot,1},\ e_1 \quad \Rightarrow \gamma_1 : \frac{\overline{\gamma_0\,\gamma_1}}{\overline{\gamma_0\ l_1}} = e_1\ ;$$
$$\Rightarrow p_1$$

New parametrisation

Ex : $k = 4$ and $d = 3$

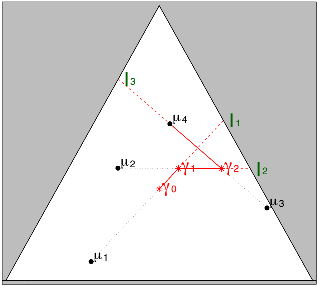


$$\mu_{.,2},\ e_2 \quad \Rightarrow \gamma_2 : \frac{\overline{\gamma_1\,\gamma_2}}{\overline{\gamma_1\,l_2}} = e_2\ ;$$

$$\Rightarrow p_2$$

New parametrisation

Ex : $k = 4$ and $d = 3$

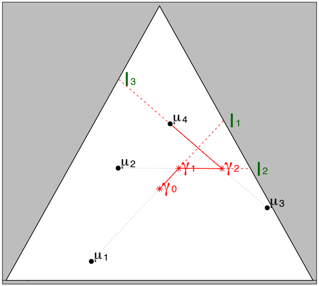


$$\boldsymbol{\mu}_{\cdot,3},\ e_3 \quad \Rightarrow \gamma_3 : \frac{\overline{\gamma_2\,\gamma_3}}{\overline{\gamma_2\,l_3}} = e_3\ ; \quad \boldsymbol{\mu}_{\cdot,4} = \gamma_3.$$

$$\Rightarrow p_3, p_4$$

New parametrisation

Ex : $k = 4$ and $d = 3$



Parametrisation of h with $\theta = (\boldsymbol{\mu}_{\cdot,1:k-1}, \mathbf{e}_{1:k-1}, \nu_{1:k})$

$$(\boldsymbol{\mu}_{\cdot,1:k-1}, e_{1:k-1}) \text{ gives } (\boldsymbol{\mu}_{\cdot,1:k}, p_{1:k})$$