Non parametric modeling of multivariate extremes with Dirichlet mixtures

Anne Sabourin

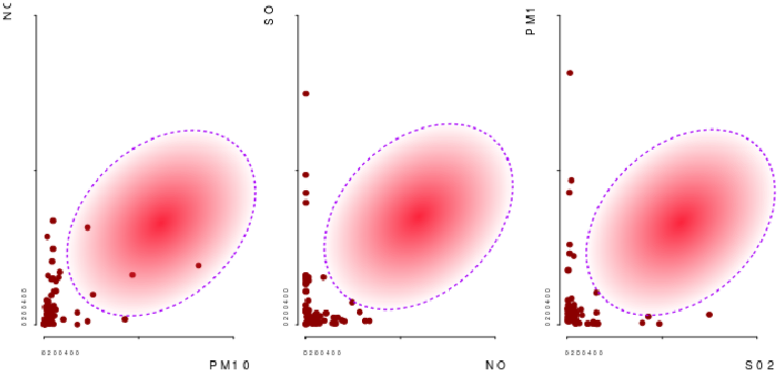
Institut Mines-Télécom, Télécom ParisTech, CNRS-LTCI

Joint work with Philippe Naveau (LSCE, Saclay), Anne-Laure Fougères (ICJ, Lyon 1), Benjamin Renard (IRSTEA, Lyon).

March 26th, 2013
Rare and Extreme workshop, Aber Wrach

Air quality

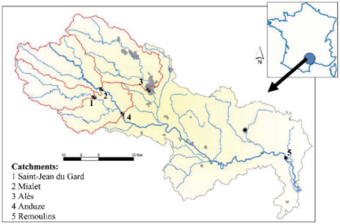
- Five air pollutants recorded in Leeds (UK), daily.
- Health issue : probability of a joint (simultaneous) excess of alert thresholds ?
- Probability of regions far from the origin ?



Censored Multivariate extremes : floods in the 'Gardons'

joint work with Benjamin Renard

- Daily streamflow at 4 neighbouring sites
  (St Jean du Gard, Mialet, Anduze, Alès).
- **Joint distribution of extremes ?**
  → Probability of simultaneous floods.
- Historical data → censored data ; few clean data.

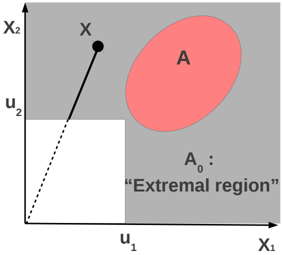


Gard river Neppel *et al.* (2010)

Multivariate extremes

- Random vectors $\mathbf{Y} = (Y_1, \ldots, Y_{d,})$; $\quad Y_j \geq 0$
- Margins : $Y_j \sim F_j$, $1 \leq j \leq d$ (any).
- **Standardization** (unit Fréchet margins) : $X_j = -1/\log F_j(Y_j)$
- Joint extremes : $\mathbf{X}$'s distribution above large thresholds ?

$P(\mathbf{X} \in A \,|\, X \in A_0)$? $\qquad (A \subset A_0, \mathbf{0} \notin A_0)$, A_0 'far from the origin'.



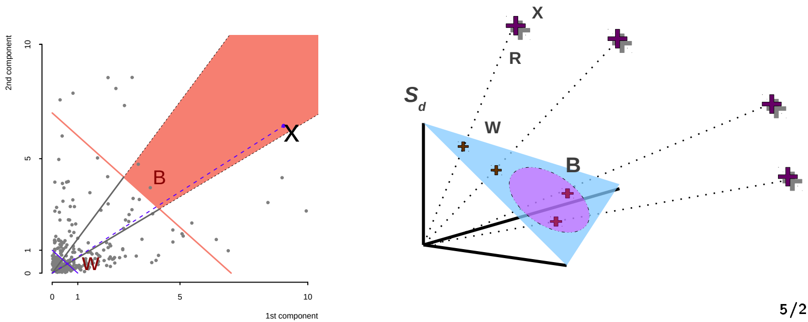
Polar decomposition

- Polar coordinates : $R = \sum_{j=1}^{d} X_j$ (L_1 norm) ; $\mathbf{W} = \frac{\mathbf{X}}{R}$.
- $\mathbf{W} \in$ simplex $\mathbf{S}_d = \{\mathbf{w} : \quad w_j \geq 0, \sum_j w_j = 1\}$.

$$\text{Characterize } \mathbb{P}(\mathbf{X} \in A \mid A \in A_0)$$

$$\Leftrightarrow$$

$$\text{Characterize } \mathbb{P}(R > r, \mathbf{W} \in B \mid R > r_0)$$

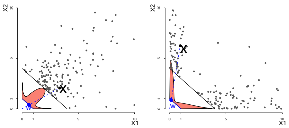


Fundamental Result, Angular distribution

- Radial homogeneity (under hypothesis of regular variation)

$$\mathbb{P}(R > r\, t, \mathbf{W} \in B \mid R \geq t) \underset{t \to \infty}{\to} \frac{1}{r} H(B)$$

- Above large thresholds r_0, $R \perp\!\!\!\perp W$;
- H (+ margins) rules the joint distribution



- One condition only for genuine H : **moments constraint**,

$$\int \mathbf{w} \, \mathrm{d}H(\mathbf{w}) = (\frac{1}{d}, \dots, \frac{1}{d}). \quad \text{Center of mass = center of simplex.}$$

- Few constraints : **non parametric** family !

Estimating the angular measure : non parametric problem

- **Non parametric estimation** (empirical likelihood, Einmahl *et al.*, 2001, Einmahl, Segers, 2009, Guillotte *et al*, 2011.) No explicit expression for asymptotic variance, Bayesian inference with $d = 2$ only, nothing for censored data.
- Compromise : **Mixture** of countably many parametric models $\rightarrow$ Infinite-dimensional model + easier Bayesian inference (handling parameters).

Dirichlet mixture model

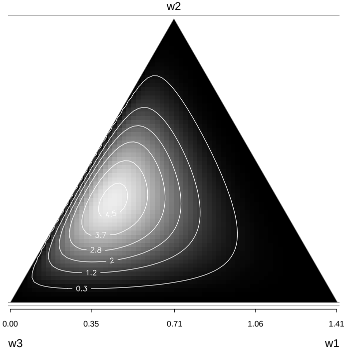
(Boldi, Davison, 2007 ; S. , Naveau, 2013)

- How to deal with the **moments constraint** on H to generate parameters / define a prior ?
- Do **MCMC** methods work in moderate dimension ($d = 5$) ?
- Does it still work with **censored data** ?

Dirichlet distribution

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

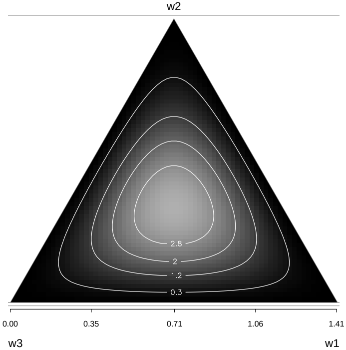
- $\boldsymbol{\mu} \in \mathring{\mathbf{S}}_d$: location parameter (point on the simplex) : 'center' ;
- $\nu > 0$: concentration parameter.



Dirichlet distribution

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

- $\boldsymbol{\mu} \in \mathring{\mathbf{S}}_d$: location parameter (point on the simplex) : 'center' ;
- $\nu > 0$: concentration parameter.



Dirichlet mixture model

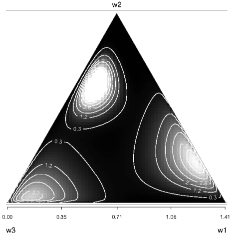
Boldi, Davison, 2007

- $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$, $\psi = (\boldsymbol{\mu}, \mathbf{p}, \boldsymbol{\nu})$,

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

- Moments constraint $\rightarrow$ on $(\boldsymbol{\mu}, p)$:

$$\sum_{m=1}^{k} p_m \,\boldsymbol{\mu}_{\cdot,m} = (\frac{1}{d}, \dots, \frac{1}{d}) \,.$$



Weakly dense family ($k \in \mathbb{N}$) in the space of admissible angular measures

Bayesian inference with non censored data

- Moments constraints $\Rightarrow$ barycenter constriant on $(\boldsymbol{\mu}, p) \Rightarrow$
  - Prior construction ?
  - Parameter generation for MCMC sampling ?

  Difficult for dimension > 2.

- Re-parametrization S. , Naveau (13) : work with unconstrained parameter
  - Weak posterior consistency
  - MCMC with reversible jumps manageable in moderate dimension ($\simeq 5$).

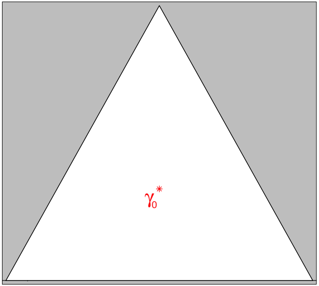
Re-parametrization

Sabourin, Naveau 2013

- How to build a prior on $(p\,,\,\boldsymbol{\mu})$?
- Constraint on **center of mass** : $\sum_j p_j\, \boldsymbol{\mu}_{\cdot,j}$
- Sequential construction : Use **associativity** properties of barycenter.
- Intermediate variables : **partial centers of mass** ; determined by **eccentricity parameters** $(e_1, \ldots, e_{k-1}) \in (0,1)^{k-1}$.
- Deduce last $\boldsymbol{\mu}_{\cdot,k}$ from first ones : **no more constraints** !

Re-parametrization : intermediate variables $(\gamma_1, \ldots, \gamma_{k-1})$, partial barycenters

ex : $k = 4$

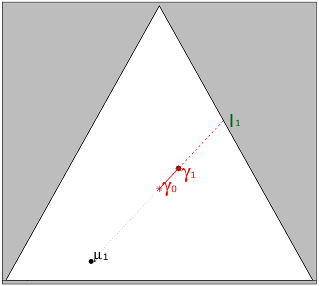


γ_m : Barycenter of kernels 'following $\boldsymbol{\mu}_{.,m}$ " : $\boldsymbol{\mu}_{.,m+1}, \ldots, \boldsymbol{\mu}_{.,k}$.

$$\gamma_m = \Big(\sum_{j>m} p_j\Big)^{-1} \sum_{j>m} p_j\, \boldsymbol{\mu}_{.,j}$$

γ_1 on a line segment : *eccentricity* parameter $e_1 \in (0,1)$.

$ex : k = 4$

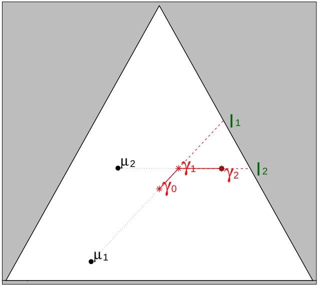


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,1} \in \mathbf{S}_d,\ e_1 \in (0,1)) \longrightarrow \gamma_1 \text{ defined by } \frac{\overline{\gamma_0\,\gamma_1}}{\overline{\gamma_0\, l_1}} = e_1 \ ;$$

$$\longrightarrow p_1 = \frac{\overline{\gamma_0\,\gamma_1}}{\overline{\boldsymbol{\mu}_{\cdot,1}\,\gamma_1}} \ .$$

γ_2 on a line segment : *eccentricity* parameter $e_2 \in (0,1)$.

ex : $k = 4$

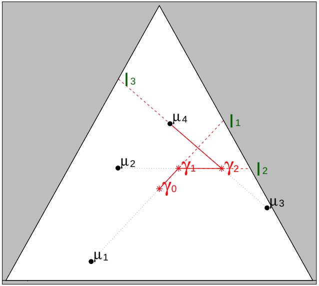


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,2},\, e_2) \longrightarrow \gamma_2 : \frac{\overline{\gamma_1\,\gamma_2}}{\overline{\gamma_1\, l_2}} = e_2$$

$$\longrightarrow p_2$$

Last density kernel = last center $\boldsymbol{\mu}_{\cdot,k}$.

$ex : k = 4$

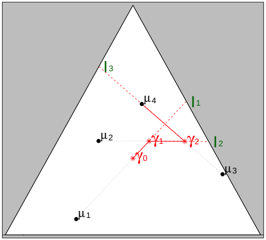


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,3},\, e_3) \quad \longrightarrow \gamma_3$$

$$\longrightarrow p_3\,,\ \boldsymbol{\mu}_{.,4} = \gamma_3.$$

$$\longrightarrow p_4$$

Summary



- Given

$$(\boldsymbol{\mu}_{.,1:k-1}, e_{1:k-1}),$$

One obtains

$$(\boldsymbol{\mu}_{.,1:k}, p_{1:k}).$$

- The density h may thus be parametrized by

$$\boldsymbol{\theta} = (\boldsymbol{\mu}_{.,1:k-1}, \mathbf{e}_{1:k-1}, \nu_{1:k}) \in \text{'rectangle', unconstrained .}$$

Bayesian model

- New parameter : $\theta_k = (\boldsymbol{\mu}_{\cdot,1:k-1}, e_{1:k-1}, \nu_{1:k})$
- Unconstrained parameter space : union of product spaces ('rectangles')

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

  - Inference : Gibbs + Reversible-jumps.

- Restriction (numerical convenience) : $k \leq 15$, $\nu < \nu_{\max}$, *etc* ...
- 'Reasonable' prior $\simeq$ 'flat' and rotation invariant. Balanced weight and uniformly scattered centers.

MCMC sampling : Metropolis-within-Gibbs, reversible jumps.

Three transition types for the Markov chain :

- *Classical (Gibbs)* : one $\boldsymbol{\mu}_{.,m}$, e_m or a ν_m is modified.

  Proposals of new Dirichlet centers depend on the data.

- *Trans-dimensional* (Green, 1995) :
  One component $(\mu_{.,k}, e_k, \nu_{k+1})$ is added or deleted.

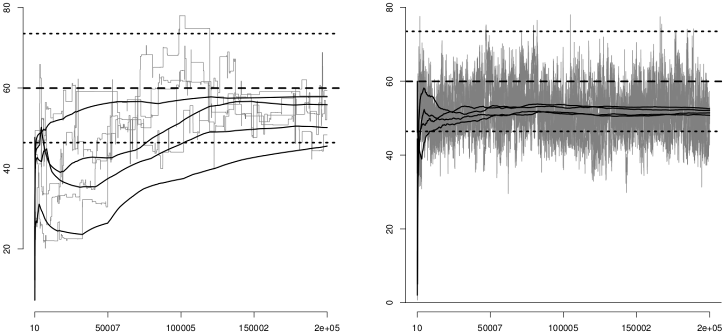
  Trans-dimensional moves are natural. Additional components again depend on the data

- *'Shuffle'* : Indices permutation of the original mixture :

  Re-allocating mass from old components to new ones.

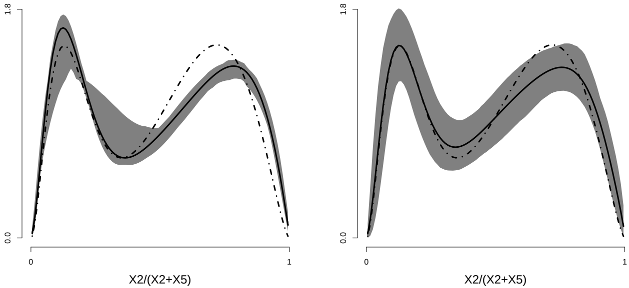
Resuts in the re-parametrized version

- Theoretically (Asymptotics) :
  - Posterior consistency : $\forall U$ weakly open in Θ, containing θ_0, $\pi_n(U) = \pi(U|\text{data}_{1:n}) \xrightarrow[n\to\infty]{} 1$.
  - Markov chain's ergodicity : $\sum_{t=1}^{T} g(\theta_t) \xrightarrow[T\to\infty]{} \mathbb{E}_{\pi_n}(g)$
- Empirically : convergence checks.
  Better mixing :



Resuts in the re-parametrized version

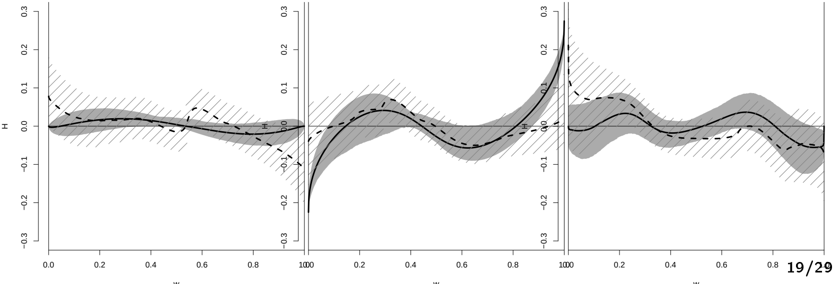
- Theoretically (Asymptotics) :
  - Posterior consistency : $\forall U$ weakly open in Θ, containing θ_0, $\pi_n(U) = \pi(U|\text{data}_{1:n}) \xrightarrow[n\to\infty]{} 1$.
  - Markov chain's ergodicity : $\sum_{t=1}^{T} g(\theta_t) \xrightarrow[T\to\infty]{} \mathbb{E}_{\pi_n}(g)$
- Empirically : convergence checks.
  Better coverage of credible sets (d=5, bivariate margins, simulated data)



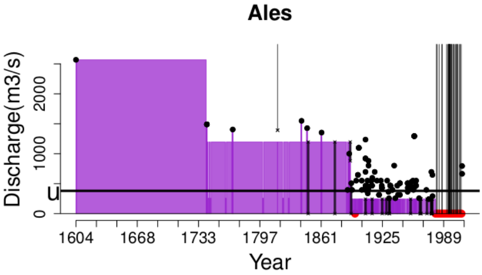
Resuts in the re-parametrized version

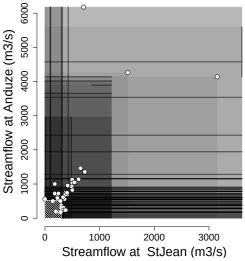
- Theoretically (Asymptotics) :
  - Posterior consistency : $\forall U$ weakly open in Θ, containing θ_0, $\pi_n(U) = \pi(U|\text{data}_{1:n}) \xrightarrow[n\to\infty]{} 1$.
  - Markov chain's ergodicity : $\sum_{t=1}^{T} g(\theta_t) \xrightarrow[T\to\infty]{} \mathbb{E}_{\pi_n}(g)$
- Empirically : convergence checks.
  As good (in dimension 2) as the bivariate non-parametric model of Guillotte et. al. (2006) (simulated data in logistic/asymmetric logistic/Dirichlet.

  Solid line : DM. dotted : alternative non parametric model)



Inference with censored data



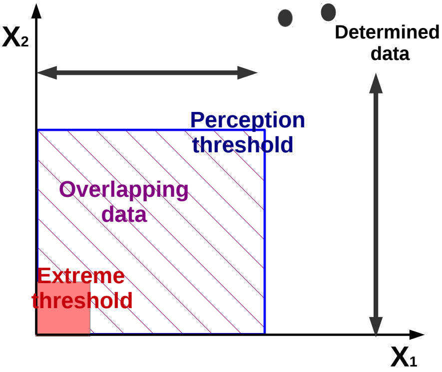


- Existing litterature : Ledford & Tawn, 1996 : censoring at threshold.
  - GEV models
  - Explicit expression for censored likelihood.

Issues

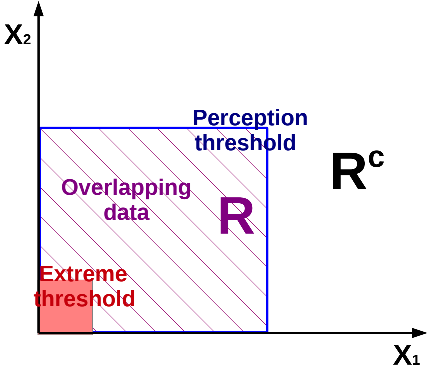
- Censored data $\neq$ points but segments or boxes in $\mathbf{R}^d$.
- Angles W_i undefined.
- Intervals overlapping threshold : extreme data or not ?
- Censored likelihood : density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.

Undetermined data (overlapping threshold)



Considering 'undetermined data' as missing ⇒ biais !

Undetermined data (overlapping threshold)

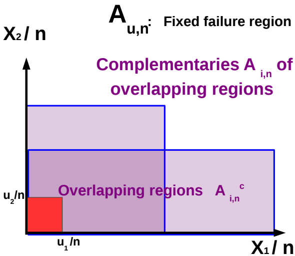


Data in region $R \Leftrightarrow$ not in region R^c ...

Well defined likelihood in a Poisson model

Poisson model

$$\left\{ \left(\frac{t}{n}, \frac{\mathbf{X}_t}{n} \right), \; 1 \leq t \leq n \right\} \sim \mathrm{PRM}(\mathrm{Leb} \times \mu_*) \text{ on } [0,1] \times A_{u,n}$$



μ_* : ' exponent measure', with Dirichlet Mixture angular component

$$\frac{\mathrm{d}\mu_*}{\mathrm{d}r \times \mathrm{d}\mathbf{w}}(r, \mathbf{w}) = \frac{d}{r^2} h(\mathbf{w}) .$$

Likelihood of overlapping data :

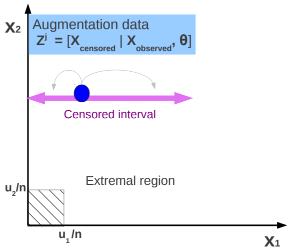
$$\mathbf{P}\left[N \left\{ (\frac{t_2}{n} - \frac{t_1}{n}) \times \frac{1}{n} A_i \right\} = 0 \right] = \exp\left[-(t_2 - t_1)\mu_*(A_i) \right]$$

'Censored' likelihood : and data augmentation

- Data augmentation : Generate missing components under univariate conditional distributions.

One more Gibbs step, no more numerical integration.

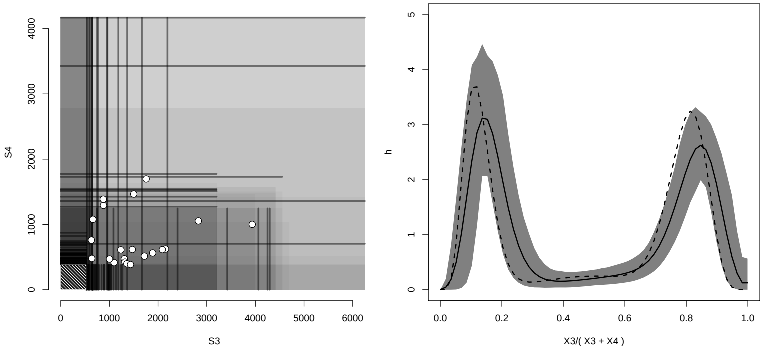
$$\mathbf{Z}^j_{1:r} \sim [X_{\text{missing}} | X_{\text{obs}}, \theta]$$



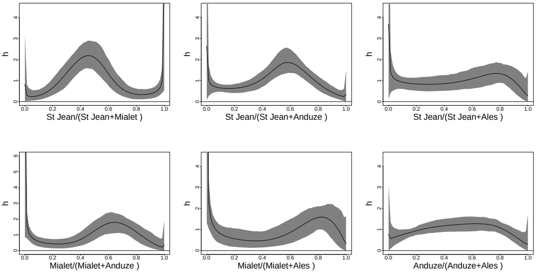
Dirichlet $\Rightarrow$ Explicit univariate conditionals
Exact sampling of censored data on censored interval

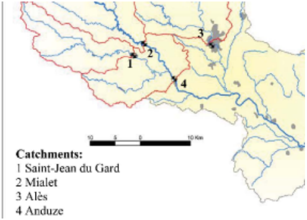
Simulated data (Dirichlet, $d = 4$, $k = 3$ components), same censoring as real data

Pairwise plot and angular measure density
(true/ posterior predictive)



Angular predictive density for Gardons data





Conclusion

- Bayesian Dirichlet model for multivariate large excesses :
  - 'non' parametric, suitable for moderate dimension, adaptable to censored data.
  - Two packages R :
    - `DiriXtremes`, MCMC algorithm for Dirichlet mixtures,
    - `DiriCens`, implementation with censored data.
- Towards high dimension (GCM grid, spatial fields)
  - Impose reasonable structure (sparse) on Dirichlet parameters ?
- Possible application : Posterior sample $\rightarrow$ Simulation of regional extremes ?

References

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

Gómez, G., Calle, M. L., and Oller, R.
Frequentist and bayesian approaches for interval-censored data.
Statistical Papers, 45(2) :139–173, 2004.

Ledford, A. and Tawn, J. (1996).
Statistics for near independence in multivariate extreme values.
Biometrika, 83(1) :169–187.

Neppel, L., Renard, B., Lang, M., Ayral, P., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., and Vinet, F. (2010).
Flood frequency analysis using historical data : accounting for random and systematic errors.
Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55(2) :192–208.

Sabourin, A., Naveau, P. (2013)
Bayesian Dirichlet mixture model for multivariate extremes : a re-parametrization.
Computation. Stat and Data Analysis

Schnedler, W. (2005).
Likelihood estimation for censored random vectors.
Econometric Reviews, 24(2) :195–217.

Van Dyk, D. and Meng, X. (2001).
The art of data augmentation.
Journal of Computational and Graphical Statistics, 10(1) :1–50.