Bayesian model mergings for multivariate extremes
Application to regional predetermination of floods with incomplete data

Anne Sabourin

PhD defense
University of Lyon 1

Committee:
Anthony Davison, Johan Segers *(Rapporteurs)*
Anne-Laure Fougères (Director), Philippe Naveau (Co-director),
Clémentine Prieur, Stéphane Robin, Éric Sauquet.

September 24th, 2013

Multivariate extreme values

- Risk management: Largest events, largest losses

- Hydrology: 'flood predetermination'.
  - Return levels (extreme quantiles)
  - Return periods (1 / probability of occurrence)

  $\rightarrow$ digs, dams, land use plans.

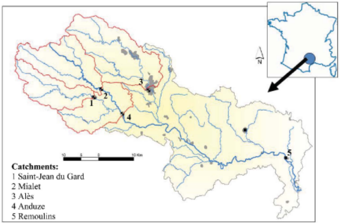
- Simultaneous occurrence of rare events can be catastrophic
  - **Multivariate extremes:**

    Probability of jointly extreme events ?

Censored Multivariate extremes: floods in the 'Gardons'

joint work with Benjamin Renard

- Daily streamflow data at 4 neighbouring sites :
  St Jean du Gard, Mialet, Anduze, Alès.
- **Joint distributions of extremes ?**
  $\rightarrow$ probability of simultaneous floods.
- Recent, 'clean' series very short
- Historical data from archives, depending on 'perception thresholds' for floods (Earliest: 1604). $\rightarrow$ censored data



Gard river   Neppel *et al.* (2010)

How to use all different kinds of data ?

Multivariate extremes for regional analysis in hydrology

- Many sites, many parameters for marginal distributions, short observation period.
- 'Regional analysis': replace time with space.
  Assume some parameters constant over the region and use extreme data from all sites.
- Independence between extremes at neighbouring sites ?
  Dependence structure ?
  - Idea: use multivariate extreme value models

Outline

Multivariate extremes and model uncertainty

Bayesian model averaging ('Mélange de modèles')

Dirichlet mixture model ('Modèle de mélange'):
a re-parametrization

Historical, censored data in the Dirichlet model

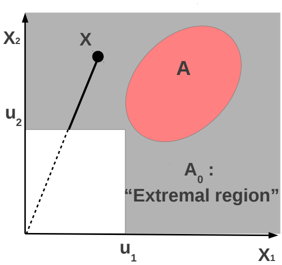
Multivariate extremes

- Random vectors $\mathbf{Y} = (Y_1, \dots, Y_{d,})$; $\quad Y_j \geq 0$
- Margins: $Y_j \sim F_j$, $1 \leq j \leq d$
  (Generalized Pareto above large thresholds)
- **Standardization** ($\rightarrow$ unit Fréchet margins)

$$X_j = -1/\log\left[F_j(Y_j)\right] \;; \quad P(X_j \leq x) = e^{-1/x}, \quad 1 \leq j \leq d$$

- Joint behaviour of extremes: distribution of $\mathbf{X}$ above large thresholds ?

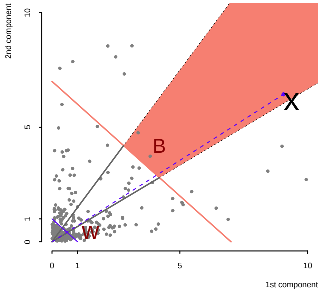
$P(\mathbf{X} \in A | X \in A_0)$? $\qquad (A \subset A_0, \mathbf{0} \notin A_0)$, A_0 'far from the origin'.

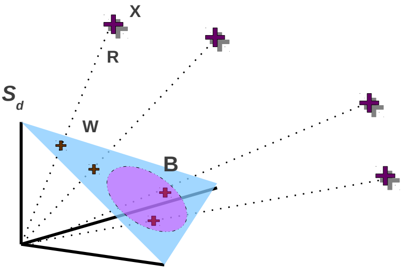


Polar decomposition and angular measure

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

$$H(B) = P(\mathbf{W} \in B) \quad (B \subset \mathbf{S}_d).$$





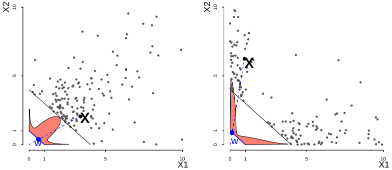
Fundamental Result

de Haan, Resnick, 70's, 80's

- Radial homogeneity (regular variation)

$$P(R > r, \mathbf{W} \in B | R \geq r_0) \underset{r_0 \to \infty}{\sim} \frac{r_0}{r} H(B) \quad (r = c\, r_0, c > 1)$$

- Above large radial thresholds, R is independent from W
- H (+ margins) entirely determines the joint distribution



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

$$\int \mathbf{w}\, \mathrm{d}H(\mathbf{w}) = (\frac{1}{d}, \dots, \frac{1}{d}).$$

Center of mass at the center of the 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.

- Restriction to **parametric family**: Gumbel, logistic, pairwise Beta . . . Coles & Tawn, 91, Cooley *et al.*, 2010, Ballani & Schlather, 2011 :

  How to take into account model uncertainty ?

Outline

Multivariate extremes and model uncertainty

Bayesian model averaging ('Mélange de modèles')

Dirichlet mixture model ('Modèle de mélange'):
a re-parametrization

Historical, censored data in the Dirichlet model

BMA: Averaging estimates from different models

Disjoint union of several parametric models

- Sabourin, Naveau, Fougères, 2013 (Extremes)
- Package R: `'BMAmevt'` , available on CRAN repositories [1].

[1] http://cran.r-project.org/

Outline

Multivariate extremes and model uncertainty

Bayesian model averaging ('Mélange de modèles')

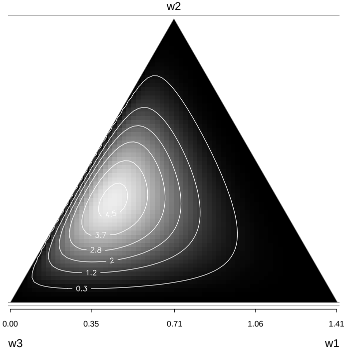
Dirichlet mixture model ('Modèle de mélange'):
a re-parametrization

Historical, censored data in the Dirichlet model

Dirichlet distribution

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

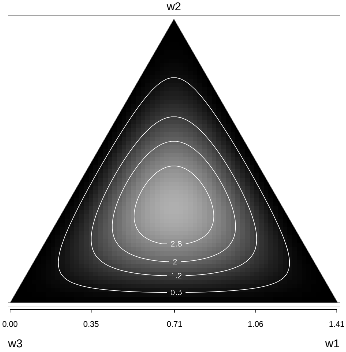
- $\boldsymbol{\mu} \in \mathring{\mathbf{S}}_d$: location parameter (point on the simplex): 'center';
- $\nu > 0$: concentration parameter.



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

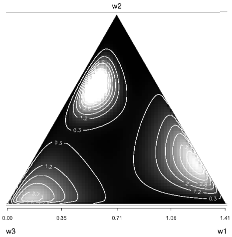
Boldi, Davison, 2007

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

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

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

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



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

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$.
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin, *under review* ; Sabourin, Renard, *in preparation*

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$.

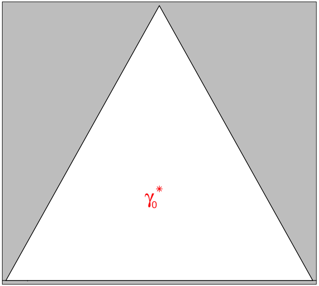
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.

  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin, *under review* ; Sabourin, Renard, *in preparation*

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

ex: $k = 4$

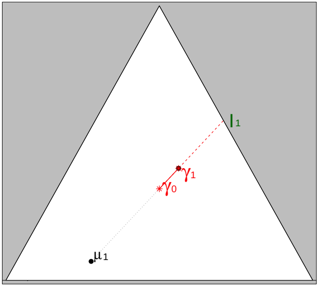


γ_m : Barycenter of kernels 'following $\boldsymbol{\mu}_{.,m}$": $\boldsymbol{\mu}_{.,m+1}, \ldots, \boldsymbol{\mu}_{.,k}$.

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

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

ex: $k = 4$

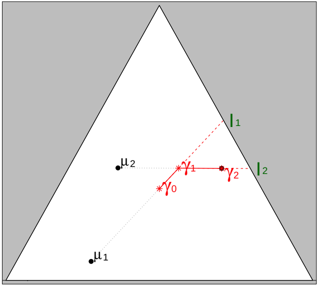


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,1} \in \mathbf{S}_d,\ e_1 \in (0,1)) \longrightarrow \gamma_1 \text{ defined by } \frac{\overline{\gamma_0\,\gamma_1}}{\overline{\gamma_0\,l_1}} = e_1\ ;$$

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

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

ex: $k = 4$

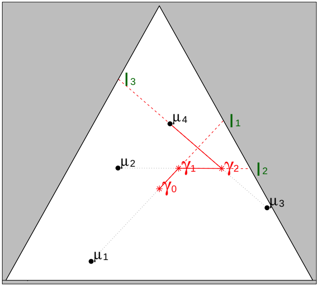


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,2},\, e_2) \longrightarrow \gamma_2 : \frac{\overline{\gamma_1\,\gamma_2}}{\overline{\gamma_1\, l_2}} = e_2$$

$$\longrightarrow p_2$$

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

ex: $k = 4$

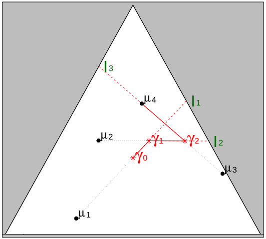


$$\text{Draw } (\boldsymbol{\mu}_{\cdot,3},\ e_3) \longrightarrow \gamma_3$$

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

$$\longrightarrow p_4$$

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

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

Bayesian model

- 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) : $k \leq 15$, $\nu < \nu_{\mathrm{max}}$, *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 $\boldsymbol{\mu}_{\cdot,m}$, e_m or a ν_m is modified.
- *Trans-dimensional* (Green, 1995):
  One component $(\mu_{\cdot,k}, e_k, \nu_{k+1})$ is added or deleted.
- *'Shuffle'*: Indices permutation of the original mixture:
  Re-allocating mass from old components to new ones.

Results: model's and algorithm's consistency

- Ergodicity: The generated MC is ϕ-irréducible, aperiodic and admits π_n (= posterior | $W_{1:n}$) as invariant distribution.
  - *Consequence*:

$$\forall g \in \mathcal{C}_b(\Theta), \quad \frac{1}{T}\sum_{t=1}^{T} g(\theta_t) \to \mathbb{E}_{\pi_n}(g)\,.$$

  - *Key point*: π_n is invariant under the 'shuffle' moves.
- Posterior consistency for π_n under 'weak conditions'[2], π-a.s., $\forall U$ weakly open containing θ_0,

$$\pi_n(U) \xrightarrow[n\to\infty]{} 1\,.$$

  - *Consequence*:

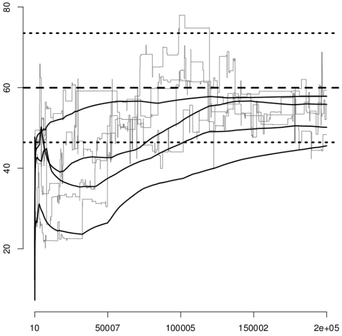
$$\mathbb{E}_{\pi_n}(g) \xrightarrow[n\to\infty]{} g(\theta_0)\,.$$

  - *Key*: The Euclidian topology is finer than the Kullback topology in this model.

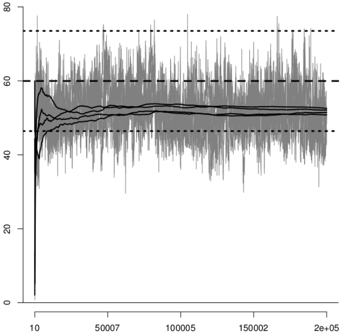
[2] If the prior grants some mass to every Euclidian neighbourhood of Θ and if θ_0 is in the Kullback-Leibler closure of Θ

Convergence checking (simulated data, $d = 5$, $k = 4$)

- θ summarized by a scalar quantity (integrating the DM density against a test function)



Original algorithm

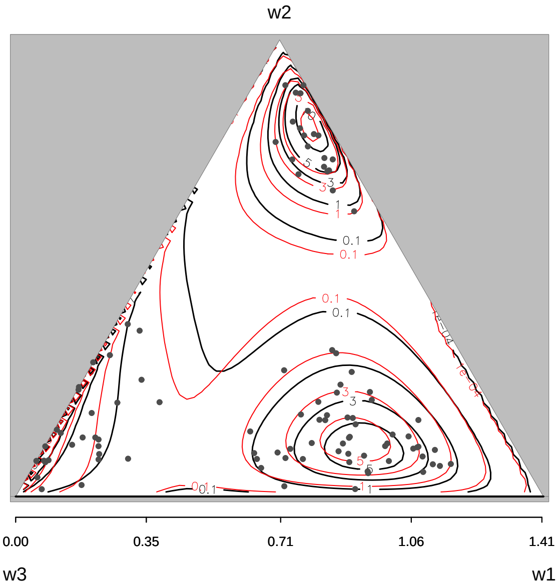


Re-parametrized version

- standard tests:
  - Stationarity (Heidelberger & Welch, 83)
  - variance ratio (inter/intra chains, Gelman, 92)

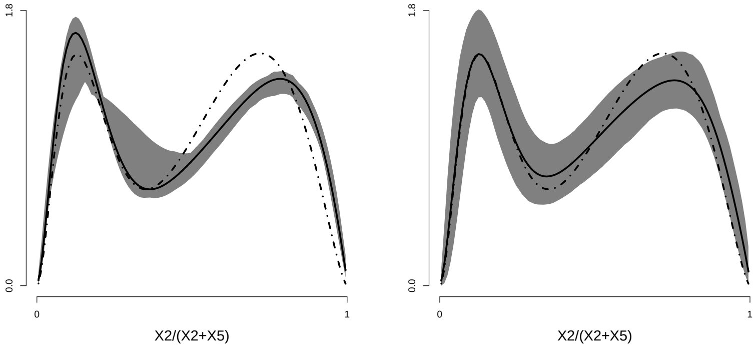
Estimating H in dimension 3, simulated data

MCMC: $T_2 = 50\,10^3$; $T_1 = 25\,10^3$.



Dimension 5, simulated data

Angular measure density for one pair $(T_2 = 200\,10^3,\ T_1 = 80\,10^3)$.



Gelman ratio: Original version: 2.18 ; Re-parametrized: 1.07.

Credibility sets (posterior quantiles): wider.

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$.
   - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
   - Intervals overlapping threshold: extreme or not ?
   - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
   - Sabourin, *under review* ; Sabourin, Renard, *in preparation*

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$.
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin, *under review* ; Sabourin, Renard, *in preparation*

Outline

Multivariate extremes and model uncertainty

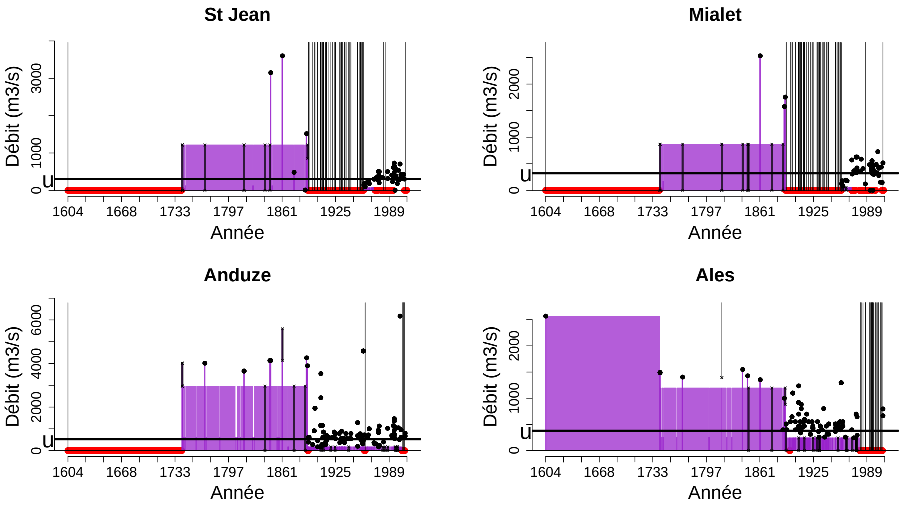
Bayesian model averaging (‘Mélange de modèles’)

Dirichlet mixture model (‘Modèle de mélange’):
a re-parametrization

Historical, censored data in the Dirichlet model

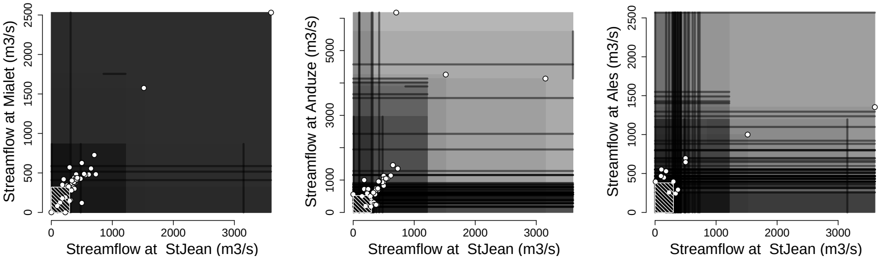
Censored data: univariate and pairwise plots

Univariate time series:



Censored data: univariate and pairwise plots

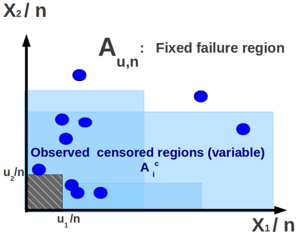
Bivariate plots:



Using censored data: wishes and reality

- Take into account as many data as possible
  $\rightarrow$ Censored likelihood, integration problems
- Information transfer from well gauged to poorly gauged sites using the dependence structure
  $\rightarrow$ Estimate together marginal parameters + dependence

Data overlapping threshold and Poisson model



How to include the rectangles overlapping threshold in the likelihood ?

$$\left\{ \left(\frac{t}{n}, \frac{\mathbf{X}_t}{n} \right), \ 1 \le t \le n \ \right\} \sim \text{Poisson Process } (\mathrm{Leb} \times \lambda) \ \text{ on } [0,1] \times A_{u,n}$$

λ: ' exponent measure', with Dirichlet Mixture angular component

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

Overlapping events appear in Poisson likelihood as

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

'Censored' likelihood: model density integrated over boxes

- Ledford & Tawn, 1996: partially extreme data censored at threshold,
  - GEV models
  - Explicit expression for censored likelihood.

- Here: *idem* + natural censoring
  - Poisson model
  - No closed form expression for integrated likelihood.

- Two terms without closed form:
  - Censored regions A_i overlapping threshold:

$$\exp\left\{-(t_2 - t_1)\lambda(A_i)\right\}$$

  - Classical censoring above threshold

$$\int_{\text{censored region}} \frac{\mathrm{d}\lambda}{\mathrm{d}\mathbf{x}}\,.$$

Data augmentation

One more Gibbs step, no more numerical integration.

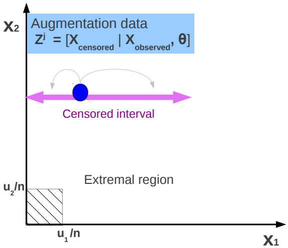
- Objective: sample $[\theta|Obs] \propto$ likelihood (censored obs)
- Additional variables (replace missing data component): $\mathcal{Z}$
- Full conditionals $[Z_i|Z_{j\neq j}, \theta, Obs], [\theta|\mathbf{Z}, Obs], \ldots$ explicit (Thanks Dirichlet): $\rightarrow$ Gibbs sampling.
- Sample $[z, \theta|Obs]_+$ (augmented distribution) on $\Theta \times \mathcal{Z}$.

Censored regions above threshold

$$\int_{\text{Censored region}} \frac{d\lambda}{dx} \, dx_{j_1:j_r} \quad :$$

Generate missing components under univariate conditional distributions

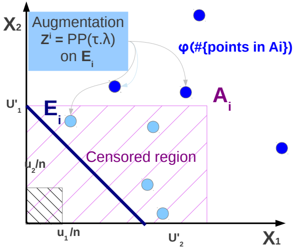
$$\mathbf{Z}^{j}_{1:r} \sim [X_{\text{missing}} | X_{\text{obs}}, \theta]$$



Dirichlet ⇒ Explicit univariate conditionals
Exact sampling of censored data on censored interval

Censored regions overlapping threshold

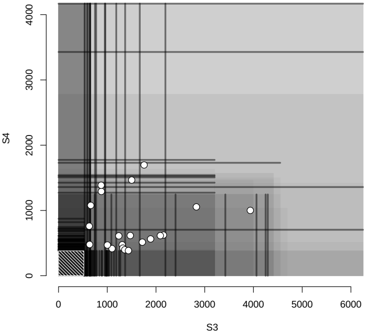
$$e^{-(t_{2,i}-t_{1,i})\lambda(A_i)} \Leftrightarrow \begin{cases} \text{augmentation Poisson process } N_i \text{ on } E_i \supset A_i. \\ + \\ \text{Functional } \varphi(N_i) \end{cases}$$

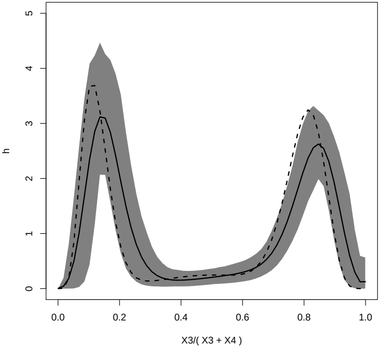


$$[z,\theta|Obs] \propto \underbrace{\cdots}_{\text{density terms, prior, augmented missing components}} [N_i]\varphi(N_i)$$

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

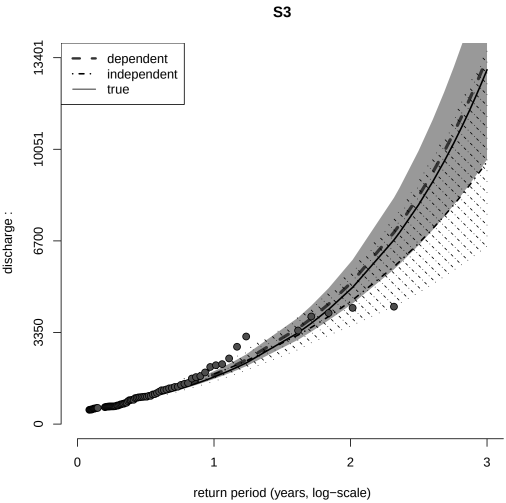
Pairwise plot and angular measure density
(true/ posterior predictive)



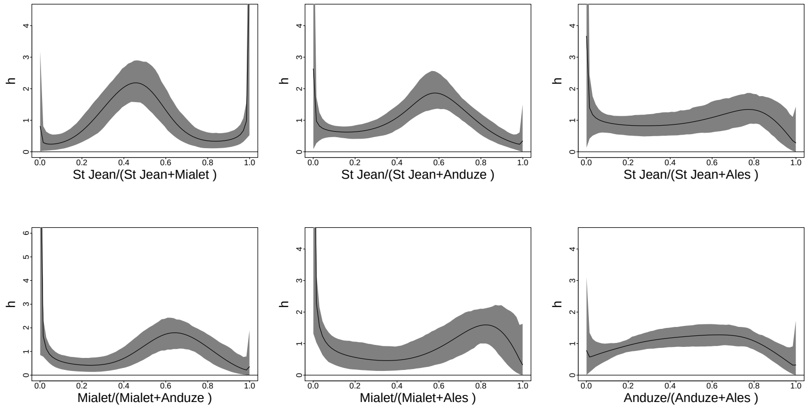


Simulated data (Dirichlet, $d = 4$, $k = 3$ components), same censoring as real data

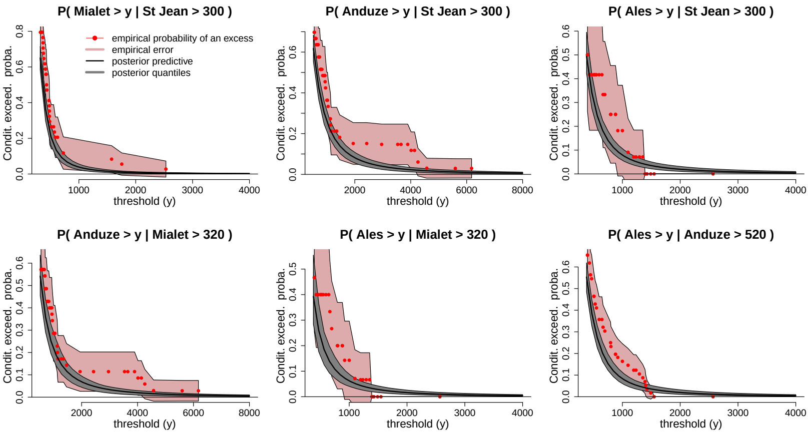
Marginal quantile curves: better in joint model.



Angular predictive density for Gardons data



Conditional exceedance probability



Conclusion

- Building Bayesian multivariate models for excesses:
  - Dirichlet mixture family: 'non' parametric, Bayesian inference possible up to re-parametrization
  - Censoring $\rightarrow$ data augmenting (Dirichlet conditioning properies)
  - Two packages `R`:
    - `DiriXtremes`, MCMC algorithm for Dirichlet mixtures,
    - `DiriCens`, implementation with censored data.
- High dimensional sample space (GCM grid, spatial fields) ?
  - Impose reasonable structure (sparse) on Dirichlet parameters
  - Dirichlet Process ? Challenges :
    Discrete random measure $\neq$ continuous framework

Bibliographie I

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

Coles, SG and Tawn, JA
Modeling extreme multivariate events
JR Statist. Soc. B, 53:377–392, 1991

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

Hosking, J.R.M. and Wallis, J.R..
Regional frequency analysis: an approach based on L-moments
Cambridge University Press, 2005.

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.

Resnick, S. (1987).
Extreme values, regular variation, and point processes, volume 4 of Applied Probability. A Series of the Applied Probability Trust.
Springer-Verlag, New York.

Bibliographie II

Sabourin, A., Naveau, P. and Fougères, A-L. (2013)
Bayesian model averaging for multivariate extremes.
Extremes, 16(3) 325–350

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.

Tanner, M. and Wong, W. (1987).
The calculation of posterior distributions by data augmentation.
Journal of the American Statistical Association, 82(398):528–540.

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

Outline

Multivariate extremes and model uncertainty

Bayesian model averaging ('Mélange de modèles')

Dirichlet mixture model ('Modèle de mélange'):
a re-parametrization

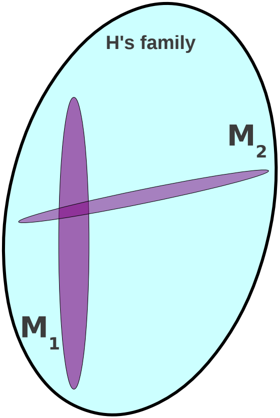
Historical, censored data in the Dirichlet model

Bayesian Model Averaging: reducing model uncertainty.

- Parametric framework: arbitrary restriction, different models can yield different estimates.
- First option: Fight !
  - Choose one model (Information criterions: BIC/ AIC /AICC)

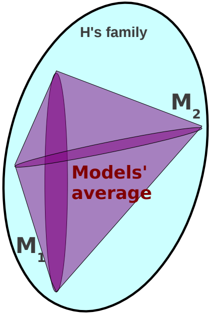
Bayesian Model Averaging: reducing model uncertainty.

- Parametric framework: arbitrary restriction, different models can yield different estimates.
- First option: Fight !
  - Choose one model (Information criterions: BIC/ AIC /AICC)



Bayesian Model Averaging: reducing model uncertainty.

- Parametric framework: arbitrary restriction, different models can yield different estimates.
- BMA = averaging predictions based on posterior model weights



- Already widely studied and used in several contexts (weather forecast . . .).

Hoeting et al. (99), Madigan & Raftery (94), Raftery *et al.* (05)

BMA: principle

- J statistical models $\mathscr{M}_{(1)}, \ldots \mathscr{M}_{(J)}$, with parametrization Θ_j, $1 \leq j \leq J$ and *priors* π_j defined on Θ_j

BMA: principle

- J statistical models $\mathscr{M}_{(1)}, \dots \mathscr{M}_{(J)}$, with parametrization Θ_j, $1 \leq j \leq J$ and *priors* π_j defined on Θ_j
- **BMA model = disjoint union:** $\tilde{\Theta} = \bigsqcup_1^J \Theta_j$,
  with ***prior*** p **on index set** $\{1, \dots, J\}$:
  $p(\mathscr{M}_j)$ = 'prior marginal model weight' for $\mathscr{M}_j$
  - ***prior* on** $\tilde{\Theta}$: $\tilde{\pi}(\bigsqcup_1^J B_j) = \sum_1^J p(\mathscr{M}_j)\, \pi_j(B_j) \quad (B_j \subset \Theta_j)$

BMA: principle

- J statistical models $\mathscr{M}_{(1)}, \ldots \mathscr{M}_{(J)}$, with parametrization Θ_j, $1 \leq j \leq J$ and *priors* π_j defined on Θ_j
- **BMA model = disjoint union:** $\tilde{\Theta} = \bigsqcup_1^J \Theta_j$,
  with ***prior*** p **on index set** $\{1, \ldots, J\}$:
  $p(\mathscr{M}_j)$ = 'prior marginal model weight' for $\mathscr{M}_j$
  - ***prior*** **on** $\tilde{\Theta}$: $\tilde{\pi}(\bigsqcup_1^J B_j) = \sum_1^J p(\mathscr{M}_j)\, \pi_j(B_j) \quad (B_j \subset \Theta_j)$
  - ***posterior*** (conditioning on data X) = weighted average

$$\tilde{\pi}(\bigsqcup_1^J B_j | X) = \sum_1^J \underbrace{p(\mathscr{M}_j \,|\, X)}_{\text{posterior marginal model weight}} \overbrace{\pi_j(B_j \,|\, X)}^{\text{posterior in } \mathscr{M}_j}$$

Key: posterior weights. (Laplace approx or standard MC ?)

$$p(\mathscr{M}_j \,|\, X) \propto p(\mathscr{M}_j) \int_{\Theta_j} \text{Likelihood}(X | \theta)\, d\pi_j(\theta)$$

BMA for multivariate extremes Sabourin, Naveau, Fougères (2013)

- *Background:* univariate EVD's of different types Stephenson & Tawn (04) or multivariate, asymptotically dependent/ independent EVD's Apputhurai & Stephenson (10)

- Our approach: averaging *angular measure* models, with angular data $\mathscr{W}$.
  - $\mathbf{F}_1, \dots, \mathbf{F}_J$ max-stable distributions $\rightarrow \sum_j p_j \mathbf{F}_j$ **not** max-stable.
  - $H_1, \dots H_J$ angular measures (moments constraint) $\rightarrow \sum_j p_j H_j$ **is** a valid angular measure ! (linearity)

BMA for multivariate extremes Sabourin, Naveau, Fougères (2013)

- *Background:* univariate EVD's of different types Stephenson & Tawn (04) or multivariate, asymptotically dependent/ independent EVD's Apputhurai & Stephenson (10)

- Our approach: averaging *angular measure* models, with angular data $\mathscr{W}$.
  - $\mathbf{F}_1, \dots, \mathbf{F}_J$ max-stable distributions $\to \sum_j p_j \mathbf{F}_j$ **not** max-stable.
  - $H_1, \dots H_J$ angular measures (moments constraint) $\to \sum_j p_j H_j$ **is** a valid angular measure ! (linearity)

Implementing and scoring BMA for Multivariate extremes

Does the BMA framework perform significantly better than selecting models based on AIC ?

- Yes, in terms of logarithmic score for the predictive density (Kullback-Leibler divergence to the truth) Madigan & Raftery (94)

In average over the union model, w.r.t prior !

- **Simulation study :** Evaluation *via* proper scoring rules (Logarithmic + probability of failure regions)
  - 2 models of same dimension: Pairwise-Beta / Nested asymmetric logistic
  - 100 data sets simulated from another model
- **Results**: The BMA framework performs
  - consistently (for all scores),
  - slightly (1/20 to 1/100),

  better than model selection.

Implementing and scoring BMA for Multivariate extremes

Does the BMA framework perform significantly better than selecting models based on AIC ?

- Yes, in terms of logarithmic score for the predictive density (Kullback-Leibler divergence to the truth) Madigan & Raftery (94)

In average over the union model, w.r.t prior !

- **Simulation study :** Evaluation *via* proper scoring rules (Logarithmic + probability of failure regions)
  - 2 models of same dimension: Pairwise-Beta / Nested asymmetric logistic
  - 100 data sets simulated from another model
- **Results**: The BMA framework performs
  - consistently (for all scores),
  - slightly (1/20 to 1/100),

  better than model selection.

Implementing and scoring BMA for Multivariate extremes

Does the BMA framework perform significantly better than selecting models based on AIC ?

- Yes, in terms of logarithmic score for the predictive density (Kullback-Leibler divergence to the truth) Madigan & Raftery (94)

  In average over the union model, w.r.t prior !

- **Simulation study :** Evaluation *via* proper scoring rules (Logarithmic + probability of failure regions)
  - 2 models of same dimension: Pairwise-Beta / Nested asymmetric logistic
  - 100 data sets simulated from another model
- **Results**: The BMA framework performs
  - consistently (for all scores),
  - slightly (1/20 to 1/100),

  better than model selection.

Implementing and scoring BMA for Multivariate extremes

Does the BMA framework perform significantly better than selecting models based on AIC ?

- Yes, in terms of logarithmic score for the predictive density (Kullback-Leibler divergence to the truth) Madigan & Raftery (94)

In average over the union model, w.r.t prior !

- **Simulation study :** Evaluation *via* proper scoring rules (Logarithmic + probability of failure regions)
  - 2 models of same dimension: Pairwise-Beta / Nested asymmetric logistic
  - 100 data sets simulated from another model
- **Results**: The BMA framework performs
  - consistently (for all scores),
  - slightly (1/20 to 1/100),

better than model selection.

Discussion

- BMA *vs* selection: Moderate gain for large sample size : Posterior concentration on 'asymptotic carrier regions' = points (parameters) of minimal KL divergence from truth
- BMA : simple if several models have already been fitted ('only' compute posterior weights)
- Way out: Mixture models for increased dimension of the parameter space. (product)