Full likelihood inference for multivariate extreme value distributions

Clément Dombry
Université de Franche-Comté, France

EM approach: M.Genton[1], R.Huser[1] and M.Ribatet[2]
Bayesian framework: S.Engelke[3] and M.Oesting[4]

[1]King Abdullah University of Science and Technology, Saudi Arabia
[2]Université de Montpellier, France [3]EPFL, Switzerland [4]Siegen University, Germany

CIRM, February 22-26, 2016.

Main motivation

- For a parametric model $\mathbf{Z} \sim F_\theta$ of multivariate **max-stable distributions**, the full LLHs are usually intractable.

Main motivation

- For a parametric model $\mathbf{Z} \sim F_\theta$ of multivariate **max-stable distributions**, the full LLHs are usually intractable.
- Direct **maximum likelihood estimation** is infeasible and one popular alternative is **pairwise composite likelihood estimation** (Padoan et al., 2010).

Main motivation

- For a parametric model $\mathbf{Z} \sim F_\theta$ of multivariate **max-stable distributions**, the full LLHs are usually intractable.
- Direct **maximum likelihood estimation** is infeasible and one popular alternative is **pairwise composite likelihood estimation** (Padoan et al., 2010).
- **Goal :** introduce two methodological frameworks that can deal with full LLHs : **EM approach** and **Bayesian setup**.
  - **improve frequentist efficiency**,
  - allow for **Bayesian methods in extremes**.

Multivariate max-stable distributions

- Max-stable distributions are mathematically justified distributions that can be used in the **block maxima method** in extreme value theory.
- A multivariate r.v. $\mathbf{Z} \sim F$ is called **simple max-stable** if the margins are **unit Fréchet** and the **max-stability property** is satisfied

$$F(\mathbf{z})^n = F(\mathbf{z}/n), \quad \mathbf{z} \in (0, +\infty)^d, \; n \geq 1.$$

- The cdf F has the particular form

$$F(\mathbf{z}) = \exp(-V(\mathbf{z})), \quad \mathbf{z} \in (0, +\infty)^d,$$

with V a -1 homogeneous function, i.e., $V(u\mathbf{z}) = u^{-1} V(\mathbf{z})$, called the **exponent function**.

Density of multivariate max-stable distributions

- **Differentiating the cdf** $F = \exp(-V)$ yields the density :

$$d = 2: \qquad f = \exp(-V)\,(-\partial_{12}V + \partial_1 V \partial_2 V)$$

$$d = 3: \qquad f = \exp(-V)\,(-\partial_{123}V + \partial_1 V \partial_{23}V + \partial_2 V \partial_{13}V + \partial_3 V \partial_{12}V - \partial_1 V \partial_2 V \partial_3 V)$$

Density of multivariate max-stable distributions

- **Differentiating the cdf** $F = \exp(-V)$ yields the density :

$$d = 2: \qquad f = \exp(-V)\,(-\partial_{12}V + \partial_1 V \partial_2 V)$$
$$d = 3: \qquad f = \exp(-V)\,(-\partial_{123}V + \partial_1 V\partial_{23}V + \partial_2 V\partial_{13}V + \partial_3 V\partial_{12}V - \partial_1 V\partial_2 V\partial_3 V)$$

- General case by **Faa-di Bruno** derivation formula :

$$f = \exp(-V) \sum_{\tau \in \mathcal{P}_d} \prod_{j=1}^{\ell} (-\partial_{\tau_j} V)$$

with $\mathcal{P}_d$ the set of partitions $\tau = (\tau_1, \ldots, \tau_\ell)$ of $\{1, \ldots, d\}$.

Density of multivariate max-stable distributions

- **Differentiating the cdf** $F = \exp(-V)$ yields the density :

$$d = 2: \qquad f = \exp(-V)\,(-\partial_{12}V + \partial_1 V \partial_2 V)$$
$$d = 3: \qquad f = \exp(-V)\,(-\partial_{123}V + \partial_1 V \partial_{23} V + \partial_2 V \partial_{13} V + \partial_3 V \partial_{12} V - \partial_1 V \partial_2 V \partial_3 V)$$

- General case by **Faa-di Bruno** derivation formula :

$$f = \exp(-V) \sum_{\tau \in \mathcal{P}_d} \prod_{j=1}^{\ell} (-\partial_{\tau_j} V)$$

with $\mathcal{P}_d$ the set of partitions $\tau = (\tau_1, \ldots, \tau_\ell)$ of $\{1, \ldots, d\}$.

- **Combinatorial explosion** of the number B_d of terms :

$$B_5 \approx 10^2, \quad B_{10} \approx 10^5, \quad B_{20} \approx 10^{13}.$$

LmB

Interpretation : partition of occurrence times

- Let $\mathbf{Z} \sim F$ and $\mathbf{X}_i$, $i = 1, 2, \ldots$ (with Fréchet margins) in its MDA :

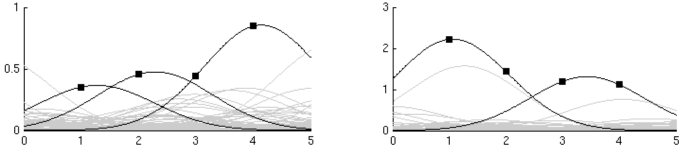
$$\mathbf{M}_n = n^{-1} \max_{1 \leq i \leq n} \mathbf{X}_i \xrightarrow{d} \mathbf{Z}.$$

Interpretation : partition of occurrence times

- Let $\mathbf{Z} \sim F$ and $\mathbf{X}_i$, $i = 1, 2, \ldots$ (with Fréchet margins) in its MDA :

$$\mathbf{M}_n = n^{-1} \max_{1 \leq i \leq n} \mathbf{X}_i \xrightarrow{d} \mathbf{Z}.$$

- Random partition $T_n \in \mathcal{P}_d$ associated to **occurrence times** of maxima, i.e., j and k in the same set if $M_{n,j}$ and $M_{n,k}$ come from same the $\mathbf{X}_i$.



Example with $d = 4$, $n = 50$: Left : $T_n = \{1, 2, 34\}$. Right : $T_n = \{12, 34\}$.

(LmB)

Stephenson-Tawn approach

- Weak convergence as $n \to \infty$: $(\mathbf{M}_n, T_n) \xrightarrow{d} (\mathbf{Z}, T)$.
- Stephenson and Tawn (2005) propose to use the **additional information of occurrence times** Π based on the **joint LLH** of $(\mathbf{Z}, T)$ with simple form

$$L(\mathbf{z}, \tau) = \exp\{-V(\mathbf{z})\} \prod_{j=1}^{\ell} \{-\partial_{\tau_j} V(\mathbf{z})\}.$$

- Wadsworth (2015) has shown that the poor approximation $T_n \approx T$ may cause bias and has proposed a **first order bias correction**.

Full LLH based inference for $\mathbf{Z} \sim F_\theta$

- Parametric model $\mathbf{Z} \sim F_\theta$, $\theta \in \Theta$, with likelihood

$$L(\mathbf{z}|\theta) = \sum_{\tau \in \mathcal{P}_d} L(\mathbf{z}, \tau \mid \theta)$$

Full LLH based inference for $\mathbf{Z} \sim F_\theta$

- Parametric model $\mathbf{Z} \sim F_\theta$, $\theta \in \Theta$, with likelihood

$$L(\mathbf{z}|\theta) = \sum_{\tau \in \mathcal{P}_d} L(\mathbf{z}, \tau \mid \theta)$$

- Treat the occurrence time T has an **unobserved latent variable**.

Observations from $\mathbf{Z}$:	$\mathbf{z_1}, \ldots, \mathbf{z_n} \in \mathbb{R}^d$
Unobserved partitions from T :	$\tau_1, \ldots, \tau_n \in \mathcal{P}_d$
Model parameters :	$\theta \in \Theta$

Full LLH based inference for $\mathbf{Z} \sim F_\theta$

- Parametric model $\mathbf{Z} \sim F_\theta$, $\theta \in \Theta$, with likelihood

$$L(\mathbf{z}|\theta) = \sum_{\tau \in \mathcal{P}_d} L(\mathbf{z}, \tau \mid \theta)$$

- Treat the occurrence time T has an **unobserved latent variable**.

Observations from $\mathbf{Z}$:	$\mathbf{z_1}, \ldots, \mathbf{z_n} \in \mathbb{R}^d$
Unobserved partitions from T :	$\tau_1, \ldots, \tau_n \in \mathcal{P}_d$
Model parameters :	$\theta \in \Theta$

- Methodology :
  - frequentist setup : **EM algorithm** for missing observations ;
  - Bayesian setup : **hierarchical model** with **prior distribution** $\pi(\theta)$.

(LmB)

Frequentist setup : EM approach

- EM algorithm is a 2-step recursive procedure :
  - **E** : given θ_t, compute the **conditional expectation**

$$Q(\theta, \theta_t) = \mathbb{E}_{\tau|\mathbf{z},\theta_t}\left[\log L(\mathbf{z}, T|\theta)\right];$$

  - **M** : compute the **maximizer**

$$\theta_{t+1} = \operatorname*{argmax}_{\theta} Q(\theta, \theta_t).$$

Frequentist setup : EM approach

- EM algorithm is a 2-step recursive procedure :
  - **E** : given θ_t, compute the **conditional expectation**

$$Q(\theta, \theta_t) = \mathbb{E}_{\tau|\mathbf{z},\theta_t}\left[\log L(\mathbf{z}, T|\theta)\right];$$

  - **M** : compute the **maximizer**

$$\theta_{t+1} = \operatorname*{argmax}_{\theta} Q(\theta, \theta_t).$$

- Use rather a stochastic EM algorithm :
  - **MC-E** : given θ_t, compute the **Monte-Carlo expectation**

$$Q(\theta, \theta_t) = \frac{1}{N}\sum_{i=1}^{N} \log L(\mathbf{z}, T^i|\theta_t) \quad \text{with} \quad (T^i)_{1\leq i\leq N} \overset{i.i.d.}{\sim} L(\tau \mid \mathbf{z}, \theta_t).$$

Frequentist setup : EM approach

- EM algorithm is a 2-step recursive procedure :
  - **E** : given θ_t, compute the **conditional expectation**

$$Q(\theta, \theta_t) = \mathbb{E}_{\tau|\mathbf{z},\theta_t} [\log L(\mathbf{z}, T|\theta)] \,;$$

  - **M** : compute the **maximizer**

$$\theta_{t+1} = \operatorname*{argmax}_{\theta} Q(\theta, \theta_t).$$

- Use rather a stochastic EM algorithm :
  - **MC-E** : given θ_t, compute the **Monte-Carlo expectation**

$$Q(\theta, \theta_t) = \frac{1}{N} \sum_{i=1}^{N} \log L(\mathbf{z}, T^i|\theta_t) \quad \text{with} \quad (T^i)_{1 \leq i \leq N} \overset{i.i.d.}{\sim} L(\tau \mid \mathbf{z}, \theta_t).$$

- Simplified notations here : n observations $\rightsquigarrow$ n partitions.

LmB

Bayesian setup : MCMC approach

- Assess the **posterior distribution**

$$L(\theta, \tau_1, \ldots, \tau_n \mid \mathbf{z_1}, \ldots, \mathbf{z_n}) \propto \pi(\theta) \prod_{i=1}^{n} L(\mathbf{z_i}, \tau_i | \theta).$$

Bayesian setup : MCMC approach

- Assess the **posterior distribution**

$$L(\theta, \tau_1, \ldots, \tau_n \mid \mathbf{z_1}, \ldots, \mathbf{z_n}) \propto \pi(\theta) \prod_{i=1}^{n} L(\mathbf{z_i}, \tau_i | \theta).$$

- MCMC approach with separate alternative updates :
  - standard **Metropolis-Hastings** for θ ;
  - **conditional sampling** for $\tau_1, \ldots, \tau_n$ according to $L(\tau_i \mid \mathbf{z_i}, \theta)$.

Bayesian setup : MCMC approach

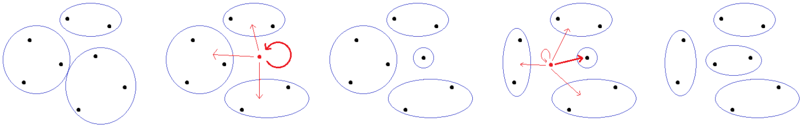
- Assess the **posterior distribution**

$$L(\theta, \tau_1, \ldots, \tau_n \mid \mathbf{z_1}, \ldots, \mathbf{z_n}) \propto \pi(\theta) \prod_{i=1}^{n} L(\mathbf{z_i}, \tau_i | \theta).$$

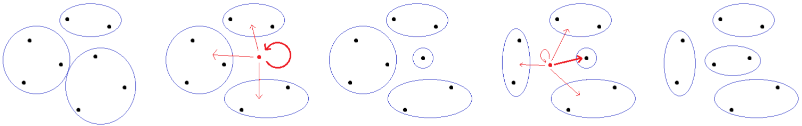
- MCMC approach with separate alternative updates :
  - standard **Metropolis-Hastings** for θ ;
  - **conditional sampling** for $\tau_1, \ldots, \tau_n$ according to $L(\tau_i \mid \mathbf{z_i}, \theta)$.
- Both approaches require :
  - **conditional sampling** : Gibbs sampler by Dombry et al. (2013) ;
  - **explicit formulas** for $V_\theta(\mathbf{z})$, $\partial_{\tau_j} V_\theta(\mathbf{z})$.

LmB

Gibbs sampler for $L(\tau \mid \mathbf{z}, \theta)$



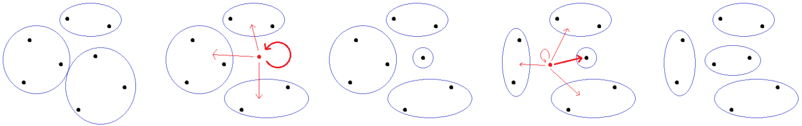
Gibbs sampler for $L(\tau \mid \mathbf{z}, \theta)$



- **Combinatorial explosion avoided** : number of possible updates $\tau^* \in \mathcal{P}_d$ such that $\tau^*_{-j} = \tau_{-j}$ is

$$\begin{cases} |\tau| & \text{if } \{x_j\} \text{ is a partitioning set of } \tau, \\ |\tau| + 1 & \text{otherwise.} \end{cases}$$

Gibbs sampler for $L(\tau \mid \mathbf{z}, \theta)$



- **Combinatorial explosion avoided** : number of possible updates $\tau^* \in \mathcal{P}_d$ such that $\tau^*_{-j} = \tau_{-j}$ is

$$\begin{cases} |\tau| & \text{if } \{x_j\} \text{ is a partitioning set of } \tau, \\ |\tau| + 1 & \text{otherwise.} \end{cases}$$

- **Proposal distribution** easily computed :

$$\mathbb{P}[T = \tau^* \mid T_{-j} = \tau_{-j}] \propto \frac{\prod_{k=1}^{|\tau^*|}\{-\partial_{\tau^*_k} V_\theta(\mathbf{z})\}}{\prod_{k=1}^{|\tau|}\{-\partial_{\tau_k} V_\theta(\mathbf{z})\}}.$$

where many terms cancel out except at most for of them.

Explicit formulas for $V_\theta(\mathbf{z})$, $\partial_{\tau_j} V_\theta(\mathbf{z})$

- In this talk, only the **simple logistic model** with $\theta \in (0,1)$:

$$
\begin{aligned}
V_\theta(\mathbf{z}) &= \left(z_1^{-1/\theta} + \cdots + z_d^{-1/\theta}\right)^\theta, \\
\partial_{\tau_j} V_\theta(\mathbf{z}) &= \theta^{1-|\tau_j|} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} \left(\sum_{i=1}^{d} z_i^{-1/\theta}\right)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta}
\end{aligned}
$$

[LmB]

Explicit formulas for $V_\theta(\mathbf{z})$, $\partial_{\tau_j} V_\theta(\mathbf{z})$

- In this talk, only the **simple logistic model** with $\theta \in (0,1)$:

$$\begin{aligned} V_\theta(\mathbf{z}) &= \left(z_1^{-1/\theta} + \cdots + z_d^{-1/\theta} \right)^\theta, \\ \partial_{\tau_j} V_\theta(\mathbf{z}) &= \theta^{1-|\tau_j|} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} \left(\sum_{i=1}^d z_i^{-1/\theta} \right)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta} \end{aligned}$$

- Many other models where (integral) formulas are available :
  - Brown-Resnick
  - Extremal-t
  - Reich-Shaby
  - Dirichlet

Simulation study : logistic model

- Simulate $n = 100$ samples $\mathbf{z}_1, \ldots \mathbf{z}_n$ from the d-dim. **max-stable logistic model** for $\mathbf{Z}$ with parameter $\theta_0 \in \{0.1, 0.7, 0.9\}$.

Simulation study : logistic model

- Simulate $n = 100$ samples $\mathbf{z}_1, \dots \mathbf{z}_n$ from the d-dim. **max-stable logistic model** for $\mathbf{Z}$ with parameter $\theta_0 \in \{0.1, 0.7, 0.9\}$.
- Run MC with **uniform prior** γ on $(0, 1)$, and take **empirical median** of posterior $L(\theta|\mathbf{z}_1, \dots \mathbf{z}_n)$ as point estimate $\hat{\theta}_{\text{Bayes}}$.

Simulation study : logistic model

- Simulate $n = 100$ samples $\mathbf{z}_1, \dots \mathbf{z}_n$ from the d-dim. **max-stable logistic model** for $\mathbf{Z}$ with parameter $\theta_0 \in \{0.1, 0.7, 0.9\}$.
- Run MC with **uniform prior** γ on $(0, 1)$, and take **empirical median** of posterior $L(\theta|\mathbf{z}_1, \dots \mathbf{z}_n)$ as point estimate $\hat{\theta}_{\text{Bayes}}$.

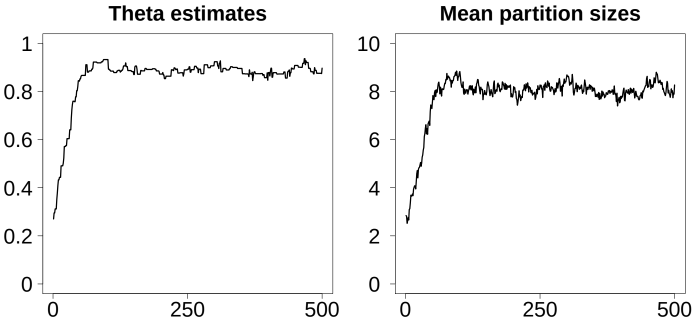


FIGURE: Theta estimates (left) and partition size (right) along the Markov Chain ; $\theta_0 = 0.9$, $d = 10$.

Simulation study : Bayesian VS Pairwise Likelihood

	$\theta_0 = 0.1$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
d	6	10	50	6	10	50	6	10	50
$\text{Bias}(\hat{\theta}_{\text{Bayes}})$	2	2	2	10	6	1	-6	-3	2
$s(\hat{\theta}_{\text{Bayes}})$	36	27	12	240	179	79	239	182	84
$\text{Bias}(\hat{\theta}_{\text{PL}})$	1	0	2	13	12	16	26	31	41
$s(\hat{\theta}_{\text{PL}})$	40	30	13	275	237	173	313	273	246

TABLE: Sample bias and standard deviation of $\hat{\theta}_{\text{Bayes}}$ and $\hat{\theta}_{\text{PL}}$, estimated from 1500 estimates ; figures multiplied by 10000.

	$\theta_0 = 0.1$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
	6	10	50	6	10	50	6	10	50
$\frac{MSE(\hat{\theta}_{\text{Bayes}})}{MSE(\hat{\theta}_{\text{PL}})}$	82	83	78	76	57	21	58	44	11

TABLE: Relative efficiencies (%) of $\hat{\theta}_{\text{Bayes}}$ compared to $\hat{\theta}_{\text{PL}}$.

Observations :

- Posterior median $\hat{\theta}_{\text{Bayes}}$ is **unbiased**.
- Substantially **reduced std. deviations and MSEs** with full LLHs.

LmB

Simulation study : marginal parameters

- Data : $n = 100$ observations from the logistic model ($d = 10$) with parameter θ_0 and $GEV(\mu_0, \sigma_0, \xi_0)$ margins.
- MSE comparison of Bayes, pairwise LLH and independence LLH estimators :

	$\theta_0 = 0.4$ (known)		
	μ_0	σ_0	ξ_0
Bayes	182	124	10
Pairwise	201	180	24
Independence	220	305	89

	$\theta_0 = 0.4$ (unknown)		
	μ_0	σ_0	ξ_0
Bayes	207	288	33
Pairwise	205	269	53
Independence	220	305	89

TABLE: MSEs of (μ_0, σ_0, ξ_0)-estimates with Bayesian approach, pairwise LLH and independence LLH, respectively ; figures multiplied by 10000.

- **Observations :**
  - Small efficiency gain for μ_0 and σ_0.
  - Larger gain for the shape parameter ξ_0.

Simulation study : in progress

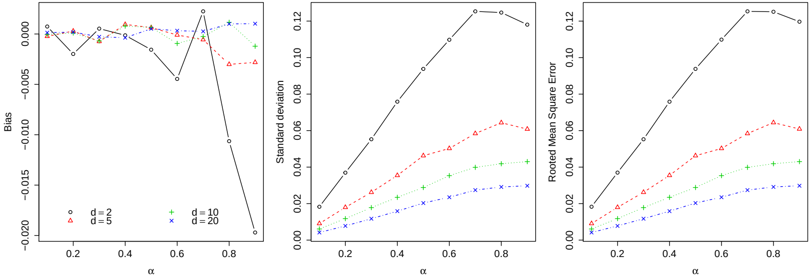
- Robustness with data in the domain of attraction simulated from an outer Clayton copula.
- Test for a linear trend in the marginal parameters :

$$Z_i \sim GEV(\mu_i, \sigma_i, \xi_i) \quad \text{with} \left\{ \begin{array}{lcl} \mu_i & = & \mu_0 + \alpha \cdot i \\ \sigma_i & = & \sigma_0 + \beta \cdot i \\ \xi_i & = & \xi_0 \end{array} \right. .$$

A bayesian test procedure that takes into account (logistic) dependence among the Z_i ?

Logistic model : EM-approach

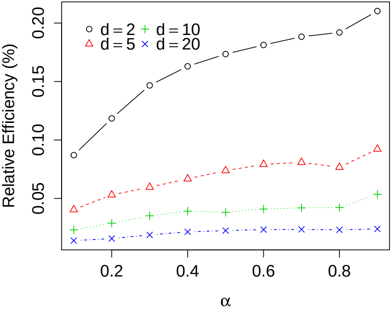
- Logistic model : $n = 20$ observations with $d \in \{2, 5, 10, 20\}$ and $\theta \in \{0.1, \ldots, 0.9\}$;
- MC-E : $N = 1000$ replicates via Gibbs sampling with d-thinning.
- EM-MLE : 50-EM iterations averaged over the last 30 iterations.
- Bias, standard deviation and rooted mean squared error :



Logistic model : EM-approach

- Comparison True-MLE VS EM-MLE : relative error

$$RE = 100 \cdot \mathbb{E}\left|\frac{\hat{\alpha}^{EM} - \hat{\alpha}^{MLE}}{\hat{\alpha}^{MLE}}\right|.$$



References

C. Dombry, F. Eyi-Minko, and M. Ribatet.
Conditional simulation of max-stable processes.
Biometrika, 100(1) :111–124, 2013.

S. A. Padoan, M. Ribatet, and S. A. Sisson.
Likelihood-based inference for max-stable processes.
J. Amer. Statist. Assoc., 105 :263–277, 2010.

A. Stephenson and J. A. Tawn.
Exploiting occurrence times in likelihood inference for componentwise maxima.
Biometrika, 92(1) :213–227, 2005.

E. Thibaud, J. Aalto, D. S. Cooley, A. C. Davison, and J. Heikkinen.
Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures.
Available from `http://arxiv.org/abs/1506.07836`, 2015.

J. L. Wadsworth.
On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions.
Biometrika, 2015.