

1) Introduction

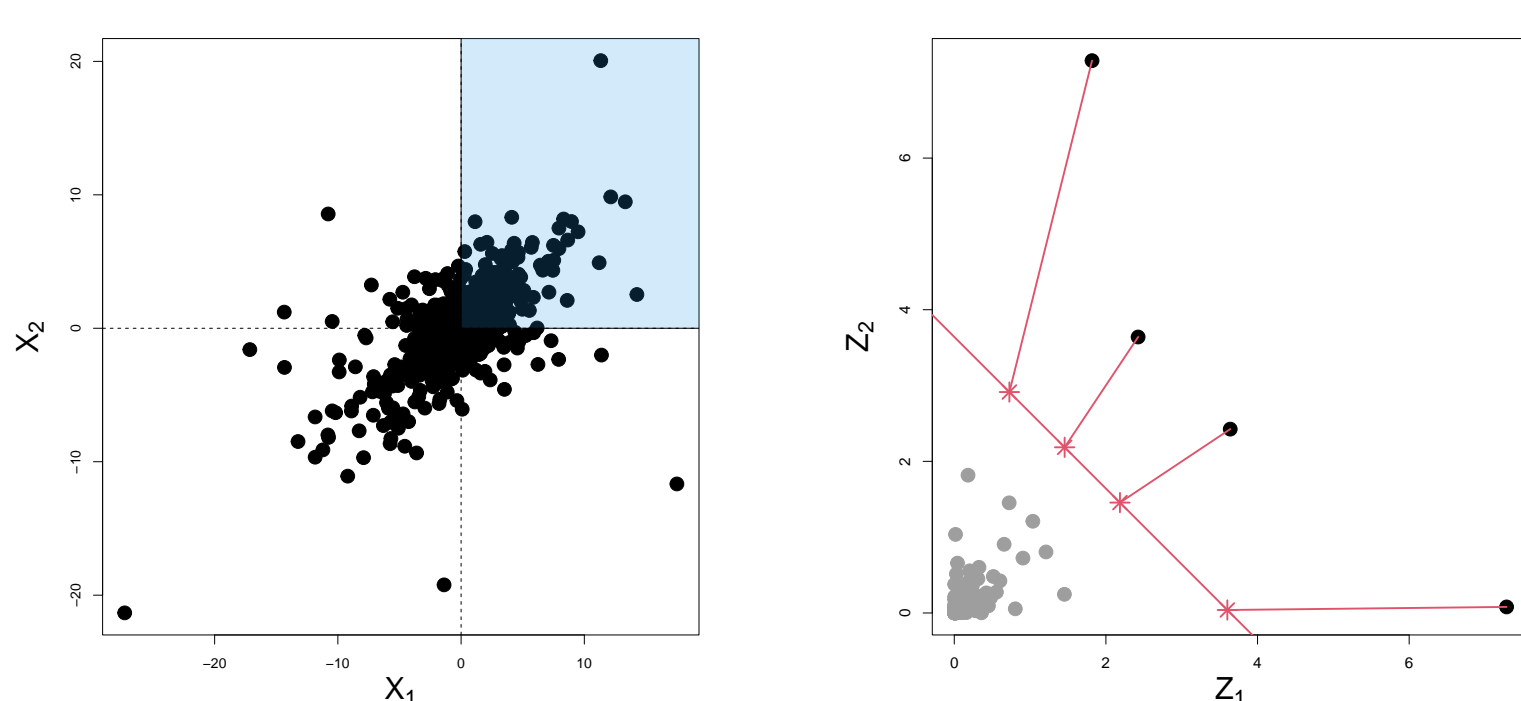
For a comprehensive understanding of extreme events, we need to estimate the probabilities of extreme values occurring jointly across multiple variables.

Objective: To model the joint behavior of extreme variables

The common procedure consists in

1. Modeling the marginal variables individually, then rescaling them to have a common standard distribution,
→ We consider a random variable $\mathbf{Z} \in \mathbb{R}_+^d$ with known unit Fréchet margins.
2. Focusing on modeling their dependence structure.
→ For this purpose, we transform to **angular variables**: $\mathbf{Z} \rightarrow (R, \mathbf{W})$,

$$R = \|\mathbf{Z}\|_1 \text{ and } \mathbf{W} = \mathbf{Z}/R.$$



- The radius and the angle are independent as $R \rightarrow \infty$.
- The asymptotic distribution $\Pr[\mathbf{Z} \geq \mathbf{z} \mid R \geq u]$ can be fully characterised by the **angular density function** h of \mathbf{W} as the threshold $u \rightarrow \infty$.

(!!) The density function h necessarily lies in

$$\mathcal{H} = \{h : \mathbb{S}_D \rightarrow [0, 1] : \int w_d h(w) dw = D^{-1}, d = 1, \dots, D\}$$

where $\mathbb{S}_D = \{w \in \mathbb{R}_+^D : \|w\|_1 = 1\}$.

2) The Dirichlet mixture model

As functions with support on the unit simplex, any angular density in \mathcal{H} can be approximated by a **constrained** Dirichlet mixture for some $K \in \mathbb{N}_+$:

$$f(\mathbf{w}) = \sum_{k=1}^K \eta_k B^{-1}(\alpha_k) \prod_{d=1}^D w_d^{\alpha_{kd}-1}, \quad \mathbf{w} \in \mathbb{S}_D, \quad (\text{cMD})$$

- $\eta = (\eta_1, \dots, \eta_K) \geq 0$, such that $\sum_{k=1}^K \eta_k = 1$,
- $\forall k = 1, \dots, K, \alpha_k \geq 0$, such that

$$\sum_{k=1}^K \eta_k \frac{\alpha_{kd}}{\sum_{p=1}^D \alpha_{kp}} = D^{-1}, \quad d = 1, \dots, D. \quad (\star)$$

However, the implementation of such models tends to be too slow especially in high dimensions.

(Boldi and Davison, 2007; Sabourin and Naveau, 2014)

3) The tilted Dirichlet mixture model

A theorem of Coles and Tawn (1991) gives a way to generate valid angular densities by transforming non-constrained ones. **The tilted Dirichlet mixture** is

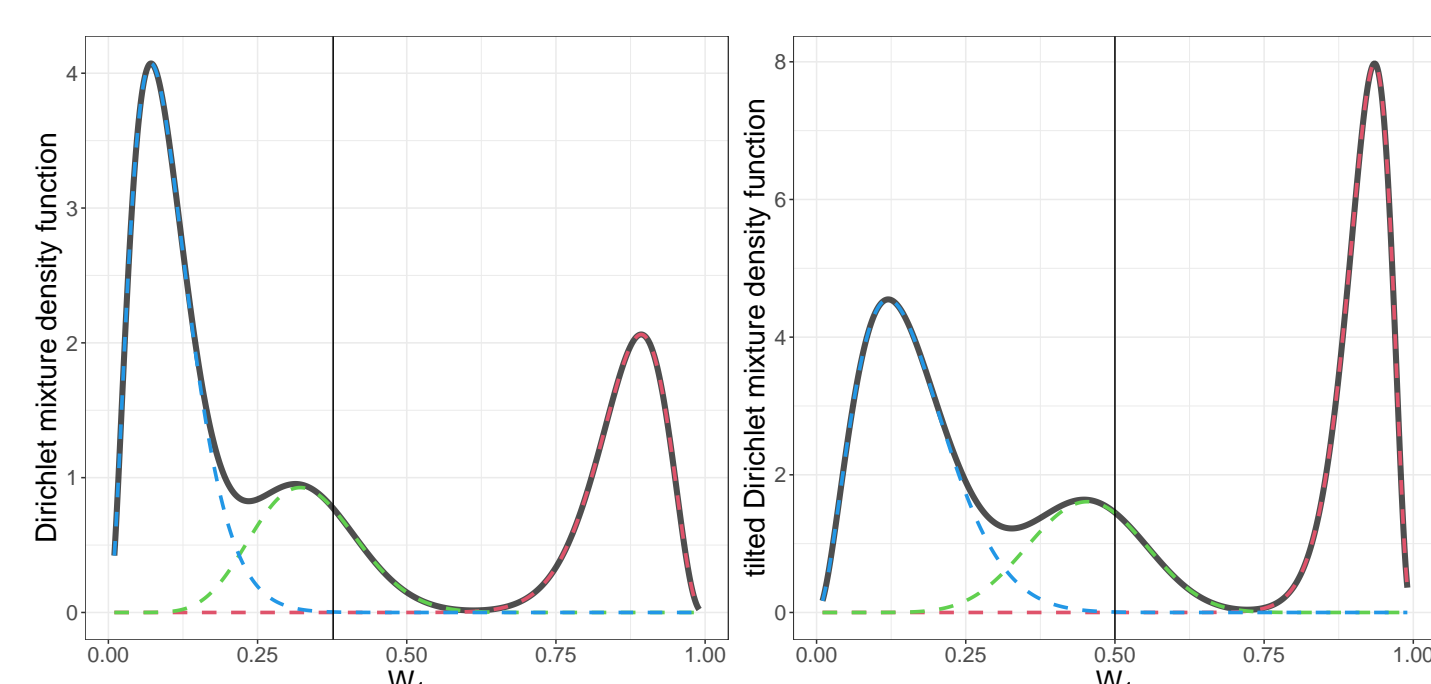
$$f(w) = \sum_{k=1}^K \eta_k B(\alpha_k)^{-1} (w' \alpha_k)^{-1' \alpha_k} \prod_{d=1}^D m_d^{\alpha_{kd}} \prod_{d=1}^D w_d^{\alpha_{kd}-1}, \quad (\text{tMD})$$

$:= f_k(w)$

for all $w \in \mathbb{S}_D$, where $m_d := \sum_{k=1}^K \eta_k \alpha_{kd} / (1' \alpha_k)$, for $d = 1, \dots, D$.

Theorem:

The class of (tDM) density functions is dense in \mathcal{H} .



Original and tilted Dirichlet mixtures

The mixture allow the probabilistic clustering of observations, so multivariate exceedances can be represented via clusters concentrated in distinct regions of the space which can be analysed separately in lower dimensions.

4) Reversible jump MCMC algorithm

Data Augmentation: We denote the parameter vector by $\theta = (K, \alpha_1, \dots, \alpha_K, \eta)$ and consider the observations $\mathbf{w}_1, \dots, \mathbf{w}_N$. We augment the data with latent allocation variables Y_1, \dots, Y_N such that

$$Y_i = k \Leftrightarrow \mathbf{w}_i \text{ generated by component } k.$$

Complete likelihood: $L(\theta) = \prod_{i=1}^N f(\mathbf{w}_i, Y_i \mid \theta)$ where

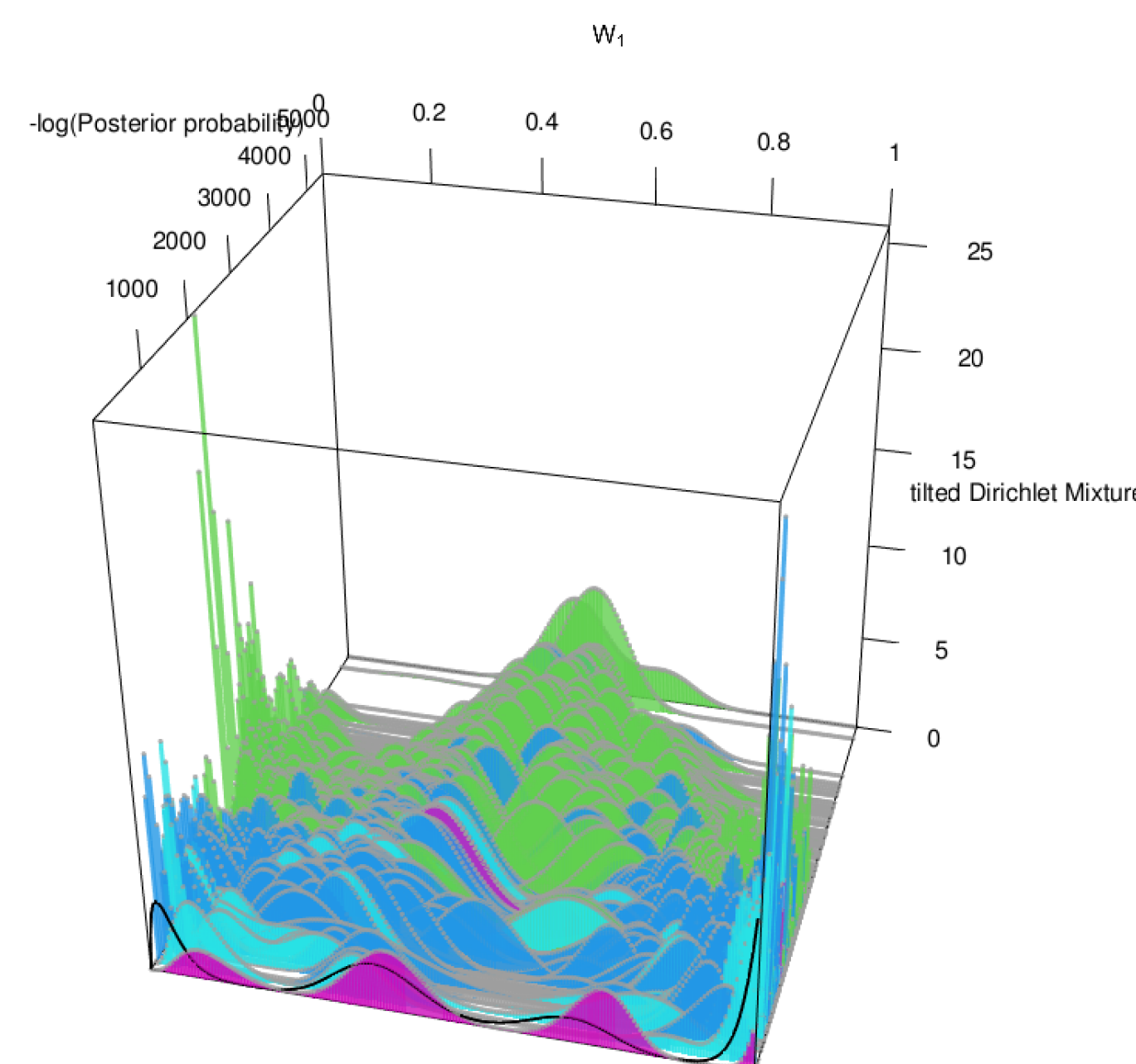
$$f(\mathbf{w}_i, Y_i \mid \theta) = \prod_{k=1}^K f_k(\mathbf{w}_i)^{1_{Y_i=k}}.$$

At each iteration, we first keep K constant and run a Metropolis–Hasting algorithm.

1. The proposed parameters do not need to satisfy (\star) , so the parameters can be proposed more easily.
2. The acceptance probability is calculated with by evaluating the *tilted* likelihood function, hence the accepted parameters are consistent with a valid angular density.

We extend each iteration with a Split/Merge step following the algorithm introduced by Richardson and Green (1997) in order to sample the number of components K in $\{K_{\min}, \dots, K_{\max}\}$.

1. We either merge two adjacent components or split one into two adjacent ones.
2. Both moves have equal probability and are defined to be the reverse of each other.
3. The probability of accepting each move is based on the *tilted* likelihood function.

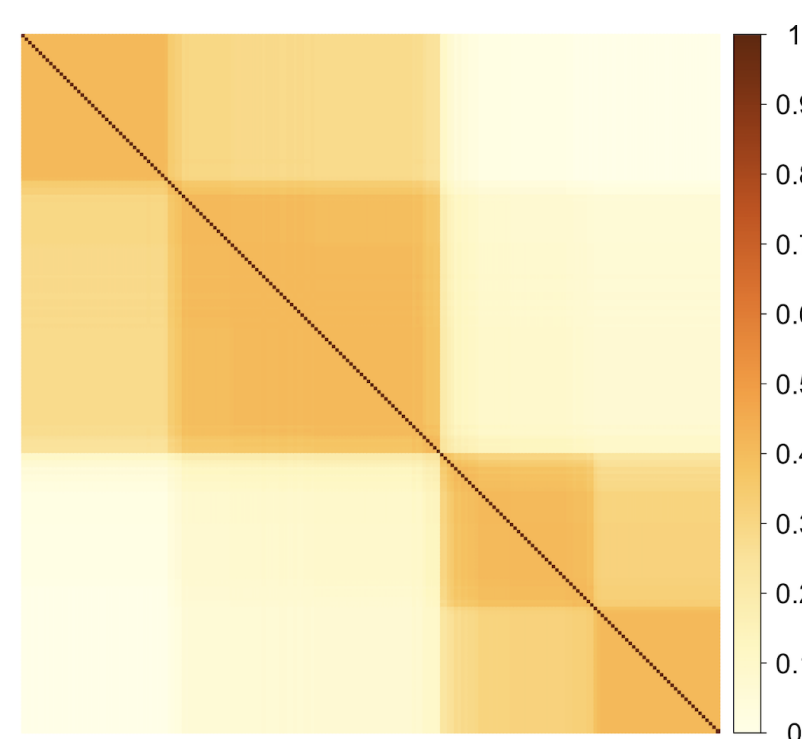


The MCMC algorithm generates a chain of density functions, colored here based on the number of components. The black curve represents the true density of the data sample.

5) Clustering the observations

The tilted Dirichlet mixture generates a soft clustering of the data into groups concentrated into distinct subspaces of the unit simplex, which we can summarize in any dimension D and for any component number K via the $N \times N$ expected posterior adjacency matrix

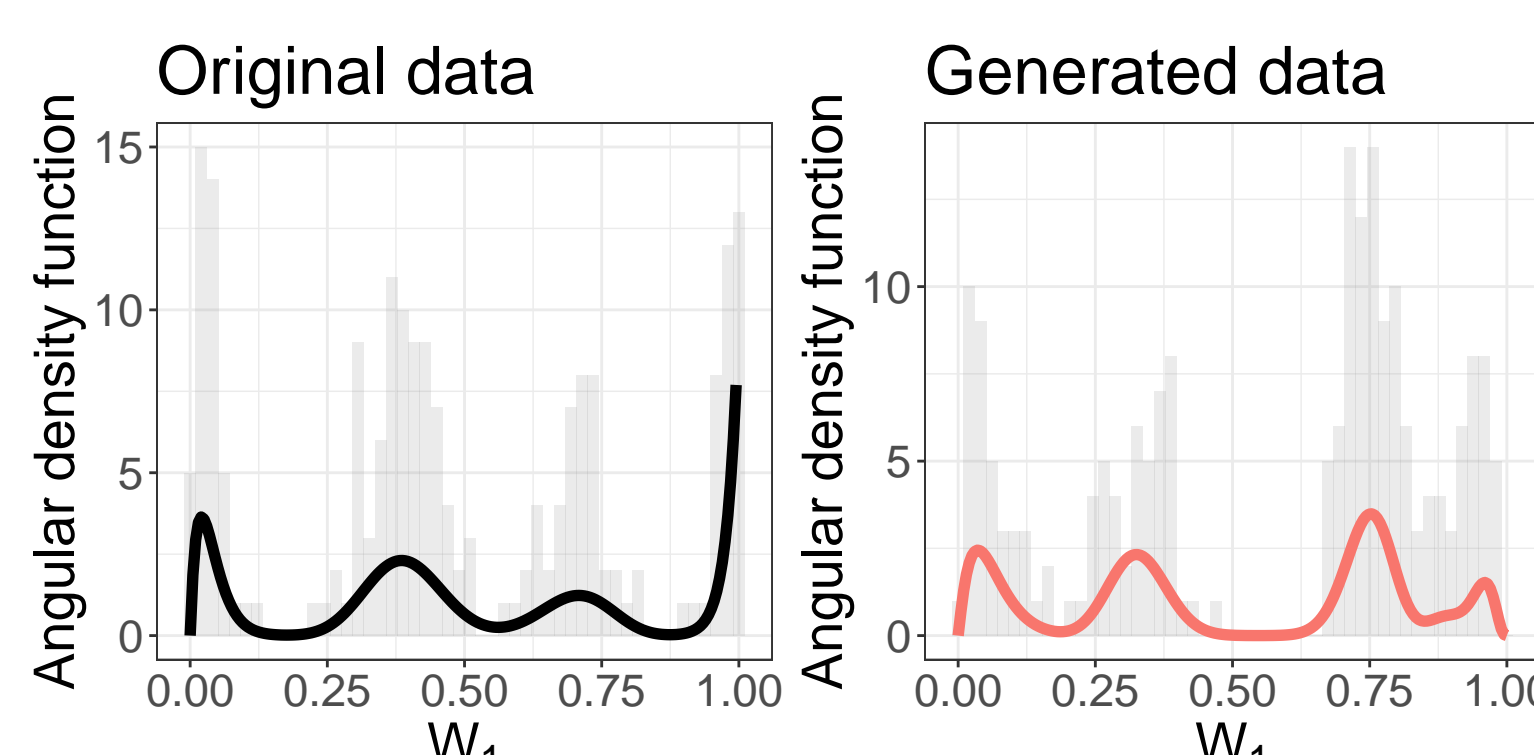
$$P_{ij} = \Pr[w_i \text{ and } w_j \text{ are clustered together}].$$



Among the chain of clusterings generated by the MCMC algorithm, we can select the one optimising a given criterion, for example minimizing Binder's distance to P

$$\hat{A} = \operatorname{argmin} \left\{ A : (N(N-1))^{-1} \sum_{i \neq j} |A_{ij} - P_{ij}| \right\}. \quad (\text{Binder})$$

We can then use the associated parameters to generate new data.



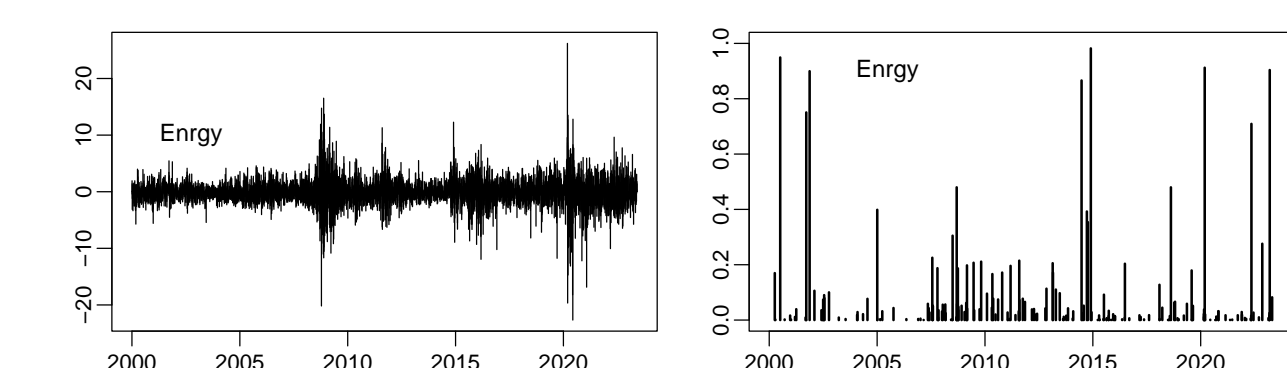
6) The running time of the algorithm

$D \setminus N$	50	200	500
5	3.75	5.09	11.20
10	3.74	5.66	12.50
20	3.05	5.74	13.31

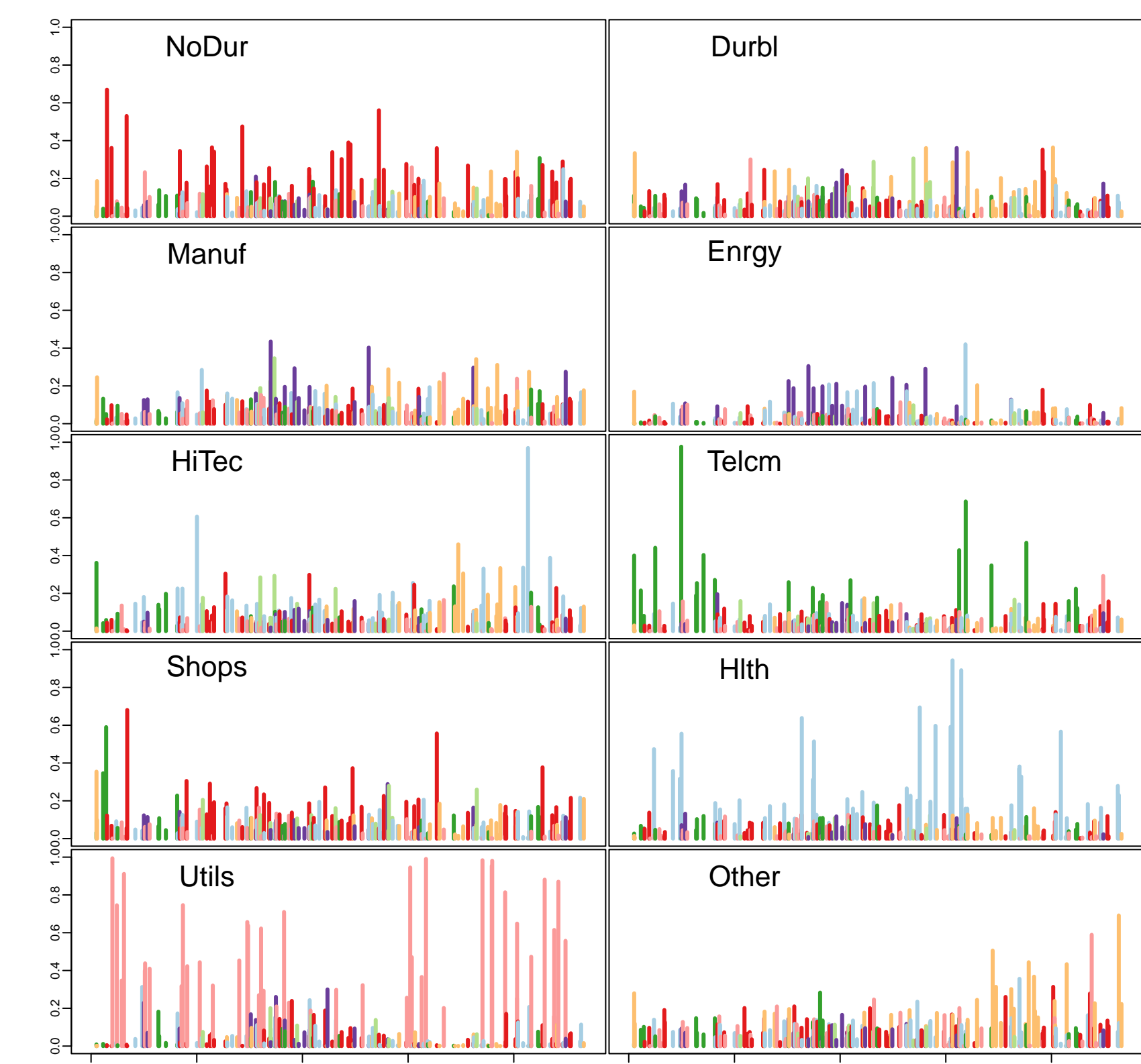
The running time in minutes of 10000 iterations increases with the sample size N but not with the dimensions D .

7) Application: Financial risk

- We consider 10 time series of daily negative returns in different sectors of the market, observed simultaneously for 20 years.
- We fit a ARMA–GARCH(1,1) model to each margin separately with t -student residuals;
- **Question:** What is the dependence structure between the extremely high residuals of each sector?
- We rescale them to unit Fréchet and select the multivariate residuals for which the sum exceeds the 0.97th quantile.



- We then apply the RJMCMC procedure and obtain a clustering of the angular variables.
- Most clusters show high residuals for one sector without the others.
- The largest clusters (each > 17%) concern healthcare, utilities and non-durable goods respectively.
- Roughly 13% of the observations belong to a cluster where several sectors tend to have relatively high residuals simultaneously.



8) Conclusion

- The angular density function is a full summary of the extremal dependence structure.
- The *tilted* Dirichlet mixture model is a valid angular density which is dense in the class of interest.
- The (tDM) model is easy to implement via a RJMCMC algorithm because its parameters are unconstrained.
→ Easy proposals for the parameters. → No fine-tuning of the initial values.
→ Higher dimensions are treated more quickly than previous models.
- This approach produces a soft clustering of the observations as an integral part of the MCMC output.
→ Modelling the heterogeneity in the extremal dependence structure ultimately allows us to analyse the clusters separately in lower dimensions.

References

- Boldi, M. O. and Davison, A. C. (2007) A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B* 69(2), 217–229.
- Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B* 53(2), 377–392.
- Sabourin, A. and Naveau, P. (2014) Bayesian dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis* 71, 542–567.