Statistical modelling of time series extremes

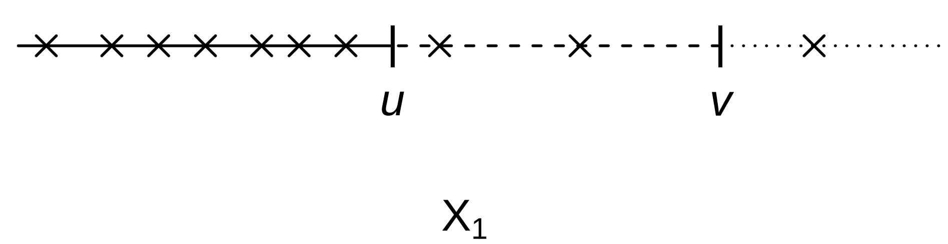
Ioannis Papastathopoulos, joint work with Graeme Auld

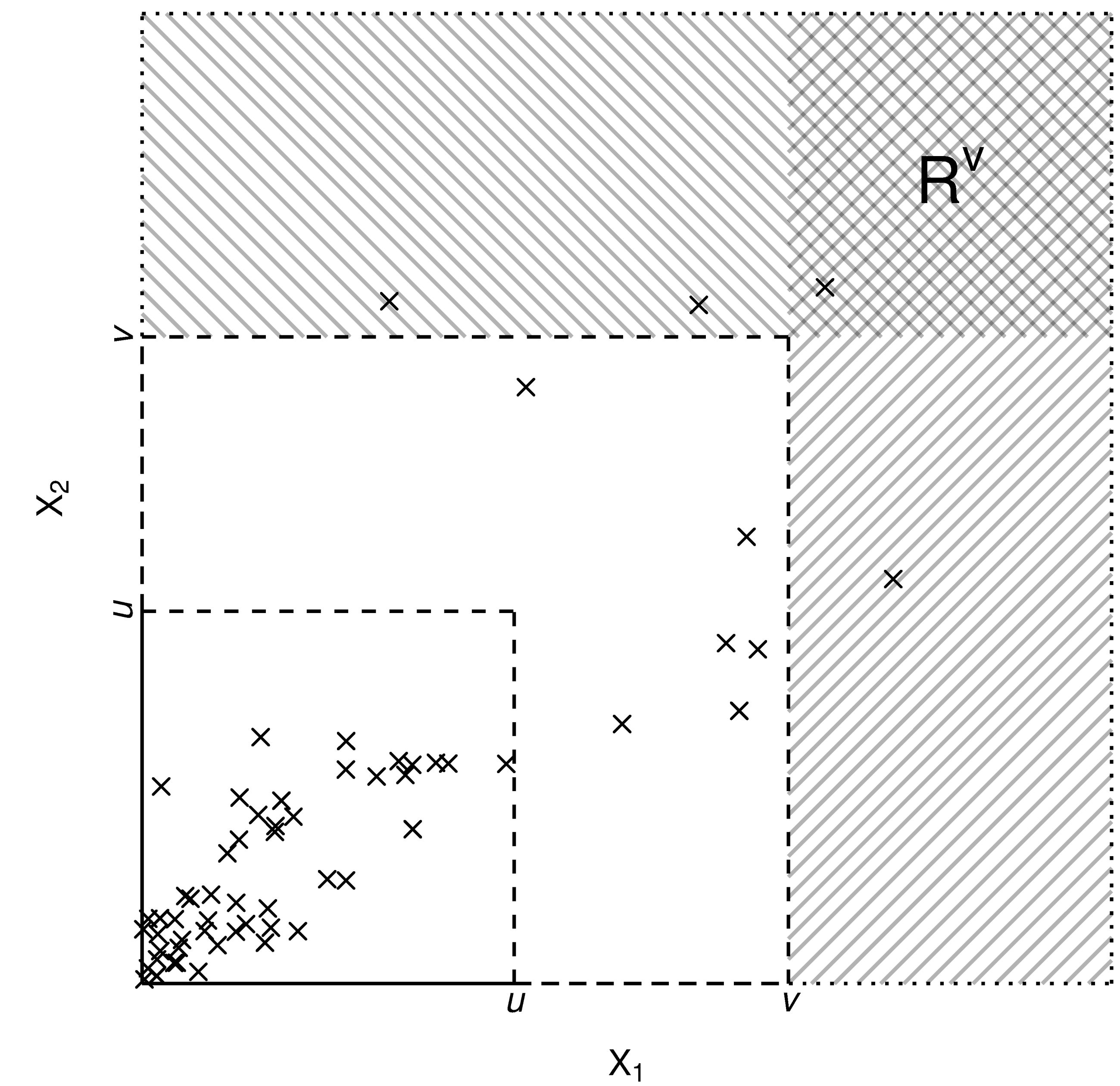
School of Mathematics

University of Edinburgh

Introduction

Threshold exceedances





Univariate exceedances

- Under mild conditions on the survival function of the random variable X

$$\Pr\left\{ \frac{X-u}{\sigma_u} > x \,\middle|\, X > u \right\} \longrightarrow \{1+\xi x\}^{-1/\xi}, \quad u \to x^*$$

i.e. the scaled excess random variable converges in distribution to the generalised Pareto distribution (Davison & Smith 1990)

- Interest often is in estimating either v_p given p, or p given v_p, where p and v_p satisfy

$$\mathbf{P}(X > v_p) = p,$$

Extrapolation

Extrapolation is dangerous. EVT provides a principled approach through asymptotically motivated *extrapolation factors*. For statistical purposes such asymptotic assumptions are taken to hold exactly over tail regions

$$\mathbf{P}\left(X > v_p\right) = \mathbf{P}\left(X > v_p \mid X > u\right)\mathbf{P}(X > u) \qquad v_p > u$$

For example, fixing p and inverting expression leads to a closed form expression for the $1/p$ return level

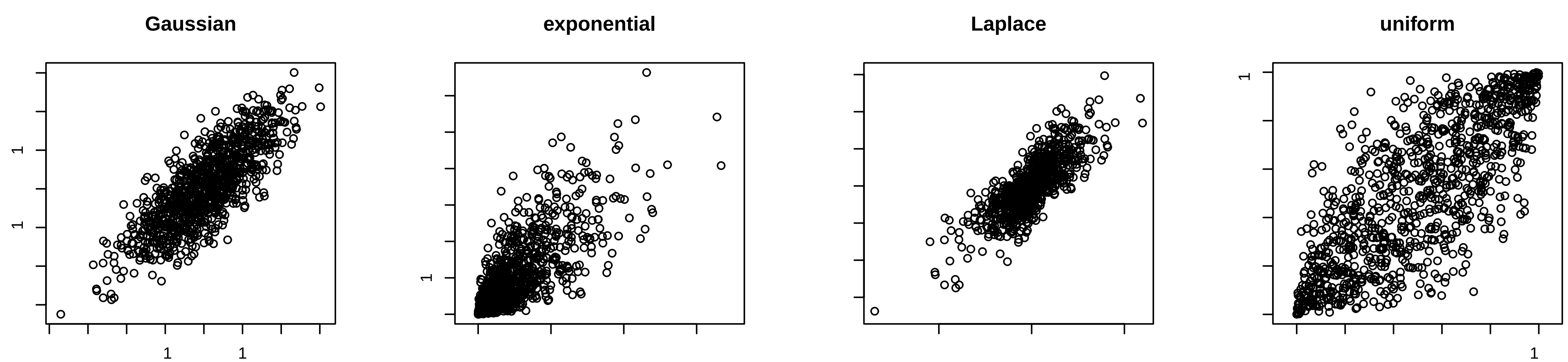
$$v_p = u + \frac{\sigma_u}{\xi}\left[\left(\frac{\pi}{p}\right)^{\xi} - 1\right],$$

where $\pi = \mathbf{P}(X > u)$.

Reduction to common margins

- In a multivariate setting, we need to characterize the dependence between random variables

- Standard practical approach is to bring marginal distributions to a common scale.

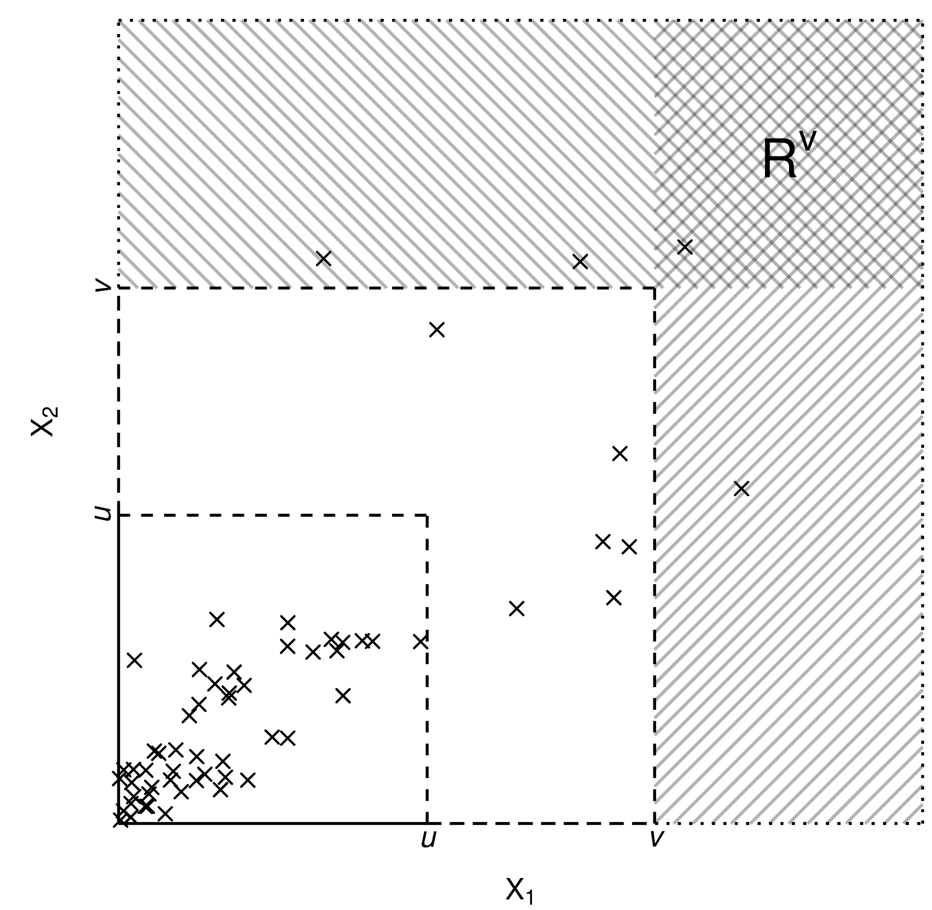


- We shall adopt this approach assuming that such standardizations are possible in practice

- this means that we can estimate the marginal distributions to a good degree

- Similarly here, we are interested in risk functionals. In particular, in estimating the probability of a subset of variables of $\boldsymbol{X}$ hitting an extreme set

$$\mathbb{P}(\boldsymbol{X} \in R^v) = \mathbb{P}(\boldsymbol{X} \in R^v \mid \boldsymbol{X} \in \mathcal{L}^u)\,\mathbb{P}(\boldsymbol{X} \in \mathcal{L}^u), \tag{1}$$



For example, $\boldsymbol{X} = (X_1, \ldots, X_d)$ could refer to a segment from a stationary time series of daily maximum temperatures. Then the risk functional would correspond to the probability of observing d days in a row where the daily max exceeds a high level v.

Asymptotic dependence

- Consider the measure of extremal dependence

$$\chi = \lim_{u \to \infty} \mathbb{P}(X_2 > u \mid X_1 > u)$$

  - $\chi > 0$: random variables are called asymptotically dependent
  - $\chi = 0$: random variables are called asymptotically independent

- Interpretation

$$\Lambda(u) = \frac{\mathbb{P}(\max\{X_1, X_2\} > u)}{\mathbb{P}(X_1 > u)} \to (2 - \chi) \qquad \text{as } u \to \infty.$$

Extremal dependence in time series

- When analysing the extremal behaviour of a **stationary time series** $\{X_t \,:\, t = 0, \pm 1, \pm 2\}$ one has to distinguish between two classes of extremal dependence.

- Assuming the time series has standard Laplace margins the lag t coefficient of asymptotic dependence is

$$\chi_t = \lim_{u \to \infty} \mathbf{P}(X_t > u \,|\, X_0 > u) \tag{2}$$

- If there exists a $t \neq 0$ such that $\chi_t > 0$ then the process is said to be *asymptotically dependent* and *asymptotically independent* otherwise.

Asymptotic dependence vs asymptotic dependence in time series

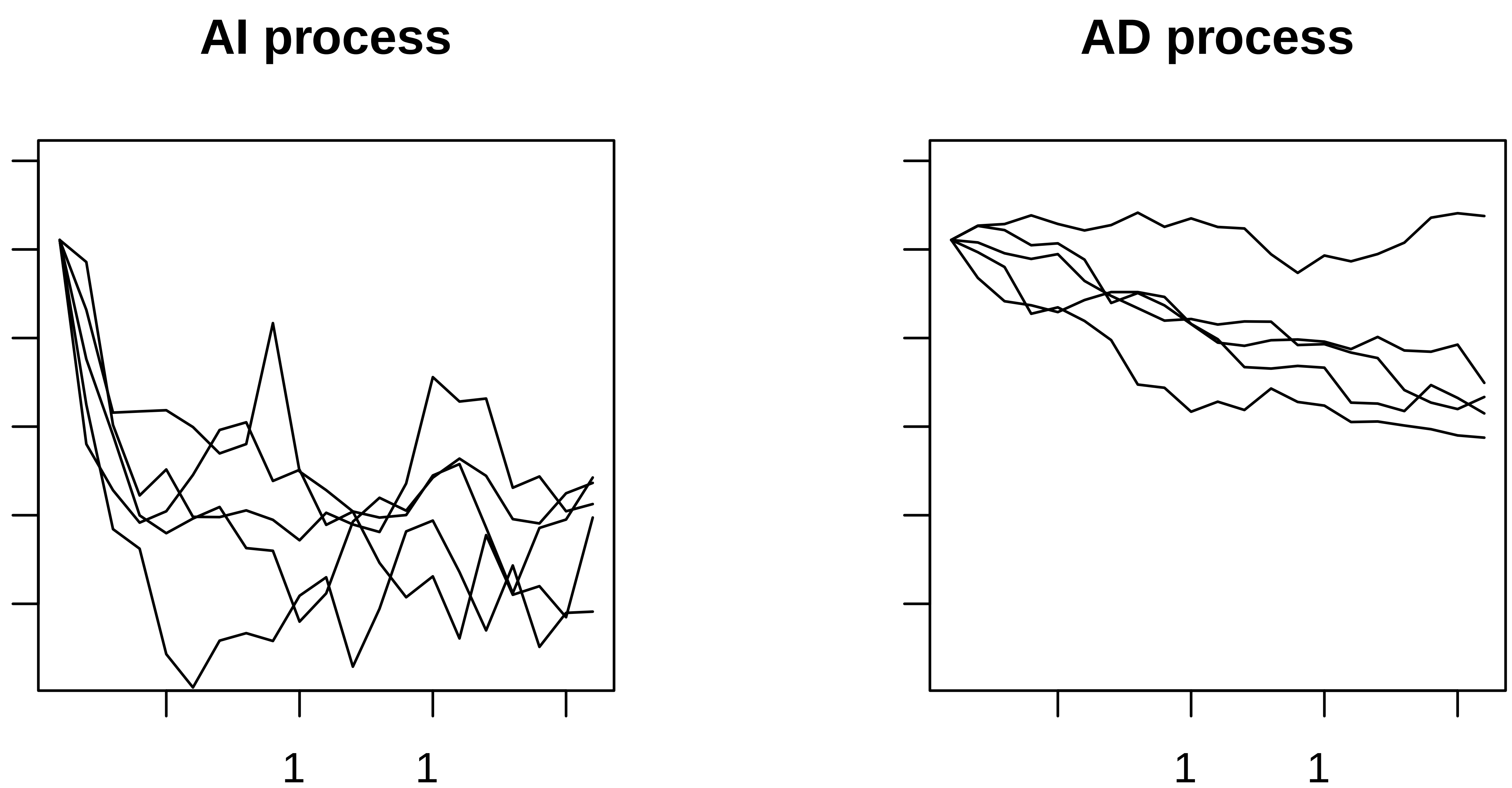


Figure 1: Simulated paths from a Gaussian autoregressive process (left) and a time series logistic process (right), conditioned on a high level relative to their marginal distribution.

Orleans daily maximum temperatures

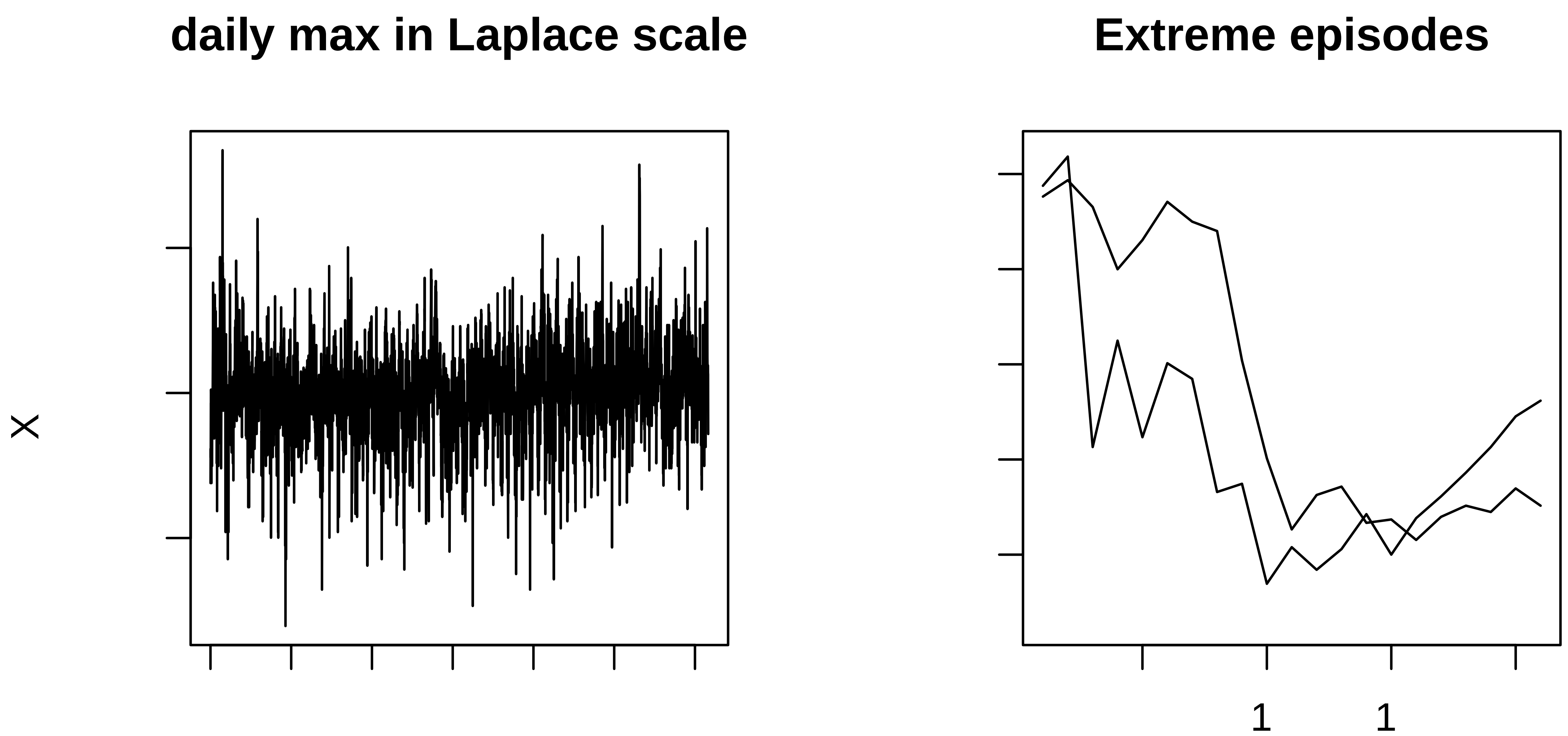


Figure 2: Left: daily maximum temperature measurements (shown in standard Laplace scale) from Orleans, France, taken during the years 1946-2012 (summer months). Right: Evolution of two events after witnessing a daily maximum temperature exceeding the 0.999 empirical quantile.

Examples of functionals of interest

$$e_1(v,d) = \mathbb{E}(\max \boldsymbol{X}_{1:d} \mid X_1 > v),$$

$$e_2(v,d) = \mathbb{E}(d^{-1} \sum_{i=1}^{d} X_i \mid X_1 > v)$$

$$e_3(v,d) = \mathbb{E}\left(\sum_{i=1}^{d} \mathbb{1}[X_i > v] \,\middle|\, X_1 > v \right)$$

$$p(v,d,r) = \mathbb{P}\left(\sum_{i=1}^{d} \mathbb{1}[X_i > v] = r \,\middle|\, X_1 > v \right),$$

See also Winter & Tawn (2016*a*,*b*)

Conditional extremes of time series

Conditional extreme value limits

- $\{X_t\}_{t\in\mathcal{T}}$ stationary process with exponentially tailed margins:
  - $\lim_{x\to\infty}\Pr(X_t > x)/e^{-x} = 1$
  - $\lim_{x\to-\infty}\Pr(X_t \le x)/e^{x} = 1$

- Assume there exist location and scale functions $a_i : \mathbb{R}\to\mathbb{R}$ and $b_i : \mathbb{R}\to\mathbb{R}_+$, such that for any $A\subset\mathcal{T}$ with $|A|<\infty$ (Heffernan & Tawn 2004)

$$\left(X_t - u, \frac{\boldsymbol{X}_A - \boldsymbol{a}_{A-t}(X_t)}{\boldsymbol{b}_{A-t}(X_t)}\right) \Bigm| X_t > u \xrightarrow{d} (E_t, \boldsymbol{Z}_A^t), \tag{3}$$

  where $E_t \perp\!\!\!\perp \boldsymbol{Z}_A^t$, $E_t\sim\exp(1)$ and $\boldsymbol{Z}_{A\setminus t}^t \sim G_{A\setminus t}$ where $G_{A\setminus t}$ has non-degenerate margins.

  - In this talk $\mathcal{T}$ is taken to be $\mathbb{Z}=\{0,\pm1,\pm2,\dots\}$. Other choices such as $\mathbb{R}$ work too with straightforward adaptations. $A-t$ stands for $\{i-t : i\in A\}$. $\boldsymbol{X}_A$ stands for $(X_i : i\in A)$.

  - The statistical method that is presented is inspired by Wadsworth & Tawn (2019) who treat the case $\mathcal{T}=\mathbb{R}^2$.

  - It is assumed that $a_0(x)=x$ and $b_0(x)=1$. This implies that $Z_t^t = 0$.

Statistical model

- Assume that $X_t \sim F_L$ where F_L denotes the standard Laplace distribution.
- Let $T_u = \{t \in \mathcal{T} : X_t > u\}$ denote the set of times where the process exceeds the level u.
- For statistical purposes, the limit relation (3) is taken to hold exactly above a sufficiently large level u. That is, we assume that for $t \in T_u$

$$X_t = u + E_t$$

$$\boldsymbol{X}_{t-k\,:\,t+k} = \boldsymbol{a}_{-k\,:\,k}(X_t) + \boldsymbol{b}_{-k\,:\,k}(X_t)\,\boldsymbol{Z}^{t}_{t-k\,:\,t+k},$$

Considerations for statistical modelling

- $k \in \mathbb{N}$ chosen sufficiently large so that extreme episodes are contained within $t-k \, : \, t+k$. Too large a k leads to increased computational burder. Context can often dictate choice of k.

- Multivariate distribution $G_{-k\,:\,k}$ admits no finite-dimensional parametric form. Require a flexible statistical model for $G_{-k\,:\,k}$ that facilitates inference and simulation (later).

- Asymptotic independence implies $a_{|i|}(x) \to 0$ and $b_{|i|}(x) \to 1$ as $i \to \infty$. So for large $|i|$ we have that $Z^t_{t+i} \sim F_L$.

- In addition to $G_{-k\,:\,k}$, there are $2 \times (2k-1)$ functions to infer ($\boldsymbol{a}_{-k\,:\,k}$ and $\boldsymbol{b}_{-k\,:\,k}$).

Statistical model for the residual vector

- Statistical model for probability density function of Z_{t+i}^t

$$f_i(z) = \frac{\delta_i}{2\sigma_i\Gamma(1/\delta_i)} \exp\left\{ - \left| \frac{z-\mu_i}{\sigma_i} \right|^{\delta_i} \right\}, \quad \mu_i \in \mathbb{R},\ \sigma_i > 0,\ \delta_i > 0. \tag{4}$$

  – $(\mu_i, \sigma_i, \delta_i) = (0, 1, 2)$: standard Gaussian

  – $(\mu_i, \sigma_i, \delta_i) = (0, 1, 1)$: standard Laplace

- In what follows we write B_k^t for the set $\{t-k \,:\, t+k\} \setminus \{t\}$.

- A versatile statistical model for $\boldsymbol{Z}_{B_k^t}^t$ is

$$\Phi^{-1}\left\{F_{B_k^t}\left(\boldsymbol{Z}_{B_k^t}^t\right)\right\} \sim \text{MVN}\left(\boldsymbol{0}, \boldsymbol{Q}^{-1}\right)$$

where $\boldsymbol{Q} = (\boldsymbol{\Sigma}^{-1})_{-(k+1),-(k+1)}$ with $\boldsymbol{\Sigma} \in \mathbb{R}^{2k+1,2k+1}$.

- Possibility for other modelling choices too, see Wadsworth & Tawn (2019).

- Wadsworth et al. (2016) show that, under the assumption of a joint density for $\boldsymbol{X}_A$, an equivalent formulation of the limit relation (3) for $A = B_k^t$ can be given

$$\lim_{u\to\infty} \Pr\left(\frac{\boldsymbol{X}_{B_k^t} - \boldsymbol{a}_{B_k^t}(u)}{\boldsymbol{b}_{B_k^t}(u)} \leq \boldsymbol{z} \,\middle|\, X_t = u \right) = G_{B_k^t}(\boldsymbol{z})$$

- Restrict this further by assuming that for any $x \in \mathbb{R}$

$$\lim_{u\to\infty} \Pr\left(\frac{\boldsymbol{X}_{B_k^t} - \boldsymbol{a}_{B_k^t}(u)}{\boldsymbol{b}_{B_k^t}(u)} \leq \boldsymbol{z} \,\middle|\, X_t = u + x \right) = G_{B_k^t}(\boldsymbol{z}; x)$$

  with $G_{B_k^t}(\boldsymbol{z}; x)$ a family of multivariate distributions with non-degenerate margins, satisfying $G_{B_k^t}(\boldsymbol{z}; 0) = G_{B_k^t}(\boldsymbol{z})$.

- Under the foregoing assumption, we have that (a_i, b_i) are necessarily extended regularly varying (Resnick & Zeber 2014, de Haan & Ferreira 2006), that is

$$\lim_{u\to\infty} \frac{a_i(u+x) - a_i(u)}{b_i(u)} = \psi_i^a(x) \quad \text{and} \quad \lim_{u\to\infty} \frac{b_i(u+x)}{b_i(u)} = \psi_i^b(x)$$

Models for norming functions

- In order to make statistical inference tractable, we are required to specify the forms of the norming functions $\boldsymbol{a}_A$ and $\boldsymbol{b}_A$.

- We consider two possibilities

$$\textbf{Model 1:} \quad a_i(x) = \alpha_i x, \quad b_i(x) = x^{\beta_i}, \tag{5}$$
$$\textbf{Model 2:} \quad a_i(x) = \alpha_i x, \quad b_i(x) = \{1 + a_i(x)\}^{\beta}, \tag{6}$$

where $\alpha_i \in [-1, 1]$, $\beta_i \in [0, 1]$, for $i = 1, \ldots, k$.

- Many results available for structured processes such as pth order Markov process (Papastathopoulos et al. 2017, Papastathopoulos & Tawn 2020)

Simplified forms for norming functions

Recurrence based approach. For example, for a pth order Markov process, $\boldsymbol{\alpha}_{1:p-1}$ are free parameters and subsequent values $\boldsymbol{\alpha}_{p:k}$ obtained via $\alpha_t = a(\boldsymbol{\alpha}_{t-p:t-1})$

- recurrence based on structure from autoregressive processes

$$\alpha_t = \theta_1\alpha_{t-1} + \theta_2\alpha_{t-2}, \quad 2 \le t \le k, \quad \text{with} \ \ \alpha_0 = 1, \alpha_1 = \theta_1/(1-\theta_2). \tag{7}$$

- recurrence based on structure from homogeneous update functions

$$\alpha_t = c\left[\sum_{i=1}^{p}\gamma_i(\gamma_i\alpha_{t-i})^{\delta}\right]^{1/\delta}, \quad d \le t \le k, \tag{8}$$

  with $c > 0$, $\delta \in \mathbb{R}$, $\boldsymbol{\gamma}_{1:p} \in S_{p-1} = \{\boldsymbol{\gamma}_{1:p} \in \mathbb{R}_+^p : \sum_{i=1}^p \gamma_i = 1\}$.

- More possibilities for modelling including general autocorrelation functions. Structure backwards in time obtained similarly.

- Remark: Laplace scale also allows modelling negative dependence and (predictable) jumps from lower to upper tail. Oscillating autocorrelation functions can be used too.

- If $\pi^t_{-k:k}(\boldsymbol{x})$ denotes the conditional density of $\boldsymbol{X}_{t-k:t+k}$ given $X_t = u$, then given an observed time series $\boldsymbol{x} = (x_1, \ldots, x_n)$

$$\log \pi^t_{-k:k}(\boldsymbol{x}) = \sum_{i=-k}^{k} \left\{ \log \delta_i - \log b_i(x_t) - \log \sigma_i - \log \Gamma\left(1/\delta_i\right) - \left| z^t_{t+i} \right|^{\delta_i} \right\}$$

$$+ 0.5 \log |\boldsymbol{Q}| - 0.5\, \boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w} + 0.5 \sum_{i=-k}^{k} \left[\Phi^{-1}\{F_{t+i}(x_{t+i})\} \right]^2 \tag{9}$$

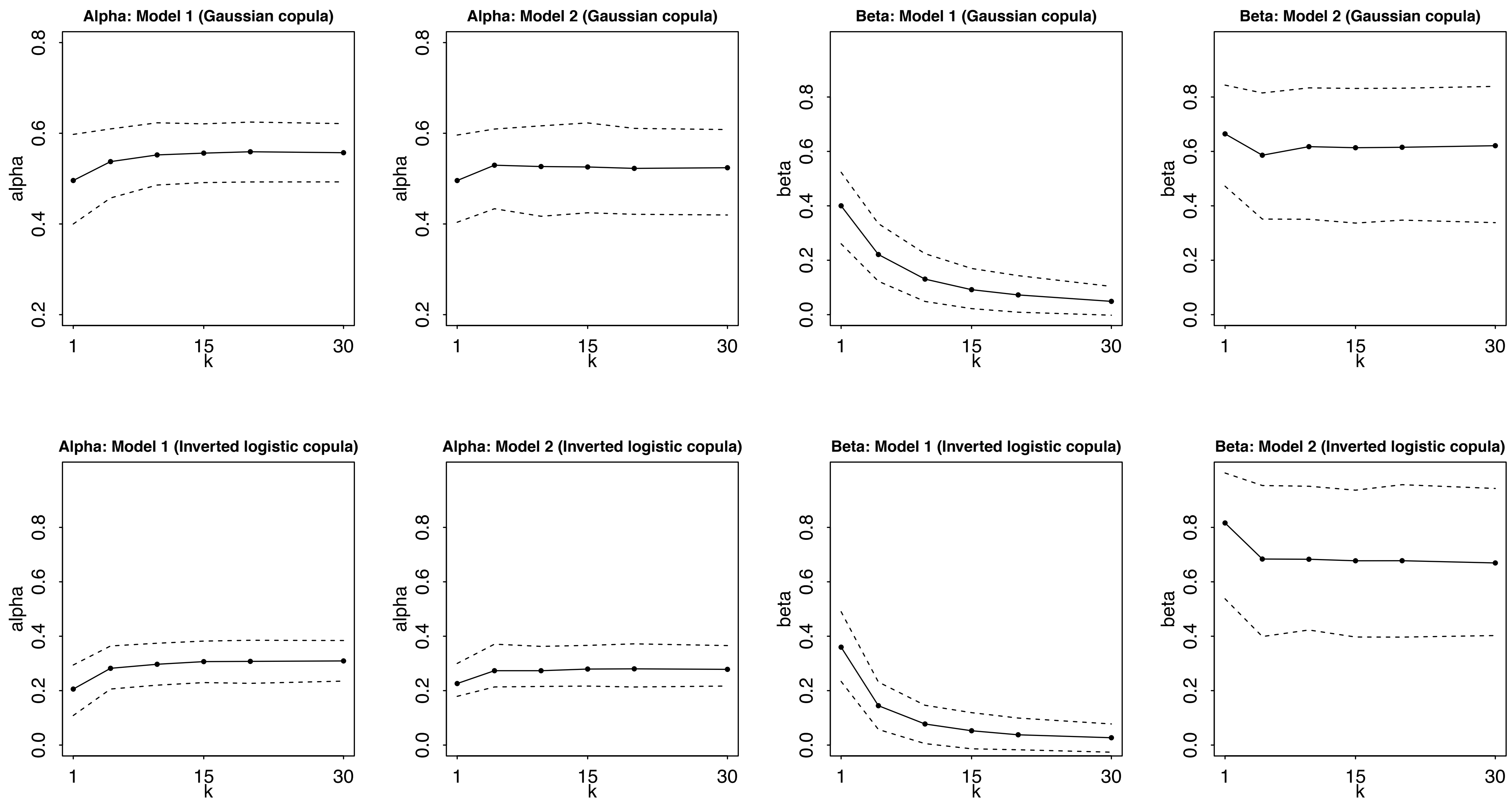
where $z^t_{t+i} = \{x_{t+i} - a_i(x_t) - b_i(x_t)\mu_i\}/\{b_i(x_t)\sigma_i\}$

- $\boldsymbol{w} = (\Phi^{-1}\{F^t_{t+1}(x_{t+1})\}, \ldots, \Phi^{-1}\{F^t_{t+k}(x_{t+k})\})$ with F^t_{t+i}

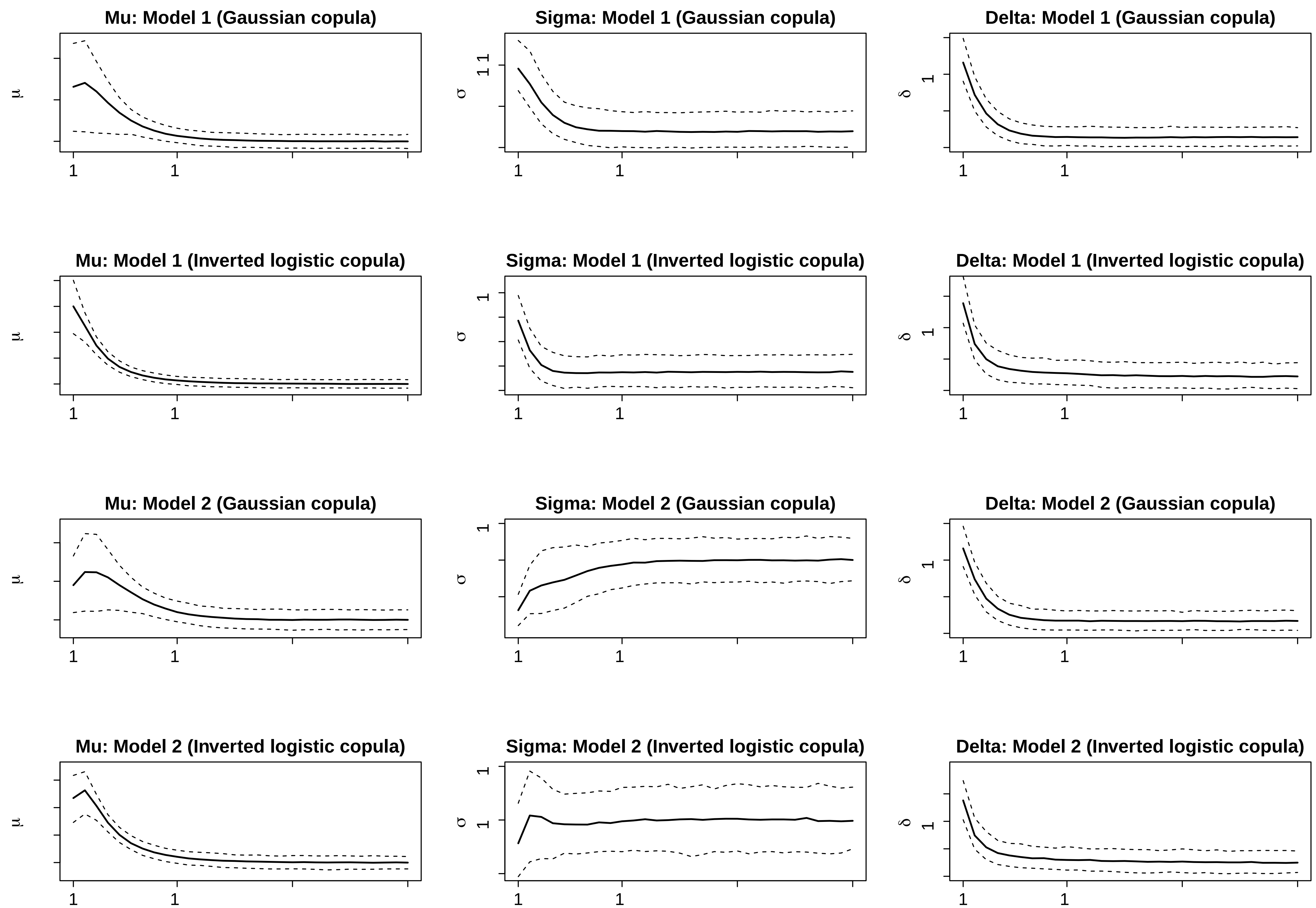
- the conditional distribution function of X_{t+i} given $X_t = u$. The composite likelihood is given by

$$\sum_{t \in T_u} \log \pi^t_{-k:k}(\boldsymbol{x})$$

Monte Carlo study



Monte Carlo study



Additional parameterizations

- Depends on the underlying data generating mechanism and exploratory analysis is useful here

- For the examples in the Monte Carlo study (and the Orleans dataset), possible parameterization under the Model 2 normings is

$$\begin{aligned} \mu_{i+1} &= Ae^{-Bi}, \\ \sigma_{i+1} &= (1 - Ce^{-Di}), \\ \delta_{i+1} &= 1 + Ee^{-Fi}, \end{aligned} \tag{10}$$

  for $A, B, C, D, E, F > 0$ and $i \geq 0$.

Rare event sampling

Forward sampling

- Interested in estimating conditionals expectations of the form

$$\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v) \qquad v > u, d \in \mathbb{N},$$

  for some function g.

- Repeatedly simulate forward from extreme event and estimate expectation using sample proportion

Forward sampling algorithm

input : Threshold $v > u$, $d, N \in \mathbb{N}$ and constants $(\hat{\boldsymbol{\alpha}}_{1:k}, \hat{\beta})$ from fitted conditional model.
output : An estimate of $\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v)$

for $i \leftarrow 1$ **to** N **do**
  simulate exceedance amount $E \sim \mathrm{Exp}(1)$;
  set $X_1^i = v + E$;
  simulate residual $\hat{\boldsymbol{Z}}_{2:d}^{(1)}$ *from fitted conditional model independently of* X_1^i ;
  set $\boldsymbol{X}_{2:d}^i = \hat{\boldsymbol{\alpha}}_{1:d-1} X_1^i + (X_1^i)^{\hat{\beta}} \hat{\boldsymbol{Z}}_{2:d}^{(1)}$;
  set $\boldsymbol{X}^i = (X_1, \boldsymbol{X}_{2:d}^i)$;
end
return $\hat{\mathbb{E}}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v) = N^{-1} \sum_{i=1}^{N} g(\boldsymbol{X}^i)$

Orleans dataset: stability plots for parameter estimates

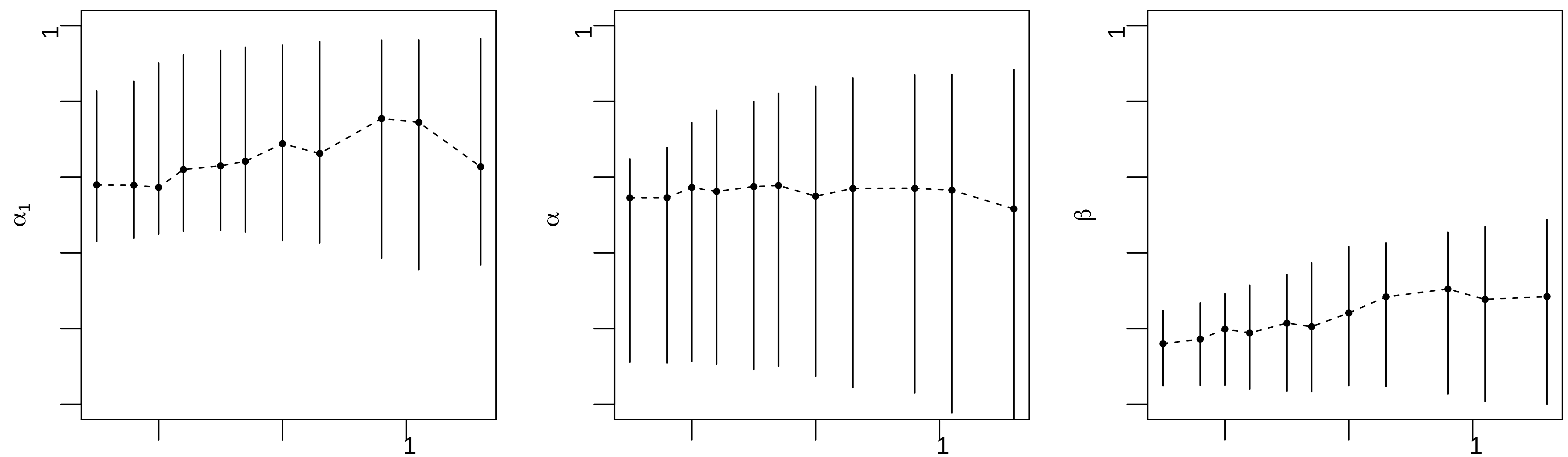
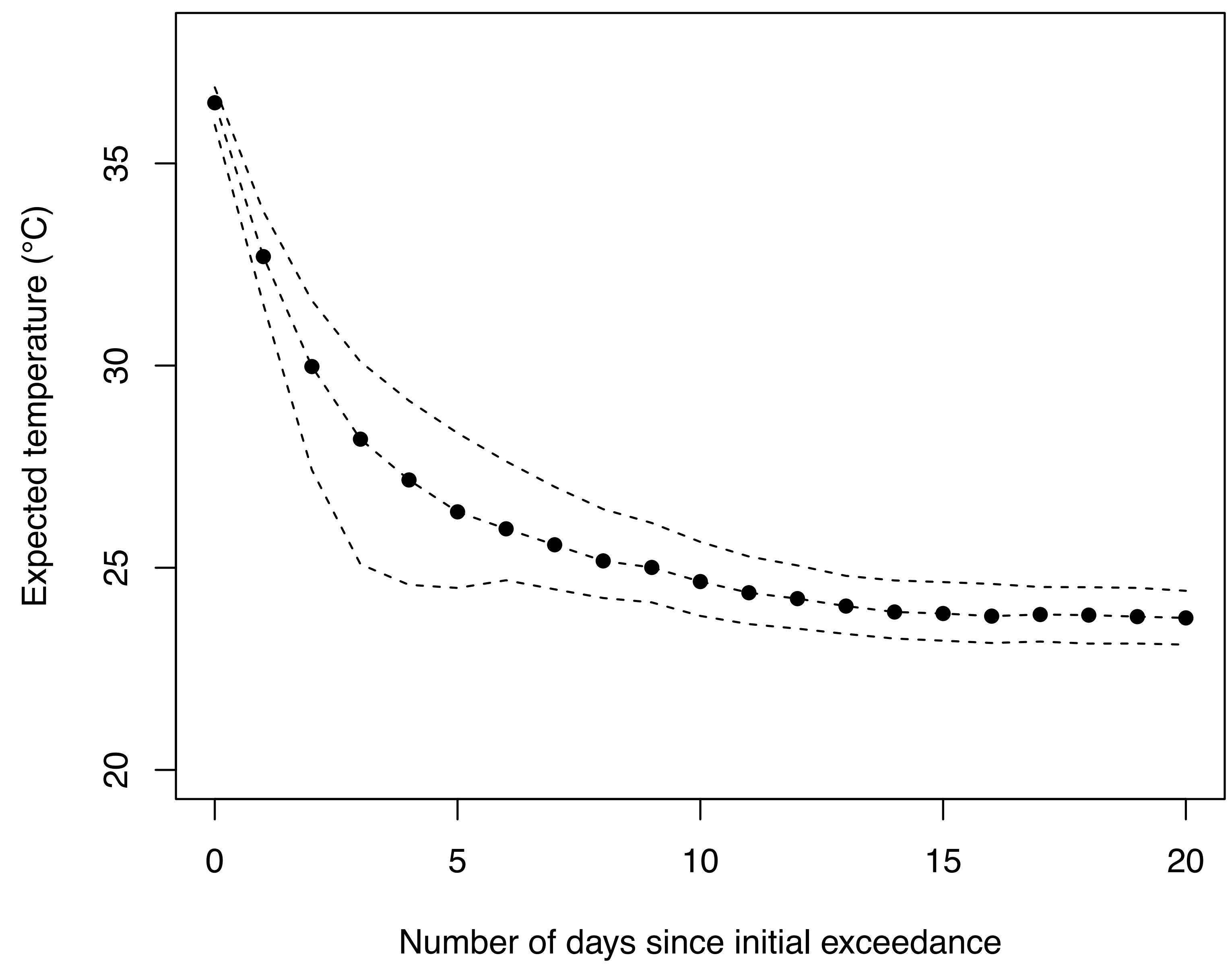


Figure 3: Plots showing parameter estimates for different thresholds used to identify exceedances.

Orleans dataset: expected evolution of extreme episode



Orleans dataset: summary statistics

- a period of three consecutive days where the mean daily maximum temperature exceeds 35°C may lead to excess mortality in Orleans between 17% and 47%.
- We estimated the probability of this event occurring during a three week period conditional on 35°C being exceeded at the start of the period. The probability is estimated as 0.309 with 95% bootstrap confidence interval (0.114, 0.420).
- Other simple summary statistics can be calculated similarly with forward sampling. For example, $\mathbb{E}\{\max(\boldsymbol{Y}_{1:21}|Y_1 > 35)\}$ is estimated as 36.76°C (36.07, 37.35)

Estimating the union of rare events

- Interested in estimating probabilities of the form

$$p = \mathbb{P}(\cup_{i=1}^d \{X_i > v_i\}) = \mathbb{E}_\pi \left[\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})\right] \qquad v_i \in (0, \infty)$$

- where $\mathcal{L} = \cup_{i=1}^d L_i$ with $L_i = \{\boldsymbol{x} \in \mathbb{R}^d : x_i > v_i\}$.

- The obvious empirical estimator

$$N^{-1} \sum_{i=1}^N \mathbb{1}_{\mathcal{L}}(\boldsymbol{X}^i),$$

of p based on N independent replications, $\{\boldsymbol{X}^i\}_{i=1}^N$, of $\boldsymbol{X}_{1:d}$, is unbiased and has variance $p(1-p)/N$.

- For each $i \in 1:d$, define π_i^* to be the conditional density of $\boldsymbol{X}_{1:d}$ given $X_i > v_i$, so that

$$\pi_i^*(\boldsymbol{x}) = \pi(\boldsymbol{x}) \mathbb{1}_{L_i}(\boldsymbol{x})/p_i, \qquad \boldsymbol{x} \in \mathbb{R}^d,$$

where $p_i = \mathbb{P}(X_i > v_i)$

Importance sampling density

- Importance sampling density (Karp & Luby 1983, Heffernan & Tawn 2004, Adler et al. 2012, Owen et al. 2019, Wadsworth & Tawn 2019)

$$\pi^* = \sum_{i=1}^{d} w_i \pi_i^*$$

where $w_i = p_i/\bar{p}$ and $\bar{p} = \sum_{i=1}^{d} p_i$ is the union bound of p.

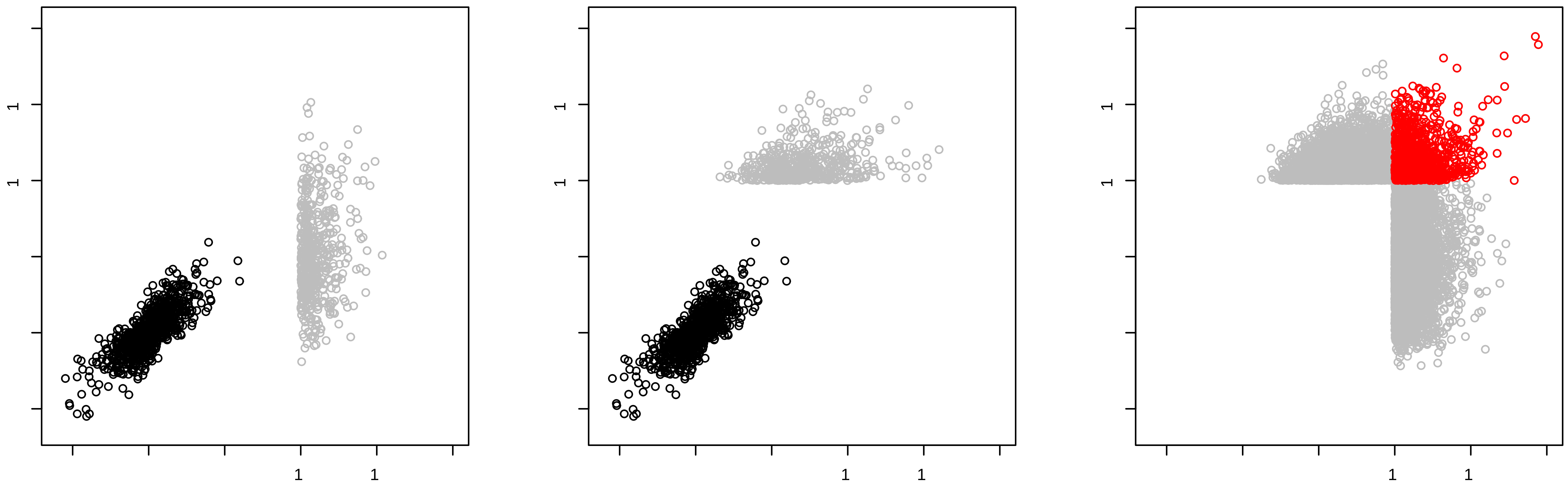
- Thus the mixture component π_i^* is sampled from with probability proportional to p_i. Since

$$p = \mathbb{E}_{\pi}\{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})\} = \mathbb{E}_{\pi^*}\left\{\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})\pi(\boldsymbol{X}_{1:d})}{\pi^*(\boldsymbol{X}_{1:d})}\right\}, \tag{11}$$

this motivates the following estimator of p

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}^i)\pi(\boldsymbol{X}^i)}{\pi^*(\boldsymbol{X}^i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}^i)\pi(\boldsymbol{X}^i)}{\sum_{j=1}^{d}\mathbb{1}_{L_j}(\boldsymbol{X}^i)\pi(\boldsymbol{X}^i)\bar{p}^{-1}}, \quad \boldsymbol{X}^i \overset{iid}{\sim} \pi^*. \tag{12}$$

Sampling importance resampling



- As $\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}^i) = 1$ when $\boldsymbol{X}^i \sim \pi^*$, estimator (12) simplifies to

$$\hat{p} = \frac{\bar{p}}{n} \sum_{i=1}^{n} \frac{1}{S(\boldsymbol{X}^i)}, \quad \boldsymbol{X}^i \overset{iid}{\sim} \pi^*, \tag{13}$$

  where $S(\boldsymbol{X}^i) = \sum_{j=1}^{d} \mathbb{1}_{L_j}(\boldsymbol{X}^i)$ counts the number of events $\{X_i > v_i\}, 1 < i < d$, that occur in the block $\boldsymbol{X}^i$ of length d.

- As $1 \leq S(\boldsymbol{X}^i) \leq d$ when $\boldsymbol{X}^i \sim \pi^*$, $\hat{p}$ is always well defined and respects the theoretical bounds $\bar{p}/d \leq p \leq \bar{p}$.

- The union bound $\bar{p}$ which appears in expression (13) is easily obtained using the assumption that the margins of the process are standard Laplace distributed.

- Owen et al. (2019) prove that $\hat{p}$ is an unbiased estimator of p and $\mathrm{Var}(\hat{p}) \leq p(\bar{p} - p)/n$, from which it follows that $\hat{p}$ is a consistent estimator of p.

- Although $\hat{p}$ may be used to estimate the probability of an arbitrary union of events, it is in the rare event setting, when p and $\bar{p}$ are small, that it is most efficient since then $p(\bar{p} - p)$ may be orders of magnitude smaller than $p(1 - p)$.

- Thus, in the rare event setting we increase the precision in estimation using importance sampling from π^* rather than π.

References

Adler, R. J., Blanchet, J. H. & Liu, J. (2012), 'Efficient Monte Carlo for high excursions of Gaussian random fields', *Ann. Appl. Probab.* **22**(3), 1167–1214.

Davison, A. C. & Smith, R. L. (1990), 'Models for exceedances over high thresholds', *J. Roy. Statist. Soc., B* **52**, 393–442.

de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Vol. 21, Springer, New York.

Heffernan, J. E. & Tawn, J. A. (2004), 'A conditional approach for multivariate extreme values (with discussion)', *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66**, 1–34.

Karp, R. M. & Luby, M. (1983), Monte-Carlo algorithms for enumeration and reliability problems, *in* '24th Annual Symposium on Foundations of Computer Science', IEEE, pp. 56–64.

Owen, A. B., Maximov, Y. & Chertkov, M. (2019), 'Importance sampling the union of rare events with an application to power systems analysis.', *Electron. J. Statist.* **13**, 231–254.

Papastathopoulos, I., Strokorb, K., Tawn, J. A. & Butler, A. (2017), 'Extreme events of Markov chains', *Advances in Applied Probability* **49**.

Papastathopoulos, I. & Tawn (2020), 'Hidden tail chains and recurrence equations for dependence parameters associated with extremes of higher-order Markov chains', *arXiv preprint arXiv:1903.04059* .

Resnick, S. I. & Zeber, D. (2014), 'Transition kernels and the conditional extreme value model', *Extremes* **17**, 263–287.

Wadsworth, J. L. & Tawn, J. (2019), 'Higher-dimensional spatial extremes via single-site conditioning', *arXiv:1912.06560* .

Wadsworth, J. L., Tawn, J. A., Davison, A. C. & Elton, D. M. (2016), 'Modelling across extremal dependence classes', *J. Roy. Statist. Soc. Ser. B, Statist. Methodol.* . doi: `10.1111/rssb.12157`.

Winter, H. C. & Tawn, J. A. (2016*a*), 'kth-order markov extremal models for assessing heatwave risks', *Extremes* pp. 1–23.

Winter, H. C. & Tawn, J. A. (2016*b*), 'Modelling heatwaves in central France: a case-study in extremal dependence', *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65**, 345–365.