Extremes of Stationary Time Series – Day 2

Dr. Ioannis Papastathopoulos (with illustrations by Lambert de Monte)

February 28, 2023

Introduction

- So far have identified[1] that dependence in time series affects the behaviour of the extreme values, the latter typically occur in clusters.

- The extremal index θ can be seen as a function of the (unknown) limiting distribution of the cluster of renormalized exceedances [Hsing et al., 1988]

$$\pi_n(j) = \mathbb{P}\left[\sum_{i=1}^{p_n} I\left(\frac{X_i - b_n}{a_n} > x\right) = j \,\middle|\, \sum_{i=1}^{p_n} I\left(\frac{X_i - b_n}{a_n} > x\right) > 0\right],$$

where $p_n = o(n)$ and $x > b_l$.

- The extremal index satisfies

$$\theta^{-1} = \sum_{j=1}^{\infty} j \lim_{n\to\infty} \pi_n(j)$$

[1] Under suitable mixing conditions such as the AIM condition.

Asymptotic dependence

- Consider the measure of extremal dependence

$$\chi = \lim_{u\to\infty} \mathbb{P}(X_2 > u \mid X_1 > u) \in [0, 1]$$

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

- Interpretation: as u approaches ∞,

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

Hence, in common margins, the probability of the maximum of two variables exceeding an extreme level is a scale factor of the probability of one of the variables exceeding this level.

  - X_1 and X_2 independent implies $\chi = 0$
  - X_1 and X_2 perfectly dependent implies $\chi = 1$
  - Reverse implications not true in general.

Another view of the extremal dependence in stationary time series

- We distinguish between two classes of extremal dependence
- The lag-t coefficient of asymptotic dependence is defined by

$$\chi_t = \lim_{u \to \infty} \mathbb{P}(X_t > u \,|\, X_0 > u) \tag{1}$$

- 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.
- Asymptotically dependent time series have $\theta \in [0, 1)$
- Asymptotically independent time series have $\theta = 1$

Another view of the extremal dependence in stationary time series

- We distinguish between two classes of extremal dependence
- The lag-t coefficient of asymptotic dependence is defined by

$$\chi_t = \lim_{u \to \infty} \mathbb{P}(X_t > u \mid X_0 > u) \tag{1}$$

- 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.
- Asymptotically dependent time series have $\theta \in [0, 1)$
- Asymptotically independent time series have $\theta = 1$

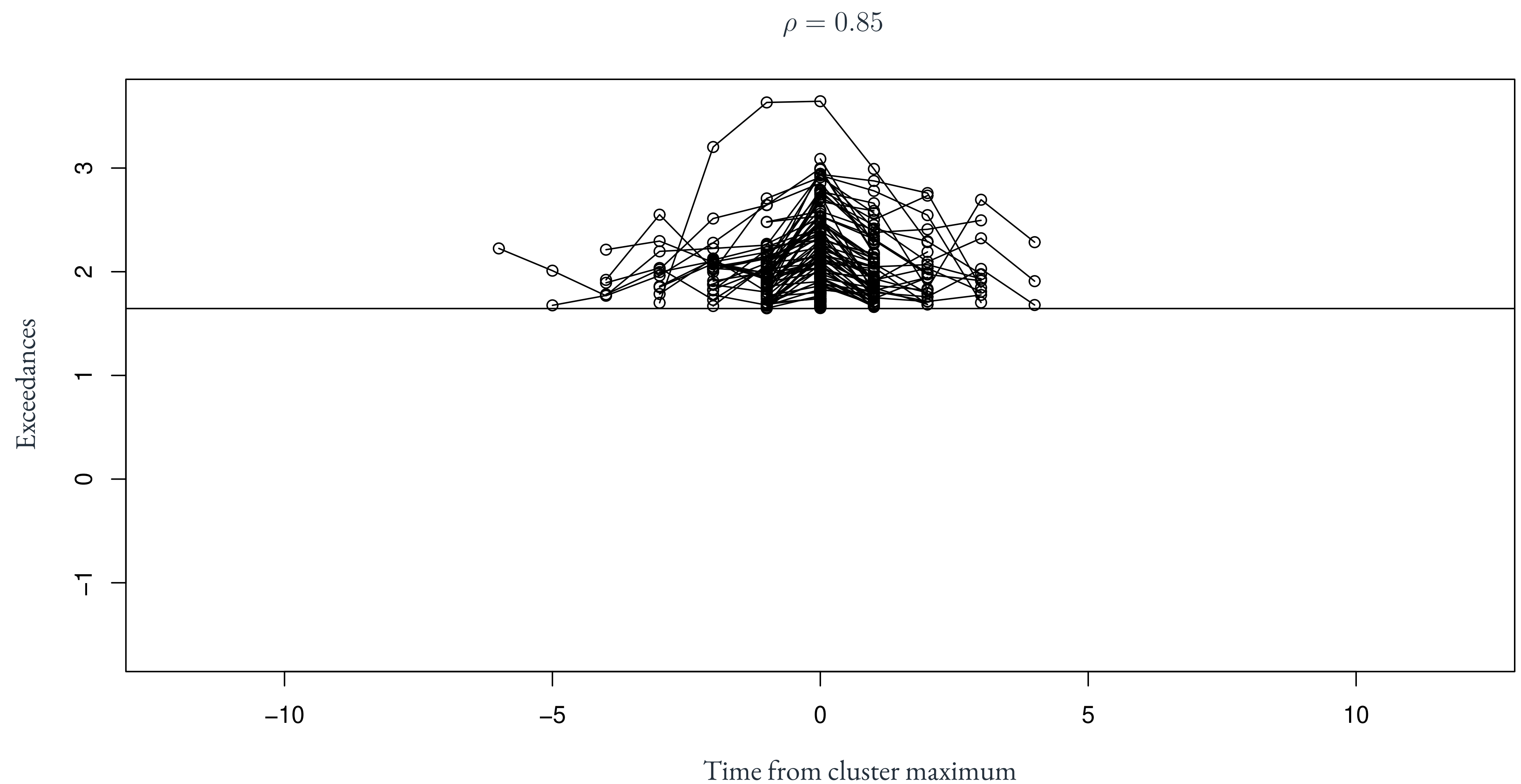
Let $X_t = \max(Y_{t-1}, Y_t)$ where Y_t IID with $F_Y(y) = \exp[-1/(2y)]$. Then

$$
\begin{aligned}
\mathbb{P}(X_2 > u \mid X_1 > u) &= \frac{\mathbb{P}(X_1 > u, X_2 > u)}{\mathbb{P}(X_1 > u)} \\
&= \frac{1 - \mathbb{P}(X_1 \le u) - \mathbb{P}(X_2 \le u) + \mathbb{P}(X_1 \le u, X_2 \le u)}{\mathbb{P}(X_1 > u)} \\
&= \frac{1 - 2\,e^{-1/u} + e^{-3/(2u)}}{1 - e^{-1/u}} \approx \frac{1 - 2(1 - \frac{1}{u}) + 1 - \frac{3}{2u}}{\frac{1}{u}} \quad \text{as } u \text{ approaches } \infty \\
&\to \frac{1}{2}
\end{aligned}
$$

Recall $\mathbb{P}(A \cap B) = 1 - \mathbb{P}((A \cap B)^c) = 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c) + \mathbb{P}(A^c \cap B^c)$

Lag-t χ for stationary Gaussian autoregressive process

- Let $X_t = \rho X_{t-1} + (1-\rho^2)^{1/2}\varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}(0,1)$ with $\rho \in (-1,1)$.
- Here $\chi_t = 0$ for all $t > 0$.
- Similarly, for this process we have $\theta = 1$, *i.e.*, extremes occur singly in the limit. Maximum is asymptotically the same as the maximum of IID variables with the same marginal distribution



Properties of extremal dependence coefficients

- Recall definition of the extremal index θ

$$\theta = \lim_{n\to\infty} \mathbb{P}(M_{2,p_n} \leq n | X_1 > n)$$

$$= \lim_{u\to\infty} \mathbb{P}(X_1 \leq u, \ldots, X_{p_u} \leq u | X_1 > u)$$

where $p_n = o(n)$, for variables with Fréchet marginals.

- The extremal index and the coefficient of asymptotic dependence are invariant to changes in the choice of the marginal distribution.

- Suppose $\{X_t\}$ is a stationary process with unit Fréchet margins satisfying the AIM(n) condition.

- Let $G(x) = 1 - \exp(-x)$, $x > 0$. Then the extremal index of the time series $\{X_t\}$ is

$$\theta = \lim_{n\to\infty} \mathbb{P}\left[X_i \le u_n : i = 2, \ldots, p_n \mid X_1 > u_n\right]$$

$$= \lim_{n\to\infty} \mathbb{P}\left[G^{\leftarrow} \circ F(X_i) \le G^{\leftarrow} \circ F(u_n) : i = 2, \ldots, p_n \mid G^{\leftarrow} \circ F(X_1) > G^{\leftarrow} \circ F(u_n)\right]$$

$$= \lim_{n\to\infty} \mathbb{P}\left[Y_i \le u_n^G : i = 2, \ldots, p_n \mid Y_1 > u_n^G\right] = \theta^G \qquad (u_n^G \sim \log n).$$

i.e., equal to the extremal index of the process $\{Y_t\}$, where Y_t follows a unit-exponential for all t.

- More generally, a monotone increasing marginal transformation does not change the value of the extremal index;

Some limitations

- Limited information when $\theta = 1$ or when $\chi_t = 0$ for all $t > 0$.

- Mean cluster-size is informative, but in applications we may be interested in other characteristics of the cluster distribution, *e.g.*,

$$T(u, d) = \mathbb{E}(g(\boldsymbol{Y}_{1:d}) \mid Y_1 > u),$$

or more generally, in the cluster-size distribution itself.

- what is the expected size of the largest element in a short period of time after seeing an exceedance:

$$T_1(u,d) = \mathbb{E}\left(\max \boldsymbol{Y}_{1:d} \,\middle|\, Y_1 > u\right),$$

- what is the expected energy content in a short period of time after seeing an exceedance:

$$T_2(u,d) = \mathbb{E}\left[\left(\frac{1}{d}\sum_{i=1}^{d} Y_i^2\right)^{1/2} \,\middle|\, Y_1 > u\right],$$

- what is the subasymptotic cluster-size distribution:

$$T_3(u,d) = \mathbb{P}\left(\sum_{i=1}^{d} \mathbf{1}[Y_i > u] = r \,\middle|\, Y_1 > u\right),$$

- and finally, how about standard measures of extremal dependence, e.g.,

$$\theta(u,d) = \mathbb{P}\big(Y_2 \le u, \ldots, Y_d \le u \mid Y_1 > u\big) \quad \text{and} \quad \chi(u,d) = \mathbb{P}(Y_{t+1} > u \mid Y_1 > u)?$$

Conditional extreme value theory

- Under mild conditions on the survival function of a random variable X, there exists $\sigma(u) > 0$

$$\mathbb{P}\left\{ \frac{X-u}{\sigma(u)} > x \,\middle|\, X > u \right\} \longrightarrow \{1+\xi x\}^{-1/\xi}, \quad u \to x^*,$$

i.e., the scaled excess random variable converges in distribution to the generalised Pareto distribution [Pickands, 1975, Davison and Smith, 1990]

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

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

- Typically, in practice v_p would correspond to a quantile of the distribution above which data have not been observed.

EVT and extrapolation

- **BEWARE:** Extrapolation is dangerous, but EVT provides a principled approach based on asymptotic theory for the behaviour of extrema.

- Fundamental premise in all statistical extreme value modelling is that we can approximate the distribution of extreme values by the limiting theoretical forms. For $v_p > u$

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

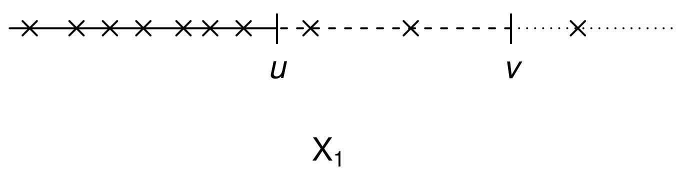
$$= \mathbb{P}\left(\frac{X-u}{\sigma(u)} > \frac{v_p - u}{\sigma(u)} \;\middle|\; X > u\right)\mathbb{P}(X > u)$$

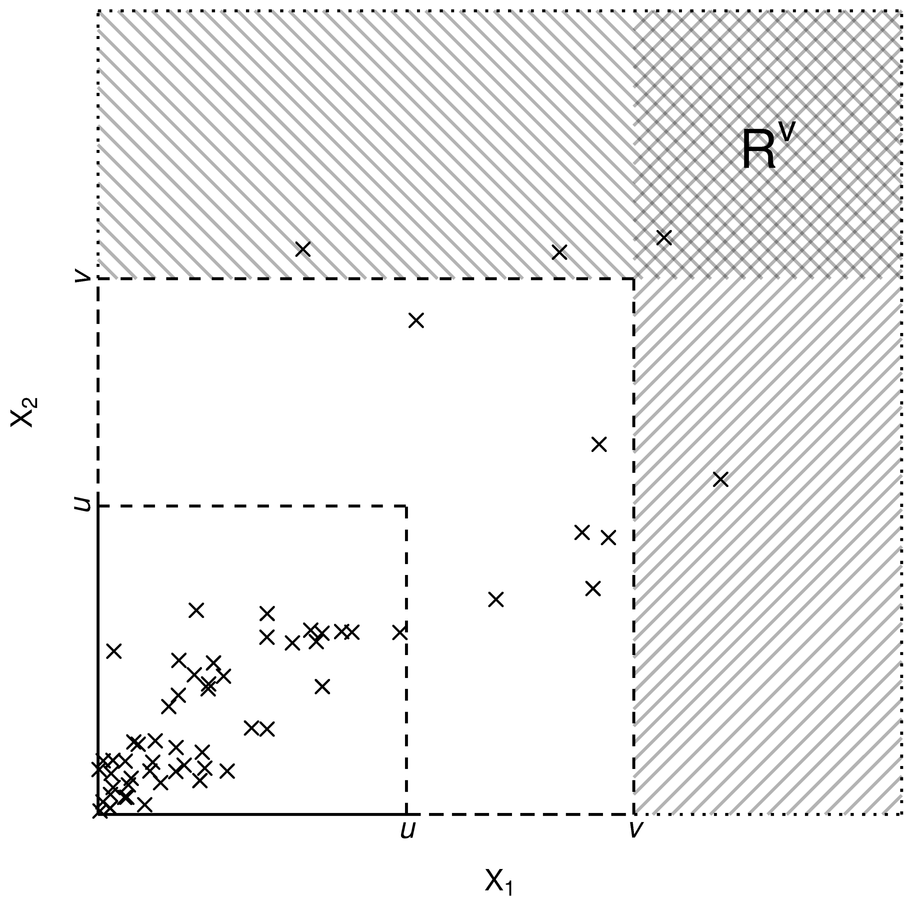
Inverting leads to a closed-form expression for the $(1/p)$-return level

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

where $\pi_u = \mathbb{P}(X > u)$.

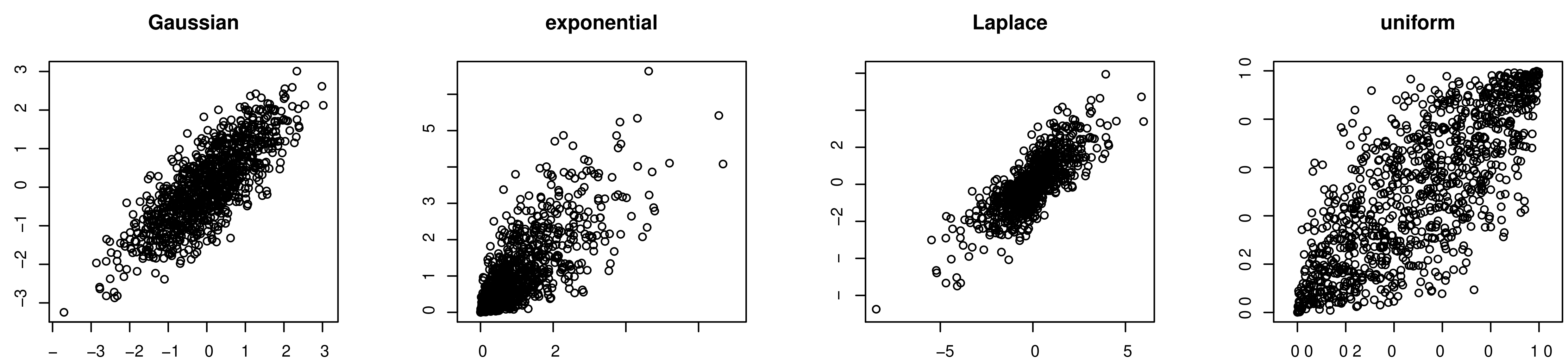
- In 2 dimensions, the notion of ordering is not the same as that in $\mathbb{R}$ due to the lack of total order.
- This brings complications in terms of defining what extreme events are.
- For example, an observation in $\mathbb{R}^2$ may be regarded as extreme if either:
  - both components exceed high thresholds;
  - at least one component exceeds a high threshold.





Reduction to common margins

- 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

Random vector (X_1, X_2) in $\mathbb{R}^2$ with standard Laplace margins.

- $a_{2|1} : \mathbb{R} \to \mathbb{R}$
- $b_{2|1} : \mathbb{R} \to \mathbb{R}_+$
- a non-degenerate distribution $G_{2|1}$ on $\mathbb{R}$

such that, as u approaches ∞,

$$\mathbb{P}\left(X_1 - u > x, \frac{X_2 - a_{2|1}(X_1)}{b_{2|1}(X_1)} < z \,\middle|\, X_1 > u\right) \to \exp(-x)G_{2|1}(z)$$

at continuity points $(x, z) \in \mathbb{R}_+ \times \mathbb{R}$ of the limit distribution.

When the joint density exists, limit formulation

$$\lim_{u\to\infty} \mathbb{P}\left(X_1 - u > x, \frac{X_2 - a_{2|1}(X_1)}{b_{2|1}(X_1)} < z \,\middle|\, X_1 > u\right) = \exp(-x) G_{2|1}(z)$$

is equivalent to

$$\begin{aligned}\lim_{u\to\infty} \frac{\frac{\partial}{\partial u}\mathbb{P}\left(X_1 > u, \frac{X_2 - a_{2|1}(X_1)}{b_{2|1}(X_1)} < z\right)}{\frac{\partial}{\partial u}\Pr(X_1 > u)} &= \lim_{u\to\infty} \boxed{\mathbb{P}\left(\frac{X_2 - a_{2|1}(X_1)}{b_{2|1}(X_1)} < z \,\middle|\, X_1 = u\right)} \\ &= G_{2|1}(z)\end{aligned}$$

In what follows, we shall use either $\{X_1 > u\}$ or $\{X_1 = u\}$, assuming

- Suppose (X, Y) follows a bivariate logistic max-stable distribution with standard Laplace margins and dependence parameter $\kappa \in (0, 1]$, that is,

$$\mathbb{P}(X \leq x, Y \leq y) = \exp\{-V\left(T(x), T(y)\right)\},$$

where $T(x) = -1/\log\left(F_L(x)\right)$,

$$V(x, y) = \left(x^{-1/\kappa} + y^{-1/\kappa}\right)^{\kappa} \qquad \text{and} \qquad V_1(x, y) = \frac{\partial}{\partial x} V(x, y).$$

- The conditional distribution for this example is

$$\mathbb{P}(Y \leq y \mid X = x) = -T(x)^2\, e^{1/T(x)}\, V_1\left\{T(x), T(y)\right\} \exp\left[-V\left\{T(x), T(y)\right\}\right].$$

- Conditioned limit law: for $z \in \mathbb{R}$,

$$\boxed{\mathbb{P}(Y - X \leq z \mid X = u) \to [1 + \exp(-z/\kappa)]^{\kappa - 1}, \qquad u \to \infty.}$$

Example: bivariate normal distribution with Laplace margins

- Suppose (X, Y) follows a bivariate normal distribution with standard Laplace margins and correlation parameter $\rho \in (-1, 1)$, that is,

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{\Phi^{-1}(F_L(x))} \int_{-\infty}^{\Phi^{-1}(F_L(y))} \frac{\mathrm{d}u\,\mathrm{d}v}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2(1-\rho^2)} u^2 + v^2 - 2\rho uv\right)$$

where Φ denotes the cdf of the standard normal distribution and Φ^{-1} is its inverse.

- The conditional distribution for this example is

$$\mathbb{P}(Y \leq y \mid X = x) = \Phi\left(\frac{\Phi^{-1}(F_L(y)) - \rho\,\Phi^{-1}(F_L(x))}{(1-\rho^2)^{1/2}}\right).$$

- Conditioned limit law: for $z \in \mathbb{R}$,

$$\boxed{\mathbb{P}\left(\frac{Y - \mathrm{sign}(\rho)\rho^2 X}{X^{1/2}} \leq z \mid X = u\right) \to \Phi\left(\frac{z}{\sqrt{2\rho^2(1-\rho^2)}}\right), \qquad u \to \infty.}$$

Example: inverted max-stable distribution with Laplace margins

- Suppose (X, Y) follows a bivariate inverted logistic max-stable distribution with unit exponential margins and dependence parameter $\kappa \in (0, 1]$, that is,

$$\mathbb{P}(X > x, Y > y) = \exp\{-V(1/x, 1/y\},$$

where

$$V(x, y) = \left(x^{-1/\kappa} + y^{-1/\kappa}\right)^{\kappa}$$

- The conditional distribution for this example is

$$\mathbb{P}(Y \leq y \mid X = x) = 1 + V_1(1, x/y)\exp(x - xV(1, x/y))$$

- Conditioned limit law: for $z \in \mathbb{R}_+$,

$$\boxed{\mathbb{P}\left(\frac{Y}{X^{1-\kappa}} \leq z \mid X = u\right) \to 1 - \exp(-\kappa z^{1/\kappa}) \qquad u \to \infty.}$$

Statistical model

The distribution of extreme values is approximated by their limiting theoretical form, for $X_1 > u$,

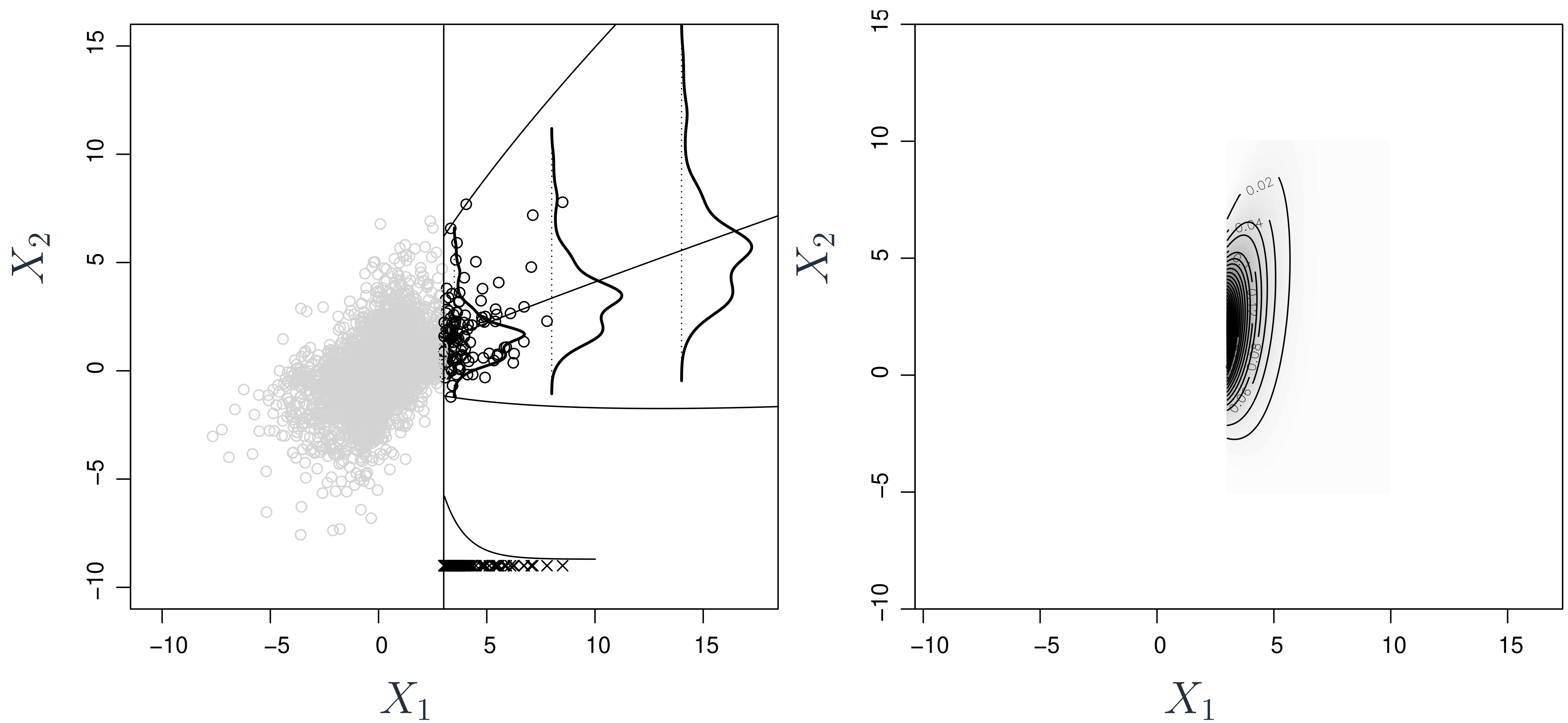
$$X_2 = a_{2|1}(X_1) + b_{2|1}(X_1) Z_{2|1}$$

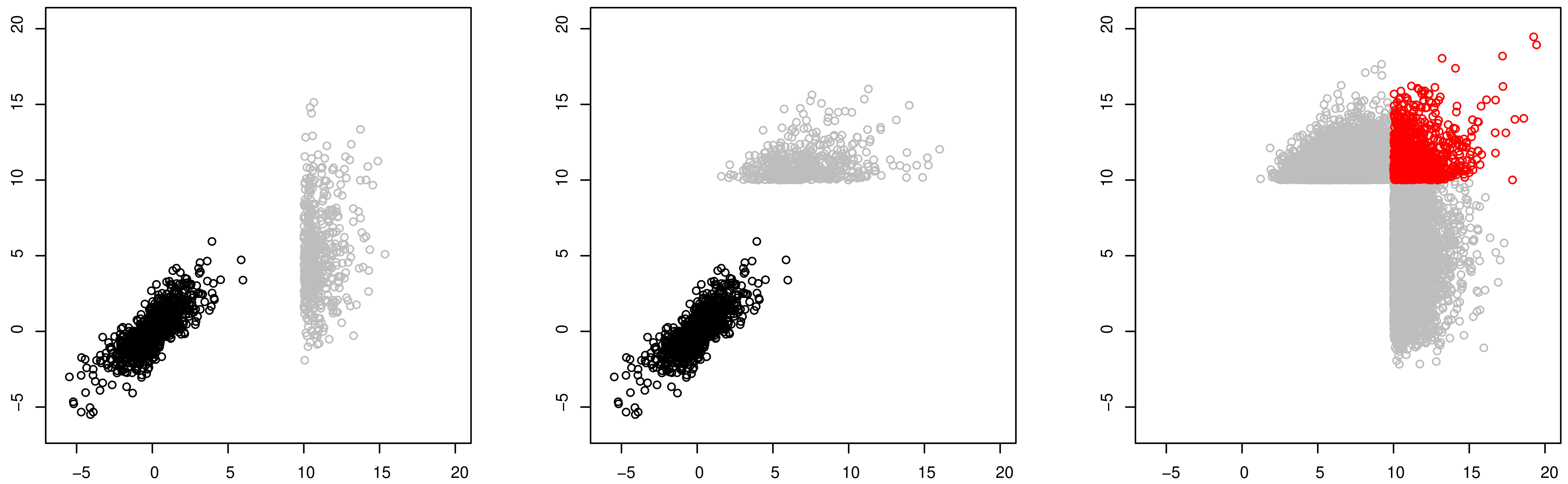
where $Z_{2|1} \sim G_{2|1}$.

- Under mild assumptions $a_{2|1} \in \mathrm{RV}_1$ and $b_{2|1} \in \mathrm{RV}_{\beta_{2|1}}$, where $\beta_{2|1} \in (0, 1)$.
- Parametric models for location-scale, simplify a and b by
  - $a_{2|1}(x) = \alpha_{2|1} x$, where $\alpha_{2|1} \in [-1, 1]$;
  - $b_{2|1}(x) = x^{\beta_{2|1}}$.
- $G_{2|1}$ does not admit a simple parametric form. False working assumption: $Z_{2|1} \sim \mathcal{N}(\mu_{2|1}, \sigma^2_{2|1})$.

A function $f : \mathbb{R} \to \mathbb{R}$ is regularly varying at ∞ with exponent ρ if $\lim_{u\to\infty} f(ux)/f(u) = x^\rho$ for all $x \geq 0$. Notation: $f \in \mathrm{RV}_\rho$.

Illustration





−5
0
5
10
15
20

Asymptotic dependence: $\alpha_{2|1} = 1$ or $\alpha_{2|1} = -1$, and $\beta_{2|1} = 0$;

Asymptotic independence: $\alpha_{2|1} \in (-1, 1)$ and $\beta_{2|1} \in (0, 1)$

Conditional extremes of stationary Markov processes

- Y_t: measurement process, assumed to be a stationary-time series with a continuous marginal distribution F_Y.
- X_t: standardized time series $X_t = F_L^{-1}\{F_Y(Y_t)\}$ where

$$F_L(x) = \begin{cases} 1 - \exp(y)/2 & y \geq 0 \\ \exp(y)/2 & y < 0 \end{cases}$$

is the standard Laplace distribution function.

- Standardization gives that $X_t \sim \text{Laplace}(0, 1)$ for all t.

Conditional extremes of a Markov process

- Suppose that F is a bivariate distribution with common standard Laplace margins admitting regular conditional probabilities.
- There exists a stationary Markov process $\{X_t : t \in \mathbb{N}\}$ such that $(X_t, X_{t+1}) \sim F$ for all t.
- Suppose further a conditioned limit law exists. For simplicity, suppose there exist $\alpha \in (0, 1]$, $\beta \in (0, 1)$ and a distribution G supported on $\mathbb{R}$ such that

$$\left(X_0 - u, \frac{X_1 - \alpha X_0}{X_0^\beta} \right) \Bigg| \{X_0 = u\} \xrightarrow{d} (E, Z_1)$$

where $E \sim \exp(1)$, $Z \sim G$, and E is independent of Z.

Recall: a process X_t is Markov if the distribution of $X_t \mid X_{t-1}, X_{t-2}, \ldots$ is equal to the distribution of $X_t \mid X_{t-1}$ for all t.
When saying that a distribution is supported on a subset A of $\mathbb{R}^k$, we do not allow the distribution to place mass at the boundary ∂A of A.

- Conditionally on $X_0 = u$ and for large u

$$X_1 \approx \alpha X_0 + X_0^{\beta} Z_1, \qquad Z_1 \sim G.$$

- For sufficiently large u, X_2 will be large too. From stationarity

$$X_2 \approx \alpha X_1 + X_1^{\beta} \varepsilon_2, \qquad \varepsilon_2 \sim G$$

$$= \alpha(\alpha X_0 + X_0^{\beta} Z_1) + (\alpha X_0 + X_0^{\beta} Z_1)^{\beta} \varepsilon_2$$

$$\approx \alpha^2 X_0 + \alpha X_0^{\beta} Z_1 + (\alpha X_0)^{\beta} \varepsilon_2 + \alpha^{\beta} \beta X_0^{2\beta-1} Z_1 \varepsilon_2$$

- After rearrangement, we see that

$$\frac{X_2 - \alpha^2 X_0}{X_0^{\beta}} \,\Big|\, \{X_0 = u\} \xrightarrow{d} (Z_1, Z_2)$$

where $Z_2 = \alpha Z_1 + \alpha^{\beta} \varepsilon_2$, where ε_2 is independent of Z_1.

- The heuristic derivations in the previous slide can be formalized under mild conditions that are realistic for practical applications.

- When appropriately renormalized, the Markov chain converges to a non-degenerate process in the sense that for all $t \geq 1$,

$$\left(X_0 - u, \frac{X_1 - \alpha_1 X_0}{X_0^{\beta_1}}, \ldots, \frac{X_t - \alpha_t X_0}{X_0^{\beta_t}} \right) \Big| \{X_0 > u\} \xrightarrow{d} (E, Z_1, \ldots, Z_t)$$

where $\alpha_t = \alpha^t$, $\beta_t = \beta$, $E \sim \exp(1)$ is independent of $(Z_1, \ldots, Z_t)$ and

$$Z_t = \alpha Z_{t-1} + \alpha^\beta \varepsilon_t$$

where $Z_0 = 0$ a.s. and $\{\varepsilon_t\}_{t=1}^{\infty}$ is a sequence of IID random variables from G.

- Classification of tail processes:
  - If $|\alpha_t| = 1$ and $\beta_t = 0$ for all t, then the tail process $\{Z_t\}$ is termed the *tail-chain* of $\{X_t\}$.
  - If $|\alpha_t| \neq 1$ for all t, then the process is termed the *hidden tail-chain* of $\{X_t\}$.

- For higher-order Markov processes where $k > 1$ we need additional initial conditions.

- Think about analogy with ordinary differential equations: the general solution to a second-order ODE requires two initial conditions, one for the value of the function at some time, say $t = 0$, and the other for the derivative of the function at $t = 0$.

- Initial conditions for stochastic recurrence relation: we can find $\boldsymbol{a}_{1\,:\,k-1} \;:\; \mathbb{R}^{k-1} \to \mathbb{R}$ and $\boldsymbol{b}_{1\,:\,k-1} \;:\; \mathbb{R}^{k-1} \to \mathbb{R}_+$ such that

$$\boxed{\frac{\boldsymbol{X}_{1\,:\,k-1} - \boldsymbol{a}_{1\,:\,k-1}(X_0)}{\boldsymbol{b}_{1\,:\,k-1}(X_0)} \,\bigg|\, \{X_0 > u\} \xrightarrow{d} \boldsymbol{Z}_{1\,:\,k-1} \qquad \text{as } u \to \infty,}$$

where each random variable Z_i has a non-degenerate distribution supported on $\mathbb{R}$.

Vector algebra is interpreted element-wise

- Consider how a complete characterization may be given for higher-order Markov processes with $k > 1$ using induction on $\mathbb{N}$. Fix a $t \geq k > 1$ and assume there exist sequences of norming functions a_i and b_i, $i = 1, \dots, t-1$, such that,

$$\frac{\boldsymbol{X}_{1:t-1} - \boldsymbol{a}_{1:t-1}(X_0)}{\boldsymbol{b}_{1:t-1}(X_0)} \,\Big|\, \{X_0 > u\} \xrightarrow{d} \boldsymbol{Z}_{1:t-1} \qquad \text{as } u \to \infty,$$

where each Z_i is a random variable with a non-degenerate distribution on $\mathbb{R}$.

- Suffices to consider marginal convergence, that is,

$$\mathbb{P}\left(\frac{X_t - a_t(X_0)}{b_t(X_0)} \leq x_t \,\Big|\, X_0 > u\right) =$$

$$\frac{1}{\overline{F}_0(u)} \int_u^{\infty} \int_{\mathbb{R}^{t-1}} \mathbb{P}\left(\boldsymbol{X}_{0:t-1} \in d\boldsymbol{x}_{0:t-1}\right) \left[\int_{-\infty}^{x_t} \mathbb{P}\left(\frac{X_t - a_t(X_0)}{b_t(X_0)} \in dz_t \,\Big|\, \boldsymbol{X}_{0:t-1} = \boldsymbol{x}_{0:t-1}\right)\right],$$

- Replace $a_t(X_0)$ by $a_t(x_0)$ by conditioning on the exact value of X_0 being equal to x_0.

- Use the Markov property so that the conditioning on all previous states is reduced to conditioning on the previous k states.

- Change variables to $z_0 = x_0 - u$ and $z_i = \{x_i - a_i(x_0)\}/b_i(x_0)$ for $i = 1, \ldots, t-1$, for any $t \geq k$, previous expression is equal to

$$\int_0^\infty \frac{F_0\{v_u(dz_0)\}}{\overline{F}_0(u)} \times \Bigg[\int_{\mathbb{R}^{t-1}} \mathbb{P}\left(\frac{\boldsymbol{X}_{1:t-1} - \boldsymbol{a}_{1:t-1}\{v_u(z_0)\}}{\boldsymbol{b}_{1:t-1}\{v_u(z_0)\}} \in d\boldsymbol{z}_{1:t-1} \,\middle|\, X_0 = v_u(z_0) \right) \times$$

$$\times \boxed{\left\{ \int_{-\infty}^{x_t} \mathbb{P}\left(\frac{X_t - a_t\{v_u(z_0)\}}{b_t\{v_u(z_0)\}} \in dz_t \,\middle|\, \frac{\boldsymbol{X}_{t-k:t-1} - \boldsymbol{a}_{t-k:t-1}\{v_u(z_0)\}}{\boldsymbol{b}_{t-k:t-1}\{v_u(z_0)\}} = \boldsymbol{z}_{t-k:t-1} \right) \right\}} \Bigg],$$

where $a_0(x) = x$, $b_0(x) = 1$ for all $x \in \mathbb{R}$ and $v_u(z_0) = u + z_0$.

Higher-order Markov processes

- To ensure marginal convergence under the induction hypothesis, we require convergence of

$$\mathbb{P}\left(\frac{X_t - a_t(u)}{b_t(u)} \le z_t \;\middle|\; \frac{\boldsymbol{X}_{t-k:t-1} - \boldsymbol{a}_{t-k:t-1}(u)}{\boldsymbol{b}_{t-k:t-1}(u)} = \boldsymbol{z}_{t-k:t-1}\right)$$

- Let $\boldsymbol{A}_t(u, \boldsymbol{z}) = \boldsymbol{a}_{t-k:t-1}(u) + \boldsymbol{b}_{t-k:t-1}(u)\,\boldsymbol{z}$ for $\boldsymbol{z} \in \mathbb{R}^k$, and rewrite probability above as

$$\mathbb{P}\left[\frac{X_t - a(\boldsymbol{X}_{t-k:t-1})}{b(\boldsymbol{X}_{t-k:t-1})} \le \frac{dz_t}{\psi^b_{t,u}(\boldsymbol{X}_{t-k:t-1})} - \psi^a_{t,u}(\boldsymbol{X}_{t-k:t-1}) \;\middle|\; \boldsymbol{X}_{t-k:t-1} = \boldsymbol{A}_t(u, \boldsymbol{z}_{t-k:t-1})\right].$$

- $a : \mathbb{R}^k \to \mathbb{R}$ and $b : \mathbb{R}^k \to \mathbb{R}_+$ are location and scale functionals and $\psi^a_{t,u}$ and $\psi^b_{t,u}$ satisfy

$$\psi^a_{t,u}(\boldsymbol{z}) = \frac{a\{\boldsymbol{A}_t(u, \boldsymbol{z})\} - a_t(u)}{b_t(u)} \qquad \text{and} \qquad \psi^b_{t,u}(\boldsymbol{z}) = \frac{b\{\boldsymbol{A}_t(u, \boldsymbol{z})\}}{b_t(u)}$$

In what follows, we assume that there exist:

- for $t = k, k+1, \ldots,$
  - location and scale functions $a_t : \mathbb{R} \to \mathbb{R}$ and $b_t : \mathbb{R} \to \mathbb{R}_+$,,
  - continuous update functions $\psi_t^a : \mathbb{R}^k \to \mathbb{R}$ and $\psi_t^b : \mathbb{R}^k \to \mathbb{R}_+$,
  - and location and scale functionals $a : \mathbb{R}^k \to \mathbb{R}$ and $b : \mathbb{R}^k \to \mathbb{R}_+$,

  such that, for all $\boldsymbol{z} \in \mathbb{R}^k$,

$$\lim_{u\to\infty} \psi_{t,u}^a(\boldsymbol{z}_u) \to \psi_t^a(\boldsymbol{z}) \qquad \text{and} \qquad \lim_{u\to\infty} \psi_{t,u}^b(\boldsymbol{z}_u) \to \psi_t^b(\boldsymbol{z})$$

  whenever $\boldsymbol{z}_u \to \boldsymbol{z} \in \mathbb{R}^k$.

- a non-degenerate distribution K supported on $\mathbb{R}$, such that for all $\boldsymbol{z} \in \mathbb{R}^k$,

$$\mathbb{P}\left[\frac{X_k - a(\boldsymbol{X}_{0:k-1})}{b(\boldsymbol{X}_{0:k-1})} \leq \varepsilon \,\middle|\, \boldsymbol{X}_{0:k-1} = \boldsymbol{A}_t(u, \boldsymbol{z}_{0:k-1})\right] \to K(\varepsilon)$$

If $\{X_t : t = 0, 1, \ldots\}$ is a stationary Markov chain satisfying the aforementioned assumptions, then

$$\left(\frac{X_0 - u}{\sigma(u)}, \frac{X_1 - a_1(X_0)}{b_1(X_0)}, \ldots, \frac{X_t - a_t(X_0)}{b_t(X_0)} \right) \Bigg| \{X_0 > u\} \xrightarrow{d} (E_0, Z_1, \ldots, Z_t), \qquad t \geq k,$$

where

(i) $E_0 \sim H_0$ and $(Z_1, Z_2 \ldots, Z_t)$ are independent,

(ii) $Z_0 = 0$ a.s., $(Z_1, \ldots, Z_{k-1}) \sim G$ and

$$Z_s = \psi_s^a(\boldsymbol{Z}_{s-k\,:\,s-1}) + \psi_s^b(\boldsymbol{Z}_{s-k\,:s-1})\,\varepsilon_s, \qquad s = k, k+1, \ldots$$

for a sequence of IID random variables $\varepsilon_s \sim K$.

Example: max-stable distribution with Laplace margins

- Suppose $\boldsymbol{X}_{0:k}$ follows a $(k+1)$-variate logistic max-stable distribution with standard Laplace margins and dependence parameter $\kappa \in (0, 1]$, that is,

$$\mathbb{P}(\boldsymbol{X}_{0:k} \leq \boldsymbol{x}_{0:k-1}) = \exp\{-V\left(T(x_0), \ldots T(x_k)\right)\},$$

where $T(x) = -1/\log\left(F_L(x)\right)$, and

$$V(x_0, \ldots, x_k) = \left(x_0^{-1/\kappa} + \cdots + x_k^{-1/\kappa}\right)^{\kappa}$$

- Define

$$a(x_0, \ldots, x_{k-1}) = -\kappa \log(e^{-x_0/\kappa} + \cdots + e^{-x_{k-1}/\kappa}) \qquad \text{and} \qquad b(x_0, \ldots, x_{k-1}) = 1.$$

- Convergence of transition probability distribution: for $\varepsilon \in \mathbb{R}$,

$$\boxed{\mathbb{P}\left(\frac{X_k - a(\boldsymbol{X}_{0:k-1})}{b(\boldsymbol{X}_{0:k-1})} \leq \varepsilon \mid \boldsymbol{X}_{0:k-1} = u\mathbf{1} + \boldsymbol{z}_{0:k-1}\right) \to [1 + \exp(-\varepsilon/\kappa)]^{\kappa-k}, \quad u \to \infty.}$$

for all $\boldsymbol{z}_{0:k-1} \in \mathbb{R}^k$

Example: inverted max-stable distribution with Laplace margins

- Suppose $\boldsymbol{X}_{0:k}$ follows a $(k+1)$-variate inverted max-stable logistic distribution with dependence parameter $\kappa \in (0,1]$ and standard Laplace margins, that is,

$$\mathbb{P}(\boldsymbol{X}_{0:k} > \boldsymbol{x}_{0:k-1}) = \exp\{-V(1/x_0, \ldots 1/x_k)\},$$

where

$$V(x_0, \ldots, x_k) = \left(x_0^{-1/\kappa} + \cdots + x_k^{-1/\kappa}\right)^{\kappa}$$

- Define

$$a(x_0, \ldots, x_{k-1}) = 0 \qquad \text{and} \qquad b(x_0, \ldots, x_{k-1}) = (x_0^{1/\kappa} + \cdots + x_{k-1}^{1/\kappa})^{\kappa(1-\kappa)}.$$

- Convergence of transition probability distribution: for $\varepsilon \in \mathbb{R}_+$,

$$\boxed{\mathbb{P}\left\{\frac{X_k - a(\boldsymbol{X}_{0:k-1})}{b(\boldsymbol{X}_{0:k-1})} \le z \mid \boldsymbol{X}_{0:k-1} = (u\, z_0, u^{1-k}\boldsymbol{z}_{0:k-1})\right\} \to 1 - \exp(-\kappa\, \varepsilon^{1/\kappa}), \quad u \to \infty.}$$

for all $\boldsymbol{z}_{0:k-1} \in \mathbb{R}_+^k$

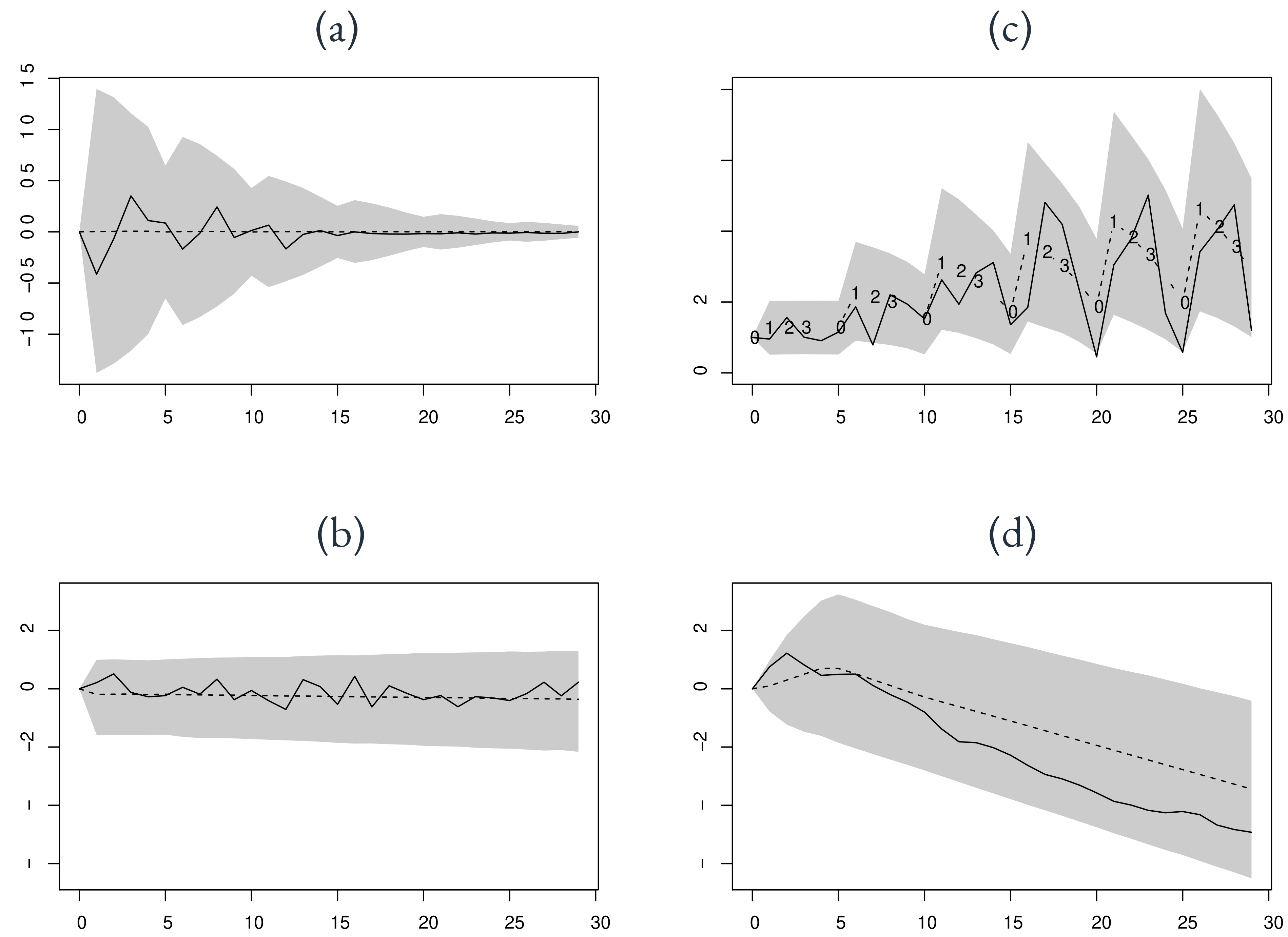


Figure 1: Pointwise 2.5% and 97.5% quantiles of the sampling distribution (shaded region), mean of the sampling distribution (dashed line) and one realization from the (hidden) tail-chain (solid line). The parameters for all copulas are chosen such that the coefficient of residual tail dependence η [Ledford and Tawn, 1997] and the extremal coefficient [Beirlant et al., 2004] are equal. The distribution of $\boldsymbol{X}_{0:k}$ that was used in each example is: (a) standard multivariate Gaussian, (b) inverted logistic, (c) max-stable logistic and (d) max-stable Hüsler–Reiss

- Asymptotically dependent process (Multivariate Regular Variation):
  - $a(\boldsymbol{x}) = \log g(e^{\boldsymbol{x}})$ where g is 1–homogeneous.
  - $b(\boldsymbol{x}) = 0$
  - $Z_t = a(\boldsymbol{Z}_{t-k:t-1}) + \varepsilon_t$

- Asymptotically independent process (Hidden Regular Variation) with a not zero:
  - a is 1–homogeneous with $a(\mathbf{1}) < 1$
  - b is β–homogeneous with $\beta \in (0, 1)$
  - $Z_t = \nabla a(\boldsymbol{\alpha}_{t-k:t-1})^\top \boldsymbol{Z}_{t-k:t-1} + b(\boldsymbol{\alpha}_{t-k:t-1})\, \varepsilon_t$

- Asymptotically independent process (Hidden Regular Variation) with a zero:
  - b is β–homogeneous with $\beta \in (0, 1)$.
  - Structure of tail process is complicated.

Properties of normalizing functions and structure of tail processes

- Asymptotically dependent process (Multivariate Regular Variation):
  - $a_t(x) = x$
  - $b_t(x) = 0$
  - $\mathbb{E}(Z_t) < 0$ for all t

- Asymptotically independent process (Hidden Regular Variation) with a not zero:
  - $a_t(x) = \alpha_t x$ where $\alpha_t = a(\boldsymbol{\alpha}_{t-k:t-1})$ and $\alpha_t \to 0$ as $t \to \infty$
  - $b_t(x) = x^\beta$ where $\beta \in (0, 1)$
  - $Z_t \xrightarrow{p} 0$ as $t \to \infty$

- Asymptotically independent process (Hidden Regular Variation) with a zero:
  - $\log \beta_t = \log \beta + \log\left(\max_{i=1,\ldots,k} \beta_{t-i}\right)$
  - Structure of tail process is complicated.

Comments on Theorem

- The entities we have referred to as tail-chains and hidden tail-chains are in fact forward tail-chains and hidden tail-chains.
- These describe the behaviour of the Markov chain only forward in time from a large observation.
- There is also the parallel interest in a backward tail/hidden tail chain, to give how the chain evolves into an extreme event, and the joint behaviour of the two, known as back-and-forth tail processes.

Back-and-forth hidden tail-chain

Statistical inference with time series conditional extremes

- Assume that $X_t \sim F_L$ where F_L denotes the standard Laplace distribution.

- Let $T_u = \{t \in \mathbb{Z} : X_t > u\}$ denote the set of times where the process exceeds the level u.

- We assume that 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 $A \subset \mathbb{Z}$ with $|A| < \infty$ and for all $t \in T_u$,

$$\left(X_t - u, \frac{\boldsymbol{X}_A - \boldsymbol{a}_{A-t}(X_t)}{\boldsymbol{b}_{A-t}(X_t)} \right) \Bigm| X_t > u \xrightarrow{d} (E_t, \boldsymbol{Z}_A^t), \tag{2}$$

where
- $E_t \sim \exp(1)$ is independent of $\boldsymbol{Z}_A^t \sim G_{A\setminus t}$;
- where $G_{A\setminus t}$ has non-degenerate margins;
- $a_0(x) = x$ and $b_0(x) = 1$. This gives $Z_t^t = 0$.

- The limit relation (2) is taken to hold exactly above a sufficiently large level u. That is, we assume that for $t \in T_u$

$$X_t = u + E_t$$

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

$$\boldsymbol{Z}^t_{t-L:t+L} \sim G_{-L:L}$$

- $L \in \mathbb{N}$ chosen sufficiently large so that extreme episodes are contained within $t-L : t+L$.
- Context often dictates the choice of L.

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

- Structure of norming functions:
  - Asymptotic dependence gives that $a_{|i|}(x) = x$ and $b_{|i|}(x) = 0$.
  - Asymptotic independence with $a_{|i|}$ not zero gives $a_{|i|}(x) \to 0$ as $i \to \infty$. So for large $|i|$ we have that $Z_{t+i}^{t} \sim F_L$.
  -

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

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

- Recurrence based on correlation model
  - Order–1 Markov chain

$$\alpha_t = \theta\, \alpha_{t-1}, \quad 1 \leq t \leq L, \quad \text{with } \alpha_0 = 1, \theta \in [-1, 1] \tag{3}$$

  - Order–2 Markov chain

$$\alpha_t = \theta_1 \alpha_{t-1} + \theta_2 \alpha_{t-2}, \quad 2 \leq t \leq k, \quad \text{with } \alpha_0 = 1, \alpha_1 = \theta_1/(1-\theta_2). \tag{4}$$

- For asymptotically independent models, the parameters θ, θ_1 and θ_2 are subject to constraints to ensure $\alpha_t \to 0$ as $t \to \infty$. In particular, we have $\theta \in (-1, 1)$ and $r_1, r_2, r_3 \in (-1, 1)^3$ where

$$\begin{aligned} \theta_1 &= r_1 - r_1 r_2 - r_2 r_3 \\ \theta_2 &= r_2 - r_1 r_3 + r_1 r_2 r_3 \\ \theta_3 &= r_3 \end{aligned}$$

Parametric forms for norming functions based on Markov chain theory

- Constraints for higher-order Markov processes may be obtained in a simpler way.
- Regarding the correlation model, we can rewrite this as

$$\alpha_t = \sum_{i=1}^{k} \theta_i \alpha_{t-i} = c \sum_i \gamma_i \alpha_{t-i}$$

where $c = \sum_i \theta_i$ and $\gamma_i = \theta_i / c$ so that $\sum_i \gamma_i = 1$.

- coefficients γ_i are free parameters on the simplex so can be reparametrized as

$$\gamma_i = \frac{\exp(\Gamma_i)}{\sum_{i=1}^{k} \exp(\Gamma_i)}, \qquad 1 \leq i \leq L$$

where $\Gamma_i \in \mathbb{R}$, subject to a sum-to-zero identifiability constraint $\sum_{i=1}^{k} \Gamma_i = 0$.

- To ensure $\alpha_t \to 0$ as $t \to \infty$, we require $|c| < 1$

Marginal model for the residual

- $\boldsymbol{Z}^t_{t-L\,:\,t+L} = (Z^t_{t-i\,:\,t+i} \,:\, i = -L, \ldots, L)$ where $Z^t_{t+i} \sim G_i$ with density

$$g_i(z) = \frac{\delta_i}{2\sigma_i \Gamma(1/\delta_i)} \exp\left\{ - \left| \frac{z - \mu_i}{\sigma_i} \right|^{\delta_i} \right\}, \quad \mu_i \in \mathbb{R},\ \sigma_i > 0,\ \delta_i > 0. \tag{5}$$

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

Dependence model for the residual

- In general, the components of the residual vector $\boldsymbol{Z}^t_{t-L\,:\,t+L}$ will be dependent.

- Simplest approach is in style with copula modelling methods. Let $\boldsymbol{\Sigma}$ be a $(2L+1) \times (2L+1)$ covariance matrix and let $\boldsymbol{Q}$ be $\boldsymbol{\Sigma}^{-1}$ without the row and column associated with Z^t_0.

- Model for residual vector $\boldsymbol{Z}^t_{t-L\,:\,t+L} \mid Z^t_0 = 0$, based on conditional normal distribution. This leads to

$$\left(\Phi^{-1}\left\{G^Z_{t-i}\left(Z^t_{t-i}\right)\right\} \,:\, i = -L, \ldots, -1, 1, \ldots, L\right) \sim \mathrm{MVN}\left(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}\right)$$

Forward sampling

- Interested in estimating conditionals expectations of the form

$$\mathbb{E}(g(\boldsymbol{X}_{0:d}) \mid X_1 > v) \qquad v > u, d \in \mathbb{N},$$

for some function g.

- Monte-Carlo based estimator by simulate forwards in time from extreme event

Forward sampling algorithm

input : Threshold $v > u, d, N \in \mathbb{N}$ and constants $(\hat{\boldsymbol{\alpha}}_{0:k}, \hat{\beta})$ from fitted conditional model.
output: An estimate of $\mathbb{E}(g(\boldsymbol{X}_{0:d}) \mid X_1 > v)$

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

References

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes, Theory and Applications*. Wiley, 2004.

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

T. Hsing, J. Hüsler, and M. Leadbetter. On the exceedance point process for a stationary sequence. *Probab. Theory Reltd Flds*, 78:97–112, 1988.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc., B*, 59:475–499, 1997.

J. Pickands. Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131, 1975.