STOR834 p.82 3/27/25

# Estimating the extremal index

1. Parametric approach
2. Blocks & runs approach
3. Ferro–Segers method

1. Idea of generalizing previous Markov chain method.

Suppose $\{X_n\}$ is a Markov chain with parametric family for $(X_n, X_{n+1})$

Example:

$$P(X_n \le x, X_{n+1} \le y\} = \exp\left[-\left(e^{-rx} + e^{-ry}\right)^{1/r}\right]$$

(used for illustration: any bivariate GVD family would suffice)

Approximate as

$$1 - F(x,y) \sim \left(e^{-rx} + e^{-ry}\right)^{1/r} \qquad x, y \text{ large}$$

Generalization:

$$1-F(x,y) \approx \left[\left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-r/\xi}+\left(1+\xi\,\frac{y-\mu}{\psi}\right)^{-r/\xi}\right]^{1/r} \qquad (*)$$

Idea of threshold approach: assume $(*)$ is exact for $x \geq u$, $y \geq u$.

Joint density of Markov chain:

$$f(X_1,\dots,X_N) = f(X_1)\cdot\prod_{i=2}^{N} f(X_i \mid X_{i-1})$$

$$= \frac{\prod_{i=2}^{N} f(X_{i-1},X_i)}{\prod_{i=2}^{N-1} f(X_i)} \qquad (**)$$

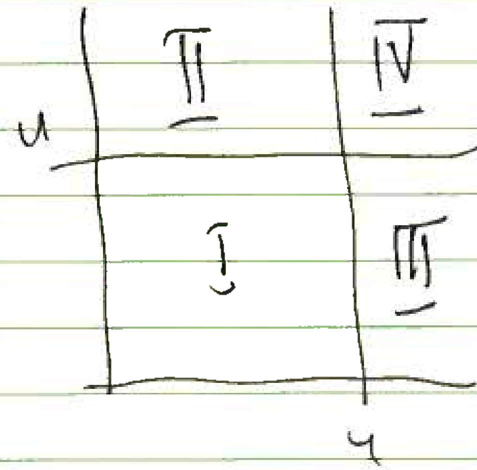
Need approximations for numerator and denominator

Denom: assume $1-F(x) = \left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}}$, $x>u$

$$\prod_{i=2}^{N-1} f(X_i) \approx \prod_{i=2}^{N-1}\left\{1-\left(1+\xi\,\frac{u-\mu}{\psi}\right)^{-\frac{1}{\xi}}\right\} I(X_i \leq u)$$

$$\cdot\ \frac{1}{\psi}\left(1+\xi\,\frac{X_i-\mu}{\psi}\right)^{-1/\xi-1} I(X_i>u)$$

For numerators in (**): divide $(X_{n-1}, X_n)$ space into 4 quadrants




On I: likelihood contribution is $F(u,u)$

On II: $\frac{\partial F(u,y)}{\partial y}$

On III: $\frac{\partial F(x,u)}{\partial x}$

On IV: $\frac{\partial^2 F(x,y)}{\partial x \partial y}$

Estimation of $\theta$: having fit model, use theoretical formula for EI (Wiener-Hopf approach or simulation)

Disadvantage: Requires specifying a dependence model and fitting numerically

## 2. Blocks and Runs Approach

Blocks approach: sample of size $n$, threshold $u_n$
Let $N_n$ be # of obs. that exceed $u_n$

Divide into blocks of length $r_n$, let $k_n = \frac{n}{r_n}$

$Z_n$: # of blocks with $\geq 1$ exceedance

$$\hat{\theta}_n = \frac{Z_n}{N_n} \quad \text{"blocks estimator"}$$

Refinement (S&Weissman 1996):

$$\tilde{\theta}_n = \frac{\log(1 - Z_n/k_n)}{r_n \log(1 - N_n/n)}$$

Runs approach: Let $W_{n,i} = 1$ if $X_i > u_n$, 0 o.w.

$$N_n = \sum_{i=1}^{n} W_{n,i} \qquad Z_n^* = \sum_{i=1}^{n-r_n} W_{n,i}(1 - W_{n,i+1}) \cdots (1 - W_{n,i+r_n})$$

$$\bar{\theta}_n = \frac{Z_n^*}{N_n}$$

Refs: ~~Led~~ Leadbetter et al (1989)
Nandagopalan (1990), S&W (1994)

## 3. Ferro-Segers method (Ferro & Segers 2003)

Objective: justify runs estimator with objective choice of r.

Let $T(u)$: interval between exceedance times of threshold $u$.

Asymptotically, $\bar{F}(u)T(u) \sim (1-\theta)\,\varepsilon_0 + \theta\,\mu_\theta$

$\varepsilon_0 \downarrow$ atom at 0   $\mu_\theta \downarrow$ exp. mean $\frac{1}{\theta}$

If RHS is $T_\theta$, then $ET_\theta = \frac{1}{\theta},\ ET_\theta^2 = \frac{2}{\theta^2}$

exploit to get estimator of $\theta$

e.g. $$\bar{\theta}_n(u) = \frac{2(N-1)}{\bar{F}_n(u)^2 \sum_1^{N-1} T_i^2}$$

where $N$: # of exceedances

$T_i$: interval between exceedances $i$ and $i+1$

$\bar{F}_n(u) \approx \frac{N}{n}$.   Doesn't require r!

Other ideas in paper:

a) Refinements of $\bar{\theta}_n(u)$

b) Use for declustering