1/14/2025 p.6

Alternative Models for Extremes

(a) Exceedances over Thresholds

(b) r-largest order statistics approach

(c) Point process approach

Consider (a) first. —— and 1/14/25

Begin 1/16/25 here

Motivation for an exceedances-based approach: given that the variable of interest is above some threshold, how high can it go?

Practical interpretation: "Let the tail speak for itself"

Consider distribution conditional on exceeding a threshold u,

$$F_u(y) = P\{Y \le u+y \mid Y > u\} = \frac{F(u+y) - F(u)}{1-F(u)}, \; y > 0.$$

Scaling result in this case:

Define $\omega_F = \sup\{x : F(x) < 1\}$, consider limits of form

$$\lim_{u \uparrow \omega_F} F_u\left(\frac{y}{\sigma_u}\right) = G(y) \qquad (**)$$

For some $\sigma_u > 0$, nondegenerate G.

Balkema–deHaan–Pickands theorem (1974, 1975)

If limit (**) exists, then $G(y)$ must be of form

$$G(y;\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{y}{\sigma}\right)_+^{-1/\xi} & \xi \neq 0 \\ 1-e^{-y/\sigma} & \xi = 0 \end{cases}$$

and the limit (**) exists if and only if limit (*) exists for the same F.

Informal motivation: Poisson-GPD model

Consider a fixed time unit (say, 1 year)

N: # of exceedances in a year, assume Poisson(λ)

Conditional on N, exceedances $Y_1,\dots,Y_N$ are GPD(σ,ξ)

$$P\left\{\max_{1\le i\le N} Y_i \le x\right\} = P(N=0) + \sum_{n=1}^{\infty} P(N=n, Y_1\le x,\dots,Y_n\le x)$$

$$= e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!}\left\{1-\left(1+\xi\frac{x-\cancel{u}}{\sigma}\right)_+^{-1/\xi}\right\}^n$$

$$= \exp\left\{-\lambda\left(1+\xi\frac{x-u}{\sigma}\right)_+^{-1/\xi}\right\}$$

which is of GEV form.

(b) r-largest order statistics approach

Formulation: $X_1, X_2, X_3, \ldots$ IID(F)

Sample size n, order $X_{n,1} \geq X_{n,2} \geq \cdots \geq X_{n,n}$

Assume $\frac{X_{n,1}-b_n}{a_n} \xrightarrow{d} Y_1$ where $Y_1 \sim H$ ($H = $ GEV)

~~Also assume c~~

Also assume convergence of densities (a little bit stronger) [do we need this?]

Then for fixed $r \geq 1$ as $n \to \infty$,

$$\left(\frac{X_{n,1}-b_n}{a_n}, \frac{X_{n,2}-b_n}{a_n}, \ldots, \frac{X_{n,r}-b_n}{a_n} \right) \xrightarrow{d} (Y_1, \ldots, Y_r)$$

where the joint density of $(Y_1, \ldots, Y_r)$ at $(y_1, \ldots, y_r)$ is

$$h(y_1, \ldots, y_r) = \psi^{-r} \prod_{i=1}^{r} \left(1 + \xi \frac{y_i - \mu}{\psi} \right)^{-1/\xi - 1} \cdot \exp\left\{ -\left(1 + \xi \frac{y_r - \mu}{\psi} \right)^{-1/\xi} \right\}$$

valid so long as $\left(1 + \xi \frac{y_i - \mu}{\psi}\right) > 0$ for each $i = 1, \ldots, r$

Origins: papers by Dwass, Lamperti, Weissman, etc. on probability derivations

Later: statistical applications (hydrology extremes, track records etc.). Caution about independence of raw observations.

Point Process Approach

Some basic notation and terminology

Sample space $\mathcal{S}$ is a complete separable metric space

For us: $\mathcal{S}$ is a measurable subset of $\mathbb{R}^d$, some $d \geq 1$

A point process on $\mathcal{S}$ is a random subset of points in $\mathcal{S}$

Defined by $N(A)$, all $A \in \mathcal{A}$ where $\mathcal{A}$ is some family of subsets of $\mathcal{S}$, e.g. all Borel sets

"Simple" means no multiple points

i.e. $N(\{x\}) = 1$ or 0 for any singleton set $A = \{x\}$.

N is a non-homogeneous Poisson process with intensity Λ if it is simple and

(i) $N(A_1), \ldots, N(A_k)$ are independent RVs for any finite sequence of disjoint sets $A_1, \ldots, A_k$ ($A_j \cap A_k = \phi$ whenever $j \neq k$)

(ii) $P\{N(A) = n\} = \dfrac{\Lambda(A)^n e^{-\Lambda(A)}}{n!}$ $n = 0, 1, 2, \ldots$

Here Λ is any measure on $\mathcal{A}$.

Now assume Λ has a density, i.e. there exists a function $\lambda(x)$, $x \in \mathcal{S}$, such that $\Lambda(A) = \int_A \lambda(x)\,dx$, all $A \in \mathcal{A}$

Notation in book: vector x because in general $\mathcal{S} \subseteq \mathbb{R}^d$ some $d \geq 1$, x is a vector of length d (but theory does apply to $d=1$)

Question: Fix a set $A \in \mathcal{A}$, points at $x_1, x_2, \ldots, x_{N(A)}$, what is joint density of $X_1 = x_1, \ldots, X_{N(A)} = x_{N(A)}$?

Assume $N(A) = n$, assume a small sphere of volume δ around each of $x_1, \ldots, x_n$

What is prob. there is exactly one point in each of $S_1, \ldots, S_n$ and no others in A?

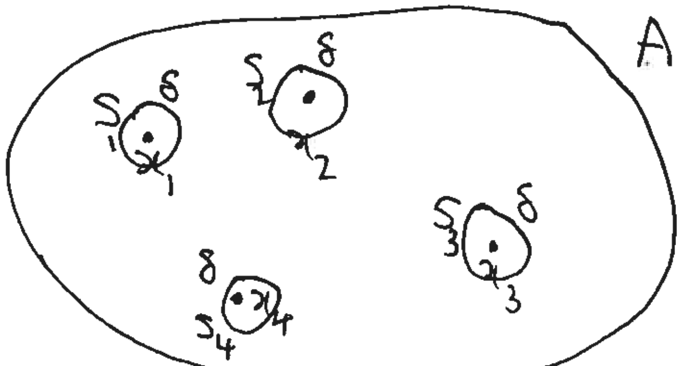

illustration with $n=4$

Answer is

$$\Lambda(S_1)e^{-\Lambda(S_1)} \cdot \Lambda(S_2)e^{-\Lambda(S_2)} \cdots \Lambda(S_n)e^{-\Lambda(S_n)} \cdot e^{-\Lambda(A \setminus S_1 \setminus S_2 \cdots \setminus S_n)}$$

$$= \prod_{i=1}^{n} \Lambda(S_i)\, e^{-\Lambda(A)} \approx \delta^n \lambda(x_1)\cdots\lambda(x_n)\, e^{-\Lambda(A)}$$

Divide by δ, let $\delta \to 0$

End 1/16/25