Generalized extreme value distribution (GEV)

Def

The generalized extreme value distribution (GEV) is given by the distr. function

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & \text{if } \xi \neq 0 \\ \exp(-\exp(-x)) & \text{if } \xi = 0 \end{cases}$$

where $1+\xi x > 0$.

parameter	support	corresponding distr.
$\xi > 0$	$x > -\xi^{-1}$	$\xi = \alpha^{-1} > 0$ Fréchet Φ_α
$\xi = 0$	$x \in \mathbb{R}$	Gumbel Λ
$\xi < 0$	$x < -\xi^{-1}$	$\xi = -\alpha^{-1} < 0$ Weibull Ψ_α

Note: H_0 can be considered as a $\xi \to 0$ limit.

Remark $\left(x \mapsto \frac{x-\mu}{\psi}\right)$
By recentering and rescaling, the support changes accordingly, but the distribution type remains the same.

Let $U(t) = F^{-1}(1 - t^{-1})$ $\quad t > 0$

Thm (Characterization of $MDA(H_\xi)$)

For $\xi \in \mathbb{R}$, the following are equivalent:

1. $F \in MDA(H_\xi)$

② $\exists$ positive measurable $a(\cdot)$ such that for $1+\xi x > 0$,

$$\lim_{t \uparrow x_F} \frac{\bar{F}(t + x\,a(t))}{\bar{F}(t)} = \begin{cases} (1+\xi x)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-x} & \text{if } \xi = 0 \end{cases}$$

③ For $x, y > 0$, $y \neq 1$ ~~and $U(t) = F^{-1}(1-t^{-1})$~~

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{U(ty) - U(y)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{if } \xi \neq 0 \\ \frac{\ln x}{\ln y} & \text{if } \xi = 0 \end{cases}$$

Remarks

- For $\xi > 0$, $a(t) \sim \xi t$, hence by $\bar{F}(t) \sim t^{-\alpha} = t^{-1/\xi}$, $\frac{\bar{F}(t + x a(t))}{\bar{F}(t)} \sim (1+\xi x)^{-\alpha} = (1+\xi x)^{-1/\xi}$
  $U(t) \sim t^{1/\alpha} = t^\xi$
- For $\xi < 0$, $\bar{F}(t) \sim (x_F - t)^\alpha = (x_F - t)^{-1/\xi}$ $a(x) \sim -\xi(x_F - x)$ $\frac{\bar{F}(t + x a(x))}{\bar{F}(t)} = (1+\xi x)^\alpha = (1+\xi x)^{-1/\xi}$ $U(t) = x_F - x^\xi$
- Probabilistic interpretation of condition ②:

$$\lim_{t \uparrow x_F} \mathbb{P}\left(\frac{X - t}{a(t)} \geq x \,\middle|\, X \geq t \right) = \begin{cases} (1+\xi x)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-x} & \text{if } \xi = 0 \end{cases}$$

rescaled excesses over a threshold t.

Def

Let X be a r.v. with distr. F and with x_F. For any $t < x_F$,

$$F_t(x) = \mathbb{P}(X - t < x \mid X \geq t) \qquad x > 0$$

is the excess distribution function of X over the threshold t.

(or residual lifetime distr after t)

Def (generalized Pareto distribution (GPD))

The generalized Pareto distribution (GPD) function is given by

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-x} & \text{if } \xi = 0 \end{cases}$$

where $x > 0$ if $\xi \geq 0$ and $x \in (0, -1/\xi)$ if $\xi < 0$.

Without changing the type of the distribution, we can recenter and rescale by replacing x by $\frac{x - \mu}{\psi}$.

Corollary

$$F \in MDA(H_\xi) \iff \lim_{t \uparrow x_F} \sup_{x \in [0, x_F - t]} \left| F_t(x) - G_\xi\left(x / a(t)\right) \right| = 0$$ for some positive function $a(\cdot)$.

Modelling

Exceedances of a high threshold: Poisson process, excess distribution: GPD.

Exceedances of levels and the kth largest maxima

$$S_n := \sum_{k=1}^{n} \mathbb{1}(X_k \geq u_n)$$

Prop

If $\{u_n\}$ is a sequence which satisfies $n\bar{F}(u_n) \to \tau \in [0,\infty]$, then

$$\mathbb{P}(S_n = k) \to e^{-\tau} \frac{\tau^k}{k!}$$

Proof: by the proof of the Poisson approx. proposition.

Let $M_n^{(k)}$ be the kth largest of $\{X_1, X_2, \dots, X_n\}$, in particular $M_n^{(1)} = M_n$.

Note that the following events are equal: $\{M_n^{(k)} < u_n\} = \{S_n < k\}$

Corollary

If the sequence $\{u_n\}$ satisfies $n\bar{F}(u_n) \to \tau \in [0,\infty]$, then

$$\mathbb{P}(M_n^{(k)} < u_n) = e^{-\tau} \sum_{j=0}^{k-1} \frac{\tau^j}{j!}$$

Corollary

If $\frac{M_n - b_n}{a_n} \overset{d}{\Longrightarrow} H$ (i.e. $\mathbb{P}(a_n(M_n - b_n) < x) \to G(x)$) where H is an extreme value distribution with distribution function $G(x)$, then for all $k = 1, 2, \dots$,

$$\mathbb{P}(a_n(M_n^{(k)} - b_n) < x) \longrightarrow G(x) \sum_{j=0}^{k-1} \frac{(-\log G(x))^j}{j!} \quad \text{where } G(x) > 0.$$

Proof: $\mathbb{P}(M_n < a_n x + b_n) \to G(x)$, hence by the Poisson approx. prop. $n\bar{F}(a_n x + b_n) \to -\log G(x)$. Corollary follows by the proposition above and the fact $\{M_n^{(k)} < u_n\} = \{S_n < k\}$.

Limit laws for records

$X_1, X_2, \dots$ iid sequence, $M_n = \max(X_1, \dots, X_n)$

$L(1) = 1$ and $L(n+1) = \inf\{k > L(n) : M_k > M_{L(n)}\}$ record times

$\{X_{L(n)}; n \geq 1\} = \{M_{L(n)}; n \geq 1\}$ record values

Let $x_F = \sup\{x : F(x) < 1\}$, $x_F^L = \inf\{x : F(x) > 0\}$; i.e. support of F is $[x_F^L, x_F]$.

and let $R(t) = -\log(1 - F(t))$ the hazard function. $R : [x_F^L, x_F] \to (0,\infty)$

Prop

① If $F(x) = 1 - e^{-x}$, $x > 0$, then $\{X_{L(n)}, n \geq 1\}$ are the points of a homogeneous PPP with intensity 1 on $(0,\infty)$, i.e. $\{X_{L(n)}, n \geq 1\} \overset{d}{=} \{\Gamma_n, n \geq 1\}$ where $\Gamma_n = E_1 + \dots + E_n$ with $E_1, E_2, \dots$ iid, standard $\mathrm{EXP}(1)$ variables.

② Suppose F is continuous. Then $\{X_{L(n)}, n \geq 1\}$ are the points of a PPP on (x_F^L, x_F) with mean measure $R(a,b) = R(b) - R(a)$ for $a, b \in (x_F^L, x_F)$. (# points between a and b $\overset{d}{=} \mathrm{Poi}(R(a,b))$.)

③ If F is continuous, then $\{(X_{L(k)}, L(k+1)-L(k)), k \geq 1\}$ are the points of a 2D PPP on $(x_F^l, x_F) \times \{1,2,\dots\}$ with mean measure $\mu((a,b)\times\{j\}) = \dfrac{F^j(b) - F^j(a)}{j}$

Proof ① $\mathbb{P}(X_{L(n+1)} > y \mid X_{L(n)} = x) = e^{-(y-x)}$ by the memoryless prop. of the EXP(1) distr.

$\mathbb{P}(\Gamma_{n+1} > y \mid \Gamma_n = x) = e^{-(y-x)}$; $X_{L(1)} \stackrel{d}{=} \Gamma_1$ and $\{X_{L(n)}, n \geq 1\} \stackrel{d}{=} \{\Gamma_n, n \geq 1\}$ as Markov Chains

② $R^{-1}(E_1) \stackrel{d}{=} X_1$ where $E_1 \sim EXP(1)$, hence $\{\max_{k=1}^{n} (R^{-1}(E_k)), n \geq 1\} \stackrel{d}{=} \{M_n, n \geq 1\}$

Since R is continuous, R^{-1} is strictly increasing, and on the left-hand side,

$\{\max_{k=1}^{n} (R^{-1}(E_k)), n \geq 1\} \stackrel{d}{=} \{R^{-1}(\max_{k=1}^{n} E_k), n \geq 1\}$

Hence $\{R^{-1}(E_{L(n)}), n \geq 1\} \stackrel{d}{=} \{X_{L(n)}, n \geq 1\}$

The points $\{E_{L(n)}, n \geq 1\}$ form a homogeneous PPP, hence their image under R^{-1} form a PPP with mean measure $\mathrm{Leb} \circ R$, because the average number of points between a and b for $\{R^{-1}(E_{L(n)}), n \geq 1\}$ is $R(b) - R(a)$.

③ $\{X_{L(n)}, n \geq 1\}$ is a PPP, $\underbrace{\{L(n+1)-L(n), n \geq 1\}}_{\text{conditionally on } \{X_{L(n)}, n\geq 1\},}$ are independent of each other and $\mathbb{P}(L(n+1)-L(n) = j \mid \{X_{L(k)}\}) = F(X_{L(n)})^{j-1}(1 - F(X_{L(n)}))$, hence $\{(X_{L(n)}, L(n+1)-L(n)), n \geq 1\}$ forms a PPP with mean measure

$$\mu((a,b)\times\{j\}) = \int_a^b \underbrace{R(dx)}_{\frac{d}{dx} -\log(1-F(x))dx} F(x)^{j-1}(1-F(x)) = \int_a^b F^{j-1}(x)\, dF(x) = \frac{F^j(b) - F^j(a)}{j}.$$

Poisson point process (PPP) on a measurable space (X, μ)

Random points in X s.t. the number of points in disjoint subsets are independent, If $A \subseteq X$, then # points in $A \stackrel{d}{=} POI(\mu(A))$.

Example

$X = \mathbb{R}$, $\mu = \mathrm{Leb}$; this can be obtained as $\{\Gamma_n, n \geq 1\}$ where $\Gamma_n = E_1 + \dots + E_n$ with $E_1, E_2, \dots$ iid. EXP(1).

Lemma

If F is a continuous distr. function, then the relative rank of X_n among $X_1, \dots, X_n$ is uniform on $\{1, 2, \dots, n\}$. In particular, $P(X_n \text{ is a record}) = \frac{1}{n}$.

Proof by symmetry.

$A_n = \{X_n \text{ is a record}\}$ $\quad \mathbb{P}(A_n) = \frac{1}{n}$

Prop $\{A_n\}_{n \geq 1}$ are independent

Proof $n_1 < n_2 < \dots < n_{k+1}$: $\mathbb{P}(A_{n_1} \cap A_{n_2} \cap \dots \cap A_{n_k} \mid A_{n_{k+1}}) = \mathbb{P}(A_{n_1} \cap \dots \cap A_{n_k})$

Let $\mu = \sum_{n=1} \delta_{L(n)}$, i.e. $\mu([1,n]) = \#$ of records in $[1,n]$.

Prop

If F is continuous, then $\dfrac{\mu([1,n])}{\log n} \to 1$ a.s.

We use

Thm (Kolmogorov) see Durrett: Probability: Theory and Examples, Chapter 1, Thm (8.3)

If $X_1, X_2, \ldots$ ~~iid~~ independent with $\mathbb{E}X_n = 0$ and $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$, then $\sum_{n=1}^{\infty} X(\omega)$ converges a.s.

Proof of Prop

$$\mu([1,n]) = \sum_{j=1}^{n} \mathbb{1}_{A_j} \qquad \mathbb{E}\,\mathbb{1}_{A_j} = \frac{1}{j} \qquad \mathrm{Var}(\mathbb{1}_{A_j}) = \frac{1}{j} - \frac{1}{j^2} = \frac{j-1}{j^2}$$

Consider the series $\sum_{j=1}^{\infty} \dfrac{\mathbb{1}_{A_j} - 1/j}{\log j}$. Then $\mathrm{Var}\left(\sum_{j=1}^{\infty} \dfrac{\mathbb{1}_{A_j} - 1/j}{\log j}\right) = \sum_{j=1}^{\infty} \dfrac{j-1}{j^2(\log j)^2} \sim \sum_{j=1}^{\infty} \dfrac{1}{j(\log j)^2} < \infty$

$\Rightarrow$ the random sum $\sum_{j=1}^{\infty} \dfrac{\mathbb{1}_{A_j} - 1/j}{\log j}$ converges a.s.

Kronecker's lemma

see Durrett: Prob. Th. and Ex. Chapter 1, Thm (8.5)

If $a_n \nearrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges, then $\dfrac{\sum_{m=1}^{n} x_m}{a_n} \to 0$

Thm

$$\frac{\mu([1,n]) - \log n}{\sqrt{\log n}} \overset{d}{\Rightarrow} \Phi \text{ standard normal}$$

Proof $\mathrm{Var}\left(\sum_{j=1}^{n} \mathbb{1}_{A_j}\right) = \sum_{j=1}^{n}\left(\frac{1}{j} - \frac{1}{j^2}\right) \sim \log n$, Lindeberg thm applies

By ~~the~~ Kronecker's lemma,

$$\frac{\mu([1,n])}{\log n} - 1 = \frac{\sum_{j=1}^{n} \mathbb{1}_{A_j} - 1/j}{\log n} + o(1) = \underbrace{\sum_{j=1}^{n} \frac{\mathbb{1}_{A_j} - 1/j}{\log j} \cdot \frac{\log j}{\log n}}_{\text{This is already the proof of Kronecker's lemma for this case.}} + o(1) \longrightarrow 0$$

Let $\mu_n = \sum_{j=1}^{\infty} \delta_{L(j)}(n\,\cdot) = \sum_{j=1}^{\infty} \delta_{n^{-1}L(j)}(\cdot)$ be the rescaled point process of records.

Thm

Let μ_∞ be a PPP on $(0,\infty)$ with mean measure $\nu(a,b) = \log b/a$. Then

$$\mu_n \overset{d}{\Rightarrow} \mu_\infty \text{ in the following sense:}$$

Tool: Laplace functional: $\Psi_\mu(f) = \mathbb{E}\exp\left(-\int_{(0,\infty)} f(x)\,d\mu(x)\right)$ for all $f \in C_0^+((0,\infty))$

positive continuous functions with compact support within $(0,\infty)$

Claim: if $\Psi_{\mu_n}(f) \to \Psi_{\mu_\infty}(f)$ $\forall f \in C_0^+((0,\infty))$, then $\mu_n \overset{d}{\Rightarrow} \mu_\infty$.

~~for the records~~:

Proof:

$$\Psi_{\mu_n}(f) = \mathbb{E}\exp\left(-\int_{(0,\infty)} f(x)\,d\mu_n(x)\right) = \mathbb{E}\exp\left(-\sum_{j=1}^{\infty} f\left(\tfrac{j}{n}\right)\mathbb{1}_{A_j}\right)$$

we use independence of $\{A_j\}$

$$= \prod_{j=1}^{\infty} \mathbb{E}\exp\left(-f\left(\tfrac{j}{n}\right)\mathbb{1}_{A_j}\right) = \prod_{j=1}^{\infty}\left(\frac{1}{j}e^{-f(\frac{j}{n})} + 1 - \frac{1}{j}\right) = \prod_{j=1}^{\infty}\left(1 - \frac{1}{j}\left(1 - e^{-f(\frac{j}{n})}\right)\right)$$

$$= \prod_{j=1}^{\infty}(1 - x_{j,n}) \quad \text{where} \quad x_{j,n} = \frac{1}{j}\left(1 - e^{-f(\frac{j}{n})}\right)$$

Two inequalities:

- $\left|\prod_{j=1}^{k} a_j - \prod_{j=1}^{k} b_j\right| \le \sum_{j=1}^{k} |a_j - b_j|$ if $|a_j|, |b_j| \le 1$, $j = 1, \ldots, k$

  Proof: by $a_1a_2 - b_1b_2 = (a_1 - b_1)a_2 + b_1(a_2 - b_2)$ and induct

- $|e^{-x} - 1 + x| \le x^2/2$ if $x \ge 0$

  Proof: $|1 - e^{-x} - x| = \left|\int_0^x (e^{-u} - 1)\,du\right| \le \int_0^x |e^{-u} - 1|\,du \le \int_0^x u\,du$

Hence $\left|\Psi_{\mu_n}(f) - \prod_{j=1}^{\infty} e^{-x_{j,n}}\right| = \left|\prod_{j=1}^{\infty}(1-x_{j,n}) - \prod_{j=1}^{\infty} e^{-x_{j,n}}\right| \le \sum_{j=1}^{\infty}\left|e^{-x_{j,n}} - 1 + x_{j,n}\right|$

$$\le \frac{1}{2}\sum_{j=1}^{\infty}(x_{j,n})^2 \le \frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{(1-e^{-f(\frac{j}{n})})}{j}\right)^2 \le \frac{1}{2}\sum_{j=1}^{\infty}\frac{1}{j^2} < \infty$$

For fixed j, $\lim_{n\to\infty} \frac{1-e^{-f(\frac{j}{n})}}{j} = 0$ since f has compact support, i.e. by dominated convergence, $\left|\Psi_{\mu_n}(f) - \prod_{j=1}^{\infty} e^{-x_{j,n}}\right| \to 0$.

~~With the notation $m_n = \sum_{j=1}^{\infty}\frac{1}{n}\delta_{\frac{j}{n}}$~~

Riemann sum

$$\Psi_{\mu_n}(f) = \prod_{j=1}^{\infty} e^{-x_{j,n}} + o(1) = \exp\left(-\sum_{j=1}^{\infty}\frac{1-e^{-f(\frac{j}{n})}}{j/n}\cdot\frac{1}{n}\right) + o(1)$$

$$\to \exp\left(-\int_0^{\infty}\frac{1-e^{-f(x)}}{x}\,dx\right)$$ since f is bounded and continuous with compact support.

Therefore, $\mu_n \overset{d}{\Rightarrow} \mu_\infty$ by the next proposition.

__Prop__

The Laplace functional of the PPP with mean measure μ is

$$\Psi_\mu(f) = \exp\left(-\int_0^{\infty}(1-e^{-f(x)})\,\mu(dx)\right)$$

__Proof__: Let $f = c\mathbb{1}_F$ where F is a measurable set. $N(F) = \#$ points in F

$$\Psi_\mu(f) = E\exp(-cN(F)) = \sum_{k=0}^{\infty} e^{-\mu(F)}\frac{(e^{-c}\mu(F))^k}{k!} = \exp\left(-(1-e^{-c})\mu(F)\right) = \exp\left(-\int_F(1-e^{-f(x)})\mu(dx)\right)$$

For step functions: by independence.

For general: by monoton convergence.