Extreme Value Theory in Civil Engineering

Baidurya Bhattacharya

Dept of Civil Engineering

IIT Kharagpur

December 2016

Homepage: www.facweb.iitkgp.ernet.in/~baidurya/

Preliminaries: Return period

- IID random variables $\{X_1, X_2, X_3, \ldots\}$ with CDF F_X
  - Occurrence (or success) = $\{X_i > x_p\}$ in i^{th} trial
  - p = P{success} = P $\{X_i > x_p\} = 1 - F_X(x_p)$
  - x_p = level corresponding to exceedance probability p
- Sequence of independent and identical Bernoulli trials:
  - Geometric random variable
  - Time between successive occurences is random
  - "Mean Return Period" associated with x_p is $1/p$ (in units of trial time interval)

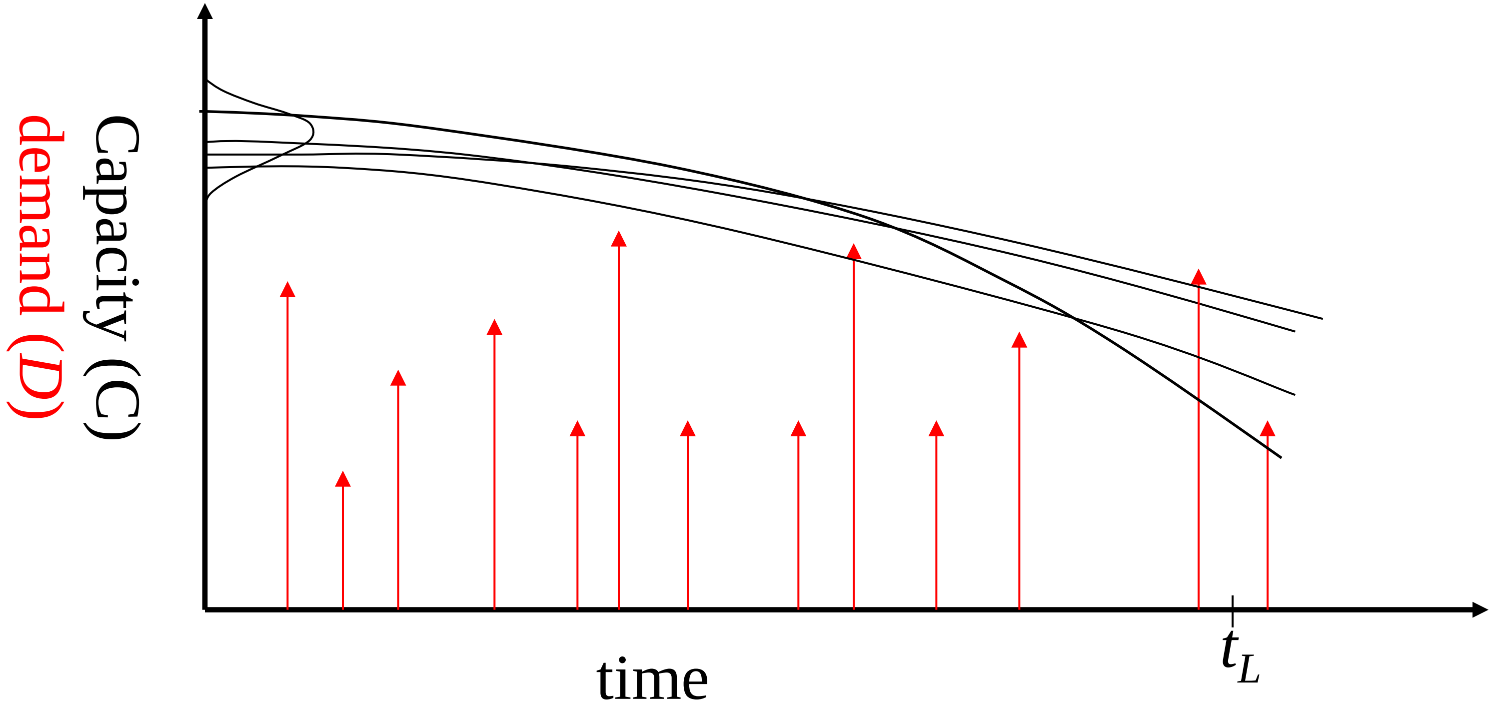
Preliminaries: Characteristic value

- IID random variables $\{X_1, X_2, X_3, \ldots\}$ with CDF F_X
  - Success = $\{X_i > x_p\}$ in i^{th} trial
  - p = P{success} = P $\{X_i > x_p\} = 1 - F_X(x_p)$
  - x_p = level corresponding to exceedance probability p
- n repeated trials – Binomial random variable
  - Mean number of occurrences = np
- x_n = Characteristic value of X_i
  - if mean number of occurrences, $np(x_n) = 1$
  - that is, $1 - F_X(x_n) = 1/n$

Motivation: Time-dependent reliability

First Passage Time: $T = \inf\left[t : C(t, \underline{x}) < D(t, \underline{x}),\ \ t \geq 0, \underline{x} \in \Omega\right]$

Probability of failure: $P_f(t_L) = P[T \leq t_L]$



Design issues

- Maximum load
- Minimum capacity
- Associated uncertainties
- Data driven

Reliability problem statement

- Time-dependent failure probability

$$P_f(t) = P[R(\tau) - D - L(\tau) \leq 0 \quad \text{for any} \quad \tau \in [0,t]]$$

- Simplification

$$P_f(t) = P[R_e - D - L_{\max,t} \leq 0]$$

- Need to estimate $L_{\max,t}$

Maximum live load estimation

- Maximum live load

$$L_{\max,t} = \max\{L_1, L_2, \ldots, L_{N(t)}\}$$

- Distribution function

$$F_{L_{\max,t}}(l) = P[L_1 \le l, L_2 \le l, \ldots, L_{N(t)} \le l]$$

- Simplifications
  - independence
  - stationarity

$$F_{L\max,t}(l) = [F_L(l)]^{N(t)}$$

Extreme value theory: problem statement

- Sequence of random variables $\{X_i\}$
- What are the limiting forms of Z_n and W_n ?

$$Z_n = \max(X_1, X_2, ..., X_n) \sim H_n$$
$$W_n = \min(X_1, X_2, ..., X_n) \sim L_n$$

- Issues
  - n unknown
  - Degeneracy of limit distributions: n infinite
  - Nature of population distributions, F_i
  - Dependence in the sequence
  - Non-stationarity of the sequence

IID (classical) case

- $\{X_i\}$ is an IID sequence
  - X_i and X_j are independent for $i \neq j$
  - $F_i = F$ are same for all i

- $H_n(x)$ and $L_n(x)$ are degenerate distributions

$$P[Z_n \le x] = H_n(x) = F^n(x)$$

$$P[W_n \le x] = L_n(x) = 1-(1-F(x))^n$$

$$\lim_{n\to\infty} H_n(x) = \begin{cases} 0, x < \omega(F) \\ 1, \text{otherwise} \end{cases}$$

$\omega(F)$ = upper end point of F

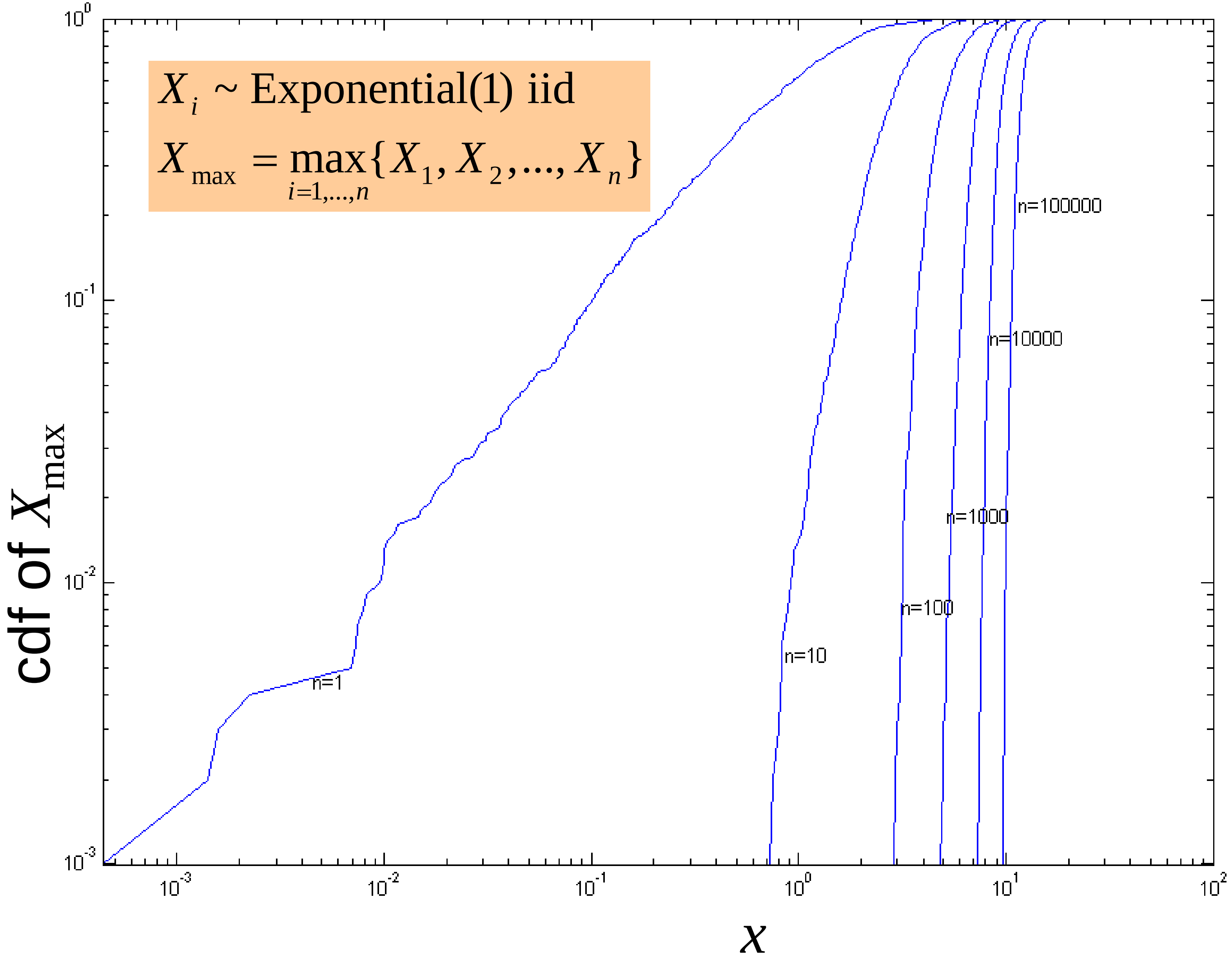
$$\lim_{n\to\infty} L_{(n)}(x) = \begin{cases} 0, x \le \alpha(F) \\ 1, \text{ otherwise} \end{cases}$$

$a(F)$ = lower end point of F

Code for CDF of Xmax vs. n

```
%This program generates a sequence of n IID exponential RVs and stores it maximum, xmax
%It then repeats the process mct times
%The distribution of xmax is plotted
%The plot is repeated for a different n
 clear all;
mct=1000;
    n=input('give n\n');
     for mcti=1:mct,
        for i=1:n,
            x(i)=-log(rand);
        end
        xmax(mcti)=max(x);
        freq(mcti)=mcti/(mct+1);
    end
     xmaxsorted=sort(xmax);
     loglog(xmaxsorted,freq);
    sizes=['n=' num2str(n)];
hold on;
 text(xmaxsorted(mct/2),.5,[sizes],'BackgroundColor',[1 1 1],'EdgeColor','black');
axis([0,100,0,1]);
```

Example: degeneracy for large n



Normalization

- Can we normalize Z_n and help matters?

$$\lim_{n\to\infty} P\left[\frac{Z_n - a_n}{b_n} \le x\right] = \lim_{n\to\infty} H_n(a_n + b_n x) = H(x)$$

- Issues
  - How many possible forms for H ?
  - How does H depend on F ?
  - How to find a_n and b_n
  - What is the speed of convergence?

Normalization

- Can we normalize $W_{(n)}$ and help matters?

$$\lim_{n\to\infty} P\left[\frac{W_{(n)} - a_n}{b_n} \le z\right] = \lim_{n\to\infty} L_{(n)}(a_n + b_n z) = L(z)$$

- Issues
  - How many possible forms for L ?
  - How does L depend on F ?
  - How to find a_n and b_n
  - What is the speed of convergence?

Normalization

As $n \to \infty$,

Instead of $F_{\max}(x) = F_X^n(x)$

Look at $F_X^n(x_n)$ where $x_n = f(x, n)$

Simplest form for x_n :

$$x_n = a_n + b_n x$$

Normalization example

$$X_i \sim \text{Exponential}(1)$$

$$F_{\max}(x) = F_X^n(x) = (1 - \exp(-x))^n$$

As $n \to \infty$,

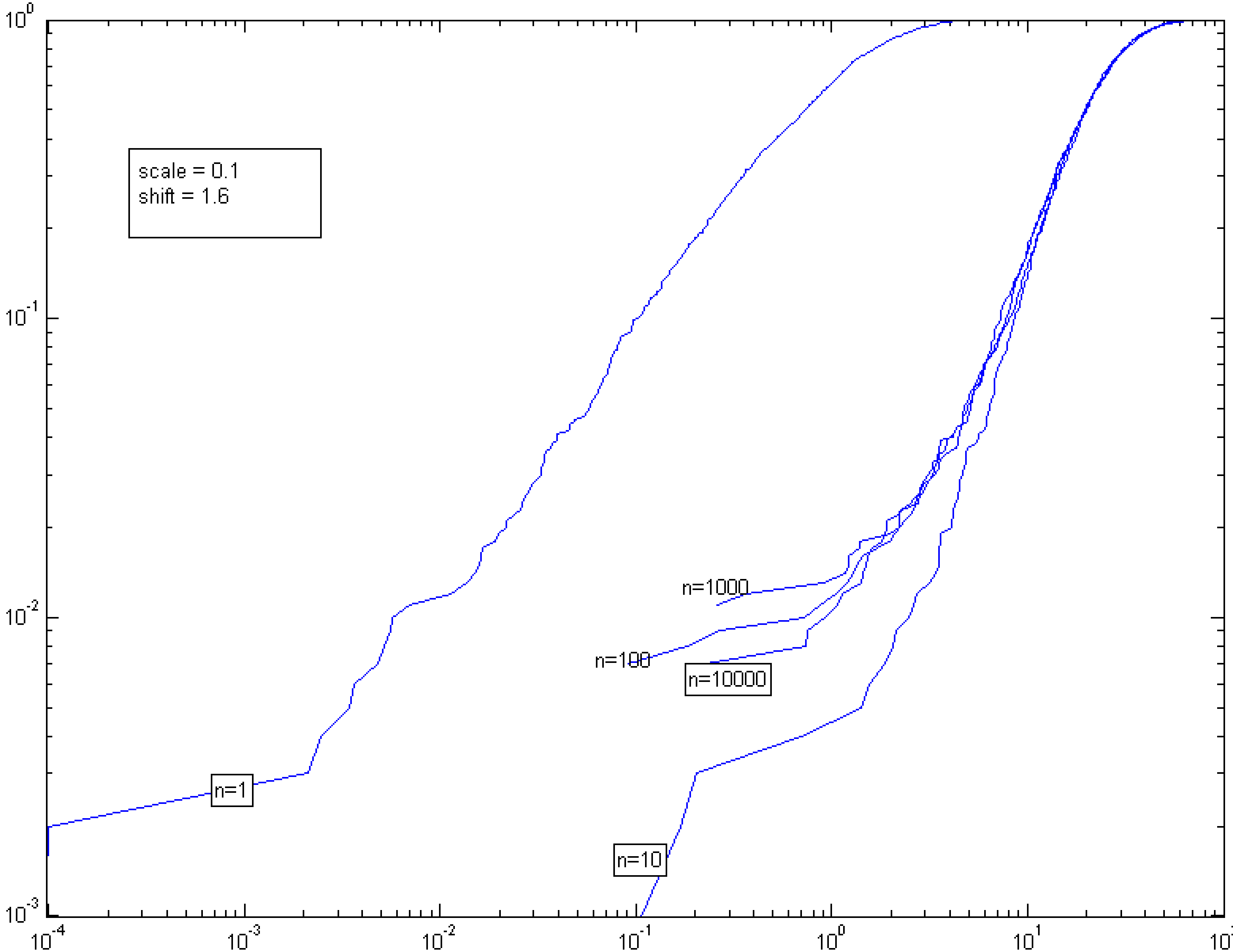
Replace $\quad x \leftarrow \alpha(x-u) + \ln(n)$

$$\text{Obtain: } F_{\max}(x) = \lim_{n\to\infty}\left(1 - \frac{\exp(-\alpha(x-u))}{n}\right)^n$$

$$= \exp(-\exp(-\alpha(x-u)))$$

Code for CDF of normalized Xmax vs. n

```
%This program generates n IID Exponentials and stores the maximum xmax.
%xmax is then recentered and scaled as a function of n
%The process is repeated n times
clear all;
mct=1000;
n=input('give n\n');
scale=input('give scale\n');
shift=input('give shift\n');
if n==1,scale=1;shift=0;end
    for mcti=1:mct,
        for i=1:n,
            x(i)=-log(rand);
        end
        xmax(mcti)=(max(x)-log(n)+shift)/scale;
        freq(mcti)=mcti/(mct+1);
    end
    xmaxsorted=sort(xmax);
    loglog(xmaxsorted,freq);
    sizes=['n=' num2str(n)];
hold on;
text(xmaxsorted(mct/2),.5,[sizes],'BackgroundColor',[1 1 1],'EdgeColor','black');
axis([0,100,0,1]);
```



scale = 0.1
shift = 1.6
n=1
n=10
n=100
n=1000
n=10000
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^0
10^1
10^2

Limit distributions in IID case

- There are only three types of non-degenerate distributions $H(x)$ for maxima
- There are only three types of non-degenerate distributions $L(x)$ for minima
- Necessary and sufficient conditions exist for $F(x)$ to yield above max or min distributions

  – Note: Two distributions F and G are of the same "type", if $F(x) = G(ax+b)$ where a, b are constants

Generalized EV distribution for maxima

In IID case, $H(z)$ must be of the same type as:

$$H_c(z) = \exp\left[-(1+cz)^{-1/c}\right],\ 1+cz>0$$

$c=0 \Rightarrow$ Gumbel (Type I) distribution: $H_G(z) = e^{-e^{-z}}, -\infty < z < \infty$

$c>0 \Rightarrow$ Frechet (Type II) distribution: $H_F(z) = \begin{cases} e^{-z^{-\gamma}}, & z>0 \\ 0 & , z \le 0 \end{cases}$

$c<0 \Rightarrow$ Weibull (Type III) distribution: $H_W(z) = \begin{cases} 1, & z>0 \\ e^{-(-z)^{-\gamma}}, & z \le 0 \end{cases}$

where $\gamma = |1/c|$

Generalized EV distribution for minima

In IID case, $L(z)$ must be of the same type as:

$$L_c(z) = 1 - \exp\left[-(1-cz)^{-1/c}\right],\ 1-cz>0$$

$c = 0 \Rightarrow$ Gumbel (Type I) distribution: $L_G(z) = 1 - e^{-e^z},\ -\infty < z < \infty$

$c > 0 \Rightarrow$ Frechet (Type II) distribution: $L_F(z) = \begin{cases} 1-e^{-(-z)^{-\gamma}}, & z \le 0 \\ 1, & z > 0 \end{cases}$

$c < 0 \Rightarrow$ Weibull (Type III) distribution: $L_W(z) = \begin{cases} 0, & z < 0 \\ 1-e^{-z^{\gamma}}, & z \ge 0 \end{cases}$

where $\gamma = |1/c|$

Domains of attraction for maxima

(A) $\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\gamma}, \qquad \gamma > 0$

(B) $\int_a^{\omega(F)} (1-F(x))dx < \infty$, any finite a, $\alpha(F), \omega(F)$ = lower and upper end points of F

(C) $\lim_{t\to\omega(F)} \frac{1-F(t+xR(t))}{1-F(t)} = e^{-x}, \qquad R(t) = E(X - t \mid X > t), t > \alpha(F)$

- $F \in D(H_F)$ if and only if $\omega(F) = \infty$ and (A) holds for F

  $$a_n = 0, b_n = \inf(x : 1 - F(x) \le 1/n)$$

- $F \in D(H_W)$ if and only if $\omega(F) < \infty$ and (A) holds for $F^*(x) = F(\omega(F) - 1/x)$

  $$a_n = \omega(F), b_n = \omega(F) - \inf(x : 1 - F(x) \le 1/n)$$

- $F \in D(H_G)$ if and only if $\omega(F) = \infty$ and (B), (C) hold

  $$a_n = \inf(x : 1 - F(x) \le 1/n), b_n = R(a_n)$$

Domains of attraction for minima

(A) $\lim_{t\to\infty}\frac{F(tx)}{F(t)} = x^{-\gamma}, \qquad \gamma > 0$

(B) $\int_{\alpha(F)}^{a} F(x)dx < \infty$, any finite a, $\alpha(F), \omega(F)$ = lower and upper end points of F

(C) $\lim_{t\to\alpha(F)}\frac{F(t+xr(t))}{F(t)} = e^{x}, \qquad r(t) = E(t - X \mid X < t), t > \alpha(F)$

- $F \in D(L_F)$ if and only if $\alpha(F) = -\infty$ and (A) holds for F

$$c_n = 0, d_n = \sup(x : F(x) \le 1/n)$$

- $F \in D(L_W)$ if and only if $\alpha(F) > -\infty$ and (A) holds for

$F^*(x) = F(\alpha(F) - 1/x), \ \ x < 0$

$$c_n = \alpha(F), d_n = \sup(x : F(x) \le 1/n) - \alpha(F)$$

- $F \in D(L_G)$ if and only if (B), (C) hold

$$c_n = \sup(x : F(x) \le 1/n), d_n = r(c_n)$$

Examples

- Cauchy
- Uniform
- Exponential
- Rayleigh

Examples

- Limit distribution for minima from Cauchy parent

$$F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}; \quad -\infty < x < \infty$$

Check Eqn (A) for minima:

$$\lim_{t \to -\infty} \frac{F(tx)}{f(t)} = \lim_{t \to -\infty} \frac{\frac{1}{2} + \frac{\arctan(tx)}{\pi}}{\frac{1}{2} + \frac{\arctan(t)}{\pi}}$$

$$= \lim_{t \to -\infty} \frac{\frac{x}{1+(tx)^2}}{\frac{1}{1+t^2}}$$

$$= \lim_{t \to -\infty} \frac{x(1+t^2)}{1+t^2x^2} = x^{-1}$$

That is, $\gamma = -1$, and Cauchy lies in the domain of attraction of Frechet for minima.

The normalizing constants are:

$$c_n = 0$$

$$d_n = \tan\left[\pi\left(\frac{1}{2} - \frac{1}{n}\right)\right]$$

Examples

- Limit distribution for maxima from uniform parent

The complemetary CDF of the uniform is:

$$F^*(x) = 1 - \frac{1}{x},\ x \le 1$$

$$\lim_{t\to\infty} \frac{1-F^*(tx)}{1-F^*(t)} = \lim_{t\to\infty} \frac{\frac{1}{tx}}{\frac{1}{t}} = x^{-1}$$

Since $\gamma = 1$, uniform gives rise to Weibull maxima.

The normalizing constants are:

$$a_n = \omega(F) = 1$$

$$b_n = \omega(F) - F^{-1}\left(1 - \frac{1}{n}\right)$$

$$= 1 - 1 + \frac{1}{n} = \frac{1}{n}$$

Examples

- Limit distribution for minima from exponential parent

$$F^*(x) = 1 - \exp\left(\frac{1}{x}\right)$$

$$\lim_{t\to-\infty} \frac{F^*(tx)}{F^*(t)} = \lim_{t\to-\infty} \frac{1-\exp\left(\frac{1}{tx}\right)}{1-\exp\left(\frac{1}{t}\right)} = \lim_{t\to-\infty} \frac{\exp\left(\frac{1}{tx}\right)\left(\frac{1}{t^2}\right)\left(\frac{1}{x}\right)}{\exp\left(\frac{1}{t}\right)\left(\frac{1}{t^2}\right)} = x^{-1}$$

Since $\gamma = 1$, exponential gives rise to Weibull minima.

$$c_n = \alpha(F) = 0$$

$$d_n = F^{-1}\left(\frac{1}{n}\right) - \alpha(F) = -\log\left(1 - \frac{1}{n}\right) \approx \frac{1}{n}$$

Domains of attraction

Domain of Attraction Type		
Distribution	For maximum	For minimum
Normal	Gumbel	Gumbel
Exponential	Gumbel	Weibull
Log-normal	Gumbel	Gumbel
Gamma	Gumbel	Weibull
Gumbel $_M$	Gumbel	Gumbel
Gumbel $_m$	Gumbel	Gumbel
Rayleigh	Gumbel	Weibull
Uniform	Weibull	Weibull
Weibull $_M$	Weibull	Gumbel
Weibull $_m$	Gumbel	Weibull
Cauchy	Frechet	Frechet
Pareto	Frechet	Weibull
Frechet $_M$	Frechet	Gumbel
Frechet $_m$	Gumbel	Frechet

M=for maximum

m=for minimum

Estimation of EV distribution parameters

- Block maxima – probability plot
- Return period plot
- Problem: not all data are utilized

Generalized Pareto Distribution

Exceedances of *X* over high threshold *u*

Define: *Y* = *X* - *u*

$$G(y) = P[Y \le y \mid Y > 0] = 1 - [1 + (cy/a)]^{-1/c}$$

$$a > 0,\ 1 + (cy/a) > 0$$

same *c* as in GEV distribution