d) Application: Lévy RV's with $\alpha < 1$

Recall that for Lévy RV's with $\alpha < 1$ the typical value of

$$S_N = \sum_{i=1}^{N} X_i \sim N^{1/\alpha} \gg N$$

On the other hand we have now shown that

$$X_{max} = \max\{X_1, .., X_N\} \sim N^{1/\alpha}$$

$\Rightarrow$ typically $X_{max}/S_N = O(1)$, i.e. the largest RV is of the same order as the sum.

To quantify this statement consider the participation ratios

$$W_N^i = \frac{X_i}{S_N} \in [0,1]$$ identically distr. but not independent!

then it can be shown that the moments $\langle W_N^k \rangle$, $k \geq 2$, have a finite limit for $N \to \infty$:

$$\lim_{N \to \infty} \langle W_N^k \rangle = \frac{\Gamma(k-\alpha)}{\Gamma(k)\Gamma(1-\alpha)}$$ $0 < \alpha < 1$, $k \geq 2$

(Derrida, 1997)

$\Rightarrow$ Problem

Lorentz center
www.lorentzcenter.nl

Also, note that

$$\sum_{i=1}^{N} W_i^k \leq W_{max}^{k-1}$$ Proof: Problems

and hence

$$\langle W_{max}^{k-1} \rangle \geq \langle W_N^k \rangle > 0 \quad \text{for } N \to \infty.$$

It follows that $W_{max} > 0$ with nonzero probability for $N \to \infty$

---

e) The peaks-over-threshold approach

For applications it is inconvenient to estimate the extreme value parameters from X_{max}, since one has only a single observation per sample. Here we consider instead all events above some threshold u, and study the limit $u \to s_+$, where s_+ is the upper limit of the support of $F(x)$ ($s_+ = \infty$ for unbounded RV's)

Lorentz center
www.lorentzcenter.nl

Define the excess probability

$$P_u(x) := \text{Prob}[X > u + x \mid X > u] =$$

$$= \frac{1 - F(u+x)}{1 - F(u)} \quad , \quad x \geq 0$$

so that $1 - P_u$ is a distribution function on $[0, s_+]$. Consider some examples:

(i) exponential: $F(x) = 1 - e^{-x} \Rightarrow P_u(x) = e^{-x}$

independent of u

(ii) Gaussian: $1 - F(x) \simeq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \to \infty$

$$\Rightarrow P_u(x) = \frac{u}{u+x} \exp\left[-\frac{1}{2}(x+u)^2 - \frac{1}{2}u^2 \right] =$$

$$= \frac{u}{u+x} \exp\left[-\frac{1}{2}x^2 - ux \right]$$

so $\lim_{u \to \infty} P_u(x) = 0$ for any x. To get a non-degenerate limit we define $y = ux$

Lorentz center
www.lorentzcenter.nl

$$\Rightarrow \quad P_u(y) = \frac{u^2}{u^2+y} \exp\left[-\frac{y^2}{u^2} - y\right] \xrightarrow[u\to\infty]{} \underline{e^{-y}}$$

(iii) <u>Pareto:</u> $F(x) = 1 - x^{-\alpha}$, $x \geq 1$

$$\Rightarrow \quad P_u(x) = \left(\frac{u}{u+x}\right)^{\alpha} \xrightarrow[u\to\infty]{} 1$$

In this case the appropriate rescaling is $\underline{y = x/u}$

$$\Rightarrow \quad P_u(y) = \left(\frac{u}{u+uy}\right)^{\alpha} = \left(\frac{1}{1+y}\right)^{\alpha} \quad \text{indep. of } u.$$

(iv) <u>Uniform:</u> $F(x) = x$, $x \in [0,1]$

$$\Rightarrow \quad P_u = \frac{1-(u+x)}{1-u} = 1 - \frac{x}{1-u}$$

which should converge for $u \to 1$. Clearly the rescaling is $y = x/1-u \Rightarrow P_u(y) = 1-y$.

These examples illustrate the limit theorem of Balkema/de Haan[1] (1974) and Pickands (1975):

[1] "Residual life time at great age"

Lorentz center
www.lorentzcenter.nl

Suppose a function $a(u)$ exists such that

$$\lim_{u \to s_+} P_u(a(u)y) = H(y)$$

exists and is non-degenerate. Then H belongs to the one-parameter family of generalized Pareto distributions (GPD)

$$H_\gamma(y) = (1+\gamma y)^{-1/\gamma}, \quad 1+\gamma y > 0$$

and $H_0(y) = e^{-y}$.

Note that $H_\gamma = -\ln G_\gamma$, and the domains of attraction are the same in both cases.

Application: Fitness distributions in experimental evolution (Rokyta et al., 2008) based on Orr (2003)

Lorentz center
www.lorentzcenter.nl

5° Theory of records

Let X_i be a sequence of continuous, i.i.d. RV.

X_k is a (upper) record if $X_k = \max(X_1, \ldots, X_k)$.

By symmetry

$$\text{Prob}[X_k = \max(X_1, \ldots, X_k)] = \frac{1}{k}$$

$\Rightarrow$ number of records $N(n)$ up to time n:

$$\langle N(n) \rangle = \sum_{k=1}^{n} \frac{1}{k} \simeq \ln(n) + 0.5772\ldots$$

Ex: Weather records (Glick)
For $n = 100$, we have $\langle N(n) \rangle \simeq 5.2$

---

a) The distribution of the number of records

(H. Glick, Amer. Math. Monthly 85, 2 (1978))

Rényi representation: (1962)

Define RV's Y_k through

$$Y_k = \begin{cases} 1 & X_k \text{ is a record} \\ 0 & \text{else} \end{cases}$$

then clearly $\langle Y_k \rangle = \frac{1}{k}$.

Correlations: let $j > i$, then

$$\langle Y_i Y_j \rangle = \text{Prob}[\, X_i = \max(X_1, \dots, X_i) \text{ and } X_j = \max(X_1, \dots, X_j)\,] =$$

(split into independent events)

$$= \text{Prob}[\, X_i = \max(X_1, \dots, X_i)\,] \times$$

$$\times \text{Prob}[\, X_j = \max(X_{i+1}, \dots, X_j)\,] \times$$

$$\times \text{Prob}[\, \underbrace{\max(X_1, \dots, X_i) < \max(X_{i+1}, \dots, X_j)}_{\hat{=}\ \max(X_1, \dots, X_j) \text{ occurs in } \{X_{i+1}, \dots, X_j\}} \,]$$

$$= \frac{1}{i} \cdot \frac{1}{j-i} \cdot \frac{j-i}{j} = \frac{1}{i} \cdot \frac{1}{j} = \langle Y_i \rangle \langle Y_j \rangle$$

This allows to compute the variance of

$$H(n) = \sum_{i=1}^{n} Y_i$$

$$\langle H(n)^2 \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle Y_i Y_j \rangle = \sum_{i=1}^{n} \frac{1}{i} + \langle H(n) \rangle^2 - \sum_{i=1}^{n} \left(\frac{1}{i} \right)^2$$

$$\Rightarrow \langle H(n)^2 \rangle - \langle H(n) \rangle^2 = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i^2} \right)$$

$$\xrightarrow{n \to \infty} \ln(n) + 0.5772 - \frac{\pi^2}{6} = \ln(n) - 1.07$$

$$\Rightarrow \lim_{n\to\infty} \frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle} = 1 \qquad \text{Poisson-like}$$

In the same way it can be shown that the Y_k are independent, and asymptotically

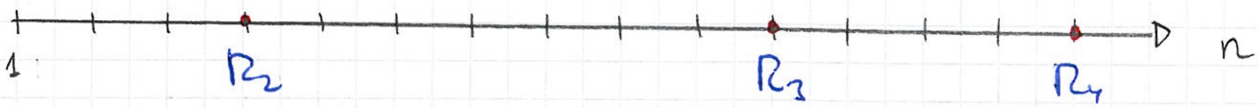
$$\text{Prob}\left[N(n) = M \right] \to \frac{1}{n} \frac{(\ln(n))^M}{M!}$$

log-Poisson distribution

---

b) Record times

Denote by R_k the "time" of the k-th record, with $R_1 = 1$ by convention.



$\Rightarrow$ a "point process" on $\mathbb{N}$

(i) Time of the second record:

$$\text{Prob}\left[R_2 > n \right] = \text{Prob}\left[X_1 = \max(X_1, \ldots, X_n) \right] =$$

$$= 1/n$$

$$\Rightarrow \text{Prob}\left[R_2 = n \right] = \text{Prob}\left[R_2 > n-1 \right] -$$

$$- \text{Prob}\left[R_2 > n \right] = \frac{1}{n-1} - \frac{1}{n} = \frac{1}{n(n-1)}, \quad n \geq 2$$

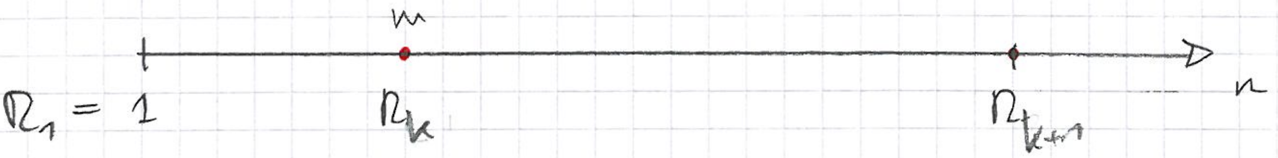
$$\sim 1/n^2, \quad n \to \infty$$

$\Rightarrow$ R_2 has a fat tail with $\alpha = 1$.

In particular, $\langle R_2 \rangle = \infty$. $\Rightarrow$ $\langle R_k \rangle = \infty$ $\forall k$.

(ii) Successive record times

Suppose the record time R_k is given. Where does record $k+1$ occur?



$$\mathrm{Prob}[R_k / R_{k+1} < \tfrac{m}{n} \mid R_k = m] =$$

$$= \mathrm{Prob}[R_{k+1} > n \mid R_k = m] =$$

$$= \mathrm{Prob}[\max\{X_1, .., X_n\} = \max\{X_1, .., X_m\}] =$$

$= m/n$ by symmetry

$\Rightarrow$ $R_k / R_{k+1} \simeq$ uniform RV on $[0, 1]$.

Theorem: (Shorrock, 1972; Resnick, 1973)
The ratios R_k / R_{k+1}, R_{k+1}/R_{k+2}, ... are asymptotically independent uniform RV's.

$$\Rightarrow \quad R_k \simeq \prod_{j=2}^{k} (1/\xi_j), \quad \xi_j \text{ uniform RV}$$

$$\Rightarrow \quad \ln R_k \simeq \sum_{j=2}^{k} \ln(1/\xi_j) = \sum_{j=2}^{k} z_j, \quad \xi_j = e^{-z_j}$$

The distribution of z_j is

$$P(z_j) = \left| \frac{d\xi_j}{dz_j} \right| = e^{-z_j}$$

$\Rightarrow$ z_j are independent exponential RV's w. mean 1

$\Rightarrow$ $\langle \ln R_k \rangle \simeq k$

$\Rightarrow$ $R_k \sim e^k$ typically

<u>Theorem</u> (Rényi, 1962)

$$\lim_{k\to\infty} (\ln R_k / k) = 1 \quad \text{with probability } 1$$

(iii) <u>Conditional expectations</u>

What is the statistics of R_k, given R_{k+1} (backward in "time") or R_{k-1} (forward in "time")?

The representation (*) implies

$$\langle R_k \mid R_{k+1} \rangle = \langle \xi_{k+1} \rangle \cdot R_{k+1} = \frac{1}{2} R_{k+1}$$

but

$$\langle R_k \mid R_{k-1} \rangle = \langle 1/\xi_k \rangle \cdot R_{k-1} = \infty$$

More precisely:

$$\mathrm{Prob}\,[R_k > n \mid R_{k-1}] = \mathrm{Prob}\,[1/\xi_k > n/R_{k-1}]$$

$$= \mathrm{Prob}\,[\xi_k < R_{k-1}/n] = \frac{R_{k-1}}{n}$$

$$\Rightarrow \quad \mathrm{Prob}[\Omega_k = n \mid \Omega_{k-1}] \simeq \frac{\Omega_{k-1}}{n^2}$$

the same qualitative behavior as for $k=2$.