Stefan Grosskinsky, Emma Southall 05.12.2019

Stochastic Modelling and Random Processes

Hand-out 6

Heavy tails, extreme value statistics

A positive random variable X with CDF F is said to have a **power-law tail** with **power** $\alpha > 0$, if

$$\overline{F}(x) := \mathbb{P}[X > x] \simeq Cx^{-\alpha} \quad \text{as } x \to \infty\ .$$

The simplest example is the **Pareto distribution** with **scale** parameter $x_m > 0$ and **power** $\alpha > 0$,

$$\text{where} \quad X \sim \text{Pareto}(x_m, \alpha) \quad \text{and} \quad \overline{F}(x) = (x_m/x)^\alpha \quad \text{for } x \geq x_m\ .$$

We have $\mathbb{E}(X) = \alpha x_m/(\alpha - 1)$ if $\alpha > 1$ and $\text{Var}(X) = \frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ if $\alpha > 2$, otherwise ∞.

More generally, X is said to have a **heavy tail** if $\frac{1}{x}\log \overline{F}(x) \to 0$ as $x \to \infty$,
which includes also Log-Normal or stretched exponential tails (and many more).

Power-laws are also called **scale-free distributions** since the power does not change under scaling, e.g. for Pareto distributions we have

$$X \sim \text{Pareto}(x_m, \alpha) \quad \text{then} \quad \lambda X \sim \text{Pareto}(\lambda x_m, \alpha) \quad \text{for } \lambda > 0\ .$$

In contrast, for $X \sim \text{Exp}(\alpha)$ has scale $\mathbb{E}[X] = 1/\alpha$ and we have $\lambda X \sim \text{Exp}(\alpha/\lambda)$.
This property is relevant in **critical phenomena in statistical mechanics**, where systems exhibit scale free distributions at points of phase transitions. Power law degree distributions in complex networks can emerge from preferential attachment-type dynamics, which is often used as an explanation for the abundance of power-law distributed observables in social or other types of networks.

A second important example are **α-stable Lévy distributions** L_α, $\alpha \in (0, 2]$, which in the symmetric case have characteristic functions

$$\chi_\alpha(t) = e^{-|c\,t|^\alpha} \quad \text{where the scale} \quad c > 0 \quad \text{determines the width}\ .$$

Note that for $\alpha = 2$ this corresponds to the centred Gaussian, and for $\alpha = 1$ it is known as the **Cauchy distribution** with PDF $f_1(x) = \frac{1}{\pi}\frac{c}{c^2+x^2}$.
In general, symmetric L_α distributions have power-law tails $\overline{F}_\alpha(x) \propto c/|x|^\alpha$ as $|x| \to \infty$.
They are the limit laws for general heavy-tailed distributions with diverging mean and/or variance.

Theorem. Generalized LLN and CLT

Let $X_1, X_2 \ldots$ be iid random variables with symmetric power-law tail $\mathbb{P}(|X_i| \geq x) \propto x^{-\alpha}$ with $\alpha \in (0, 2)$. For $S_n = \sum_{i=1}^n X_i$ we have for $\alpha \in (0, 1)$

$$\frac{1}{n}S_n \quad \text{does not converge, but} \quad \frac{1}{n^{1/\alpha}}S_n \quad \text{converges to a r.v.} \qquad \text{(modified LLN)}\ .$$

If $\alpha \in (1, 2)$, $\mathbb{E}(|X_i|) < \infty$ and we have convergence in distribution

$$\frac{1}{n}S_n \xrightarrow{D} \mu = \mathbb{E}(X_i) \qquad \text{(usual LLN)}\ ,$$

$$\frac{1}{n^{1/\alpha}}(S_n - n\mu) \xrightarrow{D} L_\alpha \qquad \text{(generalized CLT)}\ .$$

The proof follows the same idea as the usual CLT with $|t|^\alpha$ being the leading order term in the expansion of the characteristic function.

Extreme value statistics

Consider a sequence of iid random variables $X_1, X_2, \ldots$ in $\mathbb{R}$ with CDF F, and let

$$M_n = \max\{X_1, \ldots, X_n\} \quad \text{be the maximum of the first } n \text{ variables .}$$

Similarly to the CLT, the distribution of M_n converges to a (rather) universal limit distribution.

Extreme value theorem - EVT (Fisher-Tippet Gnedenko)
If there exist normalizing sequences $a_n, b_n \in \mathbb{R}$ such that $\mathbb{P}\big(\frac{M_n - b_n}{a_n} \leq x\big)$ converges to a non-degenerate CDF $G(x)$ as $n \to \infty$, then G is a **generalized extreme value distribution**

$$G(x) = \exp\bigg(- \Big(1 + k\Big(\frac{x-\mu}{\sigma}\Big)\Big)^{-1/k}\bigg) \tag{1}$$

with parameters for **location** $\mu \in \mathbb{R}$, **scale** $\sigma > 0$ and **shape** $k \in \mathbb{R}$.

The sequences and parameter values of G are related to the tail $\overline{F}$ (see e.g. [1] for details). Depending on the shape parameter k, one typically distinguishes the following three **standardized classes**:

- **Gumbel** (Type I): $k = 0$ and $G_{\text{I}}(x) = \exp\big(-e^{-x}\big)$
  limit if $\overline{F}$ has exponential tail (including actual exponential or Gaussian rv's)
- **Fréchet** (Type II): $k = 1/\alpha > 0$ and $G_{\text{II}}(x) = \begin{cases} 0 & , \ x \leq 0 \\ \exp(-x^{-\alpha}) & , \ x > 0 \end{cases}$
  limit if $\overline{F}$ has heavy tail (including e.g. power laws)
- **Weibull** (Type III): $k = -1/\alpha < 0$ and $G_{\text{III}}(x) = \begin{cases} \exp(-(-x)^{\alpha}) & , \ x < 0 \\ 1 & , \ x \geq 0 \end{cases}$
  limit if $\bar{F}$ has light tail (including bounded support such as uniform rv's)

Asymptotic tails as $x \to \infty$ of Gumbel and Fréchet are given as

$$\overline{G}_{\text{I}}(x) \simeq e^{-x} \quad \text{and} \quad \overline{G}_{\text{II}}(x) \simeq x^{-\alpha} .$$

The typical scaling (location) s_n of $\mathbb{E}(M_n)$ as a function of n can be determined relatively easily. Note that for iid random variables we have

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = \mathbb{P}(X_1 \leq x)^n = (F(x))^n .$$

Now s_n is determined by requiring that $(F(s_n))^n$ has a non-degenerate limit in $(0,1)$ as $n \to \infty$, so

$$(F(s_n))^n = \big(1 - \overline{F}(s_n)\big)^n \to e^{-c} , \ c > 0 \quad \text{which implies} \quad \overline{F}(s_n) \simeq c/n .$$

- For **exponential** $X_i \sim \text{Exp}(\lambda)$ iid random variables with tail $\bar{F}(s_n) = e^{-\lambda s_n}$ this leads to

  $$s_n \simeq (\log n - \log c)/\lambda \quad \text{which implies} \quad M_n = (\log n + \xi_n)/\lambda$$

  where ξ_n is a random variable that converges as $n \to \infty$. This implies that we may choose $b_n = \log n$ and $a_n = \lambda$ in EVT with convergence to **Gumbel**.

- For $X_i \sim \mathbf{Pareto}(x_m, \alpha)$ iid we get $s_n \simeq x_m (n/c)^{1/\alpha}$, so that $M_n = x_m n^{1/\alpha} \xi_n$ with multiplicative randomness, implying $b_n = 0$ and $a_n = x_m n^{1/\alpha}$ as a valid normalization with convergence to **Fréchet** with parameter α.
  Note that for $\alpha \in (0,1)$ with infinite mean, this implies $M_n \propto X_n \gg n$ so the sum is dominated by the largest contributions, whereas for $\alpha > 1$ we have $M_n \ll S_n \propto n$.

- For **uniform** $X_i \sim U([0,1))$ iid we expect $M_n \to 1$ as $n \to \infty$, and with $\overline{F}(x) = 1 - x$ we get $s_n \simeq 1 - c/n$ so that we can choose $b_n = 1$ and $a_n = 1/n$ with convergence to **Weibull**.

[1] J-P. Bouchaud, M. Potters, Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management, CUP 2003 2nd edition, Chapter 2

Statistics of records

Consider iid continuous random variables $X_1, X_2, \ldots$ taking values in a connected set $S \subseteq \mathbb{R}$ (e.g. $S = [0,1)$ or $S = \mathbb{R}$) with distribution function F. Define the indicators of **record events**

$$I_n := \mathbb{1}(X_n \text{ is a record}) = \begin{cases} 1 & , \text{ if } M_n = X_n,\ M_{n-1} < X_n \\ 0 & , \text{ otherwise} \end{cases} .$$

- Since the rank order of iidrv's is uniform independently of F, the record probability is

$$\mathbb{P}[X_n \text{ record}] = \mathbb{E}[I_n] = \frac{(n-1)!}{n!} = \frac{1}{n} \quad \text{and} \quad I_n \sim \text{Be}(1/n) \quad \text{with } I_1 = 1 .$$

This implies that the **number of records up to time** n,

$$R_n := \sum_{k=1}^{n} I_n \in \{1, \ldots, n\} \quad \text{has expectation} \quad \mathbb{E}[R_n] = \sum_{k=1}^{n} \frac{1}{k} \simeq \log n + \gamma + O(1/n)$$

as $n \to \infty$ with Euler constant $\gamma = 0.57721\ldots$.

- I_{n+1} and M_{n+1} depend only on M_n and X_{n+1}, and are independent of the rank order of $X_1, \ldots, X_n$ and therefore of $I_1, \ldots, I_n$. Therefore

$$\text{Var}[R_n] = \sum_{k=1}^{n} \text{Var}[I_k] = \sum_{k=1}^{n} \frac{1}{k}\Big(1 - \frac{1}{k}\Big) \simeq \log n + \gamma - \pi^2/6 + O(1/n) .$$

So $\text{STD}[R_n]/\mathbb{E}[R_n] \simeq 1/\sqrt{\log n} \to 0$ and $R_n \to \infty$ with probability 1 as $n \to \infty$.

- We can compute the **probability generating function** for $s \in [0,1]$

$$\mathbb{E}\big[s^{R_n}\big] = \mathbb{E}\big[s^{\sum_k I_k}\big] = \prod_{k=1}^{n} \mathbb{E}\big[s^{I_k}\big] = \prod_{k=1}^{n} \Big(s\frac{1}{k} + \frac{k-1}{k}\Big) = \frac{1}{n!} \prod_{k=0}^{n-1} (k+s) = \frac{\Gamma(s+n)}{\Gamma(s)\Gamma(n+1)} .$$

Then we can use Stirling's formula for asymptotics of the Gamma function to get as $n \to \infty$

$$G(s) := \mathbb{E}\big[s^{R_n - 1}\big] \simeq \frac{1}{s\Gamma(s)} n^{s-1} = \frac{1}{\Gamma(s+1)} n^{s-1} \approx e^{\log n (s-1)} ,$$

since $\Gamma(1) = \Gamma(2) = 1$ and $\Gamma(s+1)$ is close to 1 for $s \in [0,1]$. Recall that the generating function of $Y \sim \text{Poi}(\lambda)$ is $\quad G_Y(s) = \mathbb{E}\big[s^Y\big] = \sum_{k=0}^{n} (\lambda s)^k e^{-\lambda}/k! = e^{\lambda(s-1)}$.
So for large n, $R_n - 1 \in \mathbb{N}_0$ is approximately **Poisson distributed with mean** $\log n$. In particular, with probability $G(0) = 1/n$ we have $R_n = 1$ with no further record except $R_1 = 1$ until time n.

- Set $T_1 = 1$ and let $T_{i+1} = \min\big\{m > T_i : X_m \text{ is record}\big\} \in \mathbb{N}$, defining the sequence of **record times** where $T_i \uparrow \infty$ as $i \to \infty$ by definition. Since

$$\{R_n < i\} = \{T_i > n\} \quad \text{we have} \quad \mathbb{P}[T_i > n] \geq \mathbb{P}[R_n = 1] = 1/n$$

for all $i > 1$, so record times are heavy-tailed with $\mathbb{E}[T_i] = \infty$.
The **time to the next record** $T_{i+1} - T_i$ only depends on the record value X_{T_i}, and since $X_{T_1} < X_{T_2} < \ldots$ we have

$$\mathbb{P}\big[T_{i+1} - T_i > n \big| T_i = m\big] \geq \mathbb{P}\big[T_2 - 1 > n\big] \geq 1/n$$

for all $i \geq 1$ and $m \geq i$. In particular $\mathbb{E}\big[T_{i+1} - T_i \big| T_i = m\big] = \infty$.
On the other hand, given $T_i = m$ the **previous record time** T_{i-1} is uniformly distributed in $\{i-1, \ldots, m-1\}$ (modulo boundary effects close to $i-1$). In general, one can show that ratios of record times converge in distribution to a uniform random variable

$$T_{i-1}/T_i \xrightarrow{D} U\big([0,1)\big) \quad \text{as } i \to \infty .$$

Note that NONE of the above depends on the actual distribution F of the X_i!