An introduction to extreme value statistics
(part 1/2)

Nicholas R. Moloney

LONDON MATHEMATICAL LABORATORY

Dresden
October 05, 2015

Part 1 – iid/weakly correlated variables

- Maxima, exceedances.
- Iid theory, extreme value distributions.
- Point processes, Poisson approximation.
- Extensions to weakly correlated stationary sequences.
- Modeling examples.

Part 2 – examples from statistical physics

- Remarks on applications statistical physics.
- DNA replication times.
- Order parameter in percolation.
- Minimum path in tree-like structures.
- Maxima of some Gaussian processes.
- Integer partitions and the ideal Bose gas.

References

S. Coles — Introduction
An Introduction to Statistical Modeling of Extreme Values
Springer (2001)

P. Embrechts, C. Klüppelberg, T. Mikosch — Theory/modeling
Modelling Extremal Events for Insurance and Finance
Springer (1997)

M. R. Leadbetter, G. Lindgren, H. Rootzén — Theory
Extremes and Related Properties of Random Sequences and Processes
Springer (1983)

M. R. Leadbetter, H. Rootzén — Theory
Extremal theory for stochastic processes
Ann. Probab. **16**, 431 (1988)

References (in the physics literature)

J.-P. Bouchaud, M. Potters — Introduction
Theory of Financial Risks
Cambridge University Press (2000)

D. Sornette — Introduction
Critical Phenomena in Natural Sciences
Springer (2006)

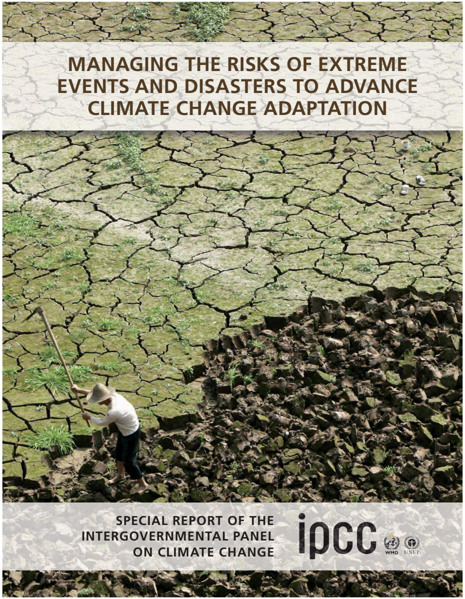
M. Clusel, E. Bertin — Review
Int. J. Mod. Phys. B **22**, 3311 (2008)

S. N. Majumdar, A. Pal — Lecture notes
arXiv:1406.6768v3 (2014)

Outline

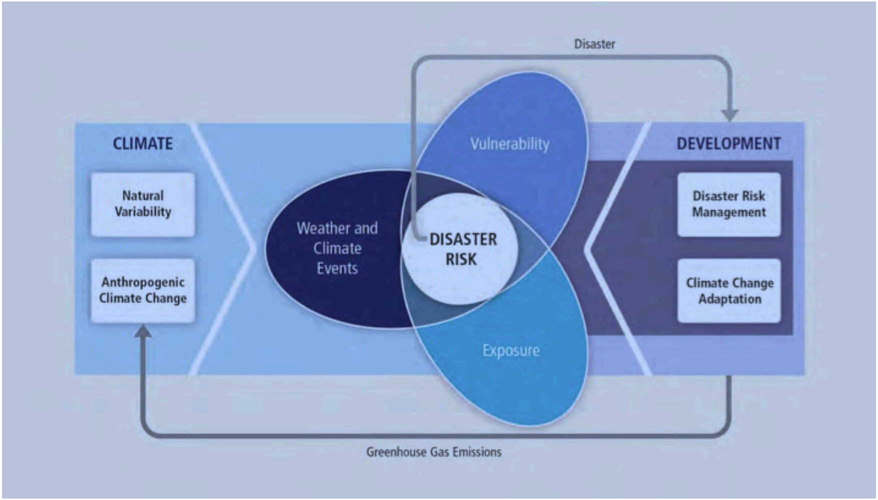
EVS and Climate Change

IPCC 2012



EVS and Climate Change

"Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation"



15th century IPCC

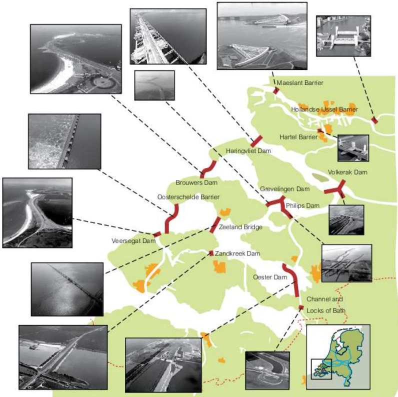
Behringer 1999; Summers 2003

Figure 3. Witches cause a hailstorm. Illustration from the title page of Ulrich Molitoris' *De laniis et phitonicis mulieribus* [Concerning Witches and Sorceresses], Cologne, 1489.

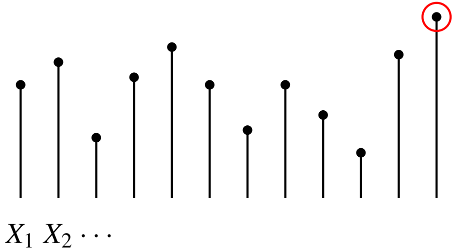
"...in some parts of Northern Germany, as well as in the provinces, townships, territories, districts, and dioceses of Mainz, Cologne,Tréves, Salzburg, and Bremen, many persons... by their incantations, spells, conjurations, and other accursed charms and crafts, enormities and horrid offences, ...have blasted the produce of the earth, the grapes of the vine, the fruits of the trees,... vineyards, orchards, meadows, pasture-land, corn, wheat, and all other cereals..."

(Pope Innocent VIII, 1484)

Deltawerken



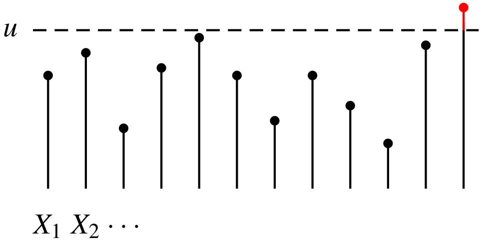
Outline



$$\mathbb{P}(\max\{X_1, \ldots, X_n\} \leq x)$$

GEV

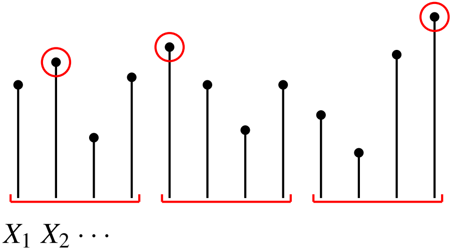
(Generalised Extreme Value)



$$\mathbb{P}(X \leq u + y | X > u)$$

GPD

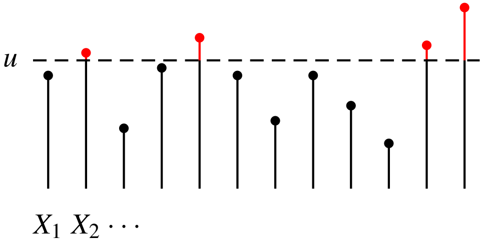
(Generalised Pareto)



$$\mathbb{P}(\max\{X_1, \ldots, X_n\} \leq x)$$

GEV

(Generalised Extreme Value)



$$\mathbb{P}(X \leq u + y | X > u)$$

GPD

(Generalised Pareto)

Statement of the problem

Draw n independent and identically distributed random variables $X_1, \ldots, X_n$ from a parent distribution $F(x)$.

What is the distribution of the maximum

$$M_n = \max\{X_1, \ldots, X_n\} \quad ?$$

For iid random variables

$$\begin{aligned}\mathbb{P}(M_n \le x) &= \mathbb{P}(X_1 \le x, \ldots, X_n \le x) \\ &= \mathbb{P}(X_1 \le x) \cdots \mathbb{P}(X_n \le x) \\ &= F^n(x).\end{aligned}$$

Limit distribution for maximum

Affine rescaling

E.g. for uniform distribution $X_i \sim U(0,1)$

$$F^n(x) = x^n \xrightarrow{n\to\infty} \begin{cases} 1 & x = 1 \\ 0 & x < 1. \end{cases}$$

Can we do better than 0-1 behaviour?

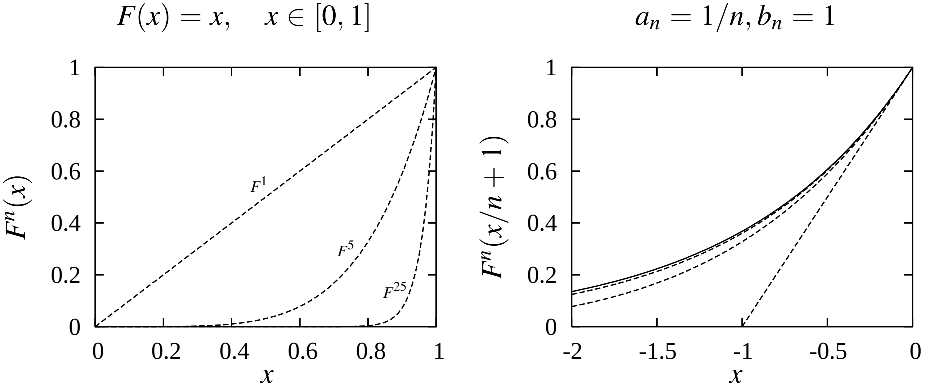
Choose a_n (scale) and b_n (location) such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$$

is a non-degenerate limiting distribution.

Example: uniform distribution

a_n (scale), b_n (location)



Example: uniform distribution

$F(x) = x$

With $a_n = 1/n, b_n = 1$:

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} \left(\frac{x}{n} + 1\right)^n = \exp(x),$$

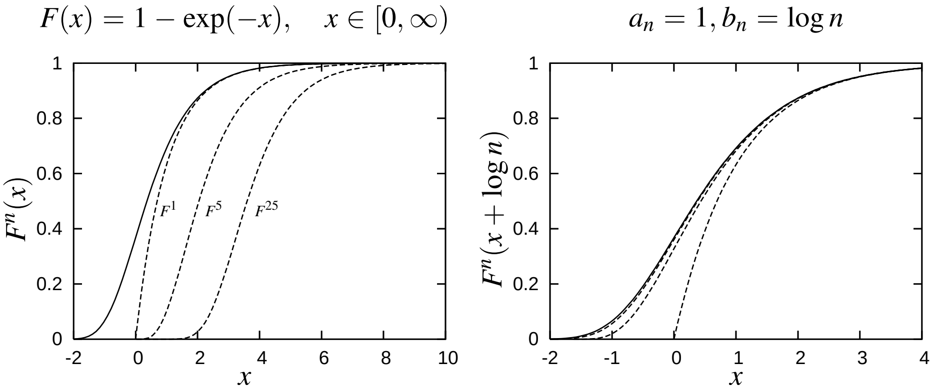
where

$$G(x) = \exp(x), \quad x \in (-\infty, 0]$$

is a Weibull distribution.

Example: exponential distribution

a_n (scale), b_n (location)



Example: exponential distribution

$F(x) = 1 - \exp(-x)$

With $a_n = 1, b_n = \log n$:

$$\lim_{n\to\infty} \{1 - \exp[-(x + \log n)]\}^n = \lim_{n\to\infty} \left[1 - \frac{\exp(-x)}{n}\right]^n$$
$$= \exp[-\exp(-x)],$$

where

$$G(x) = \exp[-\exp(-x)], \quad x \in (-\infty, \infty)$$

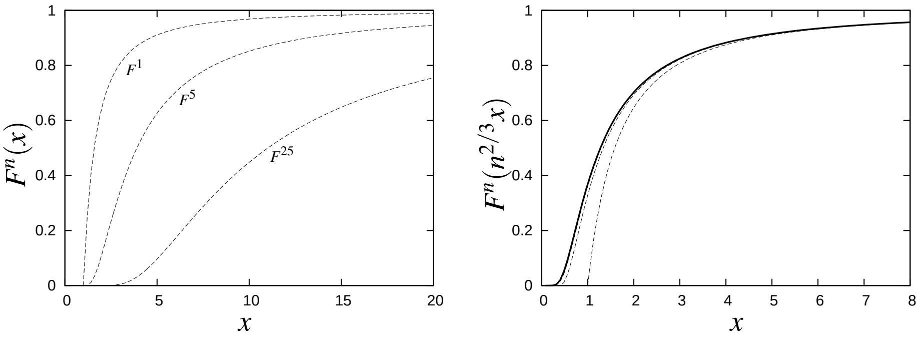
is the Gumbel distribution.

Example: a Pareto distribution

a_n (scale), b_n (location)

$$F(x) = 1 - x^{-3/2}, \quad x \in [1, \infty)$$

$$a_n = n^{2/3}, b_n = 0$$



Example: a Pareto distribution

$F(x) = 1 - x^{3/2}$

With $a_n = n^{2/3}, b_n = 0$:

$$\begin{aligned}\lim_{n\to\infty}\{1-(n^{2/3}x)^{-3/2}\}^n &= \lim_{n\to\infty}\left[1-\frac{x^{-3/2}}{n}\right]^n \\ &= \exp[-x^{-3/2}],\end{aligned}$$

where

$$G(x) = \exp[-x^{-3/2}], \quad x \in [0,\infty)$$

is a Fréchet distribution.

Generalised extreme value distribution

Jenkinson-von Mises representation

Overall maximum = maximum of block maxima:

$$\max\{\underbrace{\overbrace{\{X_{i_1},\ldots,X_{i_n}\}}^{F^n(x)\to G(a_n^{-1}(x-b_n))},\ldots,\{X_{j_1},\ldots,X_{j_n}\}}_{F^{mn}(x)}\} = \max\{\underbrace{M_{n;1},\ldots,M_{n;m}}_{G^m(a_n^{-1}(x-b_n))}\}$$

$$G^m(a_n^{-1}a_{mn}x + a_n^{-1}(b_{mn}-b_n)) \to G^m(a_m x + b_m) = G(x)$$

GEV distributions are max-stable:

$$G^m(a_m x + b_m) = G(x).$$

Solve functional equation:

$$G_\xi(x) = \exp[-(1+\xi x)^{-1/\xi}], \quad 1+\xi x > 0.$$

Generalised extreme value distribution

Fisher, Tippett 1928

Weibull, $\xi < 0$

$$G_{\xi<0}(x) = \begin{cases} \exp(-(-x)^{-1/\xi}), & x \leq 0 \\ 1, & x > 0 \end{cases}$$

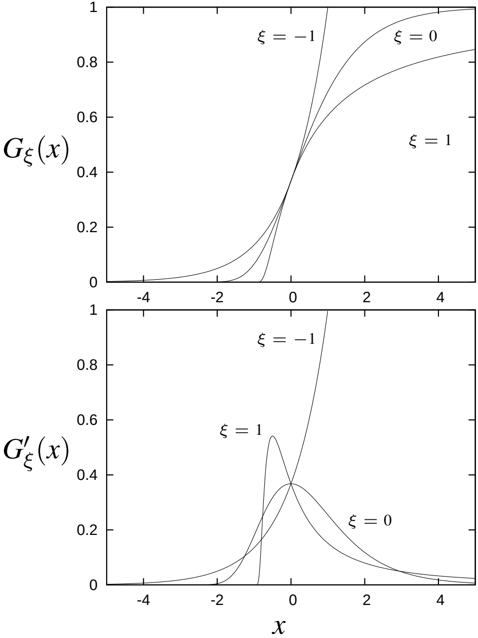
Gumbel, $\xi = 0$

$$G_{\xi=0}(x) = \exp(-\exp(-x)), \quad -\infty < x < \infty$$

Fréchet, $\xi > 0$

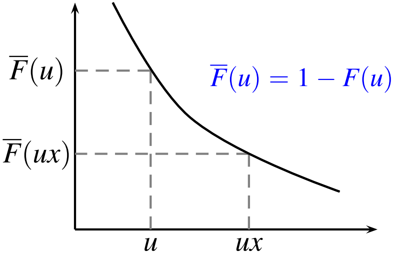
$$G_{\xi>0}(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-1/\xi}), & x > 0 \end{cases}$$

Generalised extreme value distribution



Characterising tail behaviour

Regular variation



$\overline{F} \in \mathcal{R}_{-\alpha}$ (regularly varying with exponent $-\alpha$) if

$$\lim_{u \to \infty} \frac{\overline{F}(ux)}{\overline{F}(u)} = x^{-\alpha}, \quad x > 0.$$

Domains of attraction: Fréchet

E.g. for power law $F(u) = 1 - u^{-1/\xi}, u \in [1, \infty)$,

$$\begin{aligned}\lim_{u\to\infty} \frac{\overline{F}(ux)}{\overline{F}(u)} &= \lim_{u\to\infty} \frac{(ux)^{-1/\xi}}{u^{-1/\xi}} \\ &= x^{-1/\xi}.\end{aligned}$$

Domains of attraction: Weibull

If $x_F < \infty$, check behaviour of $\overline{F}(x_F - u)$ as $u \downarrow 0$.

E.g. for uniform $F(u) = u, u \in [0, 1]$,

$$\lim_{u \downarrow 0} \frac{\overline{F}(x_F - ux)}{\overline{F}(x_F - u)} = \lim_{u \downarrow 0} \frac{ux}{u}$$
$$= x.$$

Domains of attraction: Weibull

E.g. for truncated exponential

$$F(u) = \frac{1 - \exp(-u)}{1 - \exp(-x_F)}, \quad u \in [0, x_F],$$

$$\begin{aligned} \lim_{u \downarrow 0} \frac{\overline{F}(x_F - ux)}{\overline{F}(x_F - u)} &= \lim_{u \downarrow 0} \frac{\exp(ux) - 1}{\exp(u) - 1} \\ &= x. \end{aligned}$$

Domains of attraction: Gumbel

E.g. for exponential, $F(u) = 1 - \exp(-u), u \in [0, \infty)$,

$$\begin{aligned}\lim_{u\to\infty} \frac{\overline{F}(ux)}{\overline{F}(u)} &= \lim_{u\to\infty} \exp(-u(x-1)) \\ &= \begin{cases} 0, & x > 1 \\ \infty, & x < 1. \end{cases}\end{aligned}$$

$\exp(-u) \in \mathcal{R}_{-\infty}$ (rapid variation).

Domains of attraction $\mathcal{D}(\mathcal{G}_\xi)$

$\overline{F} \in \mathcal{D}(\mathcal{G}_\xi)$ iff for some $g > 0$

$$\lim_{u \uparrow x_F} \frac{\overline{F}(u + xg(u))}{\overline{F}(u)} = (1 + \xi x)^{-1/\xi}, \quad 1 + \xi x > 0,$$

where g (non-unique) can be taken to be

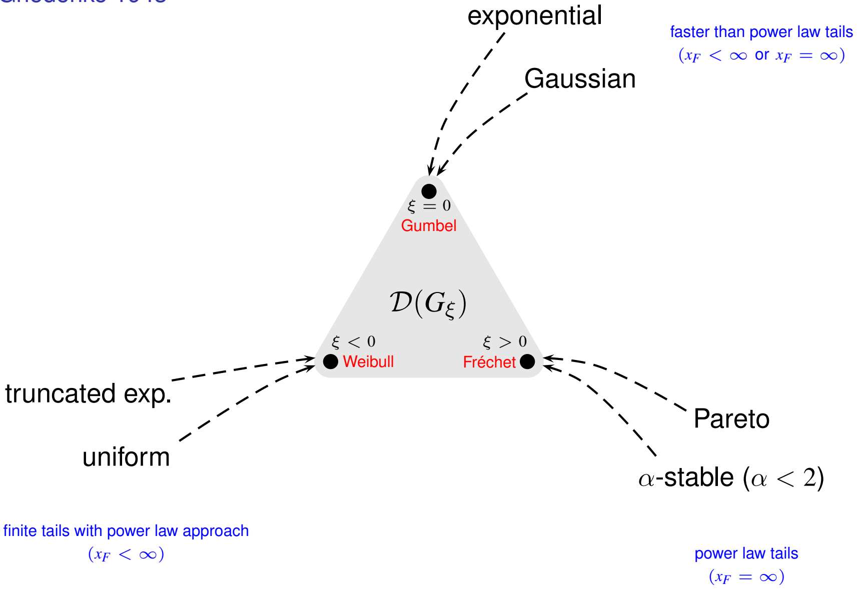
$$g(u) = \begin{cases} \xi u, & \xi > 0 \text{ (Fréchet)}, \\ -\xi(x_F - u), & \xi < 0 \text{ (Weibull)}, \\ \int_u^{x_F} \mathrm{d}t\, \overline{F}(t)/\overline{F}(u), & \xi = 0 \text{ (Gumbel)}. \end{cases}$$

Norming constants (non-unique) can be taken to be

$$\begin{array}{lll} a_n = F^{-1}(1 - n^{-1}) & b_n = 0 & \xi > 0 \text{ (Fréchet)}, \\ a_n = x_F - F^{-1}(1 - n^{-1}) & b_n = x_F & \xi < 0 \text{ (Weibull)}, \\ a_n = g(b_n) & b_n = F^{-1}(1 - n^{-1}) & \xi = 0 \text{ (Gumbel)}. \end{array}$$

Domains of attraction $\mathcal{D}(\mathcal{G}_\xi)$

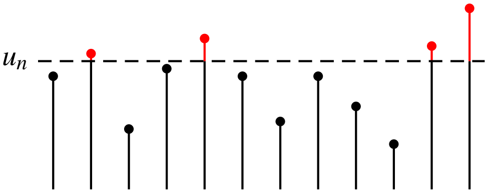
Gnedenko 1943



Outline

Poisson approximation

Number of exceedances n_u is a Poisson random variable



Poisson process

$$n_u \overset{d}{\sim} \mathrm{Bin}(n, \overline{F}(u_n))$$
$$\mathbb{E}[n_u] = n\overline{F}(u_n) \to \tau$$
$$n_u \to \mathrm{Poi}(\tau)$$

Poisson approx. to extremes $\begin{cases} n\overline{F}(u_n) \to \tau \\ \mathbb{P}(M_n \leq u_n) = \mathbb{P}(n_u = 0) \to \exp(-\tau) \end{cases}$

Generalises for kth maximum:

$$\mathbb{P}(M_n^{(k)} \leq u_n) = \mathbb{P}(n_u < k) \to \exp(-\tau) \sum_{m=1}^{k-1} \frac{\tau^m}{m!}.$$

Consistency with GEV

Refining the Poisson approximation

Since

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\xi(x),$$

$$\begin{aligned} n \log F(a_n x + b_n) &\to \log G_\xi(x) \\ n \log(1 - \overline{F}(a_n x + b_n)) &\to \log G_\xi(x) \\ n\overline{F}(a_n x + b_n) &\approx -\log G_\xi(x). \end{aligned}$$

With the choice $u_n = a_n x + b_n$,

$$\begin{aligned} \tau &= -\log G_\xi(x) \\ &= (1 + \xi x)^{-1/\xi}, \end{aligned}$$

so that

$$\mathbb{P}(M_n \leq a_n x + b_n) = \exp(-(1 + \xi x)^{-1/\xi}).$$

Connection between GEV and GPD

From $n\overline{F}(a_n x + b_n) \approx -\log G_\xi(x)$,

$$\overline{F}(x) \approx \frac{1}{n}\left[1 + \xi \frac{(x - b_n)}{a_n}\right]^{-1/\xi},$$

so excess y over threshold u is distributed as

$$\mathbb{P}(X > u + y | X > u) = \frac{\overline{F}(u+y)}{\overline{F}(u)} = \frac{\left[1 + \xi \frac{(u+y-b_n)}{a_n}\right]^{-1/\xi}}{\left[1 + \xi \frac{(u-b_n)}{a_n}\right]^{-1/\xi}}$$

$$= \left(1 + \frac{\xi}{\beta} y\right)^{-1/\xi} \quad \text{(GPD)},$$

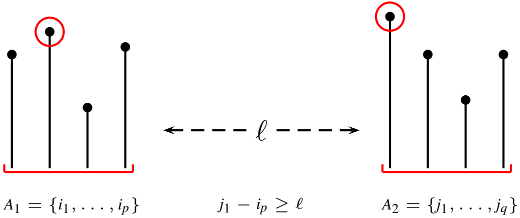
with

$$\beta = a_n + \xi(u - b_n).$$

Outline

Asymptotic independence of maxima

Loynes 1965; Leadbetter, Lindgren, Rootzén 1983



Stationary sequence X_n satisfies $D(u_n)$ (mixing) condition if

$$\left| \mathbb{P}\left(\max_{i \in A_1 \cup A_2} X_i \leq u_n \right) - \mathbb{P}\left(\max_{i \in A_1} X_i \leq u_n \right) \mathbb{P}\left(\max_{i \in A_2} X_i \leq u_n \right) \right| \leq \alpha_{n,\ell},$$

where $\alpha_{n,\ell} \to 0$ as $n \to \infty$ for some $\ell = \ell_n = o(n)$.

Berman's condition

Berman 1964

Stationary Gaussian sequence satisfies $D(u_n)$ if its autocorrelation function

$$\lim_{n \to \infty} r_n \log n = 0.$$

- AR(1) process: $r_n \sim \exp(-n)$.
- Long-range dependence: $r_n \sim n^{-\gamma}, \gamma \in (0, 1)$.

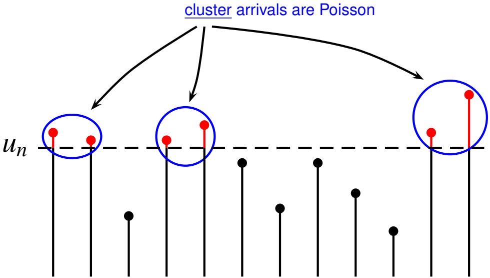
GEV/GPD limit distributions still apply!

Counterexample:

- Random walk: $r_n \sim n$ (Berman's condition fails).

Clustering & extremal index

Compound Poisson point process



$n\overline{F}(u_n) \to \tau$

$\mathbb{P}(M_n \le u_n) \to \exp(-\theta\tau)$

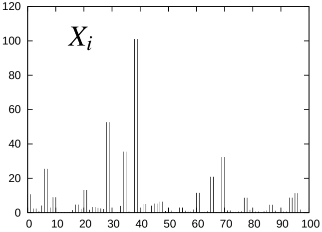
cluster size $\approx 1/\theta$

Heuristic interpretation of $\theta \in [0, 1]$:

- Inverse mean cluster size.
- Measure of loss of iid degrees of freedom.
- Multiplicity of associated compound Poisson process.

Can check for $\theta < 1$ via an anticlustering condition $D'(u_n)$.

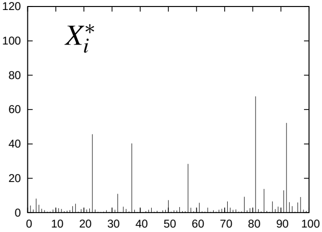
Example: $X_i = \max\{Y_i, Y_{i+1}\}$



E.g. $F_Y(y) = \exp(-1/2y), y \in [0, \infty)$

$$\mathbb{P}(X_i \le x) = \mathbb{P}(Y_i \le x, Y_{i+1} \le x) = \exp(-1/x)$$

$F_X(x) = \exp(-1/x)$ (dependent)



$F_{X^*}(x) = \exp(-1/x)$ (iid)

Example: $X_i = \max\{Y_i, Y_{i+1}\}$

For iid sequence:

$$\begin{aligned}\mathbb{P}(M_n^* \le nx) &= F_{X^*}^n(nx) \\ &= \exp(-1/x).\end{aligned}$$

For dependent sequence:

$$\begin{aligned}\mathbb{P}(M_n \le nx) &= \mathbb{P}(X_1 \le nx, \ldots, X_n \le nx) \\ &= \mathbb{P}(Y_1 \le nx, Y_2 \le nx, \ldots, Y_n \le nx, Y_{n+1} \le nx) \\ &= F_Y^{n+1}(nx) \\ &\to \exp(-1/x)^{1/2}, \quad n \to \infty.\end{aligned}$$

So

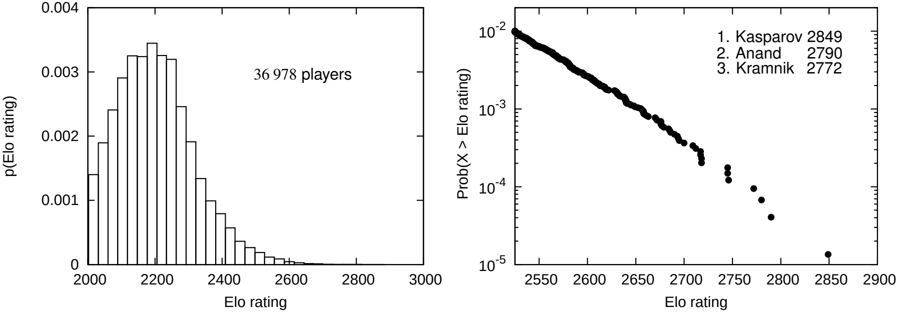
$$G_X(x) = G_{X^*}^{\theta}(x),$$

with extremal index $\theta = 1/2$.

Outline

Example: chess ratings

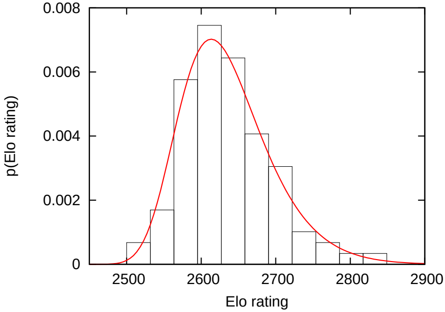
FIDE Elo ratings, January 2001



Example: chess ratings

Block maxima

$$G_{\xi;\mu,\sigma}(x) = \exp\left[-\left(1+\xi\frac{(x-\mu)}{\sigma}\right)^{-1/\xi}\right].$$

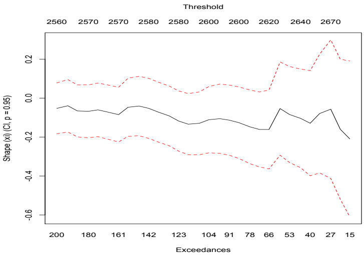


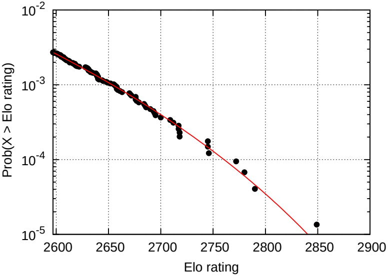
$$\xi = -0.08(6), \mu = 2609(6), \sigma = 53(4)$$

$n = 400$ (block size)

Example: chess ratings

Exceedances

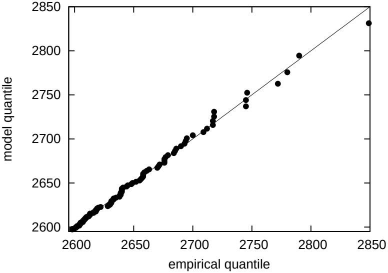




$$\overline{F}_{\xi;\beta}(y) = \frac{n_u}{n}\left[1 + \frac{\xi}{\beta}(y-u)\right]^{-1/\xi}$$

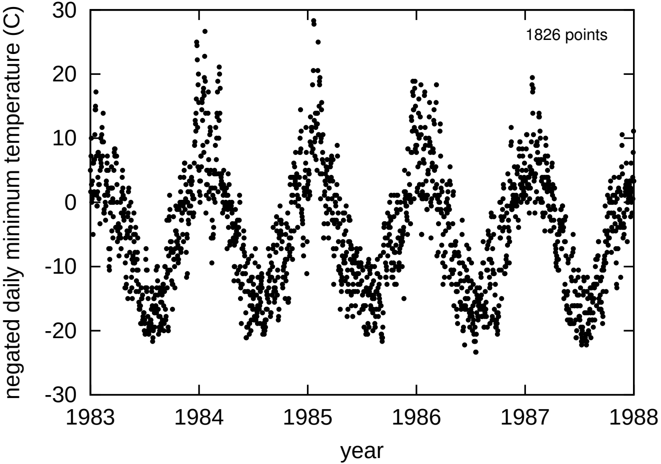
$$\xi = -0.13(9), \beta = 60(8)$$

$u = 2597$ (100 exceedances)

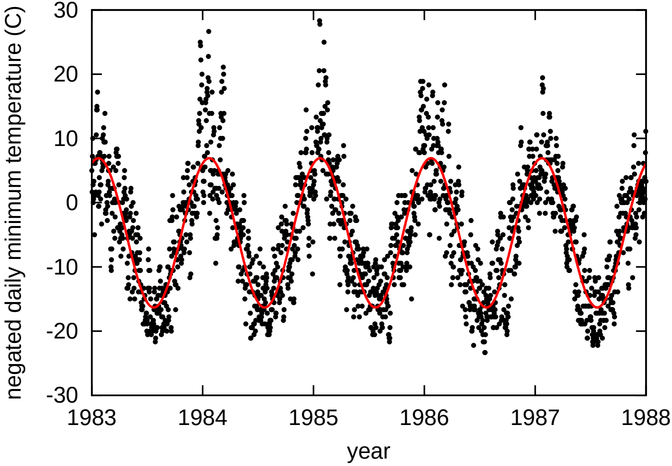


Example: Wooster daily minimum temperatures

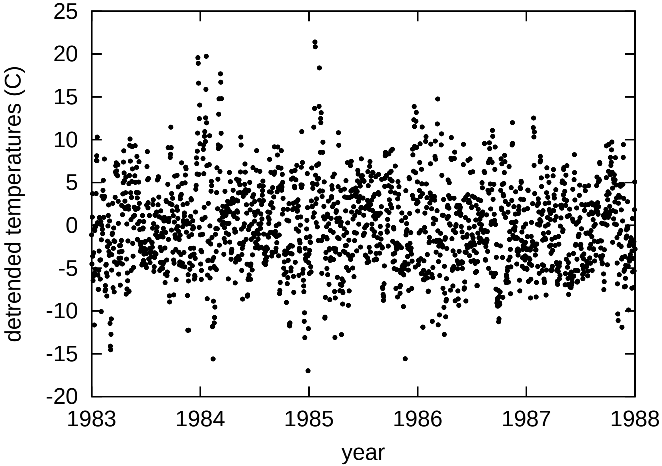
S. G. Coles, J. A. Tawn, R. L. Smith, Environmetrics **5**, 221 (1994)



Example: Wooster daily minimum temperatures

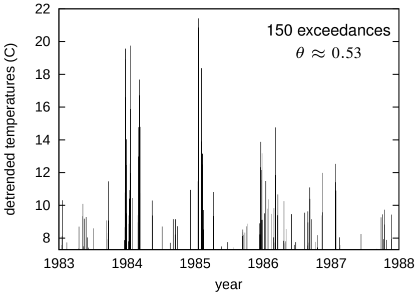


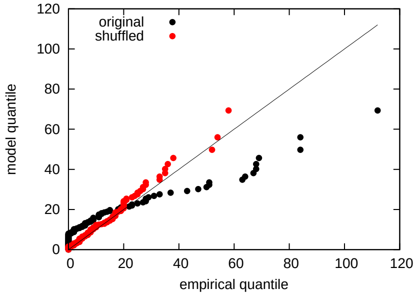
Example: Wooster daily minimum temperatures

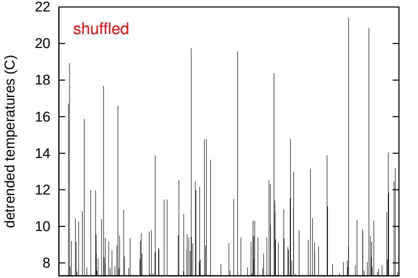


should also detrend the seasonal scale!

Exceedance arrivals (Poisson for $\theta = 1$)



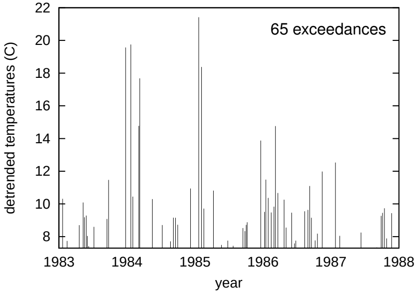


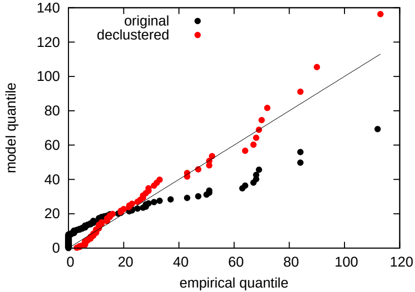


$$\mathbb{P}(\text{interval} > t) = \exp(-\lambda t)$$

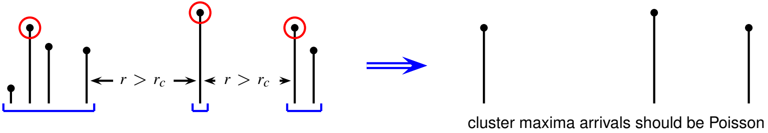
Declustering exceedances

"Runs" method (clusters separated by $r > r_c$)

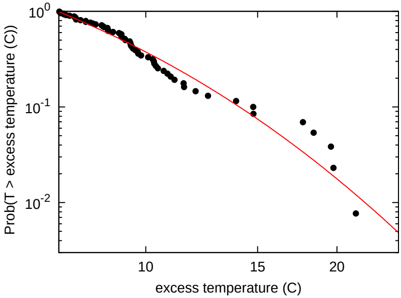


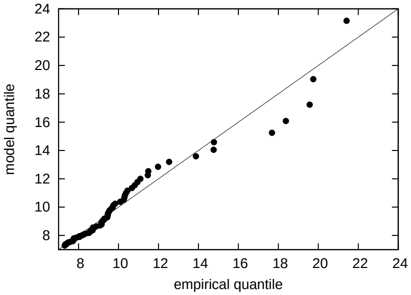


$$\mathbb{P}(\text{interval} > t) = \exp(-\lambda t)$$



Declustering exceedances



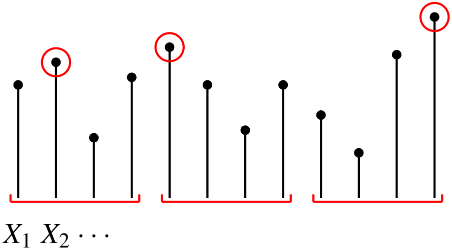


$$\overline{F}_{\xi;\beta}(y) = \frac{n_u}{n}\left[1 + \frac{\xi}{\beta}(y-u)\right]^{-1/\xi}$$

$$\xi = 0.08(14), \beta = 2.7(5)$$

(65 declustered exceedances)

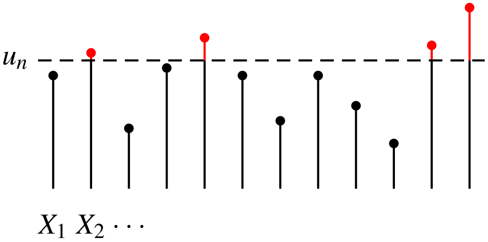
Summary



$$G_\xi(x) = \exp\left\{-\left[1+\xi\frac{(x-\mu)}{\sigma}\right]^{-1/\xi}\right\}$$

GEV

Also extends to stationary dependent sequences satisfying $D(u_n)$ conditions!



$$H_\xi(y) = 1 - \left(1+\frac{\xi}{\beta}y\right)^{-1/\xi}$$

GPD