Skoltech

Skolkovo Institute of Science and Technology

Stochastic methods in Mathematical Modelling

Lecture 4. Extreme value statistics

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Applications

- Equity risks. Day to day market risk
- Extreme weather events
- Performances in a sport (what would be the next 100 m sprint record?)
- Mutational events during evolution
- Side effects of drugs

  etc. etc.

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

We are interested in the most extreme value *M* out of the set of random variables $\{x_1, x_2, \ldots, x_N\}$.

$$\text{maximum} \quad M = \max\left(x_1, x_2, ..., x_N\right)$$

Simplification: x_1, …, x_n are i.i.d.

$$P(x_1, x_2, ..., x_N) = p(x_1)p(x_2)...p(x_N)$$

For the mean of *N* samples we have a universality law (if the variance is finite) in form of CLT

What about the maximum?

↓

Universality in the case of i.i.d.

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

We are interested in the most extreme value *M* out of the set of random variables $\{x_1, x_2, \ldots, x_N\}$.

The general case CDF of distribution of *M* reads

$$Q_N(x) = \mathrm{Prob}[M \le x, N] = \mathrm{Prob}[x_1 \le x,\ x_2 \le x, ..., x_N \le x] \ ,$$
$$= \int_{-\infty}^{x} dx_1 \int_{-\infty}^{x} dx_2 \cdots \int_{-\infty}^{x} dx_N P(x_1, x_2, \cdots, x_N) \ ,$$

Hence the pdf

$$P(M, N) = Q'_N(M) \equiv \frac{dQ_N}{dx}\Big|_{x=M}$$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

We are interested in the most extreme value *M* out of the set of random variables $\{x_1, x_2, \ldots, x_N\}$.

CDF for i.i.d. case

$$Q_N(x) = \left[\int_{-\infty}^{x} dy\ p(y)\right]^N = \left[1 - \int_{x}^{\infty} dy\ p(y)\right]^N$$

When *N* tends to ∞ $Q_N(x)$ approaches the limiting form

$$Q_N(x) \xrightarrow[z=(x-a_N)/b_N \text{ fixed}]{x\to\infty,\ N\to\infty} F\left(\frac{x-a_N}{b_N}\right)$$

$$\text{equivalently}, \quad \lim_{N\to\infty} Q_N(a_N + b_N z) = F(z)$$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Gnedenko's classical law of extremes

For the maximum statistics from i.i.d. variables only ***three*** possible ***forms*** of ***F(z)*** exist. The particular one depends only the large *x* tail of the source distribution *p*(*x*).

Fisher-Tippett-Gnedenko Theorem. Let $\{X_n\}$ be a sequence of i.i.d. random variables. If there exist norming constants $a_N > 0$, $b_N \in \mathbb{R}$ and some non-degenerate CDF *G* such that $a_N^{-1}(M_n - b_n)$ converges in distribution to *G*, then *G* is of one of the following three CDFs:

$$
\begin{aligned}
\textit{Gumbel:} \quad & G_0(x) = \exp(-e^{-x}), && x \in \mathbb{R}, \\
\textit{Fréchet:} \quad & G_{1,\alpha}(x) = \exp(-x^{-\alpha}), && x \geq 0,\ \alpha > 0, \\
\textit{Reversed Weibull:} \quad & G_{2,\alpha}(x) = \exp(-(-x)^{-\alpha}), && x \leq 0,\ \alpha < 0.
\end{aligned}
$$

$$
G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right), 1+\gamma x > 0.
$$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Uncorrelated RVs. ***Fréchet*** *distribution*

For the maximum statistics from i.i.d. variables only ***three*** possible ***forms*** of ***F(z)*** exist. The particular one depends only the large x tail of the source distribution p(x).

1) P(x) has a power-law tail $p(x) \sim A\,x^{-(1+\alpha)}$

$$F_1(z) = \begin{cases} \mathrm{e}^{-z^{-\alpha}} & \text{for } z \geq 0 \\ 0 & \text{for } z \leq 0 \end{cases}$$

Fréchet *distribution*

$$f_1(z) = F_1'(z) = \frac{\alpha}{z^{\alpha+1}} \mathrm{e}^{-z^{-\alpha}}, \qquad z \in [0, \infty)$$

TABLE 1. *A list of distributions in the Fréchet domain*

Distribution	$1 - F(x)$	Extreme value index
Pareto	$\sim Kx^{-\alpha},\ K, \alpha > 0$	$\frac{1}{\alpha}$
$F(m,n)$	$\int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \omega^{\frac{m}{2}-1} \left(1+\frac{m}{n}\omega\right)^{-\frac{m+n}{2}} d\omega$ $x > 0;\ m,n > 0$	$\frac{2}{n}$
Fréchet	$1 - \exp(-x^{-\alpha})$ $x > 0; \alpha > 0$	$\frac{1}{\alpha}$
T_n	$\int_x^\infty \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{\omega^2}{n}\right)^{-\frac{n+1}{2}} d\omega$ $x > 0; m,n > 0$	$\frac{1}{n}$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

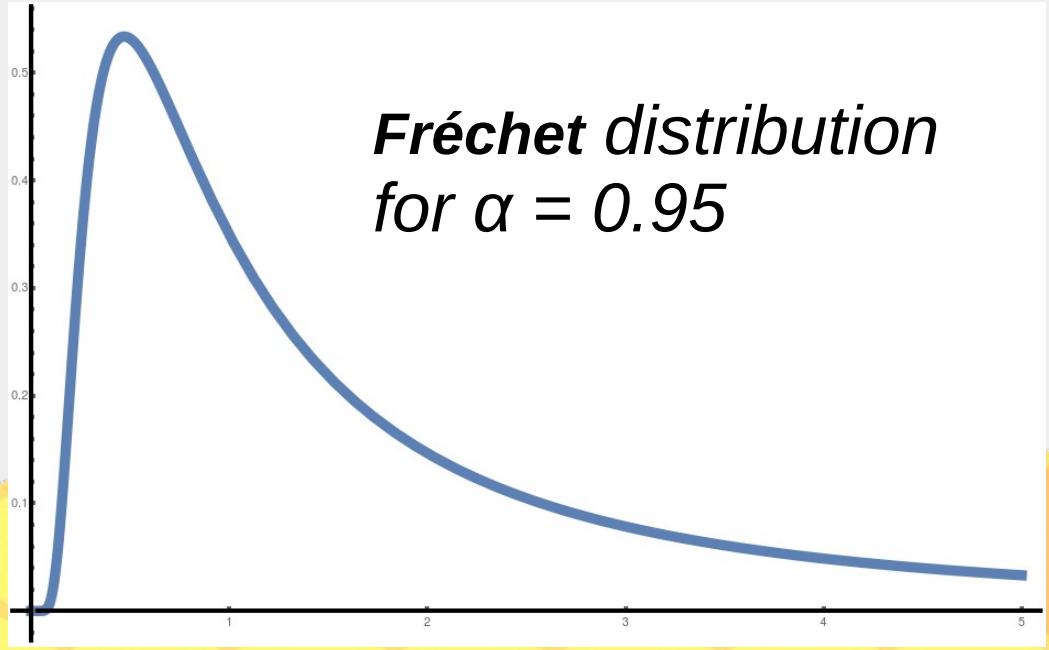
Uncorrelated RVs. ***Fréchet*** *distribution*

For the maximum statistics from i.i.d. variables only ***three*** possible ***forms*** of ***F(z)*** exist. The particular one depends only the large x tail of the source distribution p(x).

1) P(x) has a power-law tail $p(x) \sim A\,x^{-(1+\alpha)}$

$$F_1(z) = \begin{cases} \mathrm{e}^{-z^{-\alpha}} & \text{for } z \geq 0 \\ 0 & \text{for } z \leq 0 \end{cases}$$

$$f_1(z) = F_1'(z) = \frac{\alpha}{z^{\alpha+1}} \mathrm{e}^{-z^{-\alpha}}, \qquad z \in [0, \infty)$$



Fréchet *distribution*
for α = 0.95

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Uncorrelated RVs. ***Gumbel*** *distribution*

2) $p(x)$ has a tail decaying faster than a power law and unbounded

$$p(x) \sim \mathrm{e}^{-x^{\delta}} \text{ with } \delta > 0$$

$$F_2(z) = \mathrm{e}^{-\mathrm{e}^{-z}}$$

Gumbel *distribution*

$$f_2(z) = F_2'(z) = \mathrm{e}^{-z-\mathrm{e}^{-z}}, \qquad z \in (-\infty, \infty)$$

TABLE 3. *A list of distributions in the Gumbel domain*

Distribution	$1 - F(x)$
Weibull	$\exp(-\lambda x^{\tau}), \quad x > 0;\ \lambda, \tau > 0$
Exponential	$\exp(-\lambda x), \quad x > 0;\ \lambda > 0$
Gamma	$\frac{\lambda^m}{\Gamma(m)} \int_x^{\infty} u^{m-1} \exp(-\lambda u) du, \quad x > 0;\ \alpha, m > 0$
Logistic	$1/(1+\exp(x)), \quad x \in \mathbb{R}$
Normal	$\int_x^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R};\ \sigma > 0, \mu \in \mathbb{R}$
Log-normal	$\int_x^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 u}} \exp\left(-\frac{1}{2\sigma^2}(\log u - \mu)^2\right) du, \quad x > 0;\ \mu \in \mathbb{R}, \sigma > 0$

Q. Could you obtain $F_2(z)$ for the exponential distribution $p(x)=e^{-x}, x \geq 0$?

Reminder: F(z) is obtained from $Q_N(x)$ which is

$$Q_N(x) = \left[\int_{-\infty}^{x} dy\ p(y)\right]^N = \left[1 - \int_{x}^{\infty} dy\ p(y)\right]^N$$

$$Q_N(x) \xrightarrow[z=(x-a_N)/b_N \text{ fixed}]{x\to\infty,\ N\to\infty} F\left(\frac{x-a_N}{b_N}\right)$$

$$\text{equivalently}, \quad \lim_{N\to\infty} Q_N(a_N + b_N z) = F(z)$$

Gumbel law for Gaussian distr. i.i.d.

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$Q_N(x) = \left[\int_{-\infty}^{x} dy\, p(y)\right]^N = \left[1 - \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)\right]^N = \exp\left[N \ln\left(1 - \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)\right)\right]$$

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{+\infty} e^{-y^2} dy$$

Expanding the ln:

$$Q_N(x) \simeq \exp\left[-\frac{N}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)\right]$$

for $x \gg 1$

$$\operatorname{erfc}(x) \simeq \frac{e^{-x^2}}{\sqrt{\pi}\, x} \Rightarrow Q_N(x) \simeq \exp\left(-\frac{N}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x}\right)$$

$e^{-e^{-z}}$

we want to get to universal form for $z = \frac{x - a_N}{b_N}$ → a_N ? b_N ?

$$Q_N(z) = Q_N(x = a_N + b_N z) = \exp\left[-\frac{N}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}(a_N^2 + 2 b_N z a_N + b_N^2 z^2)\right)}{a_N + b_N z}\right]$$

$b_N \ll a_N$

$$Q_N(x = a_N + b_N z) = e^{-\frac{N}{\sqrt{2\pi}\, a_N} \exp\left(-\frac{1}{2}(a_N^2 + 2 b_N z a_N)\right)}$$

$$\Rightarrow \begin{cases} \frac{N}{\sqrt{2\pi}\, a_N} e^{-\frac{1}{2} a_N^2} = 1 \\ b_N a_N = 1 \end{cases}$$

$N \to \infty$

$$\begin{cases} a_N \simeq \sqrt{2 \ln N} - \frac{\ln(\ln N)}{2\sqrt{2 \ln N}} \\ b_N = 1/a_N \end{cases}$$

Gumbel law for $e^{-x} = P(x)$

$$Q_N(x) = \left[\int_{-\infty}^{x} p(x)dx\right]^N = \left[1 - e^{-x}\right]^N = e^{N \ln\left[1 - e^{-x}\right]} = e^{-Ne^{-x}} =$$

$$= e^{-e^{-(x - \ln N)}}$$

$$z = x - \ln N \longrightarrow \begin{matrix} a_N = \ln N \\ b_N = 1 \end{matrix}$$

Remark: $\forall$ distribution with $p(x) \sim c\, e^{-x^\delta}$, $\delta > 0$

$$a_N \simeq (\ln N)^{1/\delta}$$

$$b_N \simeq \frac{1}{\delta} (\ln N)^{1/\delta - 1} \rightarrow \underline{F_z(z) = e^{-e^{-z}} \quad \forall \delta}$$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Uncorrelated RVs. ***Gumbel*** *distribution*

2) p(x) has a tail decaying faster than a power law and unbounded

$$p(x) \sim \mathrm{e}^{-x^{\delta}} \text{ with } \delta > 0$$

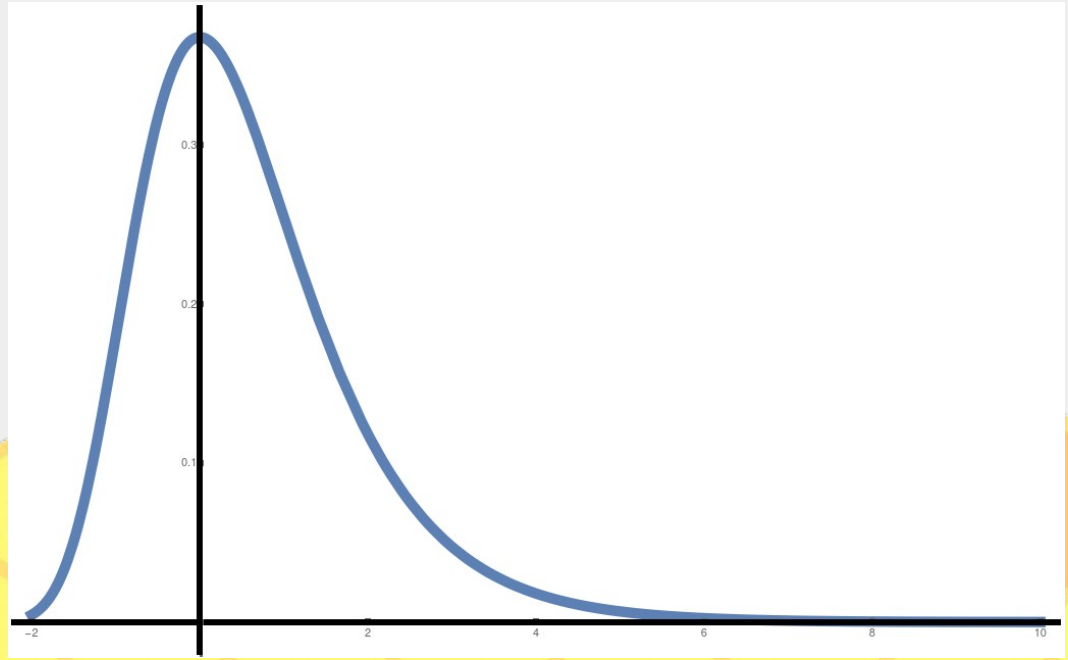
$$F_2(z) = \mathrm{e}^{-\mathrm{e}^{-z}}$$

Gumbel *distribution*

$$f_2(z) = F_2'(z) = \mathrm{e}^{-z-\mathrm{e}^{-z}}, \qquad z \in (-\infty, \infty)$$

TABLE 3. *A list of distributions in the Gumbel domain*

Distribution	$1 - F(x)$
Weibull	$\exp(-\lambda x^{\tau}), \quad x > 0;\ \lambda, \tau > 0$
Exponential	$\exp(-\lambda x), \quad x > 0;\ \lambda > 0$
Gamma	$\frac{\lambda^m}{\Gamma(m)} \int_x^{\infty} u^{m-1} \exp(-\lambda u) du, \quad x > 0;\ \alpha, m > 0$
Logistic	$1/(1+\exp(x)), \quad x \in \mathbb{R}$
Normal	$\int_x^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R};\ \sigma > 0, \mu \in \mathbb{R}$
Log-normal	$\int_x^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 u}} \exp\left(-\frac{1}{2\sigma^2}(\log u - \mu)^2\right) du, \quad x > 0;\ \mu \in \mathbb{R}, \sigma > 0$



Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Uncorrelated RVs. ***Weibull*** *distribution*

3) p(x) has a tail which is bounded, i.e. $p(x) \xrightarrow{x \to a} (a-x)^{\beta-1}$

$$F_3(z) = \begin{cases} \mathrm{e}^{-(-z)^{\beta}} & \text{for } z \le 0 \\ 1 & \text{for } z \ge 0 \end{cases}$$

Weibull *distribution*

$$f_3(z) = \beta(-z)^{\beta-1}\mathrm{e}^{-(-z)^{\beta}}, \qquad z \in (-\infty, 0]$$

TABLE 2. *A list of distributions in the Reversed Weibull domain*

Distribution	$1 - F\left(\omega(F) - \frac{1}{x}\right)$	Extreme value index
Uniform	$\frac{1}{x}$ $x > 1$	-1
Beta(p,q)	$\int_{1-\frac{1}{x}}^{1} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1}du$ $x > 1;\ p,q > 0$	$-\frac{1}{q}$
Reversed Weibull	$1 - \exp(-x^{-\alpha})$ $x > 0;\ \alpha > 0$	$-\frac{1}{\alpha}$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

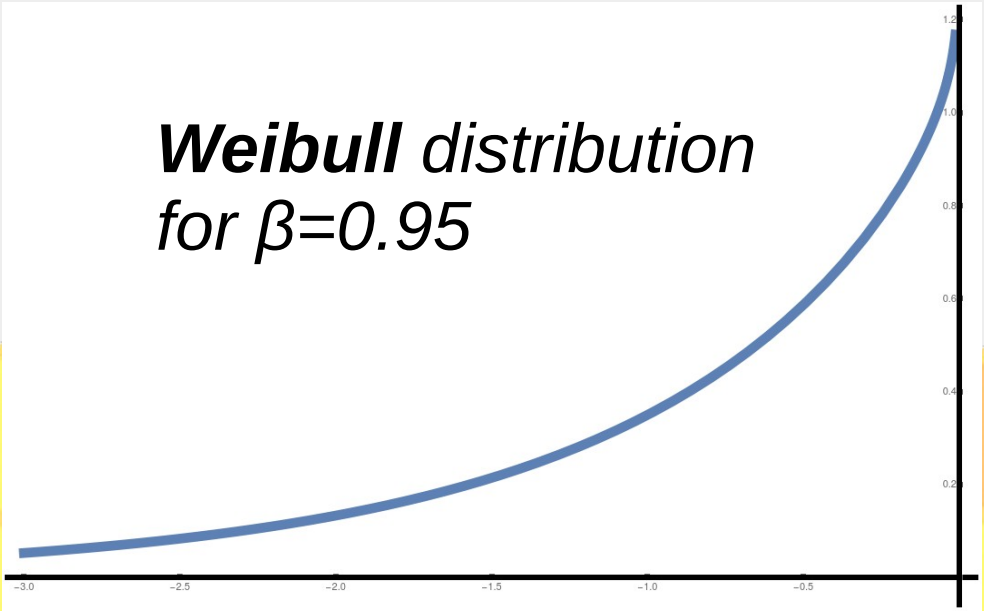
Uncorrelated RVs. ***Weibull*** *distribution*

3) p(x) has a tail which is bounded, i.e. $p(x) \xrightarrow{x \to a} (a - x)^{\beta - 1}$

$$F_3(z) = \begin{cases} \mathrm{e}^{-(-z)^{\beta}} & \text{for } z \leq 0 \\ 1 & \text{for } z \geq 0 \end{cases}$$

Weibull *distribution*

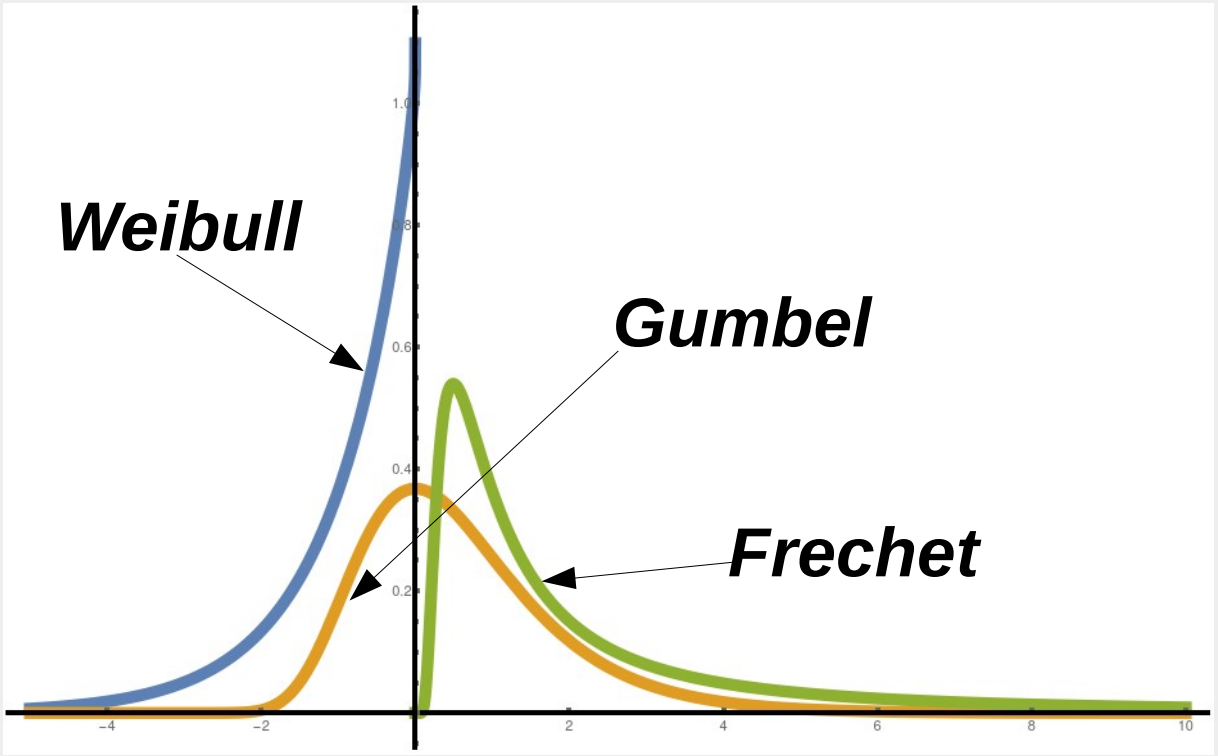
$$f_3(z) = \beta(-z)^{\beta-1}\mathrm{e}^{-(-z)^{\beta}}, \qquad z \in (-\infty, 0]$$



Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Uncorrelated RVs. ***All three universal*** *distributions*
(Frechet, Gumbel, Weibull)



Parent distribution $p(x)$	Scaling function $F(z)$	PDF of maximum $f(z)$	Nomenclature
$x^{-(1+\alpha)};\ \alpha > 0$	$\mathrm{e}^{-z^{-\alpha}}\theta(z)$	$\frac{\alpha}{z^{\alpha+1}}\mathrm{e}^{-z^{-\alpha}};\ z > 0$	**Fréchet**
$\mathrm{e}^{-x^{\delta}}$	$\mathrm{e}^{-\mathrm{e}^{-z}}$	$\mathrm{e}^{-z-\mathrm{e}^{-z}}$	**Gumbel**
$(a-x)^{\beta-1};\ \beta > 0$	$\mathrm{e}^{-(-z)^{\beta}}\theta(-z)+\theta(z)$	$\beta(-z)^{\beta-1}\mathrm{e}^{-(-z)^{\beta}};\ z < 0$	**Weibull**

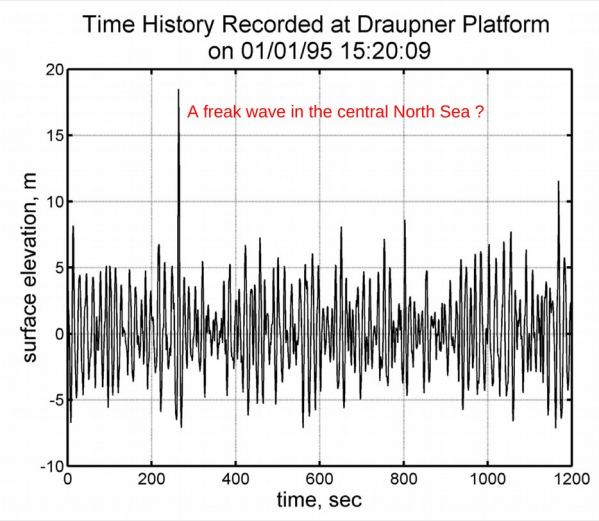
Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Example 1. Rogue (freak, killer) waves. Gumbel distr.

The Ninth Wave, I. Ayvazovsky, 1850

Rockall, 1943,
by RAF photographer
Fisher, James



From P.H. Taylor, 1995

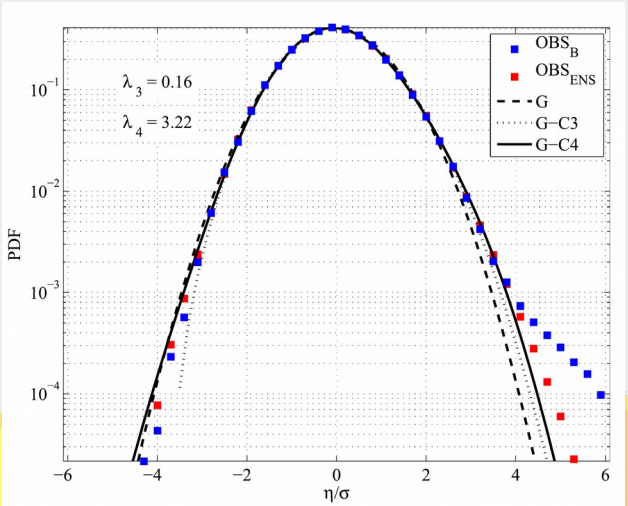


FIG. 7. PDF of dimensionless sea surface elevations of the records B_i (OBS_B) and of the space–time ensemble (OBS_{ENS}). The Gaussian distribution (G), the third- (G-$C3$), and fourth-order (G-$C4$) nonlinear corrections are displayed for reference.

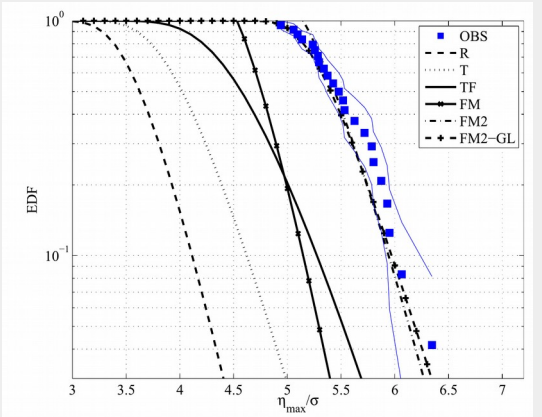


FIG. 9. Probability of exceedance (EDF) of dimensionless observed extreme crest heights $\eta_{i,\max} = \max[\eta_i(t)]$ of the records B_i (OBS). The empirical EDF stability band is plotted as blue solid line. Reference distributions of extremes: Rayleigh (R); Tayfun (T); Tayfun–Fedele (TF); Fedele (FM); second-order FM (FM2); and asymptotic Gumbel limit of FM2 (FM2-GL). The reference duration is $D = 1798$ s. Space–time extremes are computed over an area $XY = 11.2 \times 11.2\,\mathrm{m}^2 = 126\,\mathrm{m}^2$.

Benetazzo, A. et al.
Observation of extreme sea waves
in a space-time ensemble.
J. Phys. Oceanogr. 45, 2261–2275 (2015).

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Example 2. Financial markets. Frechet distribution

In financial markets the extreme price movements correspond to market corrections (ordinary periods) and stock market crashes during crisis periods.

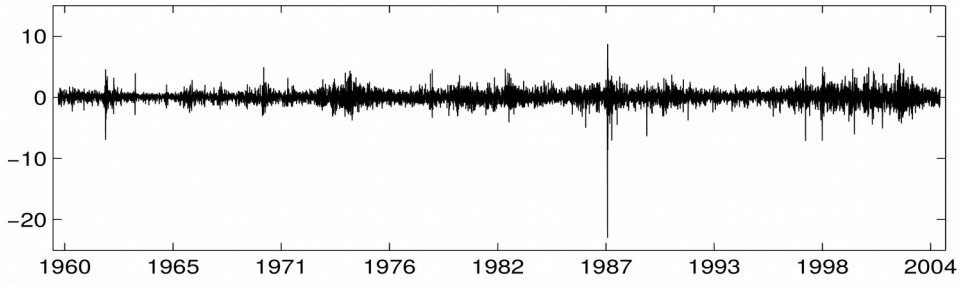


Fig. 5. Daily returns of the S&P500 index.

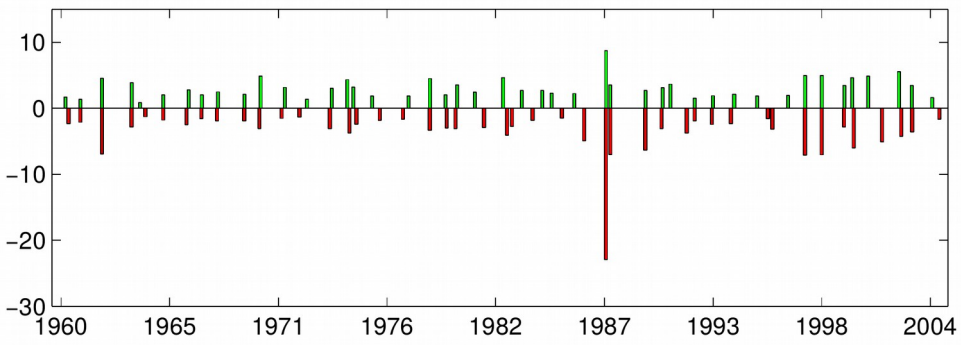


Fig. 6. Yearly minima and maxima of the daily returns of the S&P500.

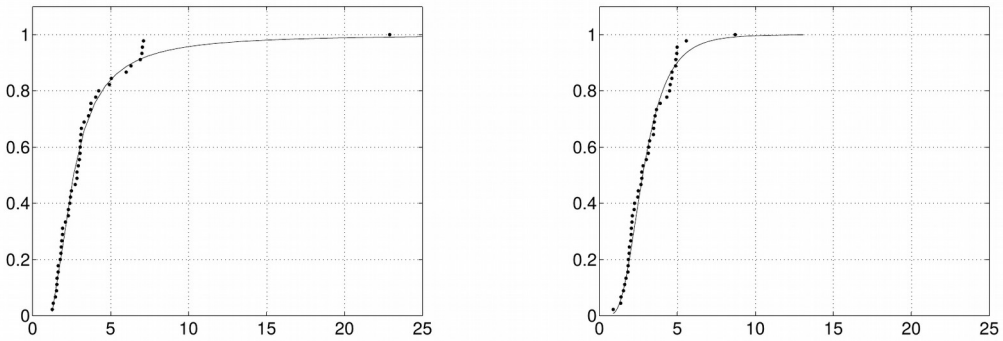


Fig. 7. Sample distribution (dots) of yearly minima (left panel) and maxima (right panel) and corresponding fitted GEV distribution for S&P500.

M. Gilli, E. Këllezi, Computational Economics 27(1), 2006, 1–23.

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Example 3. Impulse-driven surface breakdown data. Weibull distribution

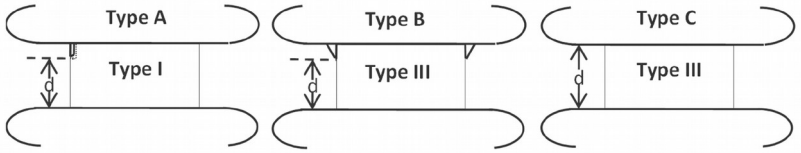


Figure 1. Sketch of the electrode/sample geometries tested and analysed (AI, BIII, and CIII).

Cumulative probability of failure

$$F(V) = 1 - \exp\left[-\left(\frac{V-\gamma}{\alpha}\right)^{\beta}\right]$$

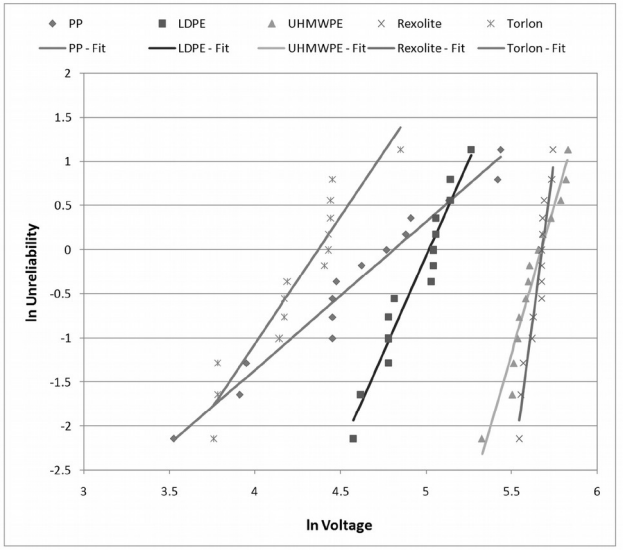


Figure 2. Weibull plots and curve fits of peak applied voltage data for type I (recess) samples tested with high-voltage electrode type A (pin).

M.P. Wilson et al., 2011 IEEE Pulsed Power Conference, https://doi.org/10.1109/PPC.2011.6191418

Skoltech
Skolkovo Institute of Science and Technology

Order statistics

What is the statistics of the succesive maxima and the gaps?

$$
\begin{aligned}
M_{1,N} &= \max\,(x_1, x_2, \cdots, x_N)\,, \\
M_{2,N} &= \text{second}\ \ \max\,(x_1, x_2, \cdots, x_N)\,, \\
&\vdots \\
M_{k,N} &= k\text{-th}\ \ \max\,(x_1, x_2, \cdots, x_N)\,, \\
&\vdots \\
M_{N,N} &= \min\,(x_1, x_2, \cdots, x_N)\,,
\end{aligned}
$$

$$
p_>(x) = \int_x^\infty p(y)dy
$$

$$
p_<(x) = \int_{-\infty}^x p(y)dy
$$

$$
Q_{k,N}(x) = \text{Prob}[M_{k,N} \le x] = \sum_{m=0}^{k-1} \binom{N}{m} [p_>(x)]^m [p_<(x)]^{N-m}
$$

$$
Q_{k,N}(x) = \text{Prob}[M_{k,N} \le x] \xrightarrow[z=(x-a_N)/b_N \ \text{fixed}]{x\to\infty,\ N\to\infty} G_k\left(\frac{x-a_N}{b_N}\right)
$$

$$
G_k(z) = F_\mu(z) \sum_{j=0}^{k-1} \frac{[-\ln F_\mu(z)]^j}{j!}
$$

$F_\mu(z)$ is Frechet, Gumbel or Weibull

G. Schehr and S.N. Majumdar, Exact record and order statistics of random walks via first-passage ideas, a book chapter in "First-Passage Phenomena and Their Applications", Eds. Metzler, Oshanin, Redner, World Scientific (2013), also arXiv: 1305:0639.

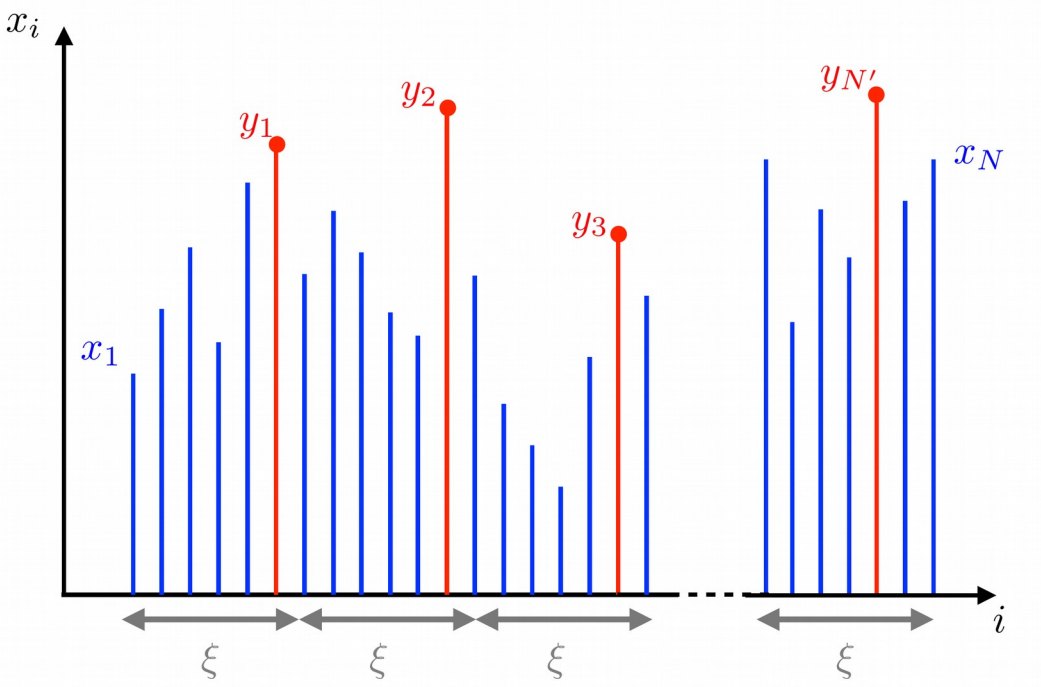
Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Weakly correlated RVs

Weak correlations = finite correlation length. As a consequence the same universality as for i.i.d.

$$C_{i,j} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \sim \mathrm{e}^{-|i-j|/\xi}$$



$$M = \max[x_1, x_2, \ldots, x_N] = \max[y_1, y_2, \ldots, y_{N'}]$$

To decide the tail of p(y) one needs to solve a strongly correlated problem

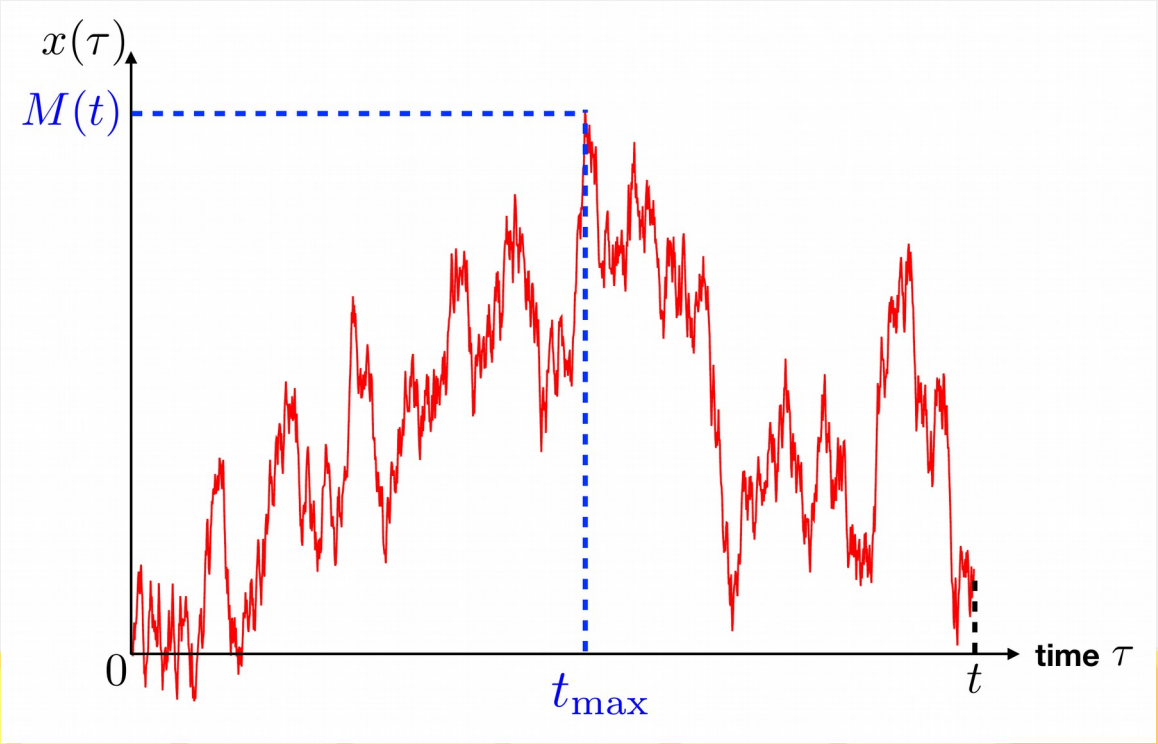
Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Strongly correlated RVs

Strong correlations
No universality discovered yet. A field with many open questions.
Some particular problems solved

Example: trajectory of Brownian motion



1D Brownian motion and RWs

$x_0 = 0$

Langevin equation $\frac{dx}{d\tau} = \eta(\tau)$ — noise

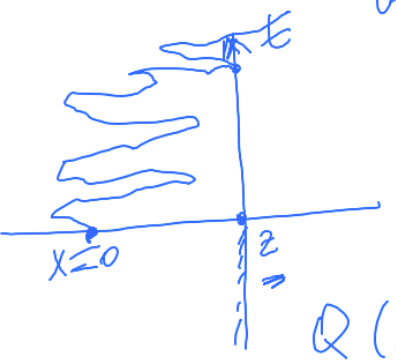
$$\begin{cases} \langle \eta(\tau) \rangle = 0 \\ \langle \eta(\tau)\eta(\tau') \rangle = 2D\delta(t-t') \end{cases}$$

white noise

$M(t) = \max_{0 \le \tau \le t} [x(\tau)]$

$x(\tau) = \int_0^\tau \eta(s)\,ds$

$\langle x(\tau) \rangle = 0$, $\langle x(\tau) \times (\tau') \rangle = 2D \min(\tau, \tau')$



CDF of the maximum

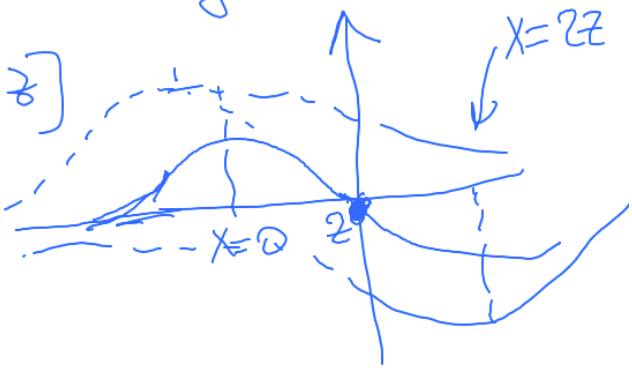
$Q(z,t) = \text{Prob}[M(t) \le z] = \text{Prob}[x(\tau) \le z,\ 0 \le \tau \le t]$

$Q(z,0) = 0$

$P(x,t|z)$ is PDF that particle is at (x,t) not exceeding z

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}$$

$$P(x,t|z) = \frac{1}{\sqrt{4\pi D t}}\left[e^{-\frac{x^2}{4Dt}} - e^{-\frac{(x-2z)^2}{4Dt}}\right] \quad , \quad x \in (-\infty, z]$$



$$Q(z,t) = \int_{-\infty}^{z} dx\, P(x,t|z) = \mathrm{erf}\left(\frac{z}{\sqrt{4Dt}}\right)$$

$$P_M(z,t) = \frac{\partial Q(z,t)}{\partial z} = \Theta(z)\frac{1}{\sqrt{\pi D t}}\exp\left(-\frac{z^2}{4Dt}\right)$$

Θ: Heaviside function $\quad \mathrm{erf} = \frac{2}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-y^2}dy$

$$\langle z \rangle = M(t) = \int_0^\infty dz\, z\, \frac{1}{\sqrt{\pi D t}} e^{-\frac{z^2}{4Dt}} = \frac{2\sqrt{Dt}}{\sqrt{\pi}} = M(t)$$

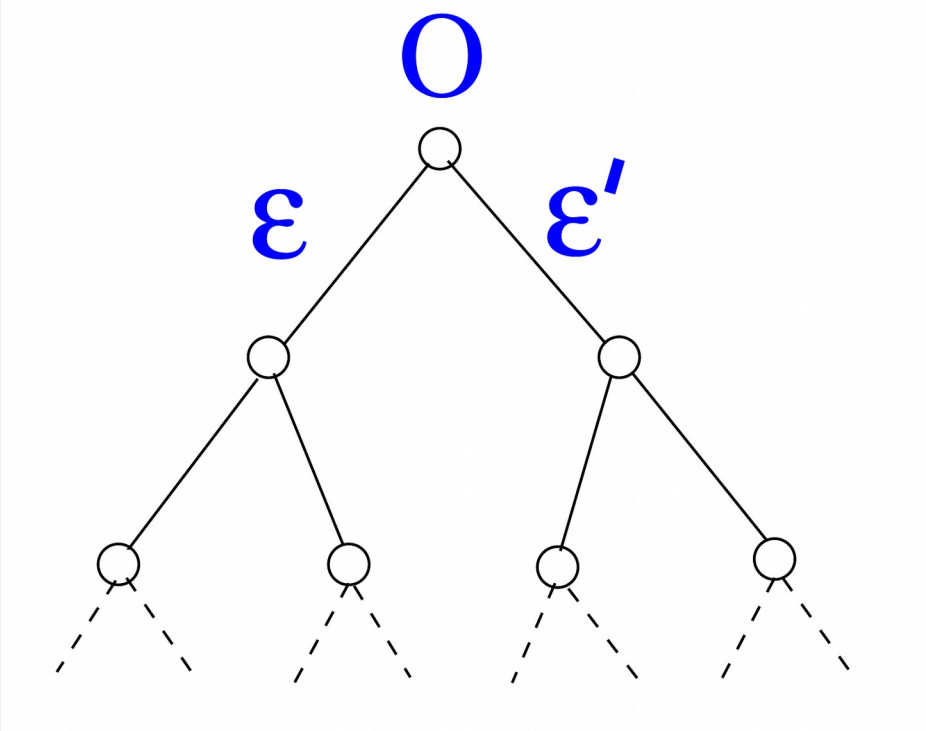
$\langle \Delta x^2(t) \rangle = 2Dt$ for BM

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Hierarchically correlated random variables

Disordered environment. Directed polymer on a Caylee tree



$$\langle E_{\min} \rangle \approx v(\lambda^*)\, n + \frac{3}{2\lambda^*} \ln n$$

$$v(\lambda) = -\frac{1}{\lambda} \ln \left[2 \int \rho(\epsilon) \mathrm{e}^{-\lambda\, \epsilon}\, d\epsilon \right]$$

Equivalent to binary search tree (BST!!!)

If the data are stored on a tree of size *N* the optimal search time scaling is O(lnN) not linear in *N*

BST height. Maximum depth of the nodes in a tree with *N* occupied nodes

For BST:

$$\langle H_N \rangle \approx \frac{1}{v(\lambda^*)} \ln N - \frac{3}{2\lambda^* v(\lambda^*)} \ln \ln N$$

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Extreme statistics in random matrix theory

$$\Pr[\{X_{i,j}\}] \propto \exp[-\frac{\beta}{2} N \mathrm{Tr}(X^2)]$$

β = 1 GOE
β = 2 GUE
β = 4 GSE

Eigenvalues, vectors are independent

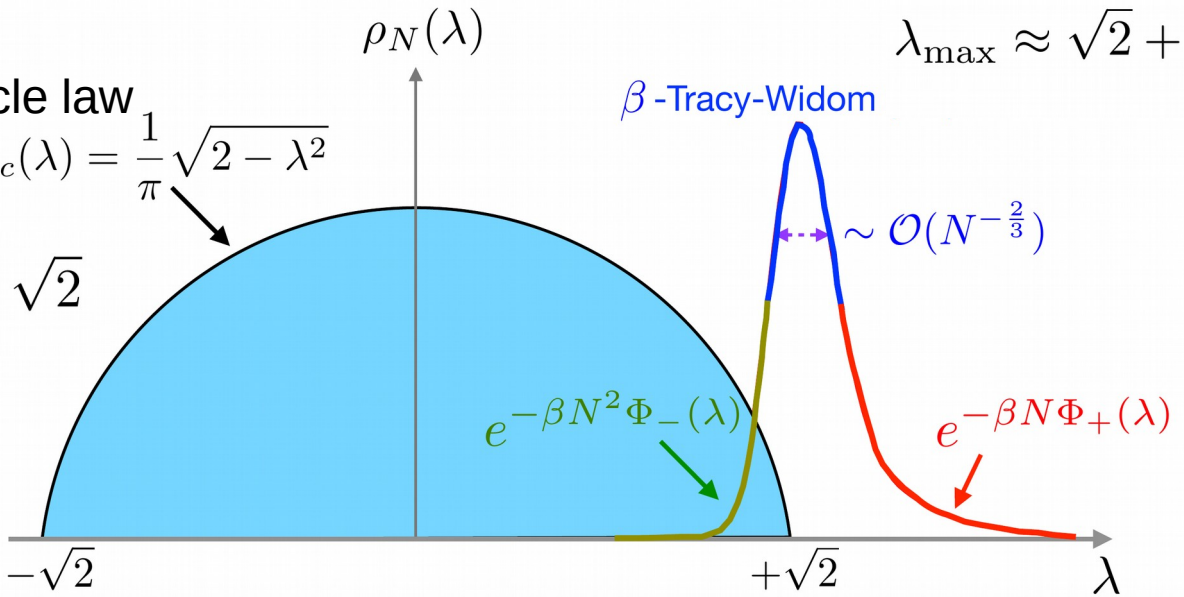
$$P_{\mathrm{joint}}(\lambda_1, \cdots, \lambda_N) = B_N(\beta) \exp\left[-\frac{\beta}{2} N \sum_{i=1}^{N} \lambda_i^2 \right] \prod_{i<j} |\lambda_i - \lambda_j|^\beta$$

Wigner semi-circle law

$$\tilde{\rho}_{sc}(\lambda) = \frac{1}{\pi}\sqrt{2-\lambda^2}$$

$$\lim_{N\to\infty} \langle \lambda_{\max} \rangle = \sqrt{2}$$

$$\lambda_{\max} \approx \sqrt{2} + \frac{1}{\sqrt{2}} N^{-2/3} \chi_\beta$$



S. N. Majumdar, G. Schehr, Top eigenvalue of a random matrix: large deviations and third order phase transition, J. Stat. Mech. P01012 (2014).

Skoltech
Skolkovo Institute of Science and Technology

Extreme value statistics

Other questions

1) Density of near-extremes

2) The time at which the max/min is reached

3) stats of records

Review of the topic with an emphasis of correlated RVs

Extreme value statistics of correlated random variables: A pedagogical review, Satya N.Majumdar, Arnab Pal, Grégory Schehr, Physics Reports, 840, 2020, 1-32 or free version at https://arxiv.org/abs/1910.10667

(And references therein!)