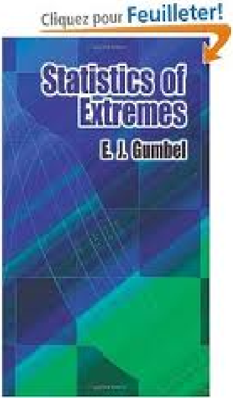
Extreme Value Statistics in Stochastic Processes

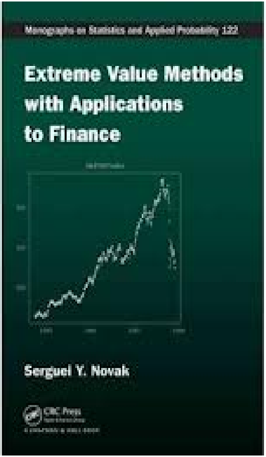
Satya N. Majumdar

Laboratoire de Physique Théorique et Modèles Statistiques,CNRS,
Université Paris-Sud, France

Extreme Events: rare but devastating

Applications

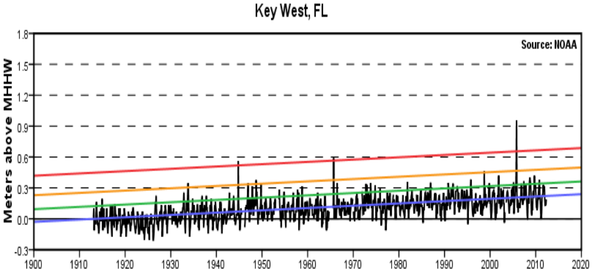


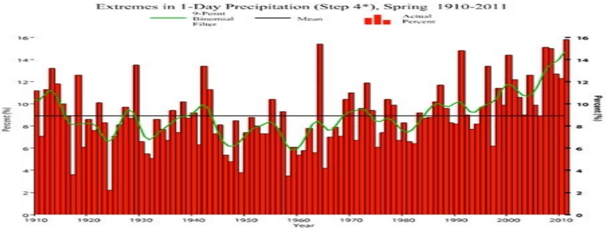


Climate studies, finance and economics, hydrology, sports,....

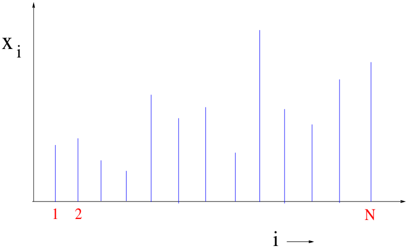
Random walks, disordered systems, random matrices, number theory, ...

Average vs. Extreme





General setting:

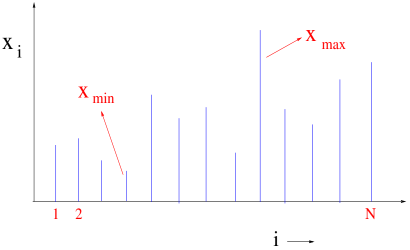


$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

General setting:



$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

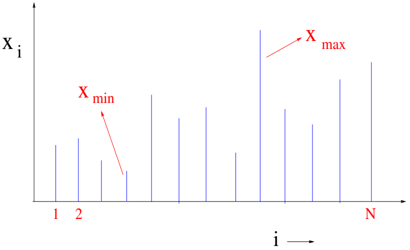
independent or correlated

Extreme Value Statistics: global maximum or minimum

$$x_{\max} = \max\{x_1, x_2, \ldots, x_N\}$$

$$x_{\min} = \min\{x_1, x_2, \ldots, x_N\}$$

General setting:



$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

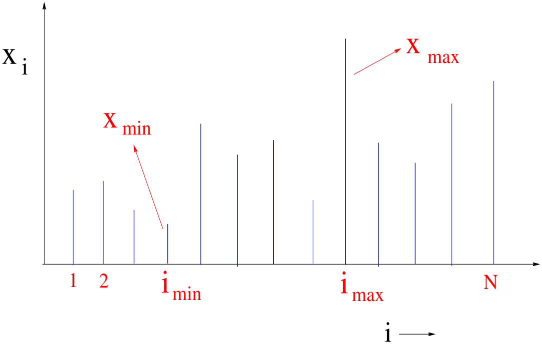
Extreme Value Statistics: global maximum or minimum

$$x_{\max} = \max\{x_1, x_2, \ldots, x_N\}$$

$$x_{\min} = \min\{x_1, x_2, \ldots, x_N\}$$

Q: Given $P(x_1, x_2, \ldots, x_N)$, what can we say about the statistics of $x_{\max}$ and $x_{\min}$?

Other extreme observables: times at which extremes occur



Q: Given $P(x_1, x_2, \ldots, x_N)$, statistics of the times $i_{\max}$ and $i_{\min}$?

Other extreme observables: Order and Gap statistics

Order the random variables: $\{x_1, x_2, \ldots, x_N\} \Rightarrow \{y_1, y_2, \ldots, y_N\}$

such that: $\{y_1 > y_2 > y_3 \ldots > y_N\}$

Other extreme observables: Order and Gap statistics

Order the random variables: $\{x_1, x_2, \ldots, x_N\} \Rightarrow \{y_1, y_2, \ldots, y_N\}$

such that: $\{y_1 > y_2 > y_3 \ldots > y_N\}$

$y_k \to$ k-th maximum $\quad k = 1, 2 \ldots, N$

Other extreme observables: Order and Gap statistics

Order the random variables: $\{x_1, x_2, \ldots, x_N\} \Rightarrow \{y_1, y_2, \ldots, y_N\}$

such than: $\{y_1 > y_2 > y_3 \ldots > y_N\}$

$y_k \to$ k-th maximum $\quad k = 1, 2 \ldots, N$

Given $P(x_1, x_2, \ldots, x_N)$:

Order statistics: statistics of the k-th maximum y_k?

Other extreme observables: Order and Gap statistics

Order the random variables: $\{x_1, x_2, \ldots, x_N\} \Rightarrow \{y_1, y_2, \ldots, y_N\}$

such that: $\{y_1 > y_2 > y_3 \ldots > y_N\}$

$y_k \rightarrow$ k-th maximum $\quad k = 1, 2 \ldots, N$

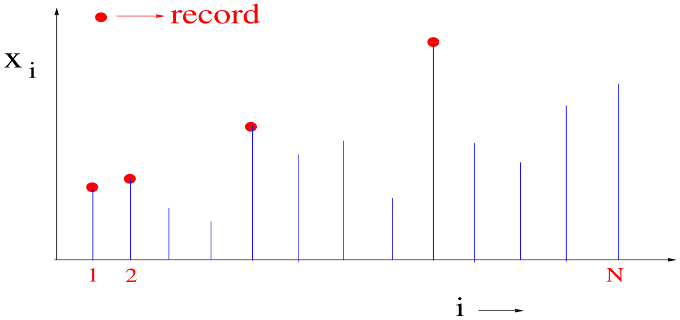
Given $P(x_1, x_2, \ldots, x_N)$:

Order statistics: statistics of the k-th maximum y_k?

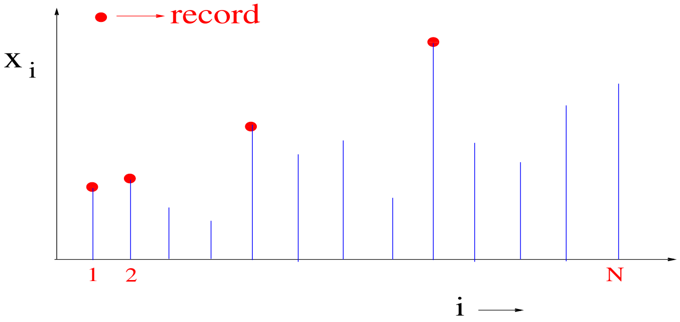
Gap statistics:

$g_k = y_k - y_{k+1} \Rightarrow$ gap between the k-th and (k+1)-th maximum

Other extreme observables: Record statistics



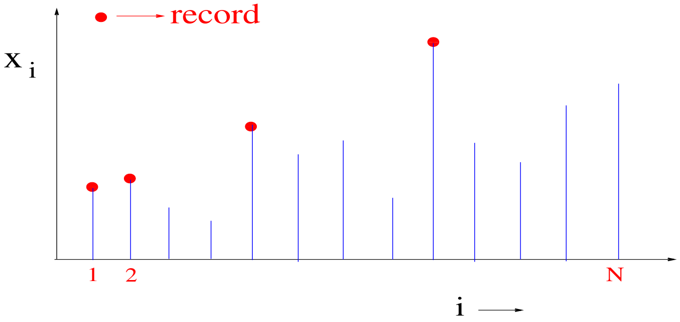
Other extreme observables: Record statistics



A record happens at step k if

$$\boxed{x_k > \{x_1, x_2, \ldots, x_{k-1}\}}$$

Other extreme observables: Record statistics

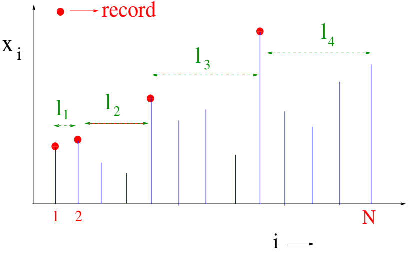


A record happens at step k if

$$\boxed{x_k > \{x_1, x_2, \ldots, x_{k-1}\}}$$

or equivalently: $x_k > \max\left[x_1, x_2, \ldots, x_{k-1}\right]$

Record statistics for a random sequence

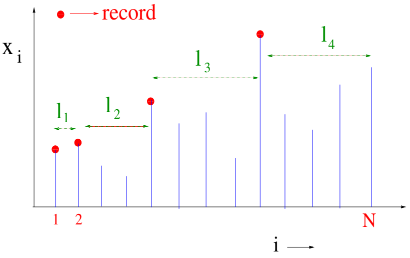


$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

Record statistics for a random sequence



$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

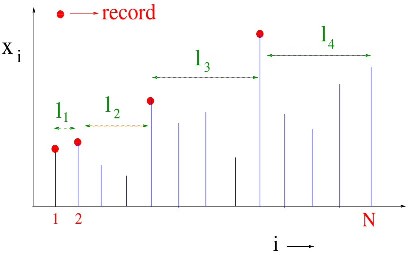
$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

- $R_N \to$ no. of records in step N;

statistics of R_N?

Record statistics for a random sequence



$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

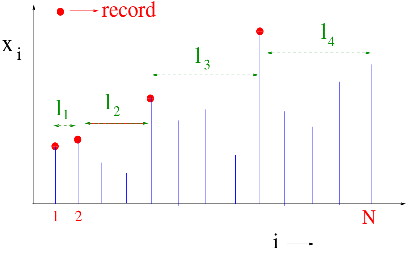
- $R_N \to$ no. of records in step N;

statistics of R_N?

- How long does a record survive?

Ages of records: $\{l_1, l_2, \ldots, l_{R_N}\}$

Record statistics for a random sequence



$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ random variables drawn from a joint pdf

$$P(x_1, x_2, \ldots, x_N)$$

independent or correlated

- $R_N \to$ no. of records in step N; statistics of R_N?
- How long does a record survive? Ages of records: $\{l_1, l_2, \ldots, l_{R_N}\}$
- Age of the longest (shortest) record?

$$l_{\max} = \max\left[l_1, l_2, \ldots, l_{R_N}\right]$$

$$l_{\min} = \min\left[l_1, l_2, \ldots, l_{R_N}\right]$$

Extreme statistics of i.i.d random variables

A particularly simple case is when

$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ set of N i.i.d random variables

each drawn from $p(x) \to P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)$

Extreme statistics of i.i.d random variables

A particularly simple case is when

$\{x_1, x_2, \ldots, x_N\} \Longrightarrow$ set of N i.i.d random variables

each drawn from $p(x) \to P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)$

In this case, several extreme observables can be computed explicitly:

- limiting laws (large N) of $x_{\max}$ and $x_{\min}$ (Fréchet, Gumbel and Weibull)
- statistics of $i_{\max}$ and $i_{\min}$
- Order and Gap statistics
- Record statistics
- ...

For a recent review, see G. Schehr and S.M., arXiv:1305.0639

Extreme value statistics of i.i.d random variables

i.i.d $\Longrightarrow$ $$\boxed{P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)}$$

Extreme value statistics of i.i.d random variables

i.i.d $\Longrightarrow$ $$\boxed{P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)}$$

Maximum: $$x_{\rm max} = \max(x_1, x_2, \ldots, x_N)$$

Extreme value statistics of i.i.d random variables

i.i.d $\Longrightarrow$

$$\boxed{P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)}$$

Maximum: $$x_{\max} = \max(x_1, x_2, \ldots, x_N)$$

Cumulative dist. of the maximum:

$$Q_N(x) = \mathrm{Prob}[x_{\max} \le x] = \mathrm{Prob}[x_1 \le x, x_2 \le x, \ldots x_N \le x]$$

Extreme value statistics of i.i.d random variables

i.i.d $\Longrightarrow$

$$\boxed{P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)}$$

Maximum: $x_{\max} = \max(x_1, x_2, \ldots, x_N)$

Cumulative dist. of the maximum:

$$Q_N(x) = \mathrm{Prob}[x_{\max} \le x] = \mathrm{Prob}[x_1 \le x, x_2 \le x, \ldots x_N \le x]$$

Independence $\Longrightarrow$ $Q_N(x) = \left[\int_{-\infty}^{x} p(x')\, dx'\right]^N = \left[1 - \int_{x}^{\infty} p(x')\, dx'\right]^N$

Extreme value statistics of i.i.d random variables

i.i.d $\Longrightarrow$ $$\boxed{P(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(x_i)}$$

Maximum: $$x_{\max} = \max(x_1, x_2, \ldots, x_N)$$

Cumulative dist. of the maximum:

$$Q_N(x) = \mathrm{Prob}[x_{\max} \le x] = \mathrm{Prob}[x_1 \le x, x_2 \le x, \ldots x_N \le x]$$

Independence $\Longrightarrow$ $Q_N(x) = \left[\int_{-\infty}^{x} p(x')\, dx'\right]^N = \left[1 - \int_{x}^{\infty} p(x')\, dx'\right]^N$

Scaling limit: N large, x large: $\boxed{Q_N(x) \to F\left[(x - a_N)/b_N\right]}$

Three universal extreme value distributions

Scale factors a_N and $b_N \Longrightarrow$ Non-universal (depends on the precise tail of $p(x)$)

Three universal extreme value distributions

Scale factors a_N and $b_N \Longrightarrow$ Non-universal (depends on the precise tail of $p(x)$)

But only 3 possible varieties of scaling functions $F(z)$ (depending only on the generic tail of $p(x)$)

$\Longrightarrow$ LAW OF EXTREMES

Three universal extreme value distributions

Scale factors a_N and $b_N \Longrightarrow$ Non-universal (depends on the precise tail of $p(x)$)

But only 3 possible varieties of scaling functions $F(z)$ (depending only on the generic tail of $p(x)$)

$\Longrightarrow$ LAW OF EXTREMES

[Fréchet (1927), Fisher and Tippet (1928), Gnedenko (1943), Gumbel (1958)...]

Three universal extreme value distributions

Scale factors a_N and b_N $\Longrightarrow$ Non-universal (depends on the precise tail of $p(x)$)

But only 3 possible varieties of scaling functions $F(z)$ (depending only on the generic tail of $p(x)$)

$\Longrightarrow$ LAW OF EXTREMES

[Fréchet (1927), Fisher and Tippet (1928), Gnedenko (1943), Gumbel (1958)...]

Several applications $\Longrightarrow$ Climate, Finance, Oceanography, Disordered Systems (Random Energy Model of Derrida),.....

Three canonical extreme value distributions

Type I (GUMBEL): If $p(x)$ is unbounded with faster than power law tail (e.g., exponential)

$$F_I(z) = \exp[-e^{-z}]$$

Three canonical extreme value distributions

Type I (GUMBEL): If $p(x)$ is unbounded with faster than power law tail (e.g., exponential)

$$F_I(z) = \exp[-e^{-z}]$$

Type II (FRÉCHET): If $p(x)$ has power law tails: $p(x) \sim x^{-(\gamma+1)}$

$$\begin{aligned} F_{II}(z) &= 0 & z \leq 0 \\ &= \exp[-z^{-\gamma}] & z \geq 0 \end{aligned}$$

Three canonical extreme value distributions

Type I (GUMBEL): If $p(x)$ is unbounded with faster than power law tail (e.g., exponential)

$$F_I(z) = \exp[-e^{-z}]$$

Type II (FRÉCHET): If $p(x)$ has power law tails: $p(x) \sim x^{-(\gamma+1)}$

$$\begin{aligned} F_{II}(z) &= 0 && z \leq 0 \\ &= \exp[-z^{-\gamma}] && z \geq 0 \end{aligned}$$

Type III (WEIBULL): If $p(x)$ is bounded: $p(x) \sim (1-x)^{(\gamma-1)}$

$$\begin{aligned} F_{III}(z) &= \exp[-|z|^{\gamma}] && z \leq 0 \\ &= 1 && z \geq 0 \end{aligned}$$

Three canonical extreme value distributions

Type I (GUMBEL): If $p(x)$ is unbounded with faster than power law tail (e.g., exponential)

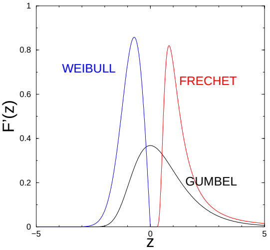
$$F_I(z) = \exp[-e^{-z}]$$

Type II (FRÉCHET): If $p(x)$ has power law tails: $p(x) \sim x^{-(\gamma+1)}$

$$\begin{aligned} F_{II}(z) &= 0 & z \leq 0 \\ &= \exp[-z^{-\gamma}] & z \geq 0 \end{aligned}$$

Type III (WEIBULL): If $p(x)$ is bounded: $p(x) \sim (1-x)^{(\gamma-1)}$

$$\begin{aligned} F_{III}(z) &= \exp[-|z|^{\gamma}] & z \leq 0 \\ &= 1 & z \geq 0 \end{aligned}$$



Extreme statistics of correlated variables

In many situations, however, the underlying random variables

$$\{x_1, x_2, \ldots, x_N\} \Rightarrow \text{correlated}$$

Joint distribution is not factorisable: $P(x_1, x_2, \ldots, x_N) \neq \prod_{i=1}^{N} p(x_i)$

Extreme statistics of correlated variables

In many situations, however, the underlying random variables

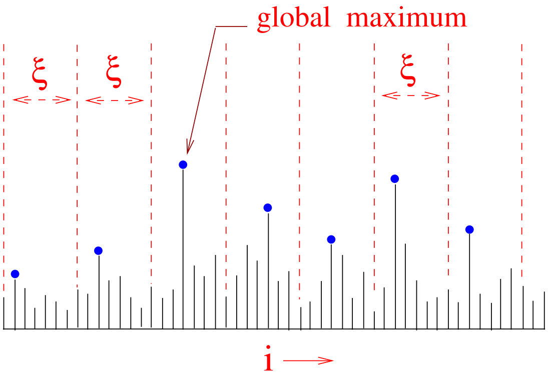
$$\{x_1, x_2, \ldots, x_N\} \Rightarrow \text{correlated}$$

Joint distribution is not factorisable: $P(x_1, x_2, \ldots, x_N) \neq \prod_{i=1}^{N} p(x_i)$

Extreme statistics of correlated variables $\Rightarrow$ nontrivial

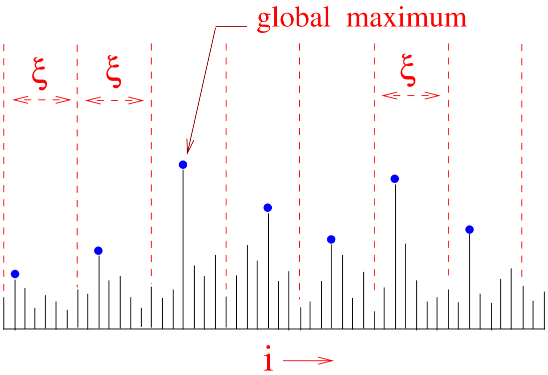
Extreme statistics in weakly correlated systems

Weakly correlated variables $\{x_1, x_2, \ldots, x_N\}$
$\rightarrow$ finite correlation length $\xi << N$



Extreme statistics in weakly correlated systems

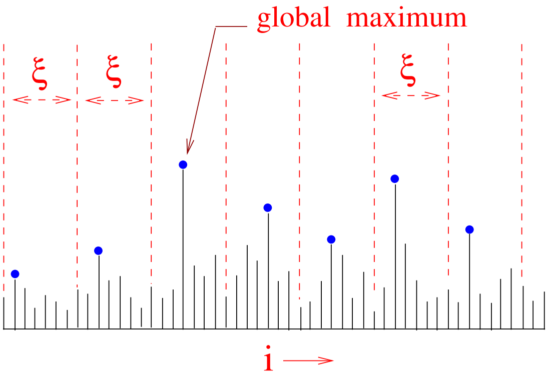
Weakly correlated variables $\{x_1, x_2, \ldots, x_N\}$
$\rightarrow$ finite correlation length $\xi << N$



- z_i → maximum in the i-th block ⇒ uncorrelated

Extreme statistics in weakly correlated systems

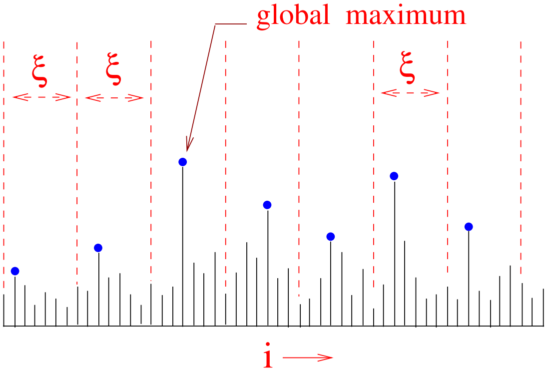
Weakly correlated variables $\{x_1, x_2, \ldots, x_N\}$
$\rightarrow$ finite correlation length $\xi << N$



- $z_i \rightarrow$ maximum in the i-th block $\Rightarrow$ uncorrelated
- Global maximum: $x_{\max} = \max(z_1, z_2, \ldots)$

Extreme statistics in weakly correlated systems

Weakly correlated variables $\{x_1, x_2, \ldots, x_N\}$
$\rightarrow$ finite correlation length $\xi << N$



- $z_i \rightarrow$ maximum in the i-th block $\Rightarrow$ uncorrelated
- Global maximum: $x_{\max} = \max(z_1, z_2, \ldots)$

$\Longrightarrow$ Fréchet, Gumbel or Weibull

Extreme statistics in strongly correlated systems

For strongly correlated $\{x_1, x_2, \ldots, x_N\}$: correlation length $\xi \sim O(N)$

Extreme statistics $\rightarrow$ nontrivial

$\Longrightarrow$ few exact results

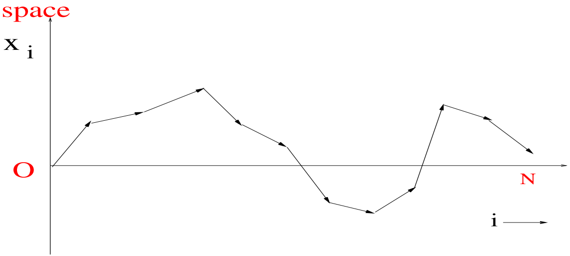
Extreme statistics in strongly correlated systems

For strongly correlated $\{x_1, x_2, \ldots, x_N\}$: correlation length $\xi \sim O(N)$

Extreme statistics $\rightarrow$ nontrivial

$\Longrightarrow$ few exact results

Example: Random Walks/Lévy flights in 1-d

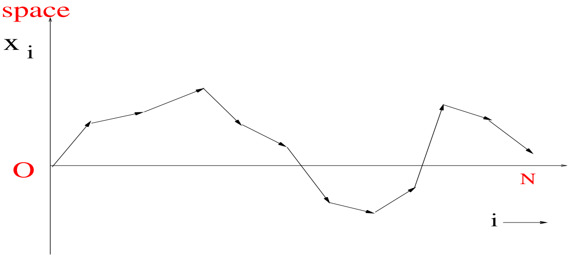


Discrete-time random walk on a line:

$$\boxed{x_i = x_{i-1} + \eta_i, \quad x_0 = 0}$$

$\eta_i \rightarrow$ i.i.d jump lengths, each drawn from a symmetric $p(\eta)$

Example: Random Walks/Lévy flights in 1-d



Discrete-time random walk on a line:

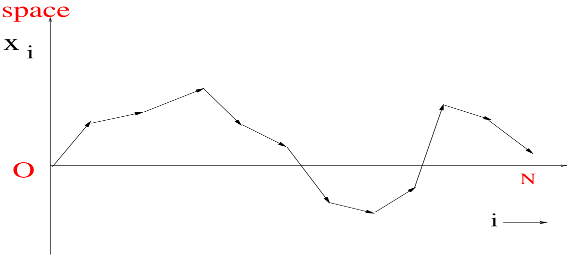
$$\boxed{x_i = x_{i-1} + \eta_i, \quad x_0 = 0}$$

$\eta_i \rightarrow$ i.i.d jump lengths, each drawn from a symmetric $p(\eta)$

if $\sigma^2 = \int \eta^2\, p(\eta)\, d\eta$ is finite $\rightarrow$ normal walk

if $p(\eta) \sim |\eta|^{-1-\mu}$ as $|\eta| \rightarrow \infty$ with $0 < \mu < 2 \rightarrow$ Lévy flight

Example: Random Walks/Lévy flights in 1-d



Discrete-time random walk on a line:

$$\boxed{x_i = x_{i-1} + \eta_i, \quad x_0 = 0}$$

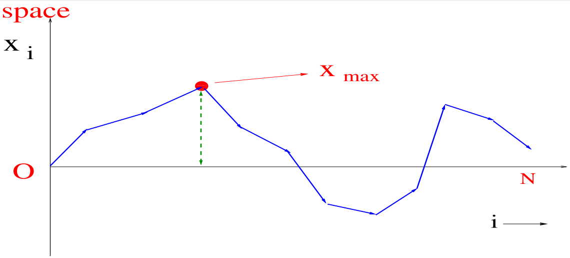
$\eta_i \rightarrow$ i.i.d jump lengths, each drawn from a symmetric $p(\eta)$

if $\sigma^2 = \int \eta^2\, p(\eta)\, d\eta$ is finite $\rightarrow$ normal walk

if $p(\eta) \sim |\eta|^{-1-\mu}$ as $|\eta| \rightarrow \infty$ with $0 < \mu < 2$ $\rightarrow$ Lévy flight

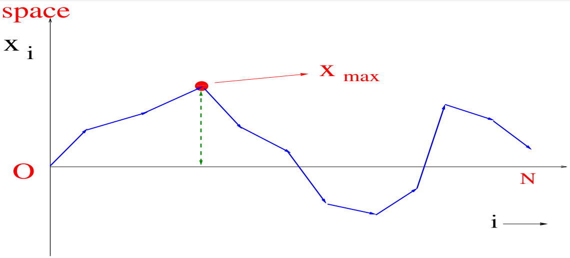
Even though the increments η_i's are uncorrelated, the position x_i's are strongly correlated

Extreme statistics of Random walks/Lévy flights



Applications: fluctuating interfaces, disordered systems (Sinai model), finance, ecology,....

Extreme statistics of Random walks/Lévy flights



Applications: fluctuating interfaces, disordered systems (Sinai model), finance, ecology,....

Extreme statistics and related questions: Record statistics, Order and Gap statistics, statistics of maximal relative height, Convex hulls in 2-d, etc. have been studied recently with many nontrivial results

[Comtet, Dumonteil, Godrèche, Kearney, Mounaix, Rosso, Randon-Furling, Sabhapandit, Schehr, Wergen, Ziff, Zoia...+ S.M.]

for a recent review see, G. Schehr and S.M., arXiv:1305.0639

Extreme statistics in physics of disordered systems

Extreme statistics is also important in disordered systems: spin glasses, polymers in disordered medium, combinatorial optimization, ...

(Bouchaud & Mézard, '97)

Extreme statistics in physics of disordered systems

Extreme statistics is also important in disordered systems: spin glasses, polymers in disordered medium, combinatorial optimization, ...

(Bouchaud & Mézard, '97)

- underlying random variables $\rightarrow$ strongly correlated

Extreme statistics in physics of disordered systems

Extreme statistics is also important in disordered systems: spin glasses, polymers in disordered medium, combinatorial optimization, ...

(Bouchaud & Mézard, '97)

- underlying random variables $\rightarrow$ strongly correlated
- Example: directed polymer in a random medium

Extreme statistics in physics of disordered systems

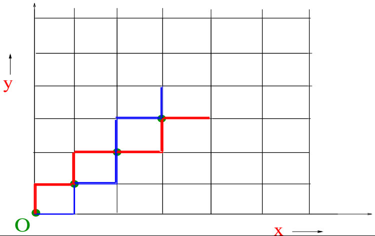
Extreme statistics is also important in disordered systems: spin glasses, polymers in disordered medium, combinatorial optimization, ...

(Bouchaud & Mézard, '97)

- underlying random variables $\rightarrow$ strongly correlated
- Example: directed polymer in a random medium

At each site i, $\epsilon_i \rightarrow$ quenched energy (each drawn independently from a distribution $\rho(\epsilon)$) $\Longrightarrow$ defines the random medium

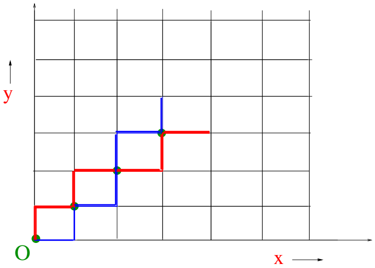
Consider now n-step directed (up-right) polymer paths starting from O



$$E_{\text{path-blue}} = \sum_{j \in \text{path-blue}} \epsilon_j$$

$$E_{\text{path-red}} = \sum_{j \in \text{path-red}} \epsilon_j$$

Directed polymer in a disordered medium



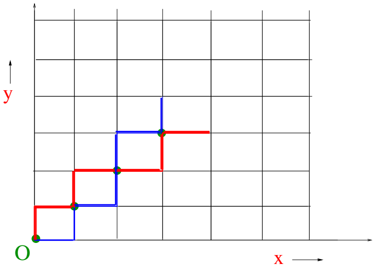
$$E_{\rm path-blue} = \sum_{j\in {\rm path-blue}} \epsilon_j$$

$$E_{\rm path-red} = \sum_{j\in {\rm path-red}} \epsilon_j$$

For fixed disorder ϵ_i's: $E_{\rm path} = \sum_{j\in {\rm path}} \epsilon_j$

Ground state energy: $E_0 = \min_{\rm paths} \left[E_{\rm path-1}, E_{\rm path-2}, ..., E_{\rm path-2^n}\right]$

Directed polymer in a disordered medium



$$E_{\rm path-blue} = \sum_{j \in {\rm path-blue}} \epsilon_j$$

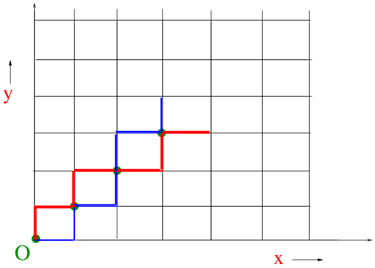
$$E_{\rm path-red} = \sum_{j \in {\rm path-red}} \epsilon_j$$

For fixed disorder ϵ_i's: $E_{\rm path} = \sum_{j \in {\rm path}} \epsilon_j$

Ground state energy: $E_0 = \min_{\rm paths} \left[E_{\rm path-1}, E_{\rm path-2}, \ldots, E_{\rm path-2^n}\right]$

Clearly, energies of different polymer paths are strongly correlated as they share some common ϵ_i's (sites common to both paths)

Directed polymer in a disordered medium



$$E_{\rm path-blue} = \sum_{j \in {\rm path-blue}} \epsilon_j$$

$$E_{\rm path-red} = \sum_{j \in {\rm path-red}} \epsilon_j$$

For fixed disorder ϵ_i's: $E_{\rm path} = \sum_{j \in {\rm path}} \epsilon_j$

Ground state energy: $E_0 = \min_{\rm paths} \left[E_{\rm path-1}, E_{\rm path-2}, \ldots, E_{\rm path-2^n} \right]$

Clearly, energies of different polymer paths are strongly correlated as they share some common ϵ_i's (sites common to both paths)

Baik, Comets, Corwin, Deift, Johansson, Johnstone, Péché, Quastel, Rains,....

Brunet, Calabrese, Derrida, Dotsenko, Le Doussal, Rambeau, Rosso, Sasamoto, Schehr, Spohn,....

Extreme statistics in correlated systems

For correlated $\{x_1, x_2, \ldots, x_N\}$ variables: $P(x_1, x_2, \ldots, x_N) \neq \prod_{i=1}^{N} p(x_i)$

Extreme statistics $\rightarrow$ nontrivial

$\Longrightarrow$ few exact results

Extreme statistics in correlated systems

For correlated $\{x_1, x_2, \ldots, x_N\}$ variables: $P(x_1, x_2, \ldots, x_N) \neq \prod_{i=1}^{N} p(x_i)$

Extreme statistics $\rightarrow$ nontrivial

$\Longrightarrow$ few exact results

Some examples that will be discussed today:

- Logarithmically correlated Gaussian random variables and Liouville field theory

  talks by Keating, Le Doussal, Rosso

- Extreme statistics in Random Matrices

  talks by Keating, Péché

- Branching Brownian motion in 1-d

  talks by Derrida, Schehr, Shi