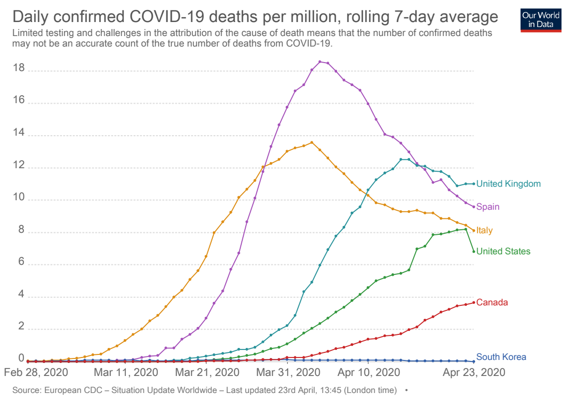
Extreme value statistics and the theory of rare events

Francesco Mori

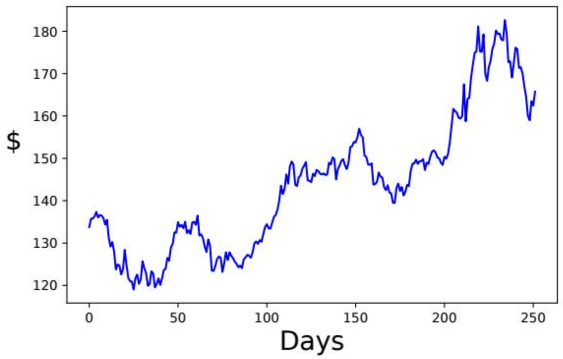
Saturday Mornings of Theoretical Physics
25/02/2023

Extreme value statistics: motivation

Applications to epidemics[1], finance [2], climate science [3]



From Our World in Data

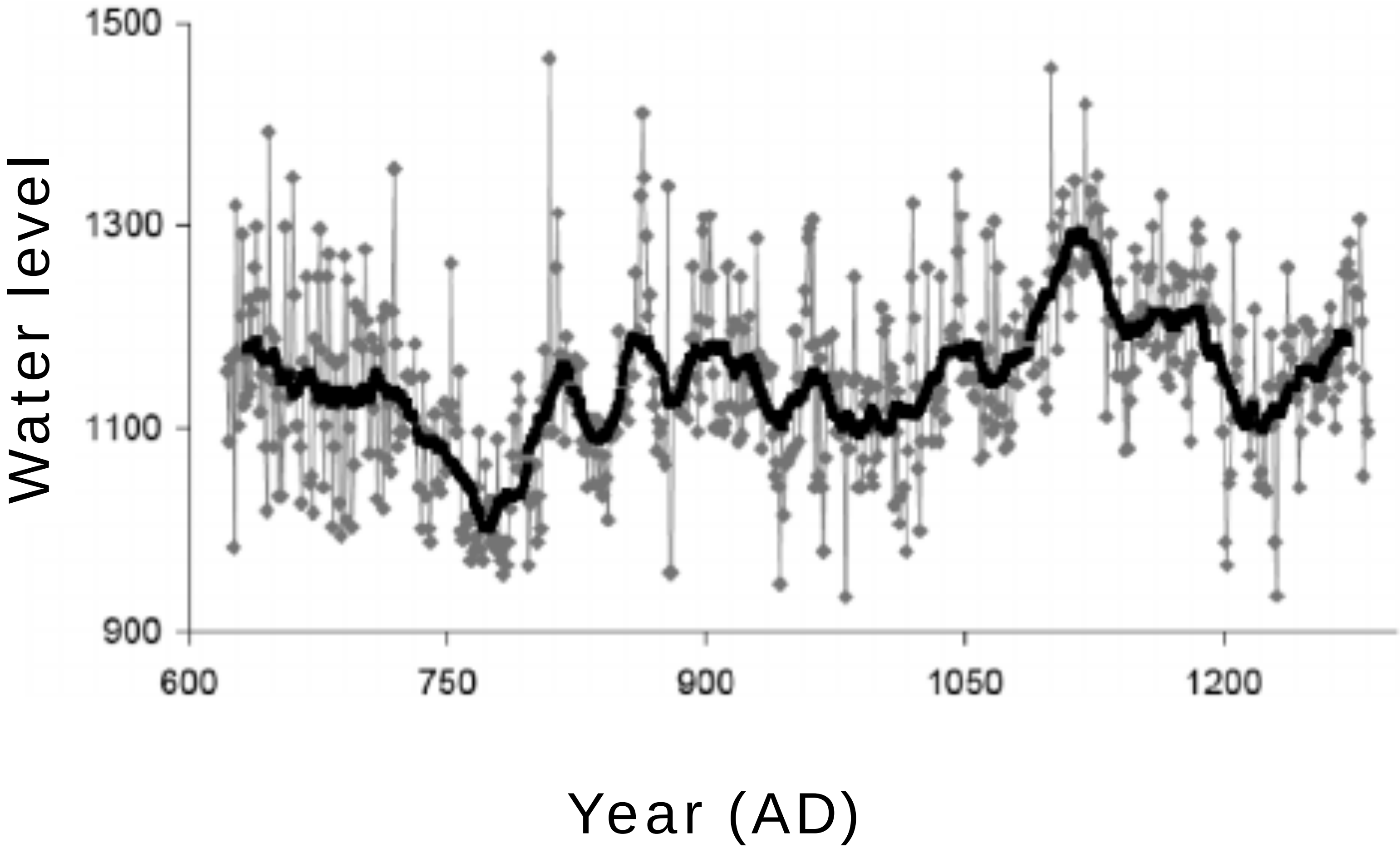


Apple price Jan 2021-Jan 2022 (from Yahoo finance)

From Wikipedia

[1] M. Marani, Marco, et al., *Proceedings of the National Academy of Sciences* **118**: e2105482118 (2021).
[2] S.N. Majumdar and J.-P. Bouchaud, Quantitative Finance 8, 753 (2008).
[3] G. Wergen and J. Krug, Europhys. Lett. 92, 30008 (2010).

Extreme value statistics: an example



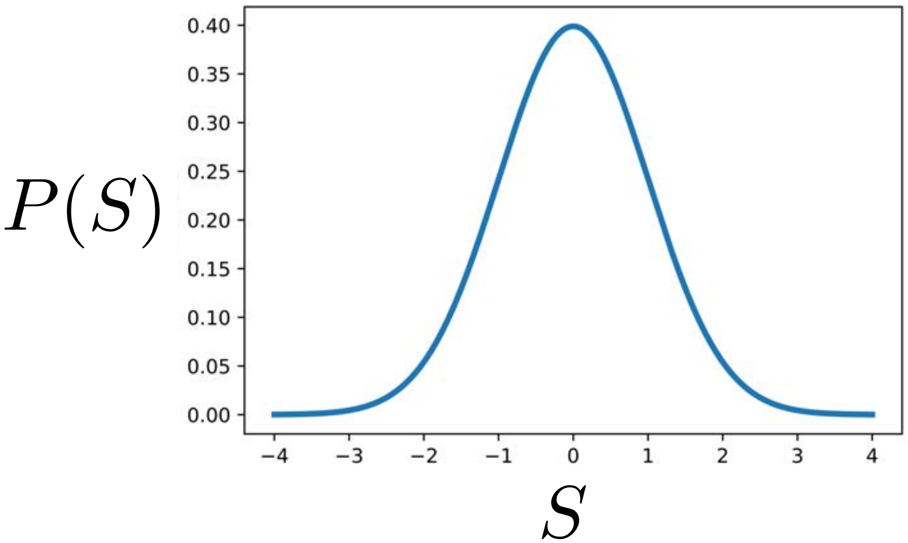
Di Baldassarre, Giuliano, et al., *Hydrological Sciences Journal,* 56.2 (2011): 199-211.

Central Limit Theorem

$$S_N = X_1 + X_2 + \ldots + X_N$$

$$P(S) = \text{Prob.}(S_N = S) \approx \frac{1}{\sqrt{2\pi\sigma^2 N}} \exp\left(-\frac{S^2}{2\sigma^2 N}\right)$$

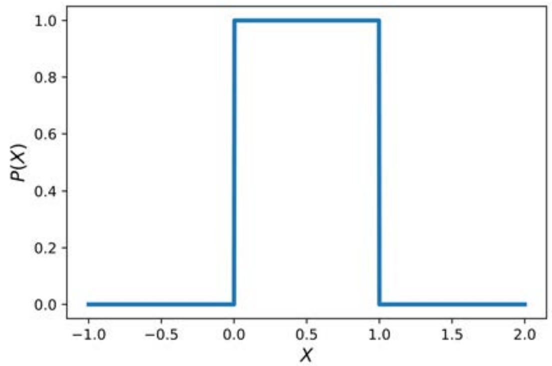
UNIVERSAL!



Gaussian distribution
(a.k.a. Normal distribution)

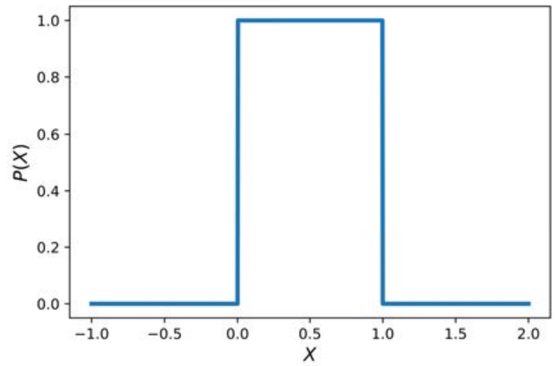
Central Limit Theorem
Numerical example

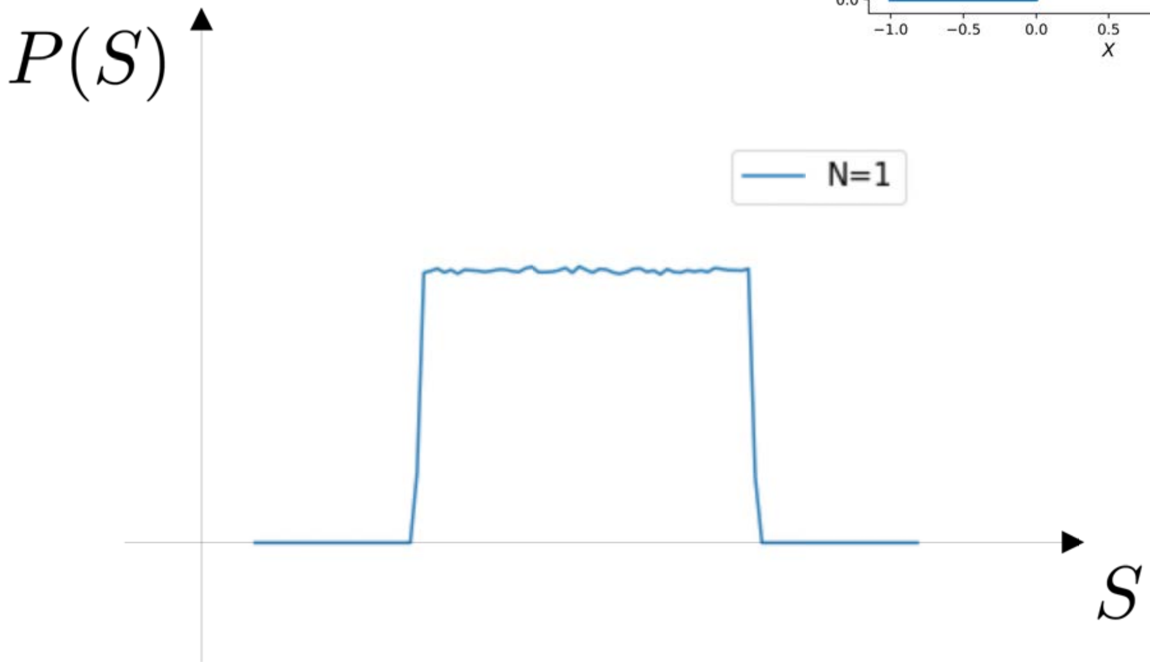
$$p(X) = \begin{cases} 1 & \text{if } 0 < x < 1\,, \\ 0 & \text{otherwise} \end{cases}$$



Central Limit Theorem
Numerical example

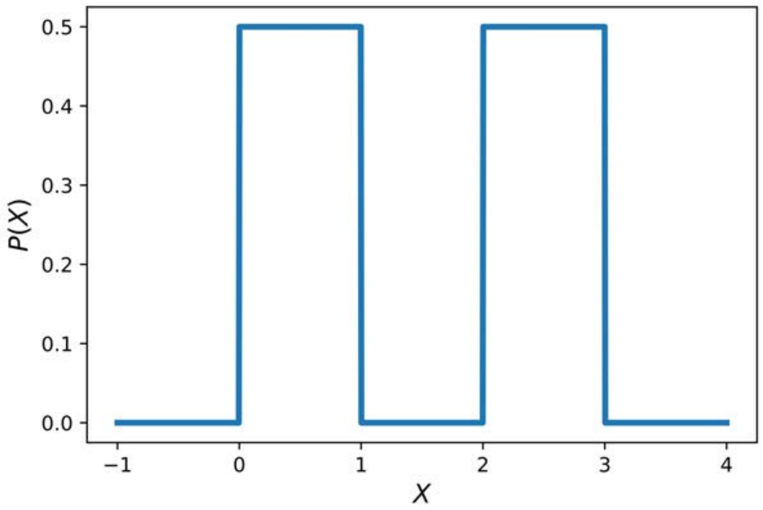
$$p(X) = \begin{cases} 1 & \text{if } 0 < x < 1\,, \\ 0 & \text{otherwise} \end{cases}$$





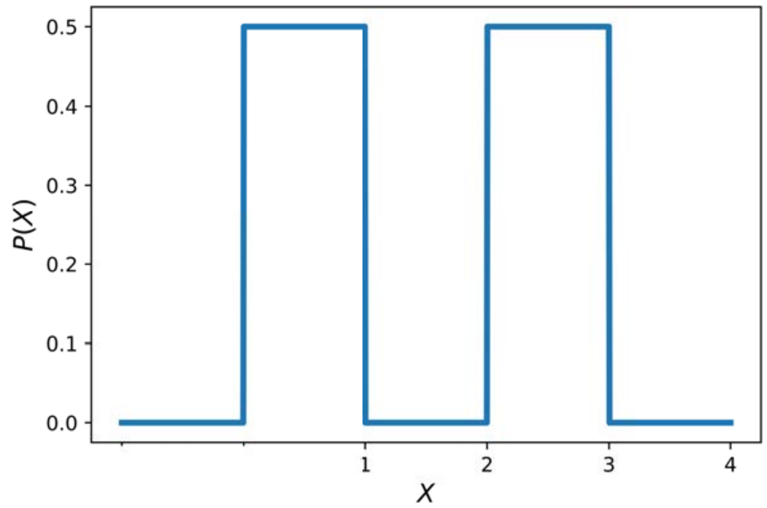
Central Limit Theorem
Numerical example

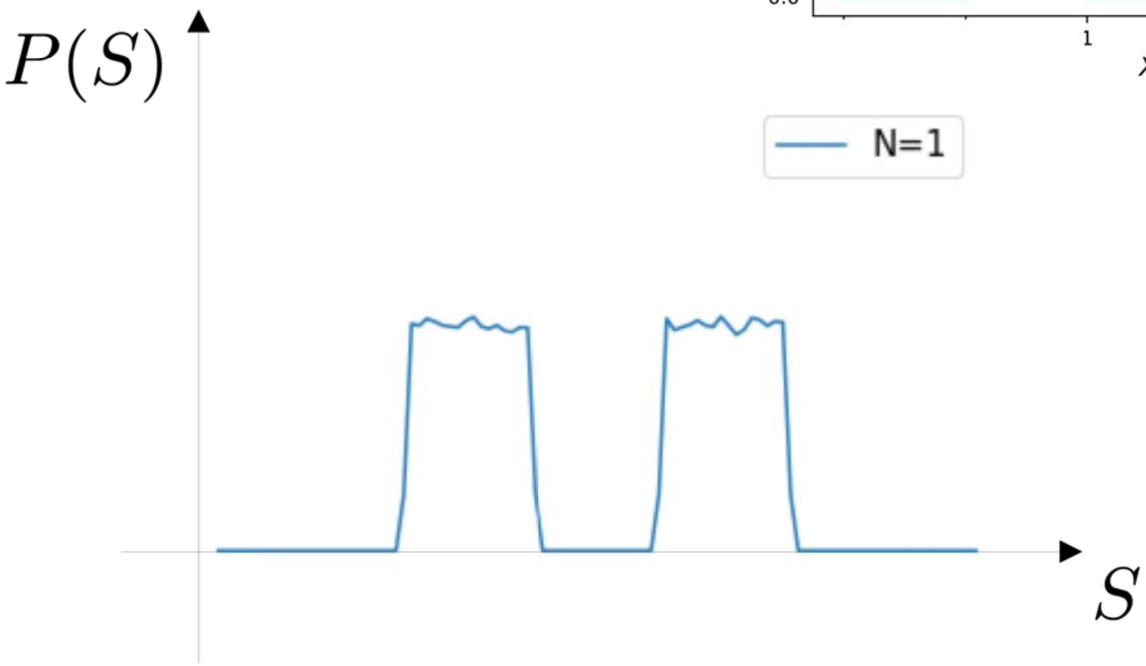
$$p(X) = \begin{cases} 1/2 & \text{if } 2 < x < 3\,, \\ 1/2 & \text{if } 0 < x < 1\,, \\ 0 & \text{otherwise} \end{cases}$$



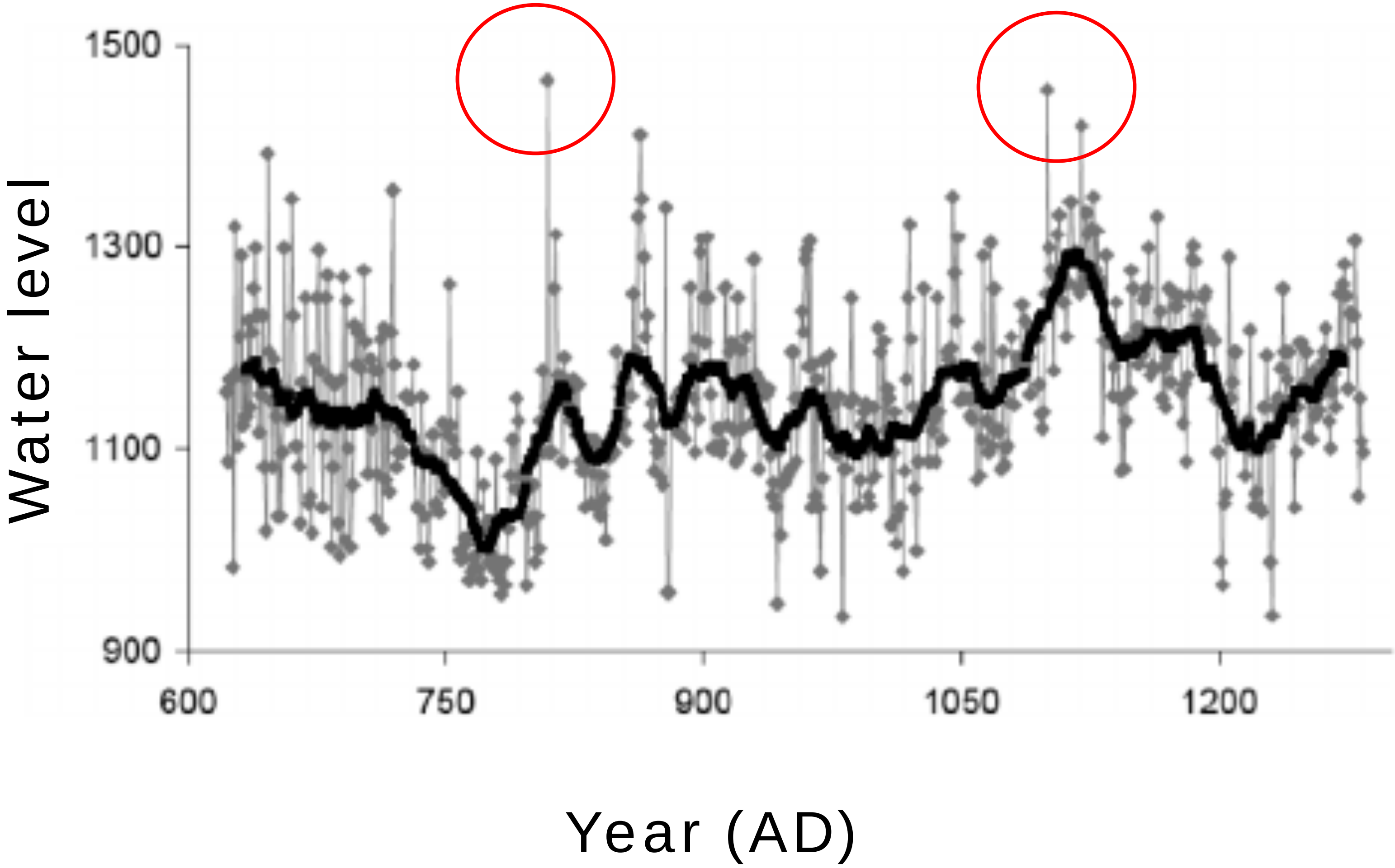
Central Limit Theorem
Numerical example

$$p(X) = \begin{cases} 1/2 & \text{if } 2 < x < 3\,, \\ 1/2 & \text{if } 0 < x < 1\,, \\ 0 & \text{otherwise} \end{cases}$$



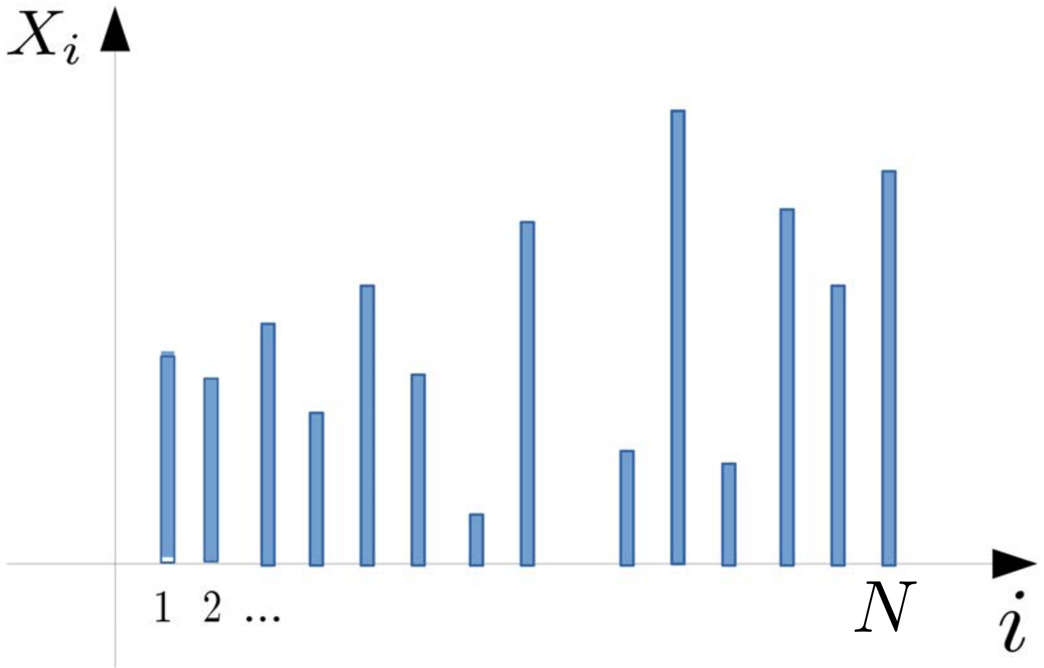


Extreme value statistics: an example



Di Baldassarre, Giuliano, et al., *Hydrological Sciences Journal,* 56.2 (2011): 199-211.

Extreme value statistics: setting



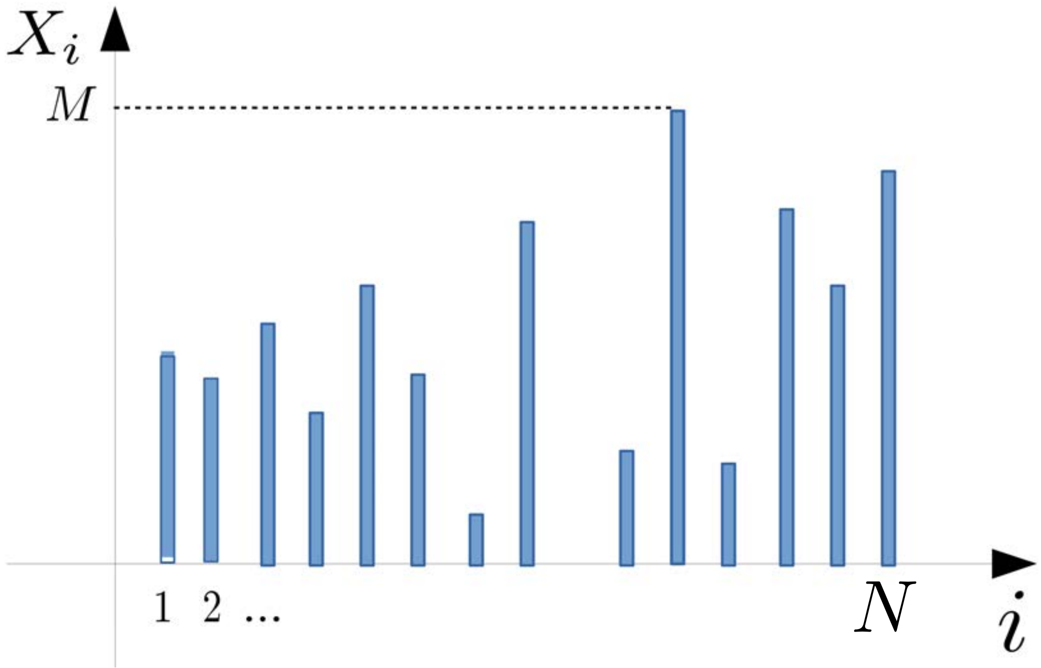
$$X_1\,, X_2\,, \ldots\,, X_N$$

random variables with distribution

$$P(X_1\,, \ldots\,, X_N)$$

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

Global Maximum

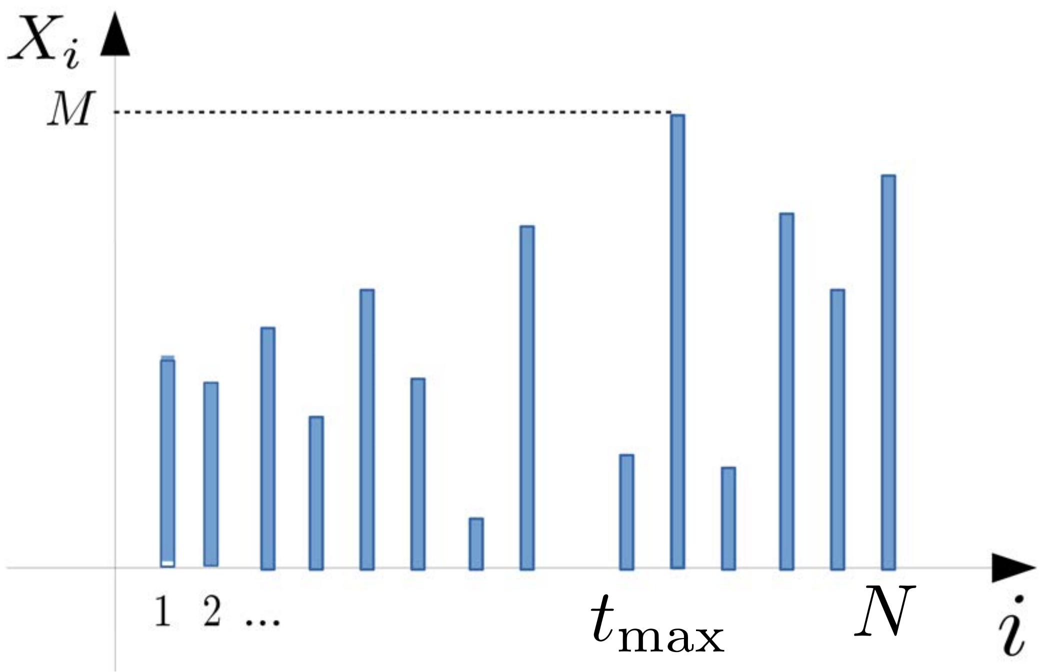


$$M = \max_{1 \le i \le N} X_i$$

Given $P(X_1, \ldots, X_N)$ what can I say about the distribution of *M*?

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

Time of the maximum



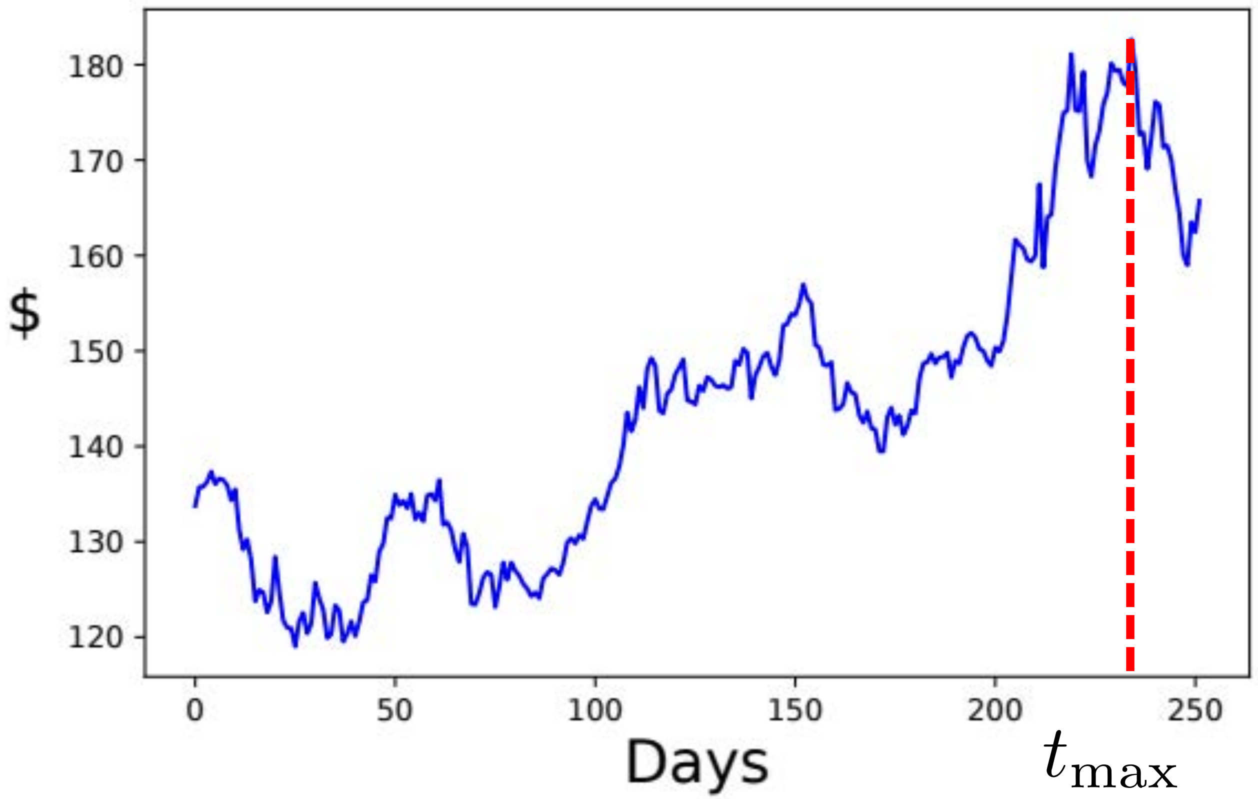
$$t_{\max} = \operatorname{argmax}_{1 \le i \le N} X_i$$

Given $P(X_1, \ldots, X_N)$ what can I say about the distribution of t_{max}?

Applications to finance, disordered systems, sports...

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

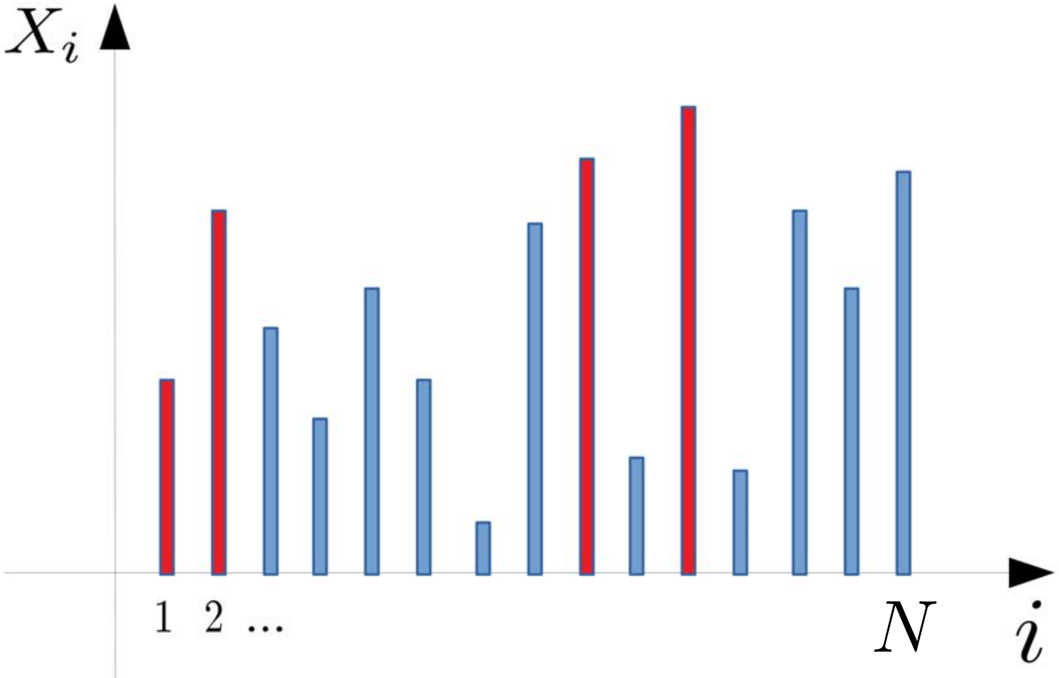
Time of the maximum

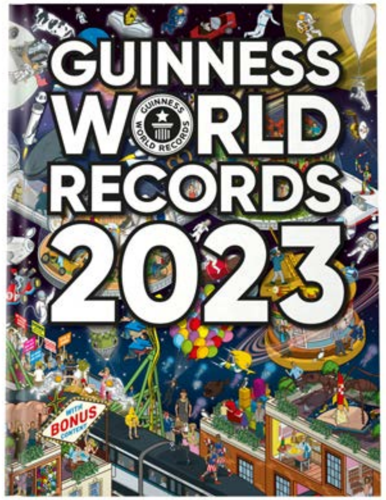


Apple price Jan 2021-Jan 2022
(from Yahoo finance)

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

Record statistics



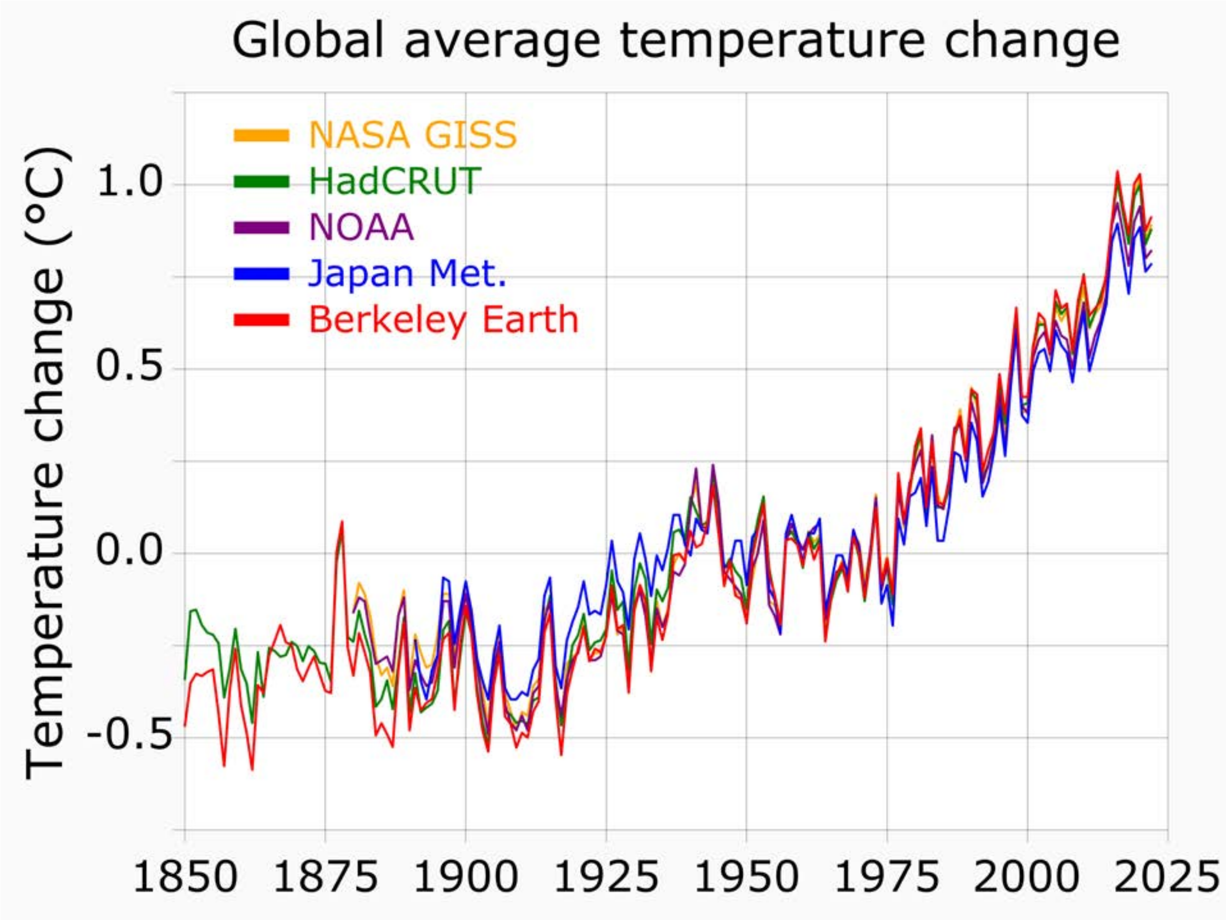


$$X_i \text{ is a record if } X_i > \{X_1\,,, X_2\,, \dots\,, X_{i-1}\}$$

Applications to climate science, sports, evolution, insurance policies,...

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

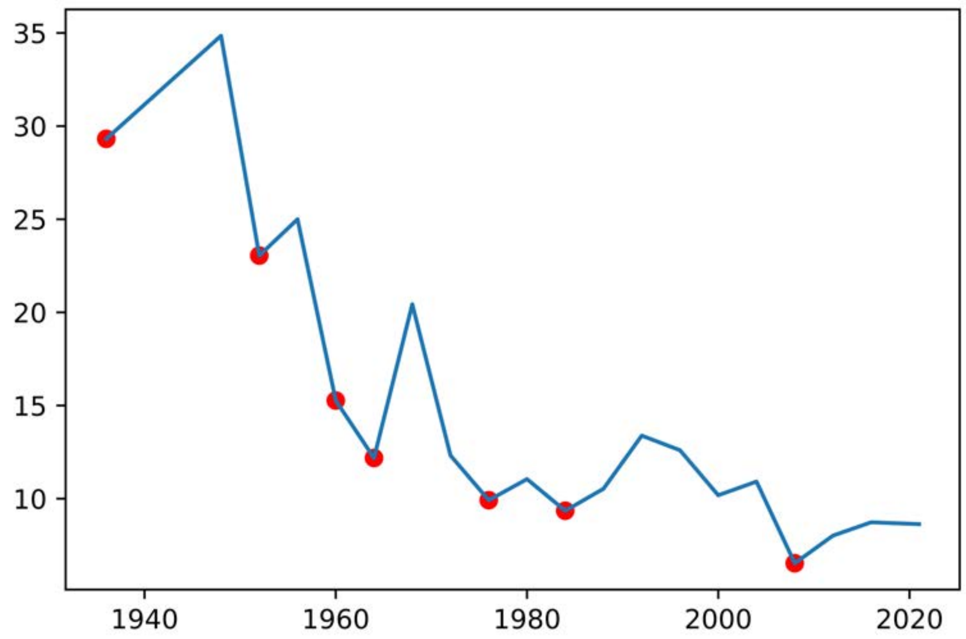
Record statistics in climate science



From Wikipedia

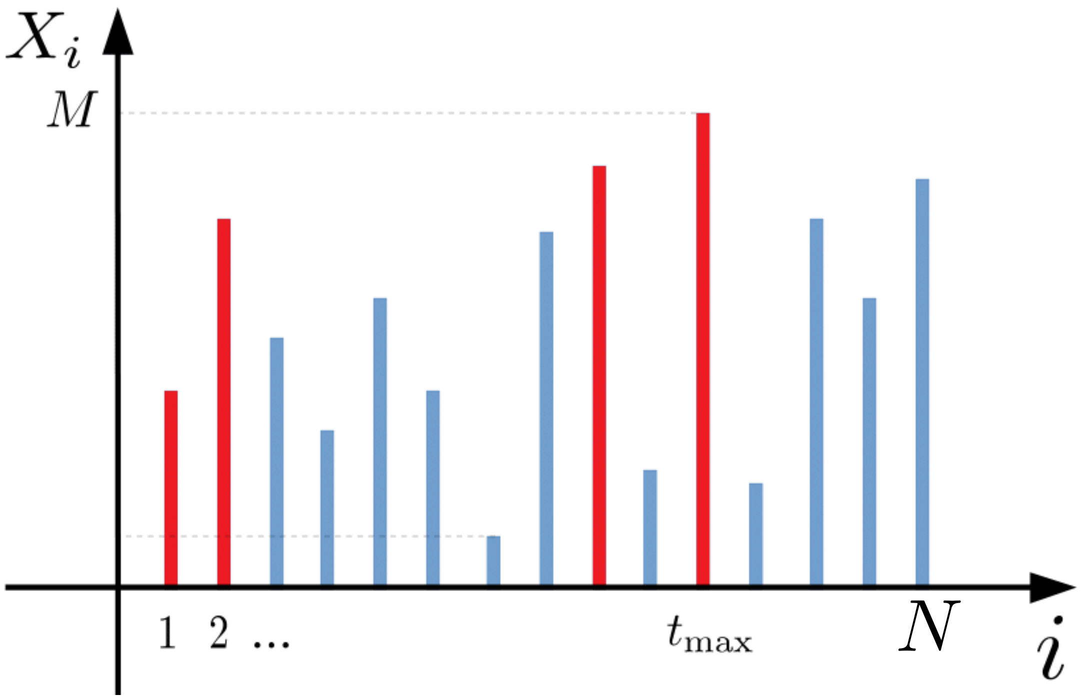
Marathon world record

Time above
2 hours [min]



Year

Extreme value statistics: setting



Global maximum

$$M = \max_{1 \le i \le N} X_i$$

Time of the maximum

$$t_{\max} = \mathrm{argmax}_{1 \le i \le N} X_i$$

Record statistics

N_R (Number of records)

[1] S. N. Majumdar, A. Pal, and G. Schehr, Physics Reports 840, 1 (2020).

Independent and identically distributed variables

$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

Joint probability distribution

Marginal probability distribution

No correlations

Example: Derrida's random energy model [1]

[1] Derrida, Bernard. "Random-energy model: An exactly solvable model of disordered systems." Physical Review B 24.5 (1981): 2613.

Extreme Value Theorem

$$M = \max_{1 \le i \le N} X_i$$

Gumbel

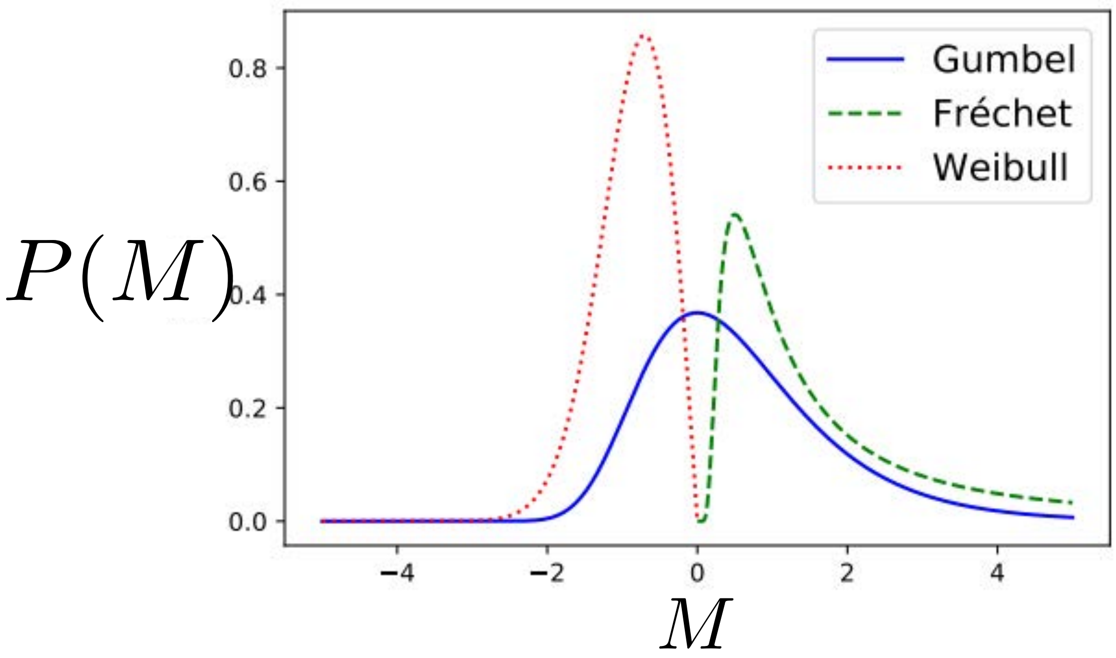
p(X) decays exponentially fast for large *x*

Fréchet

p(X) decays as a power law for large *x*

Weibull

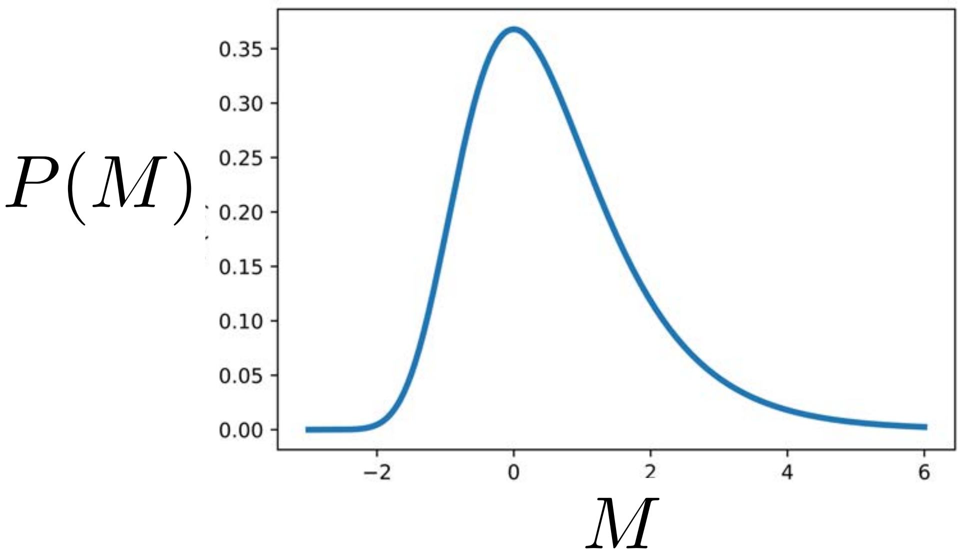
p(X) has an upper bounded support



Gumbel distribution

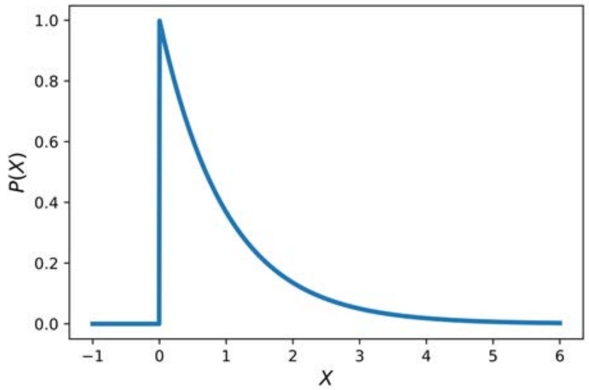
p(X) decays exponentially fast for large *x*

$$P(M) = e^{-M-e^{-M}}$$



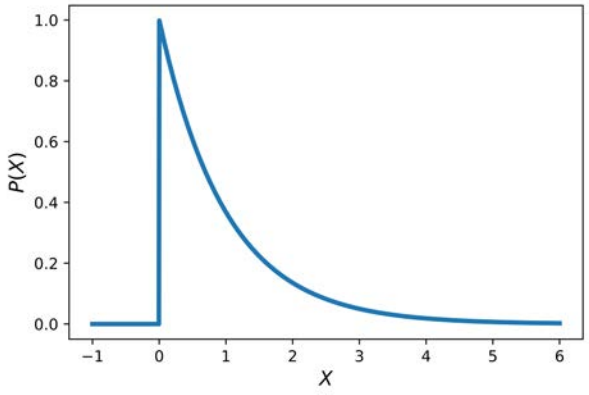
Extreme value distribution
Numerical example

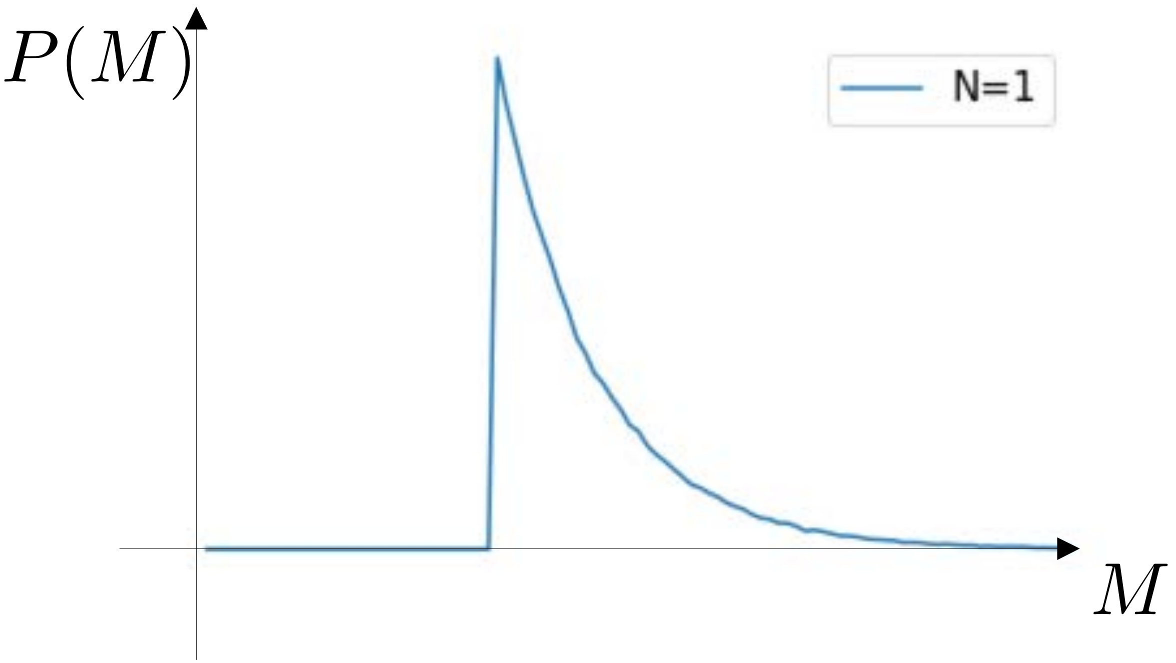
$$p(X) = \begin{cases} e^{-X} & \text{if } X > 0\,, \\ 0 & \text{otherwise} \end{cases}$$



Extreme value distribution
Numerical example

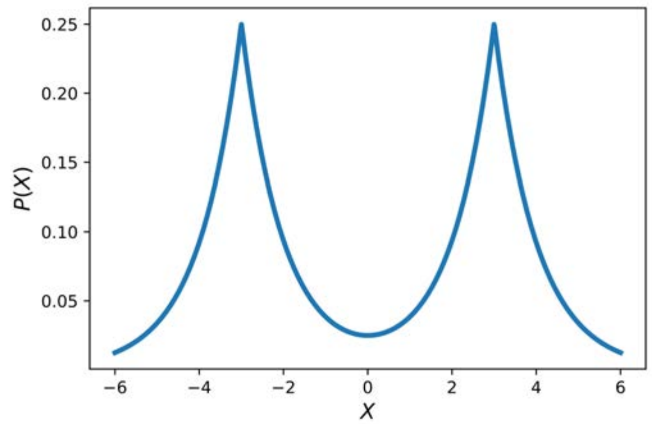
$$p(X) = \begin{cases} e^{-X} & \text{if } X > 0\,, \\ 0 & \text{otherwise} \end{cases}$$





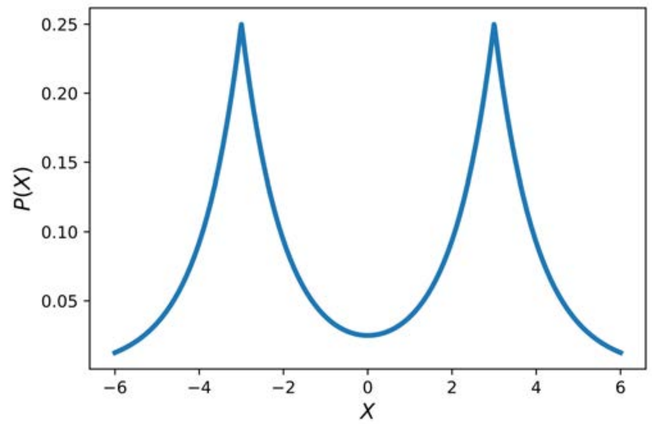
Extreme value distribution
Numerical example

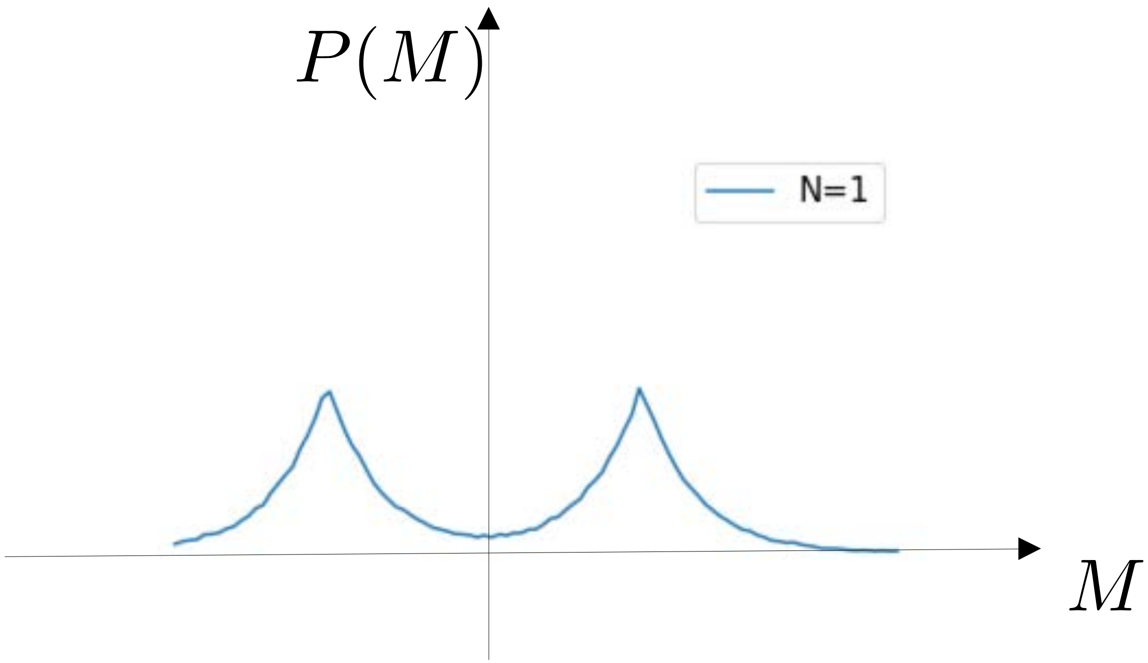
$$p(X) = \frac{1}{4}\left(e^{-|X-3|} + e^{-|X+3|}\right)$$



Extreme value distribution
Numerical example

$$p(X) = \frac{1}{4}\left(e^{-|X-3|} + e^{-|X+3|}\right)$$





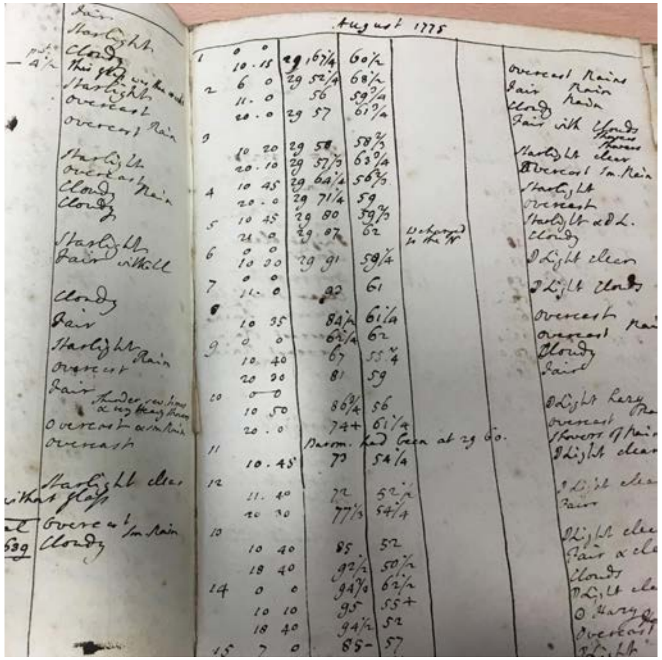
Does it work on real data?

Radcliffe Observatory

Images from Green Templeton College

Does it work on real data?

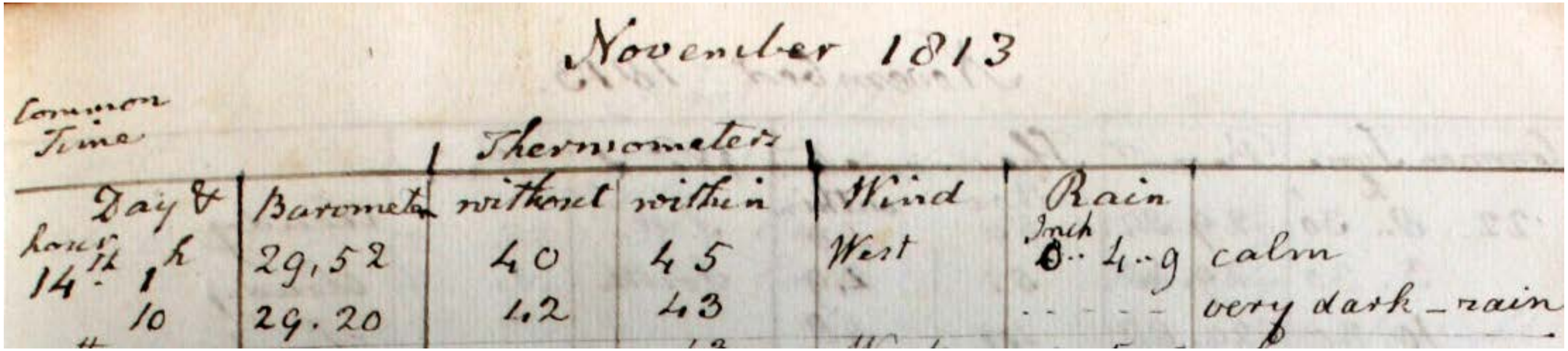
Radcliffe Meteorological Station



Images from Green Templeton College

Does it work on real data?

Radcliffe Meteorological Station



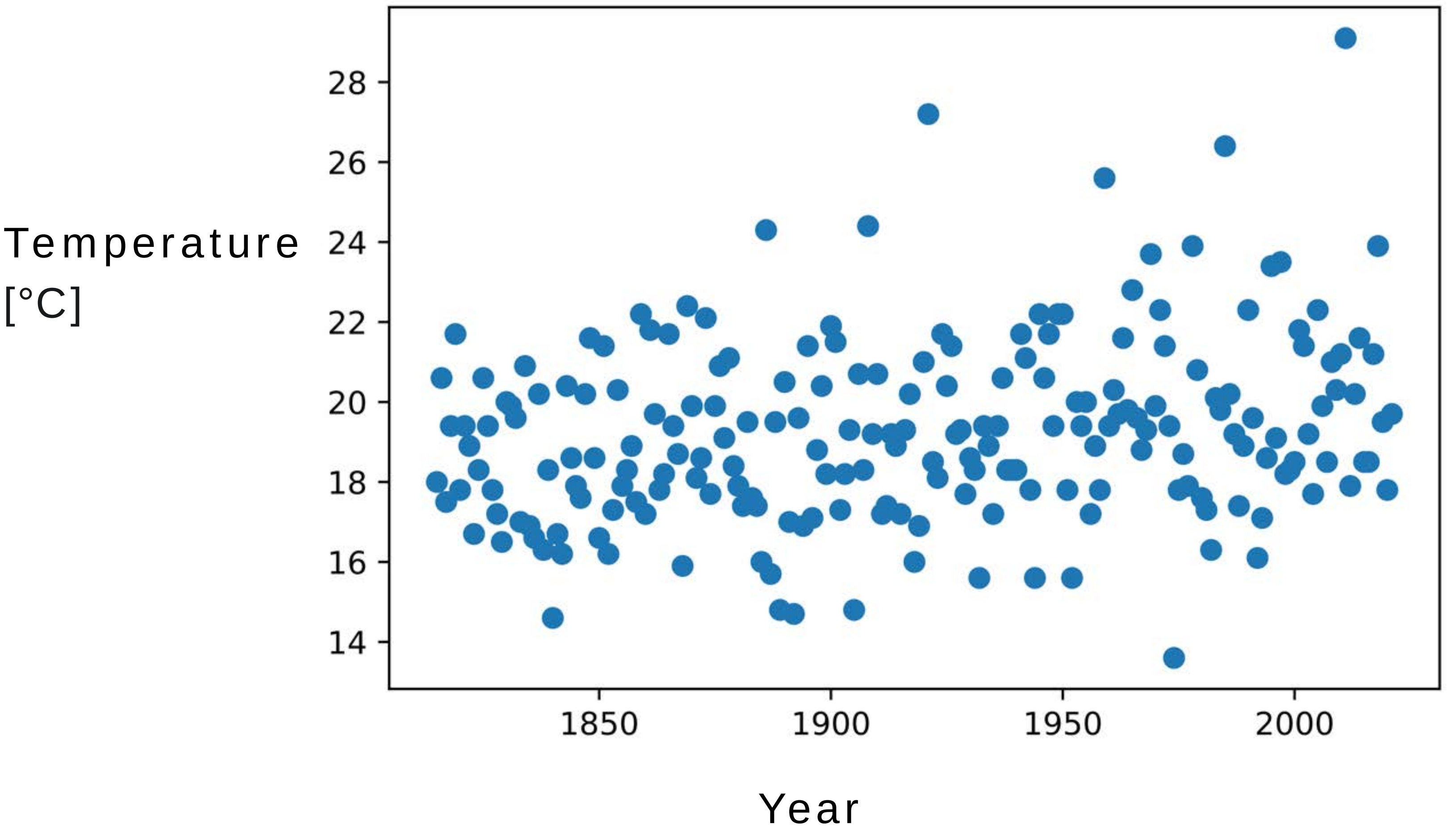
November 1813

Common Time Day & hours	Barometer	Thermometers without	Thermometers within	Wind	Rain Inch	
14th 1h	29,52	40	45	West	0..4..9	calm
10	29.20	42	43		- - - - -	very dark _ rain

YYYY	MM	DD	Tmax °C	Tmin °C	Daily Tmean °C	Daily range degC	Grass min °C	Air frost 0/1	Ground frost 0/1	Max ≥ 25.0°C	Max ≥ 30.0°C	Min ≥ 15.0 °C	Max < 0 °C
1815	1	1	6.6	-1.5	2.6	8.1		1		0	0	0	0
1815	1	2	4.9	-3.2	0.9	8.1		1		0	0	0	0
1815	1	3	2.6	-5.6	-1.5	8.2		1		0	0	0	0
1815	1	4	2.1	-6.1	-2	8.2		1		0	0	0	0
1815	1	5	1	-7.2	-3.1	8.2		1		0	0	0	0
1815	1	6	1.5	-6.6	-2.6	8.1		1		0	0	0	0
1815	1	7	-0.7	-9	-4.9	8.3		1		0	0	0	1
1815	1	8	4.9	-3.2	0.9	8.1		1		0	0	0	0
1815	1	9	1	-7.2	-3.1	8.2		1		0	0	0	0
1815	1	10	7.7	-0.4	3.7	8.1		1		0	0	0	0

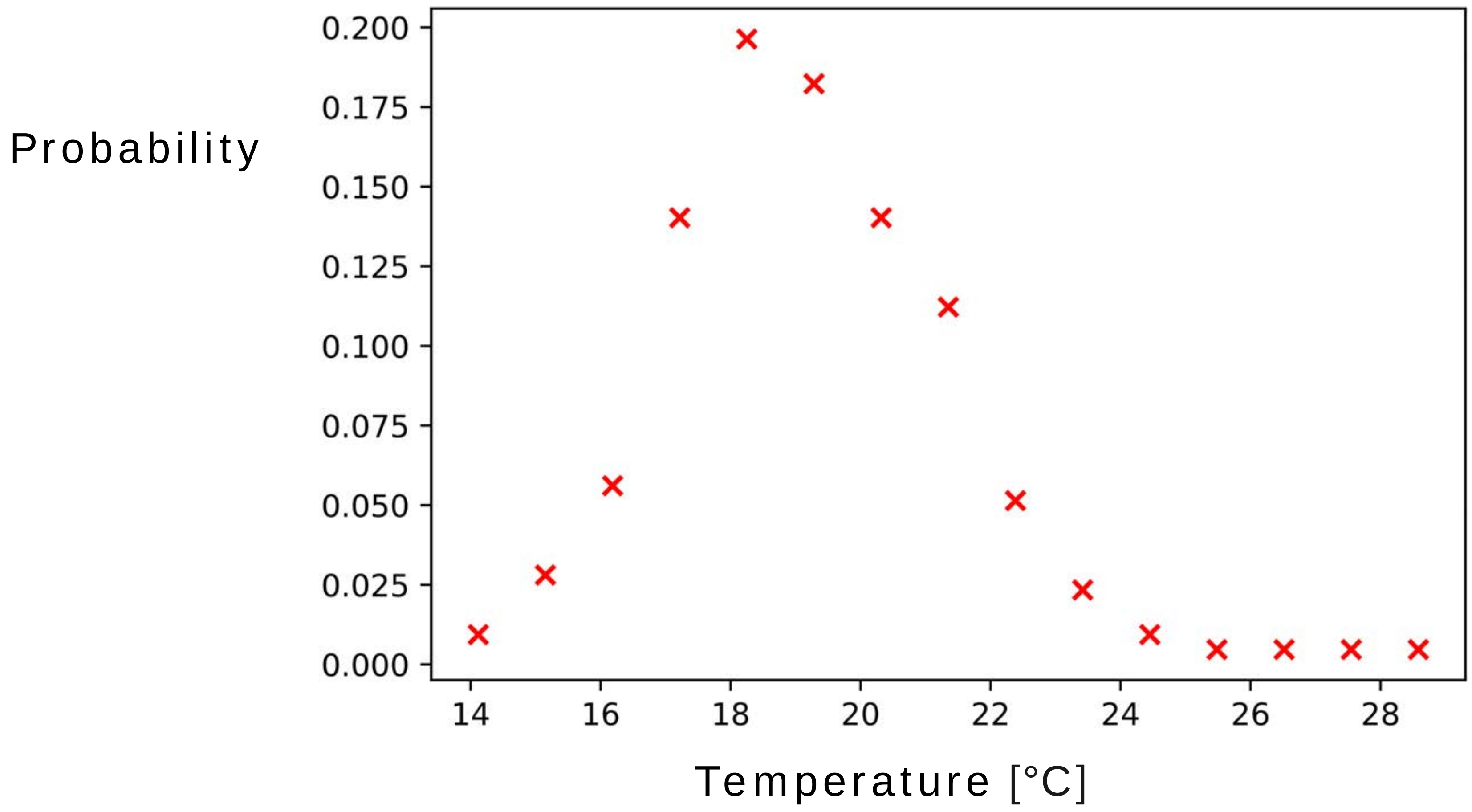
Images from the School of Geography and Environment

Maximal temperature in October



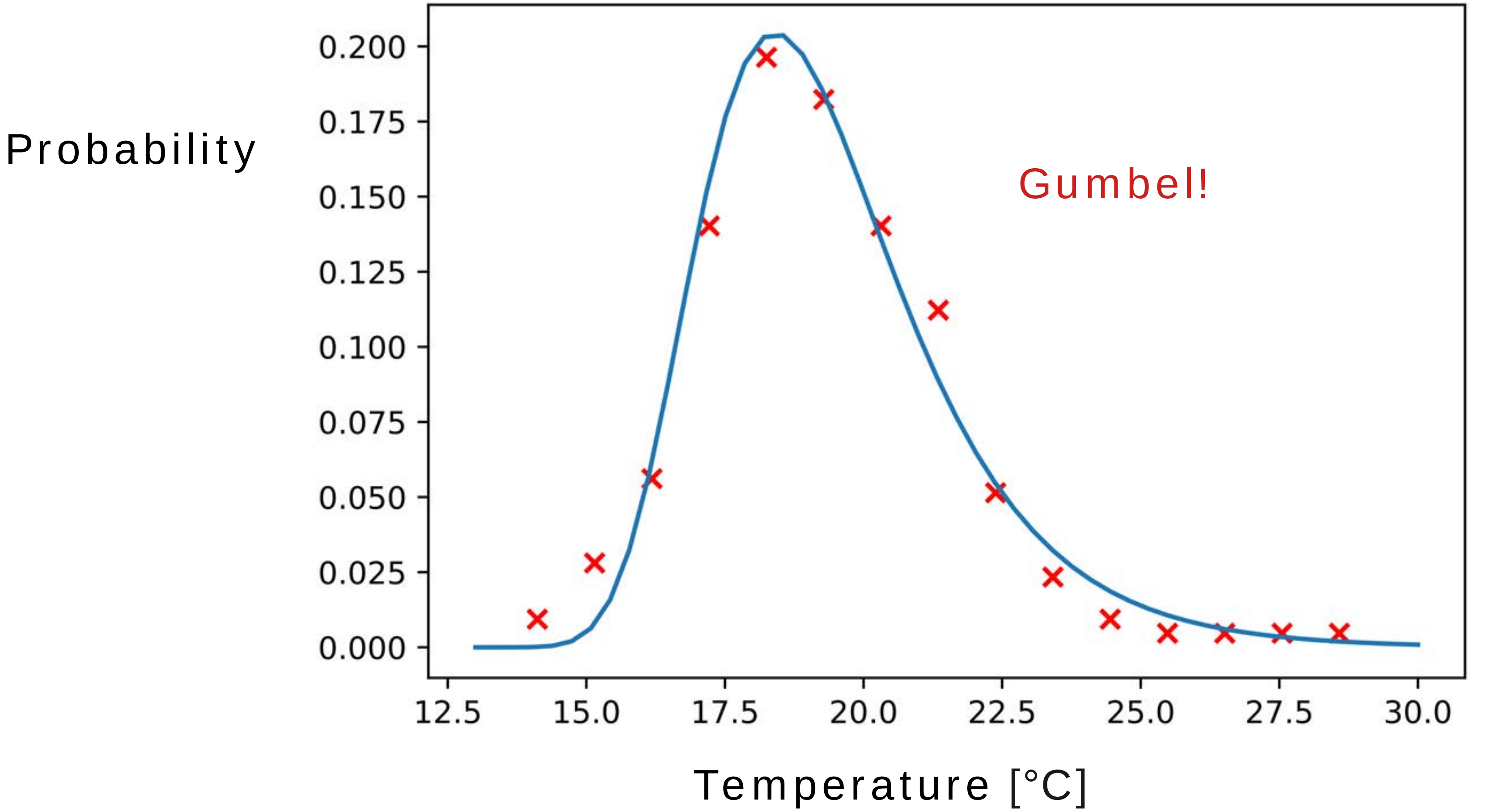
Data from https://www.geog.ox.ac.uk/

Maximal temperature in October



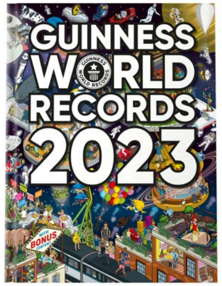
Data from https://www.geog.ox.ac.uk/

Maximal temperature in October



Data from https://www.geog.ox.ac.uk/

Independent variables: record statistics

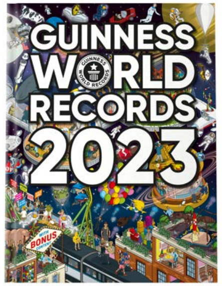


$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

$$X_i \text{ is a record if } X_i > \{X_1, , X_2, \ldots, X_{i-1}\}$$

Given a sequence of *N* random numbers, how many records do we expect to see?

Independent variables: record statistics



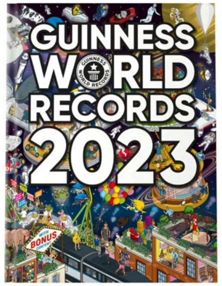
$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

$$X_i \text{ is a record if } X_i > \{X_1, , X_2, \ldots, X_{i-1}\}$$

Given a sequence of *N* random numbers, how many records do we expect to see?

$$\text{Prob.}(X_i \text{ is a record}) = \text{Prob.}(X_i > X_1, X_i > X_2, \ldots, X_i > X_{i-1}) = \frac{1}{i}$$

Independent variables: record statistics



$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

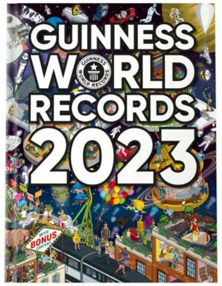
$$X_i \text{ is a record if } X_i > \{X_1, , X_2, \ldots, X_{i-1}\}$$

Given a sequence of *N* random numbers, how many records do we expect to see?

$$\text{Prob.}(X_i \text{ is a record}) = \text{Prob.}(X_i > X_1, X_i > X_2, \ldots, X_i > X_{i-1}) = \frac{1}{i}$$

$$\langle N_R \rangle = \sum_{i=1}^{N} \frac{1}{i} \approx \int_1^N \frac{1}{i} di = \log N$$

Independent variables: record statistics



$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

$$X_i \text{ is a record if } X_i > \{X_1, , X_2, \ldots, X_{i-1}\}$$

Given a sequence of *N* random numbers, how many records do we expect to see?

$$\text{Prob.}(X_i \text{ is a record}) = \text{Prob.}(X_i > X_1, X_i > X_2, \ldots, X_i > X_{i-1}) = \frac{1}{i}$$

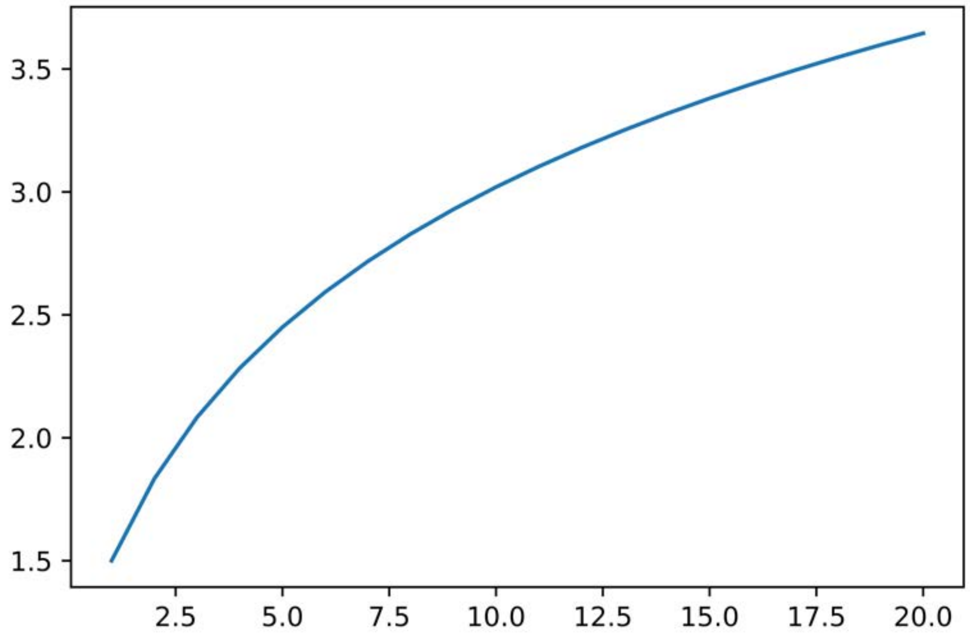
$$\langle N_R \rangle = \sum_{i=1}^{N} \frac{1}{i} \approx \int_1^N \frac{1}{i} di = \log N$$

UNIVERSAL!

Independent variables: record statistics

$$\langle N_R \rangle \approx \log(N)$$

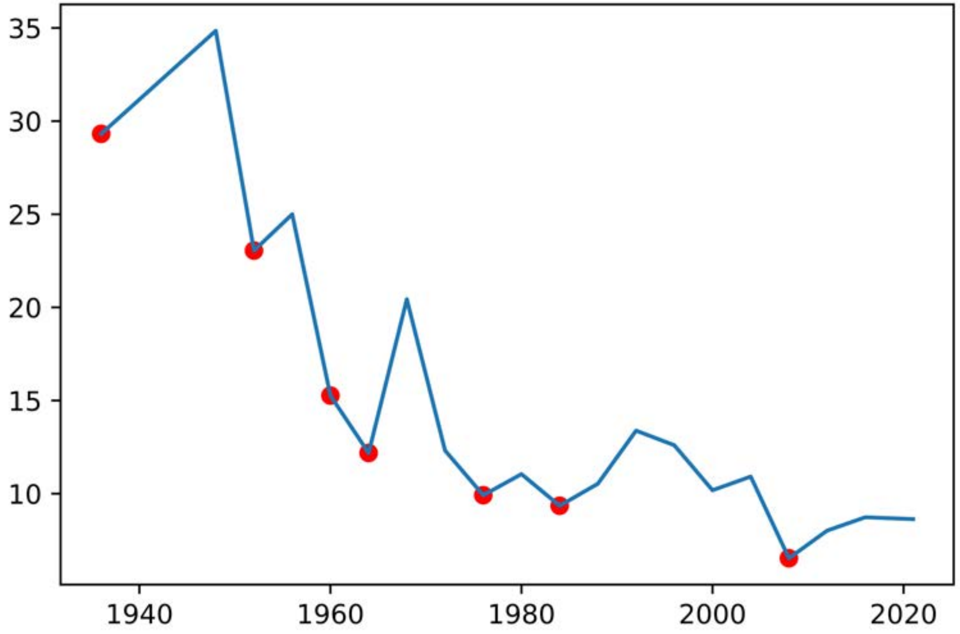
$\langle N_R \rangle$



N

Marathon record statistics

Time above
2 hours [min]

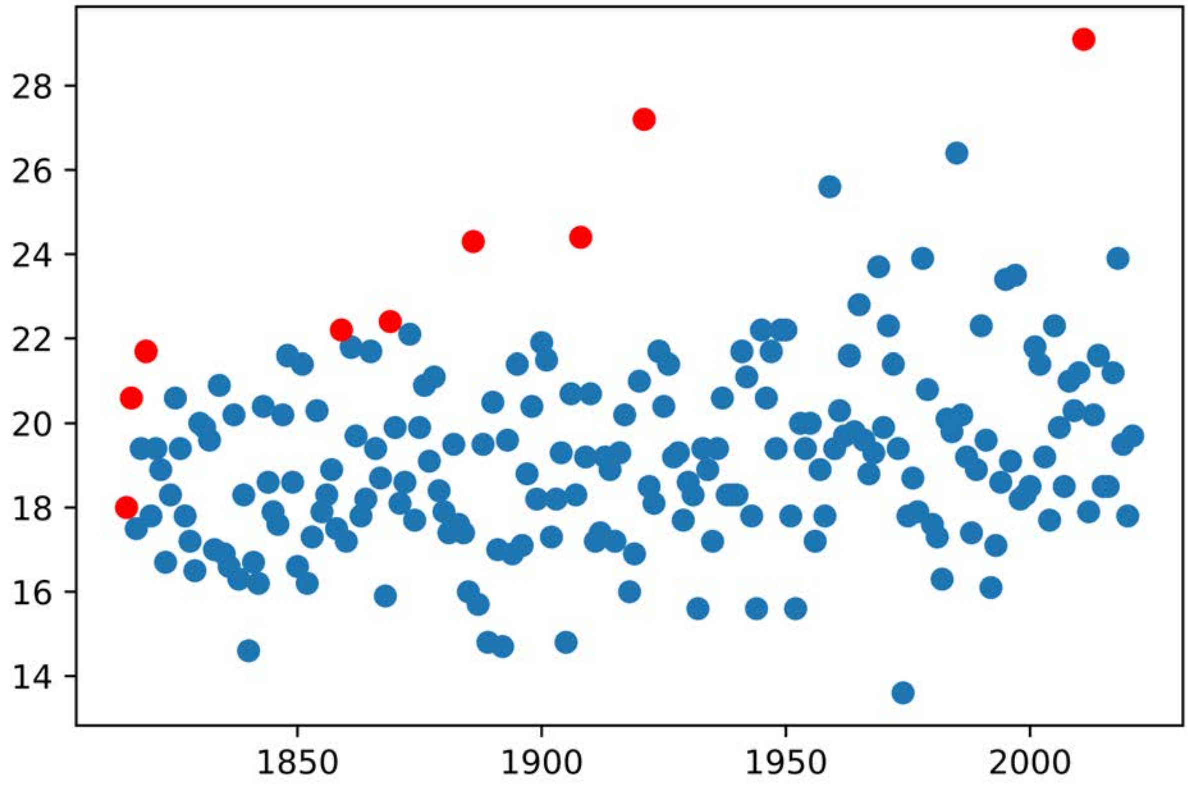


Year

In the last 20 editions of the Olympics → 7 Marathon records

Data from https://olympics.com/en/

Maximal temperature in October



Temperature [°C]

Year

In the last 200 years → 9 records

Data from https://www.geog.ox.ac.uk/

Correlated systems

Correlated random variables

$$P(X_1, \ldots, X_N) \neq \prod_{i=1}^{N} p(X_i)$$

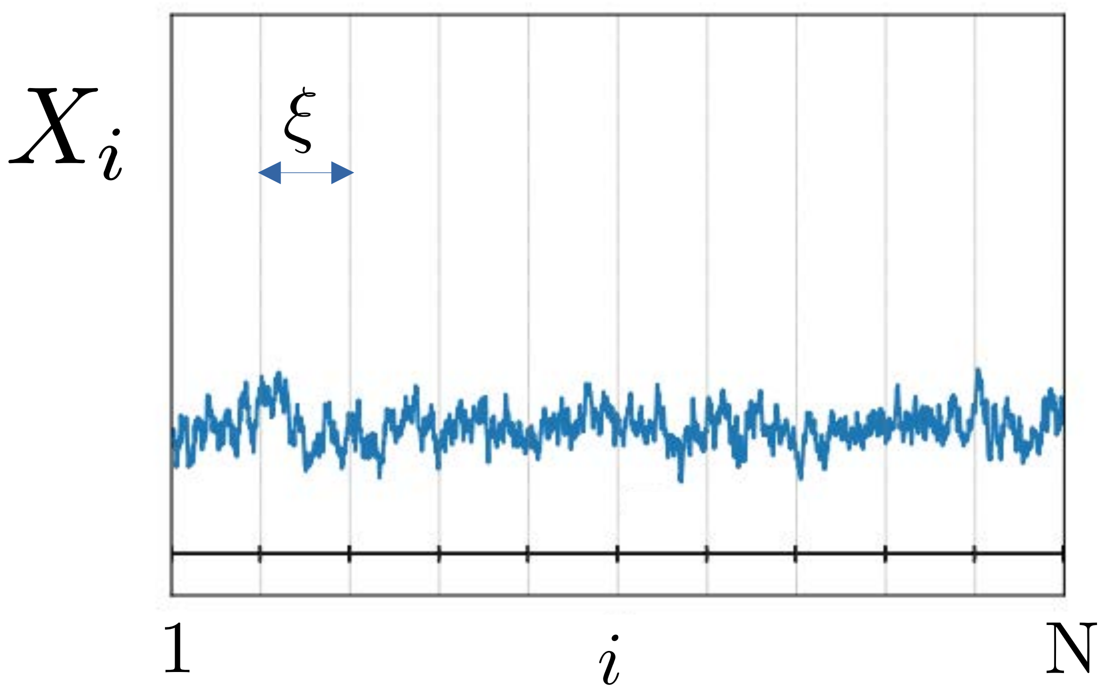
No general technique!

Weakly correlated random variables

$$N \gg \xi$$

correlations

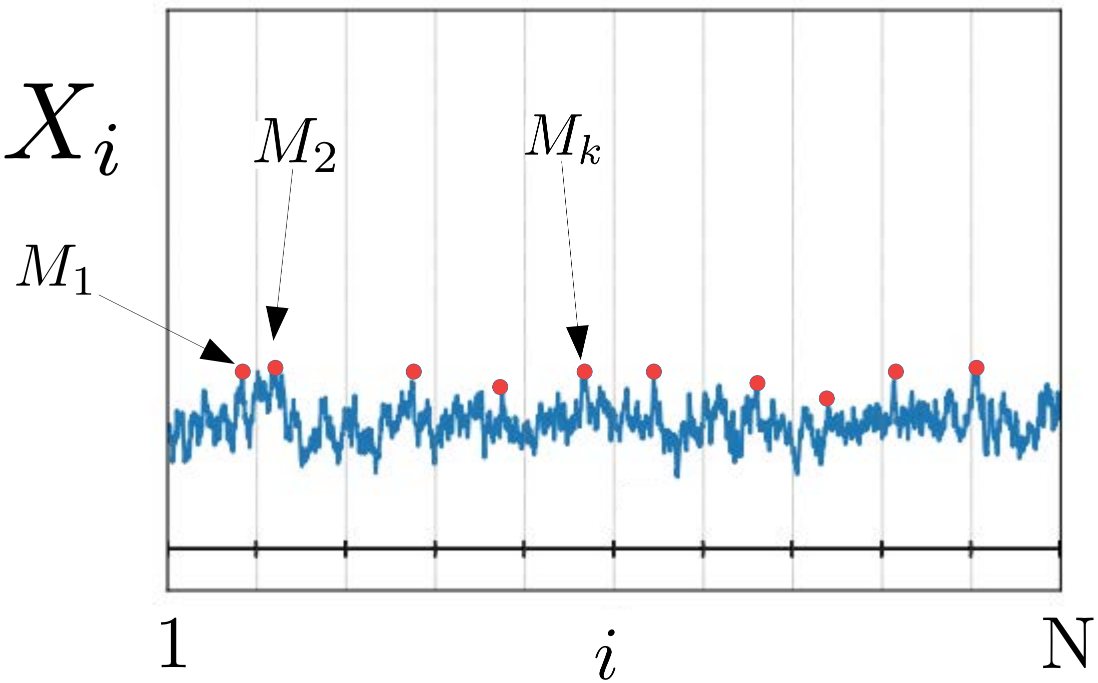
$$\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \sim e^{-|i-j|/\xi}$$



Weakly correlated random variables

i.i.d. problem

$$M = \max_k M_k$$



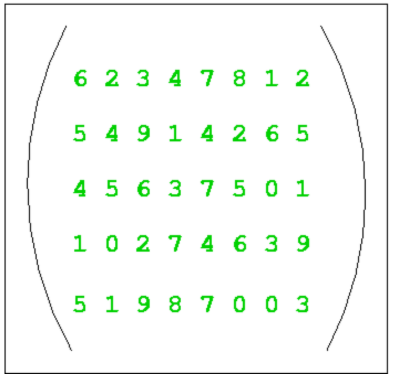
Exactly solvable models

Strongly correlated random variables

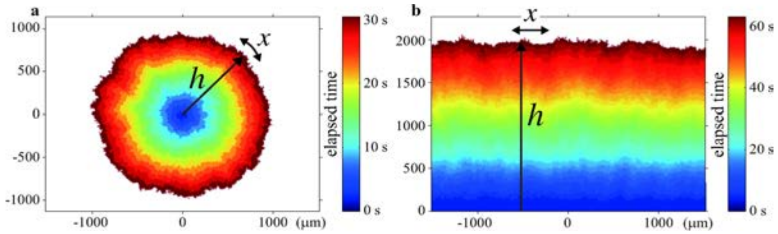
Random Matrices [1,2]

Fluctuating interfaces [3]

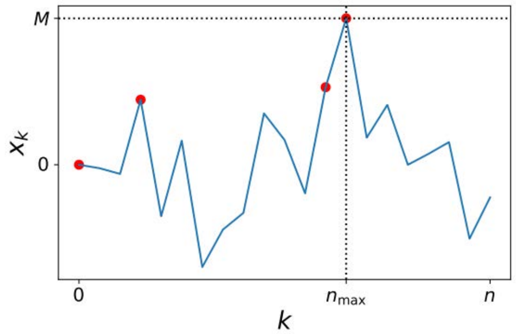
Random walks



From www.sissa.it



From Takeuchi et al., Scientific reports (2011).



[1] C. A. Tracy and H. Widom, Commun. Math. Phys. 159, 151 (1994), 177, 727 (1996).
[2] S. N. Majumdar and G. Schehr, J. Stat. Mech. Theory Exp. P01012 (2014).
[3] S. N. Majumdar and A. Comtet, *Phys. Rev. Lett.* 92, 225501 (2004).

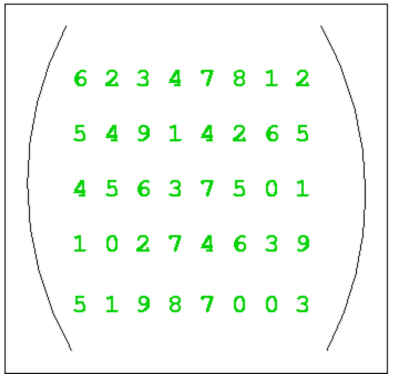
Exactly solvable models

Strongly correlated random variables

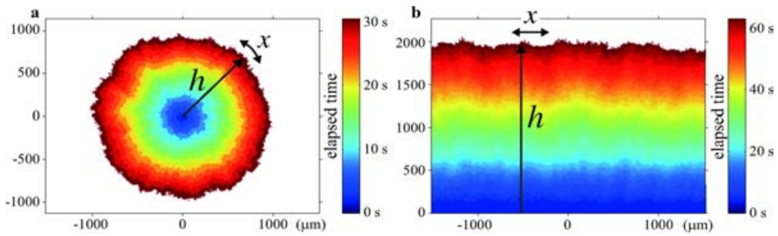
Random Matrices [1,2]

Fluctuating interfaces [3]

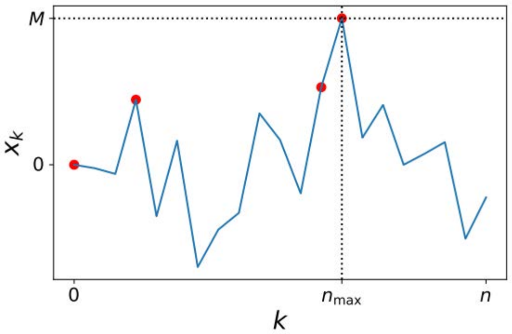
Random walks



From www.sissa.it

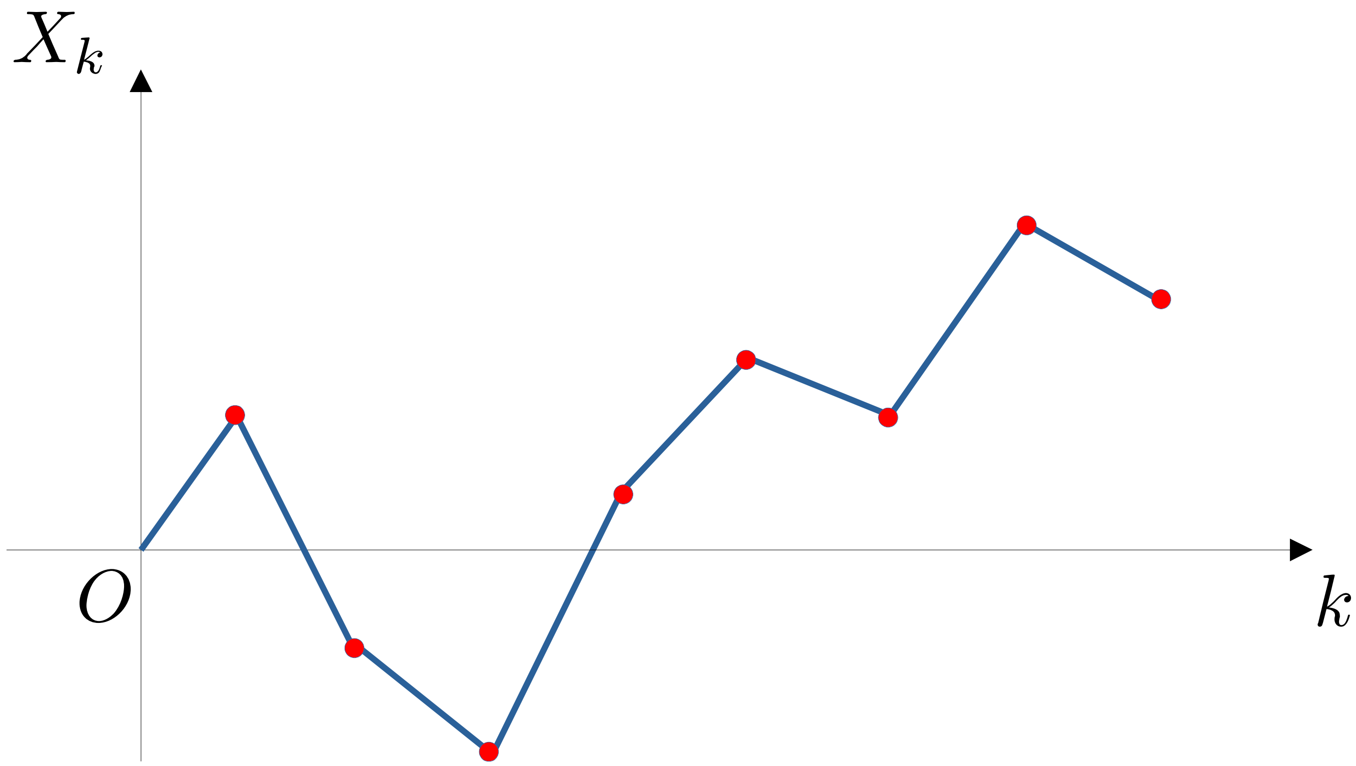


From Takeuchi et al., Scientific reports (2011).



[1] C. A. Tracy and H. Widom, Commun. Math. Phys. 159, 151 (1994), 177, 727 (1996).
[2] S. N. Majumdar and G. Schehr, J. Stat. Mech. Theory Exp. P01012 (2014).
[3] S. N. Majumdar and A. Comtet, *Phys. Rev. Lett.* 92, 225501 (2004).

Random walks

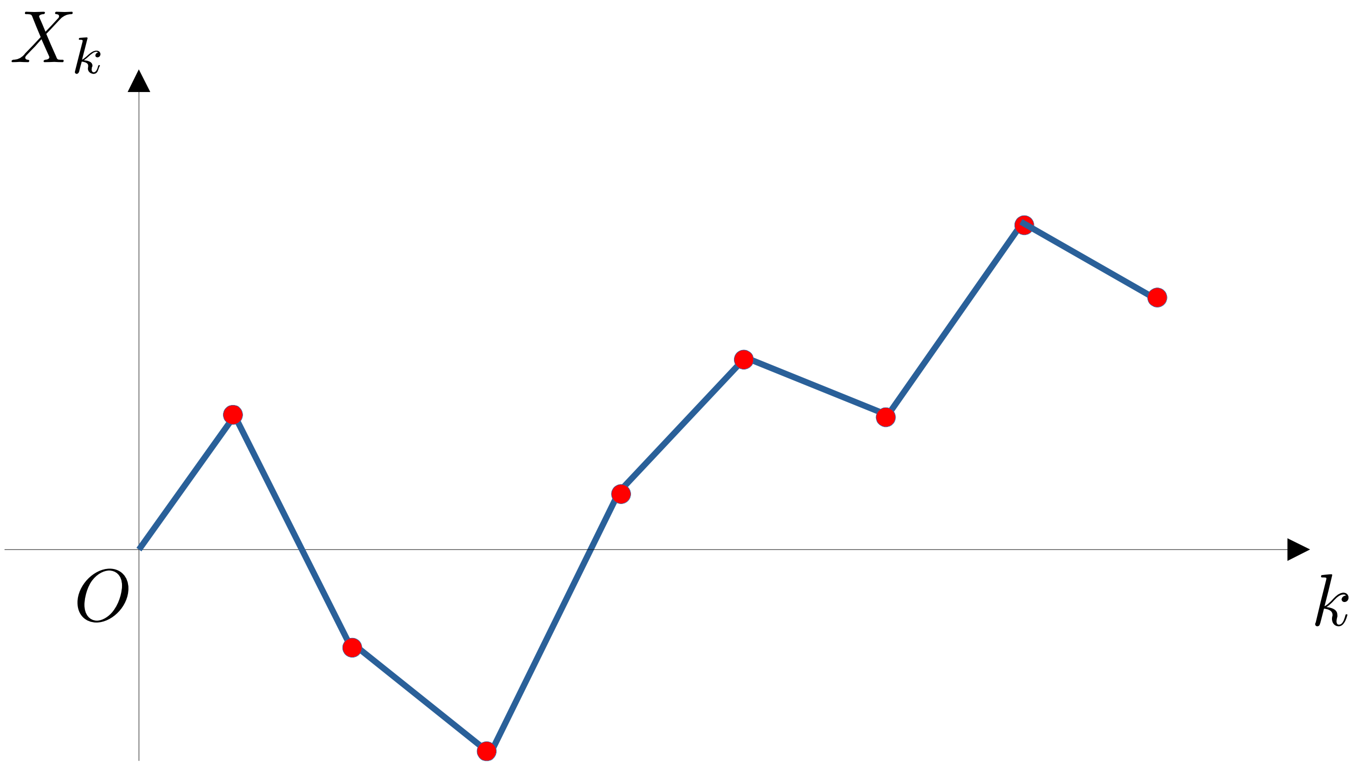


$$X_k = X_{k-1} + \eta_k$$

Independent and identically distributed jumps

$$\eta_k \sim p(\eta)$$

Random walks



$$X_k = X_{k-1} + \eta_k$$

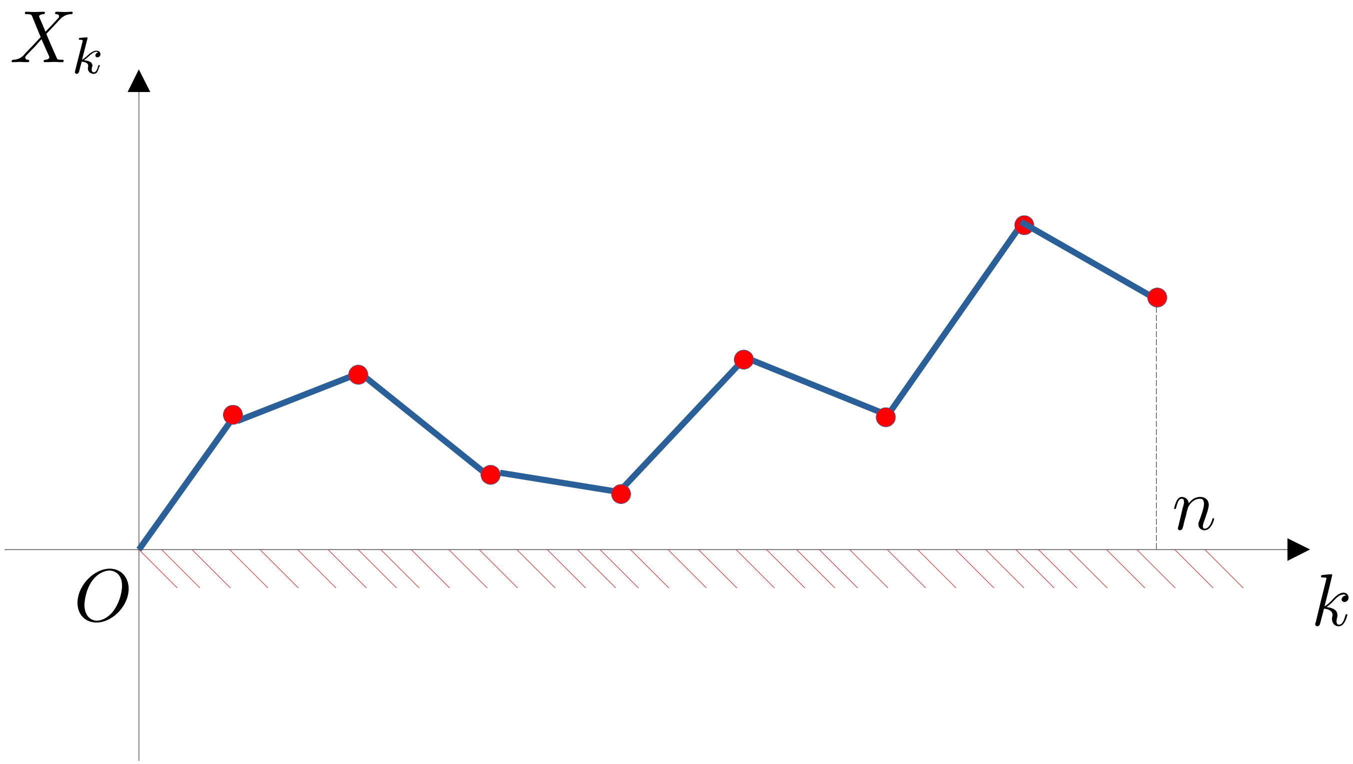
Independent and identically distributed jumps

$$\eta_k \sim p(\eta)$$

$$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N-1} p(X_{i+1} - X_i)$$

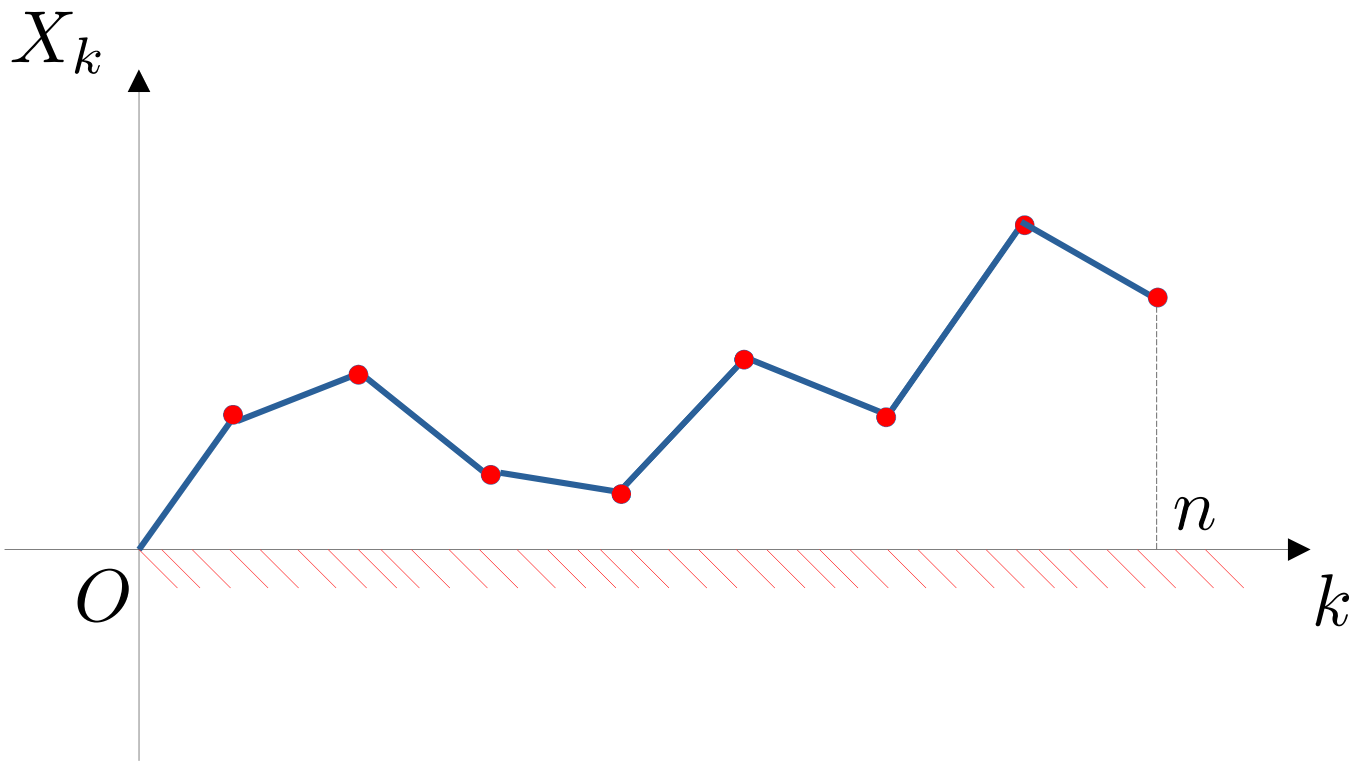
Survival probability

Survival probability



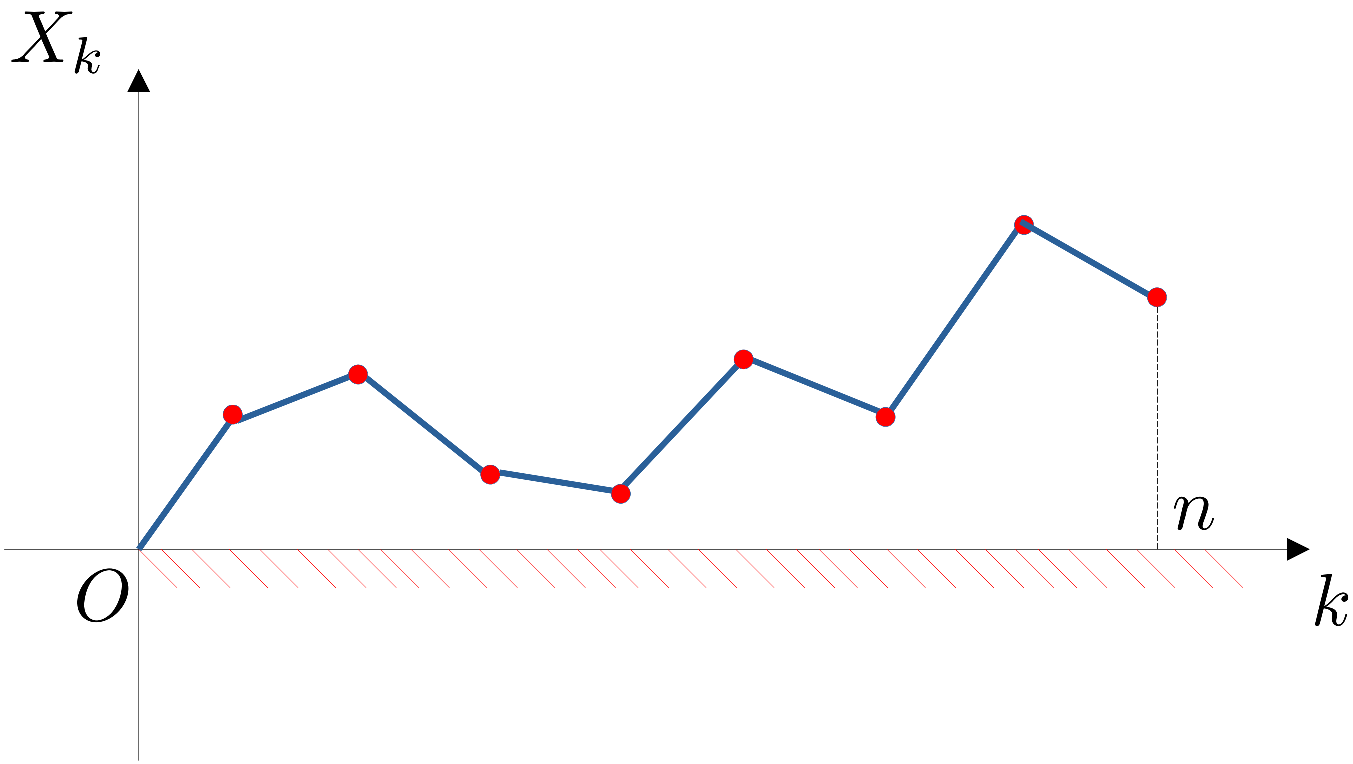
$$q_n = \text{Prob.}(X_1 \geq 0, X_2 \geq 0, \ldots, X_n \geq 0 | X_0 = 0)$$

Survival probability



$$
\begin{aligned}
q_n &= \text{Prob.}(X_1 \geq 0, X_2 \geq 0\,, \ldots\,, X_n \geq 0 | X_0 = 0) \\
&= \int_{-\infty}^{\infty} d\eta_1\,\ldots \int_{-\infty}^{\infty} d\eta_n\,\theta(\eta_1)\theta(\eta_1+\eta_2)\ldots\theta(\eta_1+\eta_2+\ldots+\eta_n)\prod_{i=1}^{n} p(\eta_i)
\end{aligned}
$$

Sparre Andersen theorem



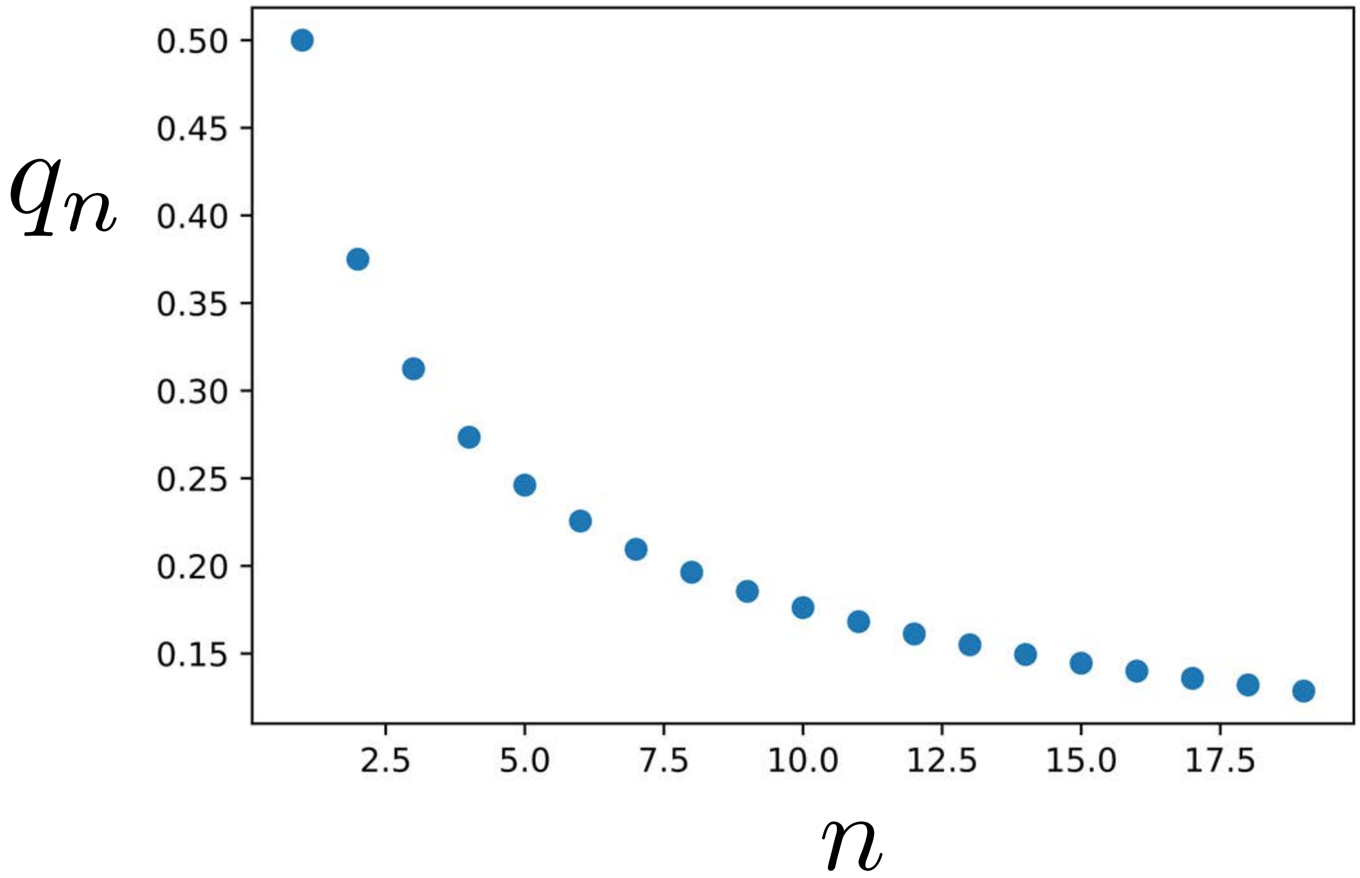
$$q_n = \binom{2n}{n} 2^{-2n}$$

UNIVERSAL!

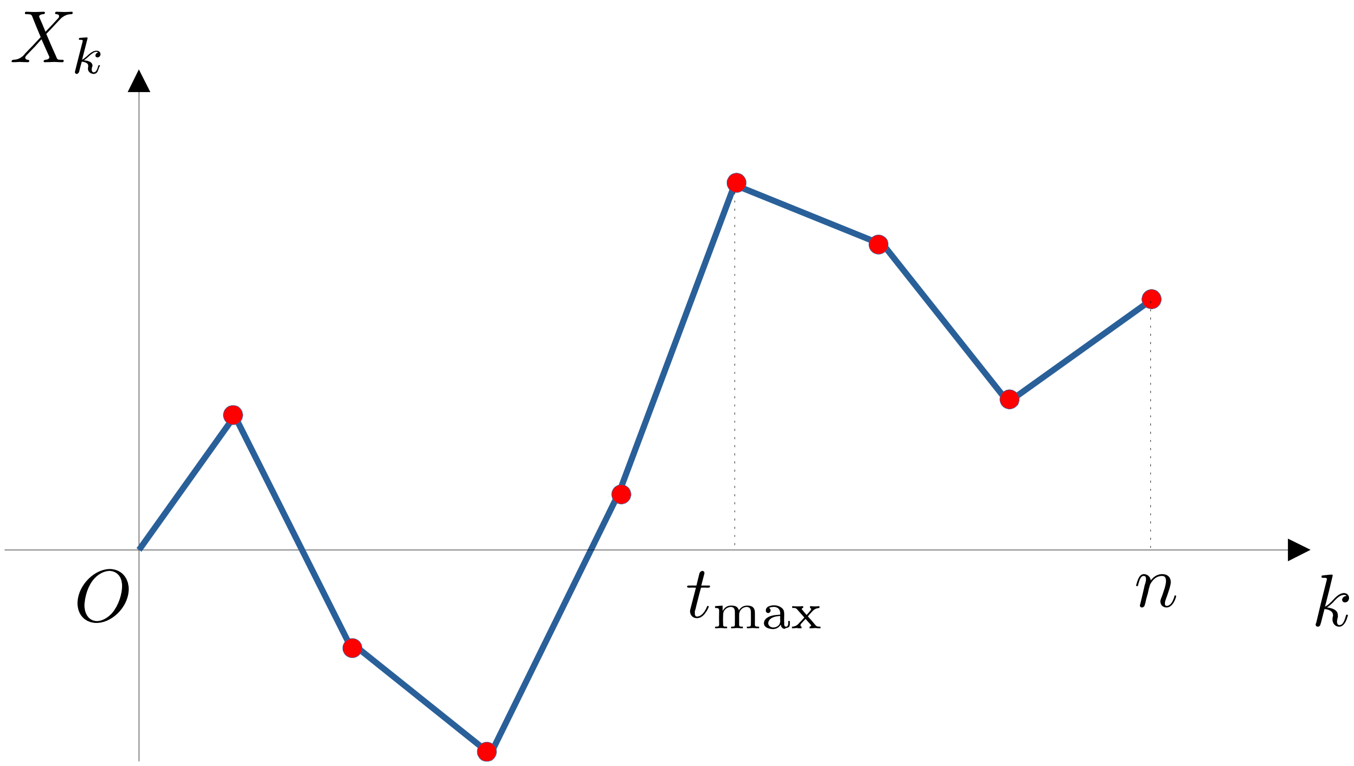
Independent of ***p(η)*** for any ***n***

[1] E. Sparre Andersen, Math. Scand. **2**, 195 (1954).

Sparre Andersen theorem



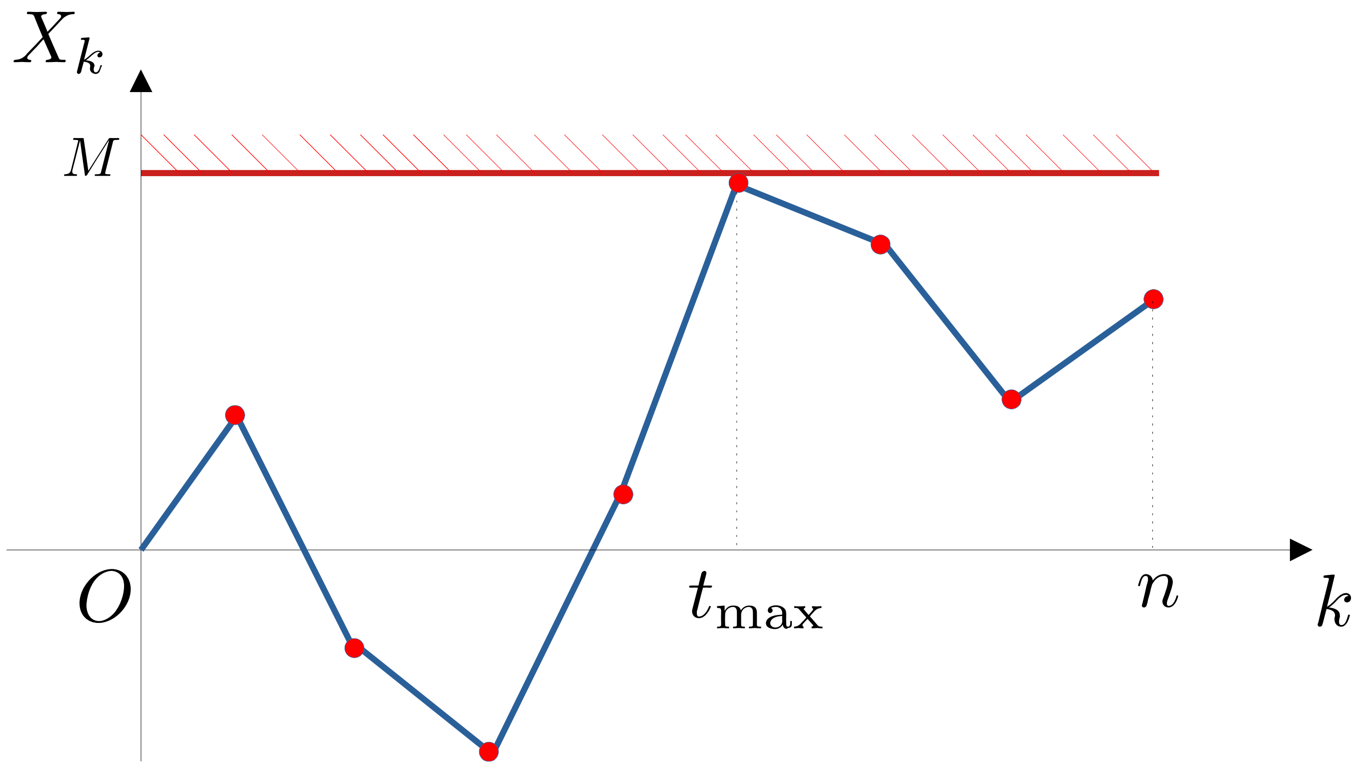
Time of the maximum



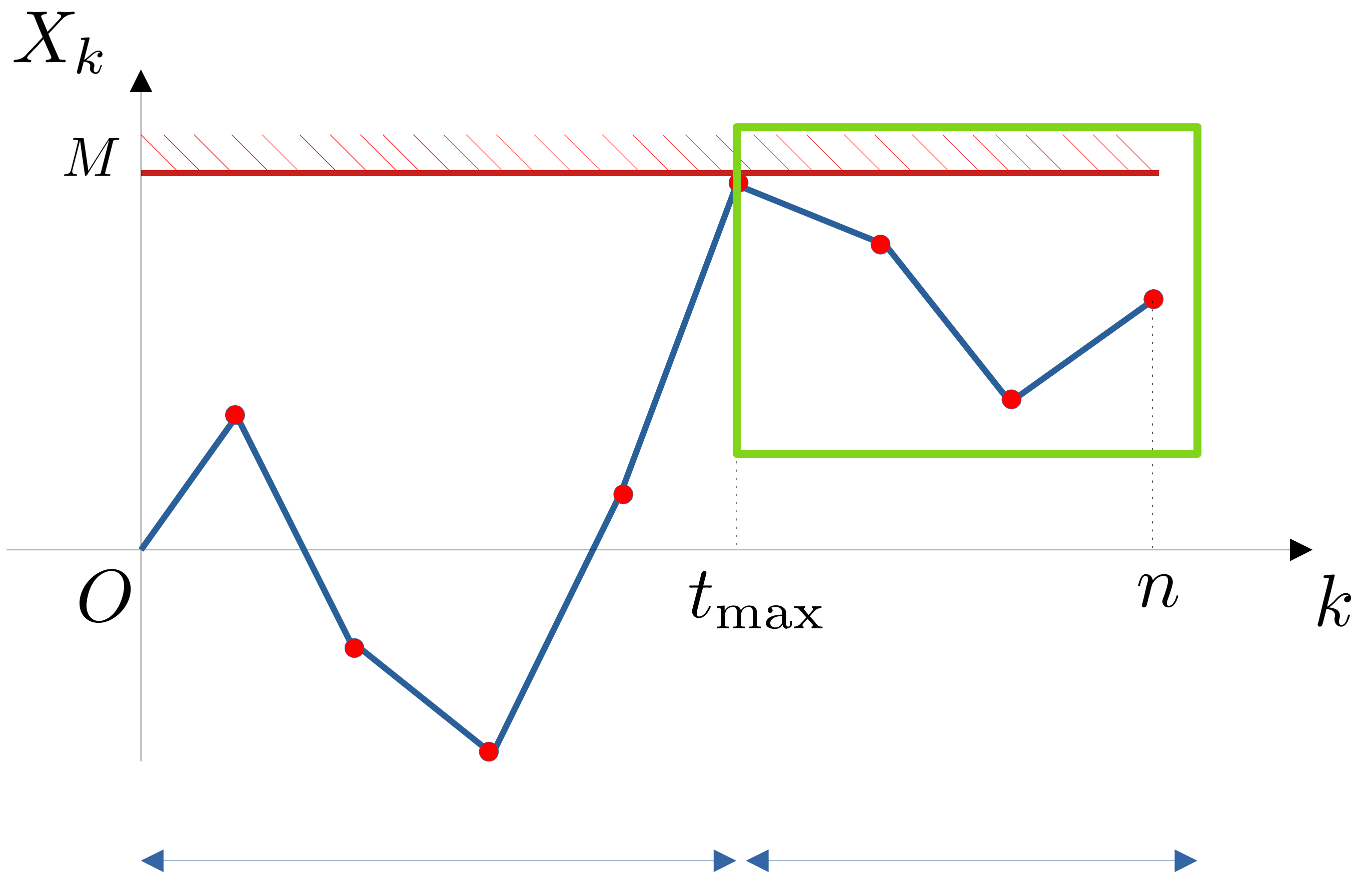
$$P(t_{\max}|n) = ?$$

Applications to finance, sports,...

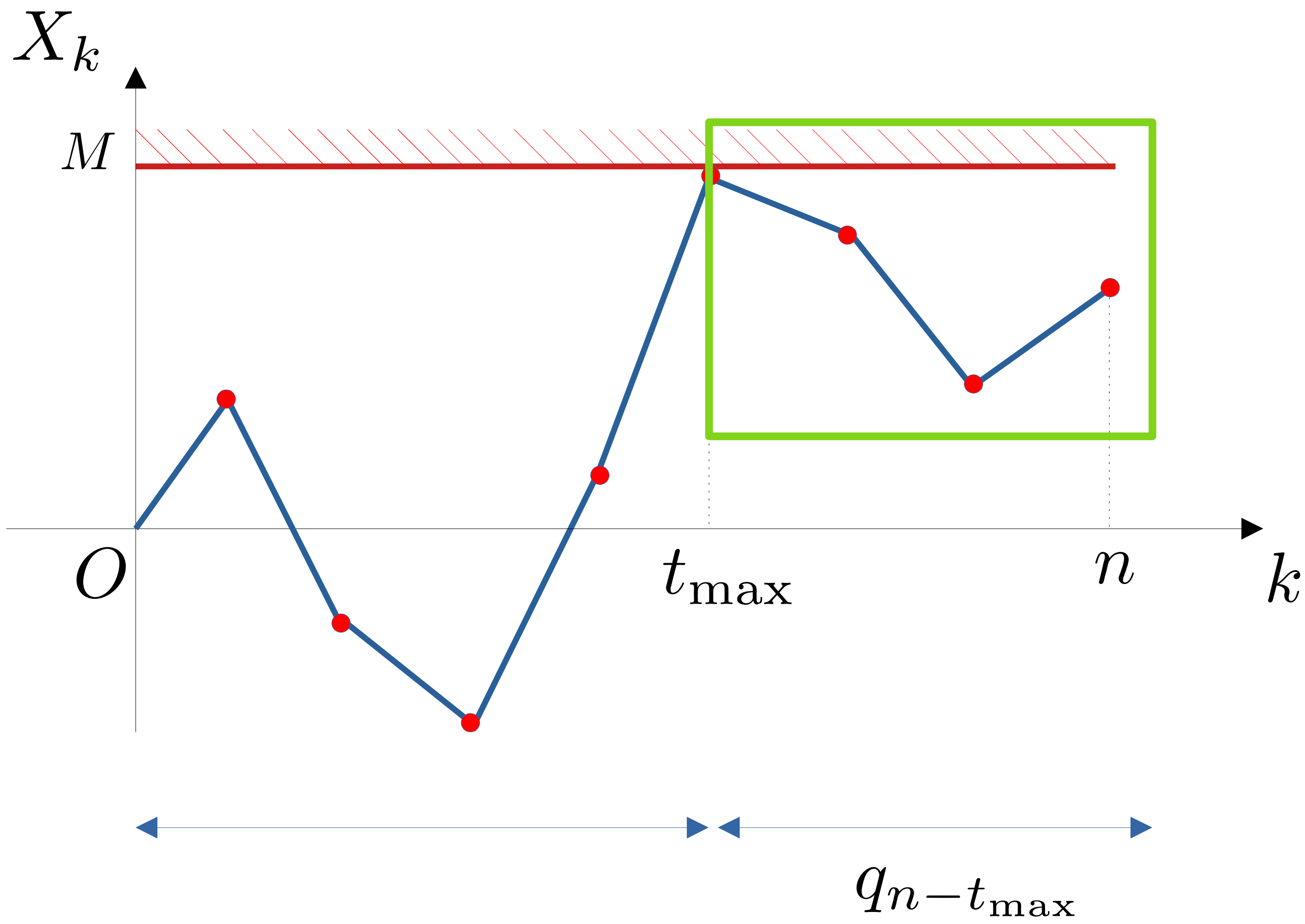
Time of the maximum



Time of the maximum

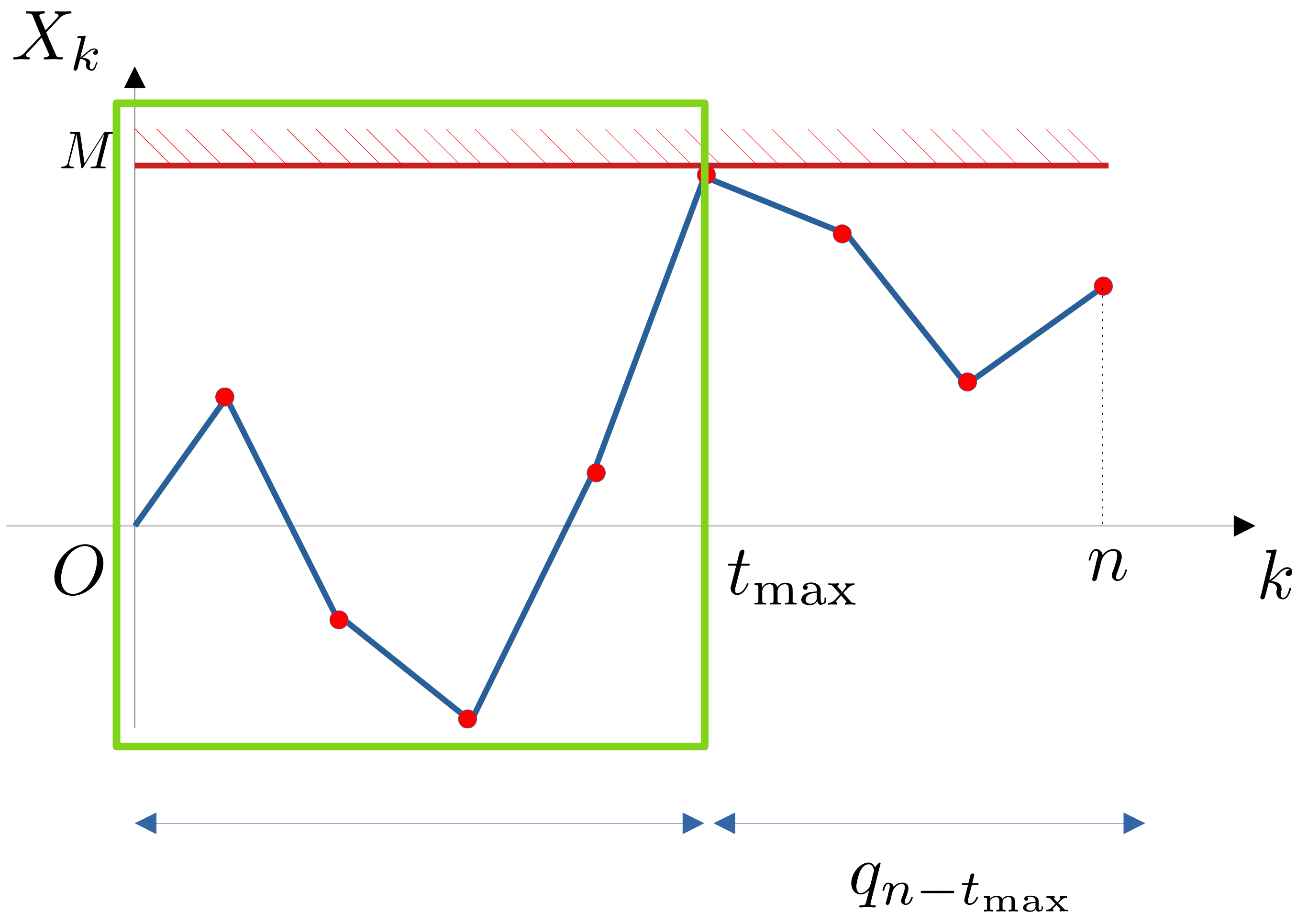


Time of the maximum



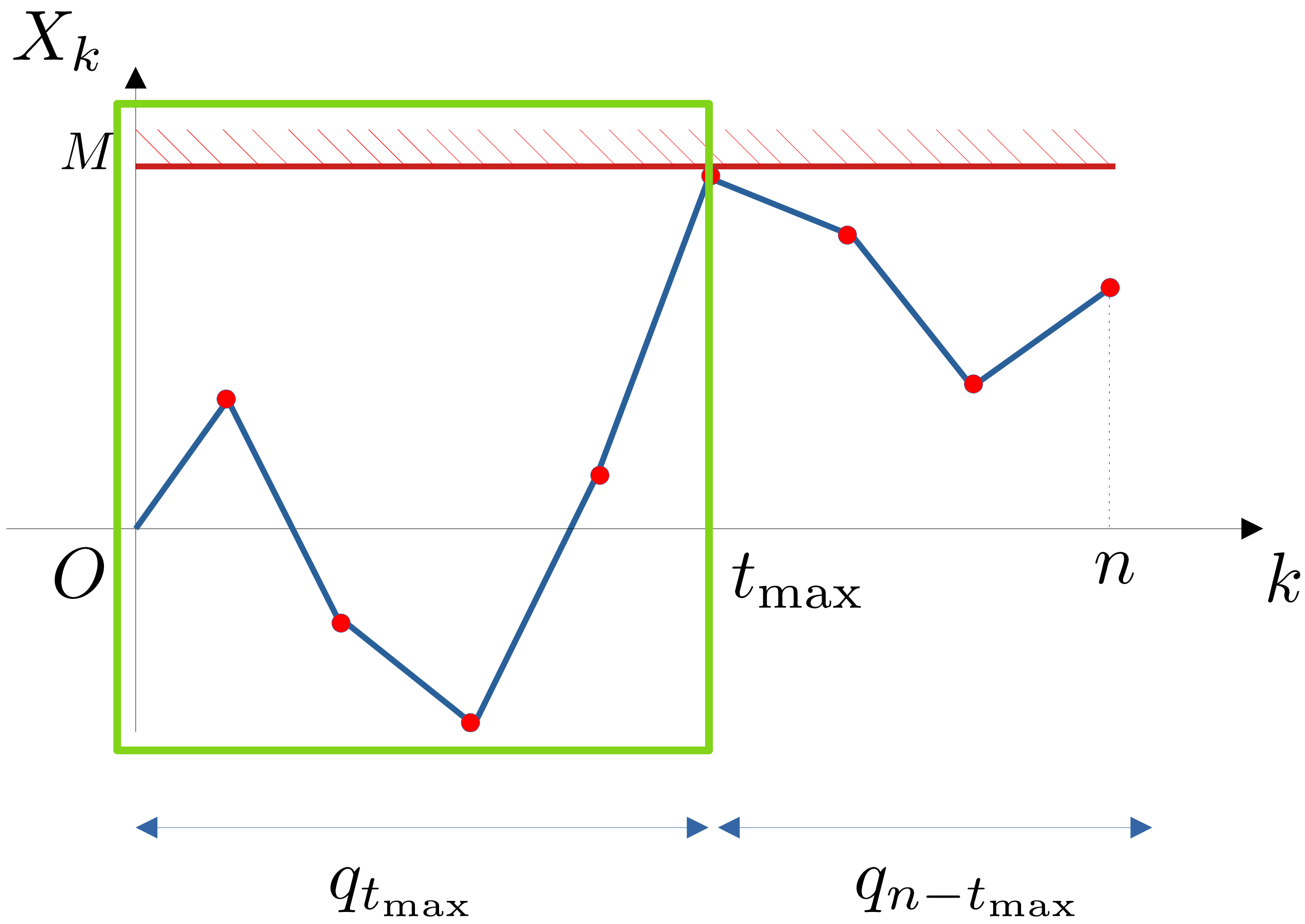
$$q_k = \binom{2k}{k} 2^{-2k}$$

Time of the maximum



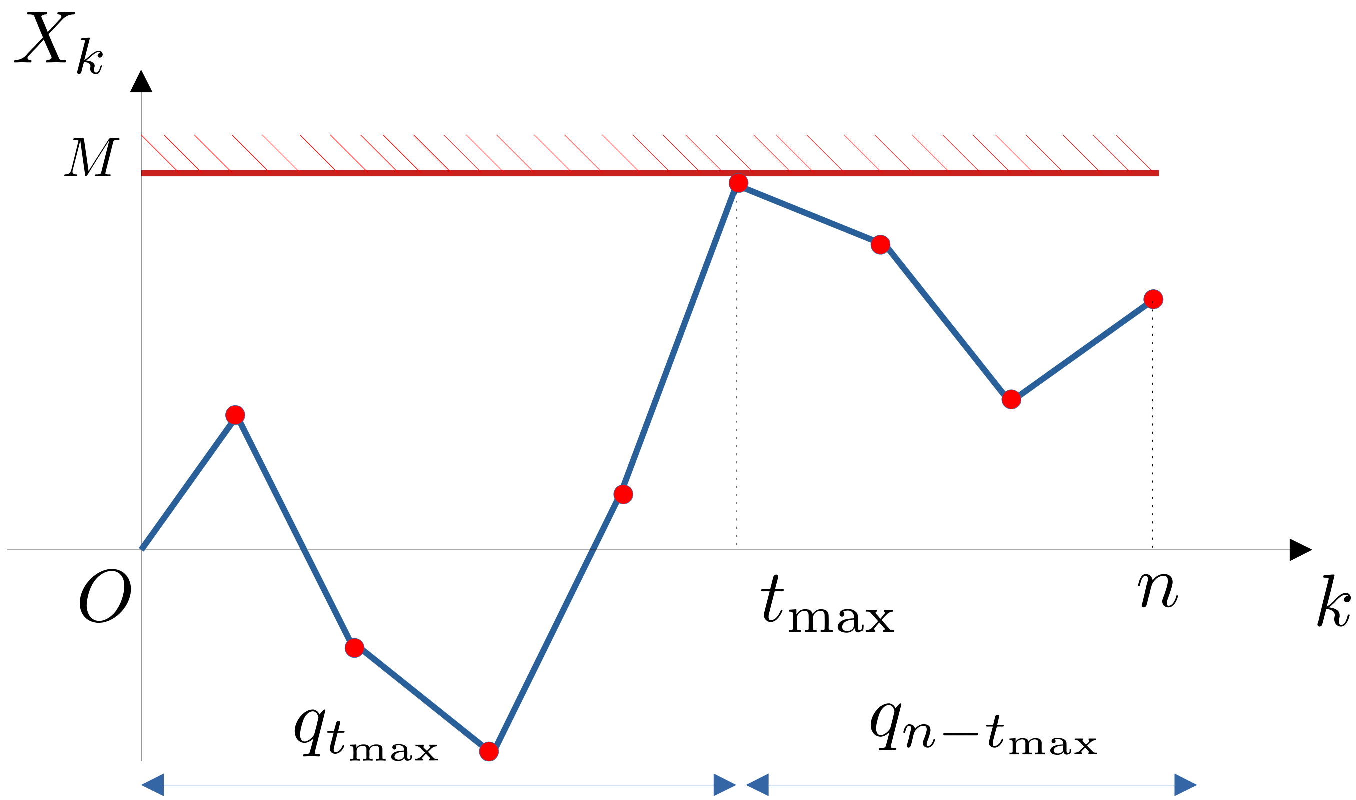
$$q_k = \binom{2k}{k} 2^{-2k}$$

Time of the maximum



$$q_k = \binom{2k}{k} 2^{-2k}$$

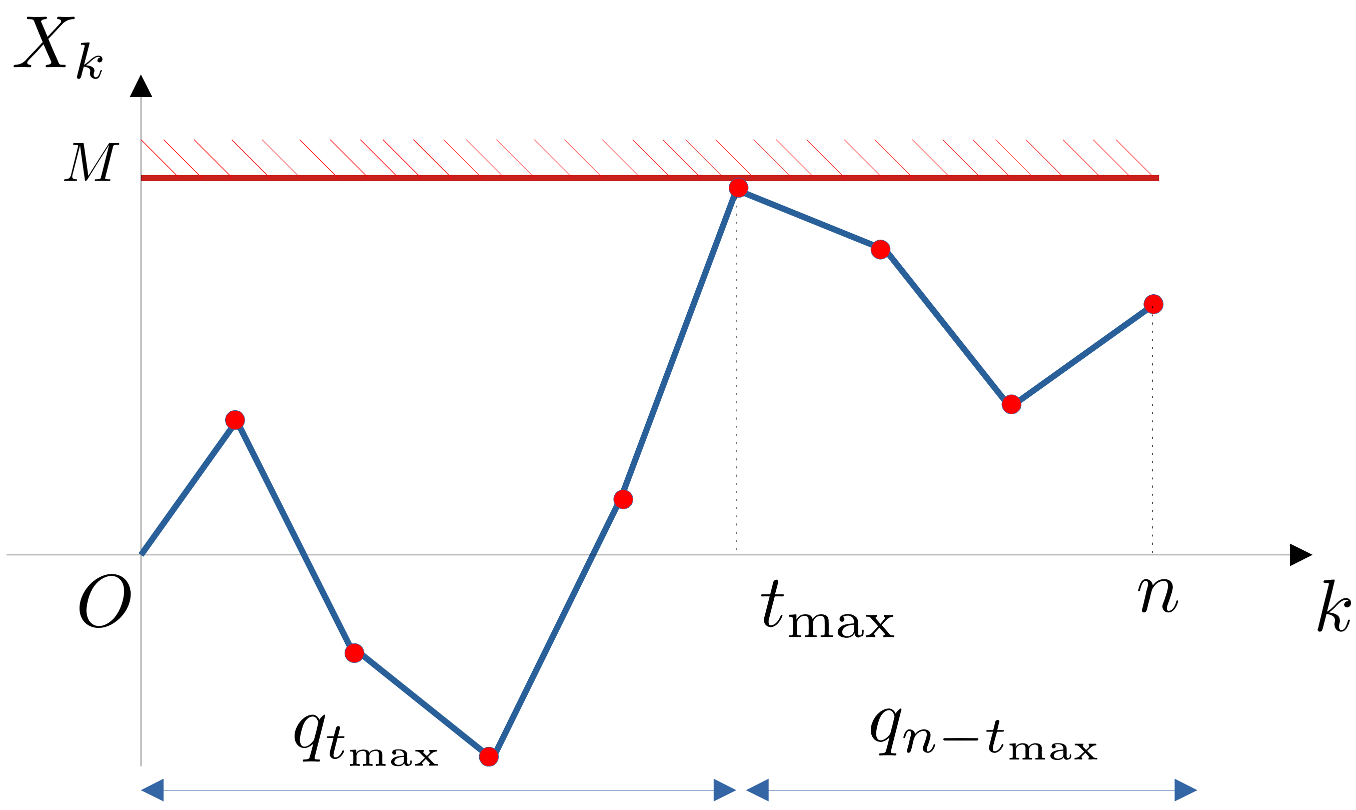
Time of the maximum



$$q_k = \binom{2k}{k} 2^{-2k}$$

$$P(t_{\max}|n) = q_{t_{\max}} q_{n-t_{\max}}$$

Time of the maximum



UNIVERSAL!

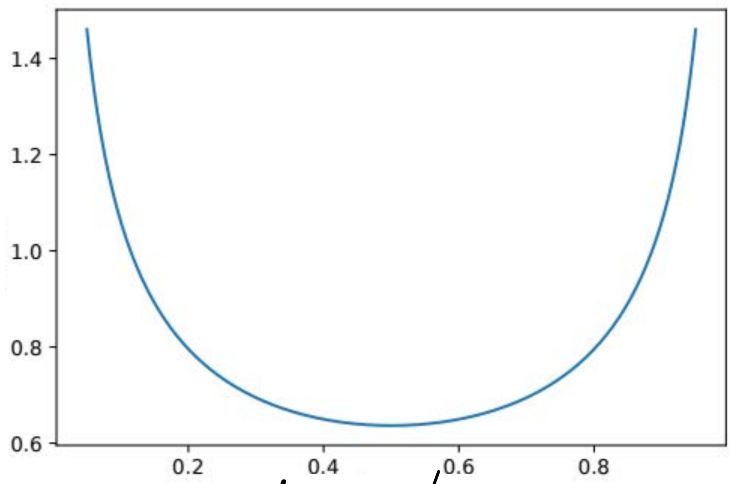
Independent of ***p(η)*** for any ***n***

$$q_k = \binom{2k}{k} 2^{-2k}$$

$$\boxed{P(t_{\max}|n) = q_{t_{\max}} q_{n-t_{\max}}}$$

Time of the maximum of a random walk

$P(t_{\max}|n)$

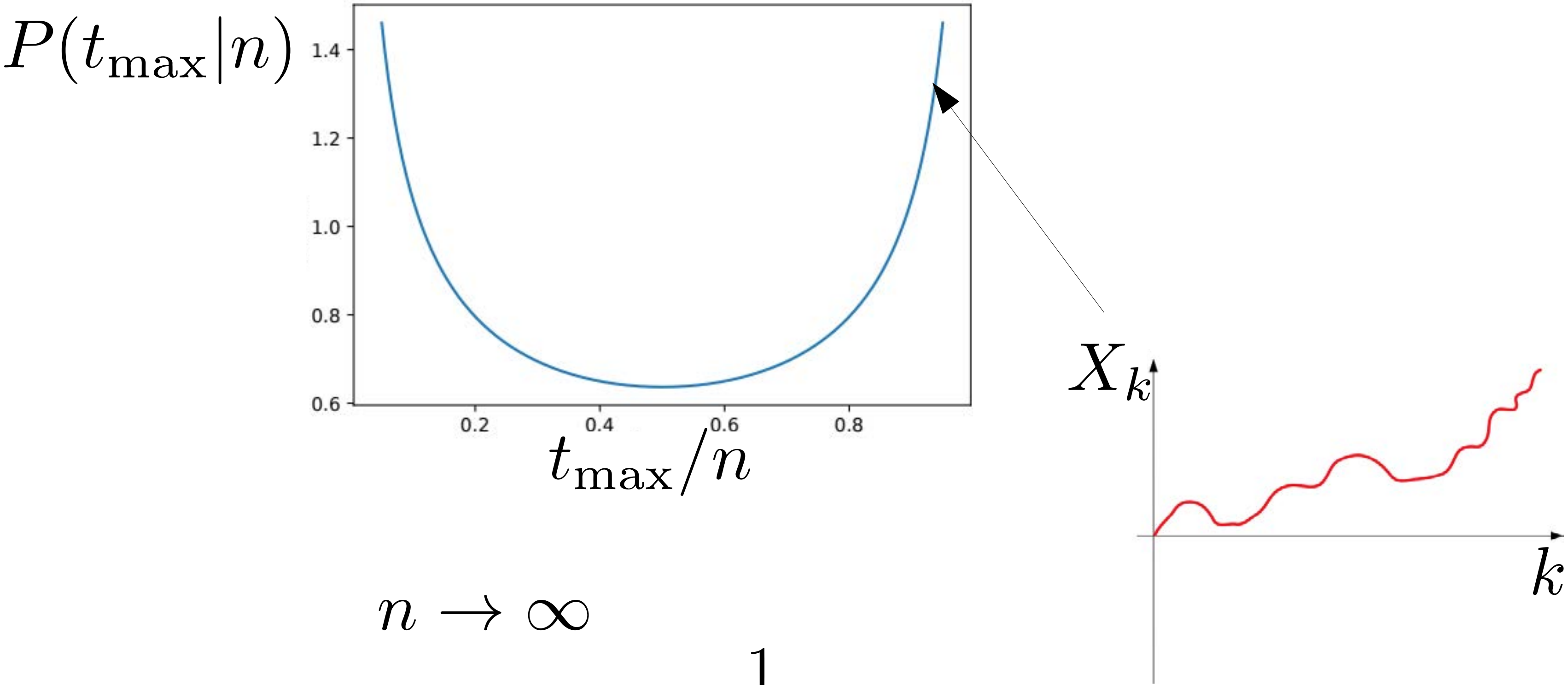


$t_{\max}/n$

$$n \to \infty$$

$$P(t_{\max}|n) \approx \frac{1}{\pi\sqrt{t_{\max}(n - t_{\max})}}$$

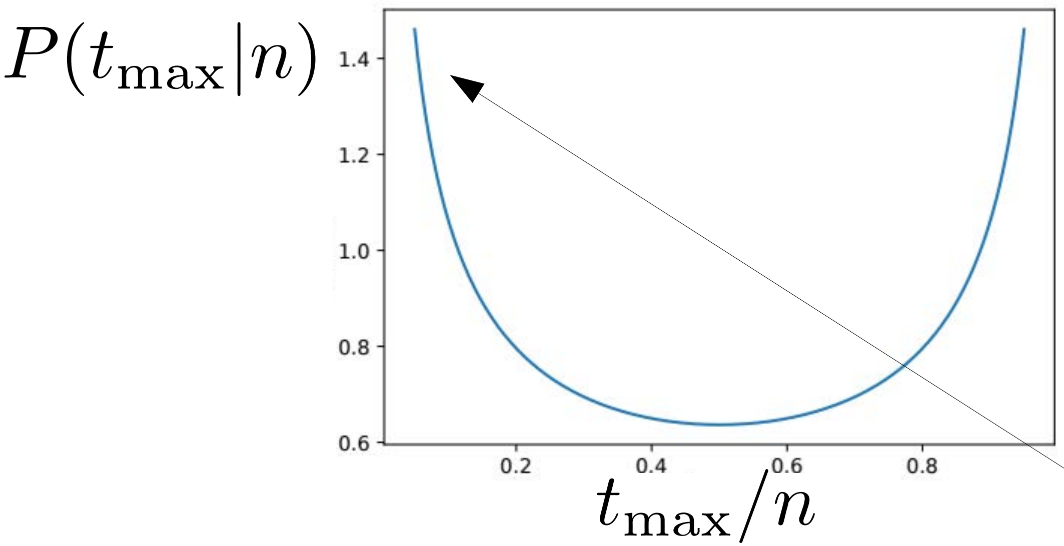
Time of the maximum of a random walk

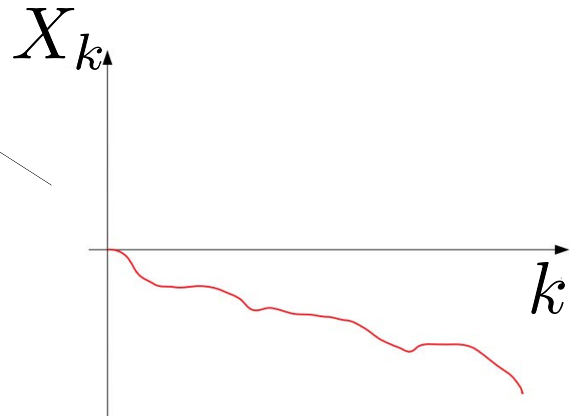


$$n \to \infty$$

$$P(t_{\max}|n) \approx \frac{1}{\pi\sqrt{t_{\max}(n - t_{\max})}}$$

Time of the maximum of a random walk

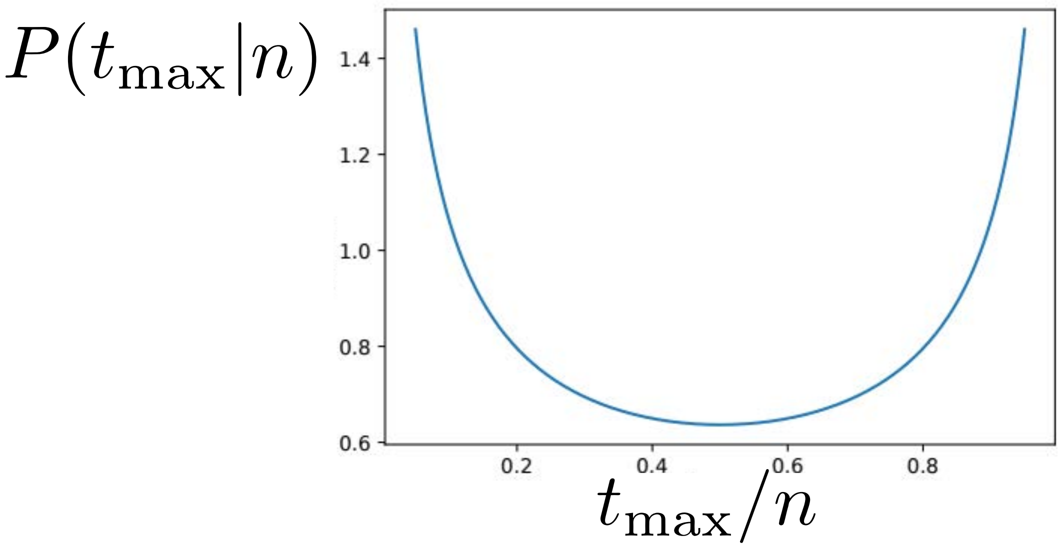




$$n \to \infty$$

$$P(t_{\max}|n) \approx \frac{1}{\pi\sqrt{t_{\max}(n - t_{\max})}}$$

Time of the maximum of a random walk

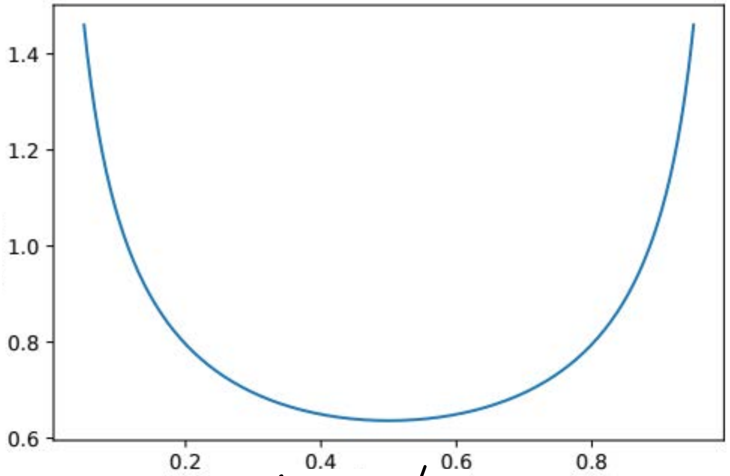


$$n \to \infty$$

$$P(t_{\max}|n) \approx \frac{1}{\pi\sqrt{t_{\max}(n - t_{\max})}}$$

Time of the minimum of a random walk

$P(t_{\min}|n)$

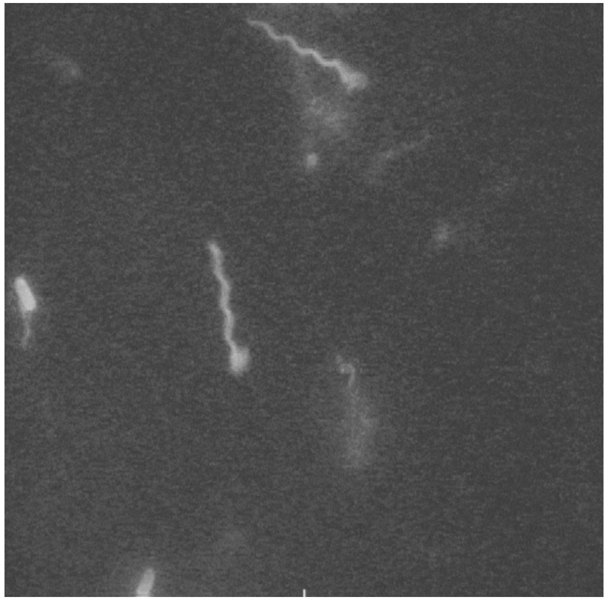


$t_{\min}/n$

$$P(t_{\min}|n) \approx \frac{1}{\pi\sqrt{t_{\min}(n - t_{\min})}}$$

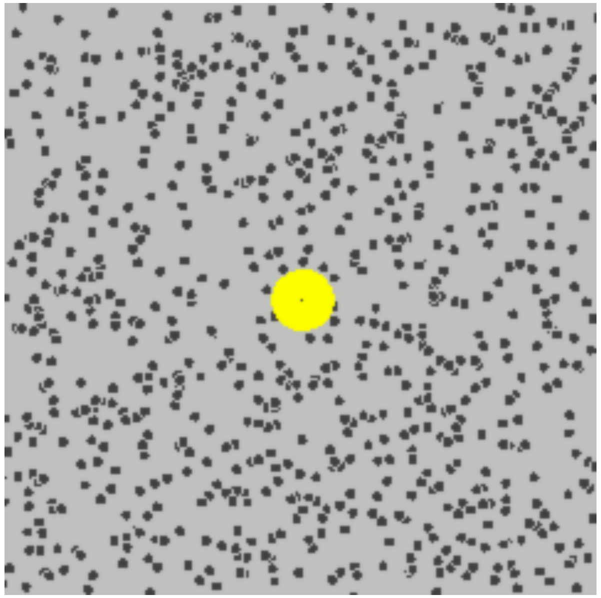
Active Particles

Active vs passive matter



Active Motion of
E. Coli bacteria
(Berg et al.)

- Persistent motion (absorbing energy from the environment)
- Out of equilibrium
- Alive

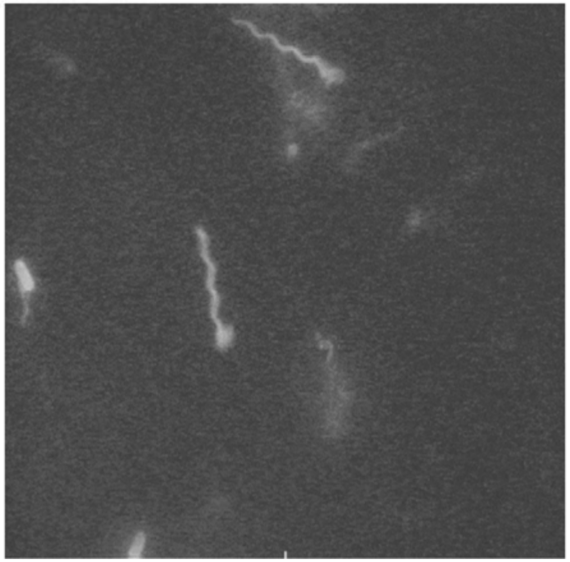


(Passive) Brownian
Motion

- Random motion (collisions with smaller molecules)
- Equilibrium
- Dead

Run-and-tumble particle (RTP) model

- Motion in d dimensions
- Random velocity $v \sim W(v)$
- Tumbling rate γ



E. Coli bacteria (Berg et al.)

[1] E. Orsingher, Random Oper. and Stoch. Equ. 3, 9 (1995).

Run-and-tumble particle (RTP) model

RTP model

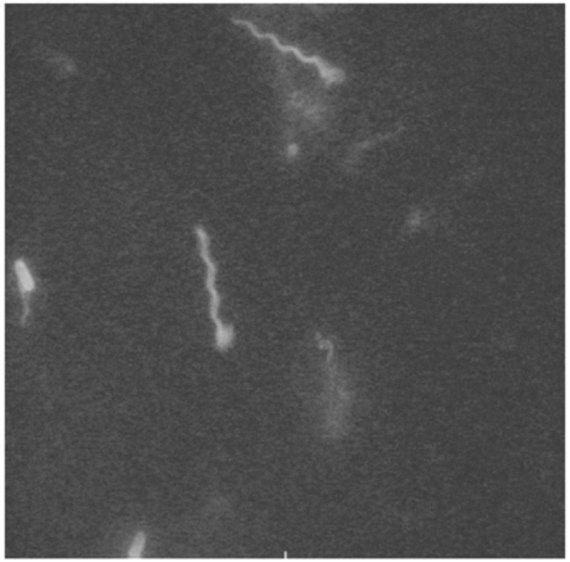
- Motion in **d** dimensions
- Random velocity **v~ W(v)**
- Tumbling rate γ

What is the **survival probability S(t)**?

S(t) = probability that the x-component of the particle does not change sign up to time t

Known for **d=1** and **constant velocity [1]**

$$S(t) = \frac{1}{2} e^{-\gamma t/2} \left(I_0(\gamma t/2) + I_1(\gamma t/2) \right)$$

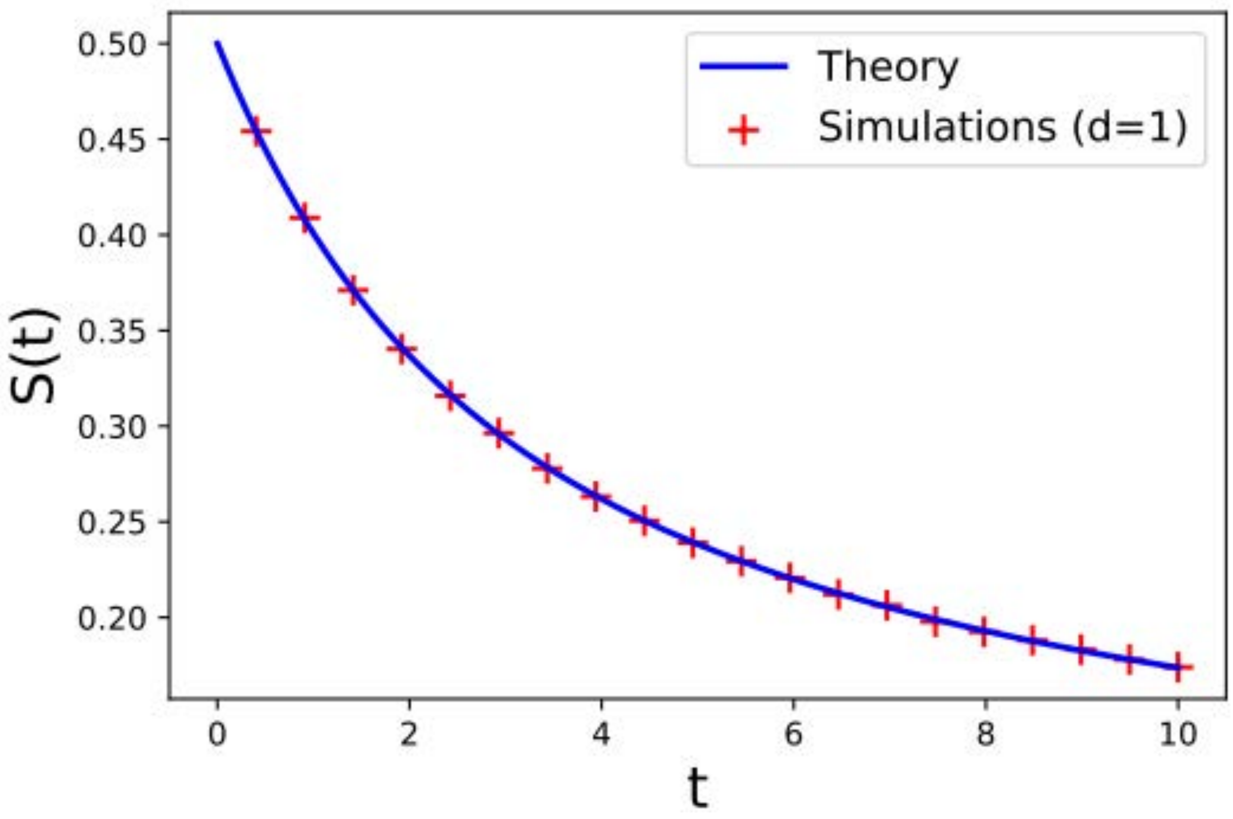


E. Coli bacteria (Berg et al.)

[1] E. Orsingher, Random Oper. and Stoch. Equ. 3, 9 (1995).

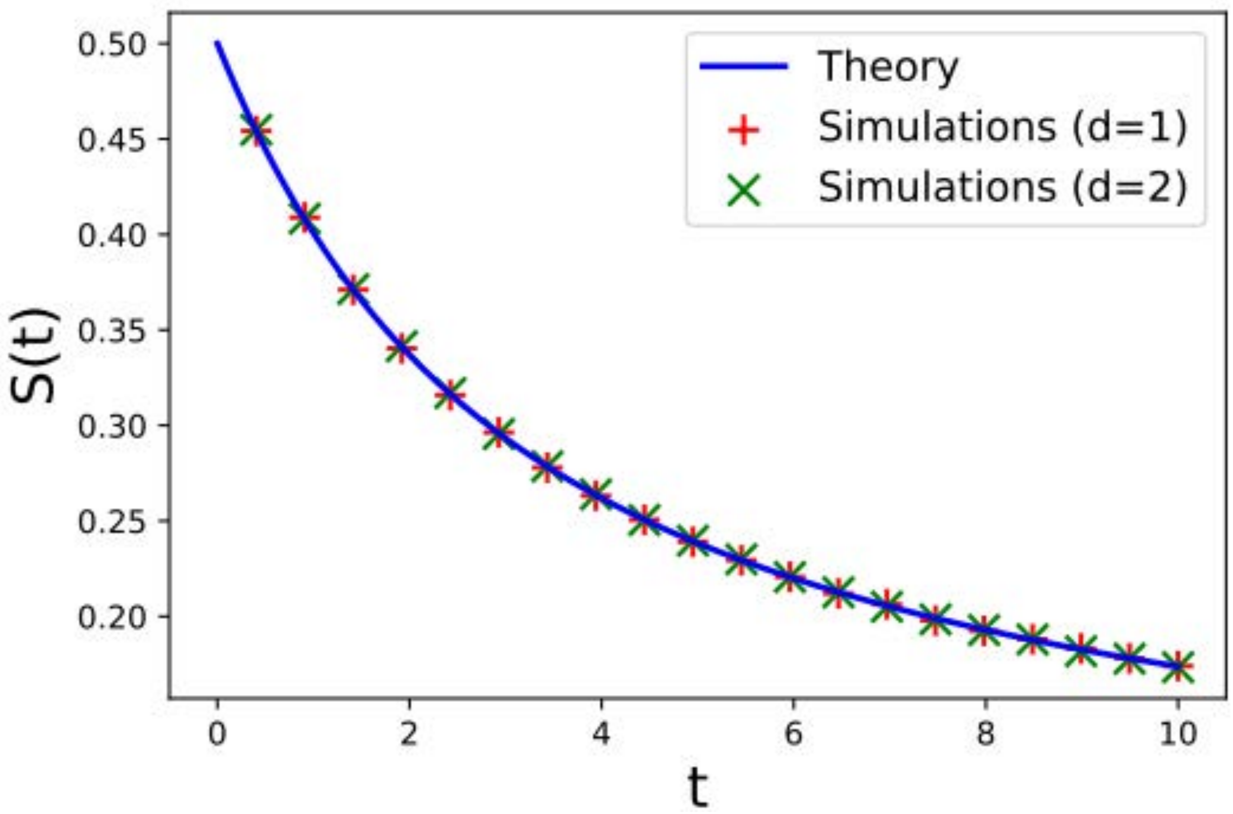
Numerical simulations

$$S(t) = \frac{1}{2}e^{-\gamma t/2}\left(I_0(\gamma t/2) + I_1(\gamma t/2)\right)$$



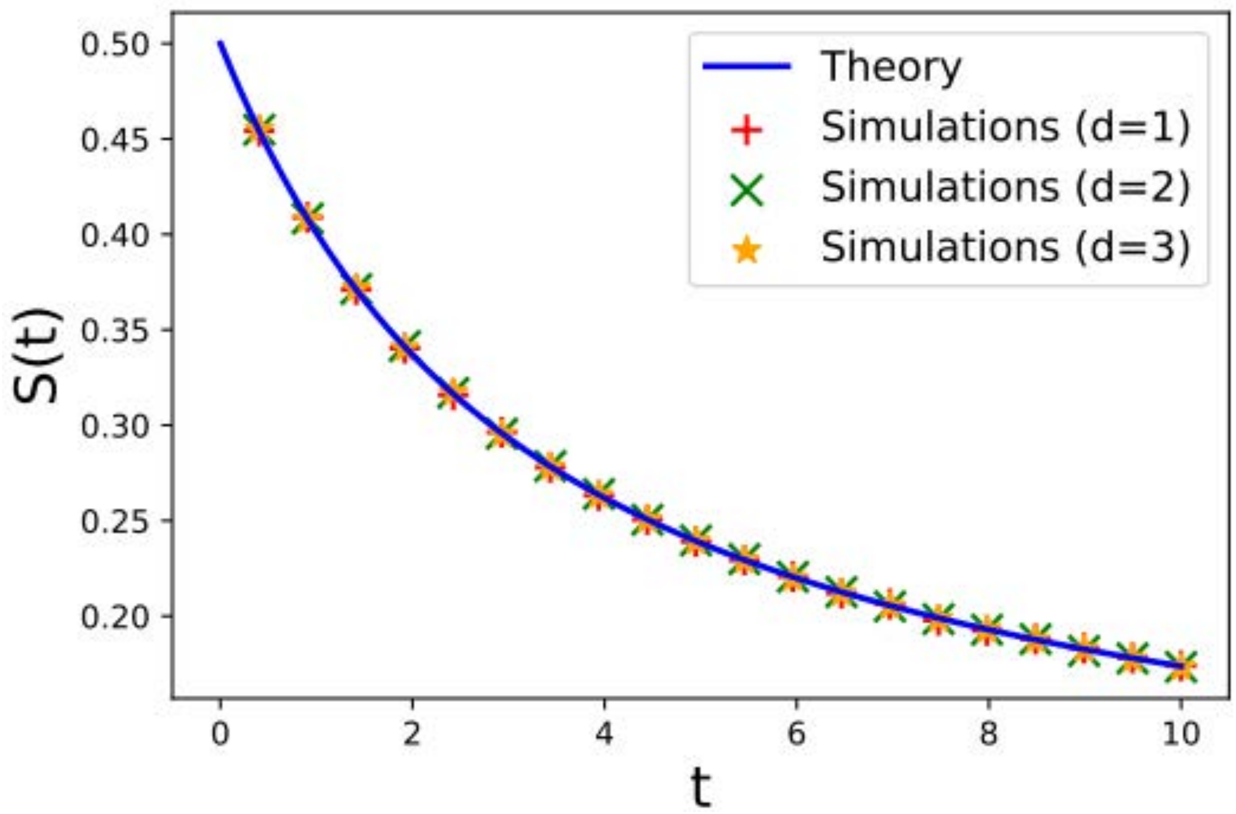
Numerical simulations

$$S(t) = \frac{1}{2} e^{-\gamma t/2} \left(I_0(\gamma t/2) + I_1(\gamma t/2) \right)$$



Numerical simulations

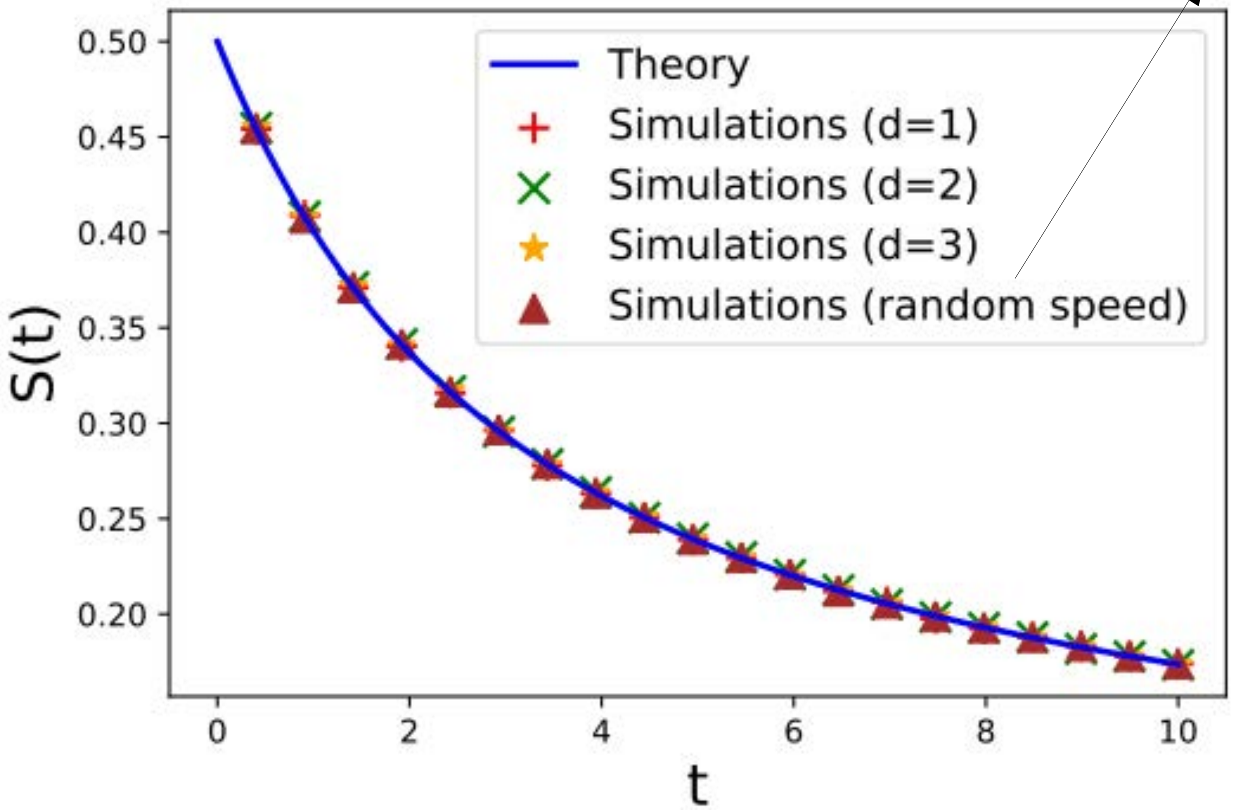
$$S(t) = \frac{1}{2} e^{-\gamma t/2} \left(I_0(\gamma t/2) + I_1(\gamma t/2) \right)$$



Numerical simulations

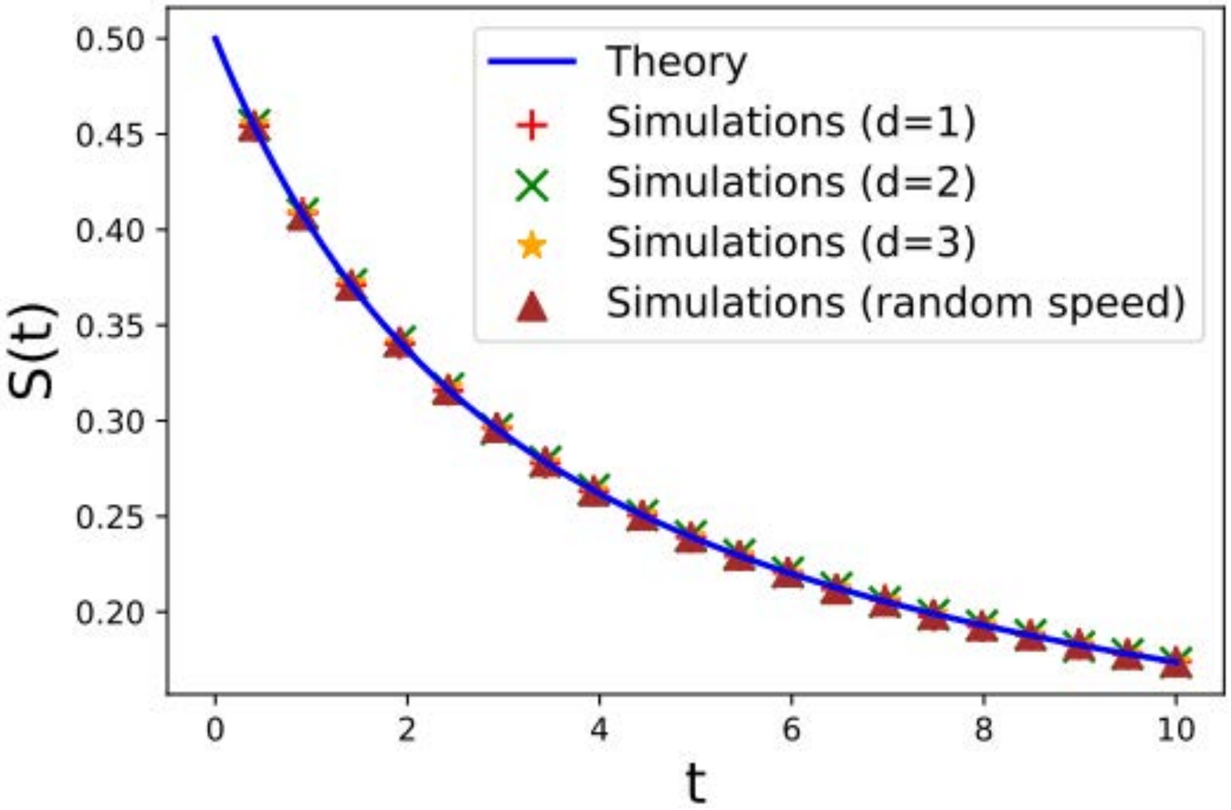
$$S(t) = \frac{1}{2} e^{-\gamma t/2} \left(I_0(\gamma t/2) + I_1(\gamma t/2) \right)$$

$$W(v) = \frac{2}{\pi(1+v^2)}$$



Numerical simulations

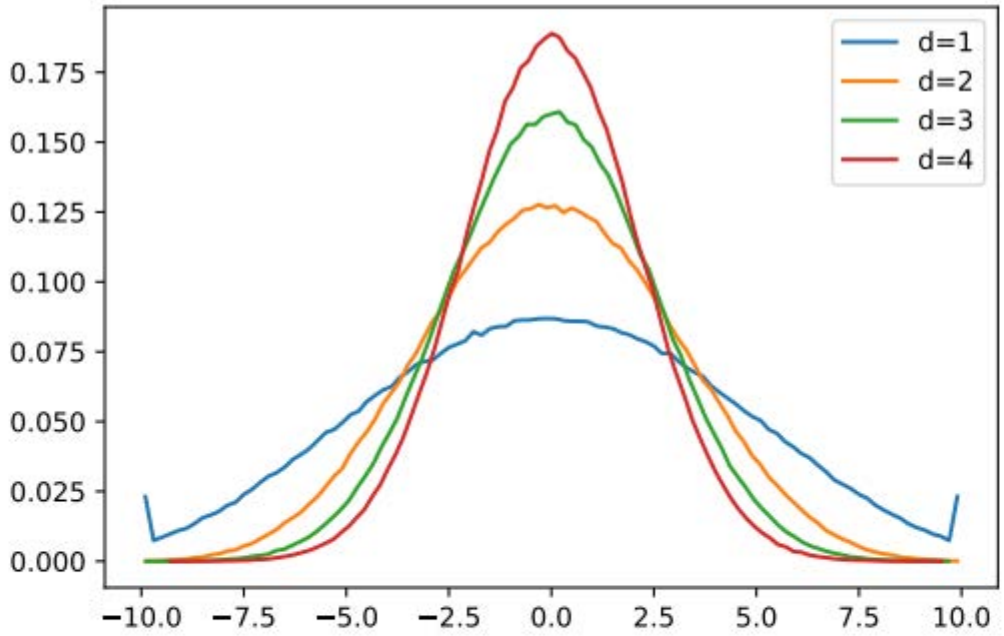
Universality



Position distribution

Not universal!

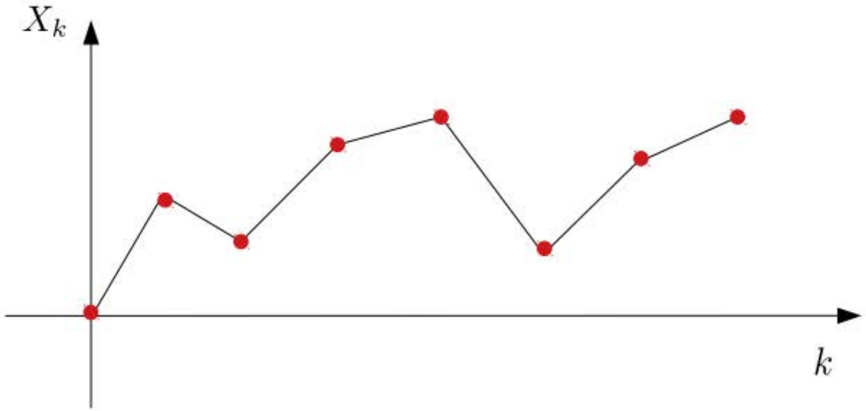
$P(X,t)$



X

$t = 10, \gamma = 1$

Sparre Andersen theorem



$$q_n = \binom{2n}{n} 2^{-2n}$$

Independent of ***p(η)*** for any ***n***

It is not obvious how to apply this result to the RTP model

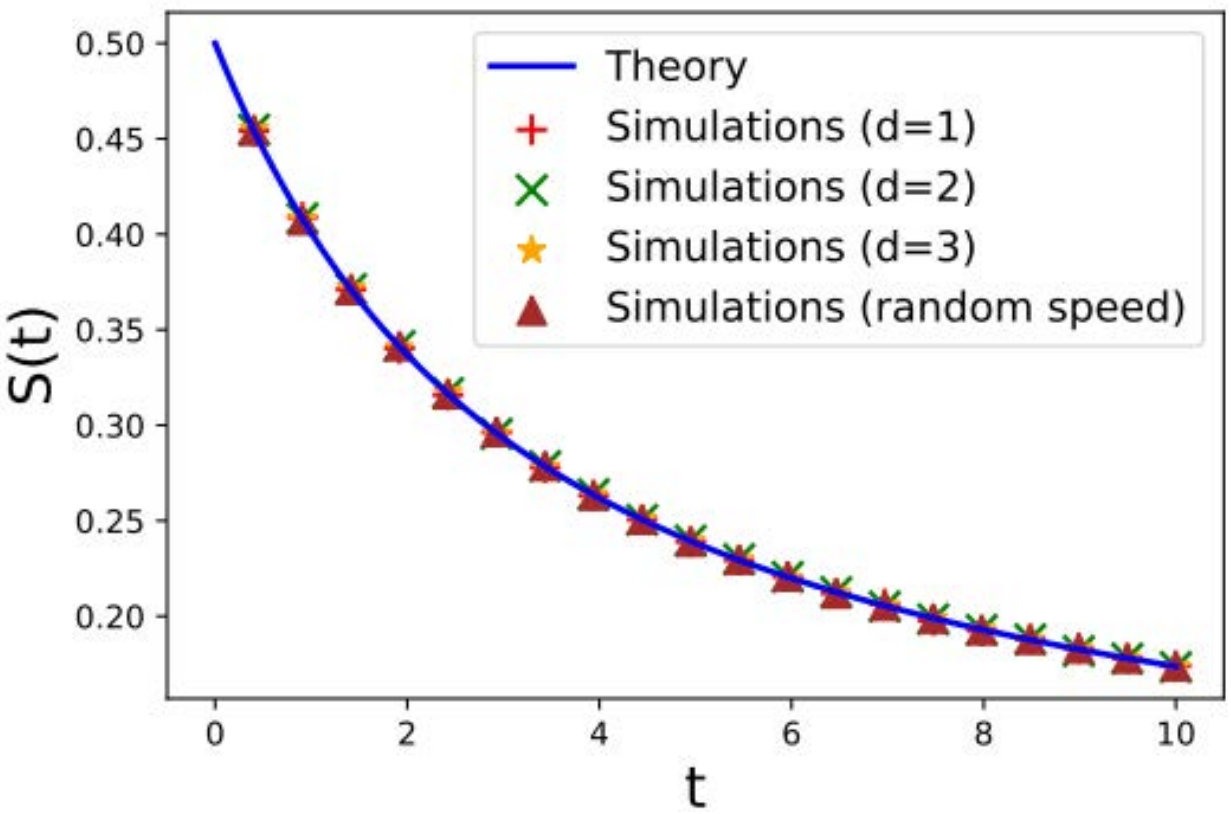
↓

Mapping to a discrete-time random walk (in Laplace space)

[1] E. Sparre Andersen, Math. Scand. **2**, 195 (1954).

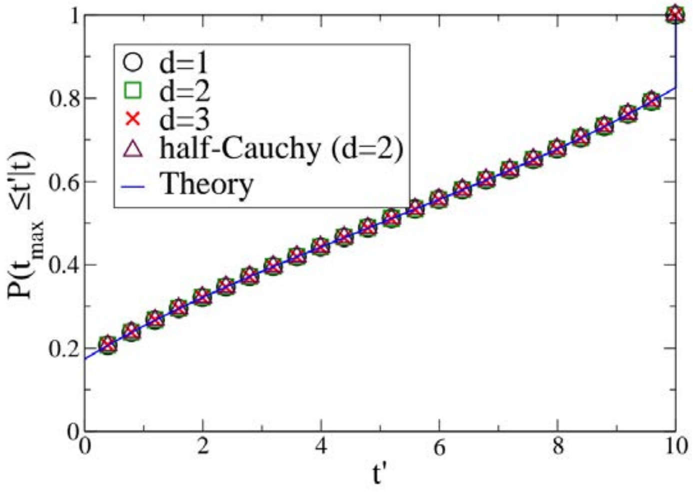
Sparre Andersen Theorem

Universality

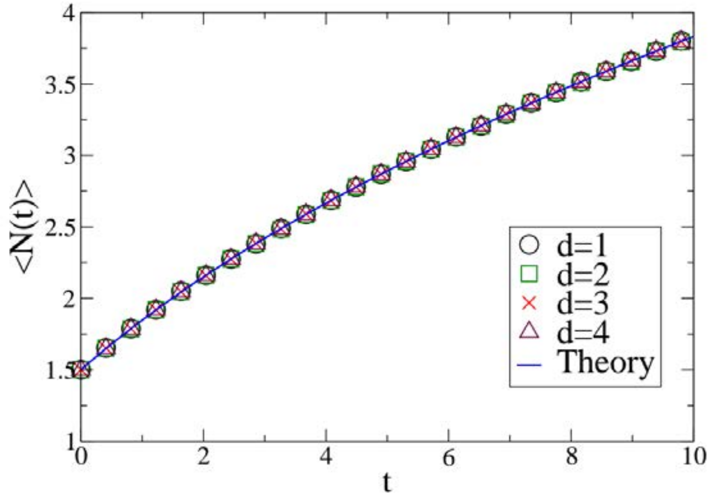


Extreme Value Statistics

Time of the maximum [1]



Average number of records [1]



Also universal!

[1] FM, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. E 102, 042133 (2020).

Extreme Value Statistics

- Using simple models, we can learn a lot about the statistics of extreme events, records, ...

- Many applications to physics, finance, evolution theory, ...

- Often we find universal results, independent of the details of the model

- Can we find a general theory of rare events in correlated systems?

Thank you for the attention