Extreme Statistics of Extreme Values

Eli Ben-Naim
Los Alamos National Laboratory

with
Paul Krapivsky (Boston University)
Pearson Miller (Yale University)

EB and P.L. Krapivsky, Phys. Rev. Lett. **113**, 030604 (2014)
P.W. Miller and EB, J. Stat. Mech. P10025 (2013)

Talk, publications available from: http://cnls.lanl.gov/~ebn

Fluctuations in Population Biology, Leiden, August 15, 2014

First-Passage Properties of Records

I. Records of correlated random variables
II. Records of uncorrelated random variables

Two different problems, similar phenomenology

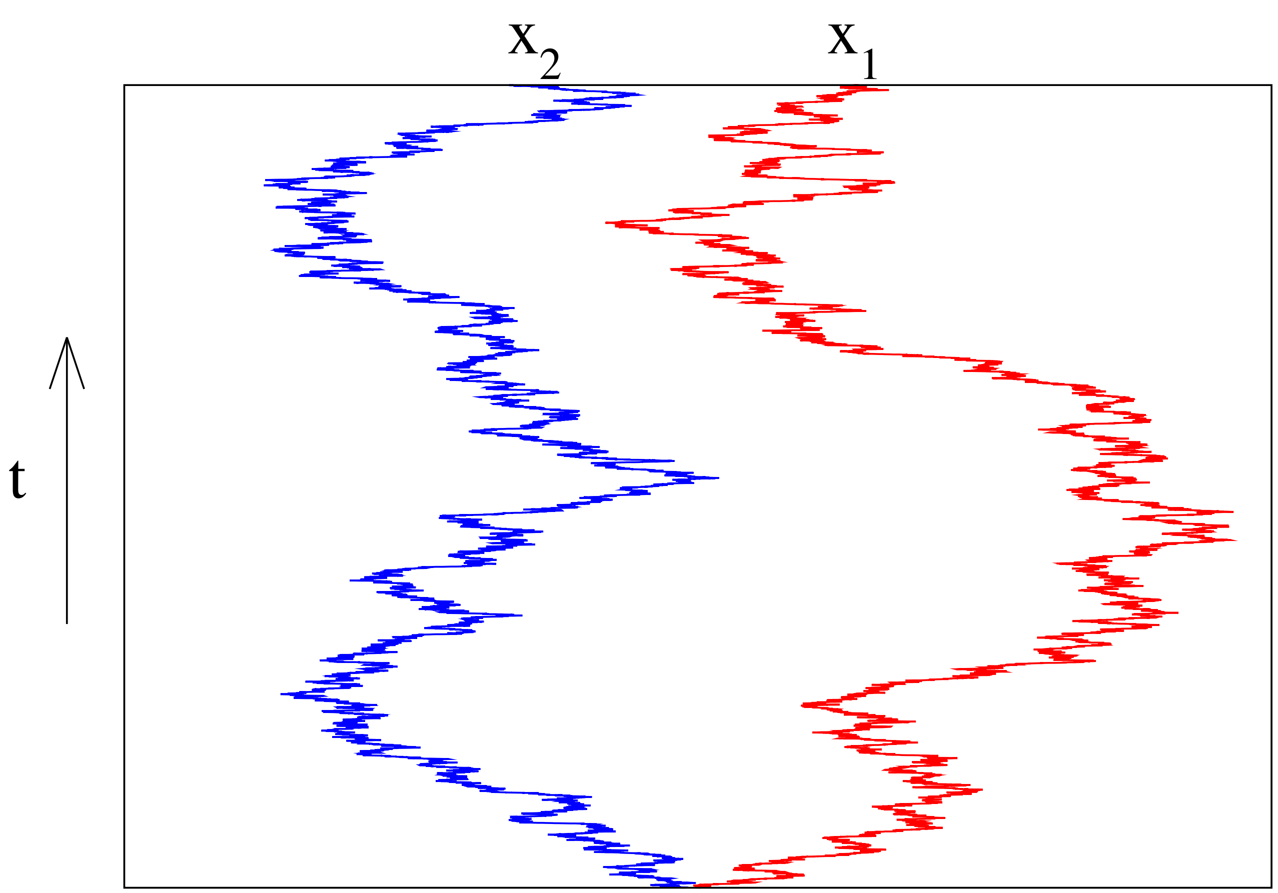
Records & Extreme value statistics

New frontier in nonequilibrium statistical physics

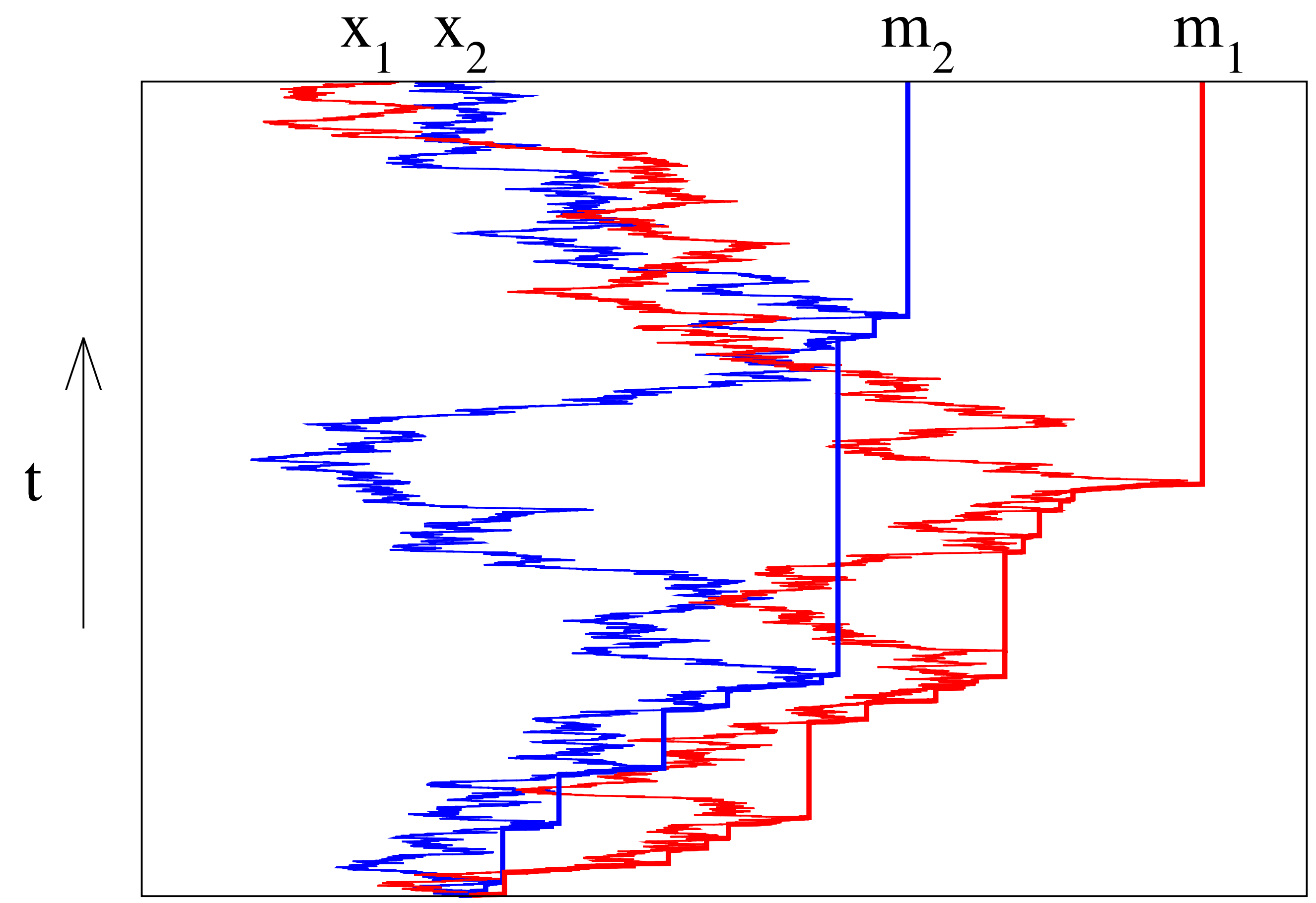
- Biological evolution — Bak, Derrida, Flyvbjerg, Jain, Krug
- Population dynamics — Kamenev, Meerson, Dykman, Doering
- Brownian motion — Comtet, Majumdar, Krug, Redner
- Surface growth — Spohn, Halpin-Healy, Majumdar, Schehr
- Transport — Mallick, Krapivsky, Derrida, Lebowitz, Speer
- Climate — Bunde, Havlin, Krug, Wergen, Redner
- Earthquakes — Davidsen, Sornette, Newman, Turcotte, EB
- Finance — Bouchaud, Stanley, Majumdar

Part I
Maxima of Brownian Particles
(records of correlated variables)

Brownian Positions

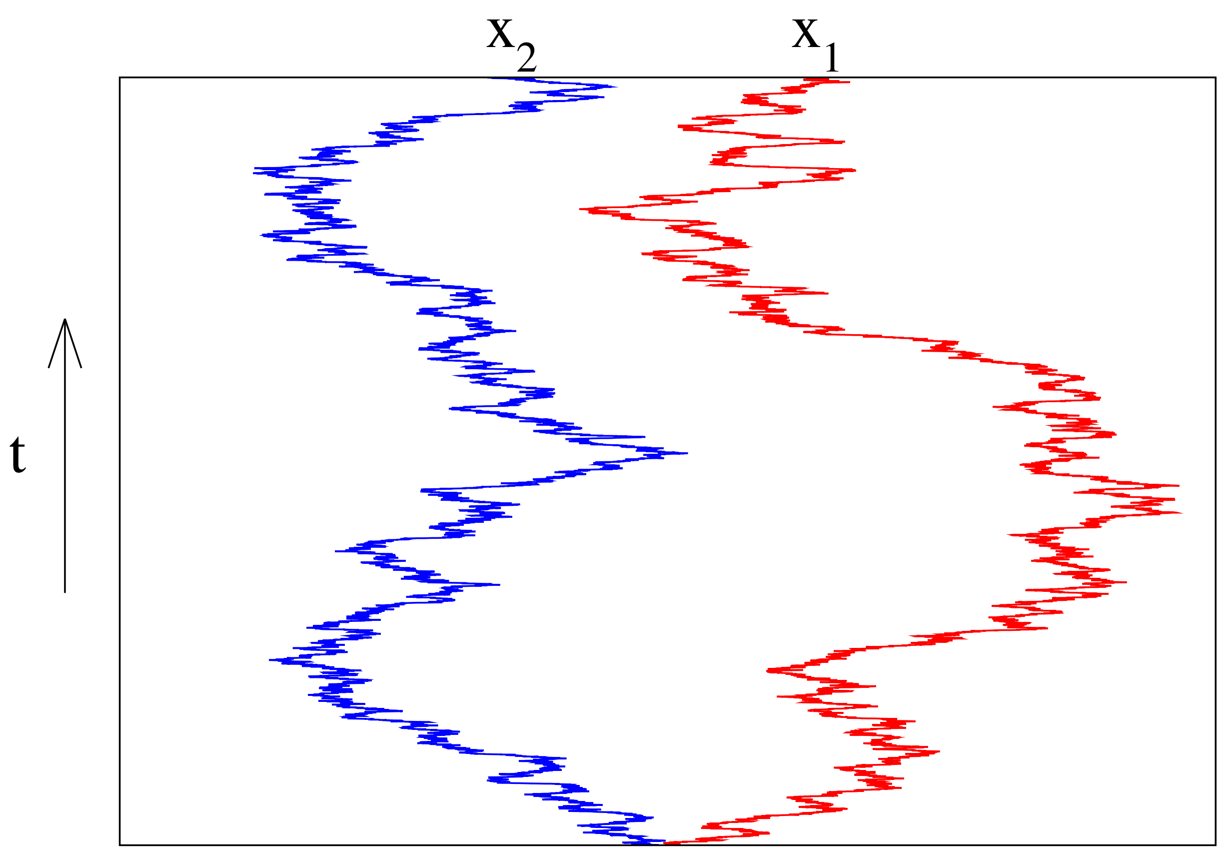


Brownian Maxima



First-Passage Kinetics: Brownian Positions

Probability two Brownian particle do not meet



- Universal probability

Sparre Andersen 53

$$S_t = \binom{2t}{t} 2^{-t}$$

- Asymptotic behavior

Feller 68

$$S \sim t^{-1/2}$$

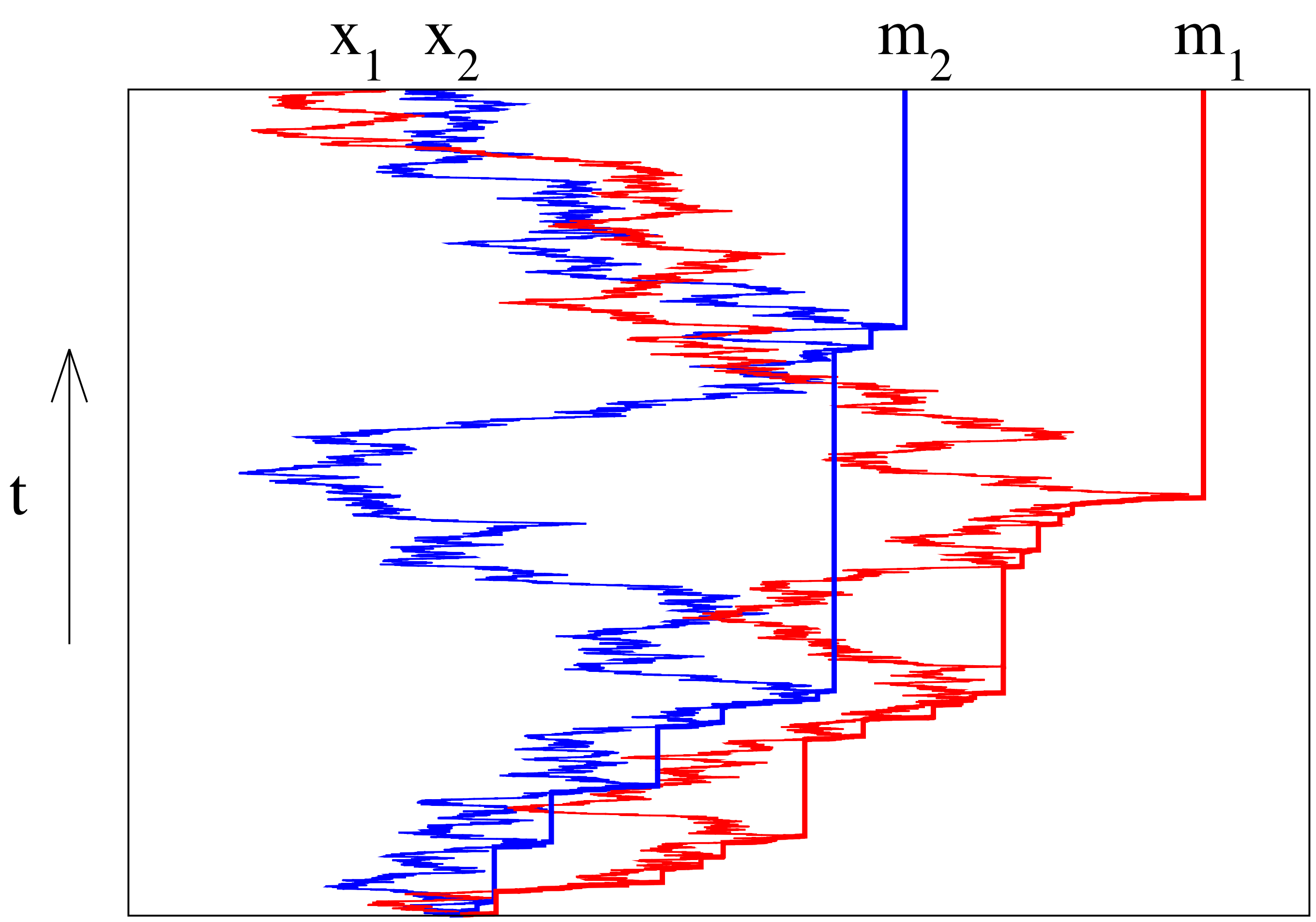
Behavior holds for Levy flights, different mobilities, etc

Universal first-passage exponent

S. Redner, A guide to First-Passage Processes 2001

First-Passage Kinetics: Brownian Maxima

Probability maximal positions remain ordered



- Numerical simulations

$$S \sim t^{-\beta}$$

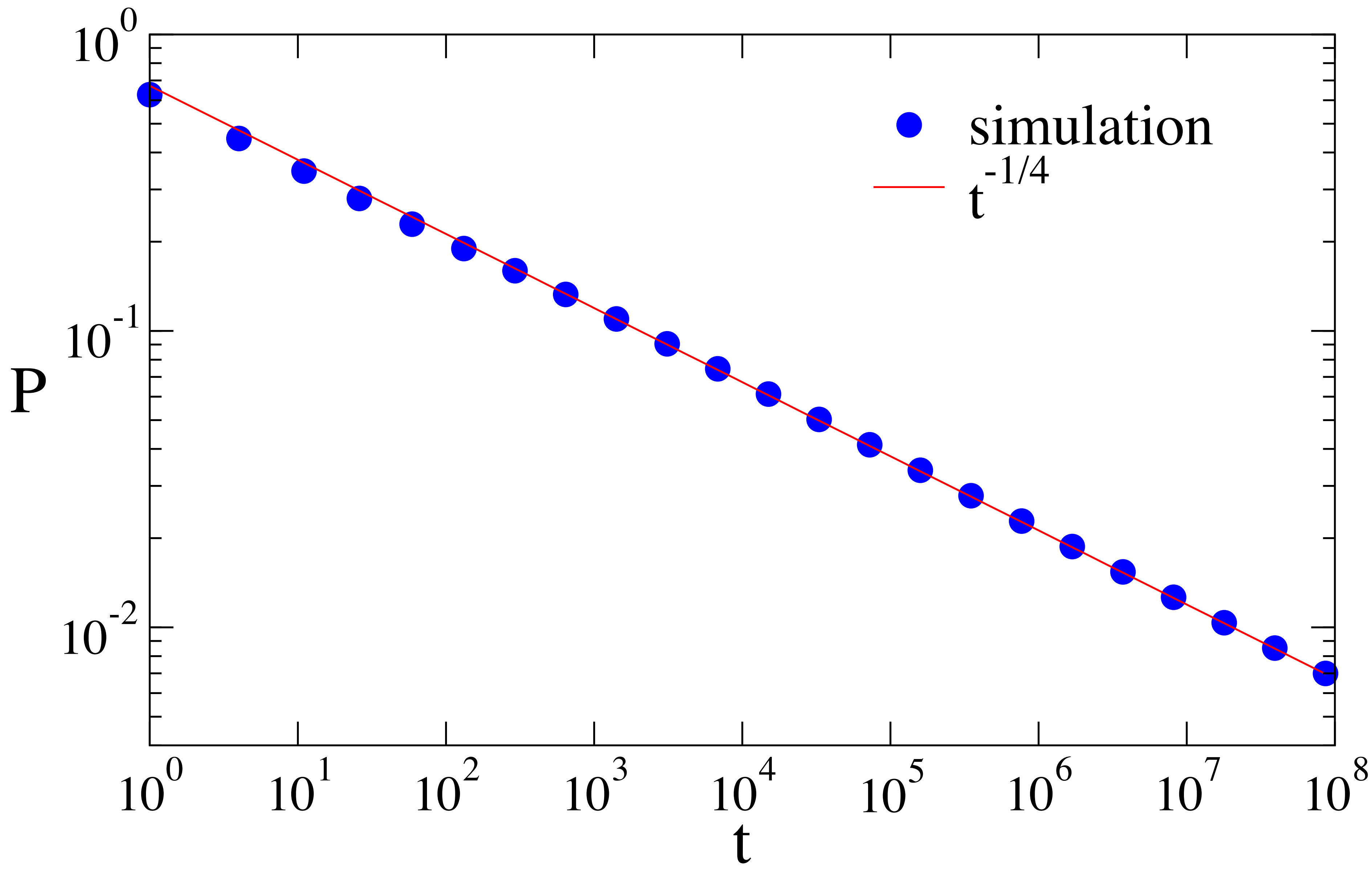
- First-passage exponent

$$\beta = 0.2503 \pm 0.0005$$

Is ¼ exact? and if it is, does ¼=½×½?

Is exponent universal?

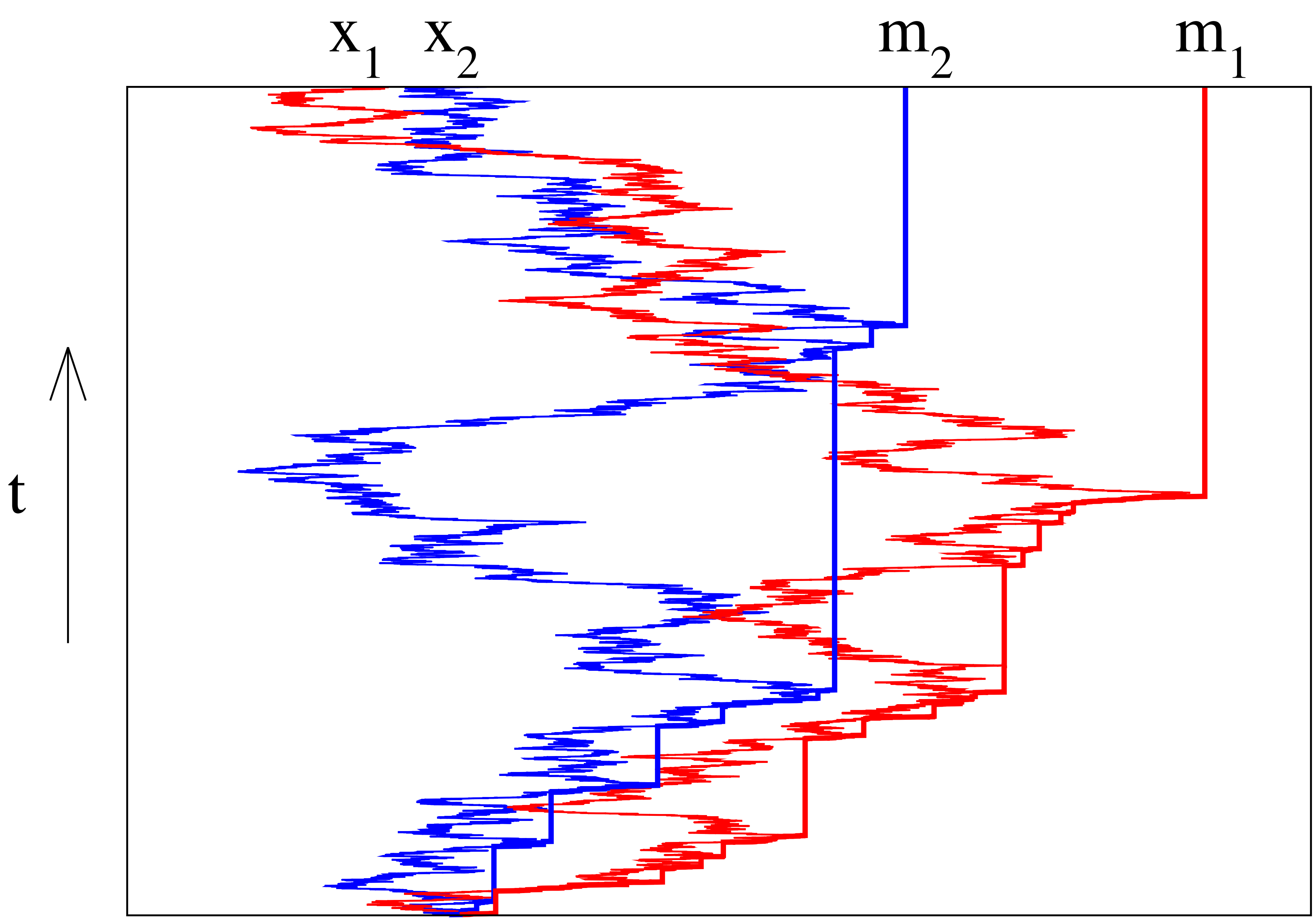
Monte Carlo simulations



Hints at a rational exponent

Inconclusive due to slow convergence

$$m_1 > m_2 \quad \text{if and only if} \quad m_1 > x_2$$



From four variables to three

- Four variables: two positions, two maxima

$$m_1 > x_1 \quad \text{and} \quad m_2 > x_2$$

- The two maxima must always be ordered

$$m_1 > m_2$$

- Key observation: trailing maximum is irrelevant!

$$m_1 > m_2 \quad \text{if and only if} \quad m_1 > x_2$$

- Three variables: two positions, one maximum

$$m_1 > x_1 \quad \text{and} \quad m_1 > x_2$$

From three variables to two

- Introduce two distances from the maximum

$$u = m_1 - x_1 \quad \text{and} \quad v = m_1 - x_2$$

- Both distances undergo Brownian motion

$$\frac{\partial \rho(u,v,t)}{\partial t} = D\nabla^2 \rho(u,v,t)$$

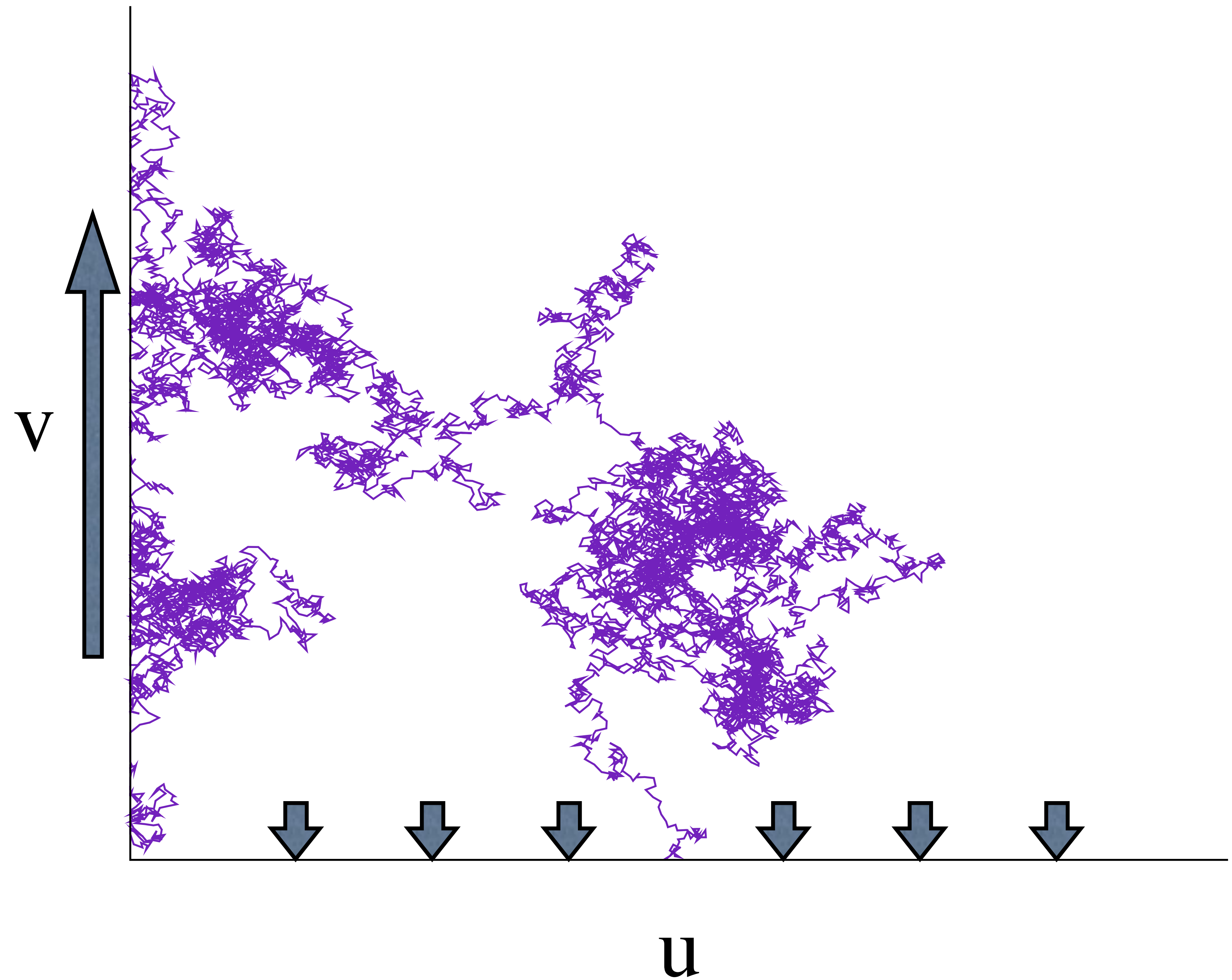
- Boundary conditions: (i) absorption (ii) advection

$$\rho\Big|_{v=0} = 0 \quad \text{and} \quad \left(\frac{\partial \rho}{\partial u} - \frac{\partial \rho}{\partial v}\right)\Bigg|_{u=0} = 0$$

- Probability maxima remain ordered

$$P(t) = \int_0^\infty \int_0^\infty du\, dv\, \rho(u,v,t)$$

Diffusion in corner geometry



"Backward" evolution

- Study evolution as function of **initial conditions**

$$P \equiv P(u_0, v_0, t)$$

- Obeys diffusion equation

$$\frac{\partial P(u_0, v_0, t)}{\partial t} = D\nabla^2 P(u_0, v_0, t)$$

- Boundary conditions: (i) absorption (ii) advection

$$P\Big|_{v_0=0} = 0 \qquad \text{and} \qquad \left(\frac{\partial P}{\partial u_0} + \frac{\partial P}{\partial v_0}\right)\Big|_{u_0=0} = 0$$

- Advection boundary condition is conjugate

Solution

- Use polar coordinates

$$r = \sqrt{u_0^2 + v_0^2} \quad \text{and} \quad \theta = \arctan \frac{v_0}{u_0}$$

- Laplace operator

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}$$

- Boundary conditions: (i) absorption (ii) advection

$$P\Big|_{\theta=0} = 0 \quad \text{and} \quad \left(r\,\frac{\partial P}{\partial r} - \frac{\partial P}{\partial \theta}\right)\Big|_{\theta=\pi/2} = 0$$

- dimensional analysis + power law + separable form

$$P(r,\theta,t) \sim \left(\frac{r^2}{Dt}\right)^{\beta} f(\theta)$$

Selection of exponent

- Exponent related to eigenvalue of <u>angular</u> part of Laplacian

$$f''(\theta) + (2\beta)^2 f(\theta) = 0$$

- Absorbing boundary condition selects solution

$$f(\theta) = \sin(2\beta\theta)$$

- Advection boundary condition selects exponent

$$\tan(\beta\pi) = 1$$

- First-passage probability

$$P \sim t^{-1/4}$$

General diffusivities

ben Avraham
Leyvraz 88

- Particles have diffusion constants D_1 and D_2

$$(x_1, x_2) \to (\widehat{x}_1, \widehat{x}_2) \quad \text{with} \quad (\widehat{x}_1, \widehat{x}_2) = \left(\frac{x_1}{\sqrt{D_1}}, \frac{x_2}{\sqrt{D_2}}\right)$$

- Condition on maxima involves ratio of mobilities

$$\sqrt{\frac{D_1}{D_2}}\, \widehat{m}_1 > \widehat{m}_2$$

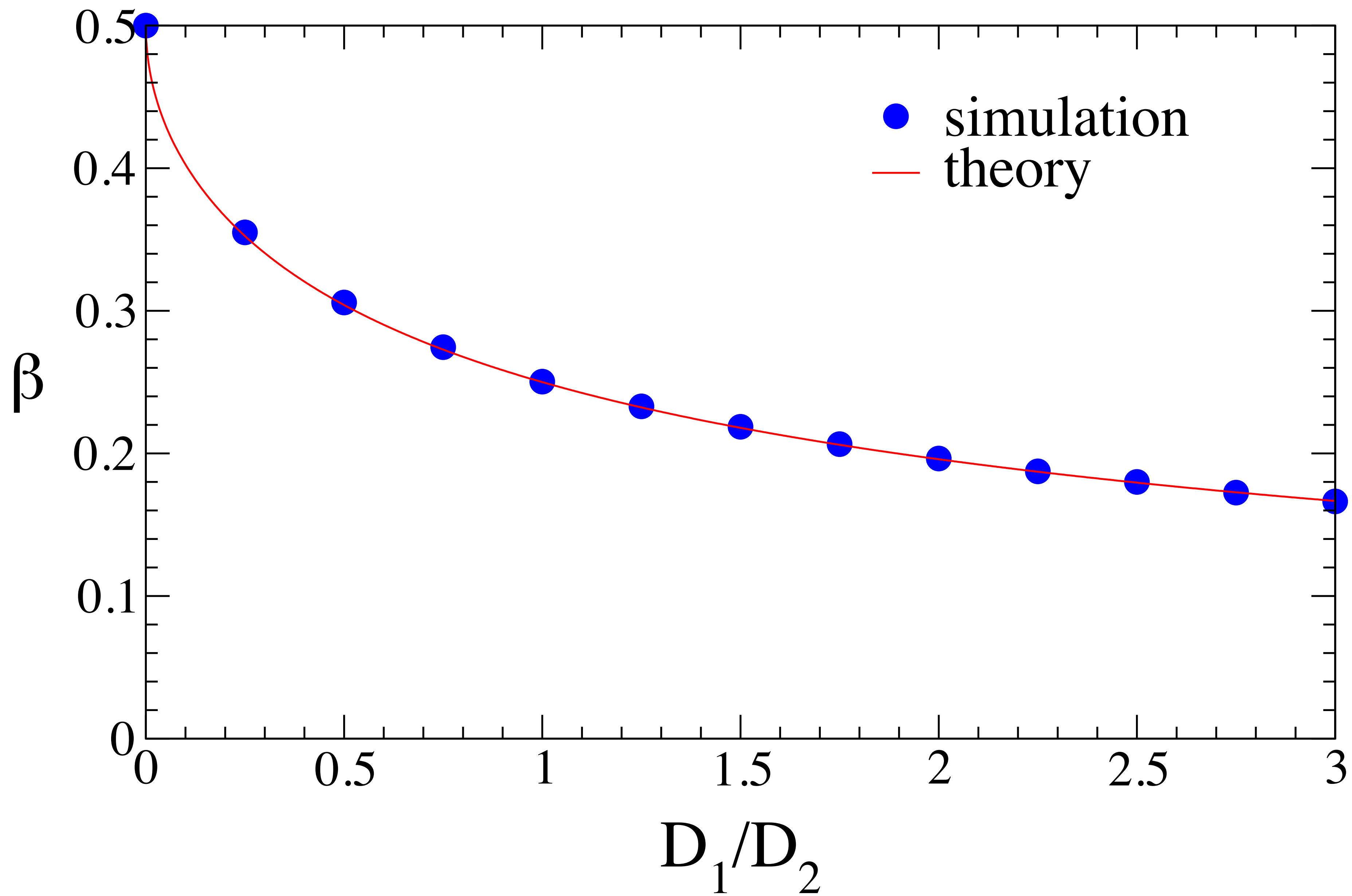
- Analysis straightforward to repeat

$$\sqrt{\frac{D_1}{D_2}} \tan\left(\beta\pi\right) = 1$$

- First-passage exponent: nonuniversal, mobility-dependent

$$\beta = \frac{1}{\pi} \arctan \sqrt{\frac{D_2}{D_1}}$$

Numerical verification



Perfect agreement

simulations can confirm a line (but not a point)

Properties

- Depends on ratio of diffusion constants

$$\beta(D_1, D_2) \equiv \beta\left(\frac{D_1}{D_2}\right)$$

- Bounds: involve one immobile particle

$$\beta(0) = \frac{1}{2} \qquad \beta(\infty) = 0$$

- Rational for special values of diffusion constants

$$\beta(1/3) = 1/3 \qquad \beta(1) = 1/4 \qquad \beta(3) = 1/6$$

- Duality: between "fast chasing slow" and "slow chasing fast"

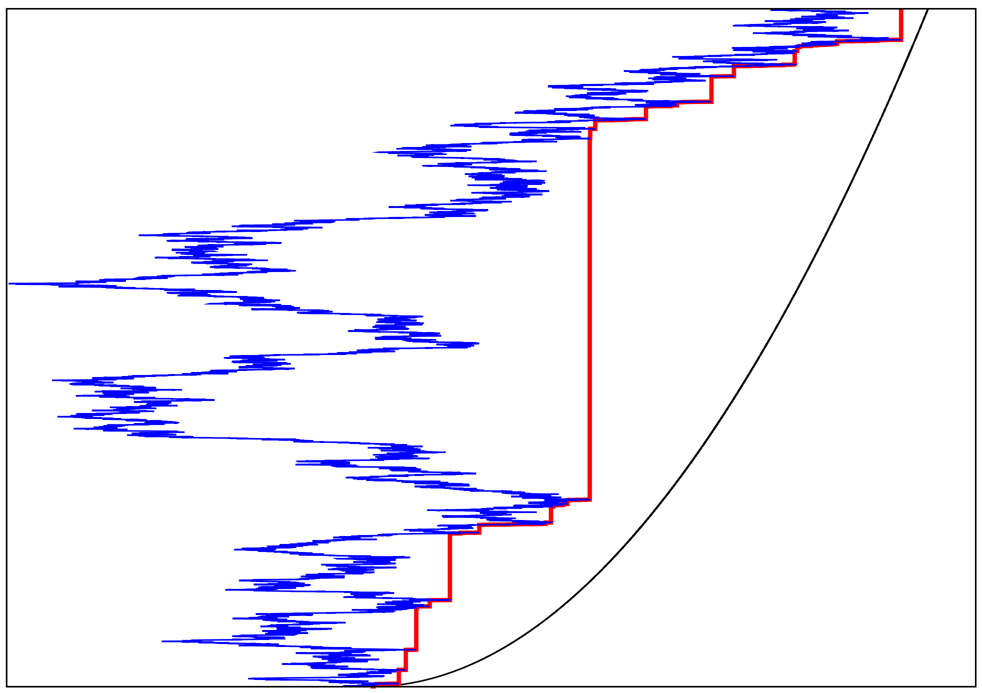
$$\beta\left(\frac{D_1}{D_2}\right) + \beta\left(\frac{D_2}{D_1}\right) = \frac{1}{2}$$

Alternating kinetics: slow-fast-slow-fast

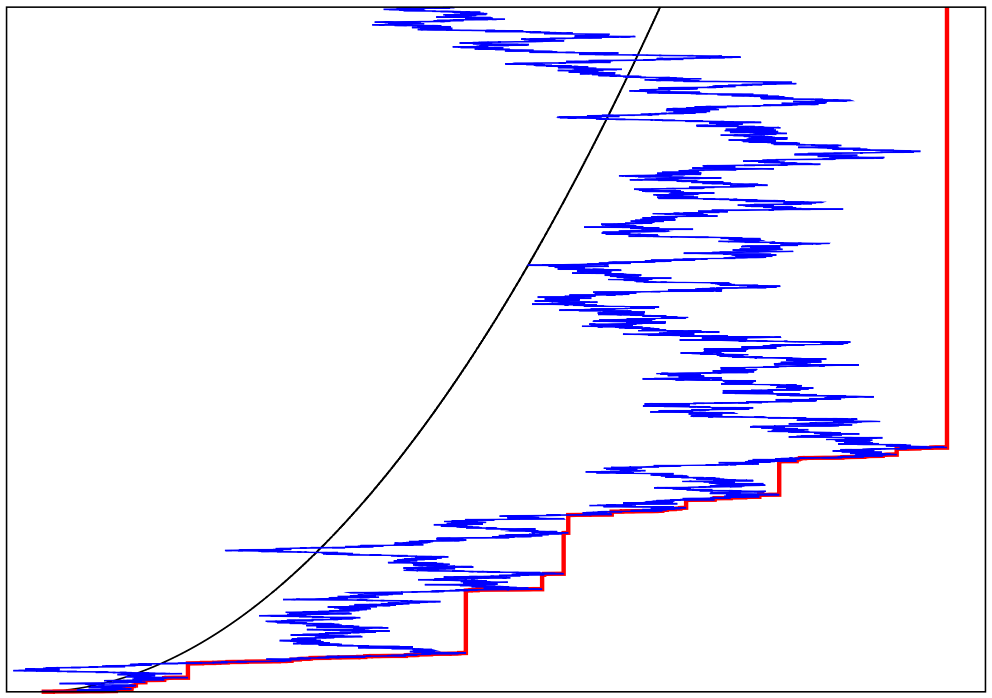
Inferior & Superior walks

Maximum is always behind or ahead of the average maximum of a Brownian particle

Krapivsky & Redner 95
EB & Krapivsky 14



$D_{2\beta}\left(-\sqrt{2/\pi}\right)=0 \quad \beta = 0.241608$



$D_{2\beta+1}\left(\sqrt{2/\pi}\right)=0 \quad \beta = 0.382258$

Different mobilities: neglect fluctuations in maximum of slower particle (represent maximum by its average) and obtain limits

$$\beta \simeq \begin{cases} \frac{1}{2} - \frac{1}{\pi}\sqrt{D_1/D_2} & D_1 \ll D_2 \\ \frac{1}{\pi}\sqrt{D_2/D_1} & D_2 \ll D_1 \end{cases}$$

Multiple particles

- All maxima perfectly ordered

Fisher & Huse 88

$$m_1 > m_2 > m_3 > \cdots > m_n$$

- Only one leader

Bramson & Griffith 91

$$m_1 > m_2 \quad m_1 > m_3 \quad \cdots \quad m_1 > m_n$$

- Only one laggard

ben Avraham & Redner 03

$$m_1 > m_n \quad m_2 > m_n \quad \cdots \quad m_{n-1} > m_n$$

- Three families of first-passage exponents

$$A_n \sim t^{-\alpha_n} \qquad B_n \sim t^{-\beta_n} \qquad C_n \sim t^{-\gamma_n}$$

Exponents are eigenvalues of
"angular" component of Laplace in n dimensions

Three families of exponents

- Simulation results: maxima vs positions

Grassberger 03
ben Avraham 03
EB & Krapivsky 10

	maxima			positions		
n	α_n	β_n	γ_n	a_n	b_n	c_n
2	1/4	1/4	1/4	1/2	1/2	1/2
3	0.653	0.432	0.335	3/2	3/4	3/8
4	1.13	0.570	0.376	3	0.91	0.306
5	1.60	0.674	0.401	5	1.02	0.265
6	2.01	0.759	0.417	15/2	1.11	0.234

- Positions: one family is known

Fisher & Huse 88

$$b_n = \frac{n(n-1)}{4}$$

- Asymptotic behavior for large number of particles

$$\alpha_n \sim n \qquad \beta_n \simeq b_n \simeq \frac{1}{4}\ln n \qquad \gamma_n \to \frac{1}{2}$$

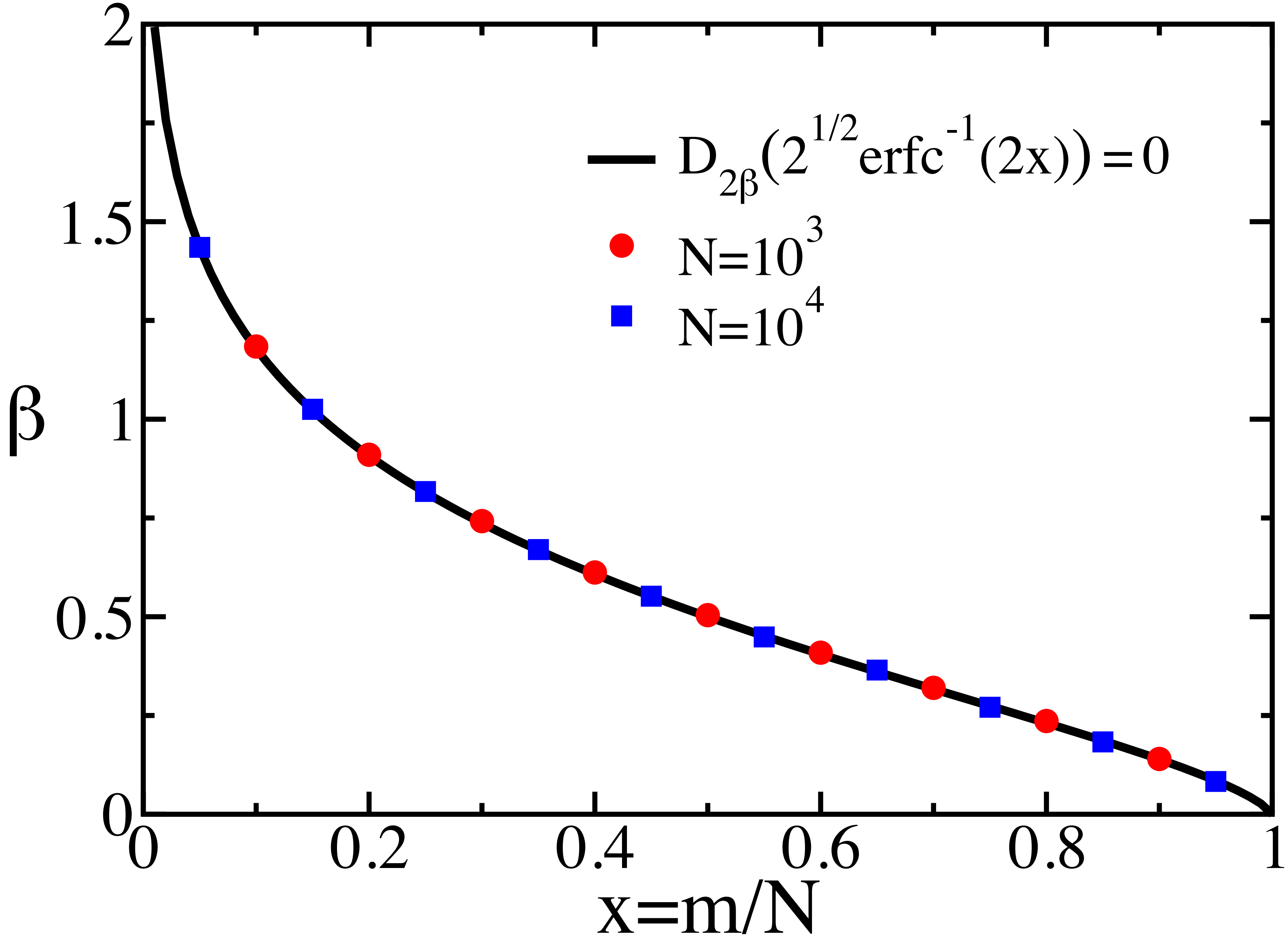
- And a conjecture!

$$\gamma_n = \frac{n-1}{2n}$$

$\gamma_1 = 0$

$\gamma_2 = 1/4$

Scaling laws for eigenvalues
in thermodynamic limit



Conclusions I

- First-passage kinetics of extremes in Brownian motion
- Problem reduces to diffusion in a two-dimensional corner with mixed boundary conditions
- First-passage exponent obtained analytically
- Exponent is continuously varying function of mobilities
- Relaxation is generally slower compared with positions
- Open questions: multiple particles, higher dimensions
- Scaling of eigenvalues in thermodynamics limit?
- "Race between maxima" as a data analysis tool

Part II
Incremental Records
(records of uncorrelated variables)

Marathon world record

Year	Athlete	Country	Record	Improvement
2002	Khalid Khannuchi	USA	2:05:38	
2003	Paul Tergat	Kenya	2:04:55	0:43
2007	Haile Gebrsellasie	Ethiopia	2:04:26	0:29
2008	Haile Gebrsellasie	Ethiopia	2:03:59	0:27
2011	Patrick Mackau	Kenya	2:03:38	0:21
2013	Wilson Kipsang	Kenya	2:03:23	0:15

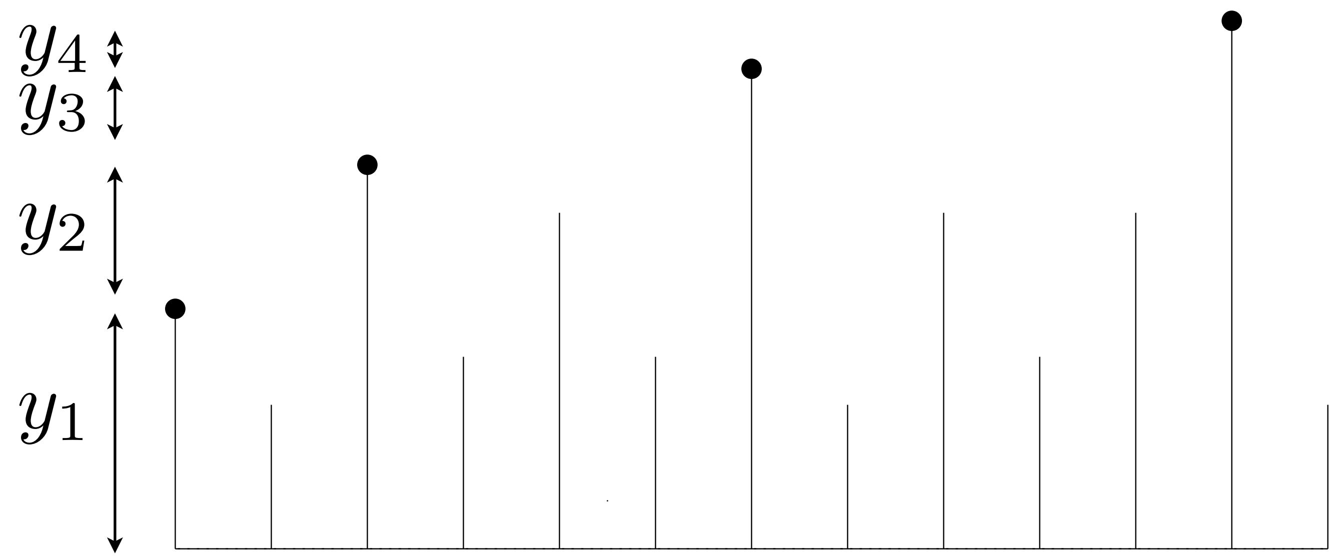
Incremental sequence of records

every record improves upon previous record by yet smaller amount

Are incremental sequences of records common?

source: wikipedia

Incremental Records



Incremental sequence of records

every record improves upon previous record by yet smaller amount

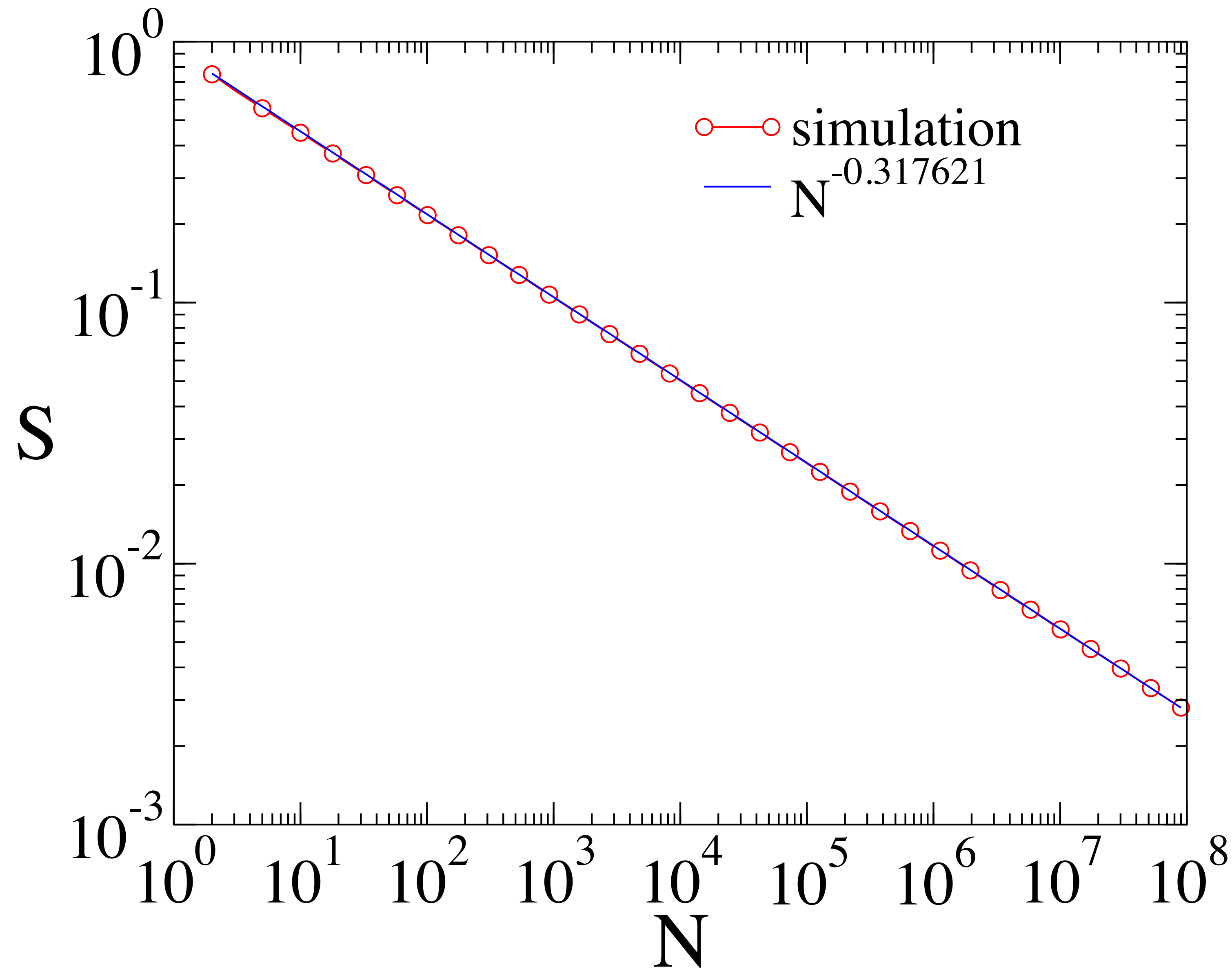
$$\text{random variable} = \{0.4, 0.4, 0.6, 0.7, 0.5, 0.1\}$$

$$\text{latest record} = \{0.4, 0.4, 0.6, 0.7, 0.7, 0.7\} \quad \uparrow$$

$$\text{latest increment} = \{0.4, 0.4, 0.2, 0.1, 0.1, 0.1\} \quad \downarrow$$

What is the probability all records are incremental?

Probability all records are incremental

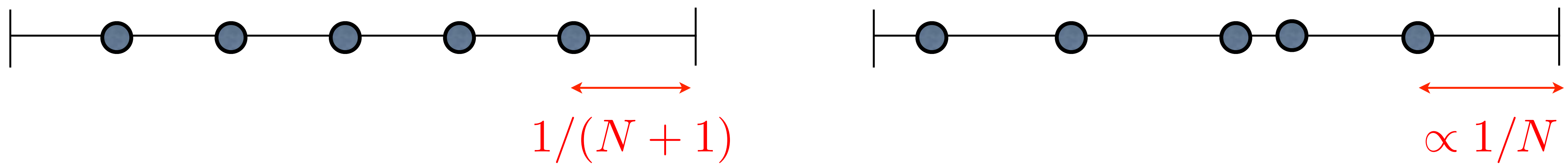


$S_N \sim N^{-\nu}$ $\quad\quad$ $\nu = 0.31762101$

Power law decay with nontrivial exponent

Question is free of parameters!

Uniform distribution



- The variable x is randomly distributed in $[0{:}1]$

$$\rho(x) = 1 \quad \text{for} \quad 0 \le x \le 1$$

- Probability record is smaller than x

$$R_N(x) = x^N$$

- Average record

$$A_N = \frac{N}{N+1} \qquad \Longrightarrow \qquad 1 - A_N \simeq N^{-1}$$

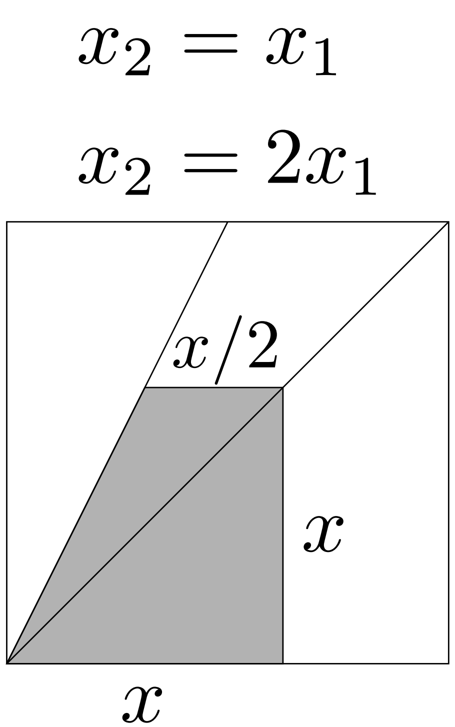
- Number of records

$$M_N = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}$$

Distribution of records

- Probability a sequence is inferior and record $< x$

$$G_N(x) \quad \Longrightarrow \quad S_N = G_N(1)$$



- One variable

$$G_1(x) = x \quad \Longrightarrow \quad S_1 = 1$$

- Two variables

$x_2 - x_1 > x_1$

$$G_2(x) = \frac{3}{4}\,x^2 \quad \Longrightarrow \quad S_2 = \frac{3}{4}$$

- In general, conditions are scale invariant $\quad x \to a\,x$
- Distribution of records for incremental sequences

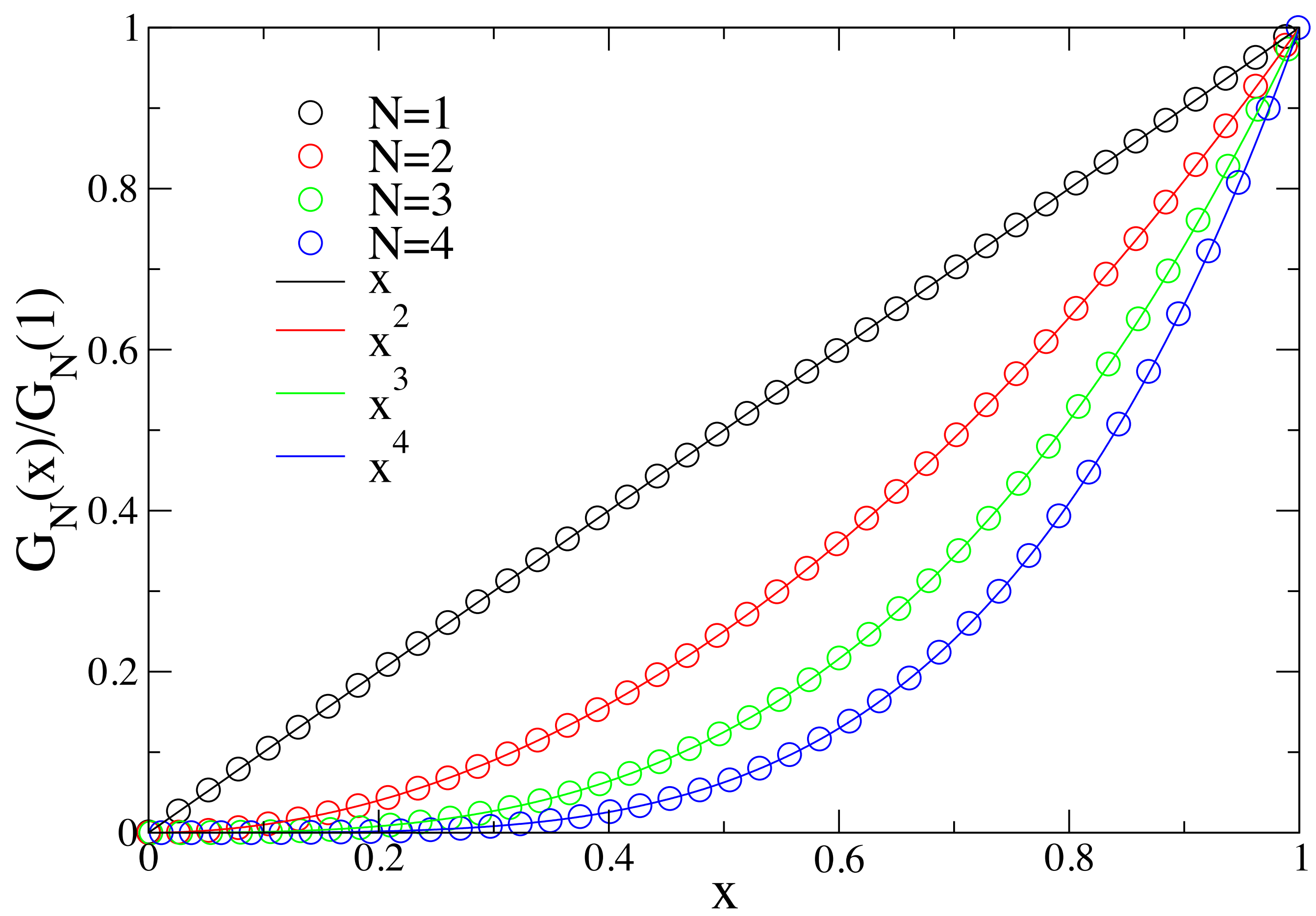
$$G_N(x) = S_N\,x^N$$

- Distribution of records for all sequences equals x^N

Statistics of records are standard

Fisher-Tippett 28
Gumbel 35

Scaling behavior



- Distribution of records for incremental sequences

$$G_N(x)/S_N = x^N = [1-(1-x)]^N \to e^{-s}$$

- Scaling variable

$$s = (1-x)N$$

Exponential scaling function

Distribution of increment+records

- Probability density $S_N(x,y)dxdy$ that:
  1. Sequence is incremental
  2. Current record is in range $(x,x+dx)$
  3. Latest increment is in range $(y,y+dy)$ with $0<y<x$
- Gives the probability a sequence is incremental

$$S_N = \int_0^1 dx \int_0^x dy\, S_N(x,y)$$

- Recursion equation incorporates memory

$$S_{N+1}(x,y) = \underbrace{x\, S_N(x,y)}_{\text{old record holds}} + \underbrace{\int_y^{x-y} dy'\, S_N(x-y,y')}_{\text{a new record is set}}$$

- Evolution equation includes integral, has memory

$$\frac{\partial S_N(x,y)}{\partial N} = -(1-x)S_N(x,y) + \int_y^{x-y} dy'\, S_N(x-y,y')$$

Scaling transformation

- Assume record and increment scale similarly

$$y \sim 1 - x \sim N^{-1}$$

- Introduce a scaling variable for the increment

$$s = (1-x)N \quad \text{and} \quad z = yN$$

- Seek a scaling solution

$$S_N(x,y) = N^2 S_N \, \Psi(s,z)$$

- Eliminate time out of the master equation

$$\left(2 - \nu + s + s\frac{\partial}{\partial s} + z\frac{\partial}{\partial z}\right)\Psi(s,z) = \int_z^\infty dz' \, \Psi(s+z, z')$$

Reduce problem from three variables to two

Factorizing solution

- Assume record and increment decouple

$$\Psi(s,z) = e^{-s}\, f(z)$$

- Substitute into equation for similarity solution

$$\left(2-\nu+s+s\frac{\partial}{\partial s}+z\frac{\partial}{\partial z}\right)\Psi(s,z) = \int_z^\infty dz'\,\Psi(s+z,z')$$

- First order integro-differential equation

$$zf'(z)+(2-\nu)f(z) = e^{-z}\int_z^\infty f(z')dz'$$

- Cumulative distribution of scaled increment $g(z)=\int_z^\infty f(z')dz'$
- Convert into a second order differential equation

$$zg''(z)+(2-\nu)g'(z)+e^{-z}g(z)=0 \qquad \begin{array}{l} g(0)=1 \\ g'(0)=-1/(2-\nu) \end{array}$$

Reduce problem from two variable to one

Distribution of increment

- Assume record and increment decouple

$$zg''(z) + (2-\nu)g'(z) + e^{-z}g(z) = 0$$

$$g(0) = 1$$
$$g'(0) = -1/(2-\nu)$$

- Two independent solutions

$$g(z) = z^{\nu-1} \quad \text{and} \quad g(z) = \text{const.} \quad \text{as} \quad z \to \infty$$

- The exponent is determined by the tail behavior

$$\nu = 0.31762101$$

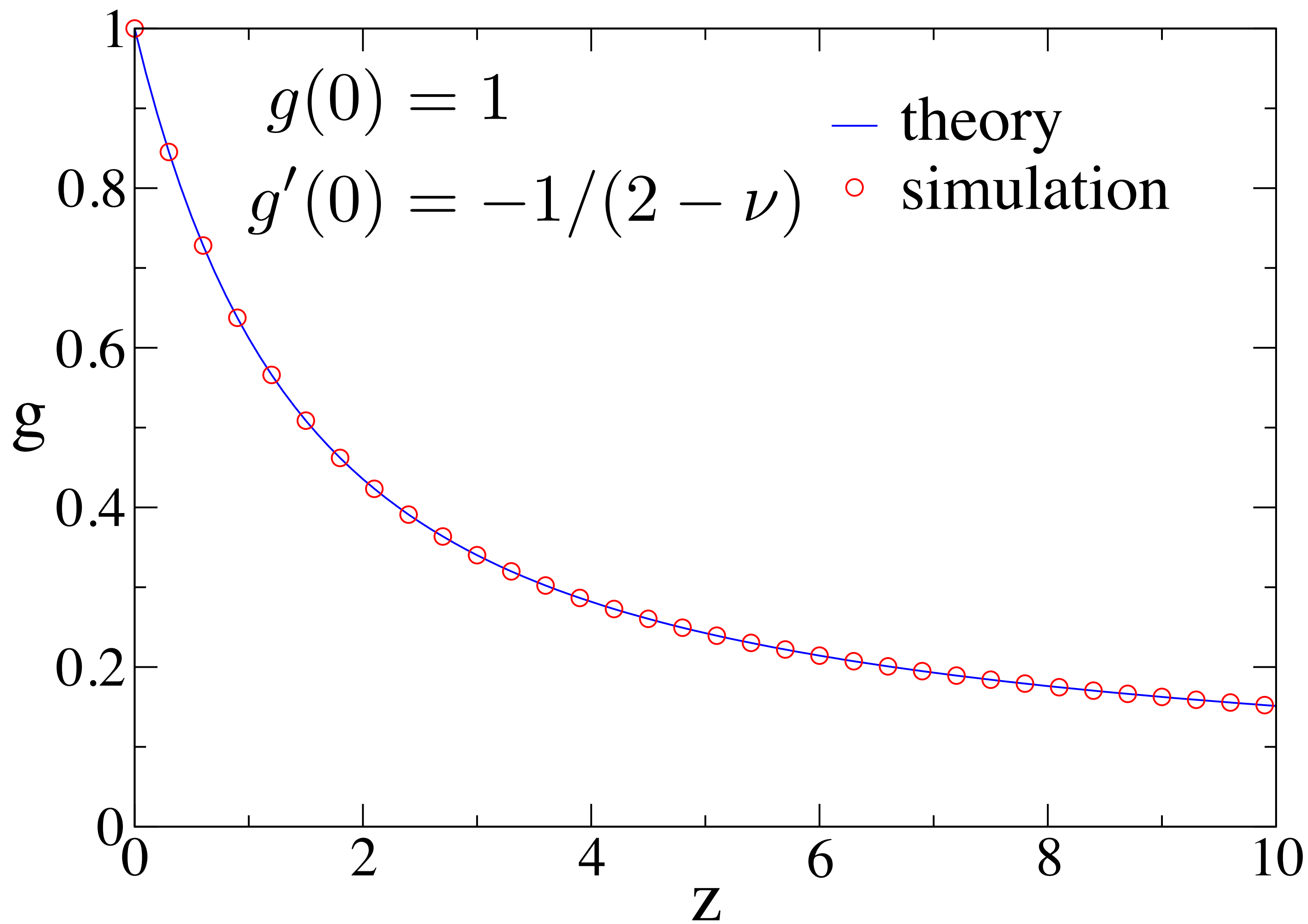
- The distribution of increment has a broad tail

$$P_N(y) \sim N^{-1} y^{\nu-2}$$

Increments can be relatively large
problem reduced to second order ODE

Numerical confirmation

Monte Carlo simulation versus integration of ODE



$$\frac{\langle sz \rangle}{\langle s \rangle \langle z \rangle} \to 1$$

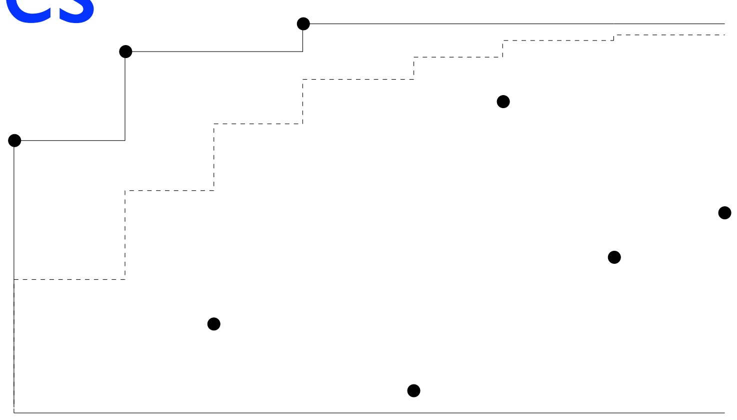
Increment and record become uncorrelated

Generalizations: Superior and Inferior Records

(records of uncorrelated variables)

Superior Records

- Start with sequence of random variables



$$\{x_1, x_2, x_3, \ldots, x_N\}$$

- Calculate the sequence of records

$$\{X_1, X_2, X_3, \ldots, X_N\} \quad \text{where} \quad X_n = \max(x_1, x_2, \ldots, x_n\}$$

- Compare with the expected average

$$\{A_1, A_2, A_3, \ldots, A_N\} = \{1/2, 2/3, 3/4, \ldots, N/(N+1)\}$$

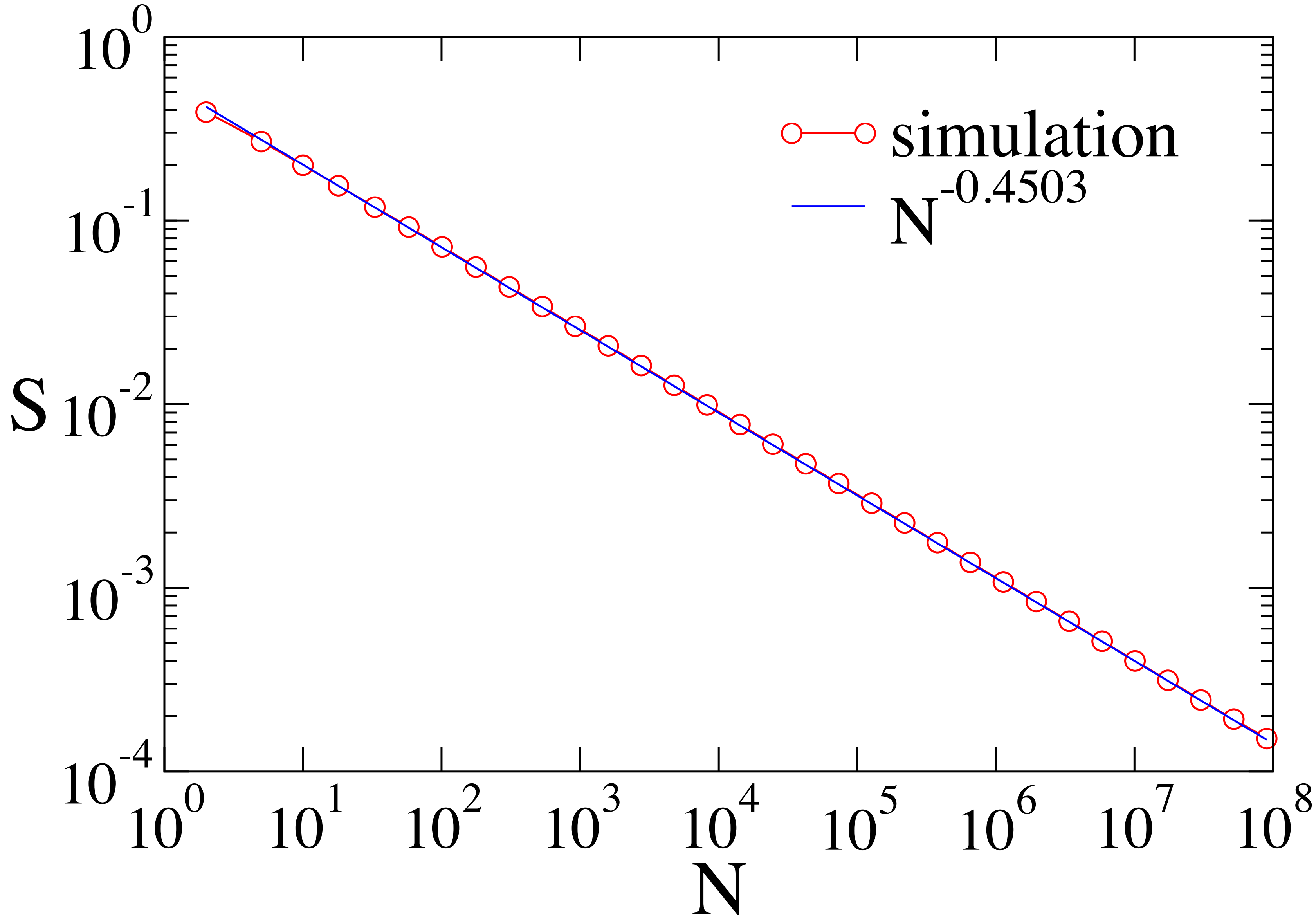
- Superior sequence = records always exceeds average

$$X_n > A_n \quad \text{for all} \quad 1 \le n \le N$$

- What fraction S_N of sequences is superior?

measure of "performance"

Numerical simulations



$$S_N \sim N^{-\beta} \qquad \beta = 0.4503 \pm 0.0002$$

Power law decay with <u>nontrivial</u> exponent

Distribution of superior records

- Cumulative probability distribution $F_N(x)$ that:
  1. Sequence is superior ($X_n > A_n$ for all n) and
  2. Current record is larger than x ($X_N > x$)
- Gives the desired probability immediately

$$S_N = F_N(A_N)$$

- Recursion equation

$$F_{N+1}(x) = x\, F_N(x) + (1-x)\, F_N(A_N) \qquad x > A_{N+1}$$

old record holds   a new record is set

- Recursive solution

$$F_1(x) = 1 - x$$
$$F_2(x) = \tfrac{1}{2}\left(1 + x - 2x^2\right)$$
$$F_3(x) = \tfrac{1}{18}\left(7 + 2x + 9x^2 - 18x^3\right)$$
$$F_4(x) = \tfrac{1}{576}\left(191 + 33x + 64x^2 + 288x^3 - 576x^4\right)$$

$$\xRightarrow{S_N = F_N(A_N)}$$

$$S_1 = \tfrac{1}{2}$$
$$S_2 = \tfrac{7}{18}$$
$$S_3 = \tfrac{191}{576}$$
$$S_4 = \tfrac{35393}{120000}$$

Scaling Analysis

- Convert recursion equation

$$F_{N+1}(x) = x\,F_N(x) + (1-x)\,F_N(A_N)$$

into a differential equation (N plays role of time!)

$$\frac{\partial F_N(x)}{\partial N} = (1-x)\,\left[F_N(A_N) - F_N(x)\right]$$

- Seek a similarity solution ($N \to \infty$ limit)

$$F_N(x) \simeq S_N \Phi(s) \quad \text{with} \quad s = (1-x)N$$

boundary conditions $\Phi(0) = 0 \quad \text{and} \quad \Phi(1) = 1$ $\left(1 - \frac{N}{N+1}\right) N \to 1$

- Similarity function obeys first-order ODE

$$\Phi'(s) + (1 - \beta\, s^{-1})\Phi(s) = 1$$

Similarity solution gives distribution of scaled record

Similarity Solution

- Equation with yet unknown exponent

$$\Phi'(s) + (1 - \beta\, s^{-1})\Phi(s) = 1$$

- General solution

$$\Phi(s) = s \int_0^1 dz\, z^{-\beta} e^{s(z-1)}$$

- Boundary condition dictates the exponent

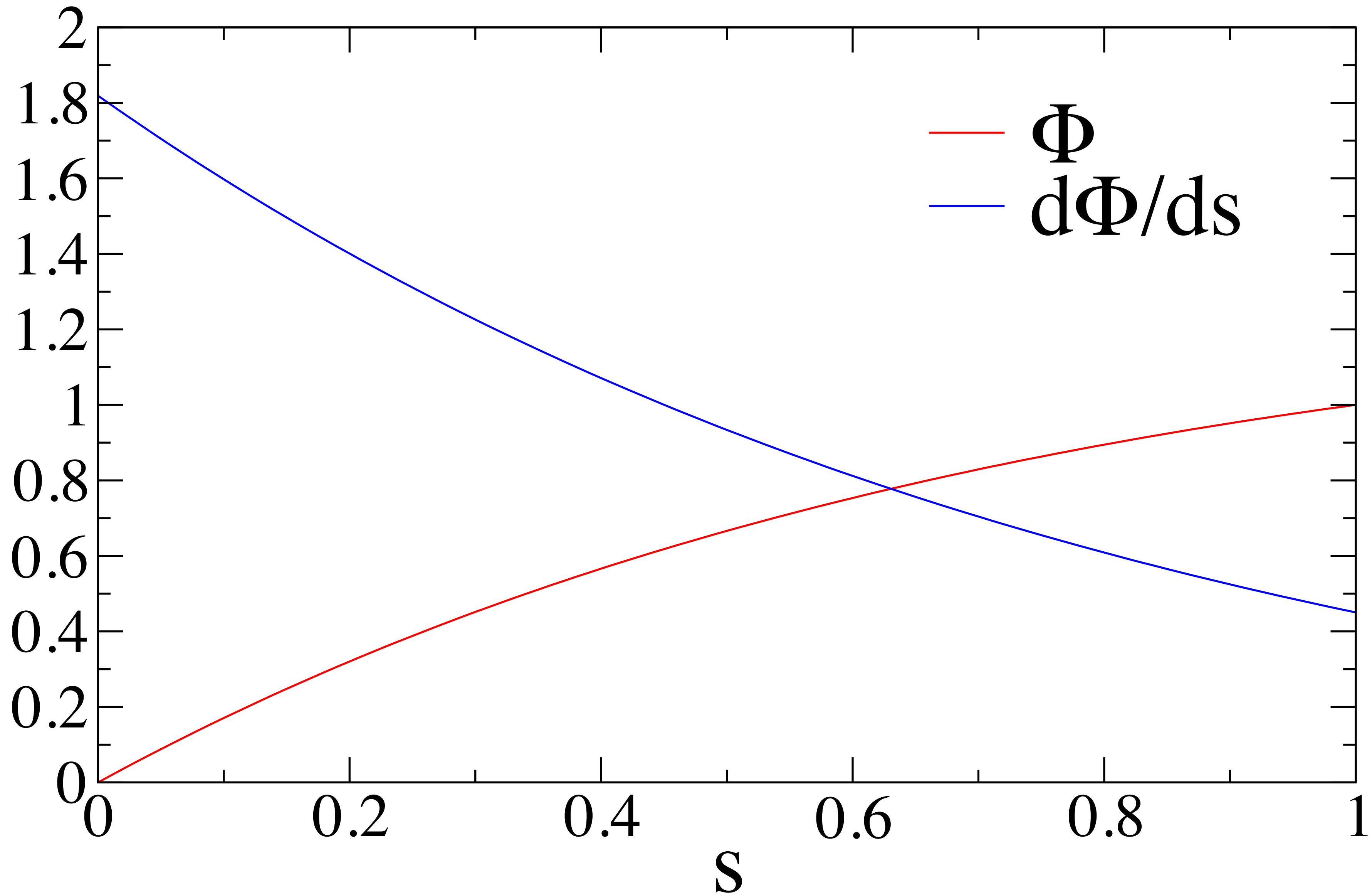
$$\int_0^1 dz\, z^{-\beta} e^{(z-1)} = 1$$

- Root is a transcendental number

$$\beta = 0.450265027495$$

Analytic solution for distribution and exponent

Distribution of records

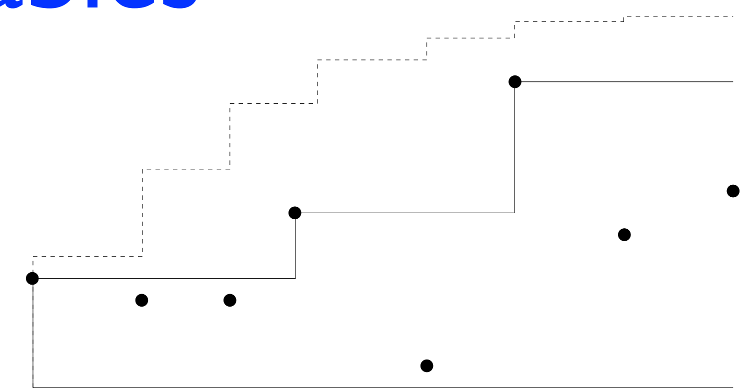


scaling variable $s = (1 - x)N$

Inferior records

- Start with sequence of random variables

$$\{x_1, x_2, x_3, \ldots, x_N\}$$



- Calculate the sequence of records

$$\{X_1, X_2, X_3, \ldots, X_N\} \quad \text{where} \quad X_n = \max(x_1, x_2, \ldots, x_n\}$$

- Compare with the expected average

$$\{A_1, A_2, A_3, \ldots, A_N\} = \{1/2, 2/3, 3/4, \ldots, N/(N+1)\}$$

- Inferior sequence = records always below average

$$X_n > A_n \quad \text{for all} \quad 1 \le n \le N$$

- What fraction of sequences are inferior?

$$I_N \sim N^{-\alpha}$$

expect power law decay, different exponent

Probability sequence is inferior

- Start with sequence of random variables

$$\{A_1, A_2, A_3, \ldots, A_N\} = \{1/2, 2/3, 3/4, \ldots, N/(N+1)\}$$

- One variable

$$x_1 < \frac{1}{2} \quad \Longrightarrow \quad I_1 = \frac{1}{2}$$

- Two variables

$$x_1 < \frac{1}{2} \quad \text{and} \quad x_2 < \frac{2}{3} \quad \Longrightarrow \quad I_2 = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$$

- Recursion equation (no interactions between variables)

$$I_{N+1} = I_N \frac{N}{N+1}$$

- Simple solution

$$I_N = \frac{1}{N+1} \qquad I_N \sim N^{-1}$$

power law decay with trivial exponent

General distributions

- Arbitrary distribution function
- Single parameter contains information about tail

$$\alpha = \lim_{N\to\infty} N \int_{A_N}^{\infty} dx\, \rho(x)$$

- Equals the exponent for inferior sequences

$$I_N \sim N^{-\alpha}$$

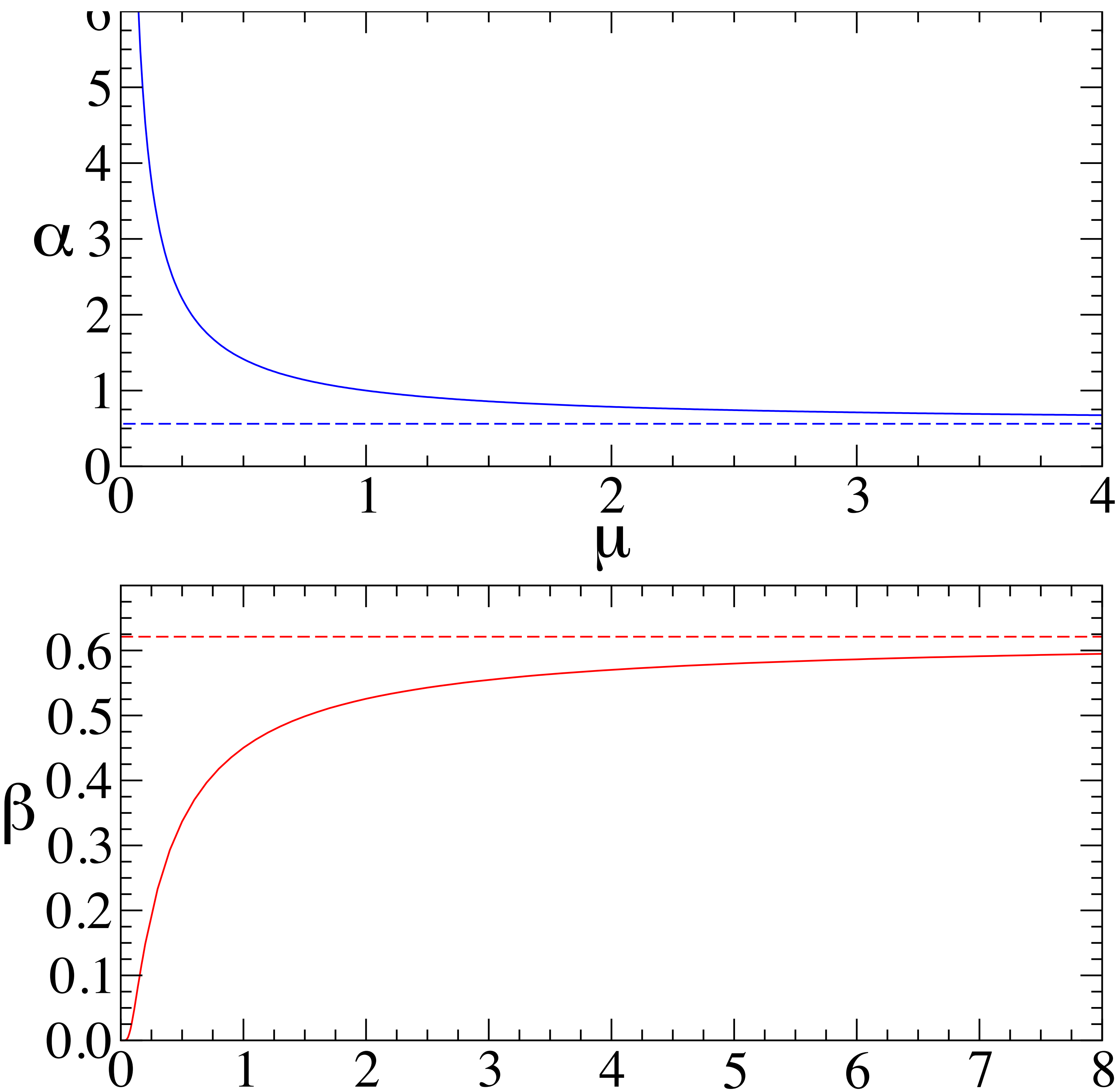
- Exponent for superior sequences

$$\alpha \int_0^1 dz\, z^{-\beta} e^{\alpha(z-1)} = 1$$

- Power-law distributions (compact support)

$$R(x) \sim (1-x)^{\mu} \quad \Longrightarrow \quad \alpha = \left[\Gamma\left(1+\tfrac{1}{\mu}\right)\right]^{\mu}$$

Continuously varying exponents



$\alpha_{\min} \leq \alpha < \infty$

$\alpha_{\min} = e^{-\gamma} = 0.561459$

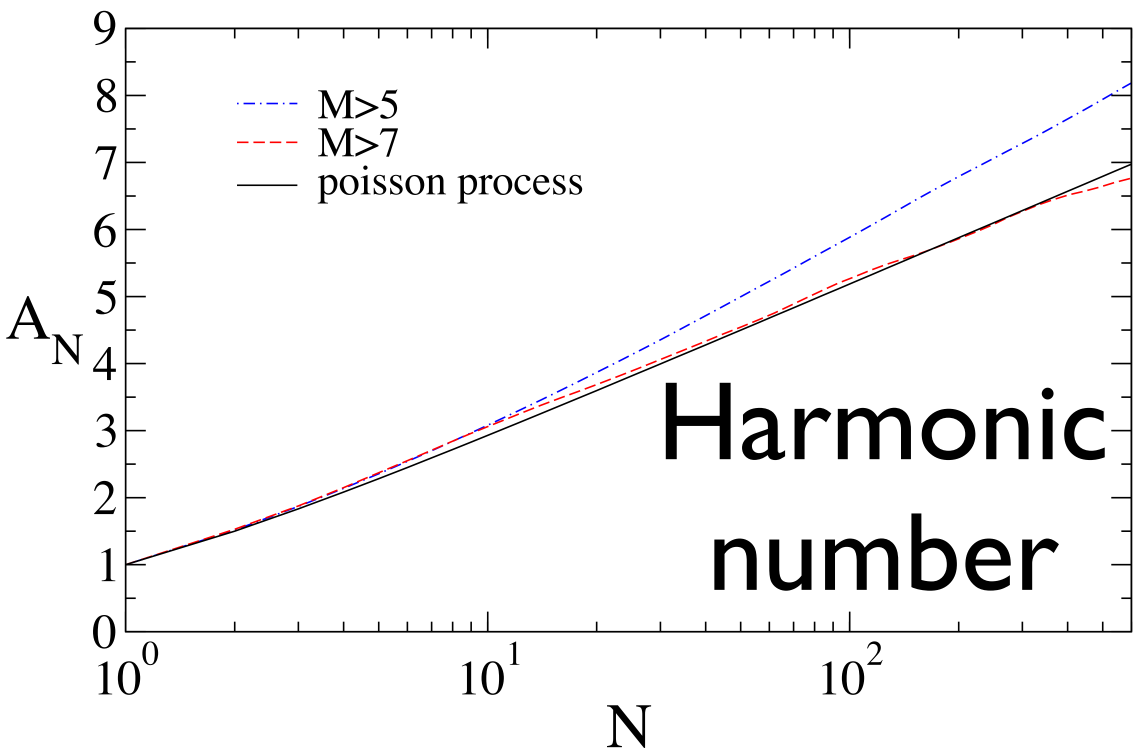
$\beta_{\max} = 0.621127$

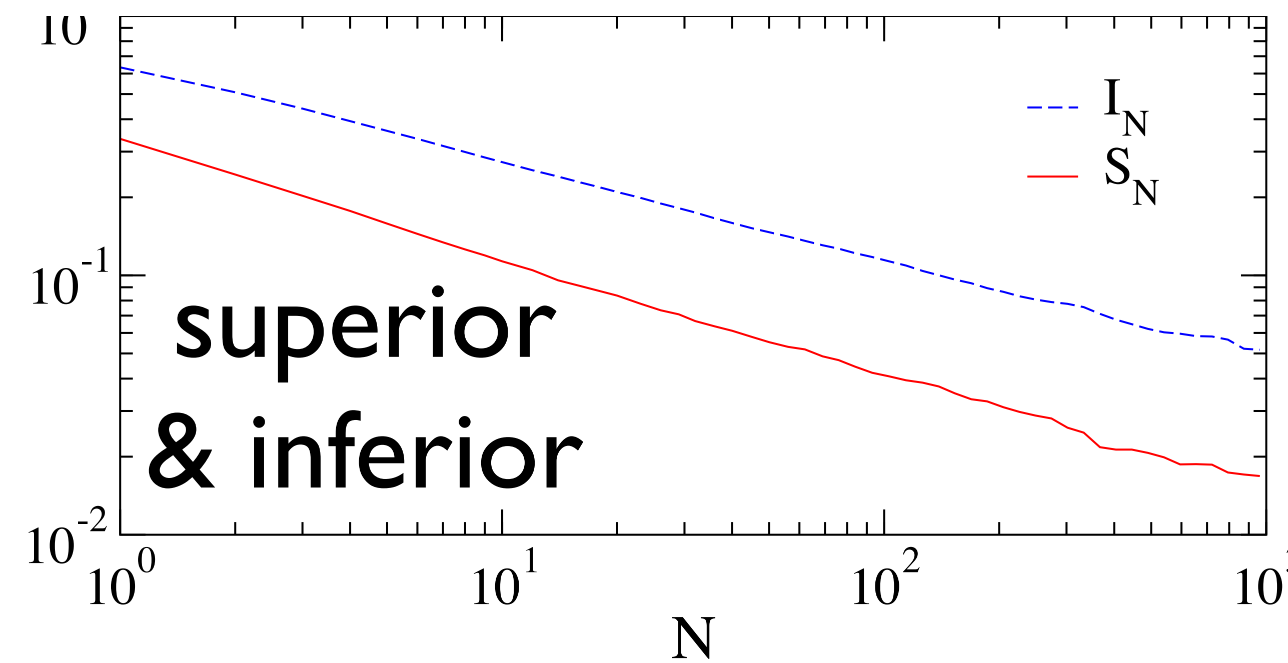
$0 < \beta \leq \beta_{\max}$

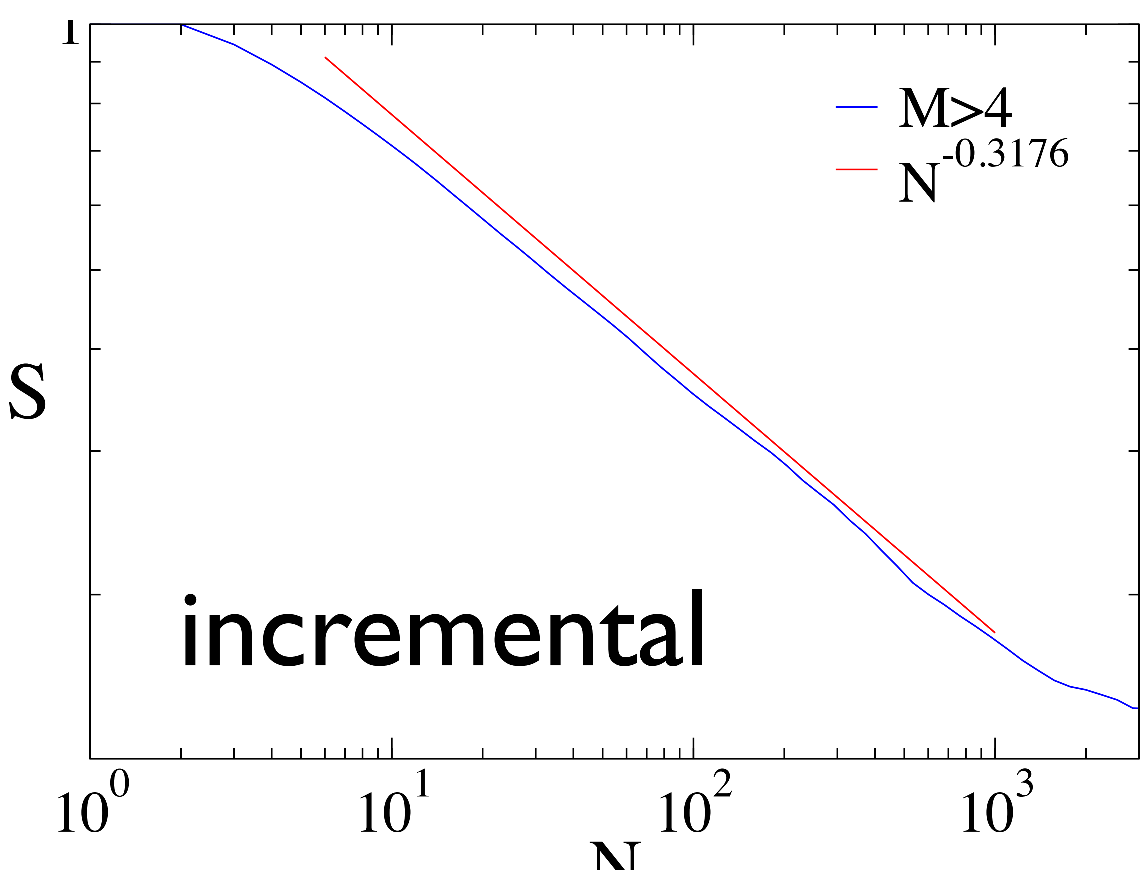
Tail of distribution function controls record statistics

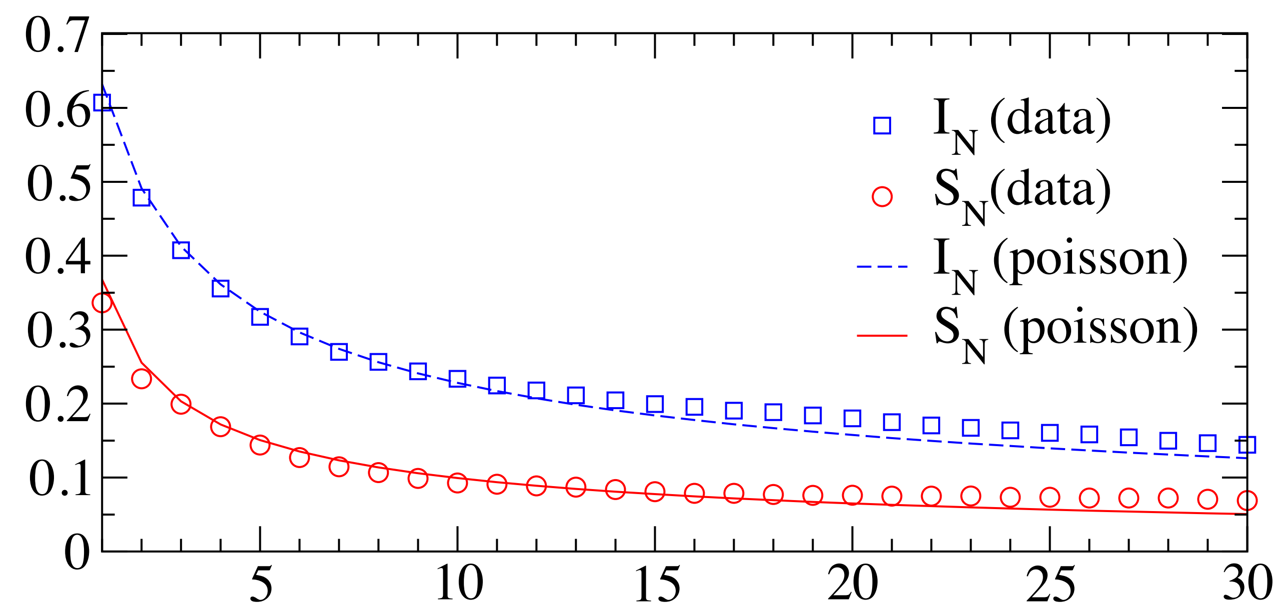
Records in earthquake data

inter-event times









good agreement with theoretical predictions

Conclusions II

- Studied persistent configuration of record sequences
- Linear evolution equations (but nonlocal/memory)
- Dynamic formulation: treat sequence length as time
- Similarity solutions for distribution of records
- Probability of persistent configuration (inferior, superior, inferior) decays as a power-law
- Power laws exponents are generally nontrivial
- Exponents can be obtained analytically
- Tail of distribution function controls record statistics

Take home message

- First-passage probabilities of records have power-law tails
- First-passage exponents are generally nontrivial

Many open problems!

Publications

1. E. Ben-Naim and P.L. Krapivsky, Phys. Rev. Lett. **113,** 030604 (2014).
2. E. Ben-Naim and P.L. Krapivsky, J. Phys. A **47,** 255002 (2014).
3. P.W. Miller and E. Ben-Naim, J. Stat. Mech. 10025 (2013).
4. E. Ben-Naim, P.L. Krapivsky, Phys. Rev. E **113,** 022145 (2013).