First-Passage Statistics of Extreme Values

Eli Ben-Naim
Los Alamos National Laboratory

with:
Paul Krapivsky (Boston University)
Nathan Lemons (Los Alamos)
Pearson Miller (Yale, MIT)

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

Random Graph Processes, Austin TX, May 10, 2016

Plan

I. **Motivation**: records & their first-passage statistics as a data analysis tool

II. **Incremental records:** uncorrelated random variables

III. **Ordered records**: uncorrelated random variables

IV. **Ordered records**: correlated random variables

I. Motivation:

records & their first-passage statistics as a data analysis tool

Record & running record

- Record = largest variable in a series

$$X_N = \max(x_1, x_2, \dots, x_N)$$

- Running record = largest variable to date

$$X_1 \leq X_2 \leq \cdots \leq X_N$$

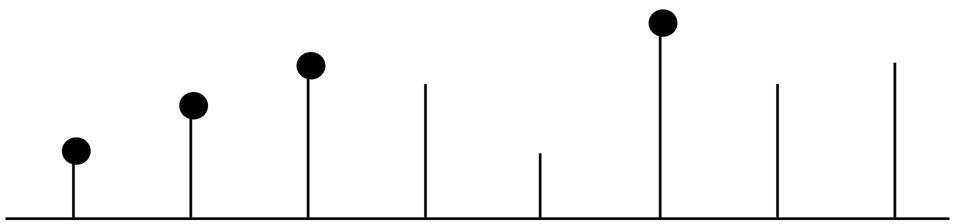
- Independent and identically distributed variables

$$\int_0^\infty dx\, \rho(x) = 1$$

Statistics of extreme values

Feller 68
Gumble 04
Ellis 05

Average number of running records



- Probability that Nth variable sets a record

$$P_N = \frac{1}{N}$$

- Average number of records = harmonic number

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

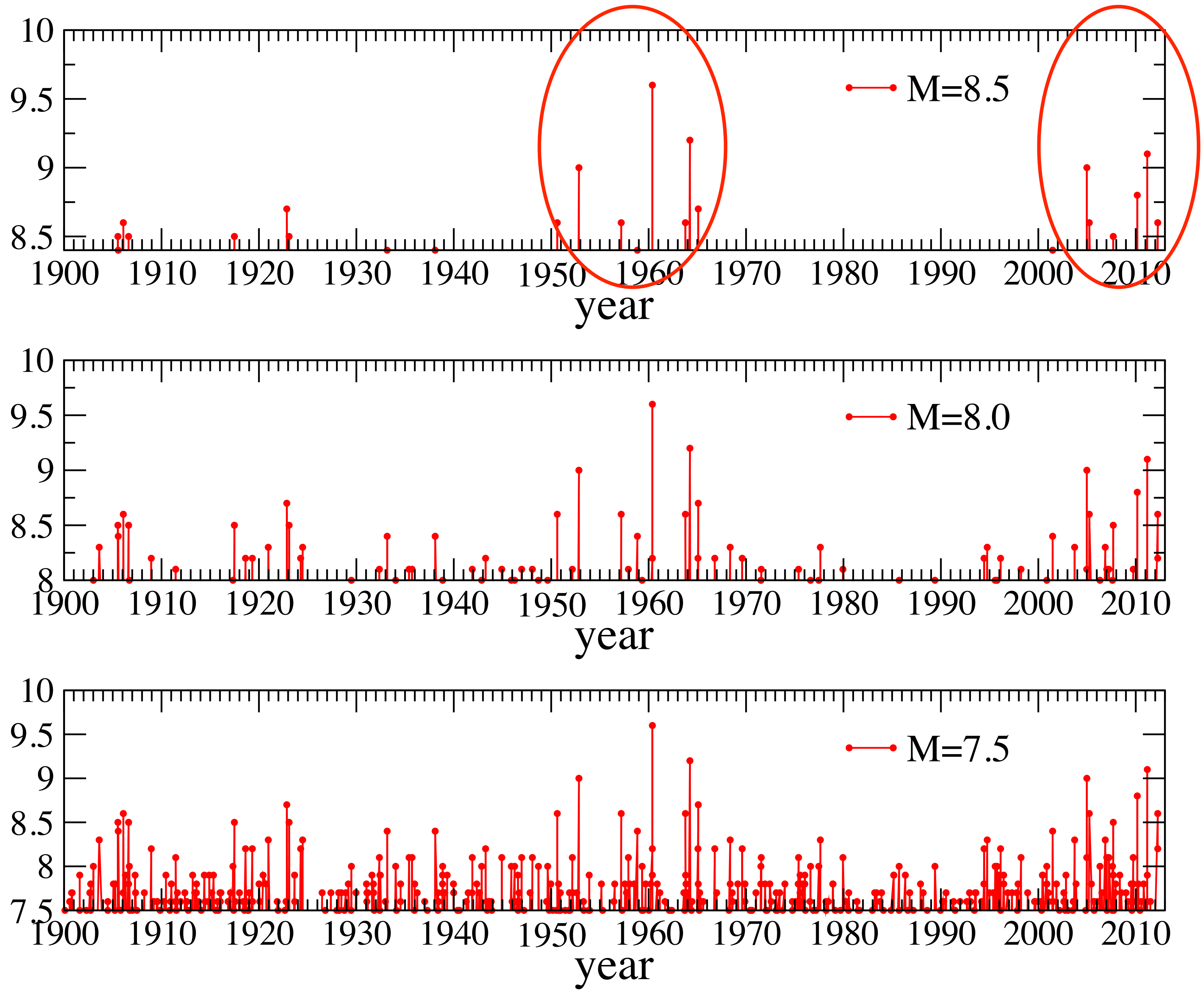
- Grows logarithmically with number of variables

$$M_N \simeq \ln N + \gamma \qquad \gamma = 0.577215$$

Behavior is <u>independent</u> of distribution function

Number of records is quite small

Time series of massive earthquakes

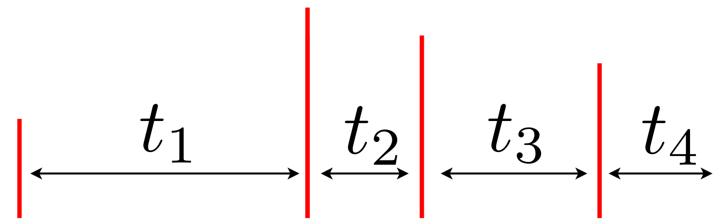


1770 M>7 events
1900-2013

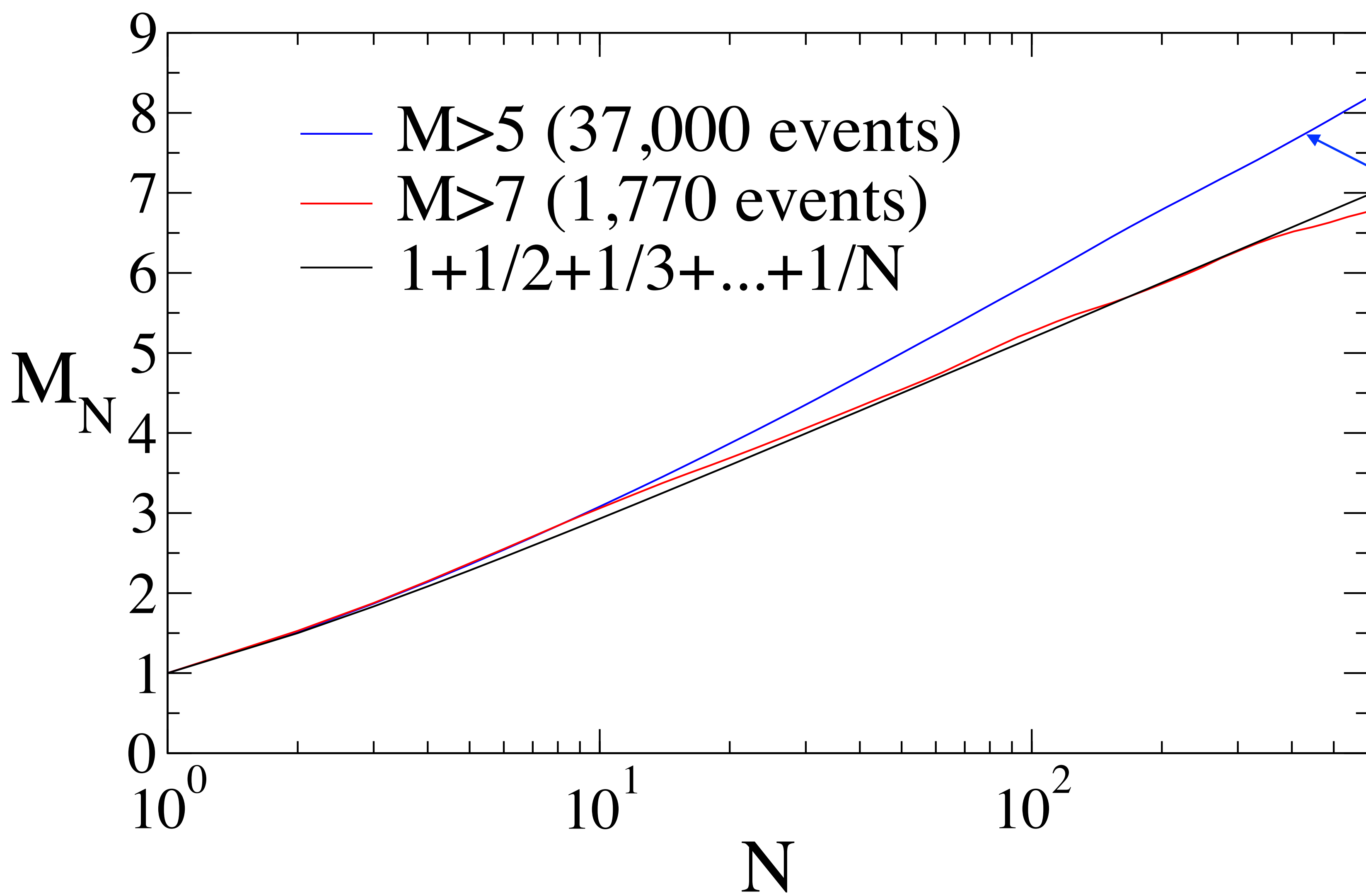
Magnitude	Annual #
9-9.9	1/20
8-8.9	1
7-7.9	15
6-6.9	134
5-5.9	1300
4-4.9	~13,000
3-3.9	~130,000
2-2.9	~1,300,000

Are massive earthquakes correlated?

Records in inter-event time statistics



Count number of running records in N consecutive events



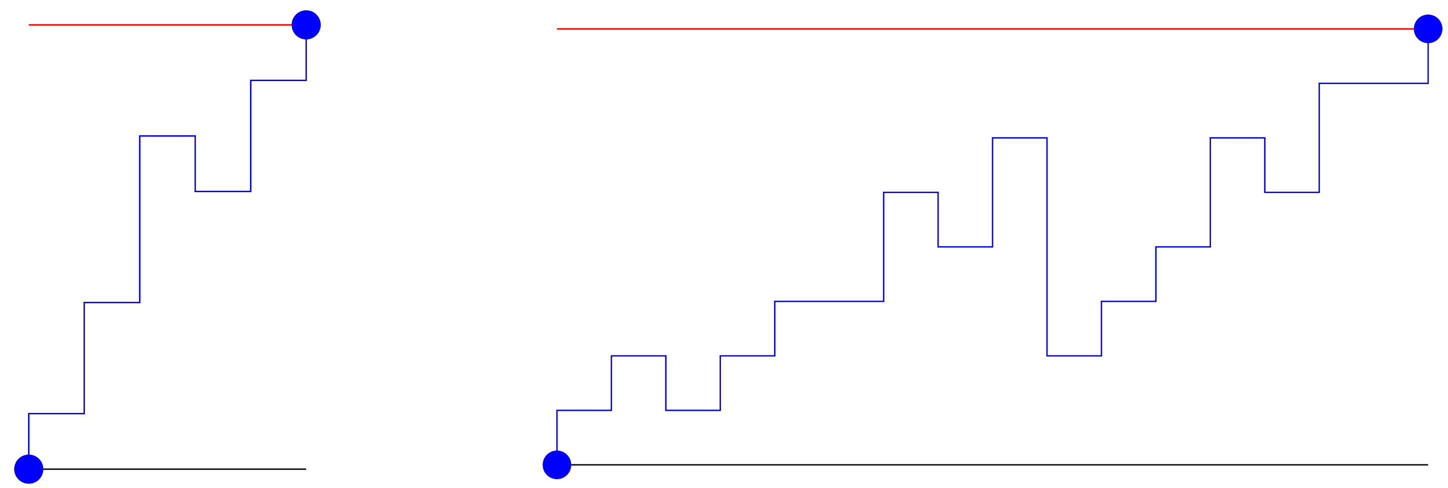
attribute deviation to aftershocks

records indicate inter-event times uncorrelated

massive earthquakes are random

EB & Krapivsky PRE 13

First-passage processes



- Process by which a fluctuating quantity reaches a threshold for the <u>first</u> time
- **First-passage probability:** for the random variable to reach the threshold as a function of time.
- **Total probability**: that threshold is <u>ever</u> reached. May or may not equal 1
- **First-passage time**: the mean duration of the first-passage process. Can be <u>finite</u> or <u>infinite</u>

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

II. Incremental records: uncorrelated random 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

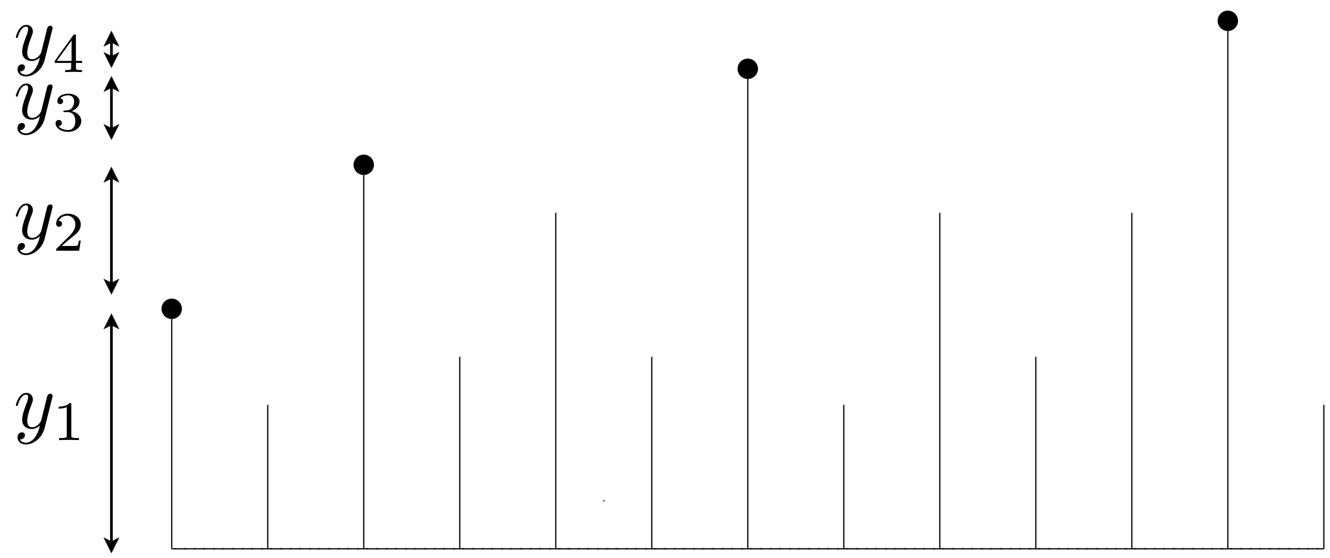
Incremental sequence of records

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

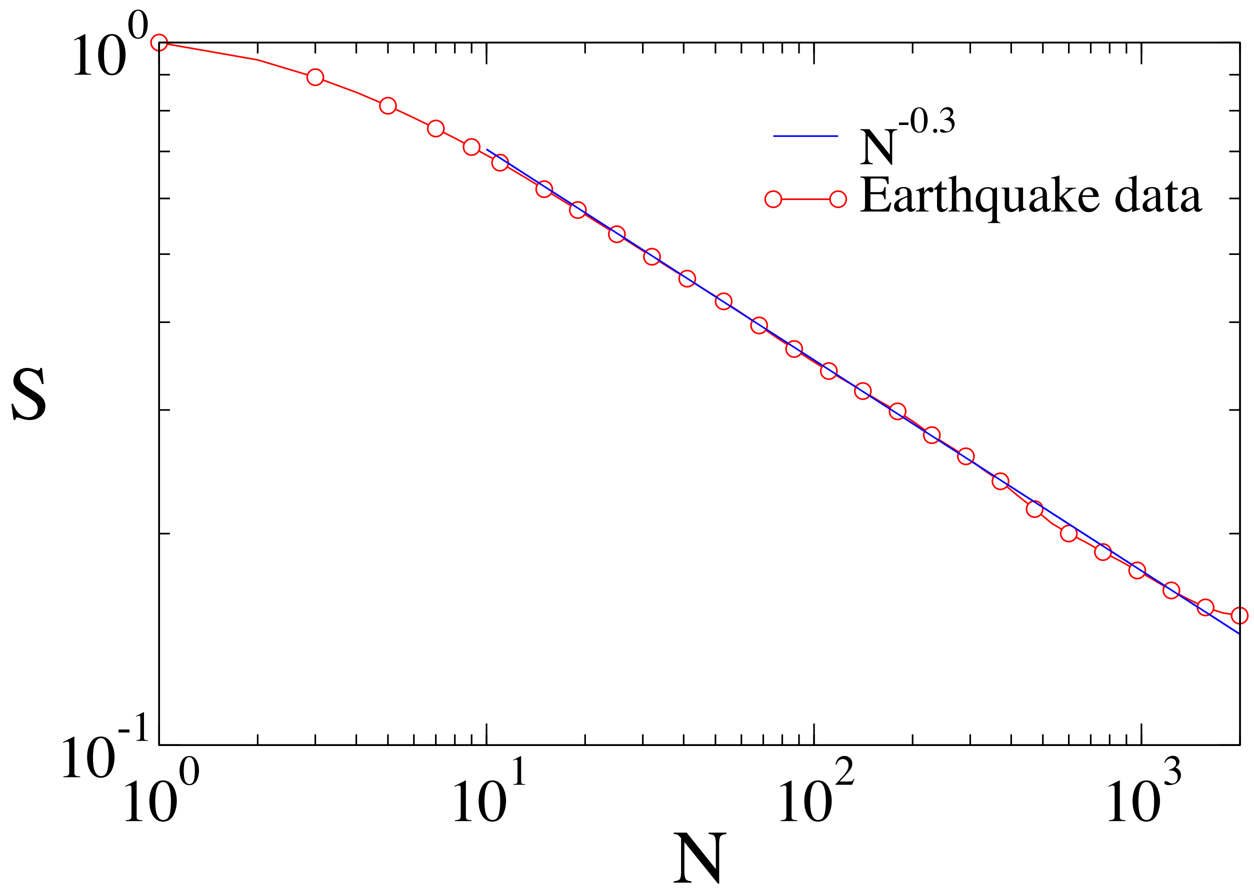
$$\text{random variable} = \{0.4, 0.4, 0.6, 0.7, 0.5, 0.1\}$$

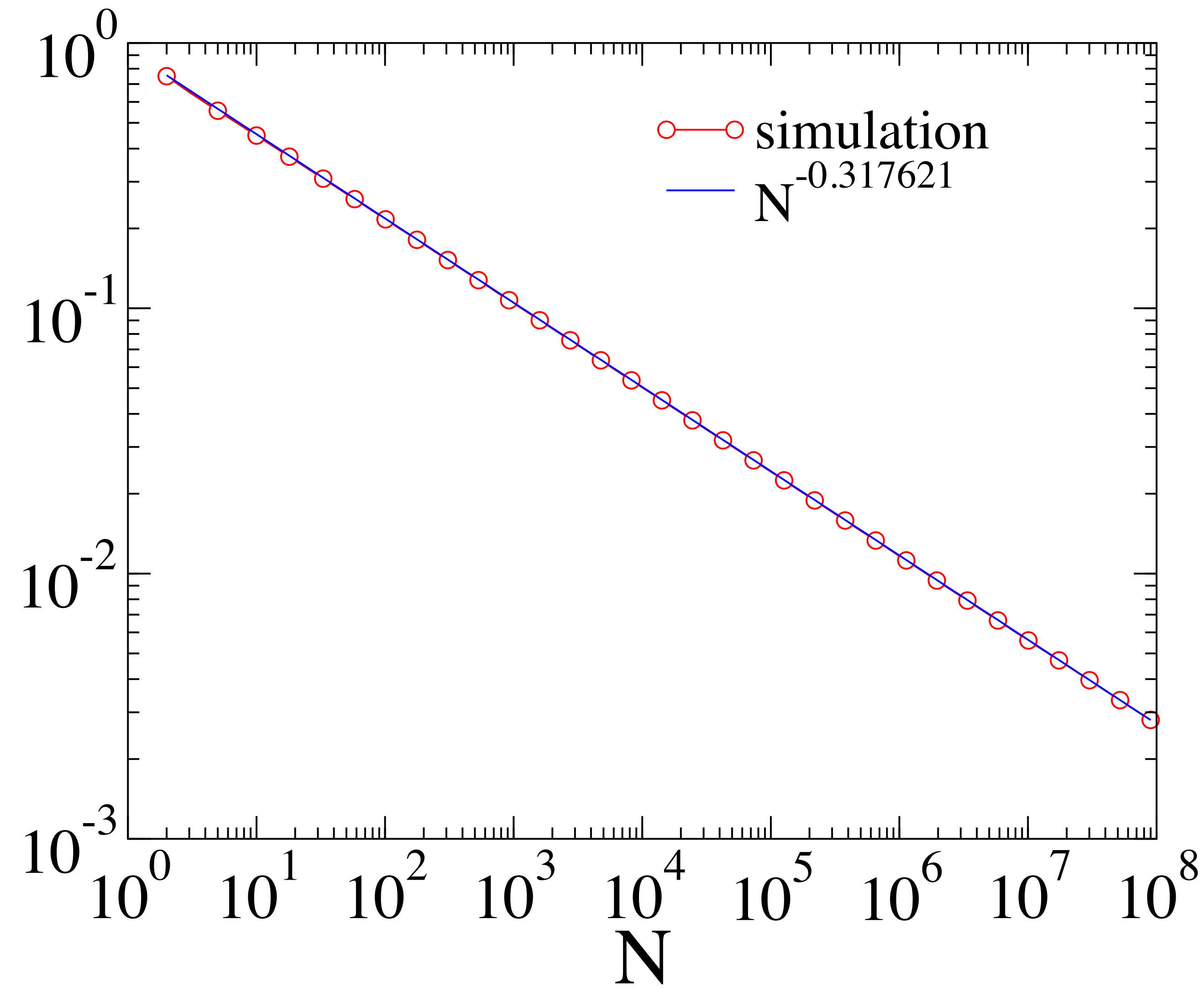
$$\text{latest record} = \{0.4, 0.4, 0.6, 0.7, 0.7, 0.7\} \quad \uparrow$$

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

What is the probability all records are incremental?

Probability all records are incremental





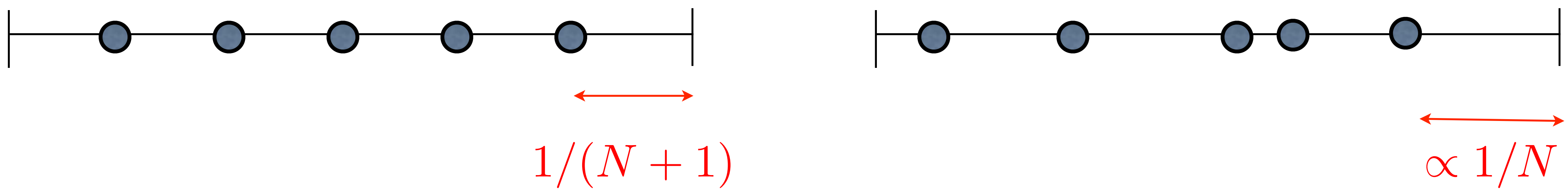
$$S_N \sim N^{-\nu} \qquad \nu = 0.31762101$$

Power law decay with nontrivial exponent

Problem is parameter-free

Miller & EB JSTAT 13

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

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

Distribution of records is purely exponential

Distribution of records

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

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

- One variable

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

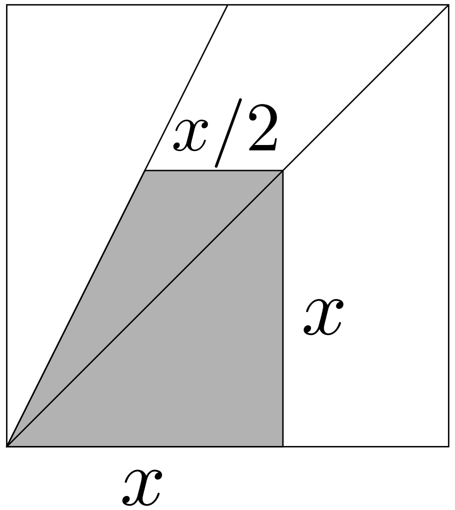
- Two variables

$x_2 - x_1 > x_1$

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

$x_2 = x_1$

$x_2 = 2x_1$



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

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

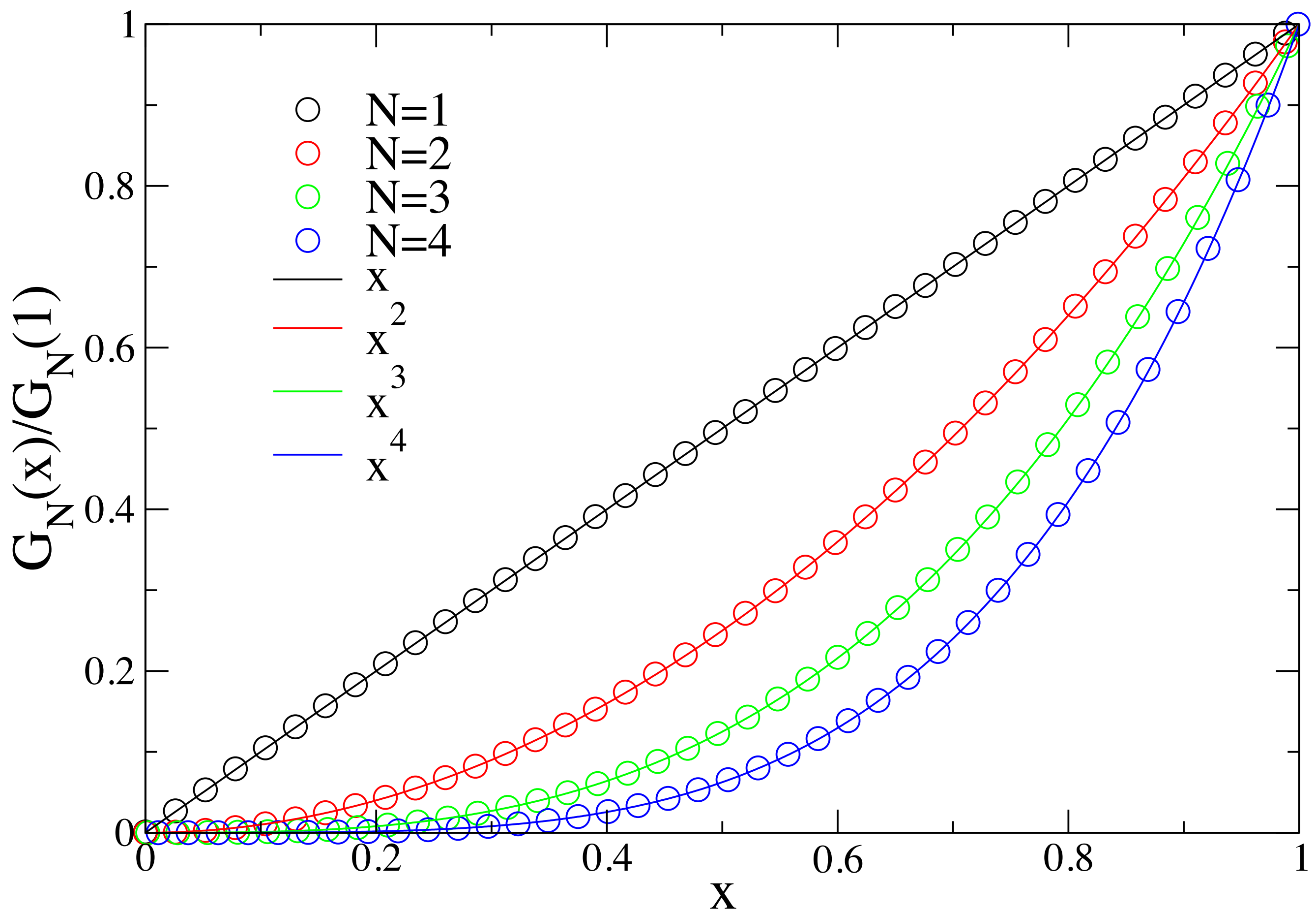
$S_1 = 1$
$S_2 = 3/4$
$S_3 = 47/72$

- 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 <u>three</u> variables to <u>two</u>

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

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

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

Reduce problem from <u>two</u> variable to <u>one</u>

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

$$\beta = 0.317621014462...$$

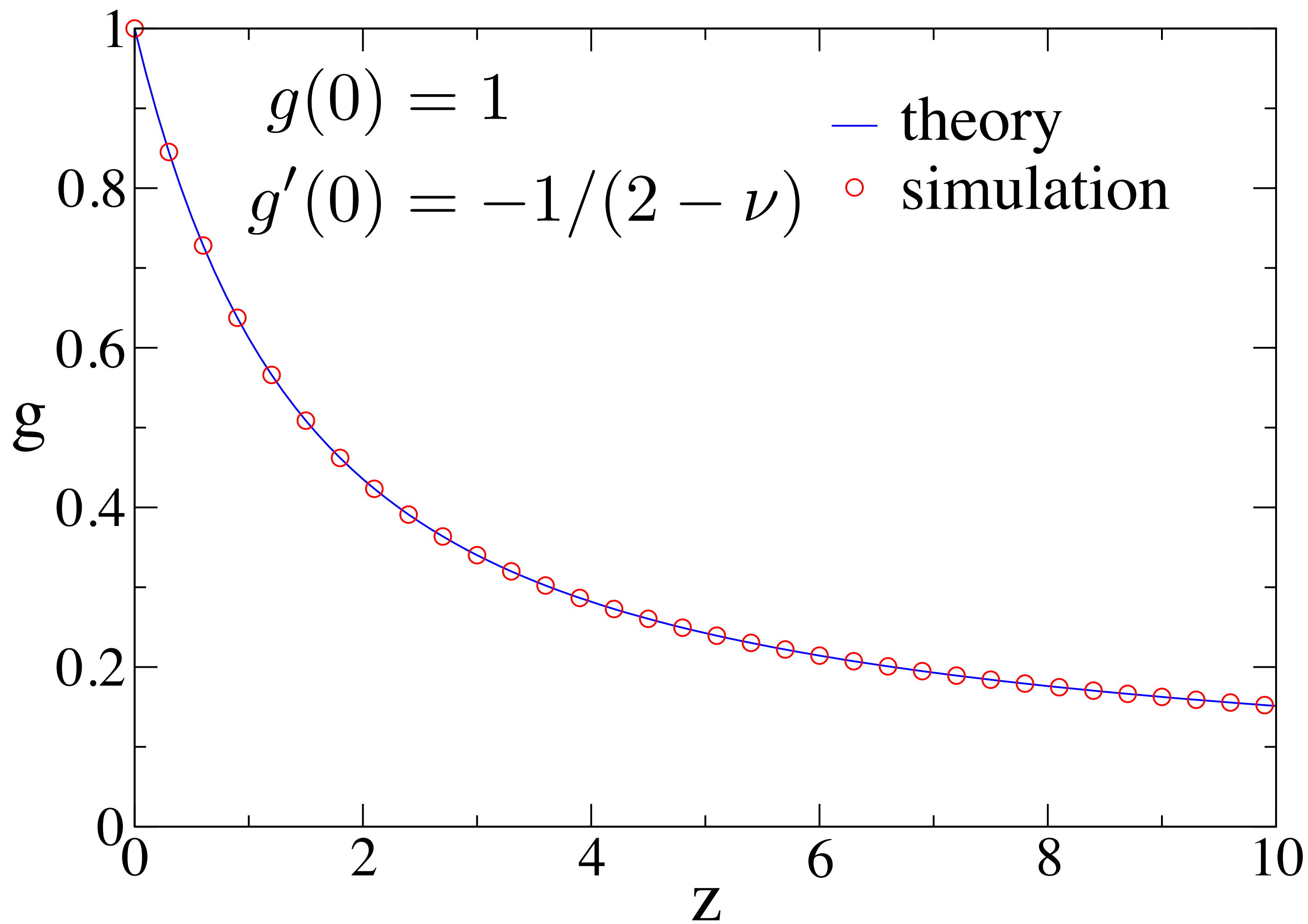
- The distribution of increment has a broad tail

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

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

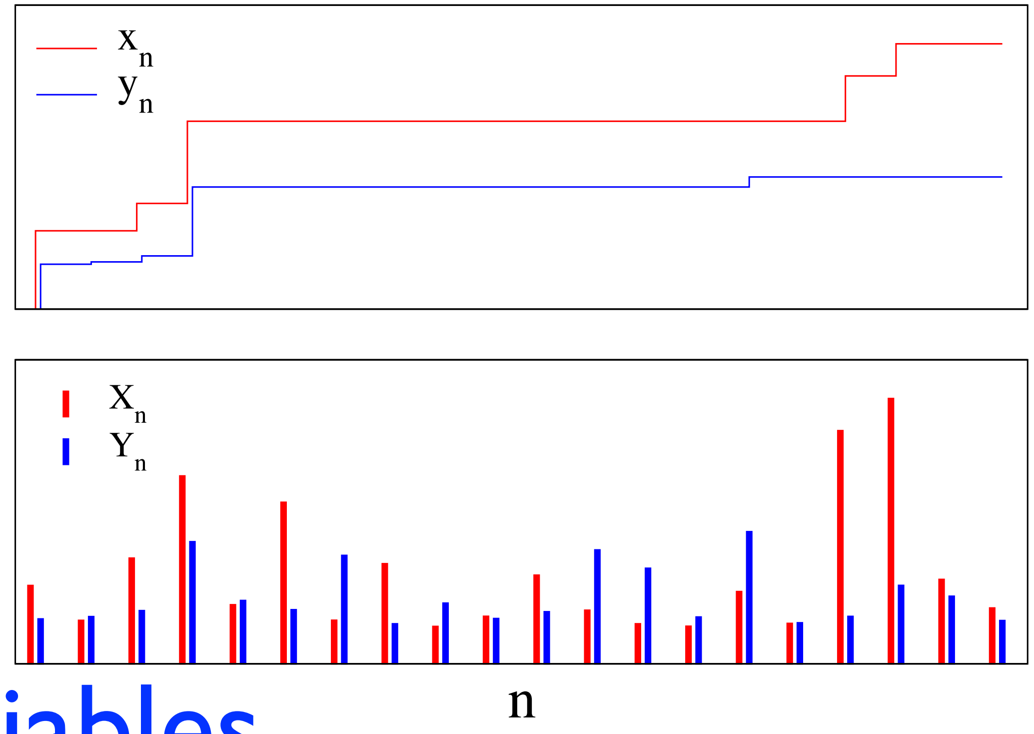
Recap II

- Probability a sequence of records is incremental
- Linear evolution equations (but with memory)
- Dynamic formulation: treat sequence length as time
- Similarity solutions for distribution of records
- Probability a sequence of records is incremental decays as power-law with sequence length
- Power-law exponent is nontrivial, obtained analytically
- Distribution of record increments is broad

First-passage properties of extreme values are interesting

III. Ordered records: uncorrelated random variables

Ordered records



- Motivation: temperature records: Record high increasing each year

- Two sequences of random variables

$$\{X_1, X_2, \ldots, X_N\} \quad \text{and} \quad \{Y_1, Y_2, \ldots, Y_N\}$$

- Independent and identically distributed variables

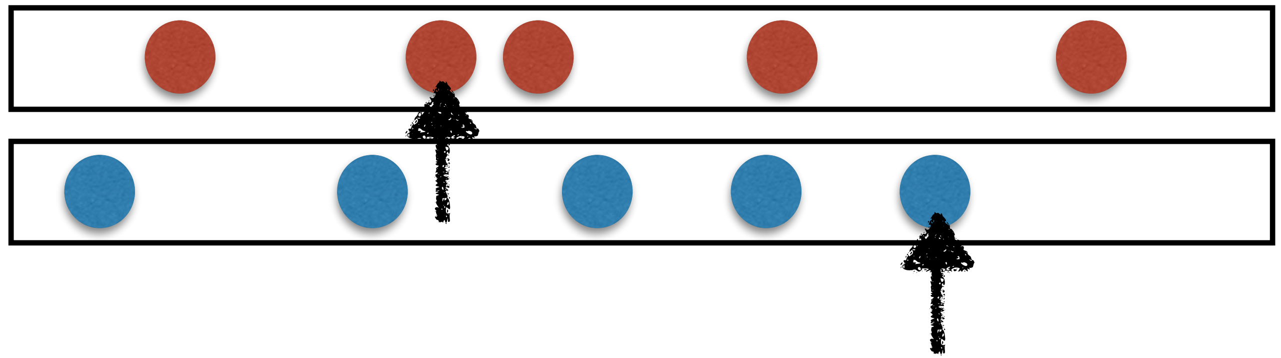
- Two corresponding sequences of records

$$x_n = \max\{X_1, X_2, \ldots, X_n\} \quad \text{and} \quad y_n = \max\{Y_1, Y_2, \ldots, Y_n\}$$

- Probability S_N records maintain perfect order

$$x_1 > y_1 \quad \text{and} \quad x_2 > y_2 \quad \cdots \quad \text{and} \quad x_N > y_N$$

Two sequences



- Survival probability obeys closed recursion equation

$$S_N = S_{N-1}\left(1 - \frac{1}{2N}\right)$$

- Solution is immediate

$$S_N = \binom{2N}{N} 2^{-2N}$$

identical to
Sparre Andersen 53!

- Large-N: Power-law decay with rational exponent

$$S_N \simeq \pi^{-1/2} N^{-1/2}$$

Universal behavior: independent of parent distribution!

Ordered random variables

- Probability P_N variables are always ordered

$$X_1 > Y_1 \quad \text{and} \quad X_2 > Y_2 \quad \cdots \quad \text{and} \quad X_N > Y_N$$

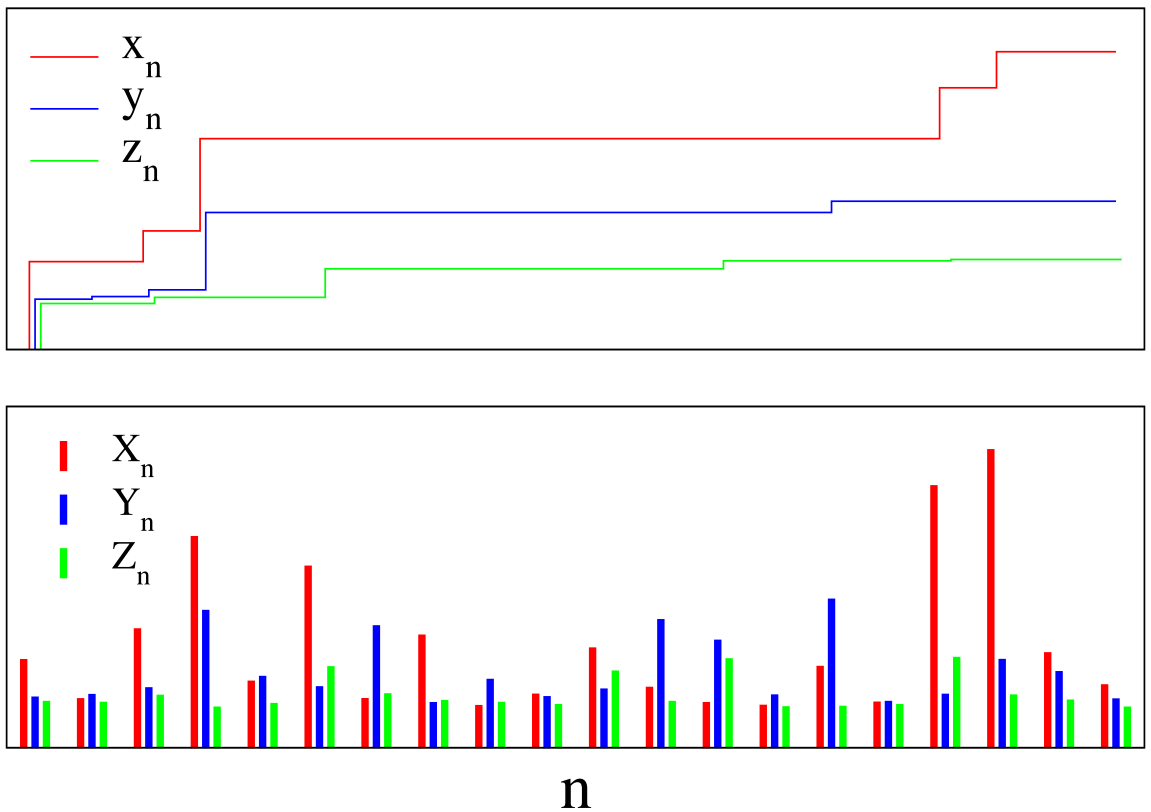
- Exponential decay

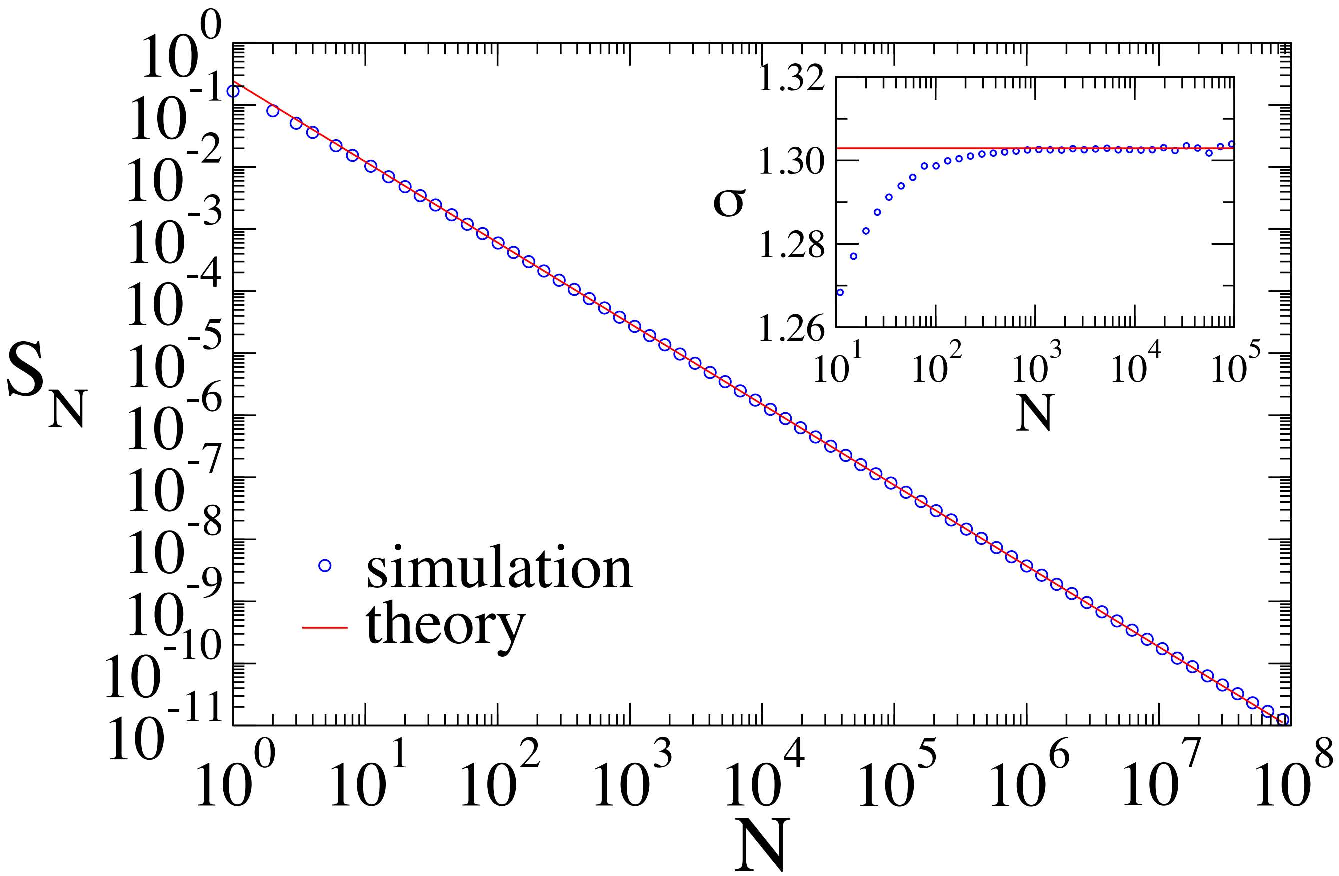
$$P_N = 2^{-N}$$

- Ordered records far more likely than ordered variables!
- Variables are uncorrelated
- Records are strongly correlated: each record "remembers" entire preceding sequence

Ordered records better suited for data analysis

Three sequences





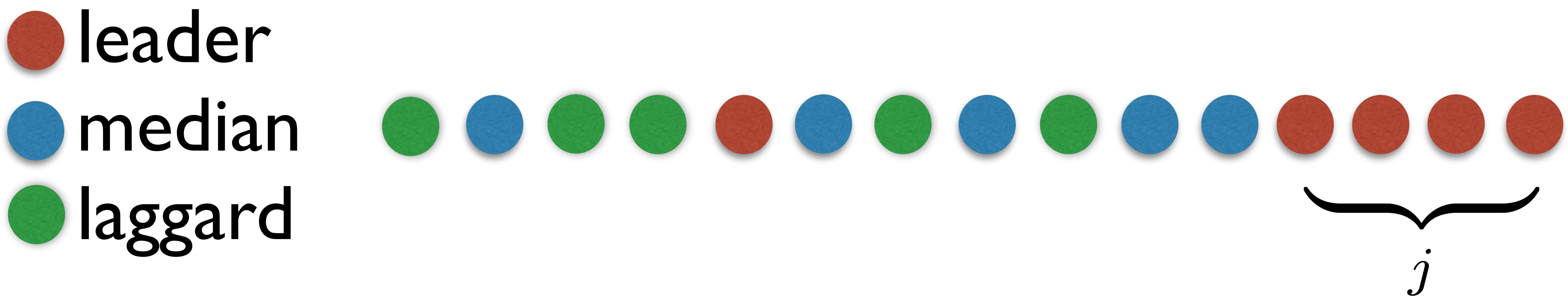
- Third sequence of random variables

$$x_n > y_n > z_n \qquad n = 1, 2, \ldots, N$$

- Probability S_N records maintain perfect order
- Power-law decay with nontrivial exponent?

$$S_N \sim N^{-\sigma} \quad \text{with} \quad \sigma = 1.3029$$

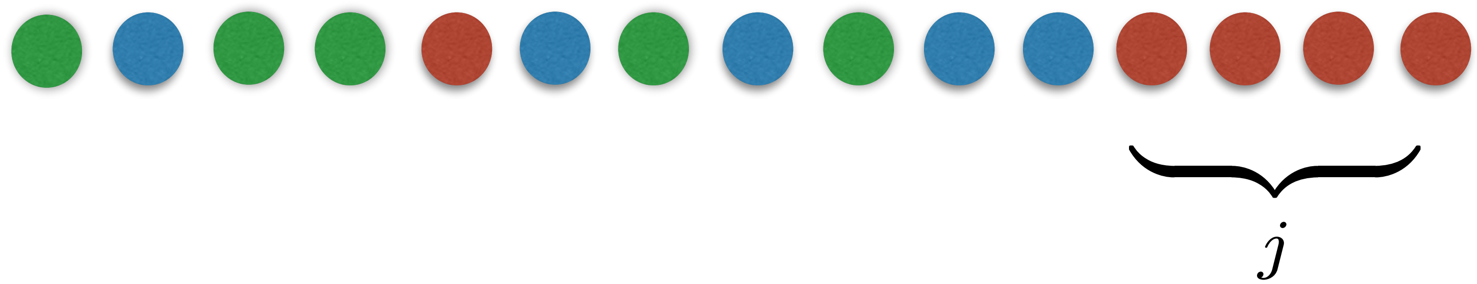
Rank of median record



- Closed equations for survival probability not feasible
- Focus on rank of the median record
- Rank of the trailing record irrelevant
- Joint probability $P_{N,j}$ that (i) records are ordered and (ii) rank of the median record equals j
- Joint probability gives the survival probability

$$S_N = \sum_{j=1}^{N} P_{N,j}$$

Closed recursion equations



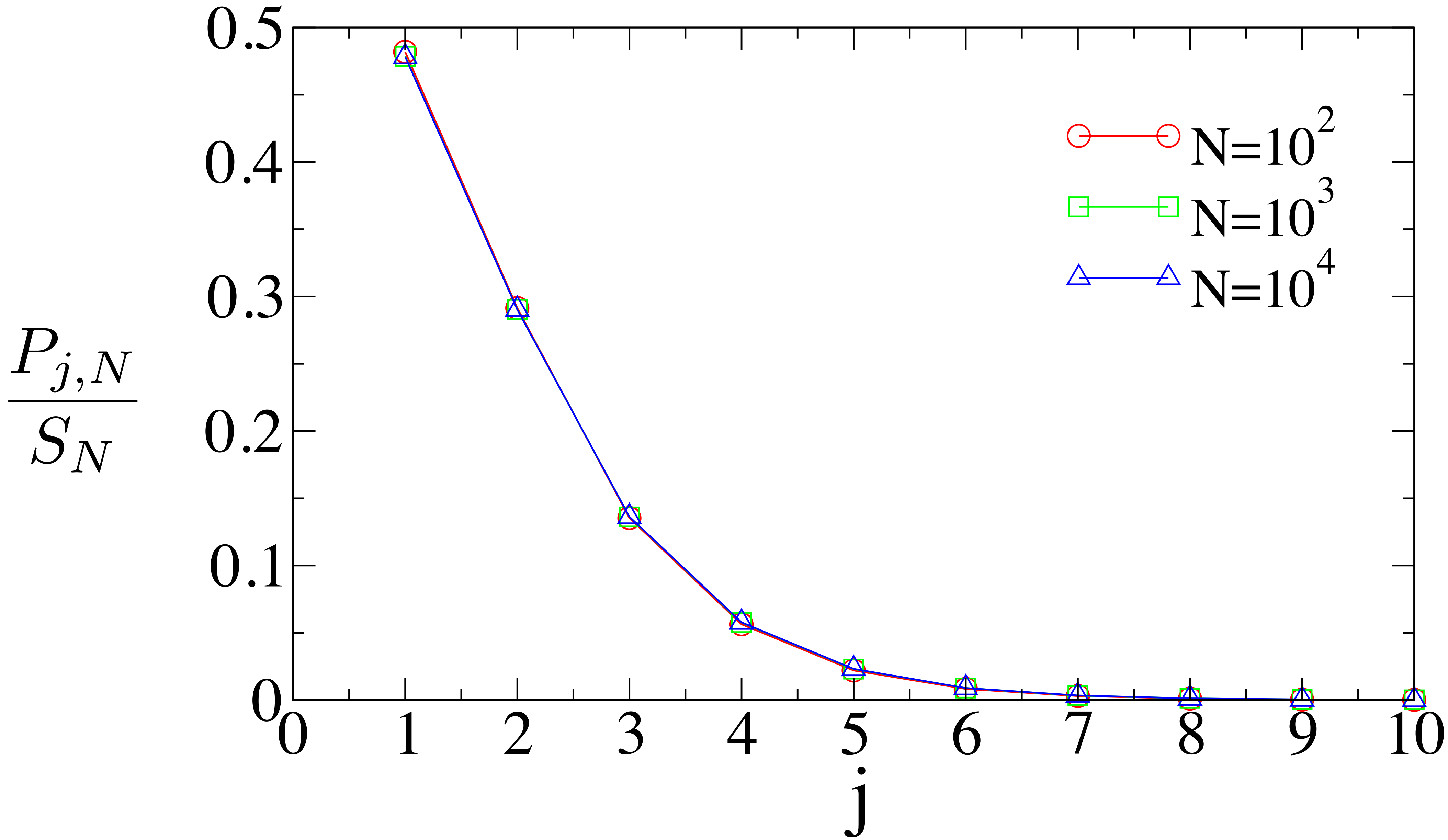
- Closed recursion equations for joint probability feasible

$$
\begin{aligned}
P_{N+1,j} = {} & \frac{3N+2-j}{3N+3}\,\frac{3N+1-j}{3N+2}\,\frac{3N-j}{3N+1}\,P_{N,j} && \mathcal{O}(N^0) \quad \text{no new records}\\
&+ \frac{3N+2-j}{3N+3}\,\frac{3N+1-j}{3N+2}\,\frac{j}{3N+1}\,P_{N,j-1} && \mathcal{O}(N^{-1}) \quad \text{new leading records}\\
&+ \frac{3N+2-j}{(3N+3)(3N+2)(3N+1)}\sum_{k=j}^{N+1}(3N-k)P_{N,k} && \mathcal{O}(N^{-1}) \quad \text{new median records}\\
&+ \frac{3N+2-j}{(3N+3)(3N+2)(3N+1)}\sum_{k=j}^{N+1}k\,P_{N,k-1} && \mathcal{O}(N^{-2}) \quad \text{two new records}
\end{aligned}
$$

- The survival probability for small N

N	S_N	$(3N)!\,S_N$
1	$\frac{1}{6}$	1
2	$\frac{29}{360}$	58
3	$\frac{4\,597}{90\,720}$	18 388
4	$\frac{5\,393}{149\,688}$	17 257 600
5	$\frac{179\,828\,183}{6\,538\,371\,840}$	35 965 636 600

Key observation



Rank of median record j and N become uncorrelated!

Asymptotic analysis

- Rank of median record j and N become uncorrelated!

$$P_{N,j} \simeq S_N\, p_j \quad \text{as} \quad N \to \infty$$

- Assume power law decay for the survival probability

$$S_N \sim N^{-\sigma}$$

- The asymptotic rank distribution is normalized

$$\sum_{j=1}^{\infty} p_j = 1$$

- Rank distribution obeys a much simpler recursion

$$\sigma\, p_j = (j+1)\, p_j - \frac{j}{3} p_{j-1} - \frac{1}{3} \sum_{k=j}^{\infty} p_k$$

Scaling exponent σ is an eigenvalue

The rank distribution

- First-order differential equation for generating function

$$(3-z)\frac{dP(z)}{dz}+P(z)\left(\frac{1}{1-z}-\frac{3\sigma}{z}\right)=\frac{z}{1-z} \qquad P(z)=\sum_{j\geq 1} p_j\, z^{j+1}$$

- Solution

$$P(z)=\sqrt{\frac{1-z}{3-z}}\left(\frac{z}{3-z}\right)^{\sigma}\int_0^z \frac{du}{(1-u)^{3/2}}\,\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}$$

- Behavior near $z=3$ gives tail of the distribution

$$p_j \sim j^{\sigma-1/2}3^{-j}$$

- Behavior near $z=1$ gives the scaling exponent

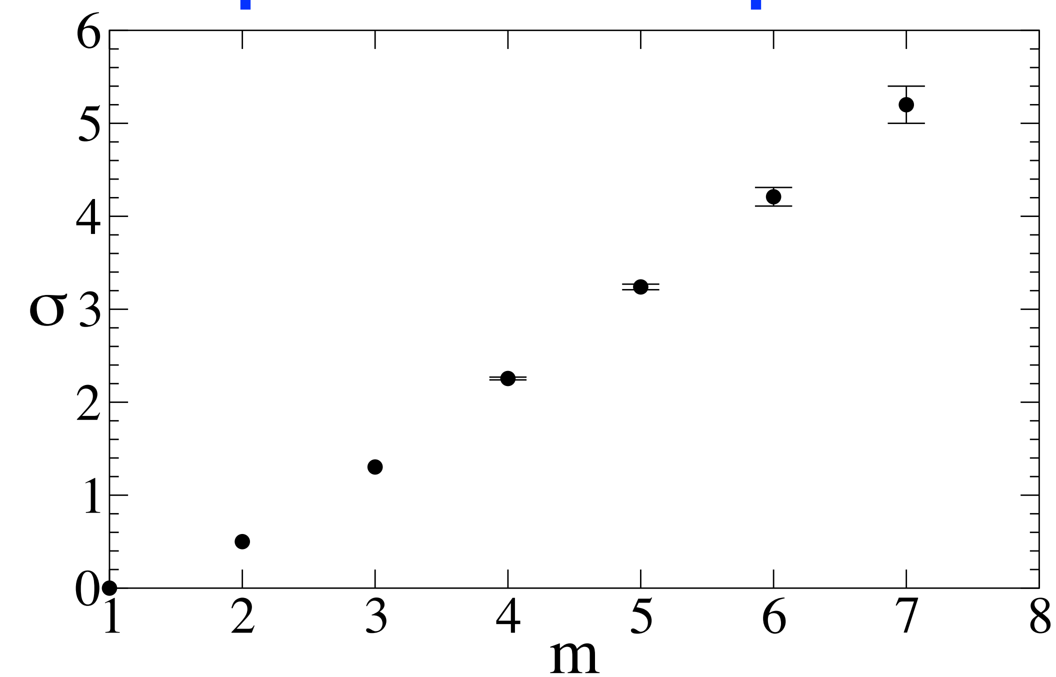
$${}_2F_1\left(-\tfrac{1}{2},\tfrac{1}{2}-\sigma;\tfrac{3}{2}-\sigma;-\tfrac{1}{2}\right)=0 \quad \Longrightarrow \quad \sigma=1.302931\ldots$$

Three sequences: scaling exponent is nontrivial

Multiple sequences

- Probability S_N that m records maintain perfect order
- Expect power-law decay with m-dependent exponent

$$S_N \sim N^{-\sigma_m}$$



- Lower and upper bounds

$$0 \leq m - \sigma_m \leq 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}$$

- Exponent grows linearly with number of sequences (up to a possible logarithmic correction)

$$\sigma_m \simeq m$$

In general, scaling exponent is nontrivial

Family of ordering exponents

- One sequence always in the lead: 1>rest

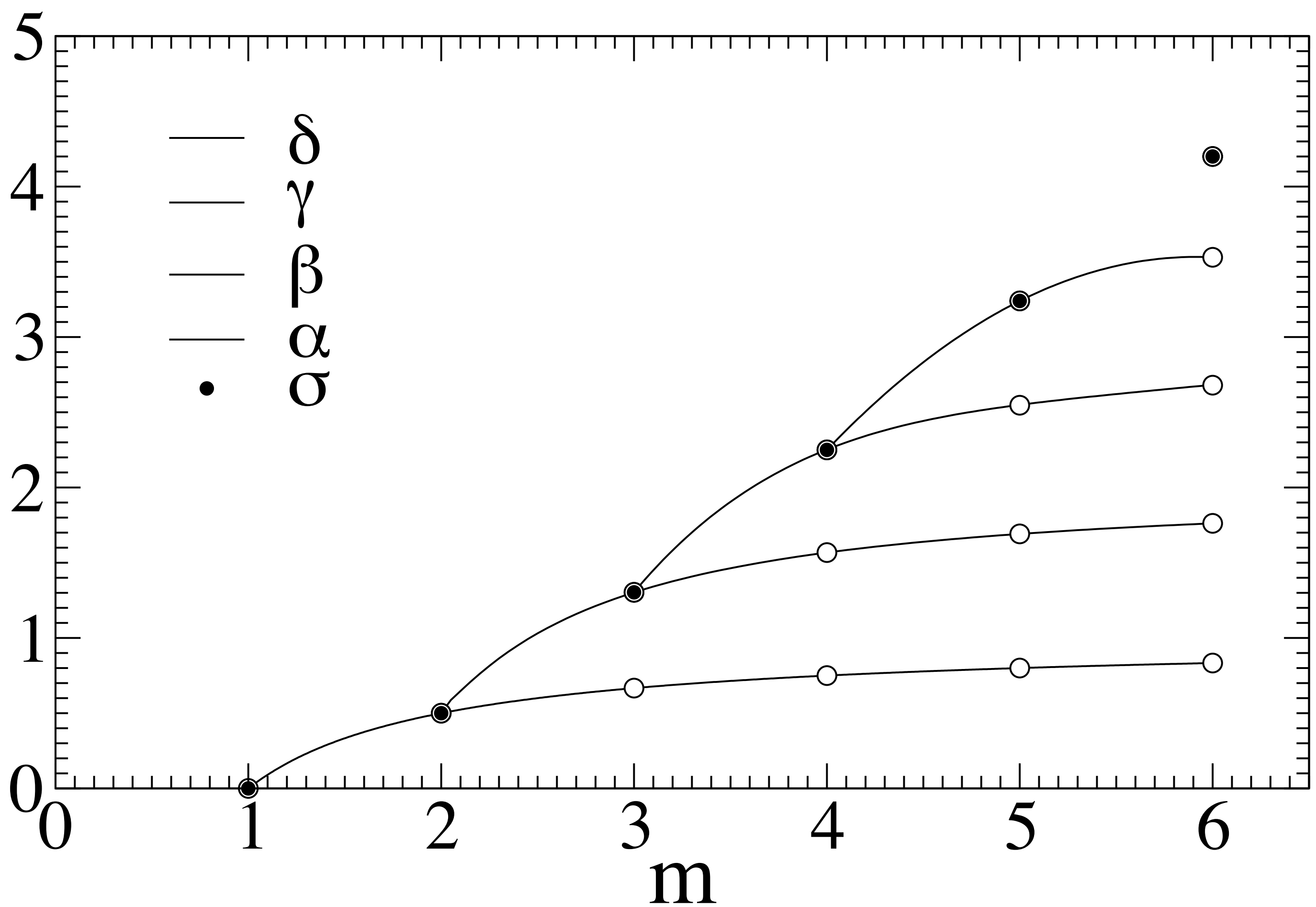
$$A_N \sim N^{-\alpha_m} \qquad \alpha_m = 1 - \frac{1}{m}$$

- Two sequences always in the lead: 1>2>rest

$$B_N \sim N^{-\beta_m} \qquad {}_2F_1\left(-\frac{1}{m-1}, \frac{m-2}{m-1} - \beta; \frac{2m-3}{m-1} - \beta; -\frac{1}{m-1}\right) = 0$$

- Three sequences always in the lead: 1>2>3>rest

$$C_N \sim N^{-\gamma_m}$$



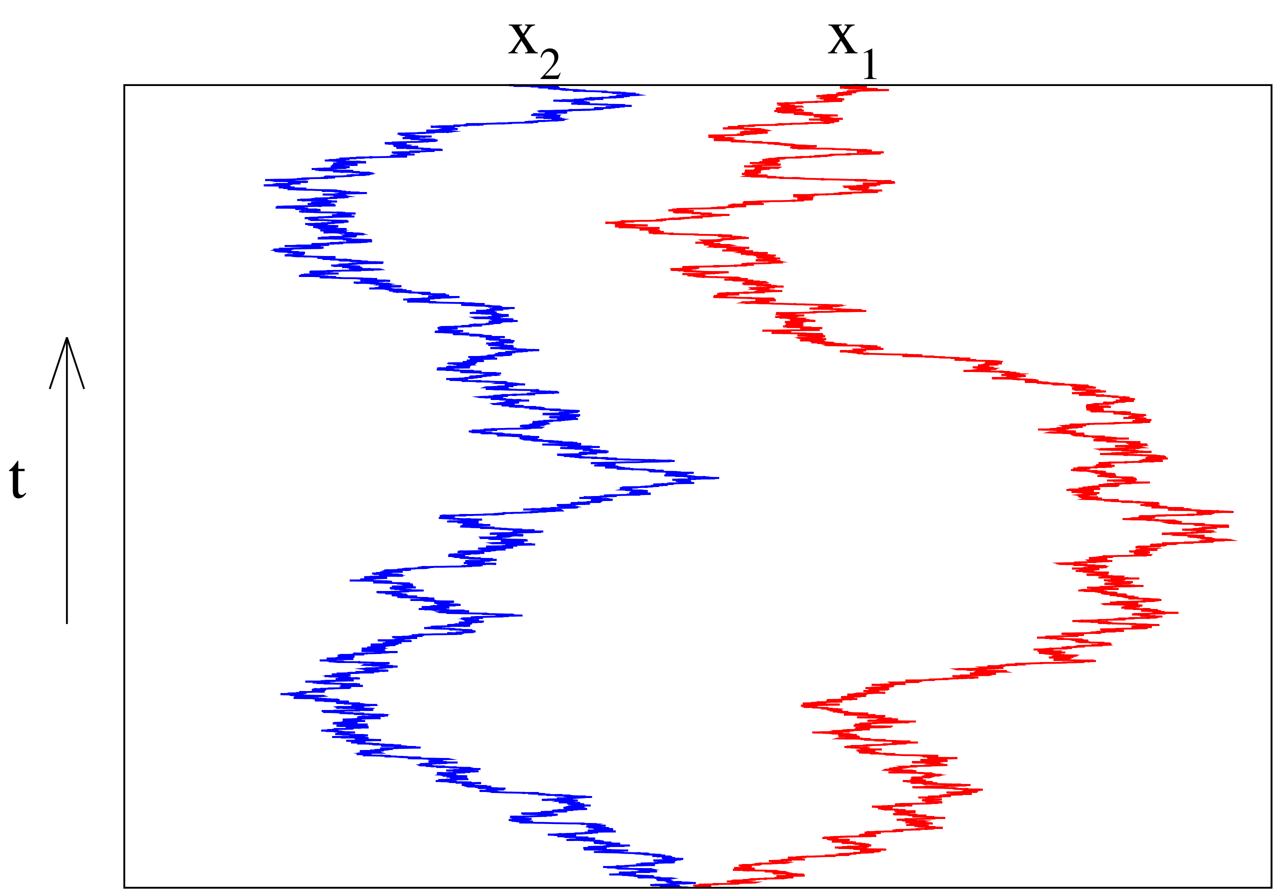
m	α	β	γ	δ
1	0			
2	1/2	1/2		
3	2/3	1.302931	1.302931	
4	3/4	1.56479	2.255	2.255
5	4/5	1.69144	2.547	3.24
6	5/6	1.76164	2.680	3.53

Recap III

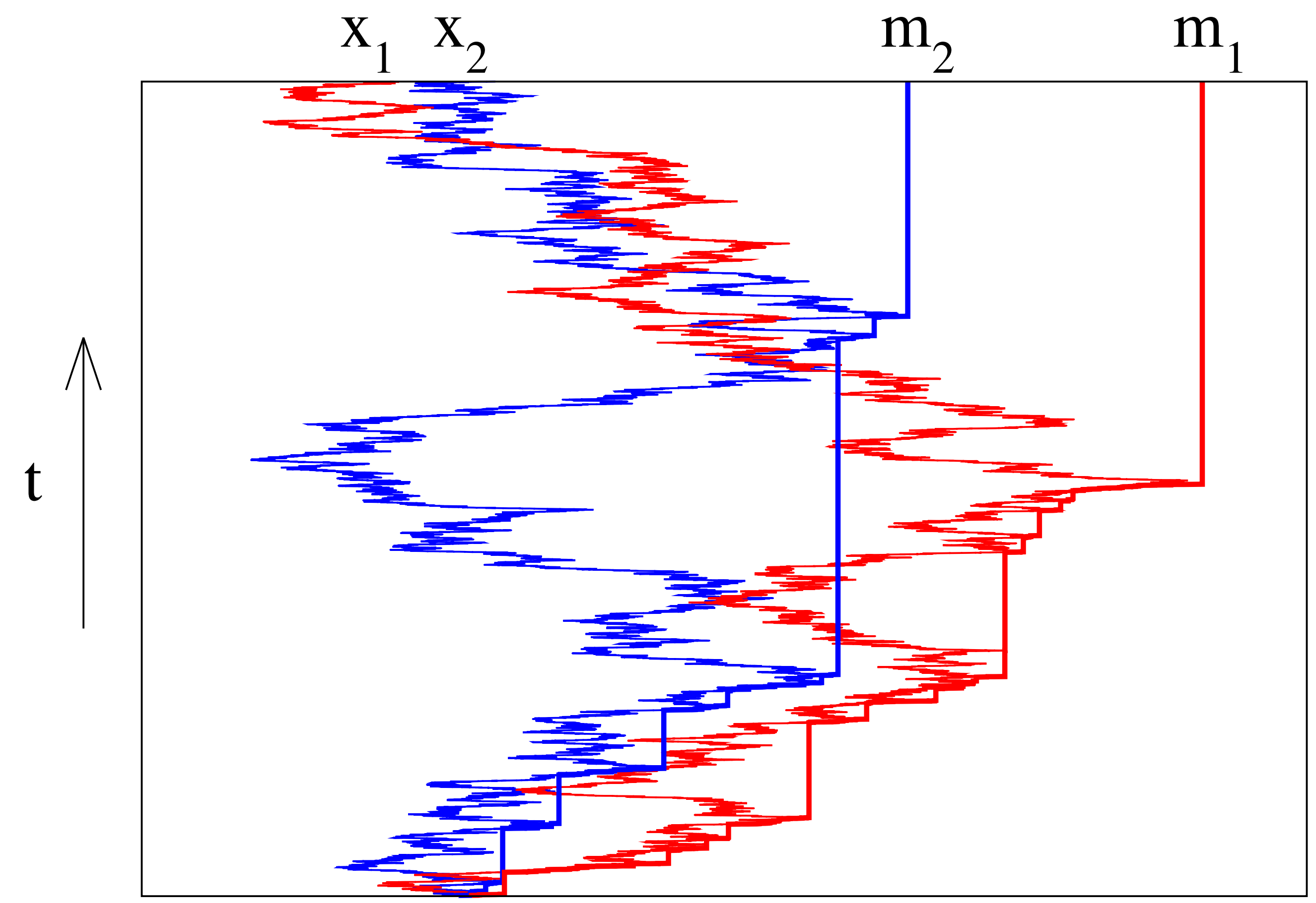
- Probability multiple sequences of records are ordered
- Uncorrelated random variables
- Survival probability independent of parent distribution
- Power-law decay with nontrivial exponent
- Exact solution for three sequences
- Scaling exponent grows linearly with number of sequences
- Key to solution: statistics of median record becomes independent of the sequence length (large N limit)
- Scaling methods allow us to tackle combinatorics

IV. Ordered records: correlated random variables

Brownian positions

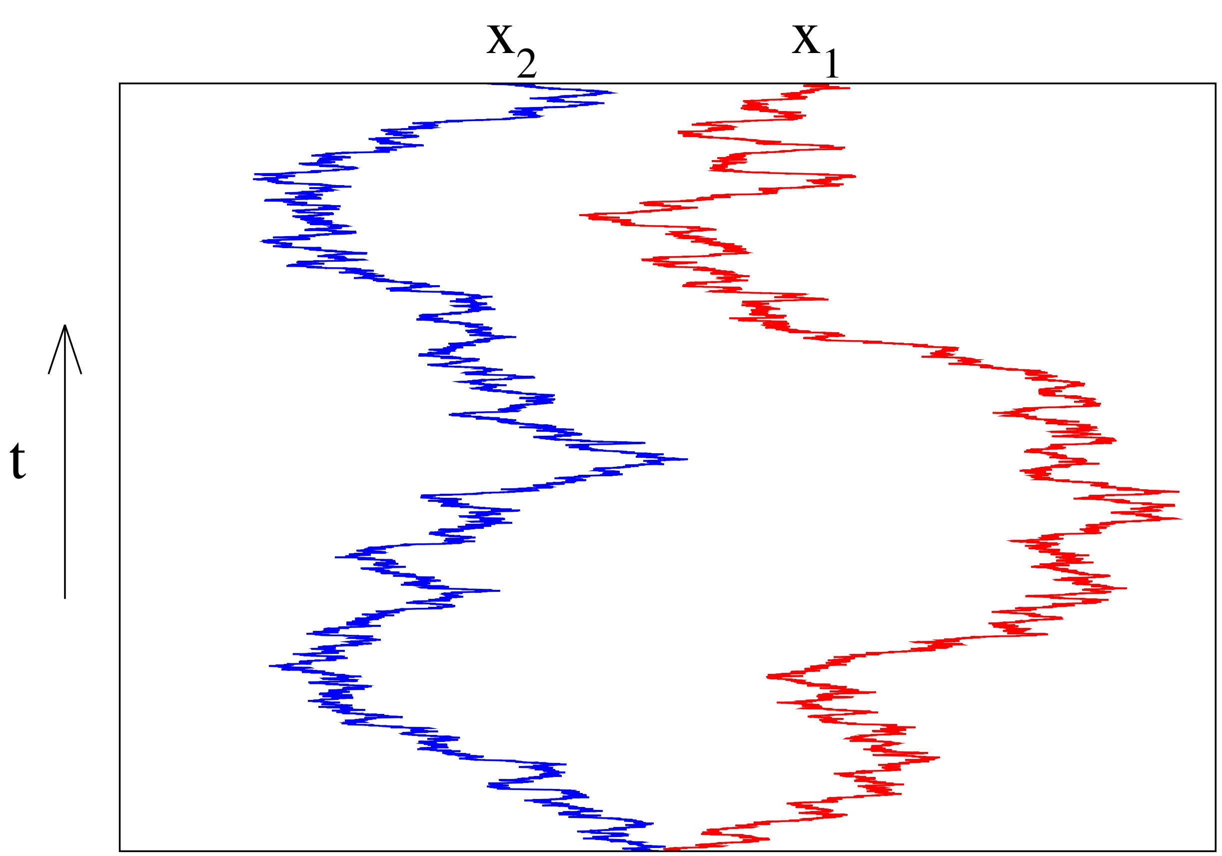


Brownian records



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

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

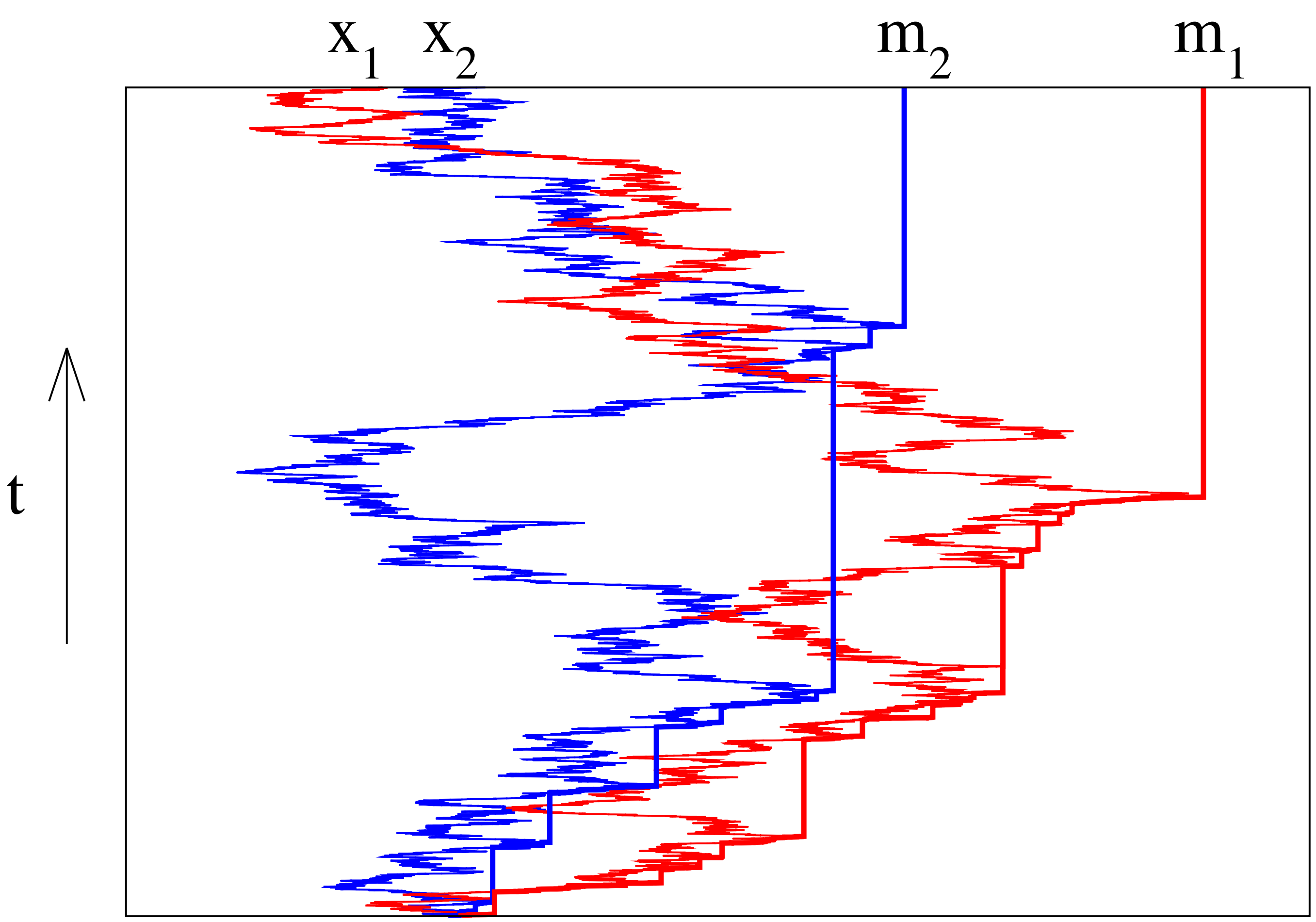
Behavior holds for Levy flights, different mobilities, etc

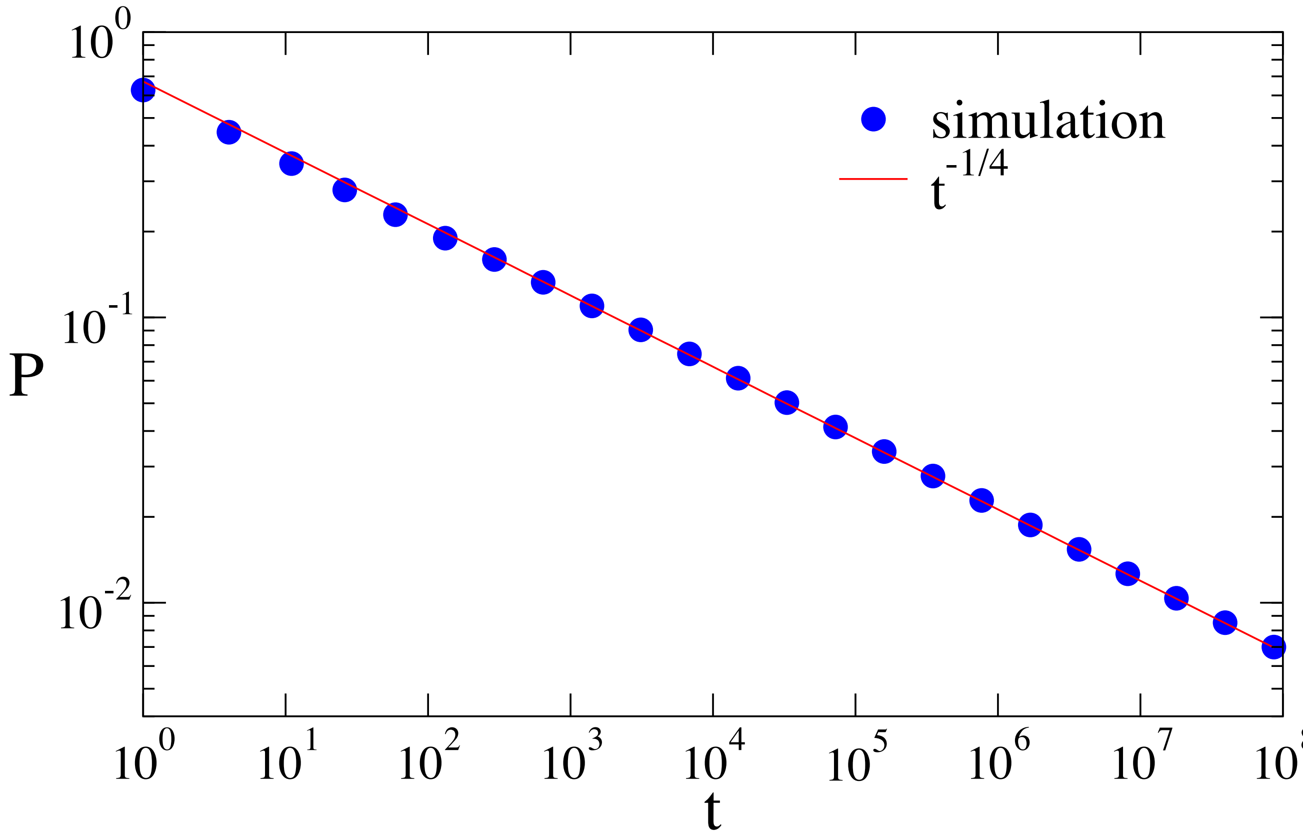
Universal first-passage exponent

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

First-passage kinetics: brownian records

Probability running records remain ordered

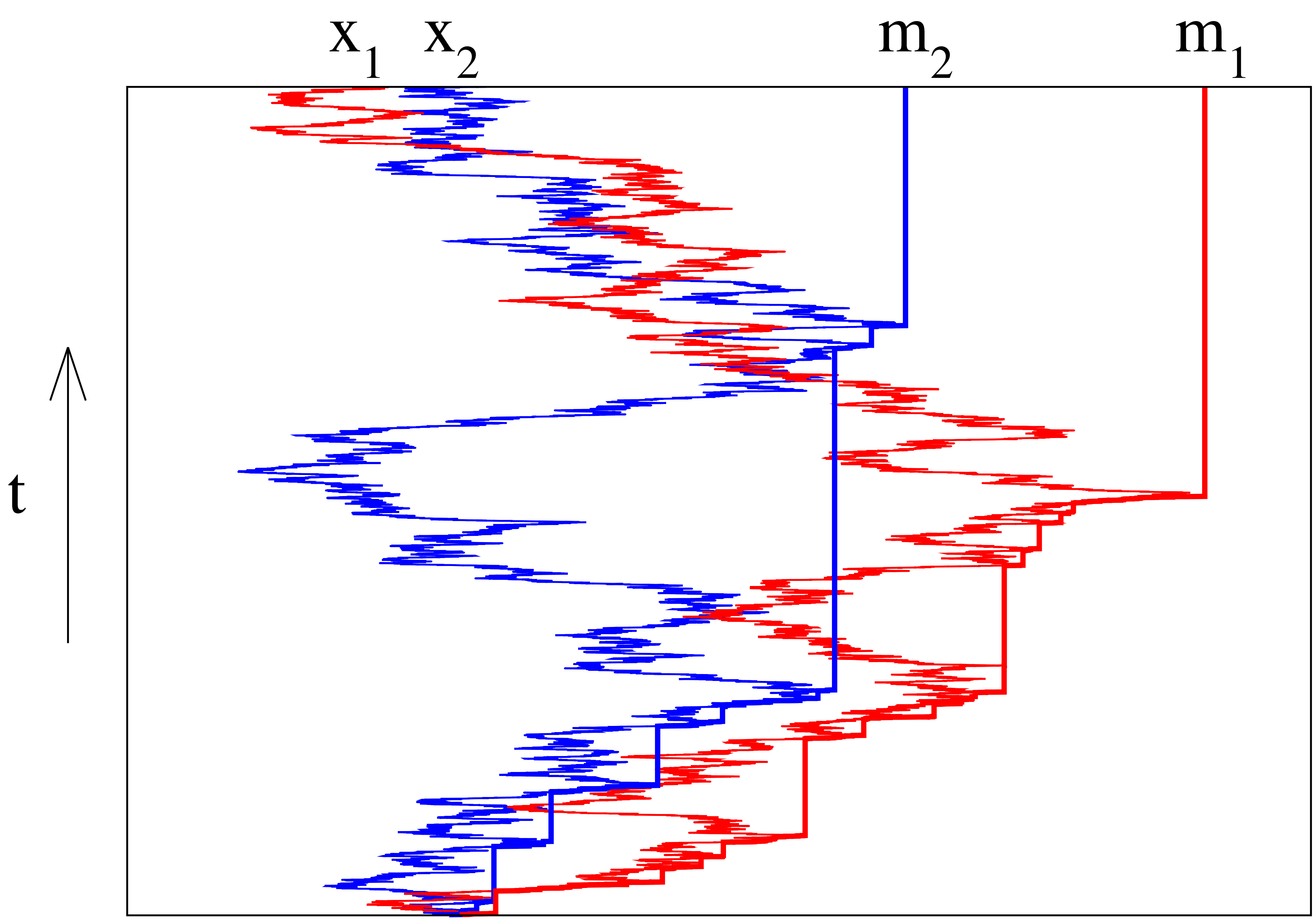




$$S \sim t^{-\beta} \qquad \beta = 0.2503 \pm 0.0005$$

Is ¼ exact? Is exponent universal?

$m_1 > m_2$ if and only if $m_1 > x_2$



From four variables to three

- Four variables: two positions, two records

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

- The two records must always be ordered

$$m_1 > m_2$$

- Key observation: trailing record is irrelevant!

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

- Three variables: two positions, one record

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

From three variables to two

- Introduce two distances from the record

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

- Both distances undergo Brownian motion

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

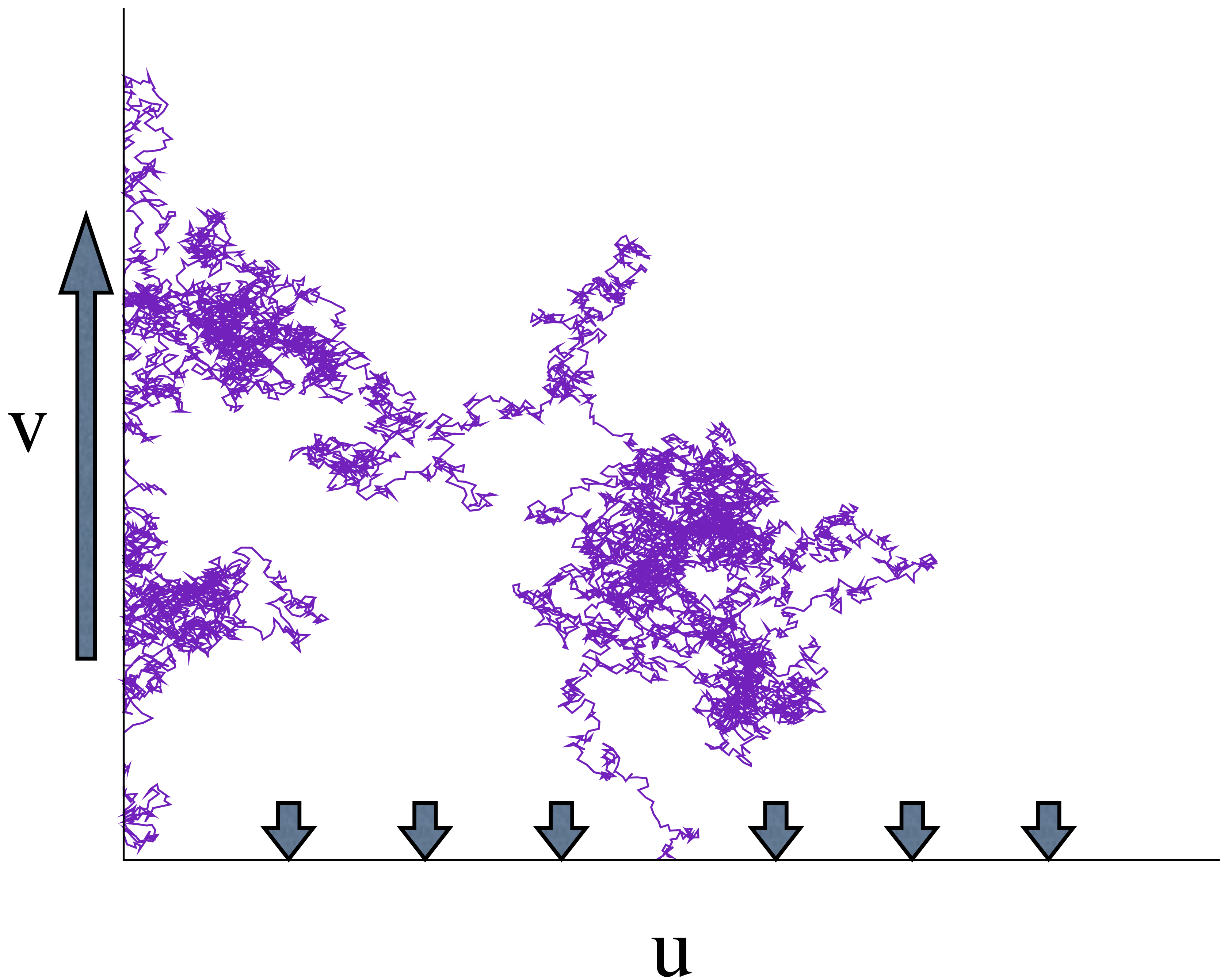
- Boundary conditions: (i) absorption (ii) advection

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

- Probability records 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 angular 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

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

EB & Krapivsky PRL 14

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 records involves ratio of mobilities

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

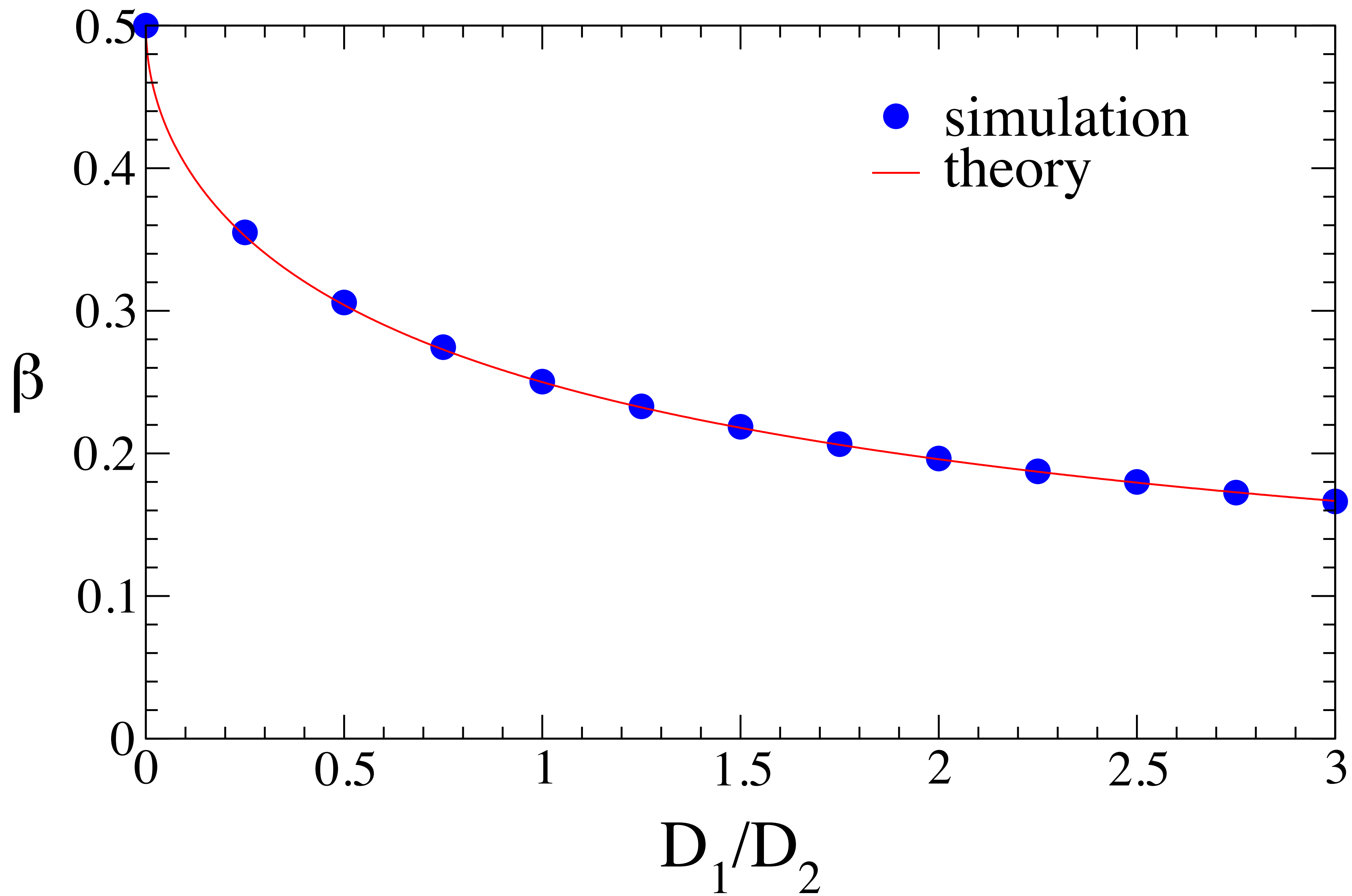
- Analysis straightforward to repeat

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

- First-passage exponent: nonuniversal, mobility-dependent

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

Numerical verification



perfect agreement

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

Multiple particles

- Probability n Brownian positions are perfectly ordered

$$P_n \sim t^{-\alpha_n} \qquad \alpha_n = \frac{n(n-1)}{4}$$

Fisher & Huse 88

- Records perfectly ordered

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

- In general, power-law decay

$$S_n \sim t^{-\nu_n}$$

n	ν_n	$\sigma_n/2$
2	1/4	1/4
3	0.653	0.651465
4	1.13	1.128
5	1.60	1.62
6	2.01	2.10

Uncorrelated variables provide an excellent approximation

Suggests some record statistics can be robust

Recap IV

- 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
- Why do uncorrelated variables represent an excellent approximation?

First-passage statistics of extreme values

- Survival probabilities decay as power law
- First-passage exponents are nontrivial
- Theoretical approach: differs from question to question
- Concepts of nonequilibrium statistical mechanics are powerful: scaling, correlations, large system-size limit
- Many, many open questions
- Ordered records as a data analysis tool

Publications

1. Scaling Exponents for Ordered Maxima,
   E. Ben-Naim and P.L. Krapivsky,
   Phys. Rev. E **92**, 062139 (2015)

2. Slow Kinetics of Brownian Maxima,
   E. Ben-Naim and P.L. Krapivsky,
   Phys. Rev. Lett. **113**, 030604 (2014)

3. Persistence of Random Walk Records,
   E. Ben-Naim and P.L. Krapivsky,
   J. Phys. A **47**, 255002 (2014)

4. Scaling Exponent for Incremental Records,
   P.W. Miller and E. Ben-Naim,
   J. Stat. Mech. P10025 (2013)

5. Statistics of Superior Records,
   E. Ben-Naim and P.L. Krapivsky,
   Phys. Rev. E **88**, 022145 (2013)