An introduction to extreme value statistics (part 2/2)

Nicholas R. Moloney



Dresden October 06, 2015 Range of applications

Gallery of correlations

Illustrations of EVS in statistical physics DNA replication times in Xenopus Laevis Order parameter in percolation EVS in trees EVS of some Gaussian processes Integer partitions and the ideal Bose gas

# Range of applications Overview/statistical physics

J.-P. Bouchaud, M. Potters *Theory of Financial Risks* Cambridge University Press (2000)

D. Sornette Critical Phenomena in Natural Sciences Springer (2006)

M. Clusel, E. Bertin Int. J. Mod. Phys. B **22**, 3311 (2008)

S. N. Majumdar, A. Pal arXiv:1406.6768v3 (2014)

J.-Y. Fortin, M. Clusel J. Phys. A: Math. Theor. **48**, 183001 (2015) Introduction

Introduction

Review

**Review** 

Lecture notes

# Range of applications Modelling

R. W. Katz, M. B. Parlange, P. Naveau Adv. Water Resour. **25**, 1287 (2002)

A. J. McNeil, R. Frey J. Empir. Financ. **7**, 271 (2000)

T. Antal *et al.* Eur. Phys. Lett. **88**, 59001 (2009)

J. Bechhoefer, B. Marshall Phys. Rev. Lett. **98**, 098105 (2007)

S. Hallerberg, J. Bröcker, H. Kantz
In Nonlinear Time Series Analysis in the Geosciences,
R. V. Donner, S. M. Barbosa (eds), Springer (2008)

Hydrology

Finance

Astronomy

Cell biology

Prediction/Forecasting

# Range of applications Disordered systems

J.-P. Bouchaud, M. Mézard J. Phys. A: Math. Gen. **30**, 7997 (1997) Random energy model

S. N. Majumdar, P. L. Krapivksy Physica A **318**, 161 (2003) **Hierarchical correlations** 

G. Biroli, J.-P. Bouchaud, M. Potters J. Stat. Mech. P07019 (2007) **Disordered systems** 

Y. V. Fyodorov Physica A **389**, 4229 (2010) Logarithmic correlations

S. N. Majumdar, G. Schehr

J. Stat. Mech. P01012 (2014)

**Random matrices** 

### Range of applications

#### Dynamical systems

C. Nicolis, V. Balakrishnan, G. Nicolis Phys. Rev. Lett. 210602 (2006)

M. Ghil *et al.* Review Nonlin. Processes Geophys. **18**, 295 (2011)

J. M. Freitas Dynamical Systems **28**, 302 (2013)

H. Aytaç, J. M. Freitas, S. Vaienti Trans. Amer. Math. Soc. **367**, 8229 (2015) Extreme events in chaotic dynamics

EVS in chaotic dynamics

EVS in deterministic and random dynamical systems

# Gallery of correlations

Time series



$$\langle x(t')x(t'+t)\rangle \sim t^{-\gamma}$$



$$g(r) \sim r^{-(d-2+\eta)}$$

Trees

 $g(i,j) \propto$  length of shared path

Random matrices

 $\begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \qquad P[\lambda_i]$ 

$$P[\lambda_i] \propto \exp\left[-\beta\left(\frac{N}{2}\sum_{i=1}^N \lambda_i^2 - \frac{1}{2}\sum_{i\neq j}\ln|\lambda_i - \lambda_j|\right)\right]$$

### Outline

Range of applications

#### Gallery of correlations

#### Illustrations of EVS in statistical physics DNA replication times in Xenopus Laevis

Order parameter in percolation EVS in trees EVS of some Gaussian processes Integer partitions and the ideal Bose gas

# DNA replication in Xenopus Laevis

Bechhoefer, Marshall 2007



FIG. 1. Schematic of DNA replication model. Space-time diagram showing multiple origins (filled circles), each expanding symmetrically at constant velocity. Domains coalesce when they meet (open circles).

 $\sim 3\times 10^9~\text{bases}$ 

- $\sim 10^5 \mbox{ replication origins}$
- $\sim 1 {\rm kb}/{\rm min}$  fork velocity
- $\sim 20$  min replication time
- $\sim 10^{-4}$  failure probability

# DNA replication in Xenopus Laevis

Bechhoefer, Marshall 2007



FIG. 2 (color online). Replication-time distribution function, fixing the mode to be  $t^* = 38$  minutes. Markers are results from Monte Carlo simulations (3000 trials per simulation); solid lines are fits to the Gumbel distribution.

### Outline

Range of applications

#### Gallery of correlations

#### Illustrations of EVS in statistical physics DNA replication times in Xenopus Laevis Order parameter in percolation EVS in trees EVS of some Gaussian processes Integer partitions and the ideal Bose gas

### Percolation



Fraction occupied by the largest cluster is the order parameter.

# Subcritical percolation ( $p \ll p_c$ )

Duxbury, Leath 1987; Bazant 2000; Borgs et al. 2001; van der Hofstad, Redig 2006



largest cluster  $\sim \log L$ 

overall largest cluster is the largest cluster from uncorrelated blocks



# Critical percolation ( $p = p_c$ )

Hovi, Aharony 1996; Sen 2001; Borgs et al. 2001



largest cluster  $\sim L^D$  (D = 91/48)

# correlations span the system! non-trivial distribution emerges



### Supercritical percolation ( $p \gg p_c$ ) Borgs et al. 2001



largest cluster  $\sim L^d$ 

largest cluster is extensive: sum up contributions from uncorrelated blocks



#### Percolation

P. M. Duxbury, P. L. Leath J. Phys. A: Math. Gen. **20**, L411 (1987) Largest cluster in subcritical percolation

J.-P. Hovi, A. Aharony Phys. Rev. E **53**, 235 (1996) Clusters in critical percolation

M. Z. Bazant Largest cluster in subcritical percolation Phys. Rev. E 62, 1660 (2000)

P. Sen Largest clusters in critical percolation J. Phys. A: Math. Gen. **34**, 8477 (2001)

C. Borgs *et al.* Finite-size scaling of largest cluster Commun. Math. Phys. **224**, 153 (2001)

R. van der Hofstad, F. Redig J. Stat. Phys. **122**, 671 (2006) Largest clusters in non-critical percolation

### Outline

Range of applications

#### Gallery of correlations

#### Illustrations of EVS in statistical physics

DNA replication times in Xenopus Laevis Order parameter in percolation

#### EVS in trees

EVS of some Gaussian processes Integer partitions and the ideal Bose gas

#### Majumdar, Krapivsky 2000



Analogy between 1-0 behaviour and wave fronts (c.f.  $\partial_t \phi = \partial_{xx} \phi + \phi - \phi^2$ )

$$\mathbb{P}(\ell_{n+1} \ge x) = F_{n+1}(x) = [(1-p)F_n(x) + pF_n(x-1)]^2$$
  
initial condition:  $F_0(x) = \begin{cases} 1, & x \le 0\\ 0, & x > 0 \end{cases}$ 



Tracking the tip of the front

$$\begin{array}{c}
q_n = F_n(0) - F_n(1) \\
= 1 - F_n(1)
\end{array}$$

Substitute  $q_n = 1 - F_n(1)$  into recursion relation for  $F_n$ :

$$q_{n+1} = 2(1-p)q_n - (1-p)^2 q_n^2$$



p < 1/2: wave front is 'pinned'

Wave front gets pinned at a finite  $\langle \ell_n \rangle$  as  $n \to \infty$ .

 $F_n(x)$  approaches a stationary distribution:

$$\mathcal{G}(x) = [(1-p)\mathcal{G}(x) + p\mathcal{G}(x-1)]^2$$

Iterate, starting with  $\mathcal{G}(0) = 1$ , to obtain  $\mathcal{G}(1), \mathcal{G}(2), \dots$ 



#### $p \ge 1/2$ : wave front is 'depinned'

Wave front travels as a fixed shape  $\mathcal{F}^2$  located at  $\langle \ell_n \rangle \approx v(p)n$ :

$$F_{n+1}(x) \to \mathcal{F}^2(x - v(n+1))$$
$$F_n(x) \to \mathcal{F}^2(x - vn)$$
$$F_n(x-1) \to \mathcal{F}^2(x - v-1)$$

Substitute into recursion relation:

$$\mathcal{F}(z-v) = (1-p)\mathcal{F}^2(z) + p\mathcal{F}^2(z-1), \quad z = x - vn.$$

Behind the front,  $1 - \mathcal{F}(z) \approx e^{\lambda z}$ . Substitute and linearise:

$$v_{\lambda}(p) = -rac{\log[2(1-p)+2pe^{-\lambda}]}{\lambda}.$$

Front selects  $v_{\lambda^{\star}}(p)$ .

Bramson 1978; Brunet, Derrida 1997; Ebert, van Saarloos 1998; Majumdar, Krapivsky 2000



(Majumdar & Krapivsky 2000)

FIG. 1. The propagating front for the cumulative distribution  $P_{n}(x)$  of the minimal length for the bimodal distribution with p = 0.8.

$$\langle \ell_n \rangle = \begin{cases} c(p), & p < 1/2\\ (\log 2)^{-1} \log \log n, & p = 1/2\\ v(p)n, & p > 1/2 \end{cases}$$

Hierarchichal structure is rather generic...



 $\mathbb{P}(\min_{k} E_{k}^{(n+1)} > x) = F_{n+1}(x) = \left[\int_{0}^{\infty} \mathrm{d}\epsilon F_{n}(x-\epsilon)\rho(\epsilon)\right]^{2}$ 

- Write down recursion relation between  $F_n$  and  $F_{n+1}$ .
- Identify conditions for travelling wave solution.
- Linearise wave equation.
- Calculate wave speeds, etc.

S. N. Majumdar, P. L. Krapivksy Phys. Rev. E **62**, 7735 (2000)

E. Ben-Naim, P. L. Krapivksy, S. N. Majumdar Phys. Rev. E **64**, 035101(R) (2001)

D. S. Dean, S. N. Majumdar Phys. Rev. E **64**, 046121 (2001)

S. N. Majumdar, P. L. Krapivksy Phys. Rev. E **65**, 036127 (2002)

D. S. Dean, S. N. Majumdar J. Phys. A: Math. Gen. **35**, L501 (2002)

S. N. Majumdar, P. L. Krapivksy Physica A **318**, 161 (2003) Extremal paths on random Cayley trees

Extremal paths on random Cayley trees

Polymers on random binary trees

Polymers, binary search trees

Fragmentation, *m*-ary search

**Overview** article

### Outline

Range of applications

#### Gallery of correlations

#### Illustrations of EVS in statistical physics

DNA replication times in Xenopus Laevis Order parameter in percolation EVS in trees EVS of some Gaussian processes Integer partitions and the ideal Bose gas



 $\langle X(s)X(t)\rangle = \min(s,t) - st$ 

(Berman's condition fails)

$$\mathbb{P}(\max_{t} x_{B}(t) \le m) = \frac{W_{m}(x_{t} = 0, t \mid x_{0} = 0, 0)}{W(x_{t} = 0, t \mid x_{0} = 0, 0)},$$

where transition probability  $W_m(x_t, t | x_0 = 0, 0)$  satisfies

$$\begin{pmatrix} \frac{\partial}{\partial t} - \frac{1}{2} \frac{\partial^2}{\partial x^2} + V_m(x) \end{pmatrix} W_m = 0 V_m(x) = \begin{cases} 0, & x < m, \\ \infty, & x \ge m. \end{cases}$$

**Reflection principle** 



**Reflection principle** 



 $W_m(0, t \mid 0, 0) = W(0, t \mid 0, 0)$  – overcounting term

**Reflection principle** 



 $W_m(0, t \mid 0, 0) = W(0, t \mid 0, 0)$  – overcounting term

**Reflection principle** 



# Mean-subtracted Brownian bridge

Watson 1961

Define a new Gaussian process by substracting the bridge's mean:

$$x_W(t) := x_B(t) - \int_0^t \mathrm{d}t' \, x_B(t').$$

'Watson' bridge has correlations

$$\langle X(s)X(t)\rangle = \min(s,t) - \frac{1}{2}(s+t) + \frac{1}{2}(s-t)^2 + \frac{1}{12}.$$

This depends only on time differences:

$$\langle X(s)X(s+t)\rangle = \frac{1}{12} \left[1 - 6|t|(1-|t|)\right].$$



Fourier decomposition

$$a_k, b_k \stackrel{d}{\sim} N(0,1)$$

$$\frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} \frac{a_k}{k} \sin \pi kt \qquad \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{1}{k} \left( a_k \sin 2\pi kt + b_k \cos 2\pi kt \right)$$

Signal construction

$$h(t) \propto \sum_{k=1}^{n} \frac{1}{k^{\alpha/2}} \left[ a_k \cos\left(\frac{2\pi k}{T}t\right) + b_k \sin\left(\frac{2\pi k}{T}t\right) \right], a_k, b_k \stackrel{d}{\sim} N(0, 1)$$
white  $1/f$  Edwards Mullins single mode mode wilkinson Herring  $mode$ 

$$\alpha = 0 \qquad 1 \qquad 2 \qquad 4 \qquad \alpha \rightarrow \infty$$
weaker correlations correlations

Maximum relative height: Raychaudhuri et al. 2001















### Extremes of Gaussian processes

S. M. Berman Ann. Math. Stat. **33**, 502 (1964)

P. Biane, J. Pitman, M. Yor Bull. Amer. Math. Soc **38**, 435 (2001)

S. N. Majumdar, A. Comtet J. Stat. Phys. **119**, 777 (2005)

G. Györgyi *et al.* Phys. Rev. E **75**, 021123 (2007)

Y. V. Fyodorov, J.-P. Bouchaud J. Phys. A: Math. Theor. **41**, 372001 (2008)

H. J. Hilhorst, P. Calka, G. Schehr J. Stat. Mech. P10010 (2008) MRH for Edwards-Wilkinson  $(1/f^2)$ 

MRH for  $1/f^{\alpha}$  noise

Berman's condition

Review

MRH for 1/f noise

Random accelaration process  $(1/f^4)$ 

Antal et al. 2001



where  $\varepsilon_k$  are standard exponential variables.

Antal et al. observed that

 $w^2 \stackrel{d}{\sim}$  Gumbel!

Rényi 1953; Bertin 2005

$$\begin{matrix} & & & \\ 0 & X_{1,n} & X_{2,n} \end{matrix}^{w^2} \\ X_{n,n} & = \max\{X_{1,n}, \dots, X_{n,n}\} \end{matrix}$$

#### Interpret the partial sums

$$X_{m,n} := \sum_{k=1}^{m} \frac{\varepsilon_k}{n-k+1}, \quad m \le n$$

as marking the positions of a collection of ordered points  $X_{m,n}$ . In particular

$$w^2 = X_{n,n} = \max\{X_{1,n}, \ldots, X_{n,n}\}.$$

#### Rényi 1953; Bertin 2005

Starting from a collection of iid standard exponentials

$$f_{\varepsilon_1,\ldots,\varepsilon_n}(y_1,\ldots,y_n) = \prod_{k=1}^n \exp(-y_k) = \exp\left(-\sum_{k=1}^n y_k\right),$$

change variables with

$$x_m = \sum_{k=1}^m \frac{y_k}{n-k+1}, \quad \left|\frac{\partial y_i}{\partial x_j}\right| = n!$$

to obtain

$$f_{X_{1,n},...,X_{n,n}}(x_1,...,x_n) = n! \exp\left(-\sum_{k=1}^n x_k\right) \mathbf{1}_{\{x_1 < \dots < x_n\}}$$
$$f_{X_1,...,X_n}(x_1,...,x_n) = \exp\left(-\sum_{k=1}^n x_k\right) = \prod_{k=1}^n \exp(-x_k).$$

Rényi 1953; Bertin 2005

Rényi's representation

$$\{X_{m,n}, m=1,\ldots,n\} \stackrel{d}{\sim} \left\{\sum_{k=1}^{m} \frac{\varepsilon_k}{n-k+1}, m=1,\ldots,n\right\}$$

implies roughness of 1/f noise

$$w^2 = \max\{X_1,\ldots,X_n\} = X_{n,n} \stackrel{d}{\sim}$$
Gumbel,

where  $X_k$ ,  $\varepsilon_k$  are standard exponentials.

A. Rényi

Acta Mathematica Scient. Hungar. IV, 191 (1953)

T. Antal *et al.* Phys. Rev. Lett. **87**, 240601 (2001)

E. Bertin Phys. Rev. Lett. **95**, 170601 (2005)

E. Bertin, M. Clusel

J. Phys. A: Math. Gen. 39, 7607 (2006)

Order statistics

Roughness of 1/f noise

Order statistics, global fluctuations

Order statistics, global fluctuations

### Outline

Range of applications

#### Gallery of correlations

#### Illustrations of EVS in statistical physics

DNA replication times in Xenopus Laevis Order parameter in percolation EVS in trees EVS of some Gaussian processes Integer partitions and the ideal Bose gas

 $\Omega(E)$  = number of ways of partitioning an integer *E* into a sum of (indistinguishable) positive integers.

E.g.  $\Omega(5) = 7$ :

 $\Omega(E)$  = number of ways of partitioning an integer *E* into a sum of (indistinguishable) positive integers.

E.g.  $\Omega(5) = 7$ : 5 4 + 13 + 23 + 1 + 12 + 2 + 12 + 1 + 1 + 11 + 1 + 1 + 1 + 1

 $\begin{array}{c|c} \epsilon_4 & & & 0 \\ \hline \epsilon_3 & \bullet & & 1 \\ \hline \epsilon_2 & & & 0 \\ \hline \epsilon_1 & \bullet & 2 \\ \hline \epsilon_0 & \text{ground state} & n_0 \end{array}$ 

 $\epsilon_k$ 

 $\epsilon_5$ 

summands expansion Young diagram

#### Bose gas

 $n_k$ 

0

 $\Omega(E, N) =$  number of ways of partitioning *E* with *N* integers.



Hardy, Ramanujan 1918; van Lier, Uhlenbeck 1937; Erdős, Lehner 1941; Auluck, Kothari 1946

• Number of partitions grows rapidly (e.g.  $\Omega(E = 1000) \approx 2.4 \times 10^{31}$ ).

$$\Omega(E) \sim \frac{1}{4\sqrt{3}E} \exp\left(\pi \sqrt{\frac{2}{3}}E^{1/2}\right).$$

▶ Probability of partitioning *E* with *N* integers converges  $(E, N \gg 1)$  to Gumbel distribution!

$$\frac{\sum_{N'=1}^{N} \Omega(E, N')}{\Omega(E)} \sim \exp\left[-\exp\left(\frac{N-b_E}{a_E}\right)\right]$$
$$a_E = \frac{\sqrt{6}}{\pi} E^{1/2}$$
$$b_E = \frac{1}{\pi} \sqrt{\frac{3}{2}} E^{1/2} \log E.$$

Auluck, Kothari 1946; Comtet, Leboeuf, Majumdar 2007

$$\Omega(E) = \sum_{\{n_k\}} \delta\left(E - \sum_{k=1}^{\infty} n_k \epsilon_k\right)$$
$$\mathcal{Z}(\beta) = \sum_E \Omega(E) e^{-\beta E} = \prod_{k=1}^{\infty} \frac{1}{1 - e^{-\beta \epsilon_k}}.$$

Saddle point approximation:

$$\Omega(E) \approx e^{\mathcal{S}(\beta_0, E)},$$

where  $\beta_0$  maximises

$$S(\beta, E) = \log \mathcal{Z}(\beta) + \beta E.$$

Using

$$E = \sum_{k=1}^{\infty} \frac{\epsilon_k}{e^{\beta_0 \epsilon_k} - 1} \approx \int_0^{\infty} \mathrm{d}\epsilon \, \frac{\epsilon}{e^{\beta_0 \epsilon} - 1} = \frac{\pi^2}{6\beta_0^2},$$

one recovers Hardy-Ramanujan:

$$\Omega(E) \approx \exp\left(\pi\sqrt{\frac{2}{3}}E^{1/2}\right).$$
<sub>53/50</sub>

Auluck, Kothari 1946; Comtet, Leboeuf, Majumdar 2007

$$\sum_{N'=1}^{N} \Omega(E, N') = \tilde{\Omega}(E, N) = \sum_{\{n_k\}} \delta\left(E - \sum_{k=1}^{\infty} n_k \epsilon_k\right) \theta\left(N - \sum_{k=1}^{\infty} n_k\right)$$
$$\mathcal{Z}(\beta, z) = \sum_{E, N} \tilde{\Omega}(E, N) e^{-\beta E} z^N = \prod_{k=1}^{\infty} \frac{1}{1 - z e^{-\beta \epsilon_k}},$$

Saddle point approximation:

$$\tilde{\Omega}(E,N) \approx e^{S(\beta_0,z_0,E,N)},$$

where  $\beta_0, z_0$  maximise

$$S(\beta, z, E, N) = \log \mathcal{Z}(\beta) + \beta E - N \log(z).$$

Saddle points fix

$$E = \sum_{k=1}^{\infty} \frac{\epsilon_k}{z_0^{-1} e^{\beta_0 \epsilon_k} - 1} \approx \int_0^{\infty} \mathrm{d}\epsilon \, \frac{\epsilon}{z_0^{-1} e^{\beta_0 \epsilon} - 1} = \frac{\mathsf{Li}_2(z_0)}{\beta_0^2}$$
$$N = \sum_{k=1}^{\infty} \frac{1}{z_0^{-1} e^{\beta_0 \epsilon_k} - 1} \approx \int_0^{\infty} \mathrm{d}\epsilon \, \frac{1}{z_0^{-1} e^{\beta_0 \epsilon} - 1} = -\frac{\log(1 - z_0)}{\beta_0}.$$

Auluck, Kothari 1946; Comtet, Leboeuf, Majumdar 2007

Take large *N* limit ( $z \uparrow 1$ ) to recover Erdős-Lehner:

$$rac{ ilde{\Omega}(E,N)}{\Omega(E)} \sim \exp\left[-\exp\left(rac{N-b_E}{a_E}
ight)
ight]$$
 $a_E = rac{\sqrt{6}}{\pi}E^{1/2}$ 
 $b_E = rac{1}{\pi}\sqrt{rac{3}{2}}E^{1/2}\log E.$ 

This calculations holds for equally spaced energy levels  $\epsilon_1 = 1, \epsilon_2 = 2, \ldots$ 

What about for  $\rho(\epsilon) = \nu \epsilon^{\nu-1}$ ?

Density of states  $\rho(\epsilon) = \nu \epsilon^{\nu-1}$  for different 1*d* potentials



Density of states  $\rho(\epsilon) = \nu \epsilon^{\nu-1}$  for different 1*d* potentials



Density of states  $\rho(\epsilon) = \nu \epsilon^{\nu-1}$  for different 1*d* potentials



partition into powers > 1

partition into integers

G. H. Hardy, S. Ramanujan Proc. London Math. Soc. **17**, 75 (1918)

C. van Lier, G. E. Uhlenbeck Physica, **4**, 531 (1937) Asymptotics of integer partitions

Ideal quantum gas, saddle point methods

P. Erdős, J. Lehner Duke Math. J. **8**, 335 (1941) Asymptotics of integer partitions, Gumbel

F. C. Auluck, D. S. Kothari Proc. Cambridge Philos. Soc. **42**, 272 (1946) Equally-spaced energy levels, Gumbel

A. Comtet, P. Leboeuf, S. N. Majumdar Phys. Rev. Lett. **98**, 070404 (2007) Power law-spaced energy levels, EVS