On the frequencies of patterns of rises and falls

Jean-Marc Luck

Institut de Physique Théorique, Saclay (CEA & CNRS)

Preprint arXiv:1309.7764

Journées ALEA 2014, March 17–21, 2014, CIRM, Marseille

Outline

- The problem
- Probabilistic approach
- Motivations
- Combinatorial approach

- Periodic patterns: the probabilistic route

  Analytical result for entropy of all families of periodic patterns

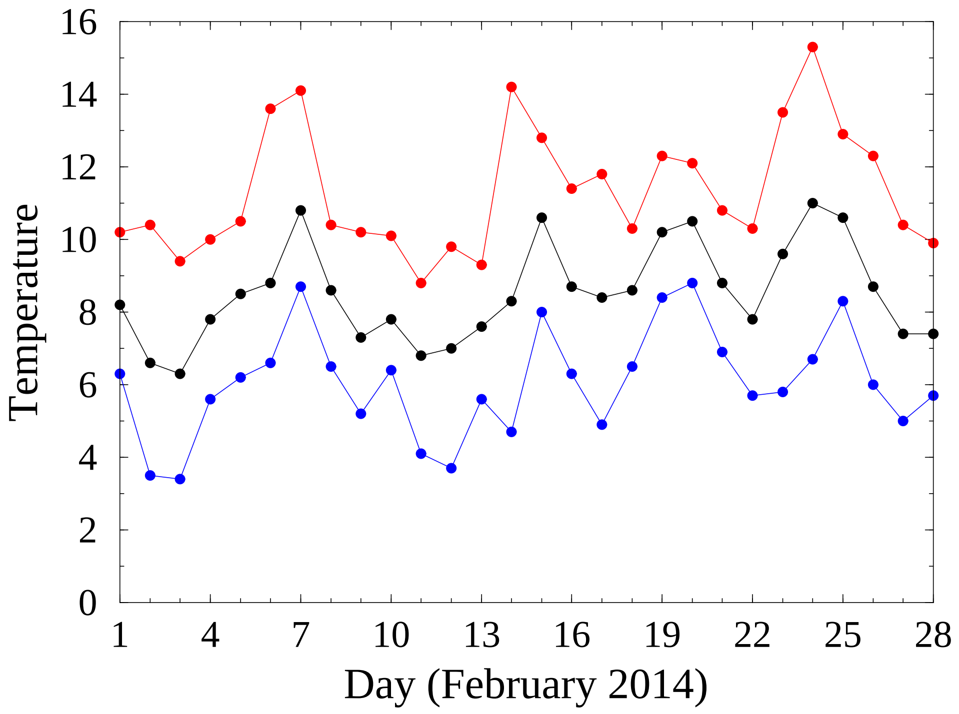
- Random patterns

  Numerical evidence for multifractal behavior

The problem

- Consider a data series

e.g. daily temperature (average, min, max) at Paris Montsouris



- What is the probability of observing a given pattern ?

e.g. 3 consecutive rises in the max series

Probabilistic approach

- Data modelled as *i.i.d. random variables* x_i
- Distribution of the x_i can be taken to be *uniform on* $[0,1]$

If $x_i > x_{i-1}$, there is a rise at the i-th place: $\varepsilon_i = +$
If $x_i < x_{i-1}$, there is a fall at the i-th place: $\varepsilon_i = -$

- What is the probability $P_n(\varepsilon_1, \cdots, \varepsilon_n)$

of observing a given pattern $\varepsilon_1, \cdots, \varepsilon_n$ of n rises and falls ?

Recursive scheme to calculate $P_n(\varepsilon_1, \cdots, \varepsilon_n)$

- Condition on *last variable*

Let $f_n(x)\,\mathrm{d}x = \mathrm{Prob}\{\varepsilon_1, \cdots, \varepsilon_n \quad \text{and} \quad x < x_n < x + \mathrm{d}x\}$

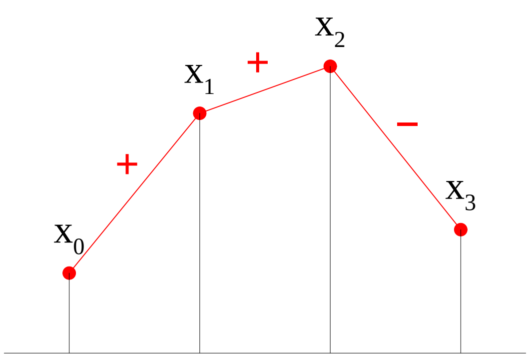
So $P_n(\varepsilon_1, \cdots, \varepsilon_n) = \int_0^1 f_n(x)\,\mathrm{d}x$

- Linear integral recursion relation *(transfer operator)*

If $\varepsilon_n = +$, then $f_n(x) = \int_0^x f_{n-1}(y)\,\mathrm{d}y$

If $\varepsilon_n = -$, then $f_n(x) = \int_x^1 f_{n-1}(y)\,\mathrm{d}y$

Example: *pattern* $++-$



$f_1(x) = x\,, \qquad f_2(x) = \frac{x^2}{2}\,, \qquad f_3(x) = \frac{1-x^3}{6}$

$$P_3(++-) = \frac{1}{8}$$

Motivations

- *Applications*

  Null model to which real data could be compared

  Recent work on microarray data in genetics (Fink et al 2007)

- *Results*

  Alternating patterns yield $P_n \sim (2/\pi)^n$ (André 1879, 1881)

  How generic is exponential law $P_n \sim e^{-\alpha n}$ *?*

  α has physical interpretation of an entropy

  $\alpha_{\min} = \ln \frac{\pi}{2} = 0.451582\cdots$ in spin chain (Derrida & Gardner 1986)

  Can α *be calculated for all (periodic) families of patterns ?*

  How is α *distributed for long pattern chosen at random ?*

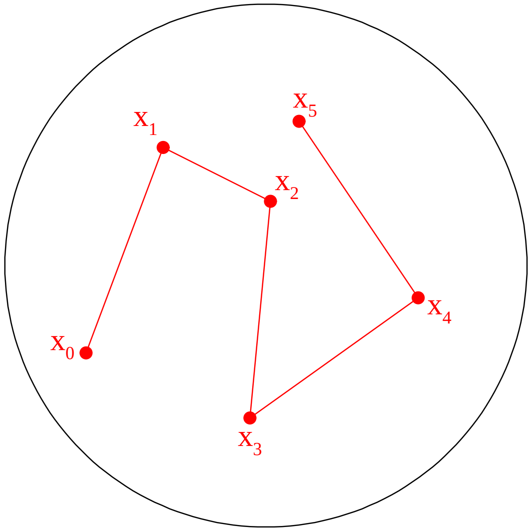
- *Technical*

Reminiscent of calculation of partition function

$$z_n = \left\langle \frac{1}{r_{0,1} r_{1,2} \cdots r_{n-1,n}} \right\rangle$$

of open chain of $n+1$ points in unit 3-dim ball $\quad (r_{i-1,i} = |\mathbf{x}_i - \mathbf{x}_{i-1}|)$

Context: *Multiple scattering of waves* (with D. Boosé and J.Y. Fortin)



$$z_0 = 1\,, \quad z_1 = \frac{6}{5}\,, \quad z_2 = \frac{51}{35}\,, \quad z_3 = \frac{62}{35}\,,$$

$$z_4 = \frac{4\,146}{1925}\,, \quad z_5 = \frac{65\,532}{25\,025}$$

$$Z(x) = \sum_{n \geq 0} z_n x^n = \frac{1}{x} \left(\frac{\tan\sqrt{3x}}{\sqrt{3x}} - 1 \right)$$

$$z_n \sim (12/\pi^2)^n$$

Combinatorial approach

- Data modelled as *uniform random permutation* σ on $n+1$ objects $\{0,1,\cdots,n\}$

  If $\sigma_i > \sigma_{i-1}$, there is a rise at the i-th place: $\varepsilon_i = +$

  If $\sigma_i < \sigma_{i-1}$, there is a fall at the i-th place: $\varepsilon_i = -$

The pattern $\varepsilon_1,\cdots,\varepsilon_n$ is the *up-down signature* of σ

(André 1879, 1881; MacMahon 1915, De Bruijn 1970, Viennot 1979 ...)

- The probability reads $P_n(\varepsilon_1,\cdots,\varepsilon_n) = \dfrac{A_n(\varepsilon_1,\cdots,\varepsilon_n)}{(n+1)!}$

  where $A_n(\varepsilon_1,\cdots,\varepsilon_n)$ is the number of permutations

  whose up-down signature is $\varepsilon_1,\cdots,\varepsilon_n$

Recursive scheme to calculate $A_n(\varepsilon_1, \cdots, \varepsilon_n)$

- Condition again on *last variable*

Let $a_{n,j}$ be the number of permutations

whose signature is $\varepsilon_1, \cdots, \varepsilon_n$ and such that $\sigma_n = j$

So $A_n(\varepsilon_1, \cdots, \varepsilon_n) = \sum_{j=0}^{n} a_{n,j}$

- Linear recursion relation (De Bruijn 1970, Viennot 1979, Atkinson 1985 ...)

If $\varepsilon_n = +$, then $\begin{cases} a_{n,0} = 0, \\ a_{n,j} = a_{n,j-1} + a_{n-1,j-1} \qquad (j = \overrightarrow{1, \cdots, n}), \end{cases}$

If $\varepsilon_n = +$, then $\begin{cases} a_{n,n} = 0, \\ a_{n,j} = a_{n,j+1} + a_{n-1,j} \qquad (j = \overleftarrow{0, \cdots, n-1}), \end{cases}$

This is a generalization of the *boustrophedon algorithm*

Alternating permutations and *boustrophedon algorithm*

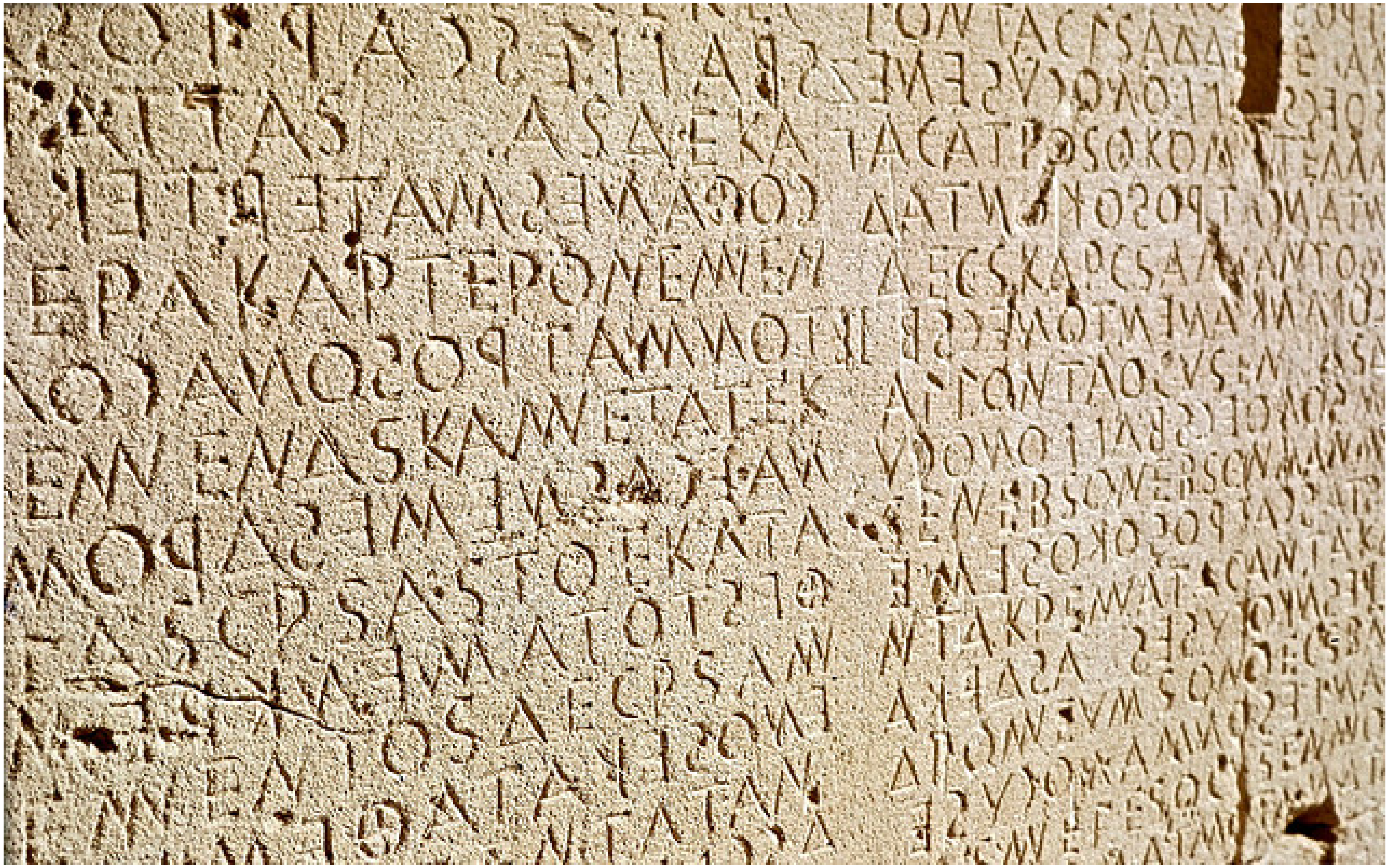
Alternating permutations $(\varepsilon_n = + - + - + - \cdots)$

$$\begin{array}{c}
0 \\
0 \rightarrow 1 \\
1 \leftarrow 1 \leftarrow 0 \\
0 \rightarrow 1 \rightarrow 2 \rightarrow 2 \\
5 \leftarrow 5 \leftarrow 4 \leftarrow 2 \leftarrow 0 \\
0 \rightarrow 5 \rightarrow 10 \rightarrow 14 \rightarrow 16 \rightarrow 16 \\
61 \leftarrow 61 \leftarrow 56 \leftarrow 46 \leftarrow 32 \leftarrow 16 \leftarrow 0
\end{array}$$

Word *boustrophedon* *("turning ox")* introduced in this context by Millar et al (1996)

Construction attributed to Seidel (1877)

Ancient boustrophedonic inscription



Gortyne Island, near Crete (5th century BC)

Explicit correspondence between both approaches

$$P_n = \frac{A_n}{(n+1)!}, \qquad P_n = \int_0^1 f_n(x)\,\mathrm{d}x, \qquad A_n = \sum_{j=0}^{n} a_{n,j}$$

- Probabilistic and combinatorial approaches complementary
- Both $f_n(x)$ and $a_{n,j}$ obey linear recursion relations
- Explicit correspondence given by

$$f_n(x) = \sum_{j=0}^{n} a_{n,j}\,\frac{x^j(1-x)^{n-j}}{j!(n-j)!}$$

The most and least probable patterns

Among the 2^n patterns $\varepsilon_1, \cdots, \varepsilon_n$ of length n

- Two most probable patterns: the alternating ones *(see below)*

  $+ - + - + - \cdots$ and $- + - + - + \cdots$

  $P_n \sim (2/\pi)^n$, $\alpha = \alpha_{\min} = \ln \frac{\pi}{2}$ (André 1879, 1881)

- Two least probable patterns: the steady ones

  i.e., the rising one $+ + + \cdots$ and the falling one $- - - \cdots$

Both routes for rising patterns

Probabilistic: $f_n(x) = \frac{x^n}{n!}$, $P_n = \frac{1}{(n+1)!}$

Combinatorial: $\sigma = I$, $a_{n,j} = \delta_{nj}$, $A_n = 1$, $P_n = \frac{1}{(n+1)!}$

$\alpha_n \approx \ln n$

Periodic patterns: the probabilistic route

Alternating patterns $(\varepsilon_n = + - + - + - \cdots)$

- Recursion relations

$$f_{2k+1}(x) = \int_0^x f_{2k}(y)\,\mathrm{d}y, \qquad f_{2k}(x) = \int_x^1 f_{2k-1}(y)\,\mathrm{d}y$$

- Generating series

$$F_0(z,x) = \sum_{k\geq 0} f_{2k}(x) z^{2k}, \qquad F_1(z,x) = \sum_{k\geq 0} f_{2k+1}(x) z^{2k+1}$$

- Integral equations

$$F_0(z,x) = 1 + z\int_x^1 F_1(z,y)\,\mathrm{d}y, \qquad F_1(z,x) = z\int_0^x F_0(z,y)\,\mathrm{d}y$$

- Differential equation

$$\frac{\partial^2 F_0}{\partial x^2} = -z^2 F_0, \qquad F_0(z,1) = 1, \qquad \frac{\partial F_0(z,0)}{\partial x} = 0$$

- Solution

$$F_0(z,x) = \frac{\cos zx}{\cos z}, \qquad F_1(z,x) = \frac{\sin zx}{\cos z}$$

- Generating series for the P_n

$$\Pi(z) = \sum_{n\geq 0} P_n z^n = \frac{1}{z}\left(F_1(z,1) + F_0(z,0) - 1\right)$$

- Result

$$\Pi(z) = \frac{\sin z + 1 - \cos z}{z \cos z} = \frac{\tan z + \sec z - 1}{z}$$

Recover thus pioneering results by André (1879, 1881)

$$\begin{aligned}\tan z &= \sum_{k\geq 0} P_{2k} z^{2k+1} = \sum_{k\geq 0} A_{2k} \frac{z^{2k+1}}{(2k+1)!} \\ \sec z &= 1 + \sum_{k\geq 0} P_{2k+1} z^{2k+2} = 1 + \sum_{k\geq 0} A_{2k+1} \frac{z^{2k+2}}{(2k+2)!}\end{aligned}$$

- The A_n are called Euler-Bernoulli numbers or Entringer numbers
- Asymptotic behavior

$$P_n \approx 2 \left(\frac{2}{\pi}\right)^{n+2}$$

- Connection with multiple-scattering problem

$$z_n = 3^{n+1} P_{2n+2}$$

p-alternating patterns: period $p \geq 2$ *ending with a single fall*

Example: for $p = 3$, $\varepsilon_n = + + - + + - + + - \cdots$

- Generating series

$$F_q(z,x) = \sum_{k \geq 0} f_{kp+q}(x) z^{kp+q} \qquad (q = 0, \cdots, p-1)$$

- Integral equations

$$F_0(z,x) = 1 + z \int_x^1 F_1(z,y)\, \mathrm{d}y, \qquad F_q(z,x) = z \int_0^x F_{q-1}(z,y)\, \mathrm{d}y \qquad (q \neq 0)$$

- Differential equation

$$\frac{\partial^p F_0}{\partial x^p} = -z^p F_0, \qquad F_0(z,1) = 1, \qquad \frac{\partial^q F_0(z,0)}{\partial x^q} = 0 \qquad (q \neq 0)$$

- Solution

$$F_q(z,x) = \frac{T_{p,q}(zx)}{T_{p,0}(z)}$$

- Result

$$\Pi(z) = \frac{1}{zT_{p,0}(z)}\left(\sum_{q=1}^{p-1} T_{p,q}(z) + 1 - T_{p,0}(z)\right)$$

- Probabilities

$$P_n \approx \mathcal{A}_n \, \mathrm{e}^{-\alpha n}$$

Smallest real positive zero z_0 of $T_{p,0}$

Entropy $\alpha = \ln z_0$

Other zeros at $z_q = z_0 \zeta^q$ for $q = 1, \cdots, p-1$

Amplitudes $\mathcal{A}_n$ periodic with period p

Generalized hyperbolic and trigonometric functions

$p \geq 2\,, \qquad q = 0, \cdots, p-1\,, \qquad \zeta = \mathrm{e}^{2\pi \mathrm{i}/p}$

$$H_{p,q}(z) = \sum_{k\geq 0} \frac{z^{kp+q}}{(kp+q)!} = \frac{1}{p}\sum_{j=0}^{p-1} \zeta^{-qj} \mathrm{e}^{\zeta^j z}$$

$$H'_{p,q} = H_{p,q-1} \quad (q \neq 0), \qquad H'_{p,0} = H_{p,p-1}$$

$$T_{p,q}(z) = \sum_{k\geq 0} (-1)^k \frac{z^{kp+q}}{(kp+q)!} = \frac{1}{p}\sum_{j=0}^{p-1} \zeta^{-q(j+1/2)} \mathrm{e}^{\zeta^{j+1/2} z}$$

$$T'_{p,q} = T_{p,q-1} \quad (q \neq 0), \qquad T'_{p,0} = -T_{p,p-1}$$

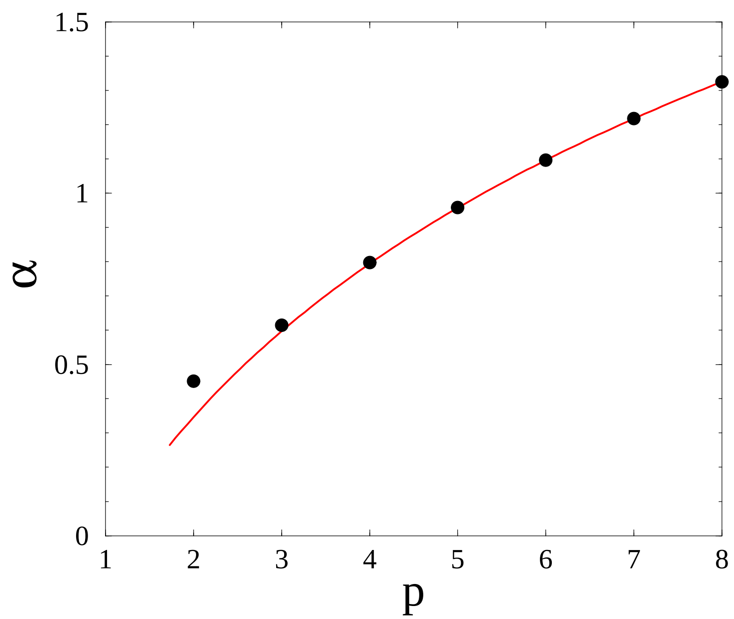
Recover thus Mendes and Remmel's (book preprint) results ... by elementary means

- $p = 2$ yields $z_0 = \frac{\pi}{2}$

$$\alpha = \alpha_{\min} = \ln \tfrac{\pi}{2} = 0.451582\cdots$$

- $p \gg 1$ yields $z_0 \approx (p!)^{1/p}$

$$\alpha \approx \tfrac{\ln p!}{p} \approx \ln p - 1$$



General case

Period $p \geq 2$

$p - \nu$ rises and ν falls per period $(1 \leq \nu \leq p-1)$, end with a fall

- Differential equation

$$\frac{\partial^p F_0}{\partial x^p} = (-1)^\nu z^p F_0$$

If $\varepsilon_{p-q} = +$, then $\partial^q F_0/\partial x^q(z,0) = 0$

If $\varepsilon_{p-q} = -$, then $\partial^q F_0/\partial x^q(z,1) = \delta_{q0}$

- Solution

$$F_0(z,x) = \sum_q C_q(z) H_{p,q}(zx) \quad \text{or} \quad \sum_q C_q(z) T_{p,q}(zx)$$

Sum over the ν indices q such that $\varepsilon_{p-q} = -$

Boundary conditions at $x = 1$ yield ν linear equations for the $C_q(z)$

- Result

$$C_q(z) = \frac{\cdots}{\Delta(z)}, \qquad \Pi(z) = \frac{\cdots}{\Delta(z)}$$

$\Delta(z)$ is $\nu \times \nu$ determinant

Entries are generalized hyperbolic or trigonometric functions

$\Delta(z)$ is entire function of z^p

- Probabilities

$$P_n \approx \mathcal{A}_n \, \mathrm{e}^{-\alpha n}$$

Smallest real positive zero z_0 of $\Delta(z)$

Entropy $\alpha = \ln z_0$

Other zeros at $z_q = z_0 \zeta^q$ for $q = 1, \cdots, p-1$

Amplitudes $\mathcal{A}_n$ periodic with period p

More explicitly

- Two falls at distances a and b $(p = a + b)$

$$\Delta(z) = \begin{vmatrix} H_{p,0}(z) & H_{p,b}(z) \\ H_{p,a}(z) & H_{p,0}(z) \end{vmatrix}$$

- Three falls at distances a , b and c $(p = a + b + c)$ [f]

$$\Delta(z) = \begin{vmatrix} T_{p,0}(z) & T_{p,c}(z) & T_{p,b+c}(z) \\ -T_{p,a+b}(z) & T_{p,0}(z) & T_{p,b}(z) \\ -T_{p,a}(z) & -T_{p,a+c}(z) & T_{p,0}(z) \end{vmatrix}$$

Duality $\nu \leftrightarrow p - \nu$ yields infinite sequence of non-linear identities

$$T_{3,0} = H_{3,0}^2 - H_{3,1} H_{3,2}$$

[f] *This is the correct form of erroneous equation (8.14) or (69) in preprint*

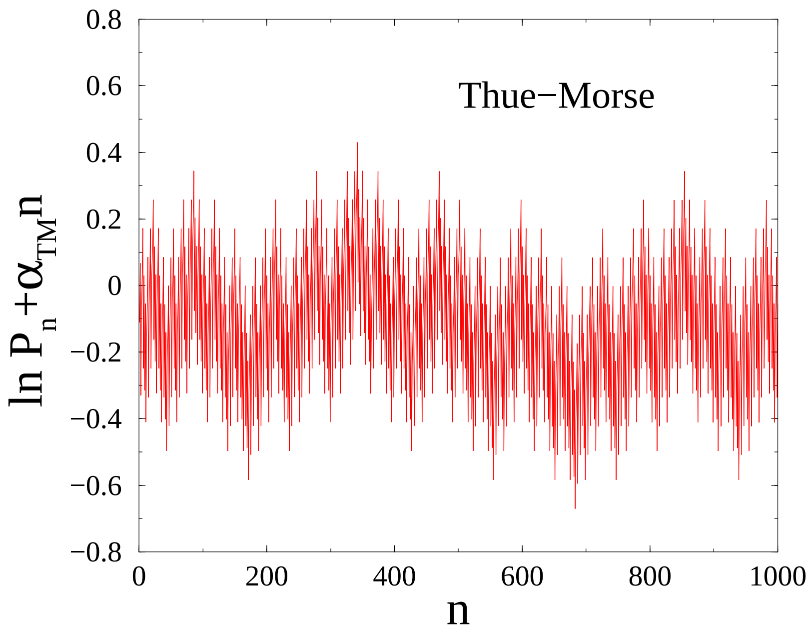
Aperiodic families of patterns

Numerical evidence for generic exponential behavior $P_n \sim e^{-\alpha n}$

i.e., non-trivial entropy α

Example: *Thue-Morse sequence* $ABBABAABBAABABBA\cdots$

Generated by substitution $S_{TM}: \begin{cases} A \rightarrow AB \\ B \rightarrow BA \end{cases}$



$\alpha_{TM} = 0.583018\cdots$

$\alpha_{Fib} = 0.562168\cdots$

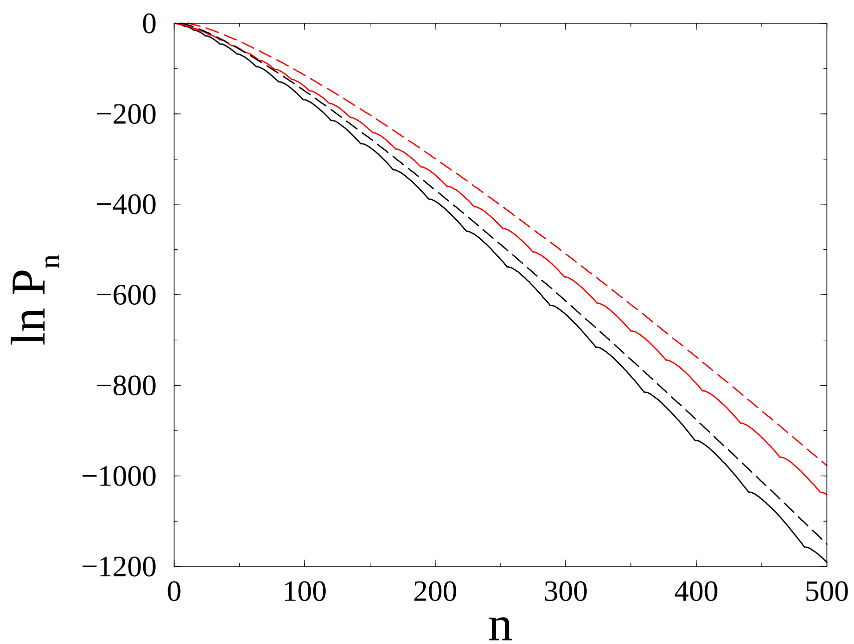
$\alpha_{RS} = 0.780693\cdots$

Chirping patterns

Patterns where rises or falls become *more and more seldom*

Examples:

- *Square chirp:* fall at place n iff $n = k^2$
- *Triangular chirp:* fall at place n iff $n = \frac{k(k+1)}{2}$



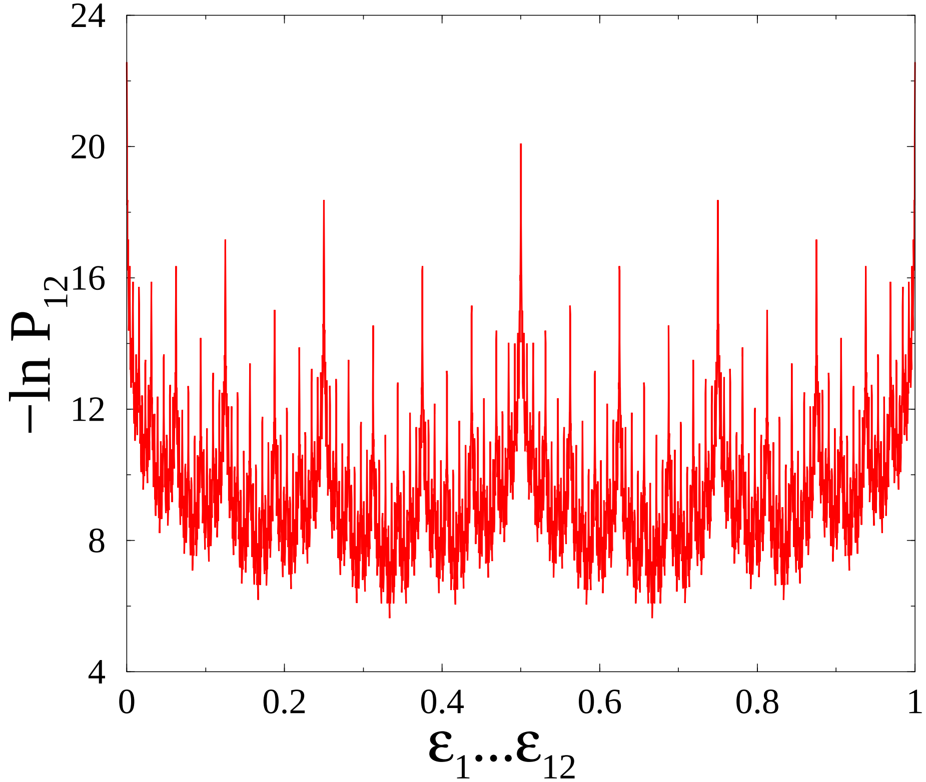
$$\ln P_n \approx -\frac{n \ln n}{2}, \quad \text{i.e.,} \quad \alpha_n \approx \frac{\ln n}{2}$$

can do $\alpha_n \approx \theta \ln n$ for any $0 \leq \theta \leq 1$

Random patterns

Consider all the 2^n patterns $\varepsilon_1, \cdots, \varepsilon_n$ of n rises and falls

- *How large is effective* $\alpha_n = -\frac{1}{n} \ln P_n(\varepsilon_1, \cdots, \varepsilon_n)$ *?*
- *Have a look at the* 4096 *patterns of length* $n = 12$



Multifractal formalism (I)

- For $q = 1, 2, \cdots$

$$Z_n(q) = \sum_{\varepsilon_1, \cdots, \varepsilon_n} P_n(\varepsilon_1, \cdots, \varepsilon_n)^q$$

is the probability that q independent uniform random permutations on $n+1$ objects have the same up-down signature

- Mallows & Shepp (1985) have proved large-deviation result

$$Z_n(q) \sim 2^{-n\tau(q)}$$

and calculated $\tau(2)$

- Interpretation: *Generalized (Rényi) dimensions* $D(q)$

$$\tau(q) = (q-1)D(q)$$

Multifractal formalism (II)

- For fixed α and $\delta \ll \alpha$, define

$$\mathcal{N}(\alpha,\delta) = \left\{ \varepsilon_1,\cdots,\varepsilon_n \;\middle|\; n\alpha < -\ln P_n(\varepsilon_1,\cdots,\varepsilon_n) < n(\alpha+\delta) \right\}$$

- Multifractal hypothesis

$$\dim \mathcal{N}(\alpha,\delta) = f(\alpha), \qquad \text{i.e.,} \qquad |\mathcal{N}(\alpha,\delta)| \sim 2^{nf(\alpha)}$$

- Correspondence between $\tau(q)$ and $f(\alpha)$

$$Z_n(q) \sim \int_0^\infty \mathrm{e}^{-qn\alpha}\, 2^{nf(\alpha)}\, \mathrm{d}\alpha \sim 2^{-n\tau(q)}$$

Legendre transform: $\qquad \tau(q) = \min_\alpha \left(\frac{q\alpha}{\ln 2} - f(\alpha) \right)$

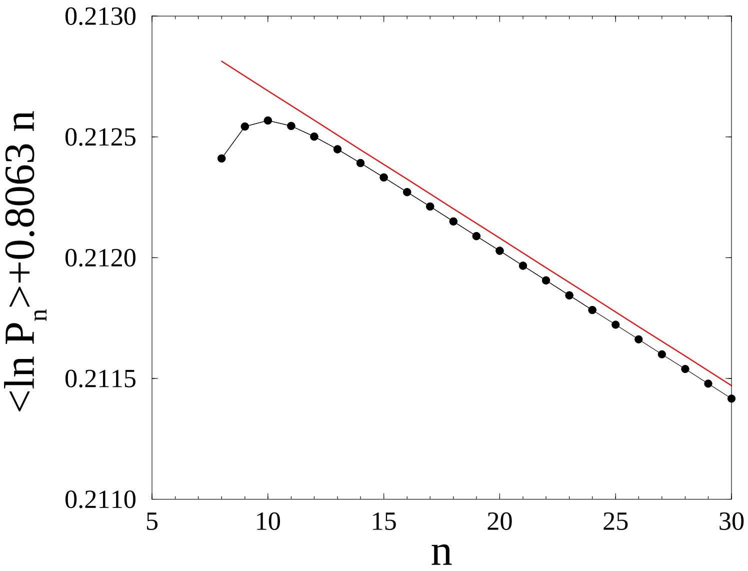
Evidence for full multifractal behavior

Full multifractal behavior means bilateral differential Legendre transform

$$\tau(q) + f(\alpha) = \frac{q\alpha}{\ln 2}, \qquad q = \ln 2\, f'(\alpha), \qquad \alpha = \ln 2\, \tau'(q)$$

Define the average $\langle X(\varepsilon_1, \cdots, \varepsilon_n) \rangle = \frac{1}{2^n} \sum_{\varepsilon_1, \cdots, \varepsilon_n} X(\varepsilon_1, \cdots, \varepsilon_n)$

- Typical behavior of P_n



$\langle \ln P_n \rangle \approx -\alpha_0 n$

$\alpha_0 = \ln 2\, \tau'(0) = 0.80636111 \cdots$

α_0 is Lyapunov exponent

Notice remarkable accuracy

- Similarly

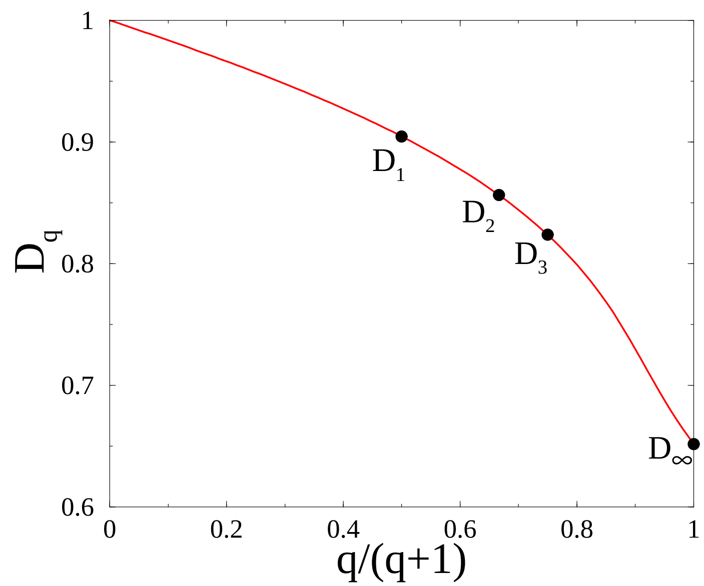
$$\left\langle (\ln P_n)^2 \right\rangle - \langle \ln P_n \rangle^2 \approx w_0 n$$

$$w_0 = -\ln 2\, \tau''(0) = 0.435600\cdots$$

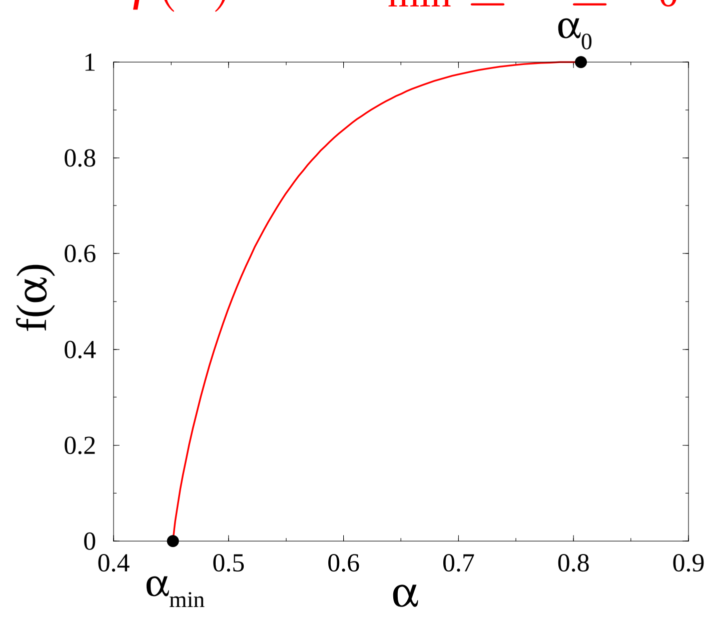
all the cumulants of $\ln P_n$ *extensive* (grow linearly in the pattern size n)

- Left half a multifractal spectrum:

$\tau(q) = (q-1)D(q)$ for $q \geq 0$



$f(\alpha)$ for $\alpha_{\min} \leq \alpha \leq \alpha_0$



Outline

- The problem
- Probabilistic approach
- Motivations
- Combinatorial approach
- Periodic patterns: the probabilistic route

  Analytical result for entropy of all families of periodic patterns
- Random patterns

  Numerical evidence for multifractal behavior

Preprint arXiv:1309.7764