# Matrix-correlated random variables: A statistical physics and signal processing duet

Florian Angeletti

NITheP National Institute for Theoretical Physics

Work in collaboration with Hugo Touchette, Patrice Abry and Eric Bertin.

13 January 2015

# Presentation

- Thesis:
  "Sums and extremes in statistical physics and signal processing"
  Advisors: Eric Bertin and Patrice Abry, Physics laboratory of ENS Lyon.
- Postdoc
  NITheP, Stellenbosch, South Africa, working with Hugo Touchette on Large deviation theory
- Themes:
  - Application of statistical physics to signal processing
  - Extreme statistics
  - Random vectors with matrix representation
  - Large deviation functions

# Out-of-equilibrium statistical physics

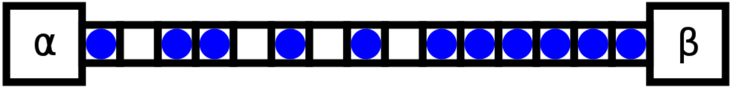
## At equilibrium:

- Microcanonical ensemble: $p(x_1, \ldots, x_n) = \mathrm{cst}$
- Canonical ensemble: $p(x_1, \ldots, x_n) = e^{-\beta \mathcal{H}(x_1, \ldots, x_n)}$

## Out-of-equilibrium:

- Constant flow of heat or particles
- Dynamic description
- Stationary distribution?

# ASEP




A simple and iconic out-of-equilibrium systems

- Asymmetric: Particles only move from left to right
- Exclusion : One particle by site
- Creation rate $\alpha$
- Destruction rate $\beta$

# Matrix-correlated random variable

How do we describe the stationary solution ?

- Matrix product ansatz (Derrida and al 1993)

$$p(x_1, \ldots, x_n) = \frac{\langle W|\, R(x_1) \ldots R(x_n)\, |V\rangle}{\langle W|\, (R(0) + R(1))^n \, |V\rangle}$$

- matrix $R(x)$
- Long range correlation
- Similar solution for 1D diffusion-reaction system
- Formal similarity with DMRG

# Objectives

## Mathematical model

$$p(x_1, \ldots, x_n) \approx \mathcal{R}(x_1) \cdots \mathcal{R}(x_n)$$

Study the statistical properties of theses models

- Hidden Markov model representation
- Signal processing application
- Topology induces correlation

- Large deviation functions
- Limit distributions for the sums
- Limit distributions for the extremes

Then go back to physical models

# Matrix representation

$$p(x_1, \ldots, x_n) = \frac{\mathcal{L}\left(\mathcal{R}(x_1) \ldots \mathcal{R}(x_n)\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

- linear form $\mathcal{L}$: $\mathcal{L}(M) = \mathrm{tr}(\mathcal{A}^T M)$
  - $\mathcal{A}$: $d \times d$ positive matrix
- $\mathcal{R}(x)$: $d \times d$ positive matrix function
  - structure matrix

$$\mathcal{E} = \int_{\mathbb{R}} \mathcal{R}(x) dx$$

  - probability density function matrix

$$\mathcal{R}_{i,j}(x) = \mathcal{E}_{i,j} \mathcal{P}_{i,j}(x)$$

- $d > 1$: Non-commutativity $\Longrightarrow$ Correlation

# Correlation

- Product structure: $p(x_1, \ldots, x_n) = \frac{\mathcal{L}(\mathcal{R}(x_1)\ldots\mathcal{R}(x_n))}{\mathcal{L}(\mathcal{E}^n)}$
- Moment matrix: $\mathcal{Q}(q) = \int x^q R(x) dx$

$$\langle X_k^p \rangle = \frac{\mathcal{L}\left(\mathcal{E}^{k-1}\mathcal{Q}(p)\,\mathcal{E}^{n-k}\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

$$\langle X_k X_l \rangle = \frac{\mathcal{L}\left(\mathcal{E}^{k-1}\mathcal{Q}(1)\,\mathcal{E}^{l-k-1}\mathcal{Q}(1)\,\mathcal{E}^{n-l}\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

$$\langle X_k X_l X_m \rangle = \frac{\mathcal{L}\left(\mathcal{E}^{k-1}\mathcal{Q}(1)\,\mathcal{E}^{l-k-1}\mathcal{Q}(1)\,\mathcal{E}^{m-l-1}\mathcal{Q}(1)\,\mathcal{E}^{n-m}\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

$$\cdots$$

# Stationarity

Translation invariance:

$$p(X_{k_1} = x_1, \dots, X_{k_l} = x_l) = p(X_{c+k_1} = x_1, \dots, X_{c+k_l} = x_l)$$

## Sufficient condition

$$[\mathcal{A}^T, \mathcal{E}] = \mathcal{A}^T\mathcal{E} - \mathcal{E}\mathcal{A}^T = 0$$

$$\forall M, \quad \mathcal{L}(M\mathcal{E}) = \mathcal{L}(\mathcal{E}M)$$

- $p(X_k = x) = \dfrac{\mathcal{L}\left(\mathcal{R}(x)\mathcal{E}^{n-1}\right)}{\mathcal{L}(\mathcal{E}^n)}$
- $p(X_k = x, X_l = y) = \dfrac{\mathcal{L}\left(\mathcal{R}(x)\mathcal{E}^{l-k-1}\mathcal{R}(y)\mathcal{E}^{n-|l-k|-1}\right)}{\mathcal{L}(\mathcal{E}^n)}$
- $p(X_k = x, X_l = y, X_m = z) = \dfrac{\mathcal{L}\left(\mathcal{R}(x)\mathcal{E}^{l-k-1}\mathcal{R}(y)\mathcal{E}^{m-l-1}\mathcal{R}(z)\mathcal{E}^{n-|m-k|-1}\right)}{\mathcal{L}(\mathcal{E}^n)}$

# Numerical generation

$$p(x_1, \ldots, x_n) = \frac{\mathcal{L}\left(\mathcal{R}(x_1)\ldots\mathcal{R}(x_n)\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

How do we generate a random vector $X$ for a given triple $(\mathcal{A}, \mathcal{E}, \mathcal{P})$?

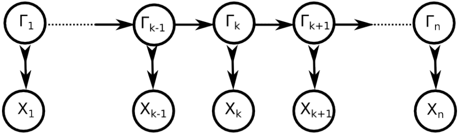
- Expand the matrix product

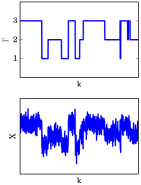
$$p(x_1, \ldots, x_n) = \frac{1}{\mathcal{L}\left(\mathcal{E}^n\right)} \sum_{\Gamma \in \{1,\ldots,d\}^{n+1}} \mathcal{A}_{\Gamma_1,\Gamma_{n+1}} \mathcal{E}_{\Gamma_1,\Gamma_2} \mathcal{P}_{\Gamma_1,\Gamma_2}(x_1) \ldots \mathcal{E}_{\Gamma_n,\Gamma_{n+}}$$

$$p(x_1, \ldots, x_n) = \sum_{\Gamma} P(\Gamma) P(\underline{X}|\Gamma)$$

- $\Gamma$, hidden Markov chain

# Hidden Markov Model




- Markov chain $\Gamma$
- Observable $\underline{X} = X_1, \ldots, X_k$
- $(X_k|\Gamma_k)$ is distributed according to the pdf $p(X_k|\Gamma_k)$

# Hidden Markov Chain representation

**Hidden Markov Chain**

$$p(\Gamma) = \frac{\mathcal{A}_{\Gamma_1,\Gamma_{n+1}}}{\mathcal{L}\left(\mathcal{E}^n\right)} \prod_k \mathcal{E}_{\Gamma_k,\Gamma_{k+1}}$$

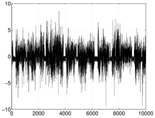
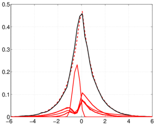
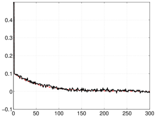
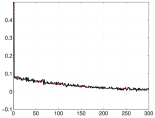
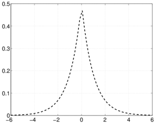
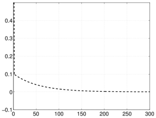
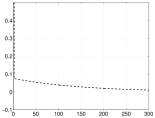
**Conditional pdf $(X|\Gamma)$**

$$p(X_k = x|\Gamma) = \mathcal{P}_{\Gamma_k,\Gamma_{k+1}}(x)$$

- $\mathcal{E}$ non-stochastic $\implies$ Non-homogeneous markov chain
- Specific non-homogeneous Hidden Markov model:
  - Hidden Markov Model $\not\Rightarrow$ Matrix representation

# Stationary time series design

- Generation of random vector with prescribed correlation and marginal distribution:
  - Matrix representation: Choice of $(\mathcal{A}, \mathcal{E}, \mathcal{P})$
  - Hidden Markov Model: Numerical generation
- Higher-order dependency structure: correlation of squares

| Realization | Marginal | Correlation | Square corr. |
|---|---|---|---|
|  |  |  |  |
| Prescribed |  |  |  |

# Dual representation

## Matrix representation

- Algebraic properties
- Statistical properties computation

## Hidden Markov Model

- 2-layer model: correlated layer + independant layer
- Efficient synthesis

# Correlation and Jordan decomposition

$$\langle X_k X_l \rangle = \frac{\mathcal{L}\left(\mathcal{E}^{k-1}\mathcal{Q}(1)\,\mathcal{E}^{l-k-1}\mathcal{Q}(1)\,\mathcal{E}^{n-l}\right)}{\mathcal{L}\left(\mathcal{E}^{n}\right)}$$

- The dependency structure of $X$ depends on the behavior of $\mathcal{E}^n$
- $\lambda_k$ eigenvalues of $\mathcal{E}$ ordered by their real parts $\Re(\lambda_1)\Re(\lambda_2) > \cdots > \lambda_r$
- $J_{k,l}$ Jordan block associated with eigeivalue $\lambda_k$

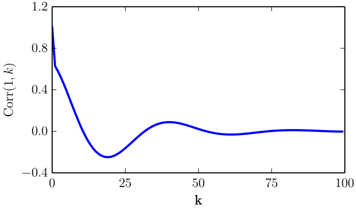
$$\mathcal{E} = B^{-1}\begin{pmatrix} J_{1,1} & 0 & 0 \\ & \ddots & \\ 0 & & J_{k,l} \end{pmatrix} B, \quad J_{k,l} = \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_k \end{pmatrix}$$

# Dependency structure

## Case 1: Short-range correlation

- $\lambda_2$ exists:

$$\langle X_k X_l \rangle - \langle X_k \rangle \langle X_l \rangle \approx \sum_{m>1} \alpha_m \frac{\lambda_m}{\lambda_1}^{|k-l|}$$

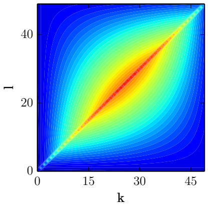



## Case 2: Constant correlation

- More than one block $J_{1,k}$: Constant correlation term

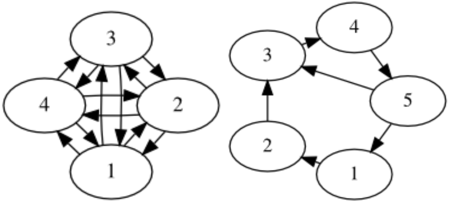
## Case 3: Long-range correlation

- $J_{1,k}$ with size $p > 1$:
  - $\langle X_k X_l \rangle - \langle X_k \rangle \langle X_l \rangle \approx P(\frac{k}{n}, \frac{k-l}{n}, \frac{l}{n}), \quad P \in \mathbb{R}[X, Y, Z]$



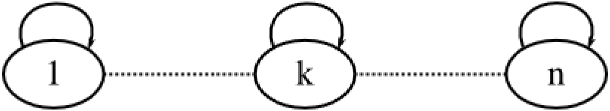
# Short-range correlation: Ergodic chain

- $\mathcal{E}$ irreducible $\iff$ $\Gamma$ ergodic Irreducible matrix $\mathcal{E}$ $\iff$ $G(\mathcal{E})$ is strongly connected
- Short-range correlation



# Constant correlation: Identity $\mathcal{E}$

- Disconnected components:

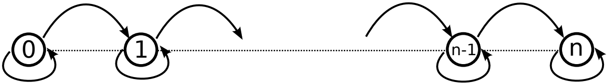



$$\mathcal{E} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

- The chain $\Gamma$ is trapped inside its starting state
- Constant correlation:
  - $\langle X_k X_l \rangle - \langle X_k \rangle \langle X_l \rangle = \frac{\mathcal{L}\left(\mathcal{Q}(1)^2\right) - \mathcal{L}(\mathcal{Q}(1))^2}{\mathcal{L}(\mathcal{E})}$

# Long-range correlation: Linear irreducible $\mathcal{E}$

- Irreversible transitions:



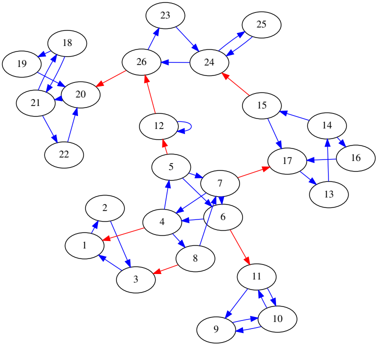

$$\mathcal{E} = \begin{pmatrix} 1 & \epsilon & 0 & \\ & \ddots & \ddots & \\ 0 & & 1 & \end{pmatrix}$$

- The chain $\Gamma$ can only stay in its current state or jump to the next
- All chains with a non-zero probability and the same starting and ending points are equiprobable
- Polynomial correlation:

$$\langle X_k X_l \rangle \approx \sum_{r+s+t=d-1} c_{r,s,t} \left(\frac{k}{n}\right)^r \left(\frac{l-k}{n}\right)^s \left(1-\frac{l}{n}\right)^t$$

# General shape of $\mathcal{E}$

$$\mathcal{E} = \begin{pmatrix} I_1 & * & T_{k,l} \\ & \ddots & * \\ 0 & & I_r \end{pmatrix}$$




- Irreducible blocks $I_k$
- Irreversibles transitions $T_{k,l}$
- Correlation: Mixture of short-range, constant and long-range correlations

# Summary

- Short-range correlation $\Longrightarrow$ Strongly connected component of size $s > 1$
- More than one weakly connected component $\Longrightarrow$ constant correlation
- Polynomial correlation $\Longrightarrow$ More than one strongly connected component

Necessary but non sufficient conditions

# Random vector sum

## Sum

$$S(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

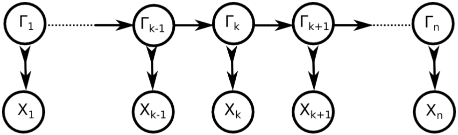
Correlated random variables

- Law of large numbers?
- Central limit theorem?
- Large deviations?

Two paths:

- Hidden Markov chain representation
- Matrix representation

# Hidden Markov path




- $S(\underline{X}|\Gamma)$: sum of sums of i.i.d. random variables:

$$S(\underline{X}|\Gamma) = \sum_{i,j} \sum_{k=1}^{n\nu_{i,j}} (X_k|i,j) \equiv S(\underline{X}|\underline{\nu})$$

- $\nu_{i,j}$ fraction of $(i \to j)$-transition:

$$\underline{\nu} = \left( \frac{\mathrm{card}\{k/\Gamma_k = i, \Gamma_{k+1} = j\}}{n} \right)_{i,j}$$

- Standard convergence theorem (law of large numbers or central limit theorem )

# Layer combination

$$p(S(\underline{X}) = s) = \sum_{\underline{\nu}} p(\underline{\nu}) p(S(\underline{X}|\underline{\nu}) = s)$$

**Limit distribution for $S(\underline{X}|\underline{\nu})$**

$$p(S(\underline{X}|\underline{\nu}) = s)$$

$+$

**Limit distribution for $\underline{\nu}$**

$$p(\underline{\nu})$$

$\Downarrow$

**Limit distribution for $S(\underline{X})$**

$$p(S(\underline{X}) = s)$$

# Distribution of $\nu$

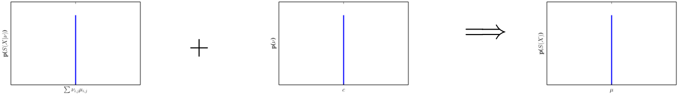
- How $\underline{\nu}$ is distributed?
- Difficulty: $\Gamma$ non-homogeneous Markov chain.

Three important subclasses:

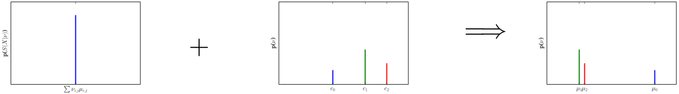
- Irreducible $\mathcal{E}$ (short-range correlation)
  - $\nu$ converges towards a dirac distribution
- Identity $\mathcal{E}$ (constant correlation)
  - $\nu$ converges towards a discrete mixture of dirac distributions
- Linear irreversible $\mathcal{E}$ (long-range polynomial correlation)
  - $\nu$ converges towards a uniform distribution on a $d$-simplex

# Limits laws for core examples

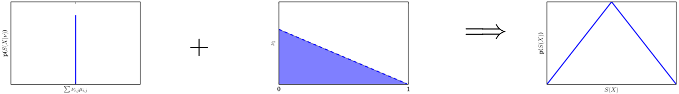
- Irreducible $\mathcal{E}$ (short-range correlations):
  - Standard limit laws



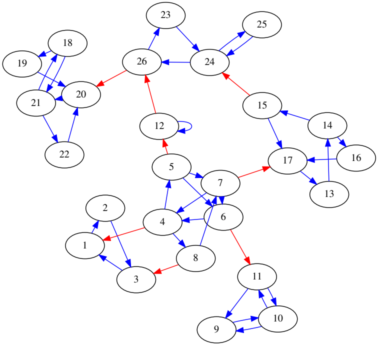
- Identity $\mathcal{E}$ (constant correlation):
  - Discrete mixture of standard limit laws:



- Linear irreversible $\mathcal{E}$ (long-range correlation):
  - Continuous mixture of limit laws



# General case




- Combinations of three core behaviors
  - Irreducible blocks: Fast convergence to the stationary state : dirac distribution
  - Separated componentes : discrete mixture
  - Irreversible transitions: continuous mixture
- Limit laws :
  - Discrete mixture of continuous mixture of standard limits distributions

# Large deviation function

## Sum

$$S(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Law of large number: existence of a concentration point
- Central limit theorem: fluctuations around this concentration point
- Large deviation: fluctuations far away of the concentrations points

## Large deviation principle

$$P(S(\underline{X})/n = s) \approx e^{-nI(s)}$$

Do we have a large principle?

# Gärtner-Ellis theorem

- Generating function

$$g_n(w) = \left\langle e^{wS_n} \right\rangle$$

- Scaled cumulant generating function $\Phi(w)$

$$\Phi(w) = \lim_{n \to \infty} \frac{\ln g_n(w)}{n}$$

**Gärtner-Ellis Theorem**

If $\Phi(w)$ exists and is differentiable, $I(s)$ exists and is

$$I(s) = \sup_{w} \{ws - \Phi(w)\}$$

# Gärtner-Ellis theorem for i.i.d. random variables

- $\Phi(w) = g(w)$
- $g(w)$ cumulant generating function

**Large deviation principle:**

$$I(s) = \sup_{w}\{ws - g(w)\}$$

# Gärtner-Ellis theorem for matrix-correlated random variables

- matrix generating function:

$$\mathcal{G}(w) = \int_{\mathbb{R}} \mathcal{R}(x) e^{-wx} dx$$

- $\Phi(w) = \ln \lambda_1(w)$
- $\lambda_1(w)$ dominant eigenvalue of $\mathcal{G}(w)$

**Large deviation principle for short-range correlation:**

- Short-range correlation: $\lambda_1(w)$ is differentiable near 0

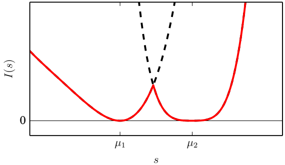
$$I(s) = \sup_{w}\{ws - \ln \lambda_1(w)\}$$

# Explicit rate function for long-range correlation

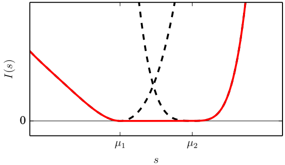
**Long-range or constant correlation**

$\Phi(w)$ is not differentiable in 0.

- Constant correlation: Non-convex rate function




- Polynomial correlation: Rate function with a flat branch



# Conclusion

- Three kind of correlation:
  - Exponential short-range correlation
  - Constant correlation
  - Polynomial long-range correlation
- Extension of the law of large numbers and the central limit theorems:
  - Long-range correlation : Continuous and discrete mixture of standard limit laws
- Large deviation principle:
  - Long-range correlation: Non-strictly convex rate function

## Perspective

- Extreme statistics
- Physical model
- Infinite dimension, higher tensor order