Random vectors with matrix representation: Insights from statistical physics

Florian Angeletti

Wednesday 8 May 2013

ENS DE LYON

Presentation

Thesis: "Sums and extremes in statistical physics and signal processing"
Advisors: Eric Bertin and Patrice Abry.

Themes: Application of statistical physics to signal processing

- Phase transition in moment estimation
- Extreme statistics
- Random vectors with matrix representation

Presentation

Thesis: "Sums and extremes in statistical physics and signal processing"
Advisors: Eric Bertin and Patrice Abry.

Themes: Application of statistical physics to signal processing

- Phase transition in moment estimation
- Extreme statistics
- Random vectors with matrix representation

Random vectors

Random vectors in signal processing

- joint probability density function: $\mathbb{P}(x_1, \ldots, x_n)$
- i.i.d. random vectors: $\mathbb{P}(x_1, \ldots, x_n) = f(x_1) \ldots f(x_n)$
- non-i.i.d. generalisation: ?

Random vectors

Random vectors in signal processing

- joint probability density function: $\mathbb{P}(x_1, \ldots, x_n)$
- i.i.d. random vectors: $\mathbb{P}(x_1, \ldots, x_n) = f(x_1) \ldots f(x_n)$
- non-i.i.d. generalisation: ?

Out-of-equilibrium physics

- Asymmetric Simple Exclusion Process model [Derrida et al., *J. Phys. A*, 1993]
- $p(x_1, \ldots, x_n) \propto \mathcal{R}(x_1) \ldots \mathcal{R}(x_n)$: f scalar $\Rightarrow \mathcal{R}$ matrix
- Product structure preserved.
- Applications in signal processing?

Outline

Matrix representation

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

- $\mathcal{L}(M) = \mathrm{tr}(\mathcal{A}^T M)$
  - $\mathcal{A}$: $d \times d$ 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 $\mathcal{P}_{i,j}(x) = \mathcal{R}(x) / \mathcal{E}_{i,j}$

- $d > 1$: Non commutativity $\Rightarrow$ Dependency

Covariance and higher-order dependency functions

Moment matrix: $M(q) = \int x^q \mathcal{R}(x) dx$

p-points moments of order q_r at positions k_r :

$$\mathbb{E}\left[X_{k_1}^{q_1} \dots X_{k_p}^{q_p}\right] = \frac{\mathcal{L}\left(\mathcal{E}^{k_1-1} M(q_1) \dots M(q_{r-1}) \mathcal{E}^{k_r - k_{r-1}-1} M(q_r) \dots M(q_p) \mathcal{E}^{n-k_p}\right)}{\mathcal{L}(\mathcal{E}^n)}.$$

Hidden random variable

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

- Expand the matrix product

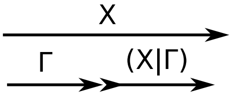
$$p(x_1, \dots, x_n) = \sum_{\underline{\Gamma}} \kappa(\Gamma)\mathbb{P}_\Gamma(\underline{X})$$

$$\kappa(\Gamma) = \frac{A_{\Gamma_0,\Gamma_N}}{\mathcal{L}\left(\mathcal{E}^n\right)}\mathcal{E}_{\Gamma_0,\Gamma_1}\dots\mathcal{E}_{\Gamma_{n-1},\Gamma_n} \quad \mathbb{P}_\Gamma(\underline{X}) = \mathcal{P}_{\Gamma_0,\Gamma_1}(x_1)\dots\mathcal{P}_{\Gamma_{n-1},\Gamma_n}(x_n)$$

- Γ: chain of index in the matrix $\mathcal{R}$
- Γ : hidden random vector

Hidden Markov Model

Separation of the randomness in 2 distinct levels.



Hidden Markov Chain Γ

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j} \frac{\left(\mathcal{E}^{n-k}\right)_{j,\Gamma_n}}{\left(\mathcal{E}^{n-k+1}\right)_{i,\Gamma_n}}.$$

$$\mathbb{P}(\Gamma_0 = i, \Gamma_n = j) = \frac{\mathcal{A}_{i,j}(\mathcal{E}^n)_{i,j}}{\mathcal{L}\left(\mathcal{E}^n\right)}.$$

Conditional pdf $(\underline{X}|\Gamma)$

$$\mathbb{P}\left(\underline{X}|\Gamma\right) = \prod_k \mathcal{P}_{\Gamma_k,\Gamma_{k+1}}(x_k)$$

$(\underline{X}|\Gamma)$ independent random variables

- Non-homogeneous Markov chain
- $\mathcal{E} \Rightarrow$ structure of the transition graph of the chain Γ

Dual representation

Matrix representation

- Useful for statistical property computation

Hidden Markov Model

- Useful for synthesis

Outline

Dependency: $\mathcal{E}$ with a dominant eigenvalue

- Eigenvalues of $\mathcal{E}$: λ_i
- Dominant eigenvalue λ_1: $\forall i \neq 1, |\lambda_1| > |\lambda_i|$
- E_i projection matrix of the associated eigenspace associated to λ_i.

$$\mathbb{E}\left[X_k X_l\right] \approx \sum_j \left(\frac{\lambda_j}{\lambda_1}\right)^{l-k-1} \frac{\mathcal{L}\left(E_1 M\left(1\right) E_j M\left(1\right) E_1\right)}{\lambda_1 \mathcal{L}\left(E_1\right)},$$

Dependency

Sum of at most $d-1$ decreasing exponentials with time scales $\tau_j = \left(\ln|\lambda_j| - \ln|\lambda_1|\right)^{-1}$

Dependency: $\mathcal{E} = \mathrm{I} + H$

- H nilpotent matrix of order $p+1$ ($H^p \neq H^{p+1} = 0$)
- Dominant eigenvalue $\lambda_1 = 1$ of multiplicity $p+1$

$$n \to +\infty, \quad \mathbb{E}\left[X_k^{q_1} X_l^{q_2}\right] \approx \sum_{i+j+m=p} \frac{(p)!}{(i)!(m)!(j)!} \left(\frac{k}{n}\right)^i \left(\frac{l-k}{n}\right)^j \left(1-\frac{l}{n}\right)^m \frac{\mathcal{L}\left(H^i M\left(q_1\right) H^j M\left(q_2\right) H^m\right)}{\mathcal{L}\left(H^p\right)}$$

Long-range (system-wide) dependency.

Correlation

- Dominant eigenvalue λ_1 of $\mathcal{E}$ of multiplicity m_{λ_1}
- If $m_{\lambda_1} = 1$: Short-range correlation
- If $m_{\lambda_1} > 1$: Associated Jordan block J_1
  - If J_1 is diagonal: constant correlation
  - If J_1 is non-diagonal: long-range polynomial correlation

Stationarity

Sufficient condition

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

Marginal distribution

$$\mathbb{P}(X_k = x) = \frac{\mathcal{L}\left(\mathcal{R}(x)\mathcal{E}^{n-1}\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

Partial joint distribution

$$\mathbb{P}(X_{k_1} = x_1, \dots X_{k_r} = x_r) = \frac{1}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

$$\mathcal{L}\left(\mathcal{R}(x_1)\mathcal{E}^{k_2-k_1-1}\mathcal{R}(x_2)\dots\mathcal{E}^{k_r-k_{r-1}-1}\mathcal{R}(x_r)\mathcal{E}^{n-(k_r-k_1)-1}\right)$$

Outline

Synthesis

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

How to choose

- d?
- $\mathcal{E}$?
- $\mathcal{P}$?
- $\mathcal{A}$?

Constraints

Classical constraints

- Marginal distribution: $\mathbb{P}_S$
- Autocovariance function: $c_{1,1} \equiv \mathbb{E}[X_0 X_t] - \mathbb{E}[X_0]\,\mathbb{E}[X_t]$
- (Higher-order) Dependency functions:
  $c_{q_1,q_2}(t) \equiv \mathbb{E}\left[X_0^{q_1} X_t^{q_2}\right] - \mathbb{E}\left[X_0^{q_1}\right]\mathbb{E}\left[X_t^{q_2}\right]$
- Limitations: sum of r exponentials with time scales θ_k d'amplitudes $\beta(q_1, q_2)$

$$c_{q_1,q_2}(t) = \sum_{k=1}^{r} \mathfrak{Re}\left\{\beta(q_1, q_2)_k \theta_k^t\right\}$$

Choice of d, $\mathcal{A}$, $\mathcal{E}$, $\mathcal{P}$

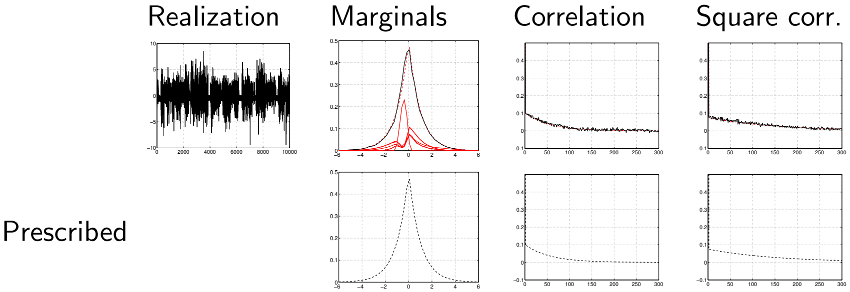
$$\mathcal{A} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}, \quad J_d = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ 1 & & & 0 \end{pmatrix}, \quad \mathcal{E} = \sum_k \alpha_k J_d^k$$

- Stationarity: $[\mathcal{A}^T, \mathcal{E}] = 0$
- Dependency functions: $\underline{\alpha} = \mathcal{F}(\underline{\tilde{\theta}})$ ($\mathcal{F}$ discrete Fourier transform)

Objectives $\Rightarrow$ Free parameters of the model

- $r \Rightarrow d$
- $\theta \Rightarrow \alpha$
- $\beta \Rightarrow M(q)$
- $\mathbb{P}_S \Rightarrow \mathcal{P}$

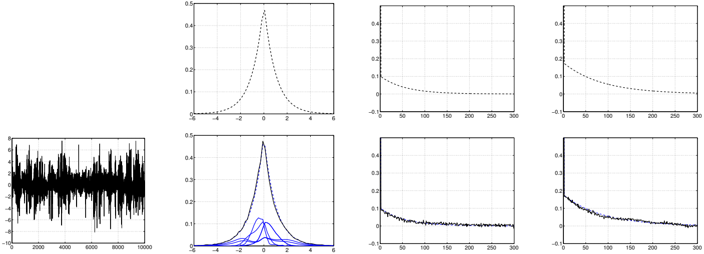
Stationary time series examples.



Stationary time series examples.

	Realization	Marginals	Correlation	Square corr.

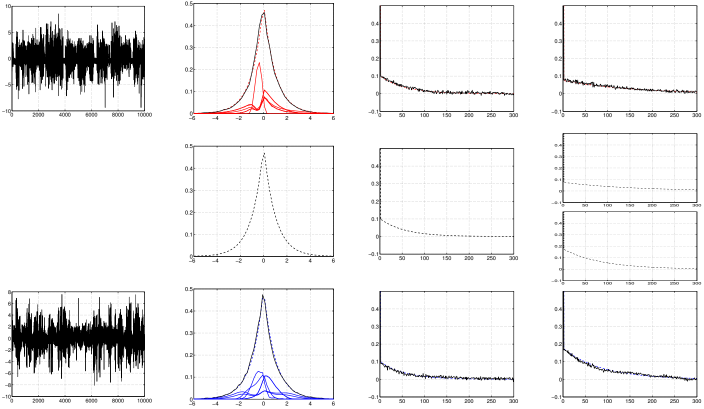
Prescribed



Stationary time series examples.

Realization | Marginals | Correlation | Square corr.

Prescribed



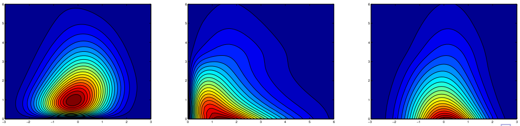
Random vector examples

- Trivariate distributions:
  - Prescribed marginal distributions:
    - p_1 gaussian
    - p_2 gamma distribution $\alpha = 2$
    - p_3 exponential distribution
  - Inter-covariance $\mathbb{E}\left[X_i X_j\right]$ prescribed a priori
- Generalization of the matrix product structure

$$p(x_1, x_2, x_3) = \frac{\mathcal{L}\left(\mathcal{R}_1(x_1)\mathcal{R}_2(x_2)\mathcal{R}_3(x_3)\right)}{\mathcal{L}\left(\mathcal{E}^n\right)}$$

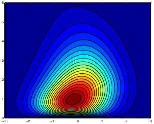
Partial bi-variate distributions :

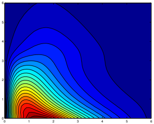
$\underline{X}$ $p_{1,2}$ $p_{2,3}$ $p_{1,3}$

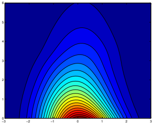


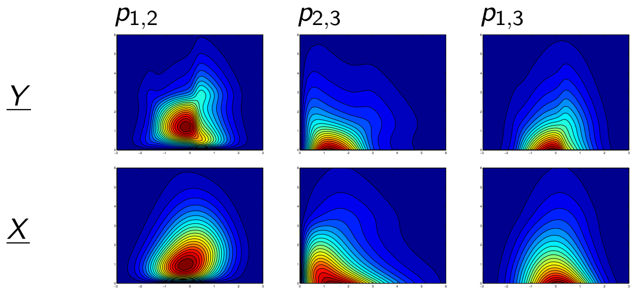
$p_{1,2}$ $p_{2,3}$ $p_{1,3}$

$\underline{X}$



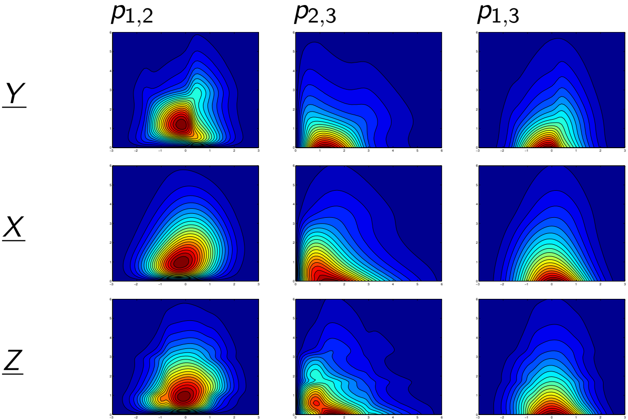






$\underline{X}$, $\underline{Y}$

- Same marginal d.
- Same correlation.
- Different square corr.



$\underline{X}$, $\underline{Y}$

- Same marginal d.
- Same correlation.
- Different square corr.
- (Different joint pdf)

$\underline{X}$, $\underline{Z}$

- Same marginal d.
- Same correlation.
- Same square corr.
- Different joint pdf

- Fine control over the dependency structure
- Stationary time series and random vectors design
- arxiv:1203.4500 (TSP in revision) , arxiv:1204.3047 (ICASSP proceedings)

Outline

Random vector sum

Sum

$$S\left(\underline{X}\right) = \sum_i X_i$$

- Sum of correlated and non-identically distributed random variables
- Equivalent to the law of large numbers?
- Equivalent to the central limit theorem?

Two paths:

- Matrix representation
- Hidden Markov Model

Random vector sum

Sum

$$S\left(\underline{X}\right) = \sum_i X_i$$

- Sum of correlated and non-identically distributed random variables
- Equivalent to the law of large numbers?
- Equivalent to the central limit theorem?

Two paths:

- Matrix representation
- Hidden Markov Model

Separation of the randomness source

- For a fixed Γ, $(\underline{X}|\Gamma)$ is independent

Conditional distribution of $S(\underline{X}|\Gamma)$

- $S(\underline{X}|\Gamma) = \sum_{i,j=1}^{d} \sum_{k=1}^{n\nu_{i,j}} X_{k;i,j}$
- $\nu_{i,j}$: fraction of the transitions from i to j:

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

- $X_{k;i,j}$ i.i.d. distributed with law $\mathcal{P}i,j$

- $S(\underline{X}|\Gamma)$ depends only on $\underline{\nu}$ and not of the fine structure of Γ

Convergence result

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

$$f_n\left(S\left(\underline{X}|\underline{\nu}\right)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \mathcal{L}_{\underline{\nu}}$$

Limit laws for ν

$$\underline{\nu} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \mathbb{P}(\underline{\nu})$$

$$\Downarrow$$

Limit laws for $S(\underline{X})$

$$f_n\left(S(\underline{X})\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \int \mathbb{P}\left(\underline{\nu}\right) \mathcal{L}_{\nu} d\underline{\nu}$$

- f_n scaling function:
  - Law of large number: $f_n(X) = \frac{X}{n}$
  - Central limit theorem: $f_n(X) = \frac{X-nm}{\sqrt{n}}$
- $\mathcal{L}_{\nu}$: pdf dependent on μ

Conditional convergence for the average

- Average:

$$\frac{S\left(X|\underline{\nu}\right)}{n} = \sum_{i,j} \nu_{i,j} \left(\frac{\sum_{k=1}^{n\nu_{i,j}} X_{k;i,j}}{n\nu_{i,j}} \right)$$

- Law of large number for i.i.d. random variables:

$$\frac{\sum_{k=1}^{\nu_{i,j}n} X_{k;i,j}}{n\nu_{i},j} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \delta\left(x - \mathbb{E}\left[X_{i,j}\right]\right)$$

- Conditional convergence of the average:

$$\frac{S(\underline{X}|\underline{\nu})}{n} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \delta\left(x - \sum_{i,j} \nu_{i,j} \mathbb{E}\left[X_{i,j}\right]\right)$$

Extension of the law of large number

Conditional law of large number

$$\frac{S(\underline{X}|\underline{\nu})}{n} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \delta\left(x - \sum_{i,j} \nu_{i,j}\mathbb{E}\left[X_{i,j}\right]\right)$$

Extension of the law of large number

$$\frac{S}{n} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \int \mathbb{P}\left(\underline{\nu}\right)\delta\left(x - \sum_{i,j} \nu_{i,j}\mathbb{E}\left[X_{i,j}\right]\right) d\underline{\nu}$$

Remark: If $\forall\left(i,j\right), \mathbb{E}\left[X_{;i,j}\right] = m$, $\frac{S}{n} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \delta\left(x - m\right)$. Henceforth, the standard law of large number is recovered.

Extension of the central limit theorem

Hypothesis: $\forall (i,j), \mathbb{E}[X_{i,j}] = m.$

Conditional central limit theorem

$$\frac{S(\underline{X}|\underline{\nu}) - mn}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{0,\sum \nu_{i,j}\sigma_{i,j}^2}$$

where $\sigma_{i,j}^2 = \mathbb{E}\left[(X_{i,j} - m)^2\right]$ and $\mathcal{N}_{\mu,\sigma^2}$ is the normal distribution of variance σ^2 and mean μ.

Extension of the central limit theorem

$$\frac{S - mn}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{D}} \int \frac{\mathbb{P}(\underline{\nu})}{\sqrt{2\pi \sum_{i,j} \nu_{i,j}\sigma_{i,j}^2}} e^{-\frac{x^2}{2\sum_{i,j}\nu_{i,j}\sigma_{i,j}^2}} d\underline{\nu}$$

Distribution of ν

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

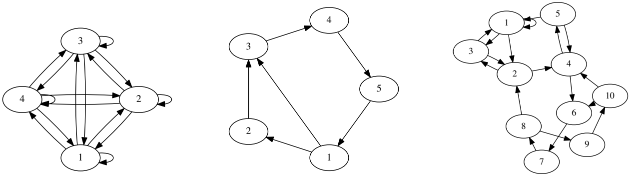
Three important subclasses:

- Irreducible $\mathcal{E}$ (short-range correlation)
- Identity $\mathcal{E}$ (constant correlation)
- Linear irreversible $\mathcal{E}$ (long-range polynomial correlation)

Irreducible $\mathcal{E}$

Irreducible matrix

$\mathcal{E}$ irreducible $\Leftrightarrow \exists k, \ \forall i,j, \ \mathcal{E}_{i,j}^k > 0$



- $\mathcal{E}$ irreducible $\Rightarrow$ one dominant eigenvalue $\Rightarrow$ Short-range correlation
- Far away from its endpoint, Γ is homogeneous.

Irreducible $\mathcal{E}$: Limit law for $\underline{\nu}$

Convergence towards the invariant measure

$$\nu_{i,j} \xrightarrow[n\to+\infty]{\text{a.s.}} c_{i,j} \text{ (deterministic)}$$

Extension of the standard law of large number and central limit theorem:

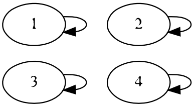
Law of large numbers $\frac{S}{n} \xrightarrow[n\to+\infty]{\mathcal{D}} \delta\left(x - \sum_{i,j} c_{i,j}\mathbb{E}\left[X\right]_{.;i,j}\right)$

Central limit theorem: $(S - nm)/\sqrt{n} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{0,\sum c_{i,j}\sigma^2_{i,j}}$

Standard limit laws.

Identity $\mathcal{E}$

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



- for all k, $\Gamma_k = \Gamma_0$
- Constant correlation
- $\mathbb{P}(\Gamma_0 = i) = \frac{\mathcal{A}_{i,i}}{\sum_k \mathcal{A}_{k,k}}$

Convergence in distribution of $\underline{\nu}$

$$\nu_{i,j} \underset{n \to +\infty}{\overset{\mathcal{D}}{\to}} \sum \mathbb{P}(\Gamma_0 = i)\, \delta(x - \mathbb{E}[X_{i,i}])$$

Identity $\mathcal{E}$: Limit laws

Extension of the Law of large numbers

$$\frac{S}{n} \xrightarrow[n\to+\infty]{\mathcal{D}} \sum_i \mathbb{P}\left(\Gamma_0 = i\right) \delta\left(x - \mathbb{E}\left[X\right]_{i,i}\right)$$

Extension of the central limit theorem

$$\frac{S - nm}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{D}} \sum_i \frac{\mathbb{P}\left(\Gamma_0 = i\right)}{\sqrt{2\pi\sigma_{i,i}^2}} e^{-x^2/(2\sigma_{i,i}^2)}$$

Non-standard limit laws: Discrete mixture of standard limit laws.

Linear irreversible $\mathcal{E}$

$$\mathcal{A}_{i,j} = 1/d \qquad \mathcal{E} = \begin{pmatrix} 1 & \epsilon & 0 & \\ & \ddots & \ddots & \\ 0 & & 1 & \end{pmatrix}$$

Statistical properties of Γ

- All acceptable Γ are equiprobable
- $i \neq j$, $\nu_{i,j} \underset{n \to +\infty}{\overset{a.s.}{\to}} 0$
- Long-range polynomial correlatin

Convergence in distribution of $\nu_{i,i}$

$$(\nu_{i,i})_{i \in 1,\dots,n} \underset{n \to +\infty}{\overset{\mathcal{D}}{\to}} \frac{1}{(d-1)!} \delta \left(\sum_i \nu_{i,i} - 1 \right)$$

Linear irreversible $\mathcal{E}$: Limit laws

Extension of the law of large number

$$\frac{S}{n} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \frac{1}{(d-1)!}\int \delta\left(\sum_i \nu_{i,i} - 1\right)\delta\left(x - \sum_{i,j}\nu_{i,j}\mathbb{E}\left[X_{i,j}\right]\right) d\underline{\nu}$$

Extension of the central limit theorem

$$\frac{S-mn}{\sqrt{n}} \underset{n\to+\infty}{\overset{\mathcal{D}}{\to}} \frac{1}{(d-1)!}\int \frac{\delta\left(\sum_i \nu_{i,i} - 1\right)}{\sqrt{2\pi\sum_i \nu_{i,i}\sigma_{i,i}^2}} e^{-\frac{x^2}{2\sum_i \nu_{i,i}\sigma_{i,i}^2}} d\underline{\nu}$$

Non-standard limit laws: Continuous mixture of standard limit laws.

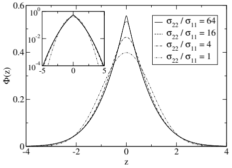
Example for $d = 2$

Extension of the law of large number

$$\frac{S}{n} \underset{n \to +\infty}{\overset{\mathcal{D}}{\to}} \frac{1}{\left|\mathbb{E}\left[X_{1,1}\right] - \mathbb{E}\left[X_{2,2}\right]\right|} \begin{cases} 1 & \text{if } x \in \left[\mathbb{E}\left[X_{1,1}\right], \mathbb{E}\left[X_{2,2}\right]\right] \\ 0 & \text{otherwise} \end{cases}$$

Extension of the central limit theorem

$$\frac{S - mn}{\sqrt{n}} \underset{n \to +\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \frac{1}{\sqrt{2\pi(\alpha\sigma_{1,1}^2 + (1-\alpha)\sigma_{2,2}^2)}} e^{-\frac{x^2}{2(\alpha\sigma_{1,1}^2 + (1-\alpha)\sigma_{2,2}^2)}} d\alpha$$



General case

- Complex mixture of the three previous behavior
- Non-standard limit laws: Discrete mixture of continuous mixtures of standard limit laws.
- Similar results for the extrema
- arxiv:1304.5406 (submitted to PRL), more articles in preparation

Outline

Perspective: Large deviation functions

- Law of large number: concentration of the average around the theoretical mean
- Central limit theorem: Fluctuations around the concentration point
- Large deviation functions: quantification of the concentration effect far away from the mean

Work in Progress

- Matrix representation more suitable
- In collaboration with Hugo Touchette (Queen Mary, University of London)
- Preliminary results:
    - $\mathcal{E}$ irreducible: standard large deviation function
    - $\mathcal{E}$ diagonal: non-convex large deviation function
    - $\mathcal{E}$ linear irreversible: large deviation function with a flat branch

Dual representation

Matrix representation

- Statistical properties
- Large deviation functions

Hidden Markov Model representation

- Synthesis
- Limits laws for sums (and maxima)