Critical order in moment estimation : insights from statistical physics

Florian Angeletti

5 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: statistical physics ∩ signal processing
  - Phase transitions in estimators
  - Extreme statistics
  - Random vectors with matrix representation
  - Large deviation functions

Moment estimation

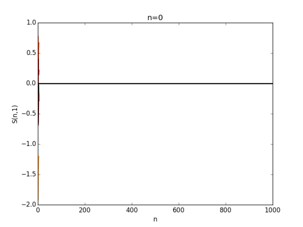
Moment : $\langle X^q \rangle$

Moment estimator :

$$S(n,q) = \frac{1}{n}\sum_{i=1}^{n} X_i^q, \quad X_i \geq 0$$

Law of large number:

X_i independent and identically distributed r.v. (i.i.d.)

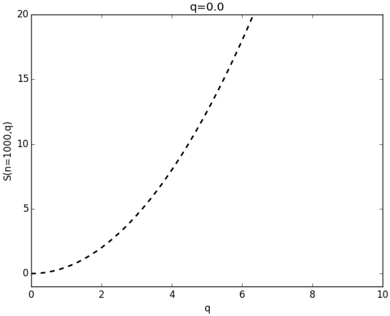


$$S(n,q) \underset{n\to+\infty}{\overset{\text{a.s}}{\to}} \langle X^q \rangle$$

Linearisation effect

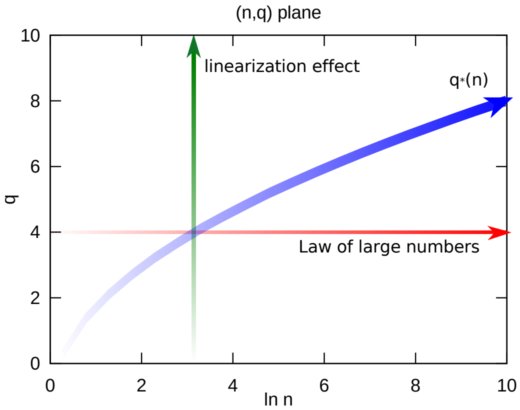
Finite sample size n?

- Fixed N
- Asymptotic behavior : $\ln S(N,q) \underset{q \to +\infty}{\sim} q \ln \max_{k=1,\dots,N}\{X_k\}$



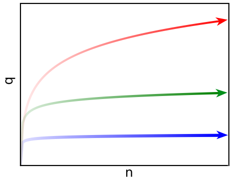
- Linearization effect in q

(n-q) plane



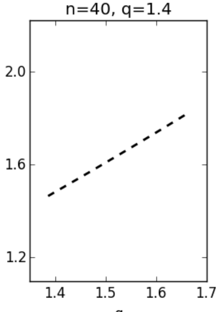
Transition line : $S(n, q^*(n))$
(inspired by [Ben Arous and al., *Probab. Theory Related Fields* , 2005])

Transition line $q^*(n)$

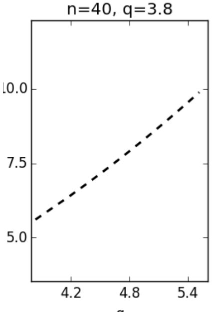


Slow $q(n) \ll q^*(n)$

$S(n, q(n)) \asymp \langle X^{q(n)} \rangle$

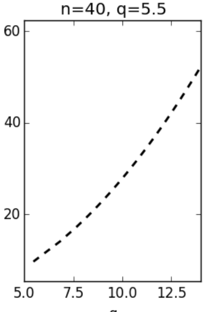


Critical $q^*(n)$



Fast $q(n) \gg q^*(n)$

$S(n, q(n)) \asymp e^{q \max\{X_i\}}$



Classes of i.i.d. random variables

Power laws

- Finite moment order q_l, $\langle X^{q_l} \rangle = +\infty$

Regular laws

- Characteristic function $\left\langle e^{\imath w X} \right\rangle$ analytic in 0.
- $\forall q > 0, \quad \langle X^q \rangle \in \mathbb{R}$
- $q^*(n) \asymp \frac{\ln n}{\ln \ln n}$ (A. Kagan and al, *Statistics & Probability Letters*, 2001)

Irregular laws

- $\forall q > 0, \quad \langle X^q \rangle \in \mathbb{R}$
- $\left\langle e^{\imath w X} \right\rangle$ non-analytic in 0.
- $q^*(n)$?

Lognormal law

- Y_i gaussian distribution:

$$P(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

- $X_i = e^{Y_i}$: lognormal distribution

Log-normal distribution

- All moments $\langle X_i^q \rangle$ are finite
- $\langle e^{\imath w Y} \rangle$ is not analytic in 0
- Infinite numbers of distributions with the same integer moments $\langle X^n \rangle$

Parametrized subclass

Exponential irregular laws

- $X_i = e^{Y_i}$
- $\mathbb{P}(Y_i > y) = e^{-y^\rho L(y)}$, with $\rho > 1$
- L slowly varying function : $L(tx) \underset{x \to +\infty}{\sim} L(x)$

ρ: tail heaviness parameter

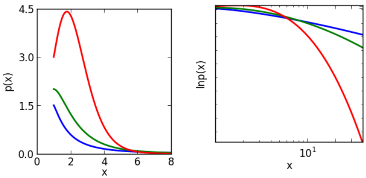
Frontiers:

- $\rho \to +\infty$: regular laws
- $\rho \to 1$: power laws

Families of test distributions

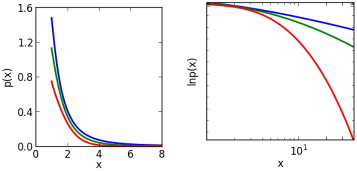
Log-weibull distribution family

- $F_Y(y) \equiv \mathbb{P}(Y > y) = e^{-y^{\rho}}$



Log-gamma distribution family

- $p_Y(y) = \frac{\rho}{2\Gamma(1/\rho)} e^{-y^{\rho}}$



$\rho = 1.5$

$\rho = 2$

$\rho = 3$

Moment estimation and statistical physics

- Moment estimator $S(n, q(n)) = \frac{1}{n} \sum_{i=1}^{n} X_i^{q(n)}$
  - $Y_i = -\ln X_i$

$$S(n, q) = \frac{1}{n} \sum_{i=1}^{n} e^{-qY_i}$$

- REM partition function:

$$Z(n = 2^m, \beta) = \sum_{i=0}^{2^m} e^{-\beta\sqrt{m}E_i}$$

Random Energy Model

A simplified spin glass model [Derrida, *Phys. Rev. B*, 1981]

- Simple disordered system
  - i.i.d. random configuration energies
- Glassy transition at finite inverse temperature β_c

How to translate arguments from the REM to moment estimation?

Towards a critical order

Competition between two effects

- Concentration effect
- Finite size effect

Concentration effect

$$\langle X^q \rangle = \int e^{qy-\phi(y)} \, \mathrm{dy} \quad \text{with } \phi = -\ln p_Y .$$

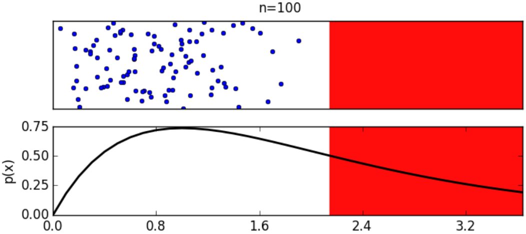
Saddle point method when $q \to +\infty$:

$$\ln \langle X^q \rangle \approx q y_{\mathrm{m}} - \phi(y_{\mathrm{m}})$$

Concentration Point $y_{\mathrm{m}}(q)$

$$y_{\mathrm{m}}(q) : \phi'(y_{\mathrm{m}}) = q$$

Finite size effect $\Rightarrow$ Cutoff



- Unreachable region : $]y^\dagger(n), +\infty[$

Cutoff point $y^\dagger$

$$y^\dagger(n) : \quad P(Y > y^\dagger(n)) = \frac{1}{n}.$$

Critical order

Truncated moments: excluding the contribution of the points beyond $y^\dagger$

$$M_t(n,q) = \int_{-\infty}^{y^\dagger} e^{qy-\phi(y)} \mathrm{d}\, y$$

- $y_\mathrm{m} < y^\dagger, \quad \ln M_t \approx \ln \langle X^q \rangle$
- $y_\mathrm{m} > y^\dagger, \quad \ln M_t \approx q y^\dagger - \phi(y^\dagger)$

Theorem:

$$\lim_{n\to+\infty} \frac{\ln S(n,q(n))}{\ln n} \overset{\mathrm{a.s}}{=} \lim_{n\to+\infty} \frac{\ln M_t(n,q(n))}{\ln n}$$

Truncated moment $\Rightarrow$ typical value of the moment estimator

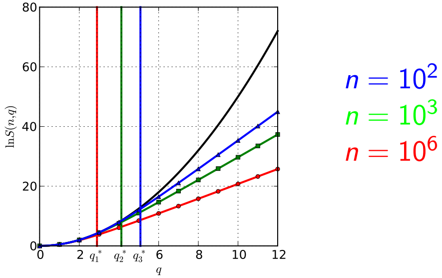
Critical order for exponential irregular laws

Critical order $q^*(n)$

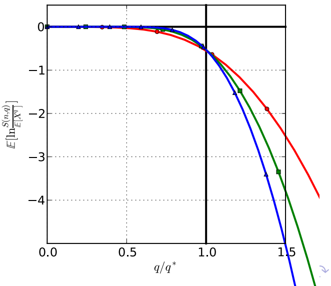
$$q^*(n) : y_{\mathrm{m}}(q^*) = y^\dagger(n)$$

$$q^*(n) \sim_{n\to+\infty} \rho \frac{\ln n}{y^\dagger(n)}$$

Prediction of the linearisation point:



Behavior of $\ln \frac{S(n,q)}{\langle X^q \rangle}$ in function of $\frac{q}{q^*}$



Asymptotic behavior of the critical order

$n \to +\infty$

$$q^*(n) \propto (\ln n)^{1-\frac{1}{\rho}}.$$

- Regular laws : $\rho \to +\infty, q^*(n) \propto \ln n$
- Log-normal law : $\rho = 2, \quad q^*(n) \propto \sqrt{\ln n}$
- Powers laws: $\rho \to 1, \quad q^*(n) \propto (\ln n)^{\rho-1}$

More details in arxiv:1204.3047

Critical order estimation

$$q^*(n) = \rho_I(y^\dagger(n)) \frac{\ln n}{y^\dagger(n)} \equiv \rho_E \theta$$

Estimation of q^*

- q^* only depends on information available from the empirical cumulative $\Rightarrow$ $q^*$is estimable
- $(1 - 1/n)$-quantile estimation $(y^\dagger(n))$
- Local power exponent at the last quantile $(\equiv \rho_I(y^\dagger(n)))$
- Theoretical and numerical analysis of the performance

θ estimation

$$\theta = \frac{\ln n}{F_Y^{-1}(\frac{1}{n})}$$

Estimation of $F_Y^{-1}(\frac{1}{n})$

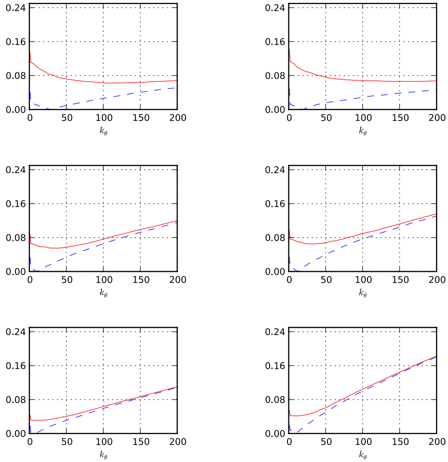
- $\max\{Y_i\}$: potentially asymptotically biased
- order statistics $Y_{1,n} \geq Y_{2,n} \geq \cdots \geq Y_{n,n}$
- $\Omega_{k_\theta} = \sum_{j=1}^{k_\theta} \alpha_j Y_{j,n}$:

$$\begin{cases} \forall i, i \neq k & \alpha_i = \frac{\sum_{l=1}^{k_\theta - 1} \frac{1}{l} - \gamma}{k_\theta - 1} \\ & \alpha_{k_\theta} = 1 - (k_\theta - 1)\alpha_1 \end{cases}$$

$\hat{\theta}_{k_\theta} = \frac{\ln n}{\Omega_{k_\theta}}$

θ: dependency on k

log-weibull log-gamma



$\rho = 1.1$

$\rho = 2$

$\rho = 3$

Relative mean square error

Relative bias

ρ_E estimation

$$\rho = \rho_I\left(F_n^{-1}\left(\frac{1}{n}\right)\right)$$

- $\partial_{\ln Y} F_Y(y) = \rho_I(y)$
- $\ln\left(-\ln \frac{i}{n}\right) \approx \rho_E \ln Y_{i,n} + \beta.$
- linear regression :
  - $C(x,y) = \overline{xy} - \overline{x}$
  - $\overline{x} = \frac{1}{k}\sum_{i=1}^{k_\rho} x_i$
- $\hat{\rho}_E = \frac{C \ln y, \ln(\ln n - \ln i))}{C(\ln y, \ln y)}$,

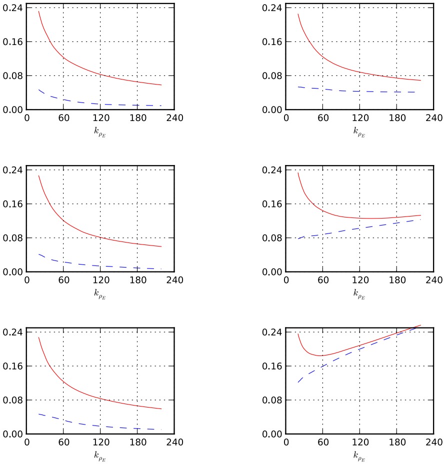
ρ_E: dependency on k

log-weibull log-gamma

$\rho = 1.1$

$\rho = 2$

$\rho = 3$



Relative mean square error

Relative bias

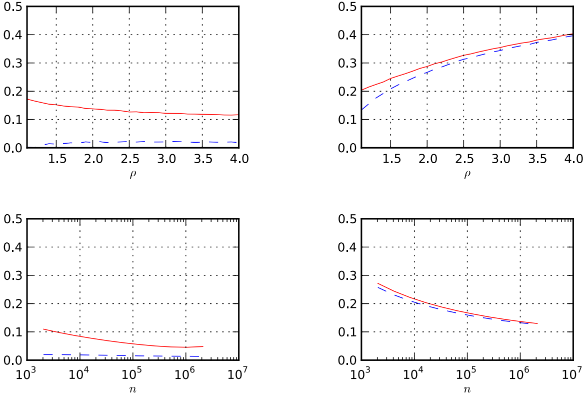
q^*

$$\hat{q}^* = \hat{\rho}_{E,k_{\rho_E}} \hat{\theta}_{k_\theta}$$

q^*: results

log-Weibull log-gamma

$n = 1000, \rho$



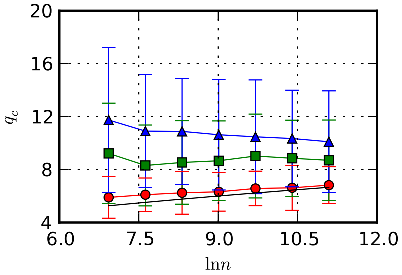
$\rho = 2, n$

Relative mean square error

Relative bias

Estimation with short-range correlations

- Exponentially correlated time series $\mathrm{Corr}(X_k X_{k+T}) \propto e^{-t/\tau}$
- $\hat{q}_c$ becomes biased

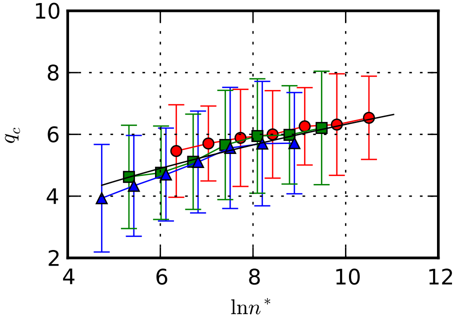


- Empirical workaround
  - Replace n with an effective number of independent samples

$$n_{\mathrm{eff}} = \frac{n}{1 + \alpha\tau}$$

  - Compute an sieved order statistics by eliminating maxima too close from each other

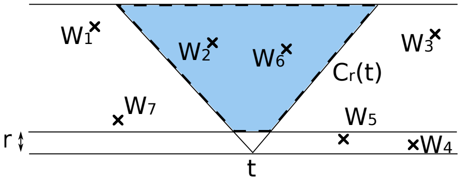
Results with correlation



Correlation $\mathrm{Corr}(t) = \exp(-t/\tau)$
$\tau = 10$,
$\tau = 50$,
$\tau = 100$
More details in arxiv:1204.3047

Multifractal process

Compound Poisson motion $Z(t)$



- $Q_r(t) = B_r(t) \prod_{p_i \in C_r(t)} W_i$
- $Z(t) = \lim_{r \to 0} \int_0^t Q_r(s) ds$
- Increments $X(a,t) = Z(t+a) - Z(t)$

Z(t): Multifractal increasing random walk

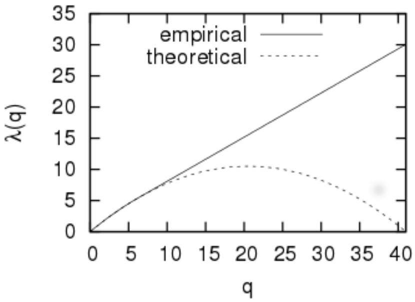
Correlation

$$\langle X(a,t) X(a,t+s) \rangle = \frac{1}{\lambda(2)\,(\lambda(2)-1)} \left(|s+a|^{\lambda(2)} + |s-a|^{\lambda(2)} - 2|s|^{\lambda(2)} \right),$$

- No characteristic scale

Linearisation effect in multifractal analysis

- Increment $X(a,t) = X(t+a) - X(t)$
- $\lambda(q)$: $\langle X(a,t)^q \rangle = C_a a^{\lambda(q)}$
- $S(a,q) = \sum_{k=1}^{L/a} X(a,ka)^q$
  - n is replaced by $|\ln a|$



- Linearisation effect at q^* independent of a

Scale dependency of the p.d.f.

Scaled variable h

$$H_a(t) = \frac{\ln X(a,t)}{\ln a} = \frac{Y(a,t)}{|\ln a|}.$$

Probability density function of H_a

Large deviation theory and Gartner-Ellis theorem $\Rightarrow$

$$p_{H_a}(h) \asymp a^{\psi(h)}$$

$\psi(h)$ Fenchel-Legendre transform of $\lambda(q)$

Scale and concentration

$$\langle X(a,t)^q \rangle \approx \int_{-\infty}^{+\infty} e^{-|\ln a|[qh+\psi(h)]}\, dh$$

Saddle point method when $a \to 0$:

$$\langle X(a,t)^q \rangle \approx -|\ln a|[q h_{\mathrm{m}} - \psi(h_{\mathrm{m}})]$$

Concentration Point $h_{\mathrm{m}}(q)$

$$h_{\mathrm{m}}(q): \quad \psi'(h_{\mathrm{m}}(q)) = q$$

- Implicit scale a dependency :

$$y_{\mathrm{m}}(a,q) = |\ln a| f(q)$$

Number of effective samples

- Cutoff point $h^\dagger$:

$$h^\dagger(a) : \quad P(\forall k \in \{0, \ldots, \frac{L}{a}\}, H_a(ka) > h^\dagger(a)) = \frac{1}{e}$$

- Effective number of independent samples n_{eff} :

$$n_{\text{eff}} : \quad P(H_a > h^\dagger(a)) = \frac{1}{n_{\text{eff}}(a)}$$

- Hypothesis :

$$n_{\text{eff}}(a) \propto \frac{L}{a}$$

Critical order

$$h^{\dagger}(a) = h_{\mathrm{m}}(q^*)$$

Critical order $q^*(a)$

$$1 + q^* \lambda'(q^*) - \lambda(q^*) = 0.$$

- q^* is independent of a !
- q^* depends only of λ : intrinsic characteristic of the cascade
- Interference between the dependency in a of p_Y and the correlation length
- known result for dyadic cascade or compound poisson cascade via different methods

More details in ¡a?¿ arxiv:1012.3688 ¡/a?¿

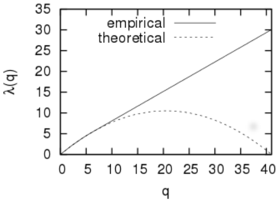
Linearisation effect

Under a reasonable conjecture:

$$\frac{\ln S(a,q)}{\ln a} \underset{\ln a \to +\infty}{\overset{\text{a.s}}{\to}} \zeta(q)$$

Linearisation effect

$$\zeta(q) = \begin{cases} \lambda(q), & -1 < q \leq q^*, \\ 1 + q\lambda'(q^*), & q > q^*. \end{cases}$$



Glassy transition

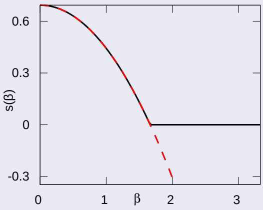
$Z \Leftrightarrow S(n,q)$

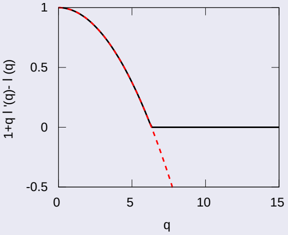
- Z fonction de partition REM $\begin{cases} \text{inverse temperature } \beta \leftrightarrow q \\ \text{energy } E_i \leftrightarrow -\ln X_i \end{cases}$

$\lambda_e(q) = \lim_{a\to 0} \ln S(a,q)/\ln a$

Analogy REM⇔ Moments

Entropy $s(\beta)$ $\Leftrightarrow$ $1 + q\lambda_e'(q) - \lambda_e(q)$





Glassy transition at β_c $\Leftrightarrow$ Moment linearisation at q^*

Conclusion

Summary

- Formal analogy between linearization effect and glassy phase transition
- A notion of critical order for moment estimation :
  - Computable
  - Estimable
- An alternative interpretation of the linearization effect in multifractal analysis

Connected questions

- Behavior of $\max_{k=1,\dots,n}\{X_k^{q(n)}\}$ (arxiv:1112.2965)
- Estimation of large deviation functions (arxiv:1409.8531)

Outlook

- A more formal analysis of critical order for correlated random variables?
  - Random vector with a matrix representation