Large deviations, moment problems and sum rules

Jan Nagel

TU Dortmund

40th Finnish Summer School on Probability and Statistics

Lammi, 23.05.2022 - 27.05.2022

Overview

1st lecture: Random moments and large deviations

2nd lecture: Large deviation for random matrices and sum rules

3rd lecture: A probabilistic proof of sum rules

The Killip-Simon sum rule

Killip, Simon (Ann. Math., 2003):

$$\mathcal{K}(\mu_{\mathrm{sc}}|\mu) + \sum_{\lambda \in E} \mathcal{F}(\lambda) = \sum_{k=1}^{\infty} \tfrac{1}{2} b_k^2 + a_k^2 - 2\log(a_k) - 1$$

Where:

- $\mathrm{supp}(\mu) = [-2, 2] \cup E$, with E at most countable,
- $\mathcal{K}(\mu_{\mathrm{sc}}|\mu) = \int \log \frac{\mathrm{d}\mu_{\mathrm{sc}}}{\mathrm{d}\mu} \mathrm{d}\mu_{\mathrm{sc}} = -\int \log(w)\, \mathrm{d}\mu_{\mathrm{sc}}$
  the Kullback-Leibler divergence, rewritten for $\mathrm{d}\mu(x) = w(x)\mathrm{d}\mu_{\mathrm{sc}}(x) + \mathrm{d}\mu_s(x)$.
- $\mathcal{F}(x) = \int_2^{|x|} \sqrt{z^2 - 4}\, \mathrm{d}z$,

And both sides are simultaneously finite or infinite.

More sum rules

Probabilistic approach to reprove the KS-SR and to obtain new sum rules:

- Choose a Hermitian $n \times n$ matrix X_n at random. It has a (random) spectral measure μ_n, supported by n points (the ev of X_n).
- Parametrizing μ_n by Jacobi coefficients $b_1, a_1, \ldots, a_{n-1}, b_n$, show that $(\mu_n)_n$ satisfies a large deviation principle with rate function $\mathcal{I}_{\mathsf{co}}$.
- Parametrizing μ_n by ev $\lambda_1, \ldots, \lambda_n$ and weights $w_1, \ldots, w_n$, show that $(\mu_n)_n$ satisfies a large deviation principle with rate function $\mathcal{I}_{\mathsf{sp}}$.
- This will imply the sum rule $\mathcal{I}_{\mathsf{sp}} = \mathcal{I}_{\mathsf{co}}$.
- The LDP in the spectral parmetrization is quite general. The LDP parametrized by coefficients is restricted to very special cases, for example the Gaussian, Laguerre or Jacobi ensemble.

Random spectral measures

- Let X_n be a random $n \times n$ matrix of the Gaussian ensemble with density

$$Z_n^{-1} \exp\left(-\tfrac{1}{2} n \operatorname{tr} X^2\right)$$

- Spectral measure: $\mu_n = \sum_{i=1}^{n} w_i \delta_{\lambda_i}$
- Density of eigenvalues is

$$c_n^{-1} \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} e^{-\frac{1}{2} n \lambda_i^2}.$$

- The density is invariant under the unitary conjugations $X \mapsto OXO^*$, which implies that the eigenvector matrix is Haar distributed on the unitary group and independent of the eigenvalues.
- $(U_{1,1}, \dots, U_{1,n})$ is then uniformly distributed on the unit sphere in $\mathbb{C}^n$. Therefore, $(w_1, \dots, w_n)$ is uniformly distributed on the unit simplex (Dirichlet distributed with parameters all 1).

Spectral LDP

- The LDP using the spectral encoding holds for more general spectral measures of random matrices, chosen according to the density

$$Z_n^{-1} \exp\left(-n \operatorname{tr} V(X)\right)$$

- We suppose that $V : \mathbb{R} \to (-\infty, \infty]$ is finite on (b^-, b^+), continuous on $[b^-, b^+]$ and infinite outside of this set ($b^\pm = \pm\infty$ is allowed), and satisfies the *confinement assumption*

$$\liminf_{|x| \to \infty} \frac{V(x)}{\log |x|} > 2.$$

- As before, $\mu_n = \sum_{i=1}^n w_i \delta_{\lambda_i}$, with weights $(w_1, \ldots, w_n)$ uniformly distributed on the standard simplex and independently, $(\lambda_1, \ldots, \lambda_n)$ have density

$$c_n^{-1} \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^n e^{-nV(\lambda_i)}.$$

Spectral LDP

Recall that $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$ is the empirical spectral measure and the eigenvalue density can be written as

$$c_n^{-1}\prod_{i<j}|\lambda_i-\lambda_j|^2\prod_{i=1}^n e^{-nV(\lambda_i)} = c_n^{-1}\exp\left\{\sum_{i\neq j}\log|\lambda_i-\lambda_j| - n\sum_{i=1}^n V(\lambda_i)\right\}$$

$$= c_n^{-1}\exp\left\{-n^2\left(\int V(x)\,\mathrm{d}\hat{\mu}_n(x) - \iint_{x\neq y}\log|x-y|\,\mathrm{d}\hat{\mu}_n(x)\mathrm{d}\hat{\mu}_n(y)\right)\right\}$$

Ben Arous, Guionnet (1997):

$(\hat{\mu}_n)_n$ satisfies the LDP with speed n^2 and rate function

$$\hat{\mathcal{I}}(\mu) = \mathcal{E}(\mu) - \inf_{\nu}\mathcal{E}(\nu),$$

with $\mathcal{E}(\mu) = \displaystyle\int V(x)d\mu(x) - \iint \log|x-y|d\mu(x)d\mu(y)$.

Spectral LDP

Ben Arous, Guionnet (1997):

$(\hat{\mu}_n)_n$ satisfies the LDP with speed n^2 and rate function

$$\hat{\mathcal{I}}(\mu) = \mathcal{E}(\mu) - \inf_{\nu} \mathcal{E}(\nu),$$

with $\mathcal{E}(\mu) = \displaystyle\int V(x)d\mu(x) - \iint \log|x-y|d\mu(x)d\mu(y).$

- $\inf_\nu \mathcal{E}(\nu) = \mathcal{E}(\mu_V)$ for a unique, compactly supported μ_V.
- This also implies $\mu_n \to \mu_V$ in probability.
- From now on, we also assume the *support condition:*

$$\mathrm{supp}(\mu_V) = S = S_1 \cup \cdots \cup S_M,$$

for S_j compact, pairwise disjoint intervals.

Spectral LDP

- Finally, we assume the *control condition:* the effective potential

$$\mathcal{J}_V(x) = V(x) - 2\int \log|x-\xi|\, d\mu_V(\xi)$$

achieves its minimal value on $\mathbb{R} \setminus \mathrm{Int}(S)$ only on the boundary.

- After reordering, assume λ_1 is the largest ev,

$$\mathbb{P}(\max_i \lambda_i \in A) = \int_A \int_{\Delta(\lambda_1)} c_n^{-1} \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^n e^{-nV(\lambda_i)}\, d\lambda^{(n-1)} d\lambda_1$$

$$= \int_A \int_{\Delta(\lambda_1)} \left(c_n^{-1} \prod_{1<i<j,} |\lambda_i - \lambda_j|^2 \prod_{i=2}^n e^{-nV(\lambda_i)} \right)$$

$$\times \exp\left\{ -nV(\lambda_1) + 2\sum_{j=2}^n \log|\lambda_1 - \lambda_j| \right\} d\lambda^{(n-1)} d\lambda_1.$$

Spectral LDP

Effective potential: $\mathcal{J}_V(x) = V(x) - 2\int \log|x-\xi|\,\mathrm{d}\mu_V(\xi)$

Ben Arous, Dembo, Guionnet (2001):

$(\max \lambda_i)_n$ satisfies the LDP with speed n and rate function

$$\mathcal{F}_V(x) = \mathcal{J}_V(x) - \inf_y \mathcal{J}_V(y)$$

if $x \geq \sup(S)$ and $\mathcal{F}_V(x) = \infty$ else.

Analogous results hold for the smallest ev or for the vector of k largest ev (Benaych-Georges, Guionnet, Maida (2012), Borot, Guionnet (2013)).

Spectral LDP

Gamboa, N, Rouault (2016,2022):

$(\mu_n)_n$ satisfies a large deviation principle with speed n and rate function

$$\mathcal{I}_{\text{sp}}(\mu) = \mathcal{K}(\mu_V|\mu) + \sum_{\lambda \in E} \mathcal{F}_V(\lambda)$$

for μ with $\text{supp}(\mu) = S \cup E$ (E at most countable), and $\mathcal{I}_{\text{sp}}(\mu) = \infty$ otherwise.

- Unlike in the LDP for $\hat{\mu}_n$, the speed is reduced to n and we allow point masses.
- In the case $|E| = \infty$, the rate is only finite if the only accumulation points of E are in ∂S.
- When $\mathrm{d}\mu = w(x)\mathrm{d}\mu_V(x) + \mathrm{d}\mu_s$, we have

$$\mathcal{K}(\mu_V|\mu) = -\int \log(w(x))\, \mathrm{d}\mu_V(x)$$

  so the rate function depends on the singular part μ_s only through its total mass and points in E.

From LDPs to sum rules

- In the GUE case $V(x) = \frac{1}{2}x^2$, we have $\mu_V = \mu_{\mathrm{sc}}$ with support $S = [-2, 2]$,

$$\mathcal{F}_V(x) = \int_2^{|x|} \sqrt{z^2 - 4}\,\mathrm{d}z.$$

- By the uniqueness of rate functions, we get the KS-SR

$$\mathcal{I}_{\mathsf{sp}}(\mu) = \mathcal{I}_{\mathsf{co}}(\mu).$$

- The new Laguerre sum rule follows by taking

$$V(x) = \tau^{-1}x - (\tau^{-1} - 1)\log(x), \qquad x > 0, \tau \in (0, 1]$$

  Dumitriu, Edelman (2002) derive the good distribution of the z_k, the LDP using this encoding is due to Gamboa, Rouault (2011).

- More generalizations: Jacobi ensemble, unitary ensembles, operator valued measures, ...

New encoding

- For the proof, assume μ_V is supported by a single interval $S = [s^-, s^+]$ (one-cut-case).
- Main difficulty: dependency between the unbounded number of outliers and the bulk.
- Suppose μ has support $I \cup E$, $I \subset [s^-, s^+]$, and $E \subset S^c$ is bounded and at most countable with accumulation points only in ∂S. This is satisfies for μ_n and any μ where the rate is finite.
- Such a measure can be written as

$$\mu = \mu_S + \sum_{i=1}^{\infty} \gamma_i^+ \delta_{\zeta_i^+} + \sum_{i=1}^{\infty} \gamma_i^- \delta_{\zeta_i^-},$$

  where μ_S is the restriction to S and

$$\begin{aligned} \zeta^- &= (\zeta_i^-)_i \in (-\infty, s^-]^{\mathbb{N}} \text{ increasing}, \\ \zeta^+ &= (\zeta_i^+)_i \in [s^+, \infty)^{\mathbb{N}} \text{ decreasing}, \\ \gamma^\pm &= (\gamma_i^\pm)_i \in [0, \infty)^{\mathbb{N}}. \end{aligned}$$

- We identify μ with the vector

$$(\mu_S, \zeta^+, \zeta^-, \gamma^+, \gamma^-).$$

Step 1: decoupling of weights

- The Dirichlet distributed weights $(w_1, \dots, w_n)$ have the same distribution as

$$\left(\frac{\omega_1}{\omega_1 + \cdots + \omega_n}, \dots, \frac{\omega_n}{\omega_1 + \cdots + \omega_n}\right),$$

  where ω_i are independent and each $\mathrm{Exp}(n)$ distributed.

- Let $\tilde{\mu}_n$ be the measure with decoupled weights,

$$\tilde{\mu}_n = \sum_{i=1}^{n} \omega_i \delta_{\lambda_i},$$

  identified with

$$(\tilde{\mu}_{n,S}, \zeta_n^+, \zeta_n^-, \gamma_n^+, \gamma_n^-).$$

- Minor annoyance: in this encoding, the outliers and weights are not independent.

Step 2: LDP for outliers and weights

- Let π_J be the projection onto the first J coordinates, then $(\pi_J(\zeta_n^+), \pi_J(\zeta_n^-))_n$ satisfies the LDP with speed n and rate function

$$\mathcal{I}_J^{\text{ev}}(z^+, z^-) = \sum_{i=1}^{J} \mathcal{F}_V(z_i^+) + \mathcal{F}_V(z_i^-)$$

  for $z^+ \in [s^+, \infty)^J$ with decreasing, $z^- \in (-\infty, s^-]^J$ with increasing coordinates.
- Proof: Follow existing proofs for a collection of J largest eigenvalues, using the explicit density (Benaych-Georges, Guionnet, Maida (2012), Borot, Guionnet (2013)).
- The sequence of $X_n \sim \text{Exp}(n)$ satisfies the LDP with speed n and rate the identity on $[0, \infty)$. Then $(\pi_J(\gamma_n^+), \pi_J(\gamma_n^-))_n$ satisfies the LDP with speed n and rate

$$\mathcal{I}_J^{\text{w}}(x^+, x^-) = ||x^+||_1 + ||x^-||_1.$$

Step 2: LDP for outliers and weights

- Although eigenvalues and weights are not independent, we can show that

$$\big(\pi_J(\zeta_n^+, \zeta_n^-, \gamma_n^+, \gamma_n^-)\big)_n$$

  satisfies the LDP with speed n and rate

$$\mathcal{I}_J^{\mathrm{ev,w}}(z^+, z^-, x^+, x^-) = \mathcal{I}_J^{\mathrm{ev}}(z^+, z^-) + \mathcal{I}_J^{\mathrm{w}}(x^+, x^-).$$

- By the Dawson-Gärtner Theorem, $(\zeta_n^+, \zeta_n^-, \gamma_n^+, \gamma_n^-)_n$ satisfies the LDP with rate

$$\begin{aligned}\mathcal{I}^{\mathrm{ev,w}}(z^+, z^-, x^+, x^-) &= \sup_J \mathcal{I}_J^{\mathrm{ev,w}}(\pi_J(z^+, z^-, x^+, x^-)) \\ &= \sum_{i=1}^{\infty} \mathcal{F}_V(z_i^+) + \mathcal{F}_V(z_i^-) + x_i^+ + x_i^-.\end{aligned}$$

Step 3: LDP for the bulk

- Let us ignore the dependence between eigenvalues in S and outliers, consider only $\tilde{\mu}_{n,S}$:

$$\begin{aligned}\mathbb{E}\left[\exp\left\{n\int f\,\mathrm{d}\tilde{\mu}_{n,S}\right\}\right] &= \mathbb{E}\left[\exp\left\{n\sum_{i=1}^{n}\omega_i f(\lambda_i)\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{\sum_{i=1}^{n}-\log(1-f(\lambda_i))\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{n\int-\log(1-f)\,\mathrm{d}\hat{\mu}_n\right\}\right].\end{aligned}$$

- Since $(\hat{\mu}_n)_n$ satisfies the LDP with speed n^2,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[\exp\left\{n\int f\,\mathrm{d}\tilde{\mu}_{n,S}\right\}\right] = -\int\log(1-f)\,\mathrm{d}\mu_V.$$

- From Rockafellar (1971), we get for $\mu = w\cdot\mu_V + \nu_s$ supported on S

$$G^*(\mu) = \sup_f\left\{\int f\,\mathrm{d}\mu - G(f)\right\} = \int(w-1-\log w)\,\mathrm{d}\mu_V + \nu_s(S).$$

Step 4: recovering the spectral measure

- This implies the LDP for $(\tilde{\mu}_{n,S}, \zeta_n^+, \zeta_n^-, \gamma_n^+, \gamma_n^-)$ with rate function

$$\mathcal{I}(\mu_S, z^+, z^-, x^+, x^-) = \mathcal{K}(\mu_V|\mu_S) + \mu_S(S) - 1 + \sum_{i=1}^{\infty} \mathcal{F}_V(z_i^+) + \mathcal{F}_V(z_i^-) + x_i^+ + x_i^- .$$

- After strengthening the topology on the weights to ℓ^1, we may map continuously

$$(\mu_S, \zeta^+, \zeta^-, \gamma^+, \gamma^-) \mapsto \mu_S + \sum_{i=1}^{\infty} \gamma_i^+ \delta_{\zeta_i^+} + \gamma_i^- \delta_{\zeta_i^-},$$

which yields by the contraction principle the LDP for $(\tilde{\mu}_n)_n$ with rate function

$$\tilde{\mathcal{I}}(\mu) = \mathcal{K}(\mu_V|\mu) + \mu(\mathbb{R}) - 1 + \sum_{\lambda \in E} \mathcal{F}_V(\lambda)$$

- Applying $\mu \mapsto \mu/\mu(\mathbb{R})$ to $\tilde{\mu}_n$, we obtain the original spectral measure and the claimed LDP.

A general coefficient LDP/sum rule?

- The LDP using the spectral encoding holds for fairly general potentials V. To obtain a sum rule, we need both the spectral LDP and the coefficient LDP, but the latter is available only in specific situations (Gaussian/Laguerre/Jacobi ensemble...).
- General density for $(b_1^{(n)}, a_1^{(n)}, \dots, b_n^{(n)})$ (Krishnapur, Rider, Virag (2016)):

$$(Z_n^V)^{-1} \exp\left\{ -n \left(\operatorname{tr} V(T_n) - 2 \sum_{k=1}^{n-1} (1 - k/n - 1/n) \log(a_k) \right) \right\},$$

where $T_n = T_n(\mu_n)$ is a tridiagonal finite-dimensional Jacobi matrix.

- Leads to the conjecture:

$$\mathcal{I}_{\text{co}}(\mu) = \lim_{n\to\infty} \left\{ \operatorname{tr}[V(T_n(\mu)) - V(T_n(\mu_V))] - \sum_{k=1}^{n-1} \log\left(\frac{a_k}{a_k(\mu_V)} \right) \right\}$$

A general sum rule?

Gamboa, N, Rouault (2022):

For V a polynomial of degree $2m$,

$$\mathcal{I}_{\text{co}}(\mu) = \lim_{n\to\infty} \left\{ \text{tr}[V(T_n(\mu)) - V(T_n(\mu_V))] - \sum_{k=1}^{n-1} \log\left(\frac{a_k}{a_k(\mu_V)}\right) + \xi_n(\mu) \right\}.$$

- The "error term" ξ_n depends only on $a_{n-m}, b_{n-m}, \dots, b_n$ and satisfies

$$|\xi_n| \leq C(V) \left(\sup_k (a_k + |b_k|)\right)^{2m} \left(|a_{n-m} - a_{n-m}(\mu_V)| + \cdots + |b_n - b_n(\mu_V)|\right).$$

- If the support of μ_V is a single interval, we have $\xi_n \to 0$ for all μ for which the rate is finite and ξ_n can be omitted.
- Interpretation of ξ_n?

Idea of proof

- The density can be decomposed using

$$\operatorname{tr} V(T_n) = \operatorname{tr} V(T_k) + M_k(a_{k-m}, \dots, b_{k+m}) + H_{k,n}(a_k, b_{k+1}, \dots, b_n).$$

- We know (from the spectral LDP) that $r^{(n)} = (b_1^{(n)}, \dots, a_{k-1}^{(n)}, b_k^{(n)})$ satisfies an LDP. The rate $\mathcal{I}_k$ satisfies

$$\mathcal{I}_k(r) = \lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(r^{(n)} \in B_\delta(r)\big).$$

- By the above decomposition, the rate is given by

$$\operatorname{tr} V(T_k) + 2 \sum_{i=1}^{k-1} \log(a_i) + \text{ standardization } + \text{ error term.}$$

The remainder term

- Consider the quartic potential

$$V(x) = \frac{x^4}{4} - \frac{\kappa x^2}{2}.$$

For $\kappa \geq 2$, the support of μ_V is

$$[-\alpha^+, -\alpha^-] \cup [\alpha^-, \alpha^+], \qquad \alpha^\pm = \sqrt{\kappa \pm 2}$$

and for $\kappa > 2$, the Jacobi coefficients are

$$b_k = 0, \qquad a_k = \frac{\alpha^+ + (-1)^{k+1}\alpha^-}{2}.$$

- Let $\bar{\mu}_V$ be the spectral measure with even/odd a_k switched.
- If the "error term" would not be present, the rate would not converge at $\bar{\mu}_V$. More interestingly, the subsequential limit for even n would yield $\mathcal{I}_{\text{co}}(\bar{\mu}_V) = 0$. But we know $\mathcal{I}_{\text{sp}}(\bar{\mu}_V) > 0$.

Sum rules

As a consequence of the "two" LDPs, $\mathcal{I}_{\mathrm{sp}}(\mu) = \mathcal{I}_{\mathrm{co}}(\mu)$:

$$\mathcal{K}(\mu_V|\mu) + \sum_{\lambda \in E} \mathcal{F}_V(\lambda)$$

$$= \lim_{n\to\infty} \left\{ \mathrm{tr}[V(T_n(\mu)) - V(T_n(\mu_V))] - \sum_{k=1}^{n-1} \log\left(\frac{a_k}{a_k(\mu_V)}\right) + \xi_n(\mu) \right\}$$

- Both sides are simultaneously finite or infinite.
- Using the bounds on ξ_n, we still obtain gems.
- Gems in different formulations were obtained in the single interval case (Nazarov et al. (2005)) and for some reference measures in the multi interval case (Yuditskii (2018)).
- Is this equality true for non-polynomial V?

Thank you for your attention!

- G. Anderson, A. Guionnet, O. Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2010
- F. Gamboa, J. Nagel, A. Rouault, *Sum rules via large deviations*, Journal of Functional Analysis 270, 2016
- J. Breuer, B. Simon, O. Zeitouni, *Large deviations and sum rules for spectral theory - a pedagogical approach*, Journal of Spectral Theory 8, 2016
- F. Gamboa, J. Nagel, A. Rouault, *Sum rules via large deviations: extension to polynomial potentials and the multi-cut regime*, Journal of Functional Analysis 282, 2022