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

Spectral measures

- Let μ be a probability measure with compact support, which contains infinitely many points.
- Then the monomials $1, x, x^2, \dots$ are linearly independent in $L^2(\mu)$, since $\int |P(x)|^2 \, \mathrm{d}\mu(x) = 0$ implies μ is supported on the zeros of P.
- By Gram-Schmidt, we obtain a sequence of monic orthogonal polynomials

$$P_k(x) = x^k + a_{k,1}x^{k-1} + \dots$$

  and orthonormal polynomials $p_k = P_k/||P_k||$ $(k \geq 0)$.
- Since $xP_k(x) \perp \{1, x, \dots, x^{k-2}\}$, the monic polynomials satisfy the recursion

$$xP_k(x) = P_{k+1}(x) + b_{k+1}P_k(x) + a_k^2 P_{k-1}(x),$$

  with recursion coefficients $b_k \in \mathbb{R}, a_k = ||P_k||/||P_{k-1}|| > 0$ $(k \geq 1)$.

Spectral measures

- The ONPs satisfy the recursion

$$xp_k(x) = a_{k+1}p_{k+1}(x) + b_{k+1}p_k(x) + a_k p_{k-1}(x)$$

  with recursion coefficients $b_k \in \mathbb{R}, a_k > 0$ $(k \geq 1)$. Furthermore, if $\operatorname{supp}(\mu) \subset [-K, K]$, then $|a_k|, |b_k| \leq K$.

- The *Jacobi matrix*

$$J = \begin{pmatrix} b_1 & a_1 & & \\ a_1 & b_2 & a_2 & \\ & a_2 & \ddots & \ddots \\ & & \ddots & \end{pmatrix}$$

  represents multiplication with x in the basis $\{p_k\}_{k\geq 0}$.

Spectral measures

- On the other hand, starting from a matrix

$$J = \begin{pmatrix} b_1 & a_1 & & \\ a_1 & b_2 & a_2 & \\ & a_2 & \ddots & \ddots \\ & & \ddots & \end{pmatrix}$$

with $a_k > 0$, $\sup_k(a_k + |b_k|) < \infty$, by the spectral theorem there is a *spectral measure* of the pair (J, e_1), satisfying

$$\int x^k \, \mathrm{d}\mu(x) = \langle e_1, J^k e_1 \rangle.$$

- The Jacobi matrix of μ is precisely J (seen by orthogonalizing $(J^k e_1)_{k \geq 0}$).

Spectral measures

- We have the one-to-one correspondence

$$\text{p.m. } \mu \text{ with bounded, infinite support}$$
$$\Leftrightarrow$$
$$\text{bounded sequences } (b_k, a_k)_{k\geq 1} \text{ with } a_k > 0, b_k \in \mathbb{R}.$$

- Obtain a_k, b_k from μ by OP recursion.
- Obtain μ from a_k, b_k by spectral theorem. Or by continued fraction expansion of the Stieltjes transform:

$$\int \frac{1}{z-x}\,\mathrm{d}\mu(x) = \langle e_1, (zI - J)^{-1} e_1 \rangle = \frac{1}{z - b_1 - \frac{a_1}{z - b_2 - \frac{a_2}{\cdots}}}$$

- A *sum rule* (coming soon) allows to relate information given in terms of a_k, b_k to spectral information.

Spectral measures

- Important example: The free Jacobi matrix

$$J_0 = \begin{pmatrix} 0 & 1 & & \\ 1 & 0 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & \end{pmatrix}$$

- Spectral measure is μ_{sc}, the semicircle law, given by

$$\mathrm{d}\mu_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\, \mathbb{1}_{[-2,2]}(x)\, \mathrm{d}x$$

- Moments are $m_{2k-1}(\mu_{\mathrm{sc}}) = 0$, $m_{2k}(\mu_{\mathrm{sc}})$ is the k-th Catalan number. Orthogonal polynomials are the Chebyshev polynomials (of the second kind).

Spectral measures

- Now assume that $|\text{supp}(\mu)| = n < \infty$.
- We can still define orthogonal monic polynomials $P_0, P_1, \dots, P_n$ and normalized polynomials $p_0, p_1, \dots, p_{n-1}$ (but $||P_n|| = 0$). This yields recursion coefficients $b_1, a_1, \dots, a_{n-1}, b_n$ with $a_k > 0$.
- On the other hand, μ is again the spectral measure of

$$J_n = \begin{pmatrix} b_1 & a_1 & & \\ a_1 & \ddots & \ddots & \\ & \ddots & & a_{n-1} \\ & & a_{n-1} & b_n \end{pmatrix}.$$

- We still have a one-to-one correspondence

$$\text{p.m. } \mu \text{ with } |\text{supp}(\mu)| = n$$
$$\Leftrightarrow$$
$$b_1, a_1, \dots, a_{n-1}, b_n \text{ with } a_k > 0, b_k \in \mathbb{R}.$$

The Killip-Simon sum rule

Killip, Simon (2003) prove the following *gem:*

- $J - J_0$ is Hilbert-Schmidt, that is, $\sum_{k=1}^{\infty} b_k^2 + (a_k - 1)^2 < \infty$,

if and only if

- $\mathrm{supp}(\mu) = [-2, 2] \cup E$ with E at most countable
- $\sum_{\lambda \in E} (|\lambda| - 2)^{3/2} < \infty$
- the decomposition $d\mu(x) = w(x)dx + d\mu_s(x)$ satisfies

$$\int_{-2}^{2} \log(w(x)) \sqrt{4 - x^2}\, dx > -\infty.$$

The Killip-Simon sum rule

Behind the gem is a *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.

The Killip-Simon sum rule

"Partial" sum rules have been obtained before under the assumption that $\text{supp}(\mu) \subset [-2, 2]$ (Szegő, Shohat, Geronimus, Krein, Nevai):

$J - J_0$ is Hilbert-Schmidt and $\sum b_k$ and $\sum (a_k - 1)$ are (conditionally) convergent if and only if a Szegő-condition holds.

Some observations:

- Both sides measure in some way a distance to μ_{sc}, resp. J_0.
- Both sides are simultaneously finite or infinite.
- LHS (and then also RHS) depends on the singular part μ_s only through its total mass!
- RHS (and then also LHS) does not depend on the order of the a_k or b_k!

The Killip-Simon sum rule: proof

- For $z \in \mathbb{D} = \{x \in \mathbb{C} : |x| < 1\}$, let $M_\mu(z) = \int \frac{1}{(z + z^{-1}) - x} \, \mathrm{d}\mu(x)$.
- M_μ is meromorphic on $\mathbb{D}$ with poles at β^{-1} when $\beta + \beta^{-1} \in E$. For almost every θ, M_μ has boundary value $M_\mu(e^{\mathrm{i}\theta}) = \lim_{r \to 1} M_\mu(re^{\mathrm{i}\theta})$.
- Let $L(z, J) = \det \left[((z + z^{-1}) - J)((z + z^{-1}) - J_0)^{-1} \right]$
- Case (1975) stated a series of sum rules $C_0, C_1, C_2, \dots$. KS prove them (for finite rank perturbations of J_0) using

$$L(z, J) = B(z) \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{\mathrm{i}\theta} + z}{e^{\mathrm{i}\theta} - z} \log |L(e^{\mathrm{i}\theta}, J)| \, \mathrm{d}\theta \right),$$

$$|L(e^{\mathrm{i}\theta}, J)|^2 \mathrm{Im} M_\mu(e^{\mathrm{i}\theta}) = \left(\prod_{k=1}^{\infty} a_k^2 \right) \sin\theta$$

$$\log L(z, J) = \sum_{n=1}^{\infty} -\frac{2}{n} z^n \, \mathrm{Tr}\left(T_n(\tfrac{1}{2} J) - T_n(\tfrac{1}{2} J_0) \right),$$

where T_n is the n-th Chebyshev polynomial (first kind).

The Killip-Simon sum rule: proof

- Taylor expansion of $\log L(z, J)$ at 0 yields the Case sum rules:

$$
\begin{aligned}
C_0: \quad & \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\left(\frac{\sin\theta}{\operatorname{Im} M_\mu(e^{i\theta})}\right) d\theta = \sum_j \log|\beta_j| - \sum_j \log|a_j| \\
C_n: \quad & \frac{-1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{\sin\theta}{\operatorname{Im} M_\mu(e^{i\theta})}\right) \cos(n\theta) d\theta + \frac{1}{n}\sum_j (\beta_j^n - \beta_j^{-n}) \\
& = \frac{2}{n}\operatorname{tr}\left(T_n(\tfrac{1}{2}J)) - T_n(\tfrac{1}{2}J_0)\right)
\end{aligned}
$$

- Key observation by KS: $C_0 + \frac{1}{2}C_2$ (the KS-SR!) has only non-negative terms.
- They removed the finite rank restriction and showed that LHS $\leq$ RHS and RHS $\leq$ LHS, even if one side is infinite.

The Gaussian unitary ensemble

- Let $X^{(n)}$ be a random complex Hermitian $n \times n$ matrix with distribution the Gaussian orthogonal ensemble (GUE) with density

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

w.r.t. Lebesgue measure in each entry on/above the diagonal. Since $\operatorname{Tr} X^2 = \sum_{i,j} |X_{i,j}|^2$, the entries $(X^{(n)}_{i,j})_{i \le j}$ are independent,

$$X^{(n)}_{i,i} \sim \mathcal{N}(0, \tfrac{1}{n}), \qquad \operatorname{Re}(X^{(n)}_{i,j}), \operatorname{Im}(X^{(n)}_{i,j}) \sim \mathcal{N}(0, \tfrac{1}{2n}) \ \ (i < j).$$

- Important property: the distribution is invariant under conjugations $X \mapsto UXU^*$ for unitary U.
- The normalization is precisely so that the *empirical eigenvalue measure*

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$$

(more about this later) converges in probability to the semicircle law

$$d\mu_{\mathrm{sc}}(x) = \frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbb{1}_{[-2,2]}(x)\, dx.$$

Random spectral measures: coefficients

- A.s., μ_n has n support points, so the first $2n-1$ recursion coefficients $b_1^{(n)}, a_1^{(n)}, \dots, a_{n-1}^{(n)}, b_n^{(n)}$ of μ_n are well-defined.
- Their distribution was found by Trotter (1984), generalized by Dumitriu, Edelman (2002).

Trotter (1984)

For μ_n as above, $b_1^{(n)}, a_1^{(n)}, \dots, a_{n-1}^{(n)}, b_n^{(n)}$ are independent and

$$b_k^{(n)} \sim \mathcal{N}(0, \tfrac{1}{n}) \quad \text{for } 1 \le k \le n,$$
$$(a_k^{(n)})^2 \sim \text{Gamma}(n-k, \tfrac{1}{n}) \quad \text{for } 1 \le k \le n-1.$$

Random spectral measures: coefficients

- To get the distribution of recursion coefficients, find a unitary H such that $H^* e_1 = e_1$ and $T = HXH^*$ is a Jacobi matrix, then

$$\langle e_1, T^k e_1 \rangle = \langle e_1, HX^k H^* e_1 \rangle = \langle e_1, X^k e_1 \rangle,$$

  so T and H have the same spectral measure and then the entries of T are the recursion coefficients.
- Will see $H = H_{n-1} \cdots H_1$
- To construct H_1, assume X is structured as

$$X = \begin{pmatrix} x_{1,1} & \tilde{x}^* \\ \tilde{x} & \tilde{X} \end{pmatrix}$$

  with $\tilde{x} \in \mathbb{C}^{n-1}$ and $\tilde{X}$ is $(n-1) \times (n-1)$.
- Choose $\tilde{H}$ unitary $(n-1) \times (n-1)$ such that $\tilde{H}\tilde{x} = ||\tilde{x}||e_1$ and let $H_1 = 1 \oplus \tilde{H}$.

Random spectral measures: coefficients

$$\text{Then} \qquad HXH^* = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{H} \end{pmatrix} \begin{pmatrix} x_{1,1} & \tilde{x}^* \\ \tilde{x} & \tilde{X} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \tilde{H}^* \end{pmatrix}$$

$$= \begin{pmatrix} x_{1,1} & \tilde{x}^* \tilde{H}^* \\ \tilde{H}\tilde{x} & \tilde{H}\tilde{X}\tilde{H}^* \end{pmatrix} = \begin{pmatrix} x_{1,1} & ||\tilde{x}|| & 0 & \dots \\ ||\tilde{x}|| & & & \\ 0 & & \tilde{H}\tilde{X}\tilde{H}^* & \\ \vdots & & & \end{pmatrix}.$$

- We can identify

$$\begin{aligned} x_{1,1} &= b_1 \sim \mathcal{N}(0, \tfrac{1}{n}), \\ ||\tilde{x}|| &= a_1 > 0 \text{ with } a_1^2 \sim \text{Gamma}(n-1, \tfrac{1}{n}) \\ \tilde{H}\tilde{X}\tilde{H}^* &\sim \text{GUE} \end{aligned}$$

  and all are independent.
- Now iterate, $H_2 = I_2 \oplus \hat{H}$, ...

LDP for coefficients

- Easy starting point: a single coefficient $b_k^{(n)} \sim \mathcal{N}(0, \frac{1}{n})$, since

$$\mathbb{P}(b_k^{(n)} \in A) = \int_A \frac{1}{\sqrt{2\pi/n}} e^{-nx^2/2}\, \mathrm{d}x = \exp\left\{ -n\left(\inf_{x \in A} \tfrac{1}{2}x^2 \right)(1 + o(1)) \right\}.$$

  So $(b_k^{(n)})_n$ satisfies the LDP with speed n and rate $\mathcal{I}_1(x) = \frac{1}{2}x^2$.

- Similarly: $(a_k^{(n)})_n$ satisfies the LDP with speed n and rate

$$\mathcal{I}_2(x) = \begin{cases} x^2 - \log(x^2) - 1 & x > 0 \\ \infty & \text{else.} \end{cases}$$

- By independence of $a_k^{(n)}, b_k^{(n)}$ and the general principles, $\left(b_1^{(n)}, a_1^{(n)}, b_2^{(n)}, \dots\right)_n$ satisfies the LDP with speed n and rate

$$\mathcal{I}(b_1, a_1, \dots) = \sum_{k=1}^{\infty} \mathcal{I}_1(b_k) + \mathcal{I}_2(a_k).$$

LDP for coefficients

- $\left(b_1^{(n)}, a_1^{(n)}, b_2^{(n)}, \dots\right)_n$ satisfies the LDP with speed n and rate

$$\mathcal{I}(b_1, a_1, \dots) = \sum_{k=1}^{\infty} \mathcal{I}_1(b_k) + \mathcal{I}_2(a_k).$$

  This LDP is in $(\mathbb{R} \times [0, \infty))^{\mathbb{N}}$ (with product topology). Not every element of this space is an admissible sequence of recursion coefficients.

- Let $\mathcal{J}_m$ be the set of recursion coefficients of measures supported by m points,

$$\mathcal{J}_m = \big\{(b_1, a_1, \dots) : a_k > 0 \text{ for } k \leq m, a_k = b_{k+1} = 0 \text{ for } k > m\big\}.$$

  The LDP can be restricted to

$$\mathcal{J} = \left(\bigcup_{m=1}^{\infty} \mathcal{J}_m\right) \cup \mathcal{J}_\infty.$$

- Then for each $(b_1, a_1, \dots) \in \mathcal{J}$, there is at least one measure μ with these coefficients.

LDP for coefficients

- The LDP for the coefficients can further be restricted to $\mathcal{J}^b$, the set of bounded sequences of recursion coefficients. They correspond to compactly supported measures, which are uniquely determined by the coefficients.
- On $\mathcal{J}^b$, the mapping $(b_1, a_1, \dots) \mapsto \mu$ is well-defined, continuous and injective. The contraction principle then leads to the following theorem.

Gamboa, Rouault (2011):

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

$$\mathcal{I}_{\mathrm{co}}(\mu) = \sum_{k=1}^{\infty} \tfrac{1}{2} b_k^2 + a_k^2 - 2\log(a_k) - 1.$$

- $\mathcal{I}_{\mathrm{co}}$ is infinite for finitely supported measures and vanishes for the measures with $b_k = 0, a_k = 1$, i.e., the semicircle law.

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 $b_1, a_1, \dots, a_{n-1}, b_n$, show that $(\mu_n)_n$ satisfies a large deviation principle with rate function $\mathcal{I}_{\mathsf{co}}$.
- Parametrizing μ_n by ev $\lambda_1, \dots, \lambda_n$ and weights $w_1, \dots, w_n$, 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.
- The Gaussian ensemble will yield the KS-SR. The Laguerre ensemble. . .

Laguerre sum rule

- Let μ be a measure with infinite, compact support in $[0, \infty)$.
- The recursion coefficients can be decomposed as

$$b_k = z_{2k-2} + z_{2k-1},$$
$$a_k^2 = z_{2k-1} z_{2k},$$

  where $z_k > 0$ and $z_0 = 0$.
- One-to-one correspondence between p.m. μ on $[0, \infty)$ with infinite, compact support and sequences $(z_k)_{k \geq 1}$ (bounded, positive).
- Central measure: Marchenko-Pastur law with parameter $\tau \in (0, 1]$

$$\mathrm{d}\mu_{\mathrm{MP}(\tau)}(x) = \frac{\sqrt{(\tau^+ - x)(x - \tau^-)}}{2\pi\tau x} \, \mathbb{1}_{(\tau^-, \tau^+)}(x) \, \mathrm{d}x, \quad \tau^\pm = (1 \pm \sqrt{\tau})^2$$

  with $z_{2k-1} = 1, z_{2k} = \tau$ for all $k \geq 1$.

Laguerre sum rule

Gamboa, N, Rouault (2016):

$$\mathcal{K}(\mu_{\mathrm{MP}(\tau)}|\mu) + \sum_{\lambda \in E} \mathcal{F}_L(\lambda) = \sum_{k=1}^{\infty} \tau^{-1} g(z_{2k-1}) + g(\tau^{-1} z_{2k}).$$

Where:

- $\mathrm{supp}(\mu) = [\tau^-, \tau^+] \cup E$, with E at most countable,
- $g(x) = x + \log(x) - 1$
- $\mathcal{F}_L(x) = \begin{cases} \int_{\tau^+}^{x} \frac{\sqrt{(t-\tau^-)(t-\tau^+)}}{t\tau} \mathrm{d}t & \text{if } x \geq \tau^+, \\ \int_{x}^{\tau^-} \frac{\sqrt{(\tau^- - t)(\tau^+ - t)}}{t\tau} \mathrm{d}t & \text{if } x \leq \tau^-. \end{cases}$

And both sides are simultaneously finite or infinite.

The Laguerre sum rule

The sum rule implies the following gem:
Let μ be a p.m. on $[0, \infty)$ with Jacobi coefficients decomposed into the $z_k > 0$, then

- $\sum_{k=1}^{\infty} (z_{2k-1} - 1)^2 + (z_{2k} - \tau)^2 < \infty$

if and only if

- $\operatorname{supp}(\mu) = [\tau^-, \tau^+] \cup E$, with E at most countable,
- $\sum_{\lambda \in E} d(\lambda, [\tau^-, \tau^+])^{3/2} < \infty$ and $0 \notin E$,
- the decomposition $\mathrm{d}\mu(x) = w(x)\mathrm{d}x + \mathrm{d}\mu_s(x)$ satisfies

$$\int_{\tau^-}^{\tau^+} \frac{\sqrt{(\tau^+ - x)(x - \tau^-)}}{x} \log(w(x))\mathrm{d}x > -\infty.$$

There is no equivalent ℓ^2-condition in terms of the a_k, b_k!

Thank you for your attention!

- R. Killip, B. Simon, *Sum rules for Jacobi matrices and their application to spectral theory*, Annals of Mathematics 158, 2003
- B. Simon, *Szegő's theorem and its descendants*, Princeton University Press, 2010
- F. Gamboa, A. Rouault, *Large deviations for random spectral measures and sum rules*, AMRX, 2011
- F. Gamboa, J. Nagel, A. Rouault, *Sum rules via large deviations*, Journal of Functional Analysis 270, 2016