Tuesday, 27 October 2020

Plan

Lecture 1

1. The β-log gas and the tridiagonal matrix models
2. Trotter's reduction from GUE/GOE to Jacobi form
3. The Jacobian determinant
4. Appl-1: Mehta integral
5. Appl-2: Wigner's semicircle law
6. Appl-3: Furedi Komlos
7. Exer: Laguerre and Jacobi matrices (to get Selberg integral?)

Lecture 2

1. Jacobi matrices and orthogonal polynomials
2. Szegö's theorem from local weak convergence
3. Kerov's theorem on asymptotic separation of roots
4. Plancherel diagram

Lecture 3

1. $n^{-2/3}$ at the edge. Statements
2. Upper bound for upper tail from Harer-Zagier, Ledoux, Jnaneshwar
3. Lower tail, via Dumaz-Virág, Ledoux-Rider,…
4. Operator limit and TW_β
5. Some heuristics

Lecture 4

1. Exponential LPP (last passage percolation)
2. Basic results and the question of LIL
3. Ledoux's result and ours
4. Sketch of the proof

Tuesday, 27 October 2020

Lecture 1

The β-log gas

Consider n particles $\lambda_1 < \ldots < \lambda_n$ on the real line with joint density proportional to

$$\exp\{-n^2\beta H(\lambda)\}d\lambda_1 \ldots d\lambda_n$$

where the energy $H = H_{n,\beta,V}$ given by

$$H(\lambda) := \frac{1}{n}\sum_k V(\lambda_k) - \frac{1}{n^2}\sum_{i<j}\log|\lambda_i - \lambda_j|.$$

The parameter $\beta > 0$. The potential function $V : \mathbb{R} \mapsto \mathbb{R}$ must grow fast at $\pm\infty$ so that the candidate density function is integrable. For example, any even degree polynomial with positive leading coefficient will do.

The first term in the energy keeps the particles from flying off to $\pm\infty$ and the second term keeps them from collapsing onto one another (*repulsion*).

- **Questions of interest as $n \to \infty$.** Where are the particles located? With what density? How are the particles separated from their neighbours? Where is the largest particle? What is its standard deviation? How is the second particle separated from the first?

Special cases of interest:

Hermite: $V(x) = x^2$ We refer to this as the *quadratic log gas*

Laguerre: $V(x) = \begin{cases} x + (a-1)\log x & \text{if } x \geq 0 \\ \infty & \text{if } x < 0 \end{cases}$ where $a > 0$

Jacobi: $V(x) = \begin{cases} (a-1)\log x + (b-1)\log(1-x) & \text{if } 0 \leq x \leq 1 \\ \infty & \text{otherwise} \end{cases}$

where $a, b > 0$

Tridiagonal matrices corresponding to the quadratic β-log gas

Let $V(x) = x^2/4$. Then the β-log gas has the same distribution as the eigenvalues of the random Jacobi matrix (symmetric tridiagonal with strictly positive entries on the sub/super-diagonal)

$$T_{n,\beta} := \frac{1}{\sqrt{\beta n}} \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & b_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & b_{n-2} & a_{n-1} & b_{n-1} \\ 0 & \dots & 0 & b_{n-1} & a_{n-1} \end{pmatrix}$$

where $a_k \sim N(0,2)$ and $b_k^2 \sim \chi^2_{\beta(n-k)}$ are all independent.

More precisely, the spectral measure of $T_{n,\beta}$ at the co-ordinate vector e_1 is given by $\mu^T_{e_1} := \sum p_k \delta_{\lambda_k}$ where $p = (p_1, \dots, p_n)$ and $\lambda = (\lambda_1, \dots, \lambda_n)$ are independent, λ has the quadratic β-log gas distribution and p has Dirichlet distribution with all parameters equal to $\beta/2$.

- Recall that the definition of the spectral measure of a Hermitian matrix A at a vector u is the measure μ^A_u whose pth moment is equal to $\langle A^p u, u \rangle$. If $A = \lambda_1 v_1 v_1^* + \dots + \lambda_n v_n v_n^*$ is the spectral decomposition of A, then explicitly,
$$\mu^A_u = \sum_{k=1}^{n} |\langle u, v_k \rangle|^2 \delta_{\lambda_k}.$$
- Key calculation: The Jacobian determinant for transforming $(a, b) \in \mathbb{R}^n \times \mathbb{R}^{n-1}_+$ to $(\lambda, p) \in \mathbb{R}^n \times \Delta_n$ (where Δ_n is the open probability simplex) is given by
$$\frac{b_1 b_2 \dots b_n}{2^{n-1} p_1 p_2 \dots p_n}.$$
- Also need the identity (valid for all Jacobi matrices):
$$b_1^{2(n-1)} b_2^{2(n-2)} \dots b_{n-1}^2 = p_1 p_2 \dots p_n \times \prod_{i<j} |\lambda_j - \lambda_i|^2.$$

Log-gas for general V

From the Jacobian determinant, it follows that if $(a, b) \in \mathbb{R}^n \times \mathbb{R}^{n-1}_+$ in $T_{n,\beta}$ has density proportional to

$$e^{-\beta n \mathrm{tr} V(T)} \prod_{k=1}^{n-1} b_k^{\beta(n-k)-1}$$

then in the spectral measure $\mu^T_{e_1}$, the variables p and λ are independent of each other, p has Dirichlet distribution (with all parameters equal to $\beta/2$) and λ has the $\beta-$log-gas density with potential V.

- If $V(x) = x^4$ then tr$V(T)$ has terms such as $b_{k-1}^2 b_k^2$ and $a_k^2 b_k^2$, $a_k a_{k+1} b_k^2$ etc. Thus, the matrix entries are not independent and this makes the analysis much harder.
- However, it is easy to see that $k \mapsto (a_k, b_k)$ forms a (time-inhomogenous) Markov chain, since there are no terms involving (a_{k-1}, b_{k-1}) and (a_{k+1}, b_{k+1}).
- More generally, if $V(x) = x^{2p} + \ldots$, then (a_k, b_k) have a Markovian structure with memory-length $p - 1$.

Laguerre ensemble

Let $B_k^2 \sim \chi^2_{\beta(n-k)}$ for $1 \le k \le n-1$and $A_k^2 \sim \chi^2_{2a+\beta(n-k)}$ for $1 \le k \le n$ be independent. Let T_n be the $2n \times 2n$ symmetric tridiagonal matrix with zeros on the diagonal and $A_1, B_1, A_2, B_2, \ldots, B_{n-1}, A_n$ on the sub-diagonal and super-diagonal. The eigenvalues of T_n are $\pm\sqrt{\lambda_k},\ 1 \le k \le n$, where λ is the Laguerre log$-$gas.

The particles are in a neighbourhood of $[-2{,}2]$

For the quadratic log-gas, we may use the associated tridiagonal matrix $T_{n,\beta}$. Observe that the eigenvalues of $T_{n,\beta}$ are bounded in absolute value by

$$\frac{1}{\sqrt{\beta n}}\left(\max_k |a_k| + 2\max_k b_k\right).$$

But by a simple union bound and the tails of the Gaussian and χ^2 distributions, we see that for any α there is a C large enough, such that with probability $1 - n^{-\alpha}$

$$\max_{1\le k\le n} a_k \le C\sqrt{\log n}, \qquad \max_{1\le k\le n} b_k \le \sqrt{\beta n}(1 + C\sqrt{\log n}).$$

Consequently, the spectrum of $T_{n,\beta}$ is contained in $[-2-\epsilon_n{,}2+\epsilon_n]$ with probability $\ge 1 - \delta_n$ where $\epsilon_n \asymp n^{-\frac{1}{2}}\sqrt{\log n}$ and $\delta_n \asymp n^{-\alpha}$.

Wigner's semicircle law

Let L_n be the empirical distribution of the particles in the β-log gas and let $\overline{L}_n$ denote the expectation of L_n. Then, $L_n, \overline{L}_n$ both converge to σ, which will denote the semicircle distribution with density $\frac{1}{2\pi}\sqrt{4-x^2}$ on $[-2,2]$.

Proof by method of moments: To a first approximation replace

$$\frac{1}{\sqrt{\beta n}} a_k \to 0 \qquad \frac{1}{\sqrt{\beta n}} b_k \to \sqrt{1-\frac{k}{n}}.$$

For fixed p and large n, we see that

$$\langle T_n^p e_1, e_1 \rangle = \sum_{i_2,\ldots,i_p} T(1,i_2)\ldots T(i_p,1) \approx \begin{cases} \binom{2q}{q}\frac{1}{q+1} & \text{if } p = 2q, \\ 0 & \text{if } p \text{ is odd.} \end{cases}$$

Thus $\mu_{e_1}^{T_n}$ converges in distribution to the *semi-circle distribution*

$$d\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\le 2}\, dx\,. \quad \blacksquare$$

On the other hand, for walks starting from k to k of length $2p$ each up/down step has weight approximately $1-\frac{k}{n}$ and the restriction that the path stays between 1 and n. Consequently, if $k \sim nt$ for some $0 < t < 1$, then

$$\langle T_n^{2q} e_k, e_k \rangle \approx \binom{2q}{q}(1-t)^q$$

which are the moments of $\xi\sqrt{1-t}$ where ξ has the arcsine distribution with density proportional to $(4-x^2)^{-1/2}$on $[-2,2]$.

> **Exercise:** For fixed k, show that the limiting moments of $\mu_{e_k}^T$ are zero for odd p and equal to $\binom{2q}{q}-\binom{2q}{q+k}$ for $p=2q$. What is the measure with these moments? Consider the special cases $k=1$ and $k\to\infty$.

Tuesday, 27 October 2020

As $L_n = \frac{1}{n}(\mu_{e_1}^{T_n} + \ldots + \mu_{e_n}^{T_n})$, one sees that $\int x^p dL_n(x) \approx \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[(\xi\sqrt{1-\frac{k}{n}})^p]$ which is $\approx \mathbb{E}[\xi^p]\int_0^1 (1-t)^{p/2}dt = \binom{2q}{q}\frac{1}{q+1}$ if $p=2q$ and zero if p is odd.

Thus we again get the semi-circle distribution, but as a superposition of scaled arcsine distributions (i.e., ξ, V are independent arc-sine and uniform, then ξV has semi-circle distribution).

Selberg integral

The Jacobian determinant computation which leads to the normalisation constant of the quadratic β-log gas gives us the identity

$$\int_{\mathbb{R}^n} \prod_{i<j} |x_i - x_j|^\beta \prod_{k=1}^{n} e^{-\frac{1}{2}x_k^2}\, dx = (2\pi)^{n/2} \prod_{k=1}^{n} \frac{\Gamma(1+\frac{\beta k}{2})}{\Gamma(1+\frac{\beta}{2})}.$$

This (sometimes known as Mehta's integral) is a limiting case of the three-parameter Selberg integral

$$\int_{[0,1]^n} \prod_{j<k} |x_j - x_k|^{2\gamma} \prod_{k=1}^{n} x_k^{\alpha-1}(1-x_k)^{\beta-1}\, dx_1 \ldots dx_n = \prod_{k=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(1+\gamma)\Gamma(\alpha+\beta+(n+j-1)\gamma)}$$

Mehta's integral can be proved from our change of variable formula. Indeed, the matrix entries are Gaussians and Chi random variables whose densities are explicit. The Jacobian determinant is also computed explicitly with all the constants. On the other side, the normalisation constant for the Dirichlet distribution (on $p = (p_1, \ldots, p_n)$ is known explicitly. From all this, it is clear that the normalisation constant for the quadratic beta log-gas can be obtained.

Can one also prove the Selberg integral? Yes, but from the tridiagonal matrix for the Jacobi ensembles. The Laguerre ensemble has two parameters and lies in between the two cases. It is a limiting case of Selberg integral and contains Mehta's integral as a limiting case.

Direct analysis of the log gas density?

While the random Jacobi matrix provides a useful and elegant approach to many aspects of the quadratic log-gas, some of the deductions can also be made directly from the joint density. This has the advantage of working for general V.

For example, here is how Wigner's semi-circle law can be derived. Observe that the exponent in the density can be written as

$$-\beta n^2 \left[\int V(x) dL_N(x) - \frac{1}{2} \int\int \mathbf{1}_{x\neq y} \log|x-y| \, dL_N(x) dL_N(y) \right].$$

From this, it is intuitively clear (but needs some argument) that the empirical distributions satisfy a large deviation principle with rate n^2 and the good rate function $\beta(I(\mu) - I(\sigma))$ where

$$I(\mu) := \int V(x) d\mu(x) - \int\int \log|x-y| \, d\mu(x) d\mu(y).$$

In particular, σ is the unique minimiser of I and the probability that the empirical distribution falls inside the set of measures μ with $I(\mu) > I(\sigma) + \epsilon$ decays like $e^{-\epsilon n^2}$. From this one can argue that the empirical distribution concentrates close to σ (in the Lévy metric on the space of probability measures on $\mathbb{R}$).

Notes

Trotter reduced the GUE/GOE matrices to tridiagonal form using Householder reflections and got the models $T_{n,\beta}$ with $\beta = 2/\beta = 1$. The realisation that one can generalise this to any β and consequently use this to study general β log gas is due to Dumitriu and Edelman. The model for general V was used in [KRV] to study universality in the edge. The large deviation principle is explained with full details in the book of Anderson-Guionnet-Zeitouni.

Zeros of orthogonal polynomials

Let μ be a probability measure on the real line that has all moments and is not supported on finitely many points. By applying Gram-Schmidt to $1, x, x^2, \ldots$ in $L^2(\mu)$, we get (normalized to be monic) polynomials $\phi_0, \phi_1, \phi_2, \ldots$ that are orthogonal with respect to μ. From the orthogonality, one can prove that there are numbers $a_n \in \mathbb{R}$, $b_n > 0$ such that $\phi_{n+1}(x) = (x - a_n)\phi_n(x) - b_{n-1}^2\phi_{n-1}(x)$ for $n \geq 0$ with the convention that $\phi_{-1} = 0$, $b_{-1} = 0$. Use these coefficients to construct the infinite tridiagonal matrix

$$T := \begin{pmatrix} a_0 & 1 & 0 & \ldots & \ldots & \ldots \\ b_0^2 & a_1 & 1 & \ldots & 0 & \ldots \\ 0 & b_1^2 & a_2 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

and let T_n denote the top-left $n \times n$ principal submatrix. The three term recurrences may be written as (with $\Phi(x) = (\phi_0(x), \phi_1(x), \ldots)^t$)

$$T\Phi(x) = x\Phi(x) \quad \forall x.$$

It is easy to see that if $\phi_n(x) = 0$, then (with $\Phi_n(x) = (\phi_0(x), \ldots, \phi_{n-1}(x))^t$)

$$T_n\Phi_n(x) = x\Phi_n(x).$$

Thus, the eigenvalues of T_n are precisely the zeros of ϕ_n. Further, from the three term recurrence, the zeros of ϕ_n interlace (strictly) with those of ϕ_{n-1}. Let $\zeta_{n,1} < \zeta_{n,2} < \ldots < \zeta_{n,n}$ denote the zeros of ϕ_n.

Question 1: How are the zeros of ϕ_n distributed?

Question 2: How are the zeros of ϕ_{n-1} interlaced with the zeros of ϕ_n? In other words, where is $\zeta_{n-1,k}$ located between $\zeta_{n,k-1}$ and $\zeta_{n,k}$?

Throughout, we assume that $b_n \sim n^\alpha$ for some $\alpha \geq 0$. For simplicity, we also assume that $a_n = 0$. Everything we say is literally correct if $a_n = o(b_n)$ but in cases where $a_n \asymp b_n$ modifications are needed to get the results. Special cases to keep in mind are Chebyshev polynomials ($\alpha = 0$) and Hermite polynomials ($\alpha = 1/2$). Laguerre polynomials are not covered by what we say because $a_n \asymp b_n \asymp n$.

Remark: The assumption that $a_n = 0$ is equivalent to μ being symmetric about 0, which is also equivalent to $\phi_n(-x) = (-1)^n \phi_n(x)$. Hence the zeros of the orthogonal polynomials are symmetrically placed about the origin.

Answer 1: (Szegö) Let L_n denote the empirical distribution of the scaled zeros $\frac{1}{b_n}\zeta_{n,k}$ for $1 \le k \le n$. Then the odd moments of L_n are zero and the even moments are given by

$$\int x^{2p} dL_n(x) = \frac{1}{n b_n^{2p}} \mathrm{Tr}[T_n^{2p}] \approx \frac{1}{b_n^{2p}} \sum_{k=1}^{n} b_k^{2p} \binom{2p}{p}$$

by approximating $T(i_1, i_2)\dots T(i_{2p}, i_1) \approx b_{i_1}^{2p}$. Then by the assumption on b_ks, this can be approximated by

$$\binom{2p}{p} \frac{1}{n} \sum_{k=1}^{n} (k/n)^{2\alpha p} \approx \binom{2p}{p} \frac{1}{1 + 2\alpha p}$$

by the usual approximation of the sum by the integral. It can also be written as $\mathbb{E}[(\xi V^\alpha)^{2p}]$ where ξ, V are independent arc-sine and uniform random variables, respectively. Thus $L_n \to \xi V^\alpha$ in distribution.

In conclusion, the unscaled zeros of $\phi_n$are mostly in $[-2b_n, 2b_n]$ and distributed within this interval according to these distributions, depending only on the parameter α.

Answer 2: (Kerov) Now we come to the second question. Fix n and let $u_k = \zeta_{n,k}/b_n$ and $v_k = \zeta_{n-1,k}/b_n$. Then $u_1 < v_1 < u_2 < \dots < v_{n-1} < u_n$. To study this, we introduce the *interlacing measure* $\tau_n = \sum_{k=1}^{n} \delta_{u_k} - \sum_{k=1}^{n-1} \delta_{v_k}$. Observe that there is no scaling by n. We compute the moments

$$\int x^{2p} d\tau_n(x) = \frac{1}{b_n^{2p}} \left(\mathrm{Tr}[(T_n)^{2p}] - \mathrm{Tr}[(T_{n-1})^{2p}] \right) = \frac{1}{b_n^{2p}} \sum_{i_1,\dots,i_{2p}} T(i_1, i_2)\dots T(i_{2p}, i_1)$$

where the last sum is over all $1 \leq i_1, \ldots, i_{2p} \leq n$ such that $i_{k+1} - i_k = \pm 1$ (including $i_{2p} - i_1 = \pm 1$) and such that at least one of the indices is equal to n (all other paths cancel). Since one of the indices is n, all indices are close to n, hence $T(i_1, i_2)\ldots, T(i_{2p}, i_1) \approx b_n^{2p}$. Consequently, the right hand side of the above display is (in the limit as $n \to \infty$) equal to the number of such tuples of indices. But that is just $\binom{2p}{p}$, since every walk on integers from the origin to itself in $2p$ steps can be lifted up in exactly one way so that the maximum is n. In conclusion τ_n converges to the arcsine measure!

How to interpret this in terms of the interlacing? For that purpose, Kerov introduced the *diagram* associated to τ_n as the function

$$\omega_n(x) = \sum_{k=1}^{n} |x - u_k| - \sum_{k=1}^{n-1} |x - v_{k-1}|$$

which can be called the *Riesz potential* of τ_n in the sense that $\frac{1}{2}\omega_n'' = \tau_n$ in the distributional sense. An alternate description of ω_n is that it is the function such that

- $\omega_n'(x) = +1$ on the intervals (u_k, v_k), (u_n, ∞) and
- $\omega_n'(x) = -1$ on the intervals (v_k, u_{k+1}), $(-\infty, u_1)$ and
- $\omega_n(x) = |x|$ on the intervals $(-\infty, u_1)$, (u_n, ∞).

From the fact that τ_n converges to the arcsine in moments, Kerov deduced that ω_n converges uniformly to

$$\Omega(x) = \begin{cases} \frac{2}{\pi}(x \arcsin(x/2) + \sqrt{4 - x^2}) & |x| \leq 2 \\ |x| & |x| > 2. \end{cases}$$

By direct calculation, one may check that $\Omega''(x)$ is the arcsine density. How does this transfer of results to diagrams help in interpretation of the interlacing? The idea is that if $v_k = \alpha u_k + (1-\alpha)u_{k+1}$ then on the interval

$$\omega(u_{k+1}) - \omega(u_k) = (2\alpha - 1)(u_{k+1} - u_k).$$

Thus, the slope of the diagram has the information about α, which tells the relative location of v_k in the interval (u_k, u_{k+1}). Of course, the result only captures this in a coarse, locally-averaged sense. The slope $\Omega'(x)$ gives the average of α for $k \approx x b_n$.

Remark: Kerov proved the theorem under more general conditions that

$$\frac{b_n}{b_{n+1}} \to 1, \quad \frac{a_{n+1} - a_n}{b_n} \to 0.$$

When $a_n = o(b_n)$ our proof can be carried through, but in the more general situation, one will have to also introduce a centering to get convergence to Ω.

Remark: Observe that the limit shape Ω (or equivalently the arcsine) does not depend on α. In this sense, this result is more universal than the one on limiting distribution of roots. The interlacing is the same for Chebyshev and Hermite polynomials, although the limiting results are different!

The largest point in the quadratic log-gas

Let us return to the quadratic log-gas and let λ_n denote the largest point. From what we have seen before (the range is within $[-2,2]$ and the limit is semi-circle), it follows that $\lambda_n \to 2$ in probability. Now we want to understand how big is $\lambda_n - 2$.

Heuristic: Since the empirical distribution of λ_ks is close to the semi-circle, perhaps λ_n is close to the $1 - \frac{1}{n}$ quantile of the semi-cirle. The latter can be calculated by setting $\frac{1}{n} = \frac{1}{2\pi} \int_{2-\epsilon_n}^{2} \sqrt{4 - x^2} dx$ which gives $\epsilon_n \asymp n^{-2/3}$, since the integral is essentially like $\int_0^{\epsilon_n} \sqrt{u} du = \epsilon_n^{3/2}$.

Precise results: We state results that not only give this $n^{-2/3}$ but also very precise estimates on the tails of $(\lambda_n - 2)n^{-2/3}$.

$$\mathbb{P}\{\lambda_n > 2 + tn^{-2/3}\} = e^{-(1+o(1))\frac{2\beta}{3}t^{3/2}}.$$
$$\mathbb{P}\{\lambda_n < 2 - tn^{-2/3}\} = e^{-(1+o(1))\frac{\beta}{12}t^{3}}.$$

That the upper and lower tails are of different exponential orders is the first feature - observe the powers $\frac{3}{2}$ on the right and the power 3 on the left. The constants in the exponent are optimal, as asserted by the $1 + o(1)$ factor.

Can we prove such estimates directly from the log-gas density? I have not seen such a proof, but like in large deviations, the idea would be to compare the optimal energy configurations under the deviant event with the unconstrained optimum (and also take into account the volumes of the events). This is naturally easier to do when the deviation is larger, for example if $t = \epsilon n^{2/3}$ where ϵ is fixed (done for the upper tail by Ben Arous—Dembo—Guionnet). What about using the GOE/GUE matrices for $\beta = 1/\beta = 2$? These matrices are unitarily invariant and give no useful information beyond the eigenvalues.

The tridiagonal matrix gives a natural approach to such a problem. Since the entries are larger at the top-left corner, it is natural to think that the eigenvector for the largest eigenvalue puts more weights on the first few entries. This is indeed true, and is used in the proofs below.

Dumaz and Virág proved these estimates in the limit, for the TW_β distribution (the limiting distribution of $(\lambda_n - 2)n^{2/3}$ but what we want is the result for finite n (this is like the difference in knowing the Gaussian tail and proving moderate deviation for sums of i.i.d. random variables). Estimates of this kind were obtained by Ledoux and Rider, but they did not care for the constants in the exponent, and we do. Our proof is an adaptation (discretization) of the Dumaz—Virág proof, with some modifications.

Proof of the upper bound on the lower tail

We now wish to prove that $\mathbb{P}\{\lambda_n \le 2 + tn^{-2/3}\} \ge e^{-(1+o(1))\frac{\beta}{12}t^3}$. By the variational formula

$$\lambda_n = \max_{v\neq 0} \frac{\langle T_n v, v\rangle}{\langle v, v\rangle}$$

It suffices to show that $\mathbb{P}\{\langle T_n v, v\rangle \le (2 + tn^{-2/3})\langle v, v\rangle\} \le e^{-(1+o(1))\frac{\beta}{12}t^3}$ for some fixed vector $v \in \mathbb{R}^n$. Because of the heuristic that was given earlier using the square-root vanishing of the semi-circle density at 2, we make the choice

$$v_k = \begin{cases} \sqrt{m-k} & k \le m, \\ 0 & k > m \end{cases}$$

for some m to be chosen later. It is also convenient to make the approximation $b_k = \sqrt{\beta(n-k)} + \xi_k$ where $\xi_k \sim N(0,\frac{1}{2})$, which is certainly valid for $k \ll n$. Then

$$\langle T_n v, v\rangle = \frac{1}{\sqrt{\beta n}}\sum_{k=1}^{n} a_k v_k^2 + \frac{2}{\sqrt{\beta n}}\sum_{k=1}^{n-1}(\sqrt{\beta(n-k)} + \xi_k)v_k v_{k+1}$$

which has a Gaussian distribution for any fixed v. For our choice of v,

- Mean is $2\sum_{k=1}^{m-1}\sqrt{1-\frac{k}{n}}\sqrt{m-k}\sqrt{m-k-1} \approx 2\sum_{k=1}^{m-1}(1-\frac{k}{2n})(m-k)$ which is to $m^2-\frac{m^3}{6n}$.
- Variance is $\frac{1}{\beta n}\sum_{k=1}^{m}(m-k)^2+\frac{4}{\beta n}\sum_{k=1}^{m-1}(m-k)(m-k-1)$ which is close to $\frac{4m^3}{3\beta n}$.
- Also $\langle v, v\rangle = \sum_{k=1}^{m}(m-k)$ which is essentially $\frac{m^2}{2}$.
- Hence $\frac{\langle T_n v, v\rangle}{\langle v, v\rangle} = 2-\frac{m}{3n}+\frac{4}{\sqrt{3\beta nm}}Z$ where $Z \sim N(0,1)$.

Thus, the right choice is $m = un^{\frac{1}{3}}$ for some u, which leads to the event

$$\mathbb{P}\left\{Z < \frac{\frac{u}{3}-t}{\frac{4}{\sqrt{3\beta u}}}\right\} \approx \exp\{-\frac{3}{32}\beta u(\frac{u}{3}-t)^2\}$$

by the usual estimate for the Gaussian tail. When optimised over $0 < u < 3t$, this leads to the choice $u = t$ and the estimate become $\frac{\beta}{24}t^3$ ∎

Remark: The lower bound on the upper tail would similarly be proved by choosing an appropriate test vector v (the square-root heuristic is not the right one, since the eigenvalues is going out of the support), since $\langle T_n v, v\rangle \geq (2+tn^{-2/3})\langle v, v\rangle$ would imply that $\lambda_n \geq 2+tn^{-2/3}$.

Upper bound on the upper tail

We could follow the Dumaz—Virág technique here, but we give a different but not self-contained proof, that also raises a question. We invoke a result of Ledoux's for $\beta = 2/\beta = 1$. We quote the result for $\beta = 1$.

Ledoux: Let A_n, B_n denote the $n \times n$ GOE/GUE matrix scaled by $\sqrt{n}$ so that the empirical distribution converges to the semi-circle. Then,

$$\mathbb{E}[\text{Tr}(A_n^{2p})] \leq C2^{2p}\left(1+\frac{p^{3/2}}{n}\right)e^{\frac{p^3}{3n^2}} \qquad \mathbb{E}[\text{Tr}(B_n^{2p})] \leq C2^{2p}e^{\frac{p^3}{12n^2}}.$$

We use this with the tridiagonal matrix to get the estimate to prove that for the tridiagonal matrix $T_{n,\beta}$ with $\beta \geq 1$,

$$\mathbb{E}[\text{Tr}(T_{n,\beta}^{2p})] \leq C2^{2p}\left(1+\frac{p^{3/2}}{n}\right)e^{\frac{p^3}{3\beta^2n^2}} \qquad ---(\#)$$

From this, the estimate follows directly, by taking $p = \beta\sqrt{t}n^{2/3}$ since (we write $2 + tn^{-2/3}$ as $2e^{\frac{t}{2n^{2/3}}}$ for convenience)

$$\mathbb{P}\{\lambda_n > 2e^{\frac{t}{2n^{2/3}}}\} \leq \mathbb{E}[\text{Tr}(T_{n,\beta}^{2p})]2^{-2p}e^{-\frac{2pt}{2n^{2/3}}} \leq C\beta^{3/2}t^{3/4}e^{\frac{p^3}{3\beta^2n^2}-\frac{2pt}{2n^{2/3}}}$$

By the choice of p, the exponent is precisely $-\frac{2}{3}\beta t^{\frac{3}{2}}$ as required.

Proof of the claim (#) [B. S. Jnaneshwar] The idea is to compare $T_{\beta,n}$ with $T_{1,\beta n}$. Observe that the diagonal entries have the same distribution in both. The $(k, k+1)$ entry (for $k \leq n-1$) are $\frac{1}{\sqrt{\beta n}}\chi_{\beta(n-k)}$ in the first case which is stochastically smaller than $\frac{1}{\sqrt{\beta n}}\chi_{\beta n-k}$ which is the corresponding entry in the second case. To use this comparison, we use the fact that $\mathbb{E}[(T_{\beta,n}^p)_{1,1}] = \frac{1}{n}\mathbb{E}[\text{Tr}(T_{\beta,n}^p)]$ (because of the independence of the eigenvalues and the weights in the spectral measure $\mu_{e_1}^T$). Hence it suffices to compare $\mathbb{E}[(T_{\beta,n}^p)_{1,1}]$ with $\mathbb{E}[(T_{1,\beta n}^p)_{1,1}]$ for $p \leq n$. Expand both of these in paths on $\mathbb{Z}_+$ (with self-loops) from origin to itself. The paths that use any loop an odd number of times vanishes. For other paths, the weight in the first matrix is smaller than in the second. Therefore, $\mathbb{E}[(T_{\beta,n}^p)_{1,1}] \leq \mathbb{E}[(T_{1,\beta n}^p)_{1,1}]$ which together with Ledoux' bound for GOE gives the claim-(#). ∎

The Tracy-Widom distributions

From the moderate deviation estimates, we see that the distributions of $(\lambda_n - 2)n^{2/3}$ are tight and hence have sub sequential limits. It is natural to expect that they actually converge to a probability measure. We give a description of this limiting distribution here.

As before, write $b_k = \sqrt{\beta(n-k)} + \xi_k$ where $\xi_k \sim N(0,\frac{1}{2})$. This will be okay as we shall only use $k \leq n^{\frac{1}{3}+\epsilon}$. Also write $\sqrt{\beta(n-k)} = \sqrt{\beta n}(1-\frac{k}{2n})$ expanding $\sqrt{1-x}$ to the first order. Then $n^{2/3}(T_{n,\beta} - 2I_n) = L_n + M_n + W_n$ where

$$L_n = n^{2/3}\begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & 1 & \ddots & \ddots \end{pmatrix}$$

$$M_n = \frac{1}{2n^{1/3}}\begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 2 & & \\ & 2 & 0 & 3 & \\ & & 3 & \ddots & \ddots \end{pmatrix}$$

$$W_n = \frac{n^{1/6}}{\sqrt{\beta}}\begin{pmatrix} a_1 & \xi_1 & & & \\ \xi_1 & a_2 & \xi_2 & & \\ 0 & \xi_2 & a_3 & \xi_3 & \\ & & \xi_3 & \ddots & \ddots \end{pmatrix}$$

Think of a vector $v \in \mathbb{R}^n$ as a discretisation of $f : \mathbb{R}_+ \mapsto \mathbb{R}$ by $v(k) = f(k/n^{1/3})$. Then

$$L_n f(k/n^{1/3}) \approx f''(k/n^{1/3}) \qquad M_n f(k/n^{1/3}) \approx \frac{k}{n^{1/3}} f(k/n^{1/3})$$

$$\sum_{k=1}^{\ell} W_n f(k/n^{1/3}) = \frac{n^{1/6}}{\sqrt{\beta}} \sum_{k=1}^{\ell} (\xi_{k-1} f((k-1)/n^{1/3}) + a_k f(k/n^{1/3}) + \xi_{k+1} f((k+1)/n^{1/3})$$

$$\approx \frac{2}{\sqrt{\beta}} \sum_{k=1}^{\ell} n^{1/6} Z_k f(k/n^{1/3}) \quad \text{(where } Z_k \sim N(0,1) \text{ are i.i.d.)}$$

$$\approx \frac{2}{\sqrt{\beta}} \int_0^x f(t) dW(t)$$ where $\ell = xn^{1/3}$ and W is standard Brownian motion. Putting these together, we see that $n^{2/3}(T_n - 2I_n)$ acts like

$$H_\beta f(x) = f''(x) + xf(x) + \frac{2}{\sqrt{\beta}} f(x) dW(x).$$

- Ramirez-Rider-Virág showed that H_β, defined on an appropriate space of functions, almost surely has a purely discrete spectrum $\theta_1 > \theta_2 > \ldots$ bounded above
- The eigenvalues $n^{2/3}(\lambda_n - 2)$ converge in distribution to θ_1
- The distribution of θ_1 is called TW_β. It is a non-symmetric distribution that has tails $\mathbb{P}\{TW_\beta > x\} = e^{-(1+o(1))\frac{2\beta}{3}x^{3/2}}$ and $\mathbb{P}\{TW_\beta < -x\} = e^{-(1+o(1))\frac{\beta}{12}x^3}$

Universality

The randomness was used in the above discussion in writing

$\frac{n^{1/6}}{\sqrt{\beta}} \sum_{k=1}^{\ell} (\xi_{k-1} f((k-1)/n^{1/3}) + a_k f(k/n^{1/3}) + \xi_{k+1} f((k+1)/n^{1/3})$ as $\frac{2}{\sqrt{\beta}} \int_0^x f(t) dW(t)$, which happens if the partial sums of ξ_k, a_k converges to Brownian motion. Hence the key thing to prove convergence to Tracy-Widom distribution is to prove that

$$\sum_{k=1}^{tn^{1/3}} (a_k + 2\xi_k) \to 2W(t) \text{ in distribution.}$$

This can be done for polynomial V that are uniformly convex.