Statistics 291: Lecture 7 (February 13, 2024)

Kac-Rice III: Second Moments and E_∞ Threshold

Instructor: Mark Sellke

Scribe: Rushil Mallarapu

1 Determinant Bounds for $\mathrm{GOE}(N)$ Matrices

To fix notation, let

$$\mathrm{Crt}_{S_N}(H_{N,p};[a,b]) = \left\{x \in S_N \mid \nabla_{\mathrm{sph}} H_{N,p}(x) = 0 \text{ and } H_{N,p}(x)/N \in [a,b]\right\} \tag{1}$$

denote the set of critical points with (dimensionless) energy $H_{N,p}(x)/N$ within a specified interval. Last lecture, we derived a formula for the annealed number of critical points with specified energy:

Theorem 1.1 (Annealed exponential growth rate for the number of critical points)**.** *For $-\infty \le a < b \le \infty$,*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\left|\mathrm{Crt}_{S_N}(H_{N,p};[a,b])\right| = \max_{E\in(a,b)} \left\{\frac{\log(p-1)}{2} - E^2\left(\frac{p-2}{4p-4}\right) + 1_{|E|\ge 2\sqrt{\frac{p-1}{p}}}\,\theta\left(E\sqrt{\frac{p}{p-1}}\right)\right\}.$$

Check the previous lecture notes for a definition of θ, and define $\Phi(E)$ to be the function in braces in Theorem 1.1. This theorem relied on an unproven bound on the expected determinant of shifted $\mathrm{GOE}(N)$ matrices, and last time we left off with proving the upper bound. In particular, we want to prove

$$\mathbf{E}\left[|\det(\mathrm{GOE}(N) - tI_N)|\right] \le \exp\left(N\psi(t) + o(N)\right). \tag{2}$$

Our goal is to use concentration of measure of the eigenvalues. We left of last time by showing:

Lemma 1.2 (Hoffman-Wielandt Inequality)**.** *Let $A_N, \tilde{A}_N$ be $N \times N$ symmetric matrices with eigenvalues $\lambda_1 \ge \dots \ge \lambda_N$, $\tilde{\lambda}_1 \ge \dots \tilde{\lambda}_N$ respectively. Then*

$$\sum_{i=1}^N |\lambda_i - \tilde{\lambda}_i|^2 \le \left\|A_N - \tilde{A}_N\right\|_F^2 = \sum_{i,j=1}^N (A_N - \tilde{A}_N)_{ij}^2.$$

As a recap, the idea was to fix the spectrum and try to rotate the eigenvectors. This has some useful corollaries:

Corollary 1.3. *If $f : \mathbf{R} \to \mathbf{R}$ is L-Lipschitz, $1 \le k \le N$, then the map*

$$A_N \mapsto \frac{1}{k}\sum_{i=1}^k f(\lambda_i)$$

is $L/\sqrt{k}$-Lipschitz.

Proof. Let $A_N, \tilde{A}_N$. Then

$$\frac{1}{k}\left|\sum_{i=1}^{k} f(\lambda_i) - f(\tilde{\lambda}_i)\right| \le \frac{L}{k}\sum_{i=1}^{k} |\lambda_i - \tilde{\lambda}_i|$$
$$\le L\sqrt{\frac{1}{k}\sum_{i=1}^{k} |\lambda_i - \tilde{\lambda}_i|^2}$$
$$\le \frac{L}{\sqrt{k}}\|A_N - \tilde{A}_N\|_F,$$

where the second-to-last inequality is Cauchy-Schwarz, and the final is Lemma 1.2. Note that we could also use any subset $i_1, \ldots, i_k$ of indices. □

Corollary 1.4. *Suppose f is L-Lipschitz, $A_N \sim \mathrm{GOE}(N)$, and let*

$$W_k = \frac{1}{k}\sum_{i=1}^{k} f(\lambda_i).$$

Then $\mathbb{P}(|W_k - \mathbf{E}W_k| \ge \epsilon) \le 2\exp\left(-\frac{Nk\epsilon^2}{16L^2}\right)$.

Proof. This should look like Lipschitz concentration! More precisely, consider $X \sim \mathcal{N}\left(0, I_{\binom{N+1}{2}}\right)$, i.e. a $\binom{N+1}{2}$-dimensional vector of i.i.d. standard Normals, and look at the map $\phi\colon \mathbf{R}^{\binom{N+1}{2}} \to \mathbf{R}^{N\times N}$ given by taking these coordinates and arranging them into a symmetric matrix (where we multiply the diagonal entries by $\sqrt{2}$). Then $\phi(X)/\sqrt{N} \sim \mathrm{GOE}(N)$. In fact, this is an isometry from $X \mapsto \mathrm{GOE}(N)$ with Lipschitz constant $\sqrt{2/N}$. Then, as the map from $\mathrm{GOE}(N)$ to W_k is $L/\sqrt{k}$-Lipschitz, so the composite map is $L\sqrt{2/Nk}$-Lipschitz, and Lipschitz concentration lets us conclude. □

It remains to prove the upper bound (2):

Proof. First, fix small $\epsilon > 0$, and upper bound $\log|x|$ by

$$\log_\epsilon|x| = \max\{\log|x| - 10\log(1/\epsilon)\}.$$

This "truncates" $\log|x|$, and usefully, has Lipschitz constant $10/\epsilon$. Now, we want to integrate with respect to the semicircle density, so it's helpful to note/take on faith that

$$\int_{\mathbf{R}} \log_\epsilon|u| - \log|u|\,\mathrm{d}u \le \epsilon.$$

In particular, since ν_{SC} has density at most 1 everywhere,

$$\int \log_\epsilon|u-t|\,\mathrm{d}\nu_{SC}(u) \le \epsilon + \int \log|u-t|\,\mathrm{d}\nu_{SC} = \epsilon + \psi(t)$$

for all $t \in \mathbb{R}$ with $\psi(t)$ defined as in last lecture. Now, since $\log_\epsilon$ is Lipschitz, we know:

(a) with high probability, by Wigner's semicircle law, we have

$$\frac{1}{N}\sum_{i=1}^{N} \log_\epsilon|\lambda_i - t| \le \epsilon + \int \log_\epsilon|u-t|\,\mathrm{d}\nu_{SC}(u) \le 2\epsilon + \psi(t),$$

as the empirical distribution converges to the semicircle distribution in the bounded Lipschitz (BL) metric;

(b) $\frac{1}{N}\sum_{i=1}^{N}\log_\epsilon|\lambda_i - t|$ concentrates with exponent N^2, which is really good! Thus, we can basically take expectations as if this was a constant. By contrast, the free energy $F_N(\beta)$ concentrates with an $O(N)$ exponent, so the behavior isn't as nice.

In particular, we have for any α

$$\mathbb{P}\left(|\det(\mathrm{GOE}(N) - tI_N| \geq e^{N(\psi(t)+2\epsilon+\alpha}\right) \leq \mathbb{P}\left(\frac{1}{N}\sum\log_\epsilon|\lambda_i - t| \geq \psi(t) + 2\epsilon + \alpha\right) \leq 2\exp\left(-\frac{\epsilon^2 N^2\alpha^2}{10^4}\right).$$

Ergo, we can conclude that

$$\begin{aligned}\mathbf{E}|\det(\mathrm{GOE}(N) - tI_N)| &\leq e^{N(\psi(t)+2\epsilon)} + \int_{e^{N(\psi(t)+2\epsilon)}}^{\infty}\mathbb{P}(|\det(\cdot)| \geq z)\,\mathrm{d}z \\ &= e^{N(\psi(t)+2\epsilon)}\left(1 + \int_0^\infty \mathbb{P}\left(|\det(\cdot)| \geq e^{N(\psi(t)+2\epsilon+\alpha)}\right) Ne^{N\alpha}\,\mathrm{d}\alpha\right).\end{aligned}$$

We get the second line by substituting $z = e^{N(\psi(t)+2\epsilon+\alpha)}$. Now, because we have such good bounds on this probability, we can bound it by

$$\leq 2\int_0^\infty Ne^{N\alpha}\exp\left(-\frac{\epsilon^2N^2\alpha^2}{10^4}\right)\mathrm{d}\alpha = 2\int_0^\infty \exp\left(\gamma - \frac{\epsilon^2\gamma^2}{10^4}\right)\mathrm{d}\gamma = O(1) \leq e^{o(N)},$$

which lets us conclude. (Here the last $O(1)$ depends on ϵ but not N.) $\square$

2 Ground-State Energies

Now that we've proved Theorem 1.1, up to the matching determinant lower bound as (2) which we'll take on faith, let's use it to understand the maximum energy of the Hamiltonian, the so-called ground-state energy:

Definition 2.1. The *ground-state energy* is

$$\mathrm{GS}_N = \frac{1}{N}\max_{x\in S_N} H_{N,p}(x).$$

Let's consider Φ from Theorem 1.1. With some thinking, we see

(a) Φ is even, with $\Phi(0) = \log(p-1)/2$.

(b) Φ is concave, as θ is very concave, and $\Phi(x) \to -\infty$ as $x \to \infty$.

Taken together, this implies Φ has a unique positive zero E_0. This is a natural threshold for guessing the ground-state energy: if we look at

$$\frac{1}{N}\log\mathbf{E}\left|\mathrm{Crt}_{S_N}(H_{N,p};[E_0+\epsilon,\infty))\right| = \max_{E\geq E_0+\epsilon}\Phi(E),$$

we see this maximum is strictly negative for small $\epsilon > 0$. Thus, by Markov, with high probability (i.e. $1-e^{-cN}$, for c depending on p,ϵ), there are no critical points above $E_0+\epsilon$, so $\mathrm{GS}_N \leq E_0+\epsilon$.

How can we get a matching lower bound? Look at 2nd moments! To compute something like

$$\mathbf{E}\left|\mathrm{Crt}_{S_n}(H_{N,p};[E_0,\infty)\right|^2,$$

we'd have to use Kac-Rice on a product of 2-spheres. This is a bit nasty, as if we fix x, y, the Gaussian-like r.v.s we had

$$H_{N,p}(x), \nabla_{\mathrm{sph}}H_{N,p}(x), \nabla^2_{\tan}H_{N,p}(x),$$

are no longer independent between x, y. However, while there are annoying correlations between, say, $\nabla_{\text{sph}} H_{N,p}(x)$ and $\nabla_{\text{sph}} H_{N,p}(y)$, it's not mathematically harder. What is hard is dealing with products of determinants, and having to compute something about correlated random matrices. It turns out that this also isn't too bad, as determinants concentrate so well that you can pretend they're constants, and commute them with expectations.

More precisely, Subag (2015), by being really careful with constants, showed

$$\frac{\mathbf{E}\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right|^2}{\mathbf{E}\left[\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right|^2\right]} = 1 + o(1). \tag{3}$$

This is fairly hard, but showing that the LHS of (3) is upper bounded by $e^{o(N)}$, which is sufficient for us, is basically a 3D calculus problem, involving the x energy, y energy, and overlap $R(x, y)$. Anyway, assuming this weaker bound, we can show

Proposition 2.2. *With notation as above,* $\lim_{N\to\infty} \mathbf{E}\,\text{GS}_N = E_0$. *More specifically,* $\mathbf{E}\,\text{GS}_N \geq E_0 - o(1)$.

Proof. The key fact here is that while the critical points don't concentrate, GS_N does, by e.g. Borell-TIS. So we know

$$\mathbb{P}(|\text{GS}_N - \mathbf{E}\,\text{GS}_N| \geq \epsilon) \leq e^{-\epsilon^2 N/8}.$$

Now, by Paley-Zygmund, we have

$$\begin{aligned}\mathbb{P}\left(\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right| \geq 1\right) &\geq \mathbb{P}\left(\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right| \geq 1/2\right)\\ &\geq \frac{1}{4}\frac{\mathbf{E}\left[\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right|^2\right]}{\mathbf{E}\left|\text{Crt}_{S_N}(H_{N,p};[E_0,\infty))\right|^2} \geq e^{-o(N)}.\end{aligned}$$

Thus, $\mathbb{P}(\text{GS}_N \geq E_0) \geq e^{-o(N)}$; the Hamiltonian has a global maximum, which is a critical point, so the ground state energy having energy above E_0 is the same as there being a critical point with energy above E_0! From here, we conclude as in the proof of the annealed free energy $F_N(\beta)$ case. □

In general, 2nd moment methods for extremal critical points works without any β threshold thing, as we saw previously. This fact should be thought of as 1-RSB behavior; there's no multilayer clustering structure!

3 Landscape Properties Beyond Maximum Values

Let $E_\infty := 2\sqrt{\frac{p-1}{p}}$, and let's look at local maxima of $H_{N,p}(x)$:

Theorem 3.1. *(a) For $E < E_\infty$, there are no local maxima:*

$$\mathbf{E}[\textit{\# of local maxima below } E] \leq e^{-cN^2}.$$

(b) For $E' \geq E \geq E_\infty$, we instead have

$$\mathbf{E}[\textit{\# of local maxima in } [E,E']] \geq (1-e^{-cN})\mathbf{E}\left|\text{Crt}_{S_N}(H_{N,p};[E,E'])\right|.$$

This theorem suggests E_∞ is a threshold for optimization algorithms, in the sense that there's a fundamental difficulty in things like gradient descent finding E_0 without getting stuck in local maxima, as above E_∞, most of the critical points are local maxima. This intuition is true, but is nontrivial, and we'll return to it later in the class.

The idea here is to show that the Kac-Rice formula gets smaller when we put conditions on the critical points. In particular, let $I \subseteq \mathbf{R}^{N-1}$ be an event for eigenvalues of $\nabla^2_{\text{sph}} H_{N,p}$. For example, for $x \in S_N$ a critical point, we know x is a local maximum iff all the eigenvalues $\lambda_i(\nabla^2_{\text{sph}} H_{N,p}(x)) < 0$ are negative, so in that case we'd take $I = (-\infty, 0)^{N-1}$.

Definition 3.2. Let $\mathrm{Crt}_{S_N}^{(I)}(H_{N,p};[a,b])$ denote the set of critical points of $H_{N,p}$ on $x \in S_N$ with

(a) $H_{N,p}(x)/N \in [a,b]$, and

(b) $(\lambda_1,\dots,\lambda_{N-1}) \in I$, for $\lambda_1 \geq \cdots \geq \lambda_{N-1}$ the eigenvalues of $\nabla^2_{\mathrm{sph}} H_{N,p}(x)$.

Proposition 3.3. *With notation as above, we have*

$$\mathbf{E}\left|\mathrm{Crt}_{S_N}^{(I)}(H_{N,p};[a,b])\right| \leq \max_{E\in[a,b]}\left\{\Phi(E)+\frac{1}{2}\log\mathbb{P}(I|E)\right\}$$

Proof. Apply Cauchy-Schwarz inside Kac-Rice:

$$\mathbf{E}\left[|\det(\mathrm{GOE}(N)-tI_N)|\cdot 1_{I(\mathrm{GOE}(N)-tI_N)}\right] \leq \mathbf{E}\left[\det(\mathrm{GOE}(N)-tI_N)^2\right]^{1/2}\sqrt{\mathbb{P}(I\mid E)}.$$

Note that the L^2 norm of the determinant is still $\exp(N\psi(t)+o(N))$; it's effectively the same proof except with an extra 2, but fundamentally it boils down to having really good concentration. Meanwhile, the square root gives the $1/2$ in the proposition; you can improve it to anything <1 by Hölder, but it's unnecessary. □

Proof of Theorem 3.1. In each case, we consider $\det(\sqrt{p(p-1)}\,\mathrm{GOE}(N-1)-pE)$. Then for $E<E_\infty$, the distribution of eigenvalues of this scaled GOE, shifted by $-pE$, cross 0, while for $E>E_\infty$, they don't. More precisely:

(a) Take
$$I=\left\{\frac{1}{\delta N}\sum_{i=1}^{\delta_N}\lambda_i \leq 0\right\}$$
for the top δN eigenvalues, where we pick δN so this is the fraction of positive eigenvalues. (Thus $\delta>0$ is a constant depending on E.) This probability is super small:
$$\mathbb{P}(\lambda_1\leq 0)\leq \mathbb{P}(I\mid E)\leq e^{-cN^2}.$$
Note that there are exponentially many $e^{\Omega(N)}$ total critical points to consider on average, but the e^{-cN^2} completely swamps this.

(b) Take $I=\{\lambda_1\geq 0\}$, which has exponentially small probability:
$$\mathbb{P}(I\mid E)\leq e^{-cN}.$$
This follows because typical GOE matrices have no outlier eigenvalues, and from the $k=1$ case of Corollary 1.4, we know λ_1 concentrates exponentially!

□

Next time, we'll discuss topological trivialization, where we introduce an external field to the p-spin Hamiltonian. We'll see that for large external field energy, there are only 2 critical points, but for lower fields, there will be many!