The circular law

Lewis Memorial Lecture / DIMACS minicourse

March 19, 2008

Terence Tao (UCLA)

Eigenvalue distributions

Let $M = (a_{ij})_{1 \leq i \leq n; 1 \leq j \leq n}$ be a square matrix. Then one has n (generalised) eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$. We are interested in the case when the a_{ij} are iid random variables (e.g. Gaussian or Bernoulli). We would also like to make as few moment assumptions as possible (ideally, just second moment).

We would like to understand the asymptotic distribution of the spectrum $\{\lambda_1, \ldots, \lambda_n\}$ in the limit $n \to \infty$. (Note that there is no natural ordering of the n eigenvalues.)

The first task is to get a rough idea of how large the eigenvalues are.

If the a_{ij} are bounded (or at least have bounded second moment), then $\mathrm{tr}(MM^*) = \sum_{1 \leq i,j \leq n} |a_{ij}|^2 = \sum_{j=1}^{n} \sigma_j^2$ has size $O(n^2)$ w.h.p. (law of large numbers). Thus we expect the singular values σ_j to be of magnitude $O(\sqrt{n})$ on average.

It turns out that there is a linear algebra inequality

$$\sum_{j=1}^{n} |\lambda_j|^2 \leq \sum_{j=1}^{2} \sigma_j^2$$

which then implies that the eigenvalues λ_j also have magnitude $O(\sqrt{n})$ (at least in a root mean square sense).

Proof of linear algebra inequality:

- By the Jordan normal form, we can write $M = AUA^{-1}$ where U is upper triangular. By the QR factorisation, we can write $A = QR$ where Q is orthogonal and R is upper triangular. We conclude that $M = QVQ^{-1}$ where V is upper triangular.
- The eigenvalues λ_j are the diagonal entries of V.
- Meanwhile, the sum $\sum_{j=1}^{2} \sigma_j^2 = \mathrm{tr}(MM^*) = \mathrm{tr}(VV^*)$ is the sum of squares of all the elements of V (including the ones above the diagonal). The claim follows.

If one assumes fourth moment control on the a_{ij}, one can obtain $\sigma_1 = O(n^{1/2})$ with high probability. From the linear algebra inequality

$$\sup_{1 \leq j \leq n} |\lambda_j| \leq \|M\| = \sigma_1$$

we thus obtain $\lambda_j = O(n^{1/2})$ for all j (not just most j). But the previous argument is more general (it only needs second moment), and for the task of controlling limiting eigenvalue distributions, it is acceptable to lose control of a minority of the eigenvalues.

Since the eigenvalues are expected to have size $O(\sqrt{n})$, it is natural to introduce the normalised spectrum $\{\frac{1}{\sqrt{n}}\lambda_1, \ldots, \frac{1}{\sqrt{n}}\lambda_n\}$. We then define the empirical spectral distribution (ESD) $\mu_n : \mathbf{R}^2 \to [0,1]$ by the formula

$$\mu_n(s,t) := \frac{1}{n}\#\{1 \leq j \leq n : \mathrm{Re}(\frac{1}{\sqrt{n}}\lambda_n) \leq s; \mathrm{Im}(\frac{1}{\sqrt{n}}\lambda_n) \leq t\}.$$

This is a sequence (in n) of (random) functions on the plane.

Question: Does the sequence μ_n converge to a limit as $n \to \infty$? If so, what is this limit?

Let $\mu : \mathbf{R}^2 \to [0,1]$ be a (deterministic) function.

Definition (Weak convergence) We say that μ_n converges weakly to μ if for every $s, t \in \mathbf{R}$, $\mu_n(s,t)$ converges to $\mu(s,t)$ in probability, i.e.

$$\lim_{n\to\infty} \mathbf{P}(|\mu_n(s,t) - \mu(s,t)| \geq \varepsilon) = 0 \text{ for all } \varepsilon.$$

Definition (Strong convergence) We say that μ_n converges strongly to μ if, with probability 1, the sequence μ_n converges uniformly to μ.

To oversimplify, weak convergence is asking for

$$\mathbf{P}(|\mu_n(s,t) - \mu(s,t)| \geq \varepsilon) = o(1)$$

while strong convergence is essentially asking for

$$\sum_{n=1}^{\infty} \mathbf{P}(|\mu_n(s,t) - \mu(s,t)| \geq \varepsilon) < \infty$$

(cf. the Borel-Cantelli lemma). Thus one needs quantitative tail bounds (better than $O(1/n)$) for strong convergence, but only qualitative tail bounds $o(1)$ for weak convergence.

[The uniformity in s, t is relatively cheap to attain, due to the monotonicity properties of μ_n.]

Wigner's semicircular law

The situation is very well understood in the self-adjoint case, in which the matrix entries a_{ij} are only iid for $i \leq j$ (and extended to $i > j$ by self-adjointness). In this case, of course, the spectrum is real (and $\mu_n(s,t)$ is really just a function of s.

The famous semi-circular law of Wigner then asserts that μ_n converges strongly to the semicircular distribution $\mu(s) = \int_{-\infty}^{s} \, d\mu$, where $\mu := \frac{1}{2\pi}(1 - s^2/4)_+^{1/2} \, ds$, at least in the case of the Gaussian ensemble.

It can be extended to the case when a_{ij} are independent for $i \leq j$, have mean zero, unit variance, and uniformly controlled second moment (thus $\mathbf{E}|a_{ij}|^2 1_{|a_{ij}|>K} \to 0$ as $K \to \infty$ uniformly in i, j).

The semicircular law is proven by the moment method, using the identity

$$\int_{-\infty}^{\infty} s^m \, d\mu_n(s) = \frac{1}{n}\sum_{j=1}^{n} \lambda_j^m = \frac{1}{n}\mathrm{tr}((\frac{1}{\sqrt{n}}M)^m)$$

for $m = 0, 1, 2, \ldots$. The key step is to show that

$$\int_{-\infty}^{\infty} s^m \, d\mu_n(s) \to \int_{-\infty}^{\infty} s^m \, d\mu(s)$$

w.h.p.

When one only has second moment control, a preliminary truncation argument to force a_{ij} to be bounded is also needed.

But this is not difficult, because the spectrum is stable in the self-adjoint case; if M' is a self-adjoint perturbation (or truncation) of a self-adjoint matrix M, then the spectrum of M' is close to that of M, as can be seen for instance from the Weyl eigenvalue inequalities.

The circular law

In the non self-adjoint case, there is an analogous law, in which the semi-circular measure $d\mu = \frac{1}{2\pi}(1 - s^2/4)_+^{1/2}$ is replaced by the circular measure $d\mu = \frac{1}{\pi} 1_{s^2+t^2 \le 1}\ dsdt$.

Conjecture (circular law). If the a_{ij} are iid with zero mean and unit variance, then μ_n converges weakly or strongly to μ.

This conjecture is strongly supported by numerics. It is not completely resolved, but there are many partial results.

For the Gaussian ensemble, the strong circular law was established in (Mehta, 1967), using an explicit formula for the joint distribution.

In (Girko, 1984), Girko proposed a method to establish the circular law in general, but certain tail bounds on least singular values were needed to make the method rigorous. In (Bai, 1997), this was accomplished, proving the strong circular law for continuous complex distributions a_{ij} with bounded sixth moment; this was relaxed to $(2+\eta)^{\text{th}}$ moment for any $\eta > 0$ in (Bai-Silverstein, 2006). Several alternate proofs of the strong and weak circular laws under similar assumptions were also given in (Girko, 2004).

In the above arguments, the continuity of the a_{ij} was needed to bound certain least singular values. But with the recent progress on these bounds in the discrete case (Rudelson, 2006; T.-Vu, 2007; Rudelson-Vershynin, 2007; T.-Vu, 2008), as discussed in the previous lecture, it became possible to remove this continuity assumption.

In (Götze-Tikhomirov, 2007) the weak circular law for discrete subgaussian ensembles was proven (using the least singular tail bounds of Rudelson); a similar result with the subgaussian hypothesis relaxed to $(4+\eta)^{\text{th}}$ moment was described in (Girko, 2004). The strong circular law under a fourth moment assumption was established in (Pan-Zhou, 2007), and relaxed to $(2+\eta)^{\text{th}}$ moment in (T.-Vu, 2008). An alternate proof of the weak circular law under $(2+\eta)^{\text{th}}$ moment assumption was given in (Götze-Tikhomirov, 2008).

Thus we are only an η away from fully resolving the circular law!

A new difficulty

There is a major new difficulty in the non-selfadjoint case, which manifests itself in two related ways:

- The spectrum can be very unstable with respect to perturbations. (In particular, truncation is now dangerous.)
- One cannot always control the spectrum effectively just from knowing a few moments.

[Note that the first few moments are very stable with respect to perturbations.]

For instance, consider the matrix

$$M := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

and the exponentially tiny perturbation

$$M' := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 2^{-n} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The matrix M has characteristic polynomial $\det(M - \lambda I) = (-\lambda)^n$ and so the spectrum is $\{0\}$ (with multiplicity n). But the perturbation M' has characteristic polynomial $\det(M' - \lambda I) = (-\lambda)^n - (-2)^{-n}$, and so the spectrum is $\{\frac{1}{2}e^{2\pi ij/n} : j = 1, \ldots, n\}$. An exponentially small change in the matrix has caused the spectrum to move by ~ 1!

Note also that $\mathrm{tr} M^k = \mathrm{tr}(M')^k$ for all $k = 0, \ldots, n-1$; the moment method cannot distinguish M from M' until the n^{th} moment, which is too difficult to compute in practice.

The instability of the spectrum near a complex number λ is closely related to the operator norm of the resolvent $(M - \lambda I)^{-1}$, or equivalently to the least singular value of $M - \lambda I$. In other words, spectral instability is caused by the presence of pseudospectrum.

Indeed, if $\|(M - \lambda I)^{-1}\|$ is bounded, then a standard Neumann series argument shows that $M' - \lambda I$ remains invertible for small perturbations M' of M, and so the spectrum cannot come close to λ.

This already indicates that the least singular values of $M - \lambda I$ will play an important role in the argument.

Girko's method

As the moment method is now unavailable, one must establish the circular law by other means. The key breakthrough was made by Girko in 1984. Instead of moments $\mathrm{tr} M^k = \sum_{j=1}^n \lambda_j^k$, the idea was to study the Stieltjes transform

$$s_n(z) := \int \frac{d\mu_n(s,t)}{s+it-z} = \frac{1}{n}\sum_{j=1}^n \frac{1}{\frac{1}{\sqrt{n}}\lambda_j - z}.$$

This is a meromorphic function with simple poles at each eigenvalue. In principle, control on s_n leads to control on μ_n, and vice versa.

Because s_n is meromorphic, it suffices to understand the real part s_{nr}. For this, one has the fundamental identity

$$s_{nr}(s+it) = \frac{1}{n}\sum_{j=1}^{n} \operatorname{Re}\frac{1}{\frac{1}{\sqrt{n}}\lambda_j - s - it}$$

$$= -\frac{d}{ds}\frac{1}{n}\sum_{j=1}^{n} \log|\frac{1}{\sqrt{n}}\lambda_j - s - it|$$

$$= -\frac{d}{ds}\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}M - (s+it)I)|.$$

Thus, in principle, it suffices to understand the quantity $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}M - (s+it)I)|$. (In practice, one controls the Fourier transform of this quantity.)

One can express $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}M-(s+it)I)|$ in terms of the eigenvalues of $\frac{1}{\sqrt{n}}M-(s+it)I$ and thus of M, but this leads us right back to where we started.

But instead, we can express this quantity in terms of the singular values $\sigma_j(\frac{1}{\sqrt{n}}M-(s+it)I)$ of $\frac{1}{\sqrt{n}}M-(s+it)I$:

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}M-(s+it)I)| = \frac{1}{n}\sum_{j=1}^{n}\log\sigma_j(\frac{1}{\sqrt{n}}M-(s+it)I)$$

The reason why this is useful is because, in contrast to the eigenvalues, the distribution of the singular values of $\frac{1}{\sqrt{n}}M - (s+it)I$ can be efficiently controlled (in principle, at least) by the moment method:

$$\frac{1}{n}\sum_{j=1}^{n}\sigma_j(\frac{1}{\sqrt{n}}M - (s+it)I)^{2k} =$$

$$\frac{1}{n}\mathrm{tr}(((\frac{1}{\sqrt{n}}M - (s+it)I)^*(\frac{1}{\sqrt{n}}M - (s+it)I))^k).$$

For instance, when $s + it = 0$ one obtains a Marchenko-Pastur law.

One can also hope to control the determinant $|\det(\frac{1}{\sqrt{n}}M - (s+it)I)|$ by a variety of other methods (e.g. base $\times$ height, cofactor expansion, etc.).

There is a remaining difficulty, which is that the function $\log : (0, +\infty) \to \mathbf{R}$ is singular at 0 and thus cannot be uniformly approximated by polynomials. (The singularity at $+\infty$ is easily dealt with, using upper bounds on $\|M\|$.) To get around this, Girko made the approximation

$$\log \sigma_j(\frac{1}{\sqrt{n}}M-(s+it)I) \approx \log \sqrt{\varepsilon^2 + \sigma_j(\frac{1}{\sqrt{n}}M - (s+it)I)^2}$$

for some $\varepsilon > 0$. The function $\log \sqrt{\varepsilon^2 + x^2}$ is now smooth at 0 and the contribution of the RHS can now be dealt with a variety of methods (after many computations).

The remaining difficulty is to show that the approximation gives an asymptotically accurate control on in the limit $\varepsilon \to 0$. For this it becomes necessary to establish lower tail bounds on the least singular value of $\frac{1}{\sqrt{n}}M - (s+it)I$.

Fortunately, with recent results we know that this singular value is usually bounded below by $n^{-O(1)}$. This allows one to take $\varepsilon = n^{-O(1)}$ in the above approximation, which eventually costs some factors of $\log \frac{1}{\varepsilon} \sim \log n$ in various estimates (which, ultimately, cause one to require $2+\eta$ moment rather than second moment).

Because of the instability of the spectrum, truncation methods are unavailable (until one has regularised the problem and reduced matters to controlling singular values instead of eigenvalues). So the case of random variables with few moment bounds is genuinely more difficult than those of higher moment bounds.

In particular, once one loses fourth moment bounds, one can no longer assume that the largest singular value σ_1 of $\frac{1}{\sqrt{n}}M - (s+it)I$ is bounded by $O(1)$; only polynomial bounds $O(n^{O(1)})$ are available. This causes some technical difficulties in this regime.

The circular law results have been extended to the case in which the entries are independent but not identical (though one always needs mean zero and unit variance), provided one has a (slightly technical) uniform control of second-moment type (which is called "κ-controllability" in (T.-Vu, 2008)). The methods also extend to sparser models in which only $n^{1+\varepsilon}$ of the n^2 coefficients are non-zero (and adjusting the normalisation accordingly); see (Götze-Tikhomirov, 2007, T.-Vu, 2008).

The case of non-zero mean is not well understood at present; the circular law here needs to be replaced by a different law, but it is not clear what the new law should be.