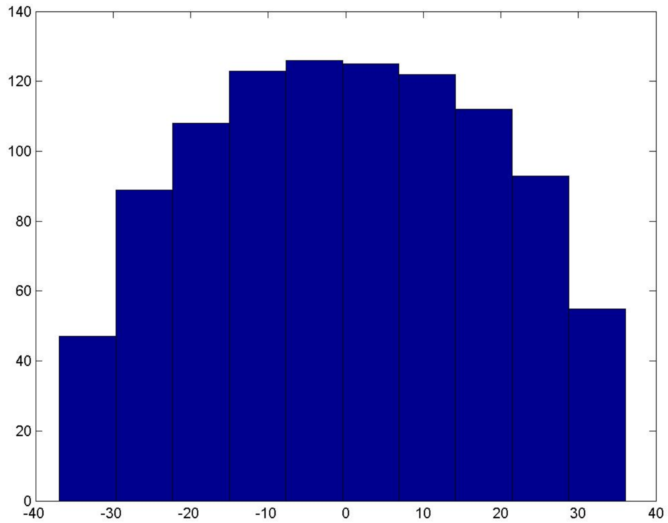
CS 683: Advanced Design & Analysis of Algorithms	March 24, 2008

Lecture 25

Lecturer: John Hopcroft — *Scribe: Ian Grayson, Sucheta Soundarajan*

If we plot the eigenvalues of a symmetric random matrix, we should get a semi-circular distribution. Conversely, if we plot the eigenvalues of a matrix and see that the distribution is semi-circular, this suggests that the matrix is random. Here is the histogram for eigenvalues of a 1000-by-1000 random matrix with entries between -1 and 1.



1 References

Reference: Wigner, E. "'On the distribution of the roots of certain symmetric matrices." Annals of Math Vol. 67, pp 325-327, 1958.

2 Introduction

Consider a symmetric matrix A, where all elements are i.i.d. from a distribution which is symmetric about 0, with variance σ^2, all moments finite. The probability of the normalized eigenvalues is $\frac{2}{\pi}\sqrt{1-\lambda^2}$ for $\lambda^2 \leq 1$. The normalization factor here is $2\sigma\sqrt{n}$. Let $\mathrm{P}(\lambda) = \lim_{n\to\infty}\mathrm{P}_n(2\lambda\sigma\sqrt{n})$.

Claim:

$$P(\lambda) = \begin{cases} \frac{2}{\pi}\sqrt{1-\lambda^2} & \text{for } \lambda^2 \le 1, \\ \\ 0 & \text{otherwise.} \end{cases}$$

3 Moments of $P(\lambda)$

We want to show that the distribution of eigenvalues equals $P(\lambda)$. First, we calculate the moments of $P(\lambda)$.

Proof by showing that the moments of $P(\lambda)$ and $\frac{2}{\pi}\sqrt{1-\lambda^2}$ are the same. Let $C(k)$ be the k^{th} moment of $\frac{2}{\pi}\sqrt{1-\lambda^2}$.

$$C_k = \begin{cases} \frac{2}{\pi}\int_{-1}^{1} \lambda^k \sqrt{1-\lambda^2}\, d\lambda & \text{for even k} \\ 0 & \text{for odd k} \end{cases}$$

Let $\lambda = \sin\theta$, then

$$\begin{aligned} C(k) &= \tfrac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^k\theta \cos^2\theta \, d\theta \qquad\qquad (d\lambda = \cos\theta\, d\theta) \\ &= \tfrac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^k\theta\, d\theta - \tfrac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+2}\theta\, d\theta \end{aligned}$$

Using the following identity from the professor's notes

(note that, in the RHS, $n = 2$ yields $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 = \pi$, which will allow us to cancel out the π in C_k)

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^n\theta\, d\theta = -\tfrac{\sin^{n-1}\theta\cos\theta}{n} + \tfrac{n-1}{n}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{n-2}\theta\, d\theta$$

$$\sin^n\theta = -\tfrac{n-1}{n}\sin^{n-2}\theta \underbrace{\cos^2\theta}_{1-\sin^2\theta} + \tfrac{\sin^n\theta}{n} + \tfrac{n-1}{n}\sin^{n-2}\theta$$

$$\int \sin^n\theta\, d\theta = \tfrac{(n-1)(n-3)\cdots 1}{n(n-2)\cdots 2}$$

we begin to reduce $C(k)$:

$$\begin{aligned} C(k) &= 2\tfrac{1\cdot 3\cdots(k-1)}{2\cdot 4\cdots k} - 2\tfrac{1\cdot 3\cdots(k+1)}{2\cdot 4\cdots(k+2)} \\ &= 2\tfrac{[1\cdot 3\cdots(k-1)](k+2-(k+1)}{2\cdot 4\cdots(k+2)} = 2\tfrac{1\cdot 3\cdots(k-1)}{2\cdot 4\cdots(k+2)} \\ &= 2\tfrac{k!}{(2\cdot 4\cdots k)^2(k+2)} \\ &= 2\tfrac{k!}{2^k(1\cdot 2\cdot 3\cdots\frac{k}{2})^2}(\tfrac{1}{k+2}) \\ &= (\tfrac{1}{k+2})(\tfrac{1}{2^{k-1}})\binom{k}{\frac{k}{2}} \end{aligned}$$

4 Moments of probability distribution

Let $m(k)$ be the moments of the probability distribution in the claim. Then $m(k) = \mathrm{E}[\frac{1}{n}\sum_{j=1}^{n}(\frac{\lambda_j}{2\sqrt{n}})^k]$.

The denominator inside the sum is the normalization factor, since $\sigma = 1$.

Then $m(k) = \frac{1}{n}\frac{1}{2^k n^{\frac{k}{2}}}\text{E}[\sum_{j=1}^{n} \lambda_j^k] = \frac{1}{n}\frac{1}{2^k n^{\frac{k}{2}}}\text{E}(\text{trace}(A^k))$, since the trace of the matrix is the sum of the eigenvalues.

Now the problem is reduced to finding the trace of A^k. We can think of diagonal elements of A^k as paths of length k from a vertex back to itself, where the value of a path is the product of the labels along the path. We can classify paths by their structure: in some paths, every edge is traversed at least twice, and in others, there is at least one edge which is traversed only once. Let us first consider paths of the second type. Then the expected value of such a path is the expected value of the product of the edge weights along the path. If the path contains edges e_1, ..., e_r occurring with frequencies f_1, ..., f_r (with at least one edge occurring only once), then the expected value of the path is $E(e_1^{f_1})...E(e_r^{f_r})$. However, since at least one f_i is 1, at least one of the elements in this product is $E(e_i)$. But this value is 0, since elements in A are sampled from $\{1, -1\}$. Thus, the expected value of paths in which at least one edge is traversed only once is 0, so we only need to consider paths in which every edge is traversed at least twice.

Moreover, we only need to consider paths in which when we traverse an edge for the first time, we go to a new vertex, since this value is asymptotically larger than other types of paths.

Such paths will see $\frac{k}{2}$ new vertices. These are depth first search trees on $\frac{k}{2}$ vertices! So now we need to count DFS trees. DFS trees are equivalent to balanced parentheses, and the number of balanced parentheses is given by the Catalan numbers, $\text{Cat(k)} = \frac{1}{k+1}\binom{2k}{k}$. (to be proven later)

Then $m(k) = \frac{1}{2^k}\frac{1}{n^{1+\frac{k}{2}}}\frac{1}{\frac{k}{2}+1}\binom{k}{\frac{k}{2}}n^{\frac{k}{2}}n$. Here, the final n represents the number of diagonal elements, and the $\frac{1}{\frac{k}{2}+1}\binom{k}{\frac{k}{2}}$ represents the number of types of paths (from the Catalan numbers). Then $m(k) = \frac{1}{k+2}\frac{1}{2^{k-1}}\binom{k}{\frac{k}{2}} = C(k)$.