Wegner's Semicircular Law

If we have a large dataset in a high dimensional space, we might ask ourselves whether the data really lie in such a high dimension, or if it is lower dimensional data hiding behind random noise, and how could we distinguish the data from the noise.

In situations like these, Wegner's semicircular law can be helpful since it will give us a picture of what the eigenvalues of a random matrix will look like, and we can use this information to filter out what is most likely the noise in a matrix with both data and noise. For example, if there are 6 outliers outside the semicircular distribution, then the data is really 6 dimensional, and the rest is just random noise.

Let A be a random $n \times n$ matrix, and a_{ij} the entry in the ith row and jth column. Assume

$$\mathrm{E}[a_{ij}] = 0$$
$$\sigma^2(a_{ij}) = 1.$$

That is, each entry of A has expected value 0 and unit variance.

Define $P_n(\lambda)$ to be the probability distribution of the normalized eigenvalues of such an $n \times n$ matrix. By normalized eigenvalues, we mean eigenvalues resized to lie in the region $[-1, 1]$. It turns out that for a random matrix with expected value 0 for each element, the eigenvalues will lie in the range $-2\sqrt{\pi}\sigma$ to $2\sqrt{\pi}\sigma$, so normalizing involves dividing the eivenvalues by $2\sqrt{\pi}\sigma$. Let

$$P(\lambda) = \lim_{n \to \infty} P_n(\lambda).$$

Think of $P(\lambda)$ as the probability distribution of eigenvalues for a large random matrix. What Wegner realized is that $P(\lambda)$ is roughly a semicircle centered at the origin with radius 1 along the x-axis, but with a slightly smaller height $(2/\pi)$ that makes the area under this semicircle 1.

Theorem (Wegner). *For $P(\lambda)$ as defined above,*

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

Proof. Idea: Compare the moments of $\frac{2}{\pi}\sqrt{1-\lambda^2}$ to the moments of $P(\lambda)$.

For simplicity, assume the elements of the matrix are

$$a_{ij} = \pm 1, \quad \Pr = 1/2.$$

Let c_k be the kth moment of $\frac{2}{\pi}\sqrt{1-\lambda^2}$.

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

To evaluate this integral, let $\lambda = \sin\theta$. Then $d\lambda = \cos\theta\ d\theta$, and as λ ranges from -1 to 1, θ ranges from $-\pi/2$ to $\pi/2$. So, for even k,

$$\begin{aligned} c_k &= \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \sin^k\theta \cos^2\theta\ d\theta \\ &= \frac{2}{\pi}\left[\int_{-\pi/2}^{\pi/2} \sin^k\theta\ d\theta - \int_{-\pi/2}^{\pi/2} \sin^{k+2}\theta\ d\theta\right]. \end{aligned}$$

Useful formulas:

$$\begin{aligned} \int \sin^n\theta\ d\theta &= \frac{-\sin^{n-1}\theta\cos\theta}{n} + \frac{n-1}{n}\int \sin^{n-2}\theta\ d\theta \\ \int \sin^2\theta\ d\theta &= \frac{1}{2}\int (1-\cos 2\theta)\ d\theta = \frac{1}{2}\theta - \frac{1}{4}\sin 2\theta. \end{aligned}$$

Our bounds of integration, $-\pi/2$ to $\pi/2$, simplify these integrals somewhat, and with a little algebra, we get

$$\begin{aligned} c_k &= \frac{2}{\pi}\left[\left(\frac{k-1}{k}\cdot\frac{k-3}{k-2}\cdots\frac{3}{4}\cdot\frac{\pi}{2}\right) - \left(\frac{k+1}{k+2}\cdot\frac{k-1}{k}\cdots\frac{3}{4}\cdot\frac{\pi}{2}\right)\right] \\ &= 2\left[\left(\frac{k-1}{k}\cdot\frac{k-3}{k-2}\cdots\frac{3}{4}\cdot\frac{1}{2}\right) - \left(\frac{k+1}{k+2}\cdot\frac{k-1}{k}\cdots\frac{3}{4}\cdot\frac{1}{2}\right)\right] \\ &= 2\left[\frac{1\cdot 3\cdots(k-1)}{2\cdot 4\cdots k\cdot(k+2)}\ (k+2-(k+1))\right] \\ &= \frac{2}{k+2}\ \frac{k!}{(2\cdot 4\cdots k)^2} \\ &= \frac{1}{(k+2)\ 2^{k-1}}\binom{k}{k/2}. \end{aligned}$$

We now go on to calculate the moments of $P(\lambda)$ itself. Let m_k be the kth moment of

$P(\lambda)$. We have

$$m_k = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\lambda_j}{2\sqrt{n}}\right)^k\right].$$

It is simply the expected average of the normalized eigenvalues raised to the kth power. Remember that we are assuming $\sigma = 1$, so we can normalize by dividing by $2\sqrt{n}$ instead of $2\sqrt{n}\sigma$.

The last expression can be rewritten as

$$m_k = \frac{1}{2^k\ n^{1+k/2}}\ \mathrm{E}\left[\sum_{j=1}^{n}\lambda_j^k\right].$$

Recall that $\mathrm{trace}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. It can also be shown that $\mathrm{trace}(A^k) = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k$.

Thus,

$$m_k = \frac{1}{2^k\ n^{1+k/2}}\ \mathrm{E}\left[\mathrm{trace}(A^k)\right].$$

You can think of A^k as representing all paths of length k in a graph described by A...

[We shall continue this proof in the next lecture.] $\square$