Random matrices and communication systems: WEEK 10

In this lecture, we first introduce the Catalan numbers and then state and prove the Wigner theorem (a slightly modified version of the original theorem, actually).

1 Preliminary: the Catalan numbers

The Catalan numbers are defined as follows:

$$c_k = \frac{1}{k+1}\binom{2k}{k} = \frac{(2k)!}{k!(k+1)!}, \quad k \geq 0$$

So the sequence starts as $1, 1, 2, 5, 14, 42, \ldots$ These numbers have multiple combinatorial interpretations. Let us give simply one here. We consider paths that evolve in discrete time over the integer numbers. These paths go either up or down by one unit in one time step. We are interested in the number of paths that start from 0 and go back to 0 in $2k$ time steps, *without hitting the negative numbers* in the interval. Such paths are called *Dyck paths* and are illustrated on Figure 1 below.

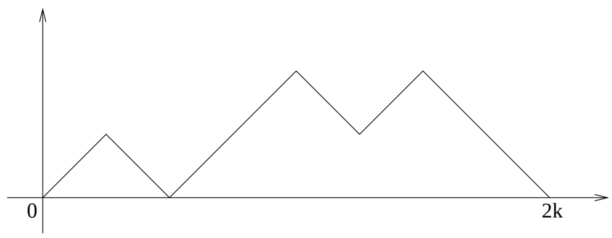

Figure 1. Dyck path

It turns out that for a given k, the number of Dyck paths of length $2k$ is equal to the Catalan number c_k. Here is the proof, also known as the *reflection principle.*

Proof. Let us first make the following trivial observation: the number of Dyck paths of length $2k$ is equal to the *total* number of paths from $(0,0)$ to $(2k,0)$, *minus* the number of paths from $(0,0)$ to $(2k,0)$ *that do hit the negative numbers* at least once in the interval.

For each of these paths hitting the negative numbers at least once, let us now define T as the first time the number -1 is hit by the path, as illustrated on Figure 2 below. From T onwards, we can draw a mirror path with respect to the horizontal axis with vertical coordinate -1, that necessarily lands in position -2 at time $2k$ (the mirror position of 0 with respect to -1).

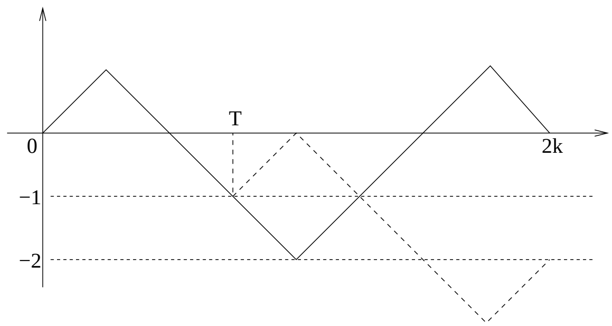

Figure 2. Reflection principle

Counting therefore the number of paths from $(0,0)$ to $(2k,0)$ hitting the negative numbers at least once is the same as counting the number of paths from $(0,0)$ to $(2k,-2)$ hitting the negative numbers at least once. But any such path *must* hit the negative numbers at some point, so the number we are computing is simply the *total* number of paths from $(0,0)$ to $(2k,-2)$.

Finally, we obtain that the number of Dyck paths of length $2k$ is equal to the total number of paths from $(0,0)$ to $(2k,0)$ minus the total number of paths from $(0,0)$ to $(2k,-2)$, which is equal to

$$\binom{2k}{k} - \binom{2k}{k-1} = \frac{(2k)!}{(k!)^2} - \frac{(2k)!}{(k-1)!\,(k+1)!} = \frac{(2k)!}{(k!)^2}\left(1 - \frac{k}{k+1}\right) = \frac{1}{k+1}\binom{2k}{k} = c_k$$

$\square$

The Catalan numbers also have the following interpretation. Let μ be the distribution with pdf

$$p_\mu(x) = \frac{1}{\pi}\sqrt{\frac{1}{x} - \frac{1}{4}}\,1_{\{0 < x < 4\}} \tag{1}$$

Then its moments are given by

$$m_k = \int_0^4 x^k p_\mu(x)\,dx = \frac{1}{k+1}\binom{2k}{k} = c_k$$

The proof is left as an exercise in the homework. Notice that as the distribution μ is compactly supported on the interval $[0,4]$, we know that

$$|m_k| \le 4^k \quad \text{(this can also be deduced directly from the definition of } c_k\text{)}$$

So the moments m_k satisfy Carleman's condition seen in the last lecture.

Remark 1.1. The above distribution μ is sometimes called the *quarter circle law.* Although this terminology is not really appropriate for such a distribution (drawing $p_\mu(x)$ as a function of x, we do not see a quarter circle...), here is the reason for this denomination. Say μ is the distribution of a (positive) random variable X. Then the distribution ν of $\sqrt{X}$ has the following pdf:

$$p_\nu(y) = p_\mu(y^2)\,2y = \frac{1}{\pi}\sqrt{\frac{1}{y^2} - \frac{1}{4}}\,2y\,1_{\{0 < y < 2\}} = \frac{1}{\pi}\sqrt{4 - y^2}\,1_{\{0 < y < 2\}} \tag{2}$$

which indeed has the form of a quarter circle.

2 Wigner's theorem

Let H be an $n \times n$ random matrix with i.i.d. complex-valued entries such that for all $1 \le j, l \le n$:

(i) $\mathbb{E}\left(|h_{jl}|^2\right) = 1$;

(ii) $\mathbb{E}\left(|h_{jl}|^k\right) < \infty$ for all $k \ge 0$;

(iii) $\mathbb{E}\left(h_{jl}^k\,\overline{h_{jl}}^{k'}\right) = 0$ if $k \neq k'$.

Notice that these last two assumptions are satisfied in particular when

(ii') the distribution of h_{jl} is compactly supported;

(iii') the distribution of h_{jl} is circularly symmetric.

Also, assumptions (i)-(iii) are satisfied when h_{jl} are i.i.d.$\sim \mathcal{N}_\mathbb{C}(0,1)$ random variables.

Let us now consider the (rescaled) Wishart random matrix $W^{(n)} = \frac{1}{n} HH^*$; this matrix is positive semi-definite, so it is unitarily diagonalizable and its eigenvalues $\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}$ are non-negative. Let also

$$\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j^{(n)}} \quad \text{i.e.} \quad \mu_n(B) = \frac{1}{n} \sharp\{1 \le j \le n \;:\; \lambda_j^{(n)} \in B\}, \quad B \in \mathcal{B}(\mathbb{R})$$

μ_n is called the *empirical eigenvalue distribution* of the matrix $W^{(n)}$. Notice that it is a *random* distribution, because for each n, the eigenvalues $\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}$ are random. For a given realization of the $\lambda^{(n)}$'s, μ_n may still be interpreted as the distribution of one of these eigenvalues picked uniformly at random.

Wigner's theorem is then the following.

Theorem 2.1. Under assumptions (i)-(iii), almost surely, the sequence $(\mu_n,\, n \ge 1)$ converges weakly towards the (deterministic) distribution μ, whose pdf is given by (1).

Before starting with the proof of this theorem, let us mention the following immediate corollary. Let $\sigma_1^{(n)}, \ldots, \sigma_n^{(n)}$ be the singular values of the matrix $H^{(n)} = \frac{1}{\sqrt{n}} H$. As $W^{(n)} = H^{(n)}\,(H^{(n)})^*$, it holds that $\sigma_j^{(n)} = \sqrt{\lambda_j^{(n)}}$. Let also

$$\nu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\sigma_j^{(n)}}$$

The above theorem can now be rephrased as:

Corollary 2.2. Under assumptions (i)-(iii), almost surely, the sequence $(\nu_n,\, n \ge 1)$ converges weakly towards the (deterministic) distribution ν, whose pdf is given by (2).

Proof of Theorem 2.1. In order to prove the result, we will use Carleman's theorem, which requires us to show that almost surely,

$$m_k^{(n)} = \int_{\mathbb{R}} x^k \, d\mu_n(x) \underset{n\to\infty}{\to} c_k \quad \forall k \ge 0 \tag{3}$$

where c_k are the Catalan numbers, that is, the moments of the distribution μ. In the sequel, we show that

$$\left| \mathbb{E}\left(m_k^{(n)}\right) - c_k \right| = O\left(\frac{1}{n}\right) \quad \forall k \ge 0 \tag{4}$$

Using similar methods (involving slightly more combinatorics, though), it can also be shown that

$$\mathrm{Var}\left(m_k^{(n)}\right) = O\left(\frac{1}{n^2}\right) \quad \forall k \ge 0 \tag{5}$$

The last two lines imply (3). Indeed, for all $\varepsilon > 0$,

$$\begin{aligned}
\sum_{n\ge1} \mathbb{P}\left(\left|m_k^{(n)} - c_k\right| > \varepsilon\right) \;&\le\; \frac{1}{\varepsilon^2} \sum_{n\ge1} \mathbb{E}\left(\left(m_k^{(n)} - c_k\right)^2\right) = \frac{1}{\varepsilon^2} \sum_{n\ge1} \mathbb{E}\left(\left(m_k^{(n)} - \mathbb{E}\left(m_k^{(n)}\right) + \mathbb{E}\left(m_k^{(n)}\right) - c_k\right)^2\right) \\
&\le\; \frac{2}{\varepsilon^2} \sum_{n\ge1} \left(\mathrm{Var}\left(m_k^{(n)}\right) + \left(\mathbb{E}\left(m_k^{(n)}\right) - c_k\right)^2\right) < \infty
\end{aligned}$$

as both terms in the series are $O(1/n^2)$ by (4) and (5). The Borel-Cantelli lemma allows then to conclude that for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|m_k^{(n)} - c_k\right| > \varepsilon \text{ infinitely often}\right) = 0$$

which is saying that $m_k^{(n)}$ converges almost surely towards c_k as $n \to \infty$.

We now set out to prove (4). Let us first develop

$$\begin{aligned}
\mathbb{E}\left(m_k^{(n)}\right) &= \mathbb{E}\left(\int_{\mathbb{R}} x^k \, d\mu_n(x)\right) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\lambda_j^{(n)}\right)^k\right) = \mathbb{E}\left(\frac{1}{n}\mathrm{Tr}\left(\left(W^{(n)}\right)^k\right)\right) \\
&= \frac{1}{n^{k+1}}\,\mathbb{E}\left(\mathrm{Tr}\left((HH^*)^k\right)\right) = \frac{1}{n^{k+1}} \sum_{j_1,l_1,\ldots,j_k,l_k=1}^{n} \mathbb{E}\left(h_{j_1,l_1}\,\overline{h_{j_2,l_1}}\cdots h_{j_k,l_k}\,\overline{h_{j_1,l_k}}\right) \qquad (6)
\end{aligned}$$

Let us first look at the cases $k=1$ and $k=2$ for simplicity. For $k=1$, we have

$$\mathbb{E}\left(m_1^{(n)}\right) = \frac{1}{n^2}\sum_{j,l=1}^{n}\mathbb{E}\left(|h_{jl}|^2\right) = 1$$

and for $k=2$, because of assumption (iii) on the matrix entries h_{jl}, as well as the independence assumption, we have

$$\begin{aligned}
\mathbb{E}\left(m_2^{(n)}\right) &= \frac{1}{n^3}\sum_{j_1,j_2,l_1,l_2=1}^{n}\mathbb{E}\left(h_{j_1,l_1}\,\overline{h_{j_2,l_1}}h_{j_2,l_2}\,\overline{h_{j_1,l_2}}\right) \\
&= \frac{1}{n^3}\left(\sum_{j,l=1}^{n}\mathbb{E}\left(|h_{jl}|^4\right) + \sum_{j,l_1\neq l_2}\mathbb{E}\left(|h_{j,l_1}|^2\,|h_{j,l_2}|^2\right) + \sum_{j_1\neq j_2,l}\mathbb{E}\left(|h_{j_1,l}|^2\,|h_{j_2,l}|^2\right)\right) \\
&= \frac{1}{n^3}\left(O\left(n^2\right) + n^2\left(n-1\right) + n^2\left(n-1\right)\right) = 2 + O\left(\frac{1}{n}\right)
\end{aligned}$$

which proves the claim for these two cases. Notice that in the second case, there are a priori n^4 terms in the sum, but only $O\left(n^3\right)$ terms bring a non-zero contribution to the overall sum.

The remainder of the proof is left to the next lecture.

Remark 2.3. Notice that equation (5), which is not proven here, could seem a priori quite unexpected. It indeed says that

$$\mathrm{Var}\left(m_k^{(n)}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\lambda_j^{(n)}\right)^k\right) = O\left(\frac{1}{n^2}\right)$$

In case the $\lambda^{(n)}$'s were n i.i.d. random variables, one would rather expect this variance to be $O(1/n)$. But the eigenvalues of a random matrix are far from being i.i.d. in general, as already observed when computing their joint distribution in Lecture 6 in the Gaussian case. They are actually n random variables built from a matrix with $O(n^2)$ i.i.d. entries, which is in concordance with the fact that the above variance is $O(1/n^2)$.