Random matrices and communication systems: WEEK 7

1 Wishart random matrices: marginal eigenvalue distribution

We only consider here the complex case, as the analysis of the real case is sensibly more difficult than what is presented below. Let H be an $n \times n$ matrix with i.i.d.$\sim \mathcal{N}_{\mathbb{C}}(0,1)$ entries and $W = HH^*$. We have seen in the previous lecture that the joint distribution of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of W is given by

$$p(\lambda_1, \ldots, \lambda_n) = c_n \prod_{j=1}^{n} e^{-\lambda_j} \prod_{j<k} (\lambda_k - \lambda_j)^2$$

where c_n is a normalization constant and where we have dropped the term "$1_{\lambda_j \geq 0}$" in the above expression in order to lighten the notation. Notice also that compared to the previous lecture, we consider here only the case $n = m$. This is meant to simplify the exposition in the sequel, but contrary to the above mentioned real case, the case $m \geq n$ can be handled with little extra effort.

Let us first give a possible reason for studying marginals of this joint distribution. The above expression allows to compute the expectation of a function of the eigenvalues $f(\lambda_1, \ldots, \lambda_n)$:

$$\mathbb{E}(f(\lambda_1, \ldots, \lambda_n)) = \int_0^\infty d\lambda_1 \cdots \int_0^\infty d\lambda_n \, p(\lambda_1, \ldots, \lambda_n) \, f(\lambda_1, \ldots, \lambda_n)$$

An example of such function is $f(\lambda_1, \ldots, \lambda_n) = \prod_{j=1}^n \lambda_j$ (which corresponds to $f(W) = \det(W)$).

If one wants now to compute the expectation of a funtion of the form $f(\lambda_1, \ldots, \lambda_n) = \sum_{j=1}^n g(\lambda_j)$, for some function g, it is of course possible to write

$$\mathbb{E}\left(\sum_{j=1}^n g(\lambda_j)\right) = \sum_{j=1}^n \int_0^\infty d\lambda_1 \cdots \int_0^\infty d\lambda_n \, p(\lambda_1, \ldots, \lambda_n) \, g(\lambda_j)$$

But notice that this may also be rewritten as

$$\mathbb{E}\left(\sum_{j=1}^n g(\lambda_j)\right) = \sum_{j=1}^n \int_0^\infty d\lambda_j \, p(\lambda_j) \, g(\lambda_j)$$

where

$$p(\lambda_j) = \int_0^\infty d\lambda_1 \cdots \int_0^\infty d\lambda_{j-1} \int_0^\infty d\lambda_{j+1} \cdots \int_0^\infty d\lambda_n \, p(\lambda_1, \ldots, \lambda_n)$$

are the first-order marginals of p. Notice in addition that the distribution $p(\lambda_1, \ldots, \lambda_n)$ is symmetric in any permutation of the λ's, so we may as well consider the eigenvalues $\lambda_1, \ldots \lambda_n$ as unordered. In this case, all the above marginals are the same, so the expression for the expectation boils down to

$$\mathbb{E}\left(\sum_{j=1}^n g(\lambda_j)\right) = n \int_0^\infty d\lambda \, p(\lambda) \, g(\lambda)$$

$p(\lambda)$ may be also interpreted here as the distribution of one of the eigenvalues $\lambda_1, \ldots, \lambda_n$ picked uniformly at random. An example where this formula applies is when $g(\lambda) = \log(\lambda)$, which corresponds to $f(\lambda_1, \ldots, \lambda_n) = \sum_{j=1}^n \log(\lambda_j)$, which corresponds in turn to $f(W) = \log\det(W)$.

Computation of the marginals. The first step for the computation of the marginal $p(\lambda)$ is to use Vandermonde's determinant formula:

$$\prod_{j<k} (\lambda_k - \lambda_j) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix}$$

Next, we introduce *Laguerre polynomials*: these are defined as

$$L_k(\lambda) = \frac{1}{k!}\, e^{\lambda}\, \frac{d^k}{d\lambda^k}\left(e^{-\lambda}\, \lambda^k\right) \quad \text{where } k \in \mathbb{N},\ \lambda \geq 0$$

So $L_0(\lambda) = 1$, $L_1(\lambda) = 1 - \lambda$, $L_2(\lambda) = \frac{1}{2}(\lambda - 2)^2 - 1$, and so on. In general, L_k is a polynomial of degree k in λ that may be written as

$$L_k(\lambda) = \gamma_k\, \lambda^k + \text{ lower order terms}, \quad \text{where } \gamma_k = \frac{(-1)^k}{k!} \neq 0$$

One can check in addition that these polynomials satisfy in the following *orthogonality relations*:

$$\int_0^\infty d\lambda\, e^{-\lambda}\, L_k(\lambda)\, L_l(\lambda) = \delta_{kl} \tag{1}$$

We now use these Laguerre polynomials to rewrite the above determinant (using the basic rules for the determinant) as

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix} = \left(\prod_{j=1}^{n-1} \frac{1}{\gamma_j} \right) \det \begin{pmatrix} L_0(\lambda_1) & L_0(\lambda_2) & \cdots & L_0(\lambda_n) \\ L_1(\lambda_1) & L_1(\lambda_2) & \cdots & L_1(\lambda_n) \\ \vdots & \vdots & & \vdots \\ L_{n-1}(\lambda_1) & L_{n-1}(\lambda_2) & \cdots & L_{n-1}(\lambda_n) \end{pmatrix}$$

In order to simplify notation, let us rewrite the matrix on the right-hand side as $\{L_{j-1}(\lambda_k)\}_{j,k=1}^n$. Combining everything together, we obtain

$$p(\lambda_1, \ldots, \lambda_n) = c_n \prod_{j=1}^n e^{-\lambda_j} \left(\prod_{j=1}^{n-1} \frac{1}{\gamma_j} \right)^2 \det\left(\{L_{j-1}(\lambda_k)\}_{j,k=1}^n \right)^2$$

Noticing that the product of γ's is yet another constant, we may simply include it in the constant c_n. Also, the above expression may be transformed into

$$p(\lambda_1, \ldots, \lambda_n) = c_n \prod_{j=1}^n e^{-\lambda_j} \det\left(\left(\{L_{j-1}(\lambda_k)\}_{j,k=1}^n\right)^T \{L_{j-1}(\lambda_k)\}_{j,k=1}^n \right) = c_n \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^n \right)$$

where $K(\lambda, \mu) = e^{-\frac{\lambda+\mu}{2}} \sum_{l=0}^{n-1} L_l(\lambda)\, L_l(\mu)$.

Remarks. - Notice that even though we are talking about the eigenvalues of the complex-valued matrix W, this matrix is Hermitian (and even positive semi-definite), so its eigenvalues are real (we therefore only have a transpose matrix above, and not a complex-conjugate transpose).

- The above trick of writing $\det(A)^2 = \det(A^T A)$ is of course made possible because of the presence of the square in the expression for the joint eigenvalue distribution. In the real case, the square is missing, which makes things much more delicate (one can always write $|\det(A)| = \sqrt{\det(A^T A)}$, but the square root creates problems later).

The Kernel K has the following nice properties, that follow from the orthogonality relations (1) for the Laguerre polynomials.

Lemma 1.1. a) $K(\mu, \lambda) = K(\lambda, \mu)$ b) $\int_0^\infty d\lambda\, K(\lambda, \lambda) = n$ c) $\int_0^\infty d\mu\, K(\lambda, \mu)\, K(\mu, \nu) = K(\lambda, \nu)$

The last property above is known as the "self-reproducing property" of the Kernel K.

Proof. a) is obvious. b) $\int_0^\infty d\lambda\, K(\lambda, \lambda) = \sum_{l=0}^{n-1} \int_0^\infty d\lambda\, e^{-\lambda}\, L_l(\lambda)^2 = n$.

c) $\int_0^\infty d\mu\, K(\lambda, \mu)\, K(\mu, \nu) = e^{-\frac{\lambda+\nu}{2}}\, L_l(\lambda)\, L_m(\nu) \sum_{l,m=0}^{n-1} \int_0^\infty d\mu\, e^{-\mu}\, L_l(\mu) L_m(\mu) = K(\lambda, \nu)$. □

From these properties follows the remarkable fact below, known as *Mehta's lemma.*

Lemma 1.2. The m^{th} order marginal of the joint eigenvalue distribution $p(\lambda_1, \ldots, \lambda_n)$ is given by

$$p(\lambda_1, \ldots, \lambda_m) = \frac{(n-m)!}{n!} \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^m\right)$$

Proof. We first give the proof for the case $m = n-1$ (the general case follows then easily). We are interested in computing

$$p(\lambda_1, \ldots, \lambda_{n-1}) = c_n \int_0^\infty d\lambda_n \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^n\right)$$

In order to lighten the notation, let us write $A = \{a_{jk}\}_{j,k=1}^n = \{K(\lambda_j, \lambda_k)\}_{j,k=1}^n$. Using the expansion formula for the determinant, we obtain

$$p(\lambda_1, \ldots, \lambda_{n-1}) = c_n \int_0^\infty d\lambda_n \det(A) = c_n \sum_{l=1}^n (-1)^{n+l} \int_0^\infty d\lambda_n\, a_{nl} \det(A(n,l))$$

where $A(n,l)$ denotes the matrix A with n^{th} row and l^{th} column suppressed. This may be further rewritten as

$$p(\lambda_1, \ldots, \lambda_{n-1}) = c_n \left(\int_0^\infty d\lambda_n\, a_{nn}\right) \det(A(n,n)) + c_n \sum_{l=1}^{n-1} (-1)^{n+l} \int_0^\infty d\lambda_n\, a_{nl} \det(A(n,l)) \qquad (2)$$

where

$$\int_0^\infty d\lambda_n\, a_{nn} = \int_0^\infty d\lambda_n\, K(\lambda_n, \lambda_n) = n$$

by property b) of Lemma 1.1, Furthermore, we have for all $1 \le l \le n-1$,

$$\begin{aligned}
&\int_0^\infty d\lambda_n\, a_{nl} \det(A(n,l)) \\
= &\int_0^\infty d\lambda_n\, a_{nl} \det \begin{pmatrix} a_{1,1} & \cdots & a_{1,l-1} & a_{1,l+1} & \cdots & a_{1,n-1} & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,l-1} & a_{n-1,l+1} & \cdots & a_{n-1,n-1} & a_{n-1,n} \end{pmatrix} \qquad (3) \\
= &\det \begin{pmatrix} a_{1,1} & \cdots & a_{1,l-1} & a_{1,l+1} & \cdots & a_{1,n-1} & \int_0^\infty d\lambda_n\, a_{1,n}\, a_{n,l} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,l-1} & a_{n-1,l+1} & \cdots & a_{n-1,n-1} & \int_0^\infty d\lambda_n\, a_{n-1,n} a_{n,l} \end{pmatrix}
\end{aligned}$$

Indeed, the integral can be brought inside the determinant for the following reason: writing down the full expansion formula for the determinant in (3), we see that each term only involves one occurrence of λ_n. We further have

$$\int_0^\infty d\lambda_n\, a_{jn}\, a_{nl} = \int_0^\infty d\lambda_n\, K(\lambda_j, \lambda_n)\, K(\lambda_n, \lambda_l) = K(\lambda_j, \lambda_l) = a_{jl}$$

by property c) of Lemma 1.1. So up to a column permutation, the matrix on the right-hand side of the above expression is $A(n,n)$, which leads to

$$\int_0^\infty d\lambda_n\, a_{nl} \det(A(n,l)) = (-1)^{n-l-1} \det(A(n,n))$$

Inserting this in equation (2) finally gives

$$\begin{aligned}
p(\lambda_1, \ldots, \lambda_{n-1}) &= c_n \left(n \det(A(n,n)) + \sum_{l=1}^n (-1)^{2n-1} \det(A(n,n)) \right) \\
&= c_n\, (n - (n-1)) \det(A(n,n)) = c_n \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^{n-1}\right)
\end{aligned}$$

which proves the claim for $m = n-1$. A similar reasoning shows that

$$p(\lambda_1, \ldots, \lambda_{n-2}) = c_n \, (n-(n-2)) \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^{n-2}\right) = 2\, c_n \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^{n-2}\right)$$

and more generally

$$p(\lambda_1, \ldots, \lambda_m) = (n-m)!\, c_n \det\left(\{K(\lambda_j, \lambda_k\}_{j,k=1}^{m}\right)$$

Finally, in order to compute the normalization constant, notice that for $m = 1$,

$$p(\lambda_1) = (n-1)!\, c_n \, K(\lambda_1, \lambda_1)$$

As we know that $\int_0^\infty d\lambda_1 \, p(\lambda_1) = 1$, this together with property b) of Lemma 1.1 implies that $n!\, c_n = 1$, i.e. $c_n = \frac{1}{n!}$, which completes the proof of the lemma. □

A word on the asymptotic analysis of the eigenvalue distribution. A particular instance of the above result is of course the case $m = 1$, which gives the first-order marginal:

$$p(\lambda) = \frac{1}{n} \, e^{-\lambda} \sum_{l=0}^{n-1} L_l(\lambda)^2$$

As mentioned above, this distribution represents the distribution of a "typical" eigenvalue of W in the "bulk" of the spectrum. Analyzing directly the behavior of $p(\lambda)$ for large values of n is not an easy task. First of all, let us mention that in order to obtain convergence, a rescaling is needed. Indeed:

$$\mathbb{E}(\lambda) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} \lambda_j\right) = \mathbb{E}\left(\frac{1}{n}\mathrm{Tr}(W)\right) = \frac{1}{n}\,\mathbb{E}(\mathrm{Tr}(HH^*)) = \frac{1}{n}\sum_{j,k=1}^{n} \mathbb{E}(|h_{jk}|^2) = n$$

We should then rather consider $p^{(n)}(\lambda)$, the distribution of an eigenvalue of $\frac{1}{n}W$ picked uniformly at random:

$$p^{(n)}(\lambda) = n\, p(n\lambda) = e^{-n\lambda} \sum_{l=0}^{n-1} L_l(n\lambda)^2$$

Studying the asymptotic behavior of this expression as n gets large requires the knowledge of fine properties of Laguerre polynomials, that we skip here. The result gives (we are going to recover this result using a different approach later in the course):

$$\lim_{n\to\infty} p^{(n)}(\lambda) = \frac{1}{\pi}\sqrt{\frac{1}{\lambda} - \frac{1}{4}}\; 1_{0<\lambda<4}$$

which is illustrated on the figures below. On the left, $p^{(n)}(\lambda)$ is represented for $n = 2$ and $n = 4$, while on the right, it is represented for $n = 8$ and $n = \infty$ (spreading the rumor that for random matrices, $8 \sim \infty$):