Scribe: Alex Wardle-Solano
Lecturer: Tim Kunisky

Lecture 12: Free Probability II

1 Notation and Moment Assumptions

Throughout this lecture we will assume that, for each k

$$\frac{1}{N}\mathbb{E}[\mathsf{Tr}(X^{(N)^k})] \to m_k = \mathop{\mathbb{E}}_{x\sim\mu}[x^k], \tag{1}$$

$$\frac{1}{N}\mathbb{E}[\mathsf{Tr}(Y^{(N)^k})] \to n_k = \mathop{\mathbb{E}}_{y\sim\nu}[y^k], \tag{2}$$

that is, that the empirical spectral distributions of $X^{(N)}$ and $Y^{(N)}$ converge in moments to probability distributions μ and ν, respectively, on $\mathbb{R}$.

2 Asymptotic Freeness and its Consequences

We give an alternative definition of asymptotic freeness, simpler than the one from the previous lecture: it suffices to show that, for all $a_1, \dots, a_\ell, b_1, \dots, b_\ell \in \mathbb{N}$, we have

$$\frac{1}{N}\mathbb{E}[\mathsf{Tr}((X^{(N)^{a_1}} - m_{a_1}I)(Y^{(N)^{b_1}} - n_{b_1}I)\dots(X^{(N)^{a_\ell}} - m_{a_\ell}I)(Y^{(N)^{b_\ell}} - n_{b_\ell}I)] \to 0. \tag{3}$$

In the previous definition, we allowed the $X^{(N)^a}$ to be arbitrary polynomials instead of just monomials (and likewise the $Y^{(N)^b}$), but it is not difficult to show that this special class of polynomials suffices.

Consider expanding this condition. We will get a main term of $\frac{1}{N}\mathbb{E}\,\mathsf{Tr}(X^{(N)^{a_1}}\cdots Y^{(N)^{b_\ell}})$, and all further terms will be either the constant $m_{a_1}\cdots n_{b_\ell}\frac{1}{N}\,\mathsf{Tr}(I) = m_{a_1}\cdots n_{b_\ell}$, or traces of lower-degree monomials in $X^{(N)}$ and $Y^{(N)}$. Therefore, we find a *recursion* for these traces, which leads to a formula like that alluded to in the last lecture:

$$\frac{1}{N}\mathbb{E}[\mathsf{Tr}(X^{(N)} + Y^{(N)})^k] \to c_k(m_1, \dots, m_k, \dots, n_1, \dots, n_k), \tag{4}$$

for some polynomial c_k.

Moreover, one can show that there is some probability distribution $\mu \boxplus \nu$ for which these polynomials satisfy

$$c_k(m_1, \dots, m_k, \dots, n_1, \dots, n_k) = \mathop{\mathbb{E}}_{z\sim\mu\boxplus\nu}[z^k]. \tag{5}$$

This is called the *additive free convolution* of μ and ν. Thus, in words, the empirical spectral distribution of $X^{(N)} + Y^{(N)}$ converges in moments to the additive free convolution of the individual limiting empirical spectral distributions of $X^{(N)}$ and $Y^{(N)}$.

3 Additive Free Convolution vs. Classical Convolution

One can compute the additive free convolution by directly working with the c_k polynomials, but an easier approach uses the Stieltjes transform and the related R-transform.

Definition 3.1 (Stieltjes transform)**.** *For a probability distribution μ,*

$$S_\mu(z) = \mathop{\mathbb{E}}_{X \sim \mu}\left[\frac{1}{X - z}\right] = \int \frac{1}{x - z} d\mu(x). \tag{6}$$

Definition 3.2 (R-transform)**.** *For a probability distribution μ,*

$$R_\mu(z) = S_\mu^{-1}(-z) - \frac{1}{z}. \tag{7}$$

The following is the reason why the R-transform is useful for free additive convolution calculations.

Theorem 3.3. *For probability distributions μ and ν,*

$$R_{\mu \boxplus \nu} = R_\mu(z) + R_\nu(z). \tag{8}$$

When considering additive free convolution, it is worth bearing scalar convolution in mind. In the scalar version, the convolution $\mu * \nu$ is the law of $X + Y$ when we draw $X \overset{\text{iid}}{\sim} \mu$, $Y \overset{\text{iid}}{\sim} \mu$. This is analogous to the effect of additive free convolution in the matrix case.

Convolution in the scalar case behaves nicely under the Laplace transform, which can also be though of as an exponential moment generating function or a characteristic function evaluated at a complex input. That is, if we define

$$\phi_\mu(z) = \mathop{\mathbb{E}}_{x \sim \mu}[e^{zx}] = \int e^{xz} d\mu(x), \tag{9}$$

then we have

$$\phi_{\mu * \nu}(z) = \phi_\mu(z)\phi_\nu(z). \tag{10}$$

We can also achieve an additive expression by taking logarithms of both sides. This gives a function known as the cumulant generating function. Cumulants are additive under convolution:

$$\psi_\mu(z) = \log \phi_\mu(z), \tag{11}$$

$$\psi_{\mu * \nu}(z) = \psi_\mu(z) + \psi_\nu(z). \tag{12}$$

For this reason, when we view the R-transform as an ordinary generating function, its coefficients are called *free cumulants*, polynomials of the moments of a distribution which are additive under additive free convolution.

4 Rederiving the Semicircle Law

Armed with our understanding of additive free convolution, we return to the result we have seen before that matrices from the Gaussian orthogonal ensemble have as their limiting empirical spectral distribution the semicircle distribution.

We begin, as before, by splitting our GOE matrix into the normalized sum of two i.i.d. GOE matrices:

$$W \overset{(d)}{=} \frac{W_1 + W_2}{\sqrt{2}} \tag{13}$$

If we had that W_1 and W_2 were asymptotically free, then we could derive the semicircle distribution via additive free convolution, as follows. Given that the empirical spectral distributions of the left and right hand sides must be equal, if ρ were the limiting empirical spectral distribution, we would we have that

$$\rho = \frac{\rho}{\sqrt{2}} \boxplus \frac{\rho}{\sqrt{2}}. \tag{14}$$

Here, $c\mu$ is the distribution of cX when $X \sim \mu$. Our first task is to find an expression for the R-transform of $\frac{\rho}{\sqrt{2}}$ in terms of the R-transform of ρ. In effect, we have to understand how the R-transform scales for constant scaling of its argument. We start from the definition of the Stieltjes transform:

$$S_{c\mu}(z) = \mathop{\mathbb{E}}_{x \sim \mu c}\left[\frac{1}{cx - z}\right] = \frac{1}{z} S_\mu\left(\frac{z}{c}\right) \tag{15}$$

Taking inverses,

$$S_{c\mu}^{-1}(z) = cS_\mu^{-1}(cz) \implies R_{c\mu}^{-1}(z) = cR_\mu^{-1}(cz). \tag{16}$$

Using the above and Theorem 2.3, we get that

$$\begin{aligned} R_\rho(z) &= 2R_{\frac{\rho}{\sqrt{2}}}(s) \\ &= 2 \cdot \frac{1}{\sqrt{2}} R_\rho\left(\frac{z}{\sqrt{2}}\right) \\ &= \sqrt{2} R_\rho\left(\frac{z}{\sqrt{2}}\right) \end{aligned} \tag{17}$$

The only nice functions that satisfy this condition are $R_\rho(z) = az$ for some $a \in \mathbb{R}$. Calculating the first two moments of the GOE, you may check that we must take $a = 1$, so that $R_\rho(z) = z$. It only remains to recover the semicircle distribution from this R-transform. We begin by substituting our condition on $R_\rho(z)$ into the definition of the R-transform.

$$S_\rho^{-1}(z) = z - \frac{1}{z}. \tag{18}$$

Inverting this,

$$S_\rho(y) = \big(z \text{ solving } z^2 + zy + 1 = 0\big) = \frac{-1 \pm \sqrt{y^2 - 4}}{2}. \tag{19}$$

Choosing the correct solution by comparing with low-order moments, we see that this is indeed the Stieltjes transform of the semicircle distribution.

So, as for the scalar Gaussian distribution, the "divisibility" property of the GOE determines its limiting empirical spectral distribution. Thus the semicircle distribution for matrix random variables plays a somewhat analogous role to the normal distribution in the scalar case.

It is worth noting that we can find analogous free distributions for a number of familiar distributions.

Example 4.1 ("Free Poisson" distribution)**.** *The limitof*

$$\underbrace{Ber\left(\frac{\lambda}{n}\right) \boxplus \cdots \boxplus Ber\left(\frac{\lambda}{n}\right)}_{n\ times} \tag{20}$$

gives the Marcenko-Pastur distribution, which arises as the limiting empirical spectral distribution of a Wishart matrix $\frac{1}{n}GG^T, G \in \mathbb{R}^{mxn}$*, where* $G_{ij} \overset{iid}{\sim} N(0,1)$ *and* $\frac{m}{n} \to \lambda$*.*

5 Condition for Asymptotic Freeness

Above we implicitly used the following, which may be proved "by hand" with some tedious moment calculations.

Theorem 5.1. *If* $W_1^{(N)}$ *and* $W_2^{(N)}$ *are i.i.d. normalized matrices drawn from the GOE, then* $W_1^{(N)}$ *and* $W_2^{(N)}$ *are asymptotically free.*

However, this is a very restricted class of matrices. We ask ourselves whether there are weaker sufficient conditions for asymptotic freeness. A natural place to start would be to ask whether independence is sufficient. The answer turns out to be *no*, which we can prove neatly by counterexample.

Consider the diagonal i.i.d. matrices $X^{(N)}$ and $Y^{(N)}$ such that $X_{ii}^{(N)}, X_{ii}^{(N)} \sim \mathsf{Unif}(\{\pm 1\})$. Asymptotic freeness requires that

$$\frac{1}{N}\mathbb{E}[\mathsf{Tr}(X^{(n)}Y^{(n)}X^{(n)}Y^{(n)})] \to 0. \tag{21}$$

However, due to the commutative property of diagonal matrices, we can write,

$$\begin{aligned} \frac{1}{N}\mathbb{E}[\mathsf{Tr}(X^{(n)}Y^{(n)}X^{(n)}Y^{(n)})] &= \frac{1}{N}\mathbb{E}[\mathsf{Tr}\,\underbrace{(X^{(N)^2}Y^{(N)^2})}_{I}] \\ &= 1 \neq 0 \end{aligned} \tag{22}$$

It turns out that, in general, asymptotic freeness between two matrices requires strong non-commutativity or eigenvector decorrelation.

Theorem 5.2 (Voiculescu)**.** *If* $X^{(N)}$ *and* $Y^{(N)}$ *are independent and* $X^{(N)}$ *is rotationally invariant, then* $X^{(N)}$ *and* $Y^{(N)}$ *are asymptotically free.*

This ensures "as much eigenvector decorrelation as possible," since the eigenvectors of $X^{(N)}$ are a uniformly random orthonormal basis. We note also that for independence it suffices to take $Y^{(N)}$ a deterministic sequence of matrices whose empirical spectral distribution converges. Examples of rotationally invariant random matrices $X^{(N)}$ include: GOE matrices, Wishart matrices,, projections to uniformly random subspaces, $Q\widetilde{X}^{(N)}Q^T$ for $\widetilde{X}^{(N)}$ arbitrary and $Q \sim \mathsf{Unif}(\mathcal{O}(N))$, where $\mathcal{O}(N)$ is the set of $N \times N$ orthogonal matrices.

6 Multiplicative Free Convolution

We briefly introduce the related concept of multiplicative free convolution, denoted $\boxtimes$. If we have $X^{(N)}$, $Y^{(N)}$ asymptotically free and $Y^{(N)} \succeq 0$, then we can compute the limiting empirical spectral distribution of matrices of the form,

$$Y^{(N)^{\frac{1}{2}}} X^{(N)} Y^{(N)^{\frac{1}{2}}} =: Z^{(N)}. \tag{23}$$

This is because

$$\mathsf{Tr}\, Z^{(N)^k} = \mathsf{Tr}((X^{(N)}Y^{(N)})^k) \tag{24}$$

after applying the cyclic property of trace, which returns us to the setting of traces of polynomials that asymptotic freeness helps analyze.

In this case, the empirical spectral distribution of $Z^{(N)}$ converges to a measure $\mu \boxtimes \nu$. We do not give a thorough account of multiplicative free convolution, but let us give a motivating example.

Example 6.1. *Consider some rotationally invariant $X^{(N)}$ and diagonal $Y^{(N)}$ such that $Y_{ii} \overset{iid}{\sim}$ $\mathsf{Ber}(\alpha)$. Given that Bernoulli random variables take only values 0 and 1, our matrix has a random set of ones and zeros in the diagonal and all-zero off-diagonal elements, and $Y^{(N)^{\frac{1}{2}}} = Y^{(N)}$. Then, we can think of the matrix $Y^{(N)^{\frac{1}{2}}} X^{(N)} Y^{(N)^{\frac{1}{2}}} = Y^{(N)} X^{(N)} Y^{(N)} = Z^{(N)}$ as being a random principal submatrix of $X^{(N)}$. Thus multiplicative free convolution with $\mathsf{Ber}(\alpha)$ can help us understand how sub-sampling affects the spectrum of a matrix; this has some useful application in signal processing and graph theory.*