8 Regular and free probability

Free probability is a generalization of probability theory to objects that do not commute, like matrices. It provides a number of useful tools that allow quick analysis of the resolvents of sums and products of random matrices if you already know the resolvent of the individual matrices.

Regular probability

Before we enter into discussion of free probability, we will review some features of regular probability theory so that we know what to expect. First, recall that two random variables are *independent* if, for any functions f and g,

$$\overline{f(X)g(Y)} = \overline{f(X)} \times \overline{g(Y)} \tag{1}$$

Alternatively, independence can be defined by saying that for any integers n and m,

$$\overline{(X^n - \overline{X^n})(Y^m - \overline{Y^m})} = 0 \tag{2}$$

This property leads to classic factorization rules, e.g.,

$$\begin{aligned} \overline{X^2Y^2} &= \overline{(X^2 - \overline{X^2})(Y^2 - \overline{Y^2})} + \overline{X^2} \times \overline{Y^2} + \overline{X^2} \times \overline{Y^2} - \overline{X^2} \times \overline{Y^2} \\ &= \overline{X^2} \times \overline{Y^2} \end{aligned} \tag{3}$$

If X is a random variable with probability density function (PDF) p_X, then its *moment generating function* is given by

$$M_X(t) = \overline{e^{tX}} = \int dx\, p_X(x) e^{tx} \tag{4}$$

It is called the moment generating function because its Taylor coefficients are the moments $m_n = \overline{X^n}$ of the distribution of X, with

$$M_X(t) = \int dx\, p_X(x) e^{tx} = \int dx\, p_X(x) \sum_{n=0}^{\infty} \frac{(tx)^n}{n!} = \sum_{n=0}^{\infty} \frac{\overline{x^n} t^n}{n!} = \sum_{n=0}^{\infty} \frac{m_n t^n}{n!} \tag{5}$$

This means that the nth derivative of M_X evaluated at zero is the nth moment of the distribution of X, with

$$M_X^{(n)}(0) = m_n \tag{6}$$

The moments of two independent random variables are not additive and, beyond the first three, neither are the central moments. The *cumulant generating function* is defined as the logarithm of the moment generating function, with

$$K_X(t) = \log M_X(t) = \log \overline{e^{tX}} \tag{7}$$

The Taylor series of the cumulant generating function gives

$$K_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} K_X^{(n)}(0) = \sum_{n=0}^{\infty} \frac{\kappa_n t^n}{n!} \tag{8}$$

where κ_n is the nth cumulant of X. The first several cumulants are

$$\kappa_1 = m_1 \qquad \kappa_2 = m_2 - m_1^2 = \mu_2 \qquad \kappa_3 = m_3 - 3m_2m_1 + 2m_1^3 = \mu_3 \tag{9}$$

$$\kappa_4 = m_4 - 4m_3m_1 - 3m_2^2 + 12m_1^2m_2 - 6m_1^4 = \mu_4 - 3\mu_2^2 \quad \kappa_5 = \mu_5 - 10\mu_3\mu_2 \tag{10}$$

where $\mu_n = \overline{(X - \overline{X})^n}$ is the nth central moment of X. So, the first few cumulants correspond to the central moments, e.g., the mean, variance, and skewness, but higher cumulants do not.

You may recall the cumulant generating function from field theory, where it plays an important role because the diagrams that contribute to its coefficients are only one-line irreducible ones, whereas the diagrams that contribute to the coefficients of the moment generating function are all of them. Besides its role in simplifying field theory calculations, the cumulants and their generating function have the important property that they are additive (cumulative) under addition of independent random variables. This is because

$$\begin{aligned} K_{X+Y}(t) &= \log M_{X+Y}(t) = \log \overline{e^{t(X+Y)}} = \log \overline{e^{tX}e^{tY}} = \log \overline{e^{tX}} \times \overline{e^{tY}} \\ &= \log(\overline{e^{tX}} \times \overline{e^{tY}}) = \log \overline{e^{tX}} + \log \overline{e^{tY}} = \log M_X(t) + \log M_Y(t) \\ &= K_X(t) + K_Y(t) \end{aligned} \tag{11}$$

Since the Taylor coefficients of the sum of two functions are the sum of the coefficients of the individual functions, we also have $\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y)$. The additivity of cumulants is an important signature of independence of random variables. You can consider the property of cumulants that

$$\kappa_n = \mu_n + (\text{polynomial of lower-order central moments}) \tag{12}$$

and that they sum for sums of independent variables to uniquely define them.

If you have two random variables and know their PDFs, you can find the PDF of their sum by taking the inverse Fourier transform of each to find φ_X and φ_Y, take their logarithm to find H_X and H_Y, sum them to find H_{X+Y}, exponentiate to find φ_{X+Y}, and finally take the forward Fourier transform to find p_{X+Y}.

Finally, note that when a random variable is multiplied by a constant, the cumulant generating function has the scaling property that

$$K_{aX}(t) = \log \overline{e^{taX}} = K_X(at) \tag{13}$$

It follows from their Taylor series definition that the cumulants are transformed by $\kappa_n(aX) = a^n\kappa(X)$.

We can use the additive properties of cumulants to quickly prove the law of large numbers and the central limit theorem. The law of large numbers states that the sample mean resulting from summing N independent and identically distributed random variables approaches the mean of their distribution. The sample mean is defined by

$$\mu_M = \frac{1}{M}\sum_{i=1}^{M} X_i \tag{14}$$

Because it is the sum of independent random variables, the cumulants of the distribution of the sample mean are given by the sum of the cumulants of its parts, or

$$\kappa_n(\mu_M) = \sum_{i=1}^{M} \kappa_n(M^{-1}X_i) = M^{-n}\sum_{i=1}^{M} \kappa_n(X_i) = M^{-n}\sum_{i=1}^{M} \kappa_n(X) = M^{-n+1}\kappa_n(X) \tag{15}$$

Therefore, for large M, $\kappa_1(\mu_M) = \kappa_1(X)$ and $\kappa_n(\mu_M) = 0$ for all $n \geq 2$. Since the first cumulant is the mean, this says that the mean of the sample mean is the same as the mean, and since all other cumulants are zero, the distribution of sample means is a δ function on its mean value.

The central limit theorem says that the average of many centered IID random variables is centered Gaussian. Specifically, it says that

$$S_M = \frac{1}{\sqrt{M}}\sum_{i=1}^{M} X_i \tag{16}$$

is Gaussian. The centered Gaussian distribution has nonzero second cumulant while every other cumulant is zero. Assuming X is centered, i.e., $\kappa_1 = 0$, then

$$\kappa_n(S_M) = \sum_{i=1}^{M} \kappa_n(M^{-\frac{1}{2}}X_i) = M^{-\frac{n}{2}+1}\kappa_n(X) \tag{17}$$

Therefore, for large M, $\kappa_2(S_M) = \kappa_2(X)$ and $\kappa_n(S_M) = 0$ for all other n. This implies that S_M is Gaussian with variance $\kappa_2(X)$.

Free probability

Free probability was developed around the property of *freeness*, which generalizes independence to noncommuntative variables, like matrices. First, define the regularized trace of an $N \times N$ matrix X as

$$\tau(X) = \sum_{N\to\infty} \frac{1}{N}\operatorname{Tr} X \tag{18}$$

Because of the self-averaging of large random matrices, this normalized trace in the large-N limit has all the properties of an expectation value of a random variable. In particular, one can write the moments of the spectral density as

$$m_n = \tau(X^n) = \lim_{N\to\infty}\frac{1}{N}\operatorname{Tr} X^n = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^n = \overline{\lambda^n} \tag{19}$$

One might be tempted to think of two matrices as being independent if, like for real-valued random variables,

$$\tau((X^n - \tau(X^n))(Y^m - \tau(Y^m))) = 0 \tag{20}$$

for all integers n and m. However, this is not really sufficient to recover the properties of independent random variables, since it has nothing to say about, e.g., $XYXY$, which is not necessarily equal to X^2Y^2. Therefore, we say that X and Y are *free* if for any set of integers $n_1, n_2, \ldots$ and $m_1, m_2, \ldots$,

$$\tau((X^{n_1} - \tau(X^{n_1}))(Y^{m_1} - \tau(Y^{m_1}))(X^{n_2} - \tau(X^{n_2}))(Y^{m_2} - \tau(Y^{m_2}))\cdots) = 0 \tag{21}$$

The factorization properties implied by this definition are more nontrivial than those implied by standard independence. For instance, while

$$\tau(XY) = \tau((X-\tau(X))(Y-\tau(Y))+\tau(X\tau(Y))+\tau(\tau(X)Y)-\tau(X)\tau(Y) = \tau(X)\tau(Y) \tag{22}$$

consider

$$\tau(XYXY) = \tau(X)^2\tau(Y^2) + \tau(X^2)\tau(Y)^2 - \tau(X^2)\tau(Y^2) \neq \tau(X^2)\tau(Y^2) = \tau(X^2Y^2) \tag{23}$$

This may seem like a very strong requirement, but importantly it is realized if X is drawn from a rotationally invariant ensemble and Y is any matrix. If X is from a rotationally invariant ensemble, then OXO^T for random orthogonal O is equivalent statistically to X. Then the first expression is equivalent to

$$\tau((X^{n_1}-\tau(X^{n_1}))O^T(Y^{m_1}-\tau(Y^{m_1}))O(X^{n_2}-\tau(X^{n_2}))O^T(Y^{m_2}-\tau(Y^{m_2}))O\cdots) \tag{24}$$

which is the normalized trace of traceless matrices with a random orthogonal matrix interspersed between them. One can show that in the large-N limit, such products are always zero.

If we want to understand how to go from properties of free matrices to properties of their sum, we want to establish something like the equivalent of the characteristic function for random matrices. A good candidate is the so-called *Harish–Chandra–Itzykson–Zuber* (HCIZ) integral. In its most general form, it is defined for two matrices X and T by

$$I(X,T) = \left\langle e^{\frac{N}{2}\operatorname{Tr} TOXO^T} \right\rangle_O \tag{25}$$

where the average is over all orthogonal matrices O. This integral naturally factorizes like the characteristic function for sums of free matrices. Consider again $O'XO'^T$ and Y for some random O'. Since the spectrum of X does not depend on O', averaging over O' or not is irrelevant. Then

$$\begin{aligned} I(O'XO'^T + Y, T) &= \left\langle e^{\frac{N}{2}\operatorname{Tr} TO(O'XO'^T+Y)O^T} \right\rangle_{O,O'} \\ &= \left\langle e^{\frac{N}{2}\operatorname{Tr} TO''XO''^T + \frac{N}{2}\operatorname{Tr} TOYO^T} \right\rangle_{O,O''} = I(X,T)I(Y,T) \end{aligned} \tag{26}$$

where we defined $O'' = OO'$, an independent orthogonal matrix from O, which factorizes the average. This result is general in T, but for our purposes we only need $T = t\boldsymbol{v}\boldsymbol{v}^T$ a rank-one matrix with $\|\boldsymbol{v}\|^2 = 1$. When this is the case, we can write $\boldsymbol{s} = \sqrt{Nt}O^T\boldsymbol{v}$, and the integral is simply

$$I_X(t) = \left\langle e^{\frac{1}{2}\boldsymbol{s}^T X \boldsymbol{s}} \right\rangle_{\|\boldsymbol{s}\|^2 = Nt} = \frac{\int d\boldsymbol{s}\, \delta(Nt - \|\boldsymbol{s}\|^2) e^{\frac{1}{2}\boldsymbol{s}^T X \boldsymbol{s}}}{\int d\boldsymbol{s}\, \delta(Nt - \|\boldsymbol{s}\|^2)} \tag{27}$$

The equivalent of the cumulent generating function would be

$$H_X(t) = \frac{2}{N} \log I_X(t) \tag{28}$$

Using the property of I shown above, $H_{X+Y} = H_X + H_Y$ for free matrices X and Y. For simple ensembles $H_X(t)$ is simple enough to compute directly, but we would like to make a more abstract calculation to connect it with the resolvent. First, exponentiate the δ function:

$$I_X(t) \propto \frac{\int d\boldsymbol{s}\, dz\, e^{\frac{1}{2}\boldsymbol{s}^T X \boldsymbol{s} + \frac{1}{2}(Ntz - z\|\boldsymbol{s}\|^2)}}{e^{\frac{N}{2}(1+\log t)}} = \frac{\int d\boldsymbol{s}\, dz\, e^{-\frac{1}{2}\boldsymbol{s}^T (zI - X)\boldsymbol{s} + \frac{1}{2}Ntz}}{e^{\frac{N}{2}(1+\log t)}} \tag{29}$$

We have also evaluated the denominator to largest order in N, which is just the value of the N-sphere of radius $\sqrt{Nt}$ we have seen now a few times. You can see the resolvant trying to appear here. Next, we can perform the Gaussian integral in $\boldsymbol{s}$, giving

$$I_X(t) \propto \frac{\int dz\, \det(zI - X)^{-\frac{1}{2}} e^{\frac{1}{2}Ntz}}{e^{\frac{N}{2}(1+\log t)}} = \frac{\int dz\, e^{\frac{N}{2}(tz - \frac{1}{N}\sum_i \log(z - \lambda_i))}}{e^{\frac{N}{2}(1+\log t)}} \tag{30}$$

where we have used the fact that the determinant is the product of the eigenvalues and then brought them into the exponential with logarithms, and λ_i are the eigenvalues of X. This is an integral in z we can evaluate with saddle-point, which gives

$$0 = \frac{\partial S}{\partial z} = t - \frac{1}{N}\sum_i \frac{1}{z - \lambda_i} = t - G_X(z) \tag{31}$$

where we have recognized the resolvant of X. The saddle-point value of z is therefore given by $z = B_X(t)$, where B_X is the *blue function* and the inverse of G_X, with $G_X(B_X(t)) = t$. Wrapping up, we have

$$H_X(t) = tz - \frac{1}{N}\sum_i \log(z - \lambda_i) - 1 - \log t = tB_X(t) - \frac{1}{N}\sum_i \log(B_X(t) - \lambda_i) - 1 - \log t \tag{32}$$

Finally, we differentiate with respect to t to arrive at

$$\begin{aligned} R_X(t) = \frac{\partial H_X(t)}{\partial t} &= B_X(t) + tB'_X(t) - \frac{1}{N}\sum_i \frac{1}{B_X(t) - \lambda_i} B'_X(t) - \frac{1}{t} \\ &= B_X(t) + tB'_X(t) - G_X(B_X(t))B'_X(t) - \frac{1}{t} = B_X(t) - \frac{1}{t} \end{aligned} \tag{33}$$

where we have defined the *R-transform*. Since $H_{X+Y} = H_X + H_Y$ for free X and Y, $R_{X+Y} = R_X + R_Y$ for free X and Y as well. Note that since $H_X(0) = 0$, we can write

$$H_X(t) = \int_0^t dt' \, R_X(t') \tag{34}$$

Let's make this a little more concrete. Recall that if X is GOE, then the resolvent was given by the equation

$$0 = \frac{1}{2} G_X(z)^2 - z G_X(z) + 1 \tag{35}$$

Since the blue function is the inverse of the resolvent, it is given by replacing every z with $B_X(t)$ and every $G_X(z)$ with t, or

$$0 = \frac{1}{2} t^2 - t B_X(t) + 1 \tag{36}$$

This linear equation is solved to give

$$B_X(t) = \frac{1}{2} t + \frac{1}{t} \tag{37}$$

and therefore the R-transform is

$$R_X(t) = B_X(t) - \frac{1}{t} = \frac{1}{2} t \tag{38}$$

and its cumulent generating function is

$$H_X(t) = \int_0^t dt' \, R_X(t') = \frac{1}{2} \int_0^t dt' \, t' = \frac{1}{4} t^2 \tag{39}$$

The R-transform is extremely useful because it is algebraically related to the resolvent. If X and Y are random matrices whose resolvents we know, we can find the resolvent of $X + Y$ and therefore the spectral density by computing R_X and R_Y, writing R_{X+Y} as their sum, and then solving back for the resolvent.

$H_X(t)$ generates so-called *free cumulants*, which like for regular random variables generalize moments in a way that is cumulative. We see from the above GOE example that the free cumulants of the GOE are $\kappa_2 = \frac{1}{2}$ for the second moment (in general different value depending on the variance of the matrix elements) and zero for all others. We see that the semicircle distribution is characteristic of matrices with such free cumulents. But, we can repeat the argument we made regarding the central limit theorem for real random variables verbatim for matrices, and the result is the same: if we sum together M free matrices with

$$S_M = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i \tag{40}$$

where the X_i are identically distributed with *any* distribution with finite moments and zero first moment, then S_M will belong to an ensemble of matrices with $\kappa_2(S_M) = \kappa_2(X)$ and all other $\kappa_n = 0$. Here, we see the universality of the semicircle: it is the limiting distribution for sums of free matrices with finite moments.