Computations of Linear Statistics for Ensembles of Random Matrices

September 2006

Introduction

We begin with two formulas.

The first formula, first proved by Andréief in 1883, says

$$\frac{1}{N!}\int\cdots\int \det(f_j(x_k))\det(g_j(x_k))dx_1\cdots dx_N$$

$$= \det\left(\int f_j(x)g_k(x)dx\right)_{j,k=1,\cdots N}$$

This can be proved using elementary properties of determinants.

The second formula comes from the Circular Unitary Ensemble (CUE) in Random Matirx Theory (RMT). By CUE we mean the group of $N \times N$ unitary matrices with probability measure being normalized Haar measure.

The Haar measure induces a probability distribution on the space of eigenangles $(\theta_1, \ldots, \theta_N)$ whose density is given by

$$\frac{1}{N!} \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2.$$

Now suppose we have a random variable of the form

$$\sum_{j=1}^{N} f(e^{i\theta_j})$$

with f real-valued.

This kind of random variable is called a linear statistic and has been studied extensively for different ensembles of random matrices.

We know that the Fourier transfom of the probability density function of the random variable is given by

$$g(\lambda) = \frac{1}{(2\pi)^N N!} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i\lambda \sum_{j=1}^{N} f(e^{i\theta_j})}$$

$$\times \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2 d\theta_1 \ldots d\theta_N$$

$$= \frac{1}{(2\pi)^N N!} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \prod_{j=1}^{N} e^{i\lambda f(e^{i\theta_j})} \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2 d\theta_1 \ldots d\theta_N$$

We notice

$$\prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2 = \prod_{j<k} (e^{i\theta_j} - e^{i\theta_k})((e^{-i\theta_j} - e^{-i\theta_k})$$

$$= \prod_{j<k} (e^{i\theta_j} - e^{i\theta_k}) \prod_{j<k} (e^{-i\theta_j} - e^{-i\theta_k}).$$

The product

$$\prod_{j<k} (e^{i\theta_j} - e^{i\theta_k})$$

is a Vandermonde determinant

$$\det((e^{i\theta_k})^{j-1})_{j,k=1}^{N}$$

as is

$$\prod_{j<k} (e^{-i\theta_j} - e^{-i\theta_k}) = \det((e^{-i\theta_k})^{j-1})_{j,k=1}^{N}.$$

Thus our formula for $g(\lambda)$ becomes

$$\frac{1}{(2\pi)^N N!}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\prod_{j=1}^{N} e^{i\lambda f(e^{i\theta_j}}$$

$$\times \det((e^{i\theta_k})^{j-1})\det((e^{-i\theta_k})^{j-1})d\theta_1\ldots d\theta_N.$$

We can incorporate a factor of the form $e^{i\lambda f(e^{i\theta_j})}$ into a column or row of either determinant in the last integral

to obtain

$$\frac{1}{(2\pi)^N N!}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\det((e^{i\theta_k})^{j-1})\det((e^{-i\theta_k})^{j-1}e^{i\lambda f(e^{i\theta_k})})d\theta_1\ldots d\theta_N.$$

We apply Andréief's formula

$$\frac{1}{N!}\int\cdots\int \det(f_j(x_k))\det(g_j(x_k))dx_1\cdots dx_N$$

$$= \det\left(\int f_j(x)g_k(x)dx\right)_{j,k=1,\cdots N}$$

to find

$$g(\lambda) = \det\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda f(\theta)}e^{ik\theta}e^{-ij\theta}d\theta\right)_{k,j=0}^{N-1}$$

or

$$g(\lambda) = \det\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda f(\theta)}e^{i(k-j)\theta}d\theta\right)_{k,j=0}^{N-1}.$$

In other words, the Fourier transform (or characteristic function) of the density function of a linear statistic for CUE is a Toeplitz determinant, that is, a determinant of a matrix whose k, j entry depends only on the difference of k and j. Much is known about the asymptotic expansion of such a determinant.

In what follows subscripts denote Fourier coefficients.

The strong Szegö limit theorem states that if the symbol ϕ defined on the unit circle has a sufficiently well-behaved logarithm then the determinant of the Toeplitz matrix

$$T_N(\phi) = (\phi_{j-k})_{j,k=0,\cdots,N-1}$$

has the asymptotic behavior

$$D_N(\phi) = \det T_N(\phi) \sim G(\phi)^N \, E(\phi) \quad \text{as } N \to \infty$$

where

$$G(\phi) = e^{(\log\phi)_0}$$

$$E(\phi) = \exp\left(\sum_{k=1}^{\infty} k\,(\log\phi)_k\,(\log\phi)_{-k}\right).$$

To obtain asymptotic information about the linear statistic we apply the Strong Szegö Limit Theorem. This shows

$$g(\lambda) \sim G(\phi)^N E(\phi), \quad \phi(e^{i\theta}) = e^{i\lambda f(e^{i\theta})}$$

where

$$G(\phi)^N = \exp\left(i\lambda\frac{N}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})d\theta\right)$$

and

$$E(\phi) = \exp\left(-\lambda^2 \sum_{k=1}^{\infty} k f_k f_{-k}\right).$$

We see that we can interpret the last formula as saying that asymptotically as $N \to \infty$: For a smooth function f the distribution of

$$S_N - N\mu$$

where

$$S_N = \sum_{j=1}^{N} f(e^{i\theta_j}),$$

and

$$\mu = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta}) d\theta$$

converges weakly to a Gaussian distribution with mean zero and variance given by

$$\sigma^2 = \sum_{1}^{\infty} k f_k f_{-k} = \sum_{1}^{\infty} k |f_k|^2$$

(The last equality holds since f is real-valued.)

For functions f that are discontinuous, we generally are not able to apply the Strong Szegö Limit Theorem. However, there are some interesting linear statistics that are important in RMT that correspond to singular symbols. For example, consider

$$f(e^{i\theta}) = \chi_I(e^{i\theta}) = \begin{cases} 1 & \text{if } e^{i\theta} \in I \\ 0 & \text{otherwise} \end{cases}$$

This random variable counts the number of eigenvalues in an arc I on the circle.

In 1968, Fisher and Hartwig stated a conjecture about a certain class of Toeplitz matrices with singular symbols. If

$$\phi_{\alpha,\beta}(e^{i\theta}) = (2 - 2\cos\theta)^{(\alpha+\beta)/2}\, e^{i(\theta-\pi)(\alpha-\beta)/2}$$

for $0 < \theta < 2\pi$ then their symbols had the form

$$\psi(e^{i\theta}) = \varphi(e^{i\theta}) \prod_{j=1}^{R} \phi_{\alpha_j,\beta_j}(e^{i(\theta-\theta_j)})$$

where φ satisfies the assumption of Szegö's theorem and $\theta_1, \cdots, \theta_R$ are distinct points on the unit circle.

They conjectured that for some range of the parameters the asymptotics had the form

$$\det T_N(\psi) \sim G(\psi)^N \, N^{\sum \alpha_j \beta_j} \, E(\varphi, \alpha_j, \beta_j, \theta_j)$$

where $E(\varphi, \alpha_j, \beta_j, \theta_j)$ is a constant (whose value they did not conjecture).

The corresponding symbol for the Toeplitz determinant representing the Fourier transform of the density for the linear statistic corresponding to counting the number of eigenvalues in an interval is $\phi(e^{i\theta}) = e^{i\lambda\chi_I}$.

This is a Fisher-Hartwig symbol.

To compute the correct α, β parameters, we note this function only has two jump discontinuities so $\alpha_j = -\beta_j, j = 1, 2$. To compute the β parameters notice that for our standard factor $\phi_{\beta,-\beta}$

$$\beta = \frac{1}{2\pi i} \log\left(\frac{\phi(1^-)}{\phi(1^+)}\right).$$

For

$$\phi(e^{i\theta}) = e^{i\lambda\chi_I}$$

with I equal to the arc from $e^{-i\gamma}$ to $e^{i\gamma}$ we have two jumps with

$$\beta_1 = \frac{1}{2\pi i}\log\left(\frac{1}{e^{i\lambda}}\right) = -\lambda/2\pi$$

at the point $e^{-i\gamma}$ and

$$\beta_2 = \frac{1}{2\pi i}\log\left(e^{i\lambda}\right) = \lambda/2\pi$$

at the point $e^{i\gamma}$. It is straight forward to check that

$$e^{i\lambda\chi_I} = e^{i\lambda\gamma/\pi}\phi_{-\lambda/2\pi,\lambda/2\pi,\gamma}\ \ \phi_{\lambda/2\pi,-\lambda/2\pi,-\gamma}.$$

If we apply the Fisher-Hartwig results directly, we have that

$$D_N(\phi) \sim e^{\frac{i\lambda N\gamma}{\pi}} N^{-\frac{\lambda^2}{2\pi^2}}(2-2\cos 2\gamma)^{\frac{\lambda^2}{4\pi^2}} G\left(1-\frac{\lambda}{2\pi}\right) G\left(1+\frac{\lambda}{2\pi}\right)$$

In the above $G(1+z)$ is the Barnes G-function satisfying $G(1+z) = \Gamma(z)G(z)$.

This does not have the nice Gaussian form as before. Notice, though,

$$N^{-\frac{\lambda^2}{2\pi^2}} = e^{-\frac{\lambda^2}{2\pi^2}\log N}$$

so that σ^2 is on the order of $(1/2\pi^2)\log N$.

Thus if we "re-scale" our random variable in a fairly natural way to be of the form

$$\frac{1}{\sqrt{\frac{\log N}{2\pi^2}}} \sum_{j=1}^{N} (\chi_I(e^{i\theta_j}) - \gamma/\pi)$$

then $g(\lambda)$ to tends to $e^{-\lambda^2}$.

Sufficient uniformity in the estimates used to prove the Fisher-Hartwig conjecture in the case of jump discontinuities is required for this result. This holds in a certain range of λ, say for $|\lambda/\sqrt{\frac{\log N}{2\pi^2}}| \leq c < 1/2$. This can be checked by a careful analysis of the estimates and convergence tools used in proving the conjecture.

Another Application of Fisher-Hartwig

Define

$$Z = \det(I - U) = \prod_{j=1}^{N} (1 - e^{i(\theta_j})$$

This is the characteristic polynomial of a random unitary matrix evaluated at the point one.

Let us also define

$$X(\theta) = \mathrm{Re}\, Z(\theta), \quad Y(\theta) = \mathrm{Im}\, Z(\theta),$$

and $\phi_N(s,t) =$

$$\frac{1}{(2\pi)^N N!} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{isX+itY} \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2 d\theta_1 \dots d\theta_N$$

One can show using the same identity of Andréief that this again a Toeplitz determinant with a Fisher-Hartwig symbol and with one singularity. The α, β parameters turn out to satisfy

$$\alpha + \beta = is \quad \alpha - \beta = t$$

Rescaling by $\sqrt{\log N/4}$ we obtain that $\phi_N(s,t)$ converges pointwise to

$$e^{-s^2-t^2}$$

Other Ensembles

Much of what has been done here can be repeated for other ensembles of random matrices.

It turns out for the Gaussian Unitary Ensemble, for the study of linear statistics it is necessary to consider a finite Wiener-Hopf operator, $W_\alpha(\phi)$, defined on $L^2(0,\alpha)$ by

$$P_\alpha \mathcal{F}^{-1} M_\phi \mathcal{F} P_\alpha$$

where P_α is multiplication by the characteristic function of $(0,\alpha)$, and $\mathcal{F}$ is the Fourier transform. For linear statistics the important quantity is

$$\det(I + W_\alpha(\phi))$$

where $\phi = e^{i\lambda f} - 1$. (The above determinant is well-defined for sufficiently nice ϕ.)

The analogue of the Strong Szegö Limit Theorem says that if $\phi = e^b - 1$ then as $\alpha \to \infty$

$$\det(I + W_\alpha(\phi)) \sim$$

$$\exp\left(\frac{\alpha}{2\pi}\int_{-\infty}^{\infty} b(x)dx + \int_0^{\infty} x\hat{b}(x)\hat{b}(-x)dx\right),$$

where $\hat{b}(x)$ is the Fourier transform of b. This formula again implies that for smooth f the linear statistics are asymptotically Gaussian. Analogues of this theorem have also recently been proved for Fisher-Hartwig type symbols.

Laguerre Ensembles and the Airy Operator

In RMT Laguerre ensembles are defined on the space of positive Hermitian matrices and for these ensembles, the study of linear statistics lead to the study of Bessel operators $B_\alpha(\phi)$ defined on $L^2(0,\alpha)$ by

$$P_\alpha \mathcal{H}_\nu M_\phi \mathcal{H}_\nu P_\alpha$$

where $\mathcal{H}_\nu$ is the Hankel transform of order ν given by

$$H_\nu(f)(x) = \int_0^\infty \sqrt{tx}\, J_\nu(tx) f(t)\, dt,$$

and J_ν is the Bessel function of order ν. If $\phi = e^b - 1$, then

$$\det(I + B_\alpha(\phi)) \sim$$

$$\exp\left(\frac{\alpha}{2\pi}\int_{-\infty}^\infty b(x)dx - \frac{\nu}{2} + \frac{1}{2}\int_0^\infty x(\hat{b}(x))^2 dx\right).$$

Some results are known in this case for singular symbols as well, but only in the case of $\nu = \pm\frac{1}{2}$.

If we scale in GUE ensemble on the edge of the spectrum, linear statistics problems reduce to the study of the Airy operators, integral operators on $L^2(0, \alpha)$ with kernel

$$A_\alpha(f)(x,y) = f(x/\alpha) \int_0^\infty A(x+z)A(z+y)dz$$

where $A(x)$ is the Airy function. (This operator also has an equivalent definition in terms of multiplication and an "Airy" transform.) The asymptotic formula reads

$$\det(I + A_\alpha(f)) \sim \exp\left(c_1 \alpha^{3/2} + c_2\right)$$

where

$$c_1 = \frac{1}{\pi} \int_0^\infty \sqrt{x} \log(1 + f(-x)) \, dx$$

and

$$c_2 = \frac{1}{2} \int_0^\infty x G(x)^2 dx$$

$$G(x) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{ixy} \log(1 + f(-y^2)) dy.$$

Results for discontinuous symbols are not yet known in the Airy case.

In all of the above cases, CUE, GUE, Laguerre, and Airy there is always a Szegö type limit theorem for a particular class of operators which implies that after scaling, linear statistics have a Gaussian or normal distribution in the limit, at least for smooth symbols. Hence there seems to exist a universality or central limit theorem for such quantities.