7. The special case of $\beta = 2$

Consider the GUE ensemble density

$$g(\lambda) = \exp\left\{-\frac{1}{4}\sum_{k=1}^{n}\lambda_k^2\right\}\prod_{i<j}|\lambda_i - \lambda_j|^2$$

where Z_n is the normalizing constant. More generally, let μ be a Borel probability measure on $\mathbb{R}$ and consider the density f on $\mathbb{R}^n$ proportional to $|\Delta(x)|^2$ with respect to the measure $\mu^{\otimes n}$. All of what we say in this section will apply to this more general density[6]. This is symmetric in $\lambda_1, \dots, \lambda_n$. The following lemma shows that it is possible to explicitly integrate out a subset of variables and get marginal densities of any subset of the co-ordinates. As we discussed earlier, this is crucial to computing local properties of the system of particles defined by the density f.

Observation: Let $p_0, p_1, \dots, p_{n-1}$ be monic polynomials such that p_k has degree k. Then,

$$\Delta(x) = \det\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix} = \det\begin{bmatrix} p_0(x_1) & p_1(x_1) & p_2(x_1)\dots & p_{n-1}(x_1) \\ p_0(x_2) & p_1(x_2) & p_2(x_2)\dots & p_{n-1}(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ p_0(x_n) & p_1(x_n) & p_2(x_n)\dots & p_{n-1}(x_n) \end{bmatrix}$$

as can be seen by a sequence of column operations. If φ_k is any polynomial with degree k and having leading coefficient c_k, then we get $\Delta(x) = C_n \det(A)$ where $a_{i,j} = \varphi_j(x_i)$ with the index i running from 0 to $n-1$ and the index j from 1 to n. The constant $C_n = (c_0 c_1 \dots c_{n-1})^{-1}$. Thus,

$$|\Delta(x)|^2 = C_n^2 \det\left(A\,A^t\right) = C_n^2 \det\left(K_n(x_i, x_j)\right)_{i,j\le n}$$

where $K_n(x,y) = \sum_{j=0}^{n-1}\varphi_j(x)\varphi_j(y)$. It turns out that choosing φ_j to be the orthogonal polynomials with respect to μ enables us to integrate out any subset of variables explicitly!

Lemma 56. *Let $(A, \mathcal{A}, \mu)$ be a measure space. Let φ_k, $1 \le k \le n$, be an orthonormal set in $L^2(\mu)$ and define $K(x,y) = \sum_{k=1}^{n}\varphi_k(x)\overline{\varphi}_k(y)$. Define $f : A^n \to \mathbb{R}$ by*

$$f(x) = (n!)^{-1}\det\left(K(x_i,x_j)\right)_{i,j\le n}.$$

(1) *For any $m \le n$ and any λ_k, $k \le m-1$, we have*

$$\int_{\mathbb{R}} \det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le m}\mu(d\lambda_m) = (n-m+1)\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le m-1}.$$

(2) *f is a probability density on A^n with respect to $\mu^{\otimes n}$. Further, if $(\lambda_1,\dots,\lambda_n)$ is a random vector in $\mathbb{R}^n$ with density f, then λ_i are exchangeable, and for any $m \le n$, the density of $(\lambda_1,\dots,\lambda_m)$ with respect to $\mu^{\otimes m}$ is given by*

$$f_k(\lambda_1,\dots,\lambda_m) = \frac{(n-k)!}{n!}\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le m}.$$

Corollary 57. *Let $\mu \in \mathcal{P}(\mathbb{R})$ have finite moments up to order $2n-2$ and let $\varphi_0,\dots,\varphi_{n-1}$ be the first n orthogonal polynomials normalized so that $\int \varphi_k\varphi_\ell d\mu = \delta_{k,\ell}$. Then, the density $f(x) = Z_n^{-1}|\Delta(x)|^2$ on $\mathbb{R}^n$ with respect to a measure $\mu^{\otimes n}$ can be rewritten as $f(x) = (n!)^{-1}\det\left(K_n(\lambda_i,\lambda_j)\right)_{i,j\le n}$ where $K_n(x,y) = \sum_{j=0}^{n-1}\varphi_j(x)\varphi_j(y)$. Further, the marginal density of any k co-ordinates is given by $\frac{(n-k)!}{n!}\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le k}$.*

[6] These are special cases of what are known as *determinantal point processes*.

The corollary trivially follows from the lemma and the observations made before the theorem. We now prove the lemma.

PROOF. (1) We need two properties of the kernel K. Both follow from orthonormality of φ_ks.

(a) The *reproducing kernel property:* $\int K(x,y)K(y,z)\mu(dy) = K(x,z)$.

(b) $\int K(x,x)\mu(dx) = n$.

By expanding the determinant

$$\int_{\mathbb{R}} \det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le m} d\lambda_m = \sum_{\pi\in\mathcal{S}_m} \operatorname{sgn}(\pi) \int_{\mathbb{R}} \prod_{i=1}^{m} K(\lambda_i,\lambda_{\pi(i)})d\lambda_m.$$

Fix π. There are two cases.

Case 1: $\pi(m) = m$. then by property (b), the term becomes

$$\prod_{i=1}^{m-1} K(\lambda_i,\lambda_{\pi(i)}) \int K(\lambda_m,\lambda_m)d\lambda_m = n\prod_{i=1}^{m-1} K(\lambda_i,\lambda_{\sigma(i)}).$$

where $\sigma\in\mathcal{S}_{m-1}$ is defined by $\sigma(i) = \pi(i)$. Observe that $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\pi)$.

Case 2: Fix $\pi(m)\neq m$. Let $p = \pi^{-1}(m)$ and $q = \pi(m)$ (thus $p,q<m$). By property (a) above,

$$\begin{aligned}\int_{\mathbb{R}} \prod_{i=1}^{m} K(\lambda_i,\lambda_{\pi(i)})d\lambda_m &= \prod_{i\neq p,m} K(\lambda_i,\lambda_{\pi(i)}) \int_{\mathbb{R}} K(\lambda_p,\lambda_m)K(\lambda_m,\lambda_q)d\lambda_m \\ &= \prod_{i\neq m} K(\lambda_i,\lambda_{\sigma(i)})\end{aligned}$$

where $\sigma(i) = \pi(i)$ for $i\neq p$ and $\sigma(p) = q$. Then $\sigma\in\mathcal{S}_{m-1}$ and $\operatorname{sgn}(\sigma) = -\operatorname{sgn}(\pi)$.

Now consider any $\sigma\in\mathcal{S}_{m-1}$. It arises from one permutation π in Case 1, and from $m-1$ distinct π in Case 2. As the $\operatorname{sgn}(\sigma)$ has opposing signs in the two cases, putting them together, we see that $\int_{\mathbb{R}} \det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le n} d\lambda_m$ is equal to

$$(n-(m-1))\sum_{\sigma\in\mathcal{S}_{n-1}} \prod_{i=1}^{n-1} K(\lambda_i,\lambda_{\sigma(i)}) = (n-m+1)\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le n-1}.$$

(2) Let $m<n$ and let $f_m(x_1,\dots,x_m) = \int_{R^{n-m}} f(x_1,\dots,x_n)d\mu(x_{m+1})\dots d\mu(x_n)$. Inductively applying the integration formula in part (2), we get

$$f_m(\lambda_1,\dots,\lambda_m) = C_n^{-1}(n-m)!\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le m}.$$

In particular, if we integrate out all variables, we get $C_n^{-1}n!$. Thus, we must have $C_n = n!$ for f to be a probability density (the positivity of f is clear because $\left(K(x_i,x_j)\right)_{i,j\le n}$ is n.n.d, being of the form AA^t).

Plugging the value of C_n back into the expression for f_m shows that

$$f_m(\lambda_1,\dots,\lambda_m) = \frac{(n-m)!}{n!}\det\left(K(\lambda_i,\lambda_j)\right)_{i,j\le k}. \blacksquare$$

These integration formulas are what make $\beta = 2$ special. None of this would work if we considered density proportional to $|\Delta(x)|^\beta$ with respect to $\mu^{\otimes n}$. As a corollary of these integration formulas, we can calculate the mean and variance of the number of points that fall in a given subset.

Proposition 58. *In the setting of Lemma 56, let $N(\cdot) = \sum_{k=1}^n \delta_{\lambda_k}$ be the unnormalized empirical measure. Let $I\subseteq A$ be a measurable subset. Then,* *(i)* $\mathbf{E}\left[(N(I))_{m\downarrow}\right] = \int_{I^m} \det\left(K(x_i,x_j)_{i,j\le k} d\mu(x)\right.$ *where* $(k)_{m\downarrow} = k(k-1)\dots(k-m+1)$.

(ii) $\mathbf{E}[N(I)] = \int_I K(x,x)d\mu(x)$ *and* $Var(N(I)) = \int\int_{I^c} |K(x,y)|^{2d\mu(y)d\mu(x)}$.

(iii) Let T_I *be the integral operator on* $L^2(I,\mu)$ *with the kernel* K. *That is* $T_I f(x) = \int_I K(x,y)f(y)d\mu(y)$ *for* $x \in I$. *Let* $\theta_1, \theta_2, \dots$ *be the non-zero eigenvalues of* T. *Then* $\theta_i \in (0,1]$ *and if* $\xi_i \sim Ber(\theta_i)$ *are independent, then* $N(I) \stackrel{d}{=} \xi_1 + \xi_2 + \dots$.

PROOF. (i) Write $N(I) = \sum_{k=1}^n \mathbf{1}_{\lambda_k \in I}$. Use the exchangeability of λ_k to write

$$\mathbf{E}\left[(N(I))_{m\downarrow}\right] = \mathbf{E}\left[\sum_{\substack{i_1,\dots,i_m \le n \\ \text{distinct}}} \mathbf{1}_{\lambda_{i_1}\in I}\mathbf{1}_{\lambda_{i_2}\in I}\dots\mathbf{1}_{\lambda_{i_m}\in I}\right] = (n)_{\downarrow m}\mathbf{P}\left[\lambda_i \in I,\ 1 \le i \le m\right].$$

Using the density of $(\lambda_1,\dots,\lambda_m)$ given in Lemma 56 we get

$$\mathbf{E}\left[(N(I))_{m\downarrow}\right] = \int_{I^m} \det\left(K(x_i,x_j)\right)_{i,j\le m} d\mu(x).$$

(ii) Apply the formula in part (i) with $m=1$ to get $\mathbf{E}[N(I)] = \int_I K(x,x)d\mu(x)$. Expressing the variance of $N(I)$ in terms of $\mathbf{E}[N(I)]$ and $\mathbf{E}[N(I)(N(I)-1)]$ one arrives at

$$\mathrm{Var}(N(I)) = \int_I K(x,x)d\mu(x) - \int_I\int_I |K(x,y)|^2 d\mu(x)d\mu(y).$$

Write the first integral as $\int_I\int_A |K(x,y)|^2 d\mu(y)$ by the reproducing property of K. Subtracting the second term give $\int_I\int_{I^c} |K(x,y)|^2 d\mu(x)d\mu(y)$.

(iii) With $I = A$, we have $T_A f = \sum_{k=1}^n \langle f, \varphi_k\rangle \varphi_k$. Thus, T is a projection operator with rank n. Clearly, $0 \le T_I \le T_A$ from which it follows that $\theta_i \in [0,1]$ and at most n of them are nonzero. If ψ_i are the corresponding eigenfunctions, then it is easy to see that $K(x,y) = \sum_i \theta_i \psi_i(x)\overline{\psi}_i(y)$.

■

Remark 59. In random matrix theory, one often encounters the following situation. Let $\mu \in \mathcal{P}(\mathbb{C})$ such that $\int |z|^{2n-2}\mu(dz) < \infty$. On $\mathbb{C}^n$ define the density $f(x) \propto |\Delta(x)|^2$ with respect $\mu^{\otimes n}$. Then we can again orthogonalize $1, z, \dots, z^{n-1}$ with respect to μ to get φ_k, $0 \le k \le n-1$ and the kernel $K(z,w) = \sum_{j=0}^{n-1} \varphi_j(z)\overline{\varphi}_j(w)$. The density can be rewritten as $f(x) = (n!)^{-1}\det\left(K(x_i,x_j)\right)_{i,j\le n}$. This is of course a special case of the more general situation outlined in Lemma 56, except that one needs to keep track of conjugates everywhere when taking inner products.

8. Determinantal point processes

Consider the density f_m of $(\lambda_1,\dots,\lambda_m)$ as described in Lemma 56. Let us informally refer to it as the chance that λ_i falls at location x_i for $1 \le i \le m$. Then the chance that $L_n := \sum_{k=1}^n \delta_{\lambda_k}$ puts a point at each x_i, $i \le m$, is precisely $(n)_{m\downarrow} f_m(x_1,\dots,x_m) = \det\left(K(x_i,x_j)\right)$.

For any random variable L taking values in the space of locally finite counting measures (eg., L_n), one can consider this chance (informally speaking), called the m^{th} joint intensity of L. If for every m, the joint intensities are given by $\det\left(K(x_i,x_j)\right)$ for some $K(x,y)$, then we say that L is a *determinantal point process*. A determinantal point process may have infinitely many points.

If a point process which has a fixed finite total number of points, then we can randomly arrange it as a vector and talk in terms of densities. But when we have infinitely many points, we cannot do this and instead talk in terms of joint intensities. Like densities, joint intensities may or may not exist. But if they do exist, they are very convenient to work with. In random matrix theory we usually get finite determinantal processes, but in the limit we often end up with infinite ones. Therefore, we shall now give precise definitions of point processes, joint intensities and determinantal processes[7]

Definition 60. Let A be a locally compact Polish space (i.e., a complete separable metric space) and let μ be a Radon measure on A. A *point process* L on A is a random integer-valued positive Radon measure on A. If L almost surely assigns at most measure 1 to singletons, we call it a *simple* point process;

Definition 61. If L is a simple point process, its *joint intensities* w.r.t. μ are functions (if any exist) $p_k : A^k \to [0,\infty)$ for $k \geq 1$, such that for any family of mutually disjoint subsets $I_1,\dots,I_k$ of A,

$$\mathbf{E}\left[\prod_{j=1}^{k} L(I_j)\right] = \int_{I_1\times\ldots\times I_k} p_k(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k). \tag{31}$$

In addition, we shall require that $p_k(x_1,\dots,x_k)$ vanish if $x_i = x_j$ for some $i \neq j$.

Definition 62. A point process L on A is said to be a *determinantal process* with kernel K if it is simple and its joint intensities with respect to the measure μ satisfy

$$p_k(x_1,\dots,x_k) = \det\left(K(x_i,x_j)\right)_{1\leq i,j\leq k}, \tag{32}$$

for every $k \geq 1$ and $x_1,\dots,x_k \in A$.

Exercise 63. When λ has density as in Lemma 56, check that the point process $L = \sum_{k=1}^{n} \delta_{\lambda_k}$ is a determinantal point process with kernel K as per the above definition.

9. One dimensional ensembles

Let $V : \mathbb{R} \to \mathbb{R}$ be a function that increases fast enough at infinity so that $\int e^{-\beta V(x)}dx < \infty$ for all $\beta > 0$. Then, define the probability measure $\mu_n(dx) = e^{-nV(x)}/Z_n$ and the let λ be distributed according to the measure $Z_{n,\beta}^{-1}|\Delta(x)|^{\beta}e^{-n\sum_{k=1}^{n} V(x_k)}$. Under some conditions on V, the empirical measure of λ converges to a fixed measure $\mu_{V,\beta}$. Then one asks about

We will now concentrate on two particular examples of $\beta = 2$ ensembles.

(1) The GUE (scaled by $\sqrt{2}$). The density is $Z_n^{-1}|\Delta(\lambda)|^2\exp\{-\sum_{k=1}^{n}\lambda_k^2/4\}$. To write it in determinant form, we define μ as the $N(0,2)$ distribution, that is $\mu(dx) = (2\sqrt{\pi})^{-1}e^{-x^2/4}dx$. Let H_k, $k \geq 0$, be the orthogonal polynomials with respect to μ obtained by applying Gram-Schmidt to the monomials $1,x,x^2,\dots$. H_k are called *Hermite polynomials*. The kernel is $K_n(x,y) = \sum_{k=0}^{n-1} H_k(x)H_k(y)$. We have chosen them to be orthonormal, $\int H_k(x)H_\ell(x)d\mu(x) = \delta_{k,\ell}$. Hermite polynomials are among the most important special functions in mathematics.

(2) The CUE (*circular unitary ensemble*). Let μ be the uniform measure on S^1 and on $(S^1)^n$ define the density $f(x) = |\Delta(x)|^2$ with respect to $\mu^{\otimes n}$. In this case $z^k = e^{ikt}$ are themselves orthonormal, but it is a bit more convenient to take $\varphi_k(t) = e^{-i(n-1)t/2}e^{ikt}$. Then, the kernel is

$$e^{-i(n-1)s/2}e^{i(n-1)t/2}\frac{1-e^{in(s-t)}}{1-e^{i(s-t)}} = D_n(s-t), \qquad D_n(u) := \frac{\sin(nu/2)}{\sin(u/2)}.$$

[7] For more detailed discussion on joint intensities, consult chapter 1 of the book **?** available at http://math.iisc.ernet.in/ manju/GAF_book.pdf. Chapter 4 of the same book has discussion and examples of determinantal point processes.

D_n is the well-known Dirichlet kernel (caution: what is usually called D_n is our D_{2n+1}). We shall later see that the eigenvalues of a random unitary matrix sampled from the Haar measure are distributed as CUE.

The GUE and CUE are similar in the local structure of eigenvalues. However, there are edge phenomena in GUE but none in CUE. However, all calculations are simpler for the CUE as the kernel is even simpler than the GUE kernel. The difficulty is just that we are less familiar with Hermite polynomials than with monomials. Once we collect the facts about Hermite functions the difficulties mostly disappear. The study of the edge is rather difficult, nevertheless.

10. Mean and Variance of linear statistics in CUE

Let λ be distributed according to CUE$_n$. Let $h : S^1 \to \mathbb{R}$ be a bounded measurable function. Let $N_n(h)$ be the linear statistic $\sum_{k=1}^n h(\lambda_k)$. Then,

$$\mathbf{E}[N_h] = \int h(x)K(x,x)d\mu(x) = n\int_0^{2\pi} h(t)\frac{dt}{2\pi}.$$

Actually this holds for any rotation invariant set of points on the circle. In particular, $\mathbf{E}[N_n(I)] = |I|/2\pi$.

The variance is considerably more interesting. Write the Fourier series of $h(t) = \sum_{k\in\mathbb{Z}} a_k e^{ikt}$ where $a_k = \int_0^{2\pi} h(t)e^{-ikt}\frac{dt}{2\pi}$. This equality is in L^2. Then,

$$\mathrm{Var}(N_n(h)) = \frac{1}{2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(h(t)-h(s))^2 K_n(s,t)K_n(t,s)\frac{dt\,ds}{4\pi^2}.$$

We write

$$(h(t)-h(s))^2 = \sum_{k,\ell\in\mathbb{Z}} a_k\bar{a}_\ell(e^{ikt}-e^{iks})(e^{-i\ell t}-e^{-i\ell s}), \qquad K_n(t,s)K_n(s,t) = \sum_{p,q=0}^{n-1} e^{i(p-q)t}e^{i(q-p)s}.$$

Hence,

$$\begin{aligned}
\mathrm{Var}(N_n(h)) &= \frac{1}{2}\sum_{k,\ell\in\mathbb{Z}}\sum_{p,q=0}^{n-1} a_k\bar{a}_\ell \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(e^{ikt-i\ell t}+e^{iks-i\ell s}-e^{iks-i\ell t}-e^{ikt-i\ell s})e^{i(p-q)t}e^{i(q-p)s}\frac{dt\,ds}{4\pi^2}\\
&= \frac{1}{2}\sum_{k,\ell\in\mathbb{Z}}\sum_{p,q=0}^{n-1} a_k\bar{a}_\ell\left\{\delta_{k-\ell+p-q}\delta_{q-p}+\delta_{p-q}\delta_{k-\ell+q-p}-\delta_{k+p-q}\delta_{-\ell+q-p}-\delta_{-\ell+p-q}\delta_{k+q-p}\right\}\\
&= \frac{1}{2}\sum_{k,\ell\in\mathbb{Z}}\sum_{p,q=0}^{n-1} a_k\bar{a}_\ell\delta_{k-\ell}\left\{\delta_{p-q}+\delta_{p-q}-\delta_{k+p-q}-\delta_{k+q-p}\right\}\\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}}|a_k|^2\sum_{p,q=0}^{n-1}\left\{2\delta_{p-q}-\delta_{k+p-q}-\delta_{k+q-p}\right\}\\
&= \sum_{k\in\mathbb{Z}}|a_k|^2\left(n-(n-|k|)_+\right).
\end{aligned}$$

Remark 64. The variance can be written as $\sum_{|k|\le n}|k||\hat{h}(k)|^2 + n\sum_{|k|>n}|a_k|^2$ where $\hat{h}(k) = a_k$. The first sum is the contribution of low frequencies in h while the second gives the contribution of the high frequencies. For smooth functions, the high frequency Fourier coefficients will be small and the first term dominates. For more wild functions, the second sum becomes significant. We shall consider two cases next.

Case 1: $h \in H^{1/2}$ which by definition means that $\|h\|^2_{H^{1/2}} := \sum_{k\in\mathbb{Z}} |k||a_k|^2 < \infty$. Observe that if $h \in C^r$, then $h^{(r)}$ has the Fourier series $\sum_{k\in\mathbb{Z}}(-ik)^2 a_k e^{-ikt}$ and hence $\sum |k|^{2r}|a_k|^2 = \|h^{(r)}\|^2_{L^2}$. Thus $H^{1/2}$ can be roughly called those functions hat have half a derivative. Indeed, one can also write the norm in a different way as

Exercise 65. Show that $\|h\|^2_{H^{1/2}} = \int\limits_{-\pi}^{\pi}\int\limits_{-\pi}^{\pi} \frac{(h(t)-h(s))^2}{(t-s)^2} \frac{dt\, ds}{4\pi^2}$.

If $h \in H^{1/2}$, we use the inequality $n - (n-k)_+ \le |k|$ to see that

$$\mathrm{Var}(N_n(h)) \le \frac{1}{2}\sum_{k\in\mathbb{Z}} |a_k|^2|k| = \|h\|^2_{H^{1/2}}.$$

This means that even as the expectation grows linearly in n, the variance stays bounded! Further, for each k fixed, $n-(n-k)_+ \to |k|$ as $n \to \infty$, and hence by DCT $\mathrm{Var}(N_n(h)) \to \|h\|^2_{H^{1/2}}$ as $n \to \infty$.

Case 2: h is the indicator of an arc $I = [a,b]$. Then $N_n(h) = N_n(I)$. We assume that the arc is proper (neither I or I^c is either empty or a singleton). The Fourier coefficients are given by $a_k = \int_a^b e^{-ikt}\frac{dt}{2\pi} = \frac{i(e^{-ikb}-e^{-ika})}{2\pi k}$. Evidently h is not in $H^{1/2}$.

We work out the special case when $I = [-\pi/2, \pi/2]$. Then $a_k = \frac{\sin(\pi k/2)}{\pi k}$ which is zero if k is even and equal to $\frac{(-1)^{j-1}}{\pi(2j+1)}$ if $k = 2j+1$. Thus,

$$\begin{aligned}\mathrm{Var}(N_n) &= 2\sum_{j=0}^{\infty}\frac{n-(n-2j-1)_+}{\pi^2(2j+1)^2} \\ &= 2\sum_{j=0}^{\lfloor\frac{n-1}{2}\rfloor}\frac{2j+1}{\pi^2(2j+1)^2} + 2\sum_{j>\lfloor\frac{n-1}{2}\rfloor}\frac{n}{\pi^2(2j+1)^2}.\end{aligned}$$

As $\sum_{j>m} j^{-2} = O(m^{-1})$, the second term is $O(1)$. The first term is easily seen to be $\frac{2}{\pi^2}\frac{1}{2}\log n + O(1)$. Thus, $\mathrm{Var}(N_n) = \frac{1}{\pi^2}\log n + O(1)$. Although it increases to infinity with n, the variance grows remarkably slower than the mean. Compare with sums of independent random variables where the mean and variance are both of order n. In the next exercise, take $I = [-\alpha, \alpha]$ without losing generality and show that the variance is asymptotically the same.

Exercise 66. Let f be a 2π-periodic function on $\mathbb{R}$ such that $\|f\|^2 := (2\pi)^{-1}\int_{-\pi}^{\pi}|f|^2$ is finite. Let $\hat{f}(k) := \int_{-\pi}^{\pi} f(t)e^{-ikt}\frac{dt}{2\pi}$ denote its Fourier coefficients. Let $f_\tau(t) = f(t-\tau)$ be the translates of f for any $\tau \in \mathbb{R}$.

(1) Use the Plancherel theorem to show that $4\sum_{k\in\mathbb{Z}}|\hat{f}(k)|^2\sin^2(k\tau) = \|f_\tau - f_{-\tau}\|^2$. [**Hint:** $\hat{f_\tau}(k) = e^{-ik\tau}\hat{f}(k)$]

(2) Let $f(t) = t$ on $[-\pi,\pi]$ and extended periodically. Show that $\hat{f}(k) = \frac{(-1)^k}{k}$ and hence conclude that for $\tau \in [0,\pi]$

$$\sum_{k=1}^{\infty}\frac{\sin^2(k\tau)}{k^2} = \tau(\pi-\tau).$$

(3) Fix $\tau \in [0,\pi]$ and let $A_n = \sum_{k=1}^n \frac{\sin^2(k\tau)}{k}$ and $B_n = \sum_{k=1}^n \frac{\cos^2(k\tau)}{k}$. Show that $A_n + B_n = \log n + O(1)$ and $B_n - A_n = O(1)$ as $n \to \infty$. Conclude that both A_n and B_n are equal to $\frac{1}{2}\log n + O(1)$.

(4) Deduce that $\mathrm{Var}(N_n(I)) = \frac{1}{\pi^2}\log n + O(1)$ as $n \to \infty$ for any proper arc I (proper means $0 < |I| < 2\pi$).

Observe that the constant in front of $\log n$ does not depend on the length of the interval. Essentially the entire contribution to the variance comes from a few points falling inside or outside the interval at the two endpoints. Points which "were supposed to fall" deep in the interior of I (or deep in I^c) have almost no chance of falling outside of I (outside of I^c, respectively) and do not contribute to the variance. This shows the remarkable rigidity of the CUE.

Proposition 67. *In the setting of the previous discussion, for any proper arc I, as $n \to \infty$,*

$$\frac{N_n(I) - \frac{|I|}{2\pi}}{\pi^{-1}\sqrt{\log n}} \xrightarrow{d} N(0,1).$$

PROOF. Fix an arc I. By part (c) of Lema 56, $N_n(I)$ is a sum of independent Bernoulli random variables. By the Lindeberg Feller CLT for triangular arrays, any sum of independent Bernoullis converges to N(0,1) after subtracting the mean and dividing by the standard deviation, provided the variance of the random variable goes to infinity. As $\mathrm{Var}(N_n(I)) \sim c\log n$, this applies to our case. ■

Next we compute the covariance between $N(I)$ and $N(J)$. We take I and J to be disjoint. Then, $\mathrm{Cov}(N(I), N(J)) = -\int_I\int_J |K(x,y)|^2 d\mu(x)d\mu(y)$.

11. Fredholm determinants and hole probabilities

Let $(A, \mathcal{A}, \mu)$ be a probability space. Let $K : A^2 \to \mathbb{R}$ or $\mathbb{C}$ be a kernel such that $\|K\| := \sup_{x,y}|K(x,y)| < \infty$. Let T be the integral operator with kernel K.

Definition 68. The *Fredholm determinant* of the operator $I - T$ which we shall also call the Fredholm determinant associated to the kernel K is defined as

$$\Delta(K) := \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int\limits_{A^m} \det\left(K(x_i,x_j)\right)_{i,j\le m} d\mu(x_1)\dots d\mu(x_m).$$

Recall the *Hadamard inequality* for matrices which says that if M is a square matrix with columns u_k, $k \le n$, then $|\det(M)| \le \prod_{j=1}^n \|u_j\|$. Therefore, $|\det\left(K(x_i,x_j)\right)_{i,j\le m}| \le (\|K\|\sqrt{m})^m$ for any m and any $x_1,\dots,x_m$. This shows that $\Delta(K)$ is well-defined for any K with $\|K\| < \infty$.

Remark 69. Let M be an $n \times n$ matrix with eigenvalues θ_j, $j \le n$. Then, we leave it as an exercise to show the identity

$$\sum_{1\le i_1<i_2<\ldots<i\le m} \theta_{i_1}\theta_{i_2}\dots\theta_{i_m} = \sum_{1\le i_1<i_2<\ldots<i\le m} \det\left(M_{i_p,i_q}\right)_{p,q\le m}$$

for any $m \ge 1$. For $m = 1$ this is just the identity $\sum\theta_i = \sum_i M_{i,i}$. For any $m \ge 1$, one can think of the identity as being exactly the same identity, applied to a different matrix. If M acts on a vector space V, then one can define the operator $M^{\wedge k}$ on the alternating tensor power $V^{\wedge k}$ as $\langle M(e_{j_1}\wedge\ldots\wedge e_{j_k}), e_{i_1}\wedge\ldots\wedge e_{i_k}\rangle = \det\left(M_{i_p,j_q}\right)_{p,q\le k}$. This has eigenvalues $\theta_{i_1}\theta_{i_2}\dots\theta_{i_k}$ where $i_1 < i_2 < \ldots < i_k$. Expressing $\mathrm{tr}(M^{\wedge k})$ in two ways gives the above identity.

Anyhow, from this identity, we get the following expression for $\det(I-M)=\prod_{j=1}^n(1-\theta_j)$.

$$\begin{aligned}\det(I-M) &= \prod_{j=1}^{n}(1-\theta_j)\\ &= 1-\sum_i \theta_i+\sum_{i<j}\theta_i\theta_j-\sum_{i<j<k}\theta_i\theta_j\theta_k+\dots\\ &= 1-\sum_i M_{i,i}+\frac{1}{2}\sum_{i,j}\det\begin{bmatrix} M_{i,i} & M_{i,j}\\ M_{j,i} & M_{j,j}\end{bmatrix}-\frac{1}{6}\sum_{i,j,k}\det\begin{bmatrix} M_{i,i} & M_{i,j} & M_{i,k}\\ M_{j,i} & M_{j,j} & M_{j,k}\\ M_{k,i} & M_{k,j} & M_{k,k}\end{bmatrix}+\dots\end{aligned}$$

Thus, $\det(I-M)$ is exactly what we defined as $\Delta(K)$, provided we take $A=[n]$ and $K(i,j)=M_{i,j}$. With the usual philosophy of regarding an integral kernel as a matrix $(K(x,y))_{x,y}$, we arrive at the definition of the Fredholm determinant. The following exercise is instructive in this respect.

Exercise 70. Let T be the integral operator with a Hermitian kernel K with $\|K\|<\infty$. Let θ_j be the eigenvalues of T. Then, for any $m\ge 1$, we have

$$\sum_{i_1<i_2<\ldots<i_m}\theta_{i_1}\theta_{i_2}\dots\theta_{i_m}=\frac{1}{m!}\int_{A^m}\det\left(K(x_i,x_j)\right)_{i,j\le m}d\mu(x_1)\dots d\mu(x_m).$$

We shall need the following simple lemma later[8].

Lemma 71. *Let K and L be two kernels on $L^2(A,\mathcal{A},\mu)$ such that $C=\max\{\|K\|,\|L\|\}<\infty$. Then,*

$$|\Delta(K)-\Delta(L)|\le\|K-L\|\left(\sum_{m=0}^{\infty}\frac{m(C\sqrt{m})^{m-1}}{m!}\right).$$

PROOF. Fix $m\ge 1$ and $x_1,\dots,x_m\in A$. Let $X_0=(K(x_i,x_j))_{i,j\le m}$ and $X_m=(L(x_i,x_j))_{i,j\le m}$. For $1\le k\le m$, let X_k be the matrix whose first k rows are those of X_0 and the rest are those of X_m. Then, $\det(X_0)-\det(X_m)=\sum_{k=1}^{m-1}\det(X_{k-1})-\det(X_k)$. Using Hadamard's inequality we see that $|\det(X_{k-1})-\det(X_k)|$ is bounded by $(C\sqrt{m})^{m-1}\|K-L\|$. Thus

$$|\det\left(K(x_i,x_j)\right)_{i,j\le m}-\det\left(L(x_i,x_j)\right)_{i,j\le m}|\le m(C\sqrt{m})^{m-1}\|K-L\|.$$

Integrate over x_js and then sum over m (after multiplying by $(-1)^{m-1}/m!$ to get the claimed result. ■

The importance of Fredholm determinants for us comes from the following expression for "hole probabilities" or "gap probabilities" in determinantal processes.

Proposition 72. *Let $(A,\mathcal{A},\mu)$ be a probability space and let K be a finite rank projection kernel (that is $K(x,y)=\sum_{j=1}^n\varphi_j(x)\overline{\varphi}_j(y)$ for some orthonormal set $\{\varphi_j\}$). Let λ have density $(n!)^{-1}\det\left(K(x_i,x_j)\right)_{i,j\le n}$. Let $I\subseteq A$ be a measurable subset of A. Then $\mathbf{P}(N(I)=0)=\Delta(K_I)$, where K_I is the kernel K restricted to $I\times I$.*

[8] We have borrowed much of this section from the book of Anderson, Guionnet and Zeitouni **?** where the reader may find more about these objects. F.Riesz and Sz. Nagy's great book on Functional analysis is another good reference for Fredholm's work in functional analysis.

PROOF. From part (c) of Lemma **??**, we know that $\mathbf{P}(N(I)=0)=\prod_j(1-\theta_j)$ where θ_j are the eigenvalues of the integral operator T_I with kernel K_I. Hence,

$$\begin{aligned}\mathbf{P}(N(I)=0) &= 1-\sum_i \theta_i+\sum_{i<j}\theta_i\theta_j-\sum_{i<j<k}\theta_i\theta_j\theta_k+\dots \\ &= \sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\int_{I^m}\det\left(K(x_i,x_j)\right)_{i,j\le m}d\mu(x_1)\dots d\mu(x_m)\end{aligned}$$

by Exercise 70. The last expression is $\Delta(K_I)$ by definition. ∎

12. Gap probability for CUE

Let λ be distributed as CUE$_n$ ensemble and unwrap the circle onto the interval $[-\pi,\pi]$. Thus λ follows the measure on $[-\pi,\pi]^n$ given by

$$\frac{1}{n!}\det\left(K_n(t_i,t_j)\right)_{i,j\le n}\frac{dt_1\dots dt_n}{(2\pi)^n}, \qquad \text{where } K_n(t,s)=\frac{\sin\left(\frac{n}{2}(s-t)\right)}{\sin\left(\frac{s-t}{2}\right)}.$$

Scale up by a factor of $n/2$ to get $\tilde{\lambda}=n\lambda/2$ which follows the measure (on $[-n\pi/2,n\pi/2]^n$)

$$\frac{1}{n!}\det\left(\tilde{K}_n(t_i,t_j)\right)_{i,j\le n}\frac{dt_1\dots dt_n}{(2\pi)^n}, \qquad \text{where } \tilde{K}_n(t,s)=\frac{2}{n}K_n\left(\frac{2t}{n},\frac{2s}{n}\right).$$

Then,

$$\tilde{K}_n(t,s)=\frac{2\sin(s-t)}{n\sin\left(\frac{s-t}{n}\right)}\to K(t,s):=\frac{2\sin(s-t)}{s-t}.$$

It is also easy to see that the convergence is uniform over (t,s) in any compact subset of $\mathbb{R}^2$. Further, $\|\tilde{K}_n\|\le 2$ and $\|K\|\le 2$. Thus, by Lemma 71, we see that $\Delta(\tilde{K}_{n,I})\to\Delta(K_I)$ for any compact interval I. By Proposition prop:holefordeterminantal this shows that for any $a<0<b$,

$$\mathbf{P}\left(\lambda_i\notin\left[\frac{2a}{n},\frac{2b}{n}\right]\ \forall i\le n\right)=\mathbf{P}\left(\tilde{\lambda}_i\notin[-a,b]\ \forall i\le n\right)\to\Delta(K_{[a,b]})$$

as $n\to\infty$. This gives the asymptotics of gap probabilities in CUE. Some remarks are due.

Of course, it is incorrect to say that we have calculated the gap probability unless we can produce a number or decent bounds for this probability. For example, we could define $F(t):=\Delta(K_{[-t,t]})$ which is the asymptotic probability that the nearest eigenvalue to 0 in CUE$_n$ is at least $2t/n$ away. Can we find $F(t)$? All we need to so is study the kernel K (called the *sine kernel*) and deduce $F(t)$ from it. This is not trivial, but has been done by ???? They show that $F(t)$ can be characterized in terms of the solution to a certain second order ODE, called the ?????? We do not prove this result in this course.

Secondly, we considered only the gap probability, but we could also consider the distributional limit of the whole point process $\tilde{L}_n:=\sum_k\delta_{\tilde{\lambda}_k}$. But then we must employ the language of Section **??**. In that language, it is not difficult to show that the convergence of $\tilde{K}_n$ to K implies that $\tilde{L}_n$ converges in distribution to L, the determinantal point process with kernel K. The latter is a stationary point process on the line (and hence has infinitely many points, almost surely). Basically this distributional convergence is the statement that all the joint intensities $\det\left(K_n(x_i,x_j)\right)_{i,j\le m}$ converge to the corresponding quantities $\det\left(K(x_i,x_j)\right)_{i,j\le m}$. However, note that the distributional convergence does not automatically imply convergence of the gap probability, because the latter is expressed as a series involving joint intensities of all orders. That is why we had to establish Lemma 71 first.

13. Hermite polynomials

Our next goal is to prove results for GUE analogous to those that we found for CUE. Additionally, we would also like to study the edge behaviour in GUE, for which there is no analogue in CUE. In this section we shall establish various results on Hermite polynomials that will be needed in carrying out this programme.

For $n \ge 0$, define $\tilde{H}_n(x) := (-1)^n e^{-x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$. It is easily seen that $\tilde{H}_n$ is a monic polynomial of degree n. It is also easy to see that the coefficients of x^{n-1}, x^{n-3} etc. are zero. Consider

$$\begin{aligned}\int H_n(x)H_m(x)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} &= (-1)^n \int H_n(x)\frac{d^n}{dx^n}e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} \\ &= \int e^{-x^2/2}\frac{d^n}{dx^n}H_n(x)\frac{dx}{\sqrt{2\pi}} \\ &= \begin{cases} 0 & \text{if } n<m \text{ because } H_n \text{ has degree only } n. \\ n! & \text{if } m=n. \end{cases}\end{aligned}$$

Thus $H_n(x) := \frac{1}{\sqrt{n!}}\tilde{H}_n(x)$ define an orthonormal sequence of polynomials with respect to N(0,1) measure called *Hermite polynomials*. Let $\psi_n(x) = (2\pi)^{-1/4}e^{-x^2/4}H_n(x)$ be the *Hermite functions*. Then $\{\psi_n : n \ge 0\}$ for an ONB for $L^2(\mathbb{R}, \text{Lebesgue})$. The following properties may be derived easily (or look up any book on special functions, for example, Andrews, Askey and Roy **?**).

Exercise 73. (1) $\left(-\frac{d}{dx}+x\right)\tilde{H}_n(x) = \tilde{H}_{n+1}(x)$ and hence also $\left(-\frac{d}{dx}+x\right)H_n(x) = \sqrt{n+1}H_{n+1}(x)$.

(2) *Hermite functions are eigenfunctions of the Hermite operator:* $\left(-\frac{d}{dx}+\frac{x}{2}\right)\psi_n(x) = \sqrt{n+1}\psi_{n+1}(x)$ and $\left(\frac{d}{dx}+\frac{x}{2}\right)\psi_n(x) = \sqrt{n}\psi_{n-1}(x)$. Consequently,

$$\left(-\frac{d^2}{dx^2}+\frac{x^2}{4}\right)\psi_n(x) = \left(n+\frac{1}{2}\right)\psi_n(x). \tag{33}$$

(3) *Three term recurrence:* $x\tilde{H}_n(x) = n\tilde{H}_{n-1}(x) + \tilde{H}_{n+1}(x)$. Consequently, $xH_n(x) = \sqrt{n}H_{n-1}(x) + \sqrt{n+1}H_{n+1}(x)$.

We now derive two integral representations for Hermite polynomials. Observe that $\frac{d^n}{dx^n}e^{-(x-w)^2/2}\big|_{w=0} = (-1)^n\frac{d^n}{dx^n}e^{-x^2/2}$. Therefore, fixing x, we get the power series expansion $e^{-(x-w)^2} = \sum_{n=0}^{\infty} H_n(x)w^n/n!$ which simplifies to $e^{xw-\frac{w^2}{2}} = \sum_{n=0}^{\infty} H_n(x)w^n/n!$. Thus,

$$H_n(x) = \frac{1}{2\pi i}\int_{\gamma}\frac{e^{xw-\frac{w^2}{2}}}{w^{n+1}}dw, \qquad \text{for any closed curve } \gamma \text{ with } \operatorname{Ind}_{\gamma}(0)=1. \tag{34}$$

A second integral representation will be obtained from the well-known identity

$$e^{-x^2/2} = \int_{\mathbb{R}} e^{-itx}e^{-t^2/2}\frac{dt}{\sqrt{2\pi}} = \int_{\mathbb{R}}\cos(tx)e^{-t^2/2}\frac{dt}{\sqrt{2\pi}}.$$

Differentiate n times with respect to x to get

$$\tilde{H}_n(x) = \begin{cases} (-1)^m \int_{\mathbb{R}}\cos(tx)x^n e^{-t^2/2}\frac{dt}{\sqrt{2\pi}} & \text{if } n=2m. \\ (-1)^{m-1}\int_{\mathbb{R}}\sin(tx)x^n e^{-t^2/2}\frac{dt}{\sqrt{2\pi}} & \text{if } n=2m-1. \end{cases} \tag{35}$$

We end the section with the Christoffel-Darboux formula.

Lemma 74. *Let μ be a probability measure on $\mathbb{R}$ with infinite support. Let p_k be the orthogonal polynomials with respect to μ normalized so that $p_n(x) = \kappa_n x^n + \dots$ with $\kappa_n > 0$. Then,*

$$\sum_{k=0}^{n-1} p_k(x)p_k(y) = \frac{\kappa_{n-1}}{\kappa_n} \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y}.$$

For $x = y$ the right hand should be interpreted as $\frac{\kappa_{n-1}}{\kappa_n}(p_n(x)p'_{n-1}(x) - p'_n(x)p_{n-1}(x))$.

PROOF. Write the three term recurrence

$$xp_n(x) = b_{n-1}p_{n-1}(x) + a_n p_n(x) + b_n p_{n+1}(x).$$

Multiply by $p_n(y)$ to get the equation

$$xp_n(x)p_n(y) = b_{n-1}p_{n-1}(x)p_n(y) + a_n p_n(x)p_n(y) + b_n p_{n+1}(x)p_n(y).$$

Write this same equation with x and y reversed and subtract from the above equation to get

$$(x-y)p_n(x)p_n(y) = -b_{n-1}(p_{n-1}(y)p_n(x) - p_{n-1}(x)p_n(y)) + b_n(p_n(y)p_{n+1}(x) - p_n(x)p_{n+1}(y)).$$

Put k in place of n and sum over $0 \le k \le n-1$ to get the identity

$$\sum_{k=0}^{n-1} p_k(x)p_k(y) = b_{n-1} \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y}.$$

In the original three term recurrence equate the coefficients of x^{n+1} to see that $b_n\kappa_{n+1} = \kappa_n$. This completes the proof. ■

Corollary 75. *For any $n \ge 1$, we have*

$$\sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n}\frac{\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y)}{x-y}.$$

The corollary follows immediately from the lemma. The importance for us is that it makes it very clear that analysis of the GUE for large n depends on understanding ψ_n (or equivalently, understanding H_n) for large n.