Motivation continues

We have seen OPP's

with random β such as

$\{\mathbb{R}, \mathbb{H}, \mathbb{C}\}$, or $\{\mathbb{R}$ or $\mathbb{H}\}$

but we will see eigenvalues at ($\beta = 2$)

classical ensembles are analogs

e.g. $\mathrm{Eig}(\underset{\text{GUE}}{N\times N}) =$ N HS Projection OPP

K is continuous kernel

indexed by $\mathbb{R}$

$\mathrm{Eig}(N\times N \text{ GUE}) \cap [a,b] =$ the N HS General OPP

K continuous indexed by $[a,b]$

Probabilities become prob densities

$$K\begin{pmatrix} i_1 & \cdots & i_N \\ j_1 & \cdots & j_N \end{pmatrix} = \det\left[K(i_j, i_k)\right]_{1\le j,k\le N}$$

$$\Rightarrow \det\left(K(x_j, x_k)\right)_{1\le j,k\le N} \quad N\times N \text{ det}$$

distribution of spacings or largest observable

reduces determinants

whose answers give fluctuations

in aztec diamonds, growth etc

How DPPs go continuous, RMT, & fluctuations in growth

Two questions:

Q1:
First $\mathcal{J}$ random from DPP
Then i uniformly random from $\mathcal{J}$ $\quad \Pr(i)$?

Q2: $\mathcal{J}$ random from DPP
$\Pr(i \in \mathcal{J})$

Not the same:

Q1: Example $n = 3$

$\{1\}, \ \frac{1}{2}\{1,2\}, \ \frac{1}{2}\{1,3\}, \ \frac{1}{3}\{1,2,3\}$

$$\Pr(i) = \sum_{i \in \mathcal{J}} (\Pr \mathcal{J}) \frac{1}{|\mathcal{J}|}$$

Q2: $\Pr(i \in \mathcal{J}) = \sum_{i \in \mathcal{J}} \Pr(\mathcal{J}) = K_{ii}$

Special Case: K = proj rank r $\Rightarrow$ $|\mathcal{J}| = r$

Q1: $\frac{1}{r} \sum_{\substack{i \in \mathcal{J} \\ |\mathcal{J}| = r}} \Pr \mathcal{J} = K_{ii}/r$

Q2: $\sum_{\substack{i \in \mathcal{J} \\ |\mathcal{J}| = r}} \Pr \mathcal{J} = K_{ii}$

(2)

Using Test Functions

"f" test function defined on $1:N$

$F(J) = \sum_{j \in J} f(j)$ e.g. $F(\{1,2,3\}) = f(1)+f(2)+f(3)$

$F(\{3\}) = f(3)$

$F(\emptyset) = 0$

$$E\left(\sum_{j \in J} f(j)\right) = \sum_{j=1}^{N} K_{j|N} f(j)$$

coefficients kind of like —

J	$f(J)$
$\emptyset$	0
$\{1\}$	$f(1)$
$\{2\}$	$f(2)$
$\{1,2\}$	$f(1)+f(2)$

$E(\sum f(i)) = f(1)\,Pr(1 \in J) + f(2)\,Pr(2 \in J)$

Multiple Subsets of size K

$f(J)$ defined for the $\binom{N}{K}$ subsets of size K

~~$F(J) = \sum$~~

$F(J) =$

$$F(J) = \sum_{\substack{S \subseteq J \\ |S| = K}} f(S)$$

e.g. $N = 4$, $K = 2$

$$F(\{1,2,3\}) = f(\{1,2\}) + f(\{1,3\}) + f(\{2,3\})$$

$$E(F(J)) = E\left(\sum_{\substack{S \subseteq J \\ |S| = K}} f(S)\right) = \sum_{\substack{|S| = K \\ S \subseteq N}} k\left(\frac{J}{S}\right) f(S)$$

~~Modified F_m~~

Modification

$f(S_1, \ldots, S_K)$ defined symm when K is random

$$F(J) = \sum_{\substack{J^K \\ \text{distinct}}} F(S_1, \ldots, S_K)$$

$$E\left(F_{mult}(J)\right) = N(N-1)\cdots(N-K-1)\, f(J)$$

Factorial Moment

④ Determinantal Point Processes

Two basic ingredients

1. Reference measure μ on X.
2. Correlation kernel $K : X \times X \to \mathbb{R}$.

ρ is determinantal with kernel K w.r.t. μ if for every nice $f : X^k \to \mathbb{R}$

$$\text{(*)} \quad \mathbb{E}\left[\sum_{\substack{(p_1,\dots,p_k) \in \rho^k \\ \text{all } p_i \text{ distinct}}} f(p_1,\dots,p_k) \right] = \int_{X^k} \det\left[K(x_i,x_j)\right] f(x_1,\dots,x_k) \, d\mu(x_1)\dots d\mu(x_k)$$

(*) implies the following pt-wise defn of K:

$$\text{(**)} \quad \lim_{\delta \to 0} \frac{\Pr\left(\text{there is one point of } \rho \text{ in } B_\delta(x_i),\ i=1,\dots,k\right)}{\prod_{i=1}^k \mu\left(B_\delta(x_i)\right)}$$

$$= \det\left[K(x_i,x_j)\right] \quad \text{for } x_1 \neq x_2 \neq \cdots \neq x_k.$$

The LHS of (**), when they exist, are called correlation functions of $\mathcal{P}$:

$$\rho_1(x),\ \rho_2(x_1, x_2),\ \rho_3(x_1, x_2, x_3),\ \ldots$$

So the correlation functions of a DPP are given by determinants of the correlation kernel.

They determine the law of $\mathcal{P}$.

The correlation functions are not probability densities. For instance,

$$\int_X d\mu(x)\, \rho_1(x) = \mathbb{E}\left[\#(\mathcal{P} \cap X)\right].$$

In general, $\int_{C^k} d\mu(x_1) \cdots d\mu(x_k)\, \rho_k(x_1, \ldots, x_k) =$

$$\mathbb{E}\left[N \cdot (N-1) \cdot (N-2) \cdots (N-k+1)\right],$$

where $N = \#(\mathcal{P} \cap C)$.

It is clear that higher correlation functions determine the lower ones.

The exact relation is that for $k \leq n$,

$$\rho_k(x_1, \dots, x_k) = \frac{1}{(n-k)!} \int_{X^{n-k}} d\mu(x_{k+1}) \cdots d\mu(x_n) \; \rho_n(x_1, \dots, x_k, x_{k+1}, \dots, x_n).$$

At the level of counting this means that if you are given the density of all n-tuples of distinct points and need to determine the density of k-tuples, you should

① Integrate out the density of the last $(n-k)$ of them.

② Divide by $(n-k)!$ to account for overcounting due to ordering.

4 The Airy process

Let us rotate counter-clockwise by 45° the set $\Lambda(t)$ from the corner growth model with exponentially distributed weights. The boundary of $\Lambda(t)$ may now be regarded as the graph of a random function $x \mapsto X_t(x)$ as in Figure 2. So for instance, $X_t(0)$ is X_t from the previous section. If $|x| > t$ then $X_t(x) = |x|$ but otherwise $X_t(x)$ is a random curve.

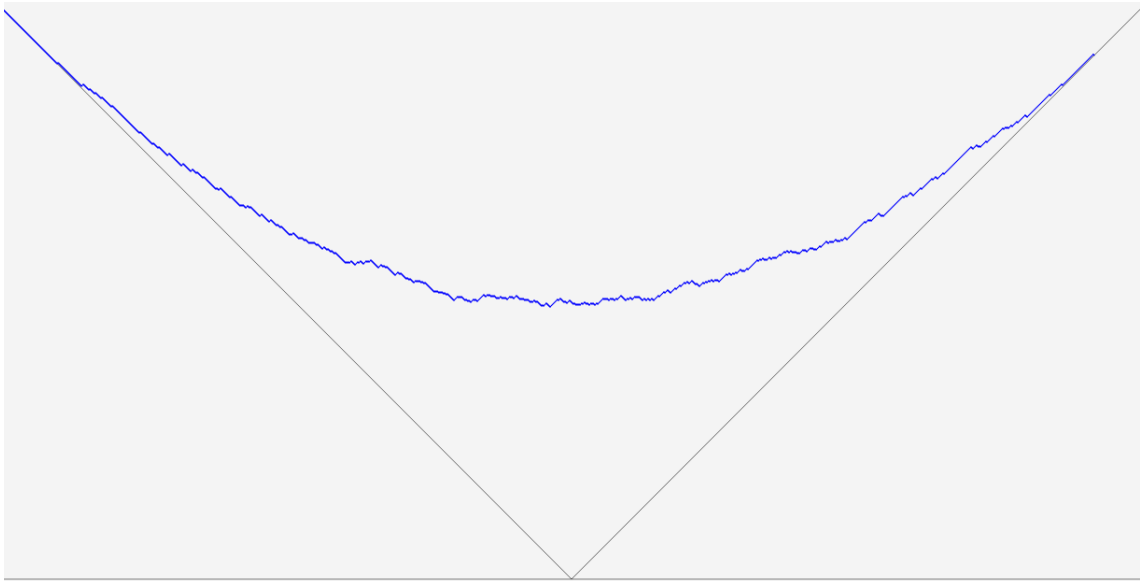

Figure 2: The blue curve is the boundary of the corner growth model rotated by 45°. In a scaling limit, it converges to a parabola minus the Airy process.

Observe that $X_t(x)$ – shaded blue in Figure 2 – resembles a parabola. This is manifestation of the fact that the corresponding boundary of the limit shape is a parabola. We would like to know the little fluctuations. Remarkably, it is possible to write a formula for the joint distribution function of $X_t(x_1), \ldots, X_t(x_m)$ for any finite set of locations $x_1, \ldots, x_m$, much like (3.1). One can access the fluctuations from these formulae.

We will not reproduce the formula here, which originally is due to Johansson citexx. We explain what is gives. Consider the joint distribution of $X_t(x_1)$ and $X_t(x_2)$. If the distance $|x_2 - x_1|$ is of order t then $X_t(x_1)$ becomes statistically independent of $X_t(x_2)$ for large times. This is intuitive because $\Lambda(t)$ has a limit shape when its coordinates are scaled by t. Much less intuitive is to figure out the exponent ξ such that $X_t(x_1)$ and $X_t(x_2)$ have non-trivial correlations when $|x_2 - x_1|$ is of order t^ξ.

It turns out that $\xi = 2/3$. So the joint fluctuations of X_t should be observed in an interval of length $t^{2/3}$ around a central point x_0. The scaling limit turns out not to depend on x_0 so long as $|x_0| << t$ (when $|x_0| \approx t$ there is an interesting transition from the flat

part of the blue curve to the jagged part). We consider $x_0 = 0$. Since $X_t(0) \approx t/4$, with fluctuations of order $t^{1/3}$, consider the rescaled function

$$H_t(u) = \frac{X_t(ut^{2/3}) - \frac{t}{4}}{-(t/16)^{1/3}}\,, \quad u \in \mathbb{R}.$$

The limit of $H_t(u)$ provides a random function $\mathcal{A}(u)$ called the *Airy process*.

Theorem 4.1 (Johansson, 2003)**.** *There is a random continuous function $u \mapsto \mathcal{A}(u)$ such that given any $u_1, \ldots, u_m$,*

$$\mathbf{Pr}\left[H_t(u_i) \le a_i,\ i = 1, \ldots, m\right] \overset{t\to\infty}{\longrightarrow} \mathbf{Pr}\left[\mathcal{A}(u_i) - u_i^2 \le a_i,\ i = 1, \ldots, m\right].$$

In other words, $H_t(u)$ converges in distribution to $\mathcal{A}(u) - u^2$ as $t \to \infty$.

The parabola u^2 appears above since the limit shape is a parabola. It appears as $-u^2$ due to the minus sign in $-(t/16)^{1/3}$. There are formulae for the distribution functions of the Airy process; see citeJohansson. The Airy process itself was discovered by Prähofer and Spohn in the context of the polynuclear growth model, a model quite close to how a coffee stain spreads on paper.

The Airy process is believed to govern the fluctuations of many growth models, similar to how Brownian motion governs fluctuations of random walks. Although, the simple random walk S_t on $\mathbb{Z}$ converges to Brownian motion in the scaling

$$u \,\to\, \frac{S_t(ut)}{t^{1/2}}\,,$$

which is markedly different from the scaling for the corner growth model. Unlike simple random walk, where the scaling parameters $t : t : t^{1/2}$ is called diffusive, the scaling parameters $t : t^{2/3} : t^{1/3}$ that appear in the corner growth model is of the Kardar-Parisi-Zhang (KPZ) "universality class". It encompasses many growth models and gives rise to new processes such as the Airy process.

There is a more direct connection between the Airy process and random matrices, which comes from stationary Dyson Brownian motion.

5 The Airy process and Dyson Brownian Motion

The Ornstein-Uhlenbeck process is a stationary, Gaussian Markov process that describes the motion of a Brownian particle in an environment with friction. By formula,

$$(\text{Ornstein-Uhlenbeck})_t = \int_0^t e^{s-t}\, dW_s\,,$$

where W_t is complex-valued Brownian motion. Consider the $N \times N$ Hermitian matrix

$$M_t = \frac{\mathrm{OU}_t + \mathrm{OU}_t^*}{2}$$

such that the entries of $\mathrm{OU_t}$ are independent Ornstein-Uhlenbeck processes. *Stationary Dyson Brownian Motion* is the joint distribution of the eigenvalues of M_t:

$$\lambda(t) = \big(\lambda_1(t) < \cdots < \lambda_N(t) \big) .$$

M_t is stationary in the sense that given positive $t_1, \ldots, t_k$ and h, the joint distribution of $(M_{t_1}, \ldots, M_{t_k})$ is the same as that of $(M_{t_1+h}, \ldots, M_{t_k+h})$. Moreover, the distribution of M_t is of an $N \times N$ GUE matrix. Therefore, $\lambda(t)$ is a stationary ensemble of N non-intersecting paths whose distribution at fixed times is the eigenvalue distribution of an $N \times N$ GUE matrix. The Airy process is the large N scaling limit of the top eigenvalue process $\lambda_N(t)$ in the same sense as Theorem 4.1:

$$u \mapsto \sqrt{2} N^{1/6} \left(\lambda_N(u N^{1/3}) - \sqrt{2} N^{1/2} \right) \overset{N \to \infty}{\longrightarrow} u \mapsto \mathcal{A}(u).$$

Observe how the scaling is of the form $N^{\frac{1}{2}} : (N^{\frac{1}{2}})^{2/3} : (N^{\frac{1}{2}})^{1/3}$. This theorem was proved by Tracy and Widom. We will describe how it comes about.

Suppose at time s we condition the N eigenvalues to be at $\lambda_i(s) = x_i$ for $x_1 < \ldots < x_N$. We want the transition density that at time $t > s$ they end up at $\lambda_i(t) = y_i$ with $y_1 < \ldots < y_N$. This can be done using the Karlon-Mcgregor formula. Let $p_t(x, y)$ be the transition density for the Ornstein-Uhlenbeck process to go from x at time 0 to y at time t. Then, as a transition density,

$$\mathbf{Pr} \left[\lambda(t) = (y_1, \ldots, y_N) \mid \lambda(s) = (x_1, \ldots, x_N) \right] = \frac{1}{Z_N} \det_{i,j} \left[p_{t-s}(x_i, y_j) \right] .$$

Using the above, Tracy and Widom gave a Fredholm determinant expression for the distribution of $\lambda_N(t)$.

Chapter 28

Stochastic Operators

Classically, many important distributions of random matrix theory were accessed through what now seems like an indirect procedure: first formulate an n-by-n random matrix, then compute an eigenvalue distribution, and finally let n approach infinity. The limiting distribution was reasonably called an eigenvalue distribution, but it did not describe the eigenvalue of any specific operator, since the matrices were left behind in the $n \to \infty$ limit.

All of that has changed with the stochastic operator approach to random matrix theory. The new framework is this:

- Select a *stochastic differential operator* such as the stochastic Airy operator

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}} W'(x),$$

  where $W(x)$ is the Wiener process.

- Compute an eigenvalue distribution.

That's it. This approach produces the same eigenvalue statistics that have been studied by the random matrix theory community for decades but in a more direct fashion. The reason: the stochastic differential operators of interest are the $n \to \infty$ continuum limits of the most-studied random matrix models, as we shall see.

The stochastic operator approach was introduced by Edelman Edelman (2003) in 2003 and developed by Edelman and Sutton Sutton (2005), Edelman and Sutton (2007a).

28.0.1 Brownian motion and white noise

We begin by discussing simple Brownian motion and its derivative, "white noise." Right away we would like to demystify ideas that almost fit the usual calculus framework, but with some differences. Readers familiar with the Dirac delta function (an infinitesimal spike) have been in this situation before.

The following simple MATLAB code produces a figure of the sort that resembles logarithmic stock market prices. Every time we execute this code we get a different random picture (shown in Figure 1 Left).

```
x = [0:h:1];   %Think of h as "Delta x"
dW = randn(length(x),1)*sqrt(h); %Think sqrt(Delta x)
W= cumsum(dW);
plot(x,W)
```

Intuitively, we break $[0, x]$ into intervals each having length Δx. For each interval, we sample ΔW which is a zero mean normal with variance equal to Δx and then sum them up. Thus, if we look at one point x, we have

$$W(x) \stackrel{d}{=} \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} \Delta W = \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} \mathrm{G} \cdot \sqrt{\Delta x}.$$

$W(x)$ is a normal with mean 0 and variance $\frac{x}{\Delta x} \times \Delta x = x$, i.e. $W(x) \sim N(0, x)$ (shown in Figure 1 Center and Right). We can write this as $W(x) = \sqrt{x} \cdot \mathrm{G}$, G denoting a standard normal. In particular, $W(1)$ is a standard normal, and $W(x) - W(y)$ has mean 0 and variance $(x - y)$, i.e. $W(x) - W(y) = \sqrt{x - y} \cdot \mathrm{G}$. $W(x)$ is known as the *Wiener process* or *standard Brownian motion*. It has the property that $W(x) - W(y)$ has the distribution $N(0, x - y)$.

A suggestive notation is

$$\mathrm{d}W = (\text{standard normal}) \cdot \sqrt{\mathrm{d}x}$$

and the corresponding Wiener process is

$$W(x) = \int \mathrm{d}W.$$

The $\sqrt{\mathrm{d}x}$ seems troubling as notation, until one realizes that the `cumsum` then has quantities that do not depend on h (or Δx) at all. Like the standard integral, mathematics prefers quantities that at least in the limit do not depend on the discretization size or method. Random quantities are the same. The $\sqrt{\mathrm{d}x}$ captures the idea that variances add when adding normals. If each increment depends on $\mathrm{d}x$ instead of $\sqrt{dx}$, then there will be no movement at all because the variance of $W(x)$ will be $\frac{x}{\Delta x} \times (\Delta x)^2 = x \times \Delta x$ which will be 0 when $\Delta x \to 0$.

The derivative $W'(x) = \frac{\mathrm{d}W}{\mathrm{d}x}$ at first seems strange. The discretization would be `dW/h` in the MATLAB code above, which is a discrete-time white noise process. At every point, it is a normal with mean 0 and variance $1/h$, and the covariance matrix is $\frac{1}{h}\mathbb{I}$.In the continuous limit, the differential form $\mathrm{d}W$ denotes a *white noise process* formally satisfying

$$\int f(x, W)W'(x)\mathrm{d}x = \int f(x, W)\mathrm{d}W.$$

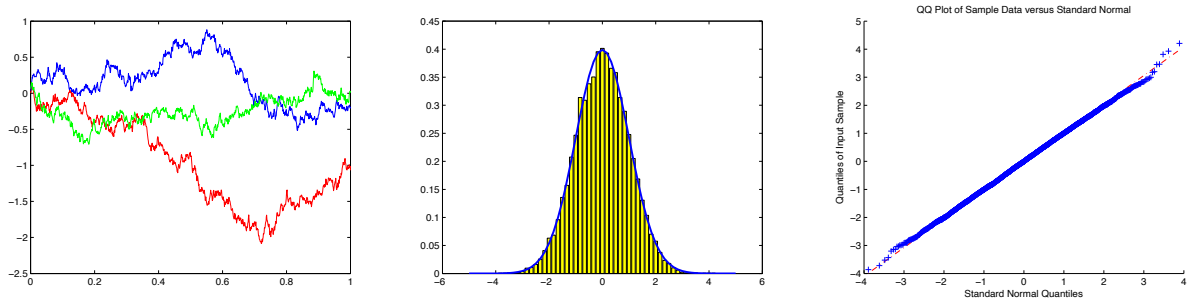

Figure 28.1: Left: Sample paths for standard Brownian motion; Center: histogram of $W(1)$ vs. the pdf of the standard normal; Right: quantile-quantile plot of $W(1)$.

Its covariance function is the Dirac delta $\mathrm{d}W_x \mathrm{d}W_y = \delta(x-y)$. We might say that $W'(x)$ has a "variance density" of 1, referring to the variance divided by the step size of the discretization.

In general we can consider integrals of the form

$$\int_0^x f(t)\mathrm{d}W = \lim_{\Delta x \to 0} \sum_{i=1}^{\left[\frac{x}{\Delta x}\right]} f(i\,[\Delta x])\Delta W,$$

which discretizes to `cumsum(f(t).*dW)`. We can think of $\mathrm{d}W$ as an operator such that $f\,\mathrm{d}W$ is a distribution—not a function in the classical sense, but able to serve as the differential in a stochastic integral. Multiplication by $\mathrm{d}W$ is called the *white noise transformation* Ross (2009).

28.0.2 Three local eigenvalue behaviors; three stochastic differential operators

The most commonly studied random matrix models over the years have been the Gaussian, Wishart, and MANOVA ensembles, also known as the Hermite, Laguerre, and Jacobi ensembles. We are primarily concerned with local eigenvalue behavior (that is, a single eigenvalue or a small number of eigenvalues rather than the entire spectrum), which depends on the location in the spectrum as well as the random matrix distribution. Remarkably, though, we see only three different local behaviors among the classical ensembles:

	Region of spectrum		
Ensemble	Left edge	Interior	Right edge
Hermite	soft edge	bulk	soft edge
Laguerre	hard edge	bulk	soft edge
Jacobi	hard edge	bulk	hard edge

In the next section, we will explore how the operator

$$\mathcal{A}_\beta = \frac{\mathrm{d}^2}{\mathrm{d}x^2} - x + \frac{2}{\sqrt{\beta}}W'(x)$$