DPP Notes (in progress) for 18.338 by Alan Edelman

September 23, 2022

Contents

Chapter 1

Introduction

Consider the n random eigenvalues of an $n \times n$ matrix, or, say, the set of those eigenvalues that are in some interval I such as $[0, 1]$. When running the second question as an experiment, we might find that maybe none of the n eigenvalues are in $[0, 1]$, or maybe some or maybe all of them are in $[0, 1]$. Thus every time we run the experiment we get a set of k numbers in $[0, 1]$, where k could be $0, 1, \ldots, n$. For important random matrices such as the GUE, this is the prototype for a determinantal point process – a random set of "points" (in this case in $\Re$) where one does not know in advance how many will show up in the set.

Determinantal Point Processes have become a big subject since the 2013 publication of

[(**?**, Alex Kulesza Ben Taskar) proposing applications to machine learning.

Very important early specific cases of DPPs arise in random matrix theory.

[In the current version of the book there is chapters 24 and the empty 25 and 26 that are related to this chapter]

[history from kuleza AND taskar is pretty good need a rewrite or something] In fact, years before Macchi gave them a general treatment, specific DPPs appeared in major results in random matrix theory [Mehta and Gaudin, 1960, Dyson, 1962a,b,c, Ginibre, 1965], where they continue to play an important role [Diaconis, 2003, Johansson, 2005b]. Recently, DPPs have attracted a flurry of attention in the mathematics community [Borodin and Olshanski, 2000, Borodin and Soshnikov, 2003, Borodin and Rains, 2005, Borodin et al., 2010, Burton and Pemantle, 1993, Johansson, 2002, 2004, 2005a, Okounkov, 2001, Okounkov and Reshetikhin, 2003, Shirai and Takahashi, 2000], and much progress has been made in understanding their formal combinatorial and probabilistic properties. The 4 term "determinantal" was first used by Borodin and Olshanski [2000], and has since become accepted as standard. Many good mathematical surveys are now available [Borodin, 2009, Hough et al., 2006, Shirai and Takahashi, 2003a,b, Lyons, 2003, Soshnikov, 2000, Tao, 2009].

- In RMT DPPs are continuous

- but there have been growth processes that are discrete? are these DPPs?

- continuous may seem mysterious but they are really not. They are great for looking at probabilities where something isn't , so spacings, largest eigenvalue (anything larger is empty space)
- the connection between continuous and discrte
- the connection to numerical linear algebra and matrix factorizations

Chapter 2

Determinantal Point Processes (DPP)

2.1 Determinant Formulas

2.1.1 $\det(A + B)$

This section really has one underlying determinant formula (Lemma 2.1) that is simple but is never shown in linear algebra classes, instead the usual advice to a first year linear algebra student who might be curious about $\det(A + B)$ is that it is not $\det(A) + \det(B)$. The story often fizzles out at that point. This section shows a great many interesting consequences of the determinant of $A + B$ formula.

Lemma 2.1.
$$\det(A + B) = \sum_{\mathcal{I} \in \mathcal{P}} \det(C_{\mathcal{I}}),$$
where $\mathcal{P}$ contains the subsets of $\{1, \ldots, n\}$ and the (column) mixture of A and B matrix $C_{\mathcal{I}}$ is defined by

$$C_{\mathcal{I}}[:, i] = \begin{cases} A[:, i], & \textit{if } i \notin \mathcal{I}. \\ B[:, i], & \textit{if } i \in \mathcal{I} \end{cases}. \tag{2.1}$$

This formula may be found, for example in row form, as formula (4) in (Marcus, 1990), and is an easy consequence of the multilinearity of determinants.

Exercise 2.1 Look up the definition of a Gray code, and show how one can swap out one column at a time from the C matrix in Equation 2.1 when summing the 2^n determinants in Julia.

The Julia function `mixAandB` creates $C_{\mathcal{I}}$ by first copying A into a matrix named C (line 8) and then overwriting the columns in $\mathcal{I}$ with the corresponding columns from B (line 9). Line 14 performs the sum over all subsets $\mathcal{I}$.

```
using LinearAlgebra,Combinatorics, Printf
N = 3

A = randn(N,N)
B = randn(N,N)

function mixAandB(A,B,ℐ)
    C = copy(A)
    C[:,ℐ] .= B[:,ℐ]
    return C
end

## Check identity
det(A+B), sum( det(mixAandB(A,B,ℐ)) for ℐ ∈ powerset(1:N) )

(1.8246027487598038, 1.8246027487598044)
```

2.1.2 Labeled sums of principal minors

Theorem 2.2 in the next paragraph displays valuable special cases of Lemma 2.1. One can start with Equation 2.7 and recognize that all of the preceding forms are special cases. Somewhat less obvious, is that Equation 2.2 alone implies all the forms that follow. We thought it would be useful for the reader to display all of these forms for easy reference. We remark that Equation 2.4 is (upon negating z) the well known formula for the characteristic polynomial of a matrix (Horn and Johnson, 2012, see (1.2.13)).

Theorem 2.2. *Let A be an $n \times n$ matrix, and $\mathcal{P}$ the set of subsets of $\{1, \ldots, n\}$. We have the following formulas for labeled sums of principal minors:*

$$\det(I + A) = \sum_{\mathcal{I} \in \mathcal{P}} A\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.2}$$

$$\det(I + zA) = \sum_{k=0}^{n} z^k \sum_{|\mathcal{I}|=k} A\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.3}$$

$$\det(zI + A) = \sum_{k=0}^{n} z^{n-k} \sum_{|\mathcal{I}|=k} A\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.4}$$

$$\det\left(I + \begin{pmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_n \end{pmatrix} A\right) = \sum_{\mathcal{I} \in \mathcal{P}} z_{\mathcal{I}} A\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.5}$$

$$\det\left(\begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} + A\right) = \sum_{\mathcal{I} \in \mathcal{P}} w_{\bar{\mathcal{I}}} A\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.6}$$

$$\det\left(\begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} + \begin{pmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_n \end{pmatrix} A\right) = \sum_{\mathcal{I} \in \mathcal{P}} z_{\mathcal{I}} w_{\bar{\mathcal{I}}} A\binom{\mathcal{I}}{\mathcal{I}}. \tag{2.7}$$

where $z_{\mathcal{I}}$ denotes $\prod_{i \in \mathcal{I}} z_i$, $w_{\bar{\mathcal{I}}}$ denotes the product over the complement of $\mathcal{I}$, $\prod_{i \notin \mathcal{I}} w_i$, and $A\binom{\mathcal{I}}{\mathcal{I}}$ is the principal minor of A formed from rows and columns in monotonic order with indices in $\mathcal{I}$ (with the convention that a 0×0 determinant is 1).

Proof. Our strategy will be to prove Equation 2.2 first, and then show that implies Equation 2.7, from which the rest are special cases.

Proof of Equation 2.2

Method 1: One can directly apply Lemma 2.1 to $\det(I+A)$ and notice that the matrices $C_{\mathcal{I}}$ have determinants that will only depend on the rows and columns with indices in $\mathcal{I}$, i.e. $\det(C_{\mathcal{I}})$ is exactly $A\binom{\mathcal{I}}{\mathcal{I}}$.

Method 2: We can avoid Lemma 2.1 by directly expanding the $n!$ determinant formula for $I+A$ by first conditioning that we will use the diagonal ones in $\bar{\mathcal{I}}$, i.e. $I_{ii}=1$, if $i \notin \mathcal{I}$, and recognizing that the sum of the remaining terms is $A\binom{\mathcal{I}}{\mathcal{I}}$.

Exercise 2.2 Write out a small case of method 2.

Method 3: Is essentially the same as Method 2 with the language of permutations. Recalling that the $n!$ formula for determinants is a sum over permutations, we first condition on the permutations with fixed points in $\bar{\mathcal{I}}$, and then proceed to sum over these permutations using the elements of A with row and column indices in $\mathcal{I}$.

Exercise 2.3 Verify 2.2 directly when A is a diagonal matrix Λ by showing $\prod(1+\lambda_i)=\sum_{\mathcal{I}\in\mathcal{P}}\lambda_{\mathcal{I}}$, where $\lambda_{\mathcal{I}}$ is defined to be $\prod_{i\in\mathcal{I}}\lambda_i$ and the empty product is 1. Extend this verification to upper triangular Λ.

Exercise 2.4 Write out the terms in the three methods of the Proof of Equation 2.2, when $n=3$.

Exercise 2.5 There is another proof of Equation 2.2 based on the Cauchy-Binet formula in chapter 12.4.1 (page 224) and $\operatorname{tr}(A^{(k)})=\sum_{|\mathcal{I}|=k}\lambda_{\mathcal{I}}=\sum_{|\mathcal{I}|=k}A\binom{\mathcal{I}}{\mathcal{I}}$.

Proof of Equation 2.7

We apply Equation 2.2, by computing

$$\det(I+\begin{pmatrix} z_1/w_1 & & & \\ & z_2/w_2 & & \\ & & \ddots & \\ & & & z_n/w_n \end{pmatrix}A),$$

and then we multiply both sides of the result by $\det\begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix}=w_1w_2\dots w_n$. Since the result is a polynomial there is no concern about dividing by 0.

2.1.3 Indicator Function Formulas

We define a diagonal matrix of 1's and 0's $I_{\mathcal{I}}$ by $(I_{\mathcal{I}})_{ii}=\begin{cases}1, & \text{if } i\in\mathcal{I}\\ 0, & \text{if } i\notin\mathcal{I}\end{cases}$. This can be thought of as the diagonal matrix defined from the indicator function of $\mathcal{I}$.

Corollary 2.3.

$$\det(I_{\bar{\mathcal{I}}_1} + I_{\mathcal{I}_2} A) = \sum_{\mathcal{I}_1 \subseteq \mathcal{J} \subseteq \mathcal{I}_2} A\binom{\mathcal{J}}{\mathcal{J}}. \tag{2.8}$$

Proof. This is a special case of Formula 2.7. We note in this case that $z_{\mathcal{I}_2} = 1$ only when $\mathcal{J} \subseteq \mathcal{I}_2$, and $w_{\bar{\mathcal{I}}_1} = 1$ exactly when $\bar{\mathcal{J}} \subseteq \bar{\mathcal{I}}_1$ which is to say $\mathcal{I}_1 \subseteq \mathcal{J}$.

We list some special cases

$$\det(I + I_{\mathcal{I}} A) = \sum_{\mathcal{J} \subseteq \mathcal{I}} A\binom{\mathcal{J}}{\mathcal{J}}, \tag{2.9}$$

$$\det(I_{\bar{\mathcal{I}}} + A) = \sum_{\mathcal{J} \supseteq \mathcal{I}} A\binom{\mathcal{J}}{\mathcal{J}}, \tag{2.10}$$

$$\det(I_{\bar{\mathcal{I}}} + I_{\mathcal{I}} A) = A\binom{\mathcal{I}}{\mathcal{I}}. \tag{2.11}$$

2.2 A discrete probability space based on determinants

2.2.1 DPPs at a glance

Notational remark: Subsets of $\{1, \ldots, n\}$ use calligraphic letters such as $\mathcal{I}, \mathcal{J}, \ldots$ and when these subsets are random, we use bold face: $\boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{J}}, \ldots$. We may abbreviate $Pr(\boldsymbol{\mathcal{I}} = \mathcal{I}), Pr(\boldsymbol{\mathcal{J}} = \mathcal{I})$ etc., as $Pr(\mathcal{I})$ when the context is clear.

what notation are we using for probability Pr, P, etc???

A discrete DPP is a random subset $\boldsymbol{\mathcal{J}}$ chosen from the 2^n subsets of $\{1, \ldots, n\}$ defined equivalently from any of the three conditions in terms of a fixed reference subset $\mathcal{I}$ and symmetric positive (semi-)definite matrices L or K where $K = L(L+I)^{-1}$ or $L = K(I-K)^{-1}$.

pdf: $p_{\mathcal{I}} \equiv Pr(\mathcal{I} = \boldsymbol{\mathcal{J}}) = L\binom{\mathcal{I}}{\mathcal{I}} / \det(I + L)$.

ccdf: $c_{\mathcal{I}} \equiv Pr(\mathcal{I} \subseteq \boldsymbol{\mathcal{J}}) = K\binom{\mathcal{I}}{\mathcal{I}}$ (complementary cdf: random $\boldsymbol{\mathcal{J}}$ includes reference $\mathcal{I}$ as subset.)

cdf: $Pr(\mathcal{I} \supseteq \boldsymbol{\mathcal{J}}) = (I - K)\binom{\bar{\mathcal{I}}}{\bar{\mathcal{I}}}$.

There is a minor technical difficulty described in the next section seeming to show that L does not exist if Pr(∅)= 0. The difficulty is indeed minor in that while L may not exist (may have eigenvalues at infinity), everything we need from L does.

If $f(\mathcal{I})$ is a function defined on the 2^n subsets of $\{1, \ldots, n\}$, then of course

$$E(f(\boldsymbol{\mathcal{I}})) = \sum_{\mathcal{I} \subseteq \{1, \ldots, n\}} p_{\mathcal{I}} f(\mathcal{I}).$$

We will see that the inclusion-exclusion principle (a special case of Möbius inversion) lets us write the same sum in terms of the $c_{\mathcal{I}}$.

In random matrix theory the ccdf is often called (somewhat confusingly) a "correlation function" following statistical mechanics nomenclature and sometimes "joint intensities" or "factorial moments" in the mathematics literature. Continuous DPPs may seem more mysterious, but they are really not much different. The K matrix becomes a continuous by continuous kernel $K(x, y)$ instead of K_{ij}. One must indicate the set on which x and y are defined, often a finite or infinite interval. One can still compute finite determinants:

maybe place later

$$K\binom{x_1, \ldots, x_k}{x_1, \ldots, x_k},$$

which now represents a sort of density in that

$K\binom{x_1,\ldots,x_k}{x_1,\ldots,x_k} dx_1 \ldots dx_k$ is the probability that if P is the random point process, a point will be found in P in intervals of size dx_i around x_i for $i = 1, \ldots, k$.

One can still compute $K(I-K)^{-1}$ if $I-K$ is invertible, and if K is a projection ($\int_z K(x,z)K(z,y) = K(x,y)$) (that is $K^2 = K$) then P always has a fixed number of points.

For example the DPP associated with the 2x2 GUE is the kernel $K(x, y) = \frac{1}{\sqrt{\pi}} e^{(-x^2-y^2)/2}(1 + 2xy)$ defined on $R \times R$. This is a continuously infinite by continuously infinite "matrix" but only has rank 2. It is a projection and so all samples have size 2. (As it must because all 2x2 matrices have two eigenvalues.) One can strict K to, say, the positive real axis cross the positive real axis. Exercise ?? explores some properties of this matrix.

should i be less sloppy about dx's? less sloppy about kernels vs operators?

reference definition of kernels

Exercise, show the eigenvalues of K are $\frac{1}{2} \pm \frac{1}{\sqrt{2\pi}}$. From there compute the eigenvalues of L and use ???? to show that the probability of eigenvalues with opposites signs is $1/2 + 1/\pi$, while the probability of both eigenvalues being positive is $1/4 - 1/(2\pi)$, and both negative is also $1/4 - 1/(2\pi)$.
Also one can compute $K[fg]$ where $f(y) = e^{-y^2/2}$ and $g(y) = e^{-y^2/2}$ as another way to compute the eigenvalues of K and also the eigenvectors. From this information it is possible to write an exact expression for $L(x, y)$.

2.2.2 The "L" Method: Assigning Probabilities with determinants

The "L approach" is the most straightforward way to explain Determinantal Point Processes, but it has a minor technical drawback that requires $\mathbb{P}(\varnothing) \neq 0$. (though it is not a problem as a limiting case as discussed at the end of this section.) The "K" method is slightly less straightforward (but can be more natural to work with) and will be discussed in Section 2.2.5.

We can define probabilities on subsets $\mathcal{I}$ of $\{1, \ldots, n\}$. We start with a positive (semi-)definite matrix L and define

$$\mathbb{P}(\mathcal{I}) = L\binom{\mathcal{I}}{\mathcal{I}} / \det(I + L). \tag{2.12}$$

As principal minors are positive (non-negative), we see from Equation 2.2 that this construction turns the power set $\mathcal{P}$ into a probability space. This is what is known as a *Determinantal Point Process* or DPP.

In the example below the $2^3 = 8$ subsets of $\{1, 2, 3\}$ are assigned probabilities that add to 1.

```julia
# Generate a random DPP
using Combinatorics
Y = randn(N,N)
L = Y'Y
for ℐ ∈ powerset(1:N)
    @printf("  %10s :", ℐ)
    @printf(" %s\n",  det(L[ℐ,ℐ])/det(L+I) )
end

print("         Sum : ")
println(sum(  det(L[ℐ,ℐ])/det(L+I)  for ℐ ∈ powerset(1:N) ))
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
    Int64[] : 0.028661374227342145
        [1] : 0.06240780300683674
        [2] : 0.1262877549061006
        [3] : 0.16726921293316652
     [1, 2] : 0.23299470329092098
     [1, 3] : 0.22273171137760317
     [2, 3] : 0.09372276635492809
  [1, 2, 3] : 0.06592467390310236
        Sum : 1.0000000000000007
```

We recall Equation 2.5 which states

$$\det(I + ZL) = \sum_{\mathcal{I} \in \mathcal{P}} z_{\mathcal{I}} L\binom{\mathcal{I}}{\mathcal{I}},$$

where $Z = \begin{pmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_n \end{pmatrix}$ which allows us to conclude

$$\frac{\det(I + ZL)}{\det(I + L)} = \sum_{\mathcal{I} \in \mathcal{P}} z_{\mathcal{I}} Pr(\mathcal{I}),$$

which we may interpret as a generating function for all the probabilities. If we write $L = K/(I - K)$, we obtain that

$$\det((I - K) + ZK) = \sum_{\mathcal{I} \in \mathcal{P}} z_{\mathcal{I}} Pr(\mathcal{I}), \tag{2.13}$$

which nicely shows that even if $(I - K)$ is not invertible, so that even if L does not exist, the quantities

$$L\binom{\mathcal{I}}{\mathcal{I}} / \det(I + L),$$

would still exist as coefficients of $z_{\mathcal{I}}$.

Alan: exercise: (a) Let $Z = (1-z)I$ in Equation 2.13 to show that

$$\det(I - zK) = \sum_{k=0}^{n}(1-z)^k \sum_{|\mathcal{I}|=k} Pr(\mathcal{I}) = \sum_{k=0}^{n}(1-z)^k Pr(|\boldsymbol{\mathcal{I}}| = k).$$

In words, the coefficients of the characteristic polynomial of K written in terms of powers of $(1-z)$ contain the probabilities there are k elements in $\boldsymbol{\mathcal{I}}$. We can therefore take derivatives to extract coefficients to show that

$$Pr(|\boldsymbol{\mathcal{I}}| = k) = \frac{(-1)^k}{k!}\frac{d^k}{dz^k}\det(I - zK)|_{z=1}.$$

(b) Use Jacobi's formula for the derivative of the determinant to derive a formula in terms of K for $Pr(|\boldsymbol{\mathcal{I}}| = 1)$. Turn this into an L formula and explain why this is obviously correct.

2.2.3 The determinant with "eigen" parameters

Suppose that $L = Q\Lambda Q^T$ is an eigendecomposition of L. We may use the Cauchy-Binet Theorem (Theorem ????) to prove that

$$L\binom{\mathcal{I}}{\mathcal{I}} = \sum_{|\mathcal{J}|=|\mathcal{I}|} \lambda_{\mathcal{I}} Y_{\mathcal{I}} \binom{\mathcal{J}}{:}^2,$$

where $Y_{\mathcal{I}} = Q[:, \mathcal{I}]$ denotes the columns of Q with indices from $\mathcal{I}$ and the colon ":" denotes all the columns. (Julia-like notation: $\det(Y_{\mathcal{I}}[\mathcal{J}, :])^2$.)

We therefore see that

$$Pr(\mathcal{I}) = \frac{L\binom{\mathcal{I}}{\mathcal{I}}}{\det(I+L)} = \sum_{|\mathcal{J}|=|\mathcal{I}|} \frac{\lambda_{\mathcal{I}}}{\prod_{i=1}^{N}(\lambda_i + 1)} Y_{\mathcal{I}} \binom{\mathcal{J}}{:}^2.$$

This formula expresses a DPP as a mixture as discussed in Section 2.4.1. It is a key formula that allows for a popular algorithm for generating samples from a DPP.

We also observe that if we have a projection DPP, (more about projection DPPs can be found in Section 2.3) so that $K = YY^T$ where $Y^TY = I_r$, then we can write

$$Pr(\mathcal{J}) = Y\binom{\mathcal{J}}{:}^2, \text{ for } |\mathcal{J}| = r.$$

Exercise 2.6 Verify directly from the Cauchy-Binet theorem and $Y^TY = I_r$ that the $\binom{N}{r}$ numbers with $|\mathcal{J}| = r$ satisfy $Y\binom{\mathcal{J}}{:}^2$ add to 1, hence they serve as probabilities. (Answered in Lemma B of Section 2.4.2, but try to do this yourself first.)

2.2.4 Conditional Probabilities

(may appear elsewhere)

2.2.5 "The K method:" The analog of a ccdf or a cdf

A question of interest for DPPs is given a $J = \{j_1, \ldots, j_k\}$, what is the probability that the random subset includes $j_1, \ldots, j_k$? An equivalent formulation is the expectation value of the product of indicator functions $\chi_{j_1}(\mathcal{I}) \times \ldots \times \chi_{j_k}(\mathcal{I})$. The answer will turn out to be $K\binom{\mathcal{I}}{\mathcal{I}}$, where $K = L(I+L)^{-1}$. Many authors prefer using K as the starting point for an exposition on DPPs.

In Section 2.2.2 we defined the probability of $\mathcal{I} \subset \{1, \ldots, n\}$ from a positive (semi-)definite matrix L by $Pr(\mathcal{I}) = L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)$. The K method works with $K = L(I+L)^{-1}$. The requirement on K is that K is symmetric positive definite with $0 \leq K \leq I$ meaning the eigenvalues of K are in $[0,1]$.

In elementary probability one defines the cdf (sometimes just called the distribution) $F(x)$ of a discrete random variable as the probability the random variable takes a value $\leq x$. If the random variable is y we write

$$F(x) = Pr(y \leq x) = \sum_{y \leq x} Pr(y).$$

mention ccdf

By analogy, we can define the cdf $F(\mathcal{I})$ on subsets of $\{1, \ldots, n\}$ by defining

$$F(\mathcal{I}) = Pr(\mathcal{J} \supseteq \mathcal{I}) = \sum_{\mathcal{J} \supseteq \mathcal{I}} Pr(\mathcal{J}),$$

the sum being over all $\mathcal{J}$ that contains $\mathcal{I}$. (We remark that the analogy would be cleaner if the inclusion convention went the other way.) In the following we will show that $F(\mathcal{I}) = K\binom{\mathcal{I}}{\mathcal{I}}$.

For a determinantal point process, then,

$$F(\mathcal{I}) = \sum_{\mathcal{J} \supseteq \mathcal{I}} Pr(\mathcal{J}) = \sum_{\mathcal{J} \supseteq \mathcal{I}} L\binom{\mathcal{J}}{\mathcal{J}} / \det(I+L) = \det(I_{\bar{\mathcal{I}}} + L)/\det(I+L),$$

using Equation 2.10.

To place this in a more useful form, we write

$$\det(I_{\bar{\mathcal{I}}} + L)/\det(I+L) = \det(I_{\bar{\mathcal{I}}}(I+L) + I_{\mathcal{I}}L)/\det(I+L) = \det(I_{\bar{\mathcal{I}}} + I_{\mathcal{I}}L(I+L)^{-1}) = K\binom{\mathcal{I}}{\mathcal{I}},$$

where we let $K = L(I+L)^{-1}$ and use Equation 2.11.

We summarize by stating

Corollary 2.4. *If $Pr(\mathcal{I}) = L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)$, then $Pr(J \supseteq \mathcal{I}) = K\binom{\mathcal{I}}{\mathcal{I}}$, where $K = L(I+L)^{-1}$ or equivalently $L = K(I-K)^{-1}$ when $I-K$ is invertible.*

For a CDF going in the "right" direction ("$J \subseteq \mathcal{I}$") we also have

Corollary 2.5. *If $Pr(\mathcal{I}) = L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)$, then $Pr(J \subseteq \mathcal{I}) = (I-K)\binom{\tilde{\mathcal{I}}}{\tilde{\mathcal{I}}}$, where $K = L(I+L)^{-1}$ or equivalently $L = K(I-K)^{-1}$ when $I-K$ is invertible.*

Proof. See Exercise 2.7.

Alan: Short Essay:
[Probably switch? Use J not calJ some DPP literature makes the random variable bold – that may be a great idea] In this section we will see that the "K" method refers to defining

$$Pr(\mathcal{I} \subseteq \mathcal{J}) = \mathcal{K}\binom{\mathcal{I}}{\mathcal{I}}.$$

This can be a little bit confusing. It can be tricky to remember what is the random variable. So let us be explicit:

It is $\mathcal{J}$ that is drawn randomly from a DPP .
It is $\mathcal{I}$ that is non-random.

We pick an $\mathcal{I}$, and ask for the probability that the random $\mathcal{J}$ includes every element of $\mathcal{I}$. Somehow two cases are easiest to remember: 1) if $\mathcal{I} = \varnothing$ then the answer 1 fits nicely (vacuously, we should say) as a check. 2) if $\mathcal{I} = \{i\}$ a singleton, then the probability that i shows up when we draw from a DPP, $Pr(i \in \mathcal{J}) = K_{ii}$.

Exercise 2.7 Prove Corollary 2.5 from Corollary 2.4 and Exercise 2.9.

2.2.6 The complementary DPP

We proceed to show that a new DPP, the complementary DPP, can be obtained from an existing DPP by any of the following (mostly) equivalent appoaches:

- Replace K with $I - K$
- $\Pr(\mathcal{J} \cap \mathcal{I} = \varnothing) = (I-K)\binom{\mathcal{J}}{\mathcal{J}}$ for all $\mathcal{J}$
- Replace L with L^{-1} if L is invertible
- $\Pr(\tilde{\mathcal{I}}) = L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)$.

It is an easy exercise (Exercise ??) using Equation 2.11 or alternatively the block LU decomposition to show that

$$M\binom{\mathcal{I}}{\mathcal{I}} = M^{-1}\binom{\bar{\mathcal{I}}}{\bar{\mathcal{I}}}\det M \tag{2.14}$$

for all invertible matrices M. This may be written for invertible L:

$$\frac{L\binom{\mathcal{I}}{\mathcal{I}}}{\det(I+L)} = \frac{L^{-1}\binom{\bar{\mathcal{I}}}{\bar{\mathcal{I}}}}{\det(I+L^{-1})}.$$

Exercise 2.8

1. Use Equation 2.11 to prove Equation 2.14.

2. Alternate derivation of Equation 2.14 : separate M into indices based on $\mathcal{I}$ and $\bar{\mathcal{I}}$. Assuming the block LU exists and is invertible, write LU and $U^{-1}L^{-1}$ and conclude Equation 2.14 by taking determinants. Use continuity to complete the argument if block LU does not exist.

We see that given a DPP defined by L, it is sensible to define a complementary DPP defined by L^{-1} which interchanges the probabilities for $\mathcal{I}$ and $\bar{\mathcal{I}}$. Thus for the complementary DPP

$$Pr_{comp}(\mathcal{I}) = \frac{L\binom{\bar{\mathcal{I}}}{\bar{\mathcal{I}}}}{\det(I+L)}.$$

We can write out the generating function

$$\sum_{\mathcal{I}\in\mathcal{P}} z_{\mathcal{I}} \frac{L\binom{\bar{\mathcal{I}}}{\bar{\mathcal{I}}}}{\det(L+I)} = \sum_{\mathcal{I}\in\mathcal{P}} z_{\bar{\mathcal{I}}} \frac{L\binom{\mathcal{I}}{\mathcal{I}}}{\det(L+I)} = \frac{\det(Z+L)}{\det(I+L)} = \frac{\det(I+ZL^{-1})}{\det(I+L^{-1})},$$

diectly using Equation 2.6.

Exercise 2.9 Show that we can switch K and $I-K$ in Equation 2.13 to obtain the generating function for the complementary DPP. In summary we can interchange L with L^{-1} or K and $I-K$ to get the complementary DPP.

2.2.7 Relationship to Möbius Inversion & Inclusion-Exclusion

Alan: hopefully there will be a möbius inversion chapter in combinatorics, maybe this will be part of combinatorics, or maybe not

Consider the upper triangular ζ matrix with rows and colums labeled by $\mathcal{P}$, the power-set of $1,\ldots,n$, with entries defined by $\begin{cases} \zeta(A,B)=1 & \text{if } A\subseteq B \\ \zeta(A,B)=0 & \text{otherwise} \end{cases}$. Exercise ???? computes the inverse of this matrix to be $\begin{cases} \mu(A,B)=(-1)^{|B-A|} & \text{if } A\subseteq B \\ \mu(A,B)=0 & \text{otherwise} \end{cases}$. ζ and μ are shown in Figure ???? for $n=3$.

Exercise 2.10 Show that the upper triangular ζ matrix is a reordering of the n-fold Kronecker product of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and therefore the inverse is the n-fold Kronecker product of $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$. (If the rows and columns are labeled $0,1,\ldots,2^{n-1}$ then an element of $\mathcal{P}$ is encoded by the 1s in the binary representation.) Use this to deduce that

$$\begin{cases} \mu(A,B) = (-1)^{|B-A|} & \text{if } A \subseteq B \\ \mu(A,B) = 0 & \text{otherwise} \end{cases}.$$

We recall that ζ serves to compute a sum/integral or a cumulative probability function from the probability mass function, and μ goes the other way, essentially acting as a difference/derivative.

MobiusSubsetPoset.jl > ...

```julia
using Combinatorics, LinearAlgebra, NamedArrays
#worth seeing Rota: https://www.maths.ed.ac.uk/~v1ranick/papers/rota1.pdf
# or Stanley: http://www-math.mit.edu/~rstan/ec/ec1.pdf#page=303
# also wikipedia: https://en.wikipedia.org/wiki/M%C3%B6bius_inversion_formula#On_posets
# and wikipedia on inclusion-exclusion: https://en.wikipedia.org/wiki/Inclusion%E2%80%93exclusion_principle#Other_forms

𝒫(N) = collect(powerset(1:N))
ζ(N) = Int.([ A ⊆ B for A∈𝒫(N), B∈𝒫(N) ])
μ(N) =   Int.(inv(ζ(N)))

N = 3
labels = string.(𝒫(N));labels[1] = "∅"
display(NamedArray( ζ(N) , (labels,labels) ) )
display(NamedArray( μ(N) , (labels,labels) ) )

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

A \ B	∅	[1]	[2]	[3]	[1, 2]	[1, 3]	[2, 3]	[1, 2, 3]
∅	1	1	1	1	1	1	1	1
[1]	0	1	0	0	1	1	0	1
[2]	0	0	1	0	1	0	1	1
[3]	0	0	0	1	0	1	1	1
[1, 2]	0	0	0	0	1	0	0	1
[1, 3]	0	0	0	0	0	1	0	1
[2, 3]	0	0	0	0	0	0	1	1
[1, 2, 3]	0	0	0	0	0	0	0	1

8×8 Named Matrix{Int64}

A \ B	∅	[1]	[2]	[3]	[1, 2]	[1, 3]	[2, 3]	[1, 2, 3]
∅	1	-1	-1	-1	1	1	1	-1
[1]	0	1	0	0	-1	-1	0	1
[2]	0	0	1	0	-1	0	-1	1
[3]	0	0	0	1	0	-1	-1	1
[1, 2]	0	0	0	0	1	0	0	-1
[1, 3]	0	0	0	0	0	1	0	-1
[2, 3]	0	0	0	0	0	0	1	-1
[1, 2, 3]	0	0	0	0	0	0	0	1

Specializing to DPP's we can recast the formula in Equation 1.2.2 to read

$$\left(K\binom{\mathcal{I}}{\mathcal{I}}\right)_{\mathcal{I}\in\mathcal{P}} = \zeta * \left(L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)\right)_{\mathcal{I}\in\mathcal{P}} \tag{2.15}$$

$$\left(L\binom{\mathcal{I}}{\mathcal{I}}/\det(I+L)\right)_{\mathcal{I}\in\mathcal{P}} = \mu * \left(K\binom{\mathcal{I}}{\mathcal{I}}\right)_{\mathcal{I}\in\mathcal{P}} \tag{2.16}$$

We can write this formula in a number of ways some of which emphasize the inclusion-exclusion principle:

$$Pr(\mathcal{I}) = \sum_{\mathcal{J}\supseteq\mathcal{I}} (-1)^{|\mathcal{J}-\mathcal{I}|} Pr(\mathcal{J}\supseteq\mathcal{I}) \tag{2.17}$$

$$= L\binom{\mathcal{I}}{\mathcal{I}} / \det(I+L) = \sum_{\mathcal{J} \supseteq I} (-1)^{|\mathcal{J}-\mathcal{I}|} K\binom{\mathcal{I}}{\mathcal{I}} \tag{2.18}$$

$$= (-1)^{|\mathcal{I}|} \det(I_{\bar{\mathcal{I}}} - K). \tag{2.19}$$

There is something wonderful that inclusion-exclusion in this case can be verified with a determinant formula. (Exercise ??).

Exercise 2.11

1. The inclusion-exclusion proof: Use the formula for inclusion-exclusion as in https://en.wikipedia.org/wiki/Inclusion%E2%80%93exclusion_principle#Other_forms to verify $Pr(\mathcal{I}) = \sum_{\mathcal{J} \supseteq \mathcal{I}} (-1)^{|\mathcal{J}-\mathcal{I}|} Pr(\mathcal{J} \supseteq \mathcal{I})$.

2. The linear algebra proof: Use Formula 2.7 to verify

$$(-1)^{|\mathcal{I}|} \det(I_{\bar{\mathcal{I}}} - K) = L\binom{\mathcal{I}}{\mathcal{I}} / \det(I+L),$$

   where $L = K(I-K)^{-1}$. (Hint: Use $K = (I_{\bar{\mathcal{I}}} + I_{\mathcal{I}})K$).

3. Notice that (1.13)?? is true by inclusion-exclusion even if the probabilities are not determinantal. However (1.14) and (1.15) may be shown to be true as a linear algebra fact, or it may be seen as a special case of inclusion-exclusion when the probabilities are determinantal.

We demonstrate the equalities in (1.14) and (1.15) ??? in the following code:

```
using Combinatorics, LinearAlgebra, NamedArrays

𝒫(N) = collect(powerset(1:N))

N = 3
Y = randn(N,N)
L = Y'Y
K = L/(L+I)
Id(𝒥) = Diagonal([i∈𝒥 for i=1:N])
Īd(𝒥) = I - Id(𝒥)

PDF1 = [det(L[𝒥,𝒥])/det(L+I) for 𝒥∈𝒫(N)]
PDF2 = [ (-1)^length(𝒥) * det( Īd(𝒥) - K) for 𝒥∈𝒫(N)]
PDF3 = [ sum( (-1)^length(setdiff(ℑ,𝒥)) *det(K[ℑ,ℑ])  for ℑ∈𝒫(N) if ℑ ⊇ 𝒥 ) for 𝒥∈𝒫(N)   ]
[PDF1 PDF2 PDF3]
```

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL

8×3 Matrix{Float64}:
 0.141389   0.141389   0.141389
 0.0928025  0.0928025  0.0928025
 0.294837   0.294837   0.294837
 0.0897117  0.0897117  0.0897117
 0.17036    0.17036    0.17036
 0.0269284  0.0269284  0.0269284
 0.171184   0.171184   0.171184
 0.0127874  0.0127874  0.0127874
```

2.2.8 correlation functions or joint intensities

Theorem 2.6. *The following two relationships are equivalent:*

1) $\sum_{\mathcal{I}\supseteq J} p_{\mathcal{I}} = c_J$ *for all* J *with* $|J| = k$.

2)

$$E_{\mathcal{I}} \sum_{\substack{J\subseteq\mathcal{I}\\|J|=k}} f_k(J) = \sum_{|J|=k} c_J f_k(J) \tag{2.20}$$

for all functions f_k *of subsets of* $\{1,\ldots,n\}$ *of size* k.

Proof. In our notation J always denotes a set of size k while $\mathcal{I}$ has no such restriction. We observe in Equation 2.20 that inside the expectation of the lhs the sum is over the $\binom{\mathcal{I}}{k}$ subsets of $\mathcal{I}$ of size k. If $\mathcal{I}$ has fewer than k elements, this sum is empty.

Proof of Theorem 2.6:

($2 \implies 1$) Given a $\hat{J}$ with $|\hat{J}| = k$, let f_k be the indicator function $\chi_{\hat{J}}$ which is 1 only when $J = \hat{J}$. Then the lhs of Equation (2.20) which in general is

$$\sum_{\mathcal{I}} p_{\mathcal{I}} \sum_{\substack{J\subseteq\mathcal{I}\\|J|=k}} f_k(J),$$

simplifies to

$$\sum_{\mathcal{I}\supseteq\hat{J}} p_{\mathcal{I}}$$

because each $\mathcal{I}$ that contains $\hat{J}$ gets counted once for the k element subset that is exactly $\hat{J}$, while those that do not contain $\hat{J}$ are not counted at all. The rhs side simplifies to $c_{\hat{J}}$ which implies (1).

($1 \implies 2$) For the rhs of (1) multiply by $f_k(J)$ and sum over all J of size k. For the lhs we have to recognize that every J of size k is counted once for every $\mathcal{I}$ that contains it with value $p_{\mathcal{I}} f_k(J)$. Equivalently each $\mathcal{I}$ is counted exactly $\binom{\zeta}{\mathcal{I}})k$ times once for each subset of size k. □

There is a more elegant way to see Theorem 2.6. label

Obviously the two sets

$$\{(\mathcal{I},\mathcal{J}) : \mathcal{I}\supseteq\mathcal{J}\} \text{ and } \{(\mathcal{I},\mathcal{J}) : \mathcal{J}\subseteq\mathcal{I}\}$$

are identical to the set $\{(\mathcal{I},\mathcal{J}) : \zeta_{\mathcal{J},\mathcal{I}} = 1\}$.

One can use this to interchange summations:

$$\sum_{\mathcal{J}} \sum_{\mathcal{I}\supseteq\mathcal{J}} g(\mathcal{I},\mathcal{J}) = \sum_{\mathcal{I}} \sum_{\mathcal{J}\subseteq\mathcal{I}} g(\mathcal{I},\mathcal{J}) = \sum_{\mathcal{I},\mathcal{J}} \zeta_{\mathcal{J},\mathcal{I}} g(\mathcal{I},\mathcal{J}).$$

The key point is the first summation is over all subsets in the power set and the second summation runs over larger subsets ($\mathcal{I}$ is bigger than $\mathcal{J}$) in the lhs and in the rhs over smaller subsets ($\mathcal{J}$ is smaller than $\mathcal{I}$) .

A very important special case can be expressed with linear algebra. If f and p are vectors of length 2^N, the expression $f^T \zeta p$ when expanded into components can be written as such a double summation. Preferring matrix-vector notation, the two summations amount to

$$(f^T \zeta)p = f^T(\zeta p).$$

In particular, if the vectors are chosen so that $p_{\mathcal{I}} = Pr(\mathcal{I})$, and $c_{\mathcal{I}} = Pr(\mathcal{I} \subseteq \mathcal{J})$, then by definition $c = \zeta p$ and the above identity reads

$$E((f^T \zeta)_{\mathcal{I}}) = f^T c, \tag{2.21}$$

the left hand side is the expected value of a component of the vector $f^T \zeta$, where the component $\mathcal{I}$ is chosen at random with probability $pr(\mathcal{I})$.

2.3 Projection Matrix DPP's

We explore the interesting case when K is a projection matrix. These may be called "Projection DPPs" or sometimes in the literature "Elementary DPPs." We will see that when K is a rank r projection matrix, $Pr(\mathcal{I})$ can only be nonzero if $|\mathcal{I}| = r$. Thus we have $\binom{N}{r}$ index sets to consider rather than 2^N. There is no finite L matrix corresponding directly to the K but we will see that the DPP can be chosen as a limit of L matrices.

2.3.1 Projection Matrices

Suppose $L = \alpha P$ is a multiple of a rank r projection matrix, meaning that P is rank r, $P^2 = P$ and $P = P^T$. The eigenvalues of P are 0 and 1, with exactly r of them being 1. The corresponding $K = L(I + L)^{-1} = \frac{\alpha}{1+\alpha} P$. Notice that as $\alpha \to \infty$, L blows up, but K tends towards P.

The corresponding DPP from Equation 2.12 is

$$Pr(\mathcal{I}) = L\binom{\mathcal{I}}{\mathcal{I}} / \det(I + L).$$

It is easy to see from the eigenvalues of L that $\det(I + L) = (1 + \alpha)^r$, so we have

$$Pr(\mathcal{I}) = \frac{\alpha^{|\mathcal{I}|}}{(1 + \alpha)^r} P\binom{\mathcal{I}}{\mathcal{I}}, \tag{2.22}$$

which is 0 if $|\mathcal{I}| > r$ because P has rank r. Furthermore, as $\alpha \to \infty$, $Pr(\mathcal{I})$ tends to 0 if $|\mathcal{I}| < r$. The only important case is $|\mathcal{I}| = r$, with $Pr(\mathcal{I}) \to P\binom{\mathcal{I}}{\mathcal{I}}$ as $\alpha \to \infty$.

We summarize that Equation 2.22 takes the form in the limit as $\alpha \to \infty$

$$Pr(\mathcal{I}) = \begin{cases} P\binom{\mathcal{I}}{\mathcal{I}} & \text{if } |\mathcal{I}| = r \\ 0 & \text{otherwise} \end{cases}. \tag{2.23}$$

What we see is that only the $\binom{N}{r}$ subsets of size exactly r can have non-zero probability.

Alan: make an exercise out of non-symmetric projection matrices

2.3.2 Janossy Densities: A way to turn a projection DPP into ordinary DPPs

Suppose K is a rank k projection matrix and $\mathcal{J}$ is a given fixed subset of $\{1, \dots, N\}$. From the previous section, K defines a projection DPP. The random $\mathcal{I}$ of size k from the projection DPP allows us to consider the random subset $\mathcal{I} \cap \mathcal{J}$ which may be of size any integer from 0 to k. Interestingly this $\mathcal{I} \cap \mathcal{J}$ is an ordinary DPP as we are about to prove. In a way, we might say there are 2^N vanilla DPPs arising, one for each subset $\mathcal{J}$ of $\{1, \dots, N\}$.

2.4 Sampling DPPs

The code to sample from a DPP is remarkably short. The `randDPP` code below takes input the eigenvectors Y and eigenvalues Λ (as a vector) and returns a random subset of $1 : N$ with probability $L\binom{\mathcal{I}}{\mathcal{I}} / \det(L + I)$.

We will explain why this algorithm works by first explaining lines 14-17 which says that a general DPP is a mixture of projection DPP's (Section 2.3) and lines 3-12 explicitly sample from a projection DPP.

```
1   using LinearAlgebra, Combinatorics, Distributions, StatsBase
2
3   function randprojDPP(Y)
4       n = size(Y,2)
5       𝒥 = fill(0,n)
6       for k=1:n
7           p = mean(abs.(Y).^2, dims=2)
8           𝒥[k] = rand(Categorical(p[:]))
9           Y=(Y*qr(Y[𝒥[k],:]).Q )[:,2:end]
10      end
11      return(sort(𝒥))
12  end
13
14  function randDPP(Y,Λ)
15      mask = rand.(Bernoulli.(Λ./(Λ.+1)))
16      return(randprojDPP(Y[:,mask]))
17  end
```

2.4.1 Mixtures

In probability a mixture is a combination of random variables; a common case would be linear combinations.

To be precise let p be a probability vector with p_i denoting the probability of the index i (which can range over any index set). The random variable i is sometimes called a categorical random variable. Let j_i be a random variable for every i in the index set. Then if we pick j_i with probability p_i we say we have a mixture of the j_i. It is easy to see if $Pr(j_i) = M_{ij}$ then the probability of the mixture is $Pr(j) = \sum_i p_i M_{ij}$. The matrix M is known as a Markov matrix. It is often not constructed explicitly.

In this chapter, the random index will be $\mathcal{I}$ a subset of $1 : N$. We illustrate a relevant example for this chapter below with Julia. The first example rolls 4 fair coins (six times) and observes heads or tails. The second example rolls 4 unfair coins (six times) with probabilities $1/2, 2/3, 3/4, 4/5$ respectively. Julia prints the result as a sequence of bits, with the corresponding $\mathcal{I}$ appearing in yellow for illustration.

```
julia> Λ = [1 1 1 1]; [ rand.(Bernoulli.(Λ./(Λ.+1))) for i=1:6]
6-element Vector{BitMatrix}:
 [1 0 1 1] 1,3,4
 [0 1 1 0] 2,3
 [1 1 0 1] 1,2,4
 [0 0 0 1] 4
 [1 0 0 1] 1,4
 [0 0 1 0] 3

julia> Λ = [1 2 3 4]; [ rand.(Bernoulli.(Λ./(Λ.+1))) for i=1:6]
6-element Vector{BitMatrix}:
 [0 1 1 1] 2,3,4
 [1 0 1 1] 1,3,4
 [0 0 1 1] 3,4
 [1 0 1 0] 1,3
 [0 0 1 1] 3,4
 [0 0 0 1] 4
```

Generalizing suppose we have N numbers $\lambda_i \geq 0$. We will now imagine N unfair coins with probability of heads for the ith coin being $\frac{\lambda_i}{\lambda_i+1}$ and tails $\frac{1}{\lambda_i+1}$. Then taking products we see that $p_{\mathcal{I}} = \frac{\lambda_{\mathcal{I}}}{\prod_{i=1}^N (1+\lambda_i)}$, where $\lambda_{\mathcal{I}} = \prod_{i \in \mathcal{I}} \lambda_i$.

The reader could check that $1 = \sum_{\mathcal{I} \in \mathcal{P}} p_{\mathcal{I}}$ by expanding $\prod_{i=1}^N (1 + \lambda_i)$ into 2^N terms, we get $\sum_{\mathcal{I} \in \mathcal{P}} \lambda_{\mathcal{I}}$. Example: $(1 + \lambda_1)(1 + \lambda_2) = 1 + \lambda_1 + \lambda_2 + \lambda_1\lambda_2$.

We can form a mixture model from an orthogonal matrix Q as follows. Let $Y_{\mathcal{I}} = Q[:, \mathcal{I}]$, the columns of Q with indices in $\mathcal{I}$. Then if we sample from the projection matrix DPP $K_{\mathcal{I}} = Y_{\mathcal{I}} Y_{\mathcal{I}}^T$ with probability $p_{\mathcal{I}} = \frac{\lambda_{\mathcal{I}}}{\prod_{i=1}^N (1+\lambda_i)}$, we have created a mixture of 2^N projection DPPs.

To be clear, the mixture would have $Pr(\mathcal{J}) = \sum_{\substack{\mathcal{I} \in \mathcal{P} \\ |\mathcal{I}|=|\mathcal{J}|}} p_{\mathcal{I}} K_{\mathcal{I}}\binom{\mathcal{J}}{\mathcal{J}}$.

Exercise 2.12 What is the 2^N by 2^N Markov matrix for the mixture above?

2.4.2 Sampling from a Projection DPP

Consequences of Cauchy-Binet

Alan: This maybe goes in the Cauchy-Binet Section, perhaps already there

LEMMA A: If $K = YY^T$ with $Y \in \Re^{n,p}$ then Cauchy-Binet immediately tells us that

$$K\begin{pmatrix} i_1 \dots i_p \\ i_1 \dots i_p \end{pmatrix} = Y\begin{pmatrix} i_1 \dots i_p \\ 1 \dots p \end{pmatrix}^2.$$

LEMMA B: If $I_p = Y^T Y$ then Cauchy-Binet immediately tells us that

$$\det I_p = 1 = \sum_{i_1 < \ldots < i_p} Y\begin{pmatrix} i_1 \dots i_p \\ 1 \dots p \end{pmatrix}^2.$$

LEMMA C: Generalizing LEMMA A, if $K = YY^T$ with $Y \in \Re^{n,p}$ then for $k = 1, \ldots, p$

$$K\begin{pmatrix} i_1 \dots i_k \\ i_1 \dots i_k \end{pmatrix} = \sum_{1 \le j_1 < \ldots < j_k \le p} Y\begin{pmatrix} i_1 \dots i_k \\ j_1 \dots j_k \end{pmatrix}^2.$$

LEMMA D: (Orthogonal Invariance): If Y is replaced by $\tilde{Y} = YQ$, where Q is an orthogonal $p \times p$ matrix, then for $k = 1, \ldots, p$

$$K\begin{pmatrix} i_1 \dots i_k \\ i_1 \dots i_k \end{pmatrix} = \sum_{1 \le j_1 < \ldots < j_k \le p} Y\begin{pmatrix} i_1 \dots i_k \\ j_1 \dots j_k \end{pmatrix}^2 = \sum_{1 \le j_1 < \ldots < j_k \le p} \tilde{Y}\begin{pmatrix} i_1 \dots i_k \\ j_1 \dots j_k \end{pmatrix}^2.$$

Proof: This follows from LEMMA C and Cauchy-Binet, since $K = YY^T = \tilde{Y}\tilde{Y}^T$.

LEMMA E: (Normalization): Generalizing LEMMA B, if $I_p = Y^T Y$ then for $k = 1, \ldots, p$

$$\sum_{i_1 < \ldots < i_k} K\begin{pmatrix} i_1 \dots i_k \\ i_1 \dots i_k \end{pmatrix} = \sum_{\substack{1 \le i_1 < \ldots < i_k \le p \\ 1 \le j_1 < \ldots < j_k \le p}} Y\begin{pmatrix} i_1 \dots i_k \\ j_1 \dots j_k \end{pmatrix}^2 = \binom{p}{k}.$$

Proof: This follows from LEMMA C and the fact that K has p eigenvalues equal to 1.

Sampling from a Projection DPP

Lines 3 through 12 above sample from the projection DPP with $K = YY^T$. One naive way to sample from a rank p projection DPP is to generate all the $\binom{N}{p}$ determinants $K\left(\begin{smallmatrix} i_1 \dots i_p \\ i_1 \dots i_p \end{smallmatrix}\right)$ perhaps by using Lemma A. Lemma B confirms that these determinants are probabilities, that is they are non-negative numbers adding to 1.

This naive approach can be considerably improved upon for efficiency by sequentially generating

$i_1, \ldots, i_p$. In particular we can calculate

$$Pr(i_k \in \mathcal{I} | i_1, \ldots, i_{k-1} \in \mathcal{I}) = Pr(i_1, \ldots, i_k)/Pr(i_1, \ldots, i_{k-1}) = K\binom{i_1 \ldots i_k}{i_1 \ldots i_k} / K\binom{i_1 \ldots i_{k-1}}{i_1 \ldots i_{k-1}}.$$

Let us see this step by step. To start, we consider $j = 1$, which gives by LEMMA C and LEMMA E

$$Pr(i_1 \in \mathcal{I}) = K_{i_1 i_1} = \sum_{j=1}^{p} Y_{i_1,j}^2.$$

We recall that $\mathrm{tr}(K) = p$ corresponding to the p numbers in each $\mathcal{I}$, so if we sample from a density with

$$\frac{1}{p} K_{i_1 i_1} = \mathrm{mean}(Y_{i_1,1:p}^2),$$

we obtain our first entry in the sample. This may be thought of as being equivalent to choosing the p-element sample and taking one of the p elements uniformly at random.

At the next step, we consider $j = 2$, which gives by Lemma D

$$Pr(i_2|i_1) = \sum_{1 \le j < j_2 \le p} Y\binom{i_1, i_2}{j_1, j_2}^2 / \sum_{j=1}^{p} Y_{i_1,j}^2.$$

Inspired by numerical linear algebra, we will pick a judicious choice of Q to simplify the above expression, We will create $\tilde{Y}$ from Y by multiplying by a Householder reflector that zeros out row i_1 in columns 2 through p.

We illustrate with a 4x4 rank 3 projection, where $i_1 = 2$:

$$Y = \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix} \to \tilde{Y} = \begin{pmatrix} \times & \times & \times \\ \times & 0 & 0 \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}.$$

With this judicious choice it is easy to see that $Pr(i_2|i_1) = \sum_{1 \le j < j_2 \le p} \tilde{Y}\binom{i_1,i_2}{j_1,j_2}^2 / \sum_{j=1}^{p} \tilde{Y}_{i_1,j}^2$ which simplifies to

$$Pr(i_2|i_1) = \sum_{j=2}^{p} \tilde{Y}_{i_2,j}^2$$

thanks to the zeros introduced by the Householder transformation.

Every subsequent step is treated accordingly.

On line 9, We make use of the fact that `QR(v,Q)` computes a Householder matrix with first column `v`.

2.5 Correlation Functions, Level Density

The preceding sections of this chapter on Determinantal Point Processes took as random objects the subsets of $1 : N$, i.e., a collection of numbers chosen from $1, \ldots, N$. It is not required that the subset be a fixed length.

In this section we ask what happens if we generate a random $\mathcal{I}$ from a DPP and then a random point or a random subset of fixed length from $\mathcal{I}$.

The probabilities related to correlation functions, with the 1-point correlation function being known as the level density.

2.5.1 A uniform random "point" from a DPP

One can generate a random i by choosing i uniformly from $\mathcal{I}$. The way to do this, is pick $\mathcal{I}$ randomly from the DPP, and if it is not empty, choose $i \in \mathcal{I}$ uniformly with probability $1/|\mathcal{I}|$.

For example if $N = 3$, the probability for $\{1\}$ is

$$Pr\{1\} + \frac{1}{2}Pr\{1,2\} + \frac{1}{2}Pr\{1,3\} + \frac{1}{3}Pr\{1,2,3\}.$$

We point out by contrast that

$$K_{11} = Pr\{1\} + Pr\{1,2\} + Pr\{1,3\} + Pr\{1,2,3\},$$

and thus one should not confuse K_{ii} with the probability of i in general.

We point out in Exercise ?? that if K is a rank-r projection matrix then

$$K_{ii} = \Pr(i)/r.$$

Exercise 2.13 Show that K is a rank-r projection matrix,

$$K_{ii} = \Pr(i)/r.$$

Is this an "if and only if?" (I don't know the answer.)

Further show that if $K = YY^T$, where $Y^TY = I_r$, then $K_{ii} = \sum_{j=1}^{r} Y_{ij}^2$. In the continuous case, this formula is related to sums of squares of orthogonal polynomials and the Christofel-Darboux formula.

Exercise 2.14 Show that if K is not a projection matrix, then the probability is a mixture of those from projection matrices. Write out this mixture.

2.5.2 Random Subsets

As a generalization consider the r-point functions, which can be defined as the probability obtained by generating $\mathcal{I}$ from a DPP, and then a random subset uniformly from $\mathcal{I}$ if $|\mathcal{I}| \geq r$.

Exercise 2.15 Show that if K is a rank-r projection matrix, then

$$K\binom{\mathcal{I}}{\mathcal{I}}/\binom{r}{k}$$

is the k-point function if $\mathcal{I} = k$ assuming $k \leq r$.

Exercise 2.16 Show that if K is not a rank-r projection matrix, then the k-point function is a mixture.

Alan: Need to check if this terminology is used in the discrete case, and if the constants are standard

2.5.3 Level Densities, Correlation Functions

The examples of uniform points or subsets, is often presented with a different scale factor and is known as level densities or correlation functions. They are also known as factorial moment measures or simply moment measures.

For a random $\mathcal{I}$ from a DPP consider the Bernoulli random variable that is 1 if $i \in \mathcal{I}$. We see that the probability of this random variable is indeed K_{ii}. This function $\rho(i)$ is the level-density or one point correlation function. Notice that this is not a probability measure on $1 : N$, however if K is a projection, then $\rho(i)/r$ is a probability measure.

Generalizing, we can define the $k-point$ correlation function $\rho_k(i_1, i_2, \ldots, i_k)$ as the probability that

$$\{i_1, i_2, \ldots, i_k\} \subseteq \mathcal{I}.$$

Evidently this is $K\binom{i_1,i_2,\ldots,i_k}{i_1,i_2,\ldots,i_k}$. Again this is not a probability measure on sized k subsets, but if M is a rank-r projector, the Exercise ??? 1.15 shows that dividing by $\binom{r}{k}$ turns ρ_k into a probability measure.

It is important to realize that for non-projections, the uniformly chosen index or set does not have as nice an expression as the probability that a point or a subset has of being in a randomly chosen $\mathcal{I}$.

Exercise: Show that for any DPP $\sum_{|\mathcal{I}|=k}(K\binom{\mathcal{I}}{\mathcal{I}}) = E\binom{|\mathcal{I}|}{k}$
In particular, when $k = 1$ we have $\mathrm{tr}(K) = E(|\mathcal{I}|)$. In other words, the expected number of points is $\mathrm{tr}(K)$. As another special case, show that if K is a rank-r projection matrix, we have $\binom{r}{k}$ by arguing algebraically or simply noting that $|\mathcal{I}| = r$ for a projection DPP.

2.6 Going Continuous

In general, the term "process" or more specifically "point process" in mathematics refers to a set of points randomly chosen on, say, the real line, or some more general space. If we talk about the N eigenvalues of a random symmetric $N \times N$ matrix, then we have a known fixed number of points, namely N. If we talk about those eigenvalues that are in the interval $[0, 1]$, then now the number might vary from 0 to N.

We might like to ask for the probability the set of points includes $x_1, \ldots, x_k$. The continuous way to do that is to have a B_i some kind of ball of radius δ around x_i and allowing for slightly more generality than being on the line we can assume we have a volume measurement for the ball. We then would ask for

$$\lim_{\delta \to 0} \frac{Pr(\text{ one point appears in } B_i for i = 1 : k)}{\prod_{i=1}^{k} \text{Vol}(B_i)}.$$

Intuition on Continuous Definition of DPP

Many students begin their study of probability with discrete random variables whose probabilities add to 1. Then continuous random variables appear whereby one looks at the probability of the random variable (a point!) landing in a small interval and then doing the usual calculus thing of taking the limit of the probability divided by the size of the interval.

Turning to the context of DPPs, the discrete situation considers a set of indices $\mathcal{I}$ and a probability that the random set of points $\mathcal{J}$ includes $\mathcal{I}$.

Going continuous, we will, in a sense, "thicken" the $\mathcal{I}$ and take a limit. Specifically, we will replace the indices in $\mathcal{I}$ with a collection of intervals B_i. We will then generate our random points, and ask for the probability that each B_i is hit.

* We recommend looking at discretizations of K for some examples.

We recall Equation 2.21

$$E((f^T \zeta)_{\mathcal{I}}) = f^T c,$$

and consider the special case where a k is chosen from $0, 1, \ldots, N$ and $f(\mathcal{I}) = 0$ unless $\mathcal{I} = k$, then we have explicitly

$$E_{\mathcal{I}} \left(\sum_{\substack{\mathcal{J} \subseteq \mathcal{I} \\ |\mathcal{J}| = k}} f(\mathcal{J}) \right) = f(\mathcal{I}) c_{\mathcal{I}}.$$

The function that takes $\mathcal{I}$ to $c_{\mathcal{I}}$ is called a correlation function because it says how likely the elements of $\mathcal{I}$ might appear together, even if in a bigger superset.

For a determinantal point process we have explicitly that $c_{\mathcal{I}}$ is a determinant:

$$E_{\mathcal{I}}\,\Big(\sum_{\substack{\mathcal{J}\subseteq\mathcal{I}\\|\mathcal{J}|=k}} f(\mathcal{J})\Big) = \sum_{|\mathcal{I}|=k} f(\mathcal{I})K\binom{\mathcal{I}}{\mathcal{I}}.$$

Show that taking $f = 1$ reproduces the result of the previous exercise by arguing that every $\mathcal{I}$ is counted $\binom{|\mathcal{I}|}{k}$ times.

It is helpful to write this out explicitly when $N = 3$ for $k = 1$ and $k = 2$.

We have for $k = 1$,

$$\begin{array}{rcl}
\underline{f^T(\zeta p)} & & \underline{\sum_i f_i K_{ii}} \\
f_1(p_1 + p_{12} + p_{13} + p_{123}) & = & f_1 Pr(1 \in \mathcal{I}) \\
f_2(p_2 + p_{12} + p_{23} + p_{123}) & = & f_2 Pr(2 \in \mathcal{I}) \\
f_3(p_3 + p_{13} + p_{23} + p_{123}) & = & f_2 Pr(3 \in \mathcal{I})
\end{array}$$

$$\begin{array}{llll}
= & p_1 & (f_1 \quad\quad\quad) & \\
+ & p_2 & (\quad f_2 \quad) & \\
+ & p_3 & (\quad\quad\quad f_3) & \\
+ & p_{12} & (f_1 + f_2 \quad) & = \underline{p^T(\zeta^T f)} = \underline{E_{\mathcal{I}} \sum_{i\in\mathcal{I}} f_i} \\
+ & p_{13} & (f_1 + \quad f_3) & \\
+ & p_{23} & (\quad f_2 + f_3) & \\
+ & p_{123} & (f_1 + f_2 + f_3) &
\end{array}$$

It is helpful to also see $k = 2$:

$$\begin{array}{rcl}
\underline{f^T(\zeta p)} & & \underline{\sum_{i<j} f_{ij} K\binom{i,j}{i,j}} \\
f_{12}(p_{12} + p_{123}) & = & f_{12} Pr(\{1,2\} \subseteq \mathcal{I}) \\
f_{13}(p_{13} + p_{123}) & = & f_{13} Pr(\{1,3\} \subseteq \mathcal{I}) \\
f_{23}(p_{23} + p_{123}) & = & f_{23} Pr(\{2,3\} \subseteq \mathcal{I})
\end{array}$$

$$\begin{array}{llll}
= & p_{12} & (f_{12} \quad\quad\quad\quad) & \\
+ & p_{13} & (\quad f_{13} \quad) & = \underline{p^T(\zeta^T f)} = \underline{E_{\mathcal{I}} \sum_{\{i,j\}\subseteq\mathcal{I}} f_{ij}} \\
+ & p_{23} & (\quad\quad\quad\quad f_{23}) & \\
+ & p_{123} & (f_{12} + f_{13} + f_{23}) &
\end{array}$$

This can be written in unordered fashion:

$$E_{\mathcal{I}}\Big(\sum_{\substack{\{j_1,\ldots,j_k\}\in\mathcal{I}^k\\ j_i \text{ distinct}}} f(j_1,\ldots,j_k)\Big) = \sum_{\substack{\{i_1,\ldots,i_k\}\in N^k\\ i_i \text{ distinct}}} f(i_1,\ldots,i_k)K\binom{i_1,\ldots,i_k}{i_1,\ldots,i_k},$$

where f is a symmetric function of k variables.

The unordered version multiplies the "set version" on both sides by $k!$ This unordered version leads to the continuous version very nicely now.

$$E(\sum f(x_1, \ldots, x_k)) = \int_{\Re^k} f(x_1, \ldots, x_k) \det K(x_i, x_j)_{1 \le i,j \le k} dx$$

We can turn these formulas into a definition.

Definition: A discrete DPP is a probability measure on the power set of $1 : N$ such that for any function f,

$$E_{\mathcal{I}} \,(\sum_{\substack{\mathcal{J} \subseteq \mathcal{I} \\ |\mathcal{J}| = k}} f(\mathcal{J})) = \sum_{|\mathcal{I}| = k} f(\mathcal{I}) K \binom{\mathcal{I}}{\mathcal{I}}.$$

A continuous point process is a random discrete set of points $\mathcal{A}$. We can assume $\mathcal{A} \subset \Re$ or maybe C . One can get fancy and talk about DPP's on any locally compact Polish space, which is just the technicality that is required for a clean generalization, but we do not need this here.

C

Maybe see Guionnet, Zeitouni,etc

We can say that the process $\mathcal{A}$ has a kth correlation function $c_k(\mathcal{A})$ if

$$E_{\mathcal{A}} \sum_{\substack{x_1, \ldots, x_k \in \mathcal{A} \\ x_i \text{ distinct}}} f(x_1, \ldots, x_k) = \int_{R^k} f(x_1, \ldots, x_k) c_k(x_1, \ldots, x_k) dx_1, \ldots, x_k.$$

We say that the process $\mathcal{A}$ is a DPP if $c_k(x_1, \ldots, x_k) = K\binom{x_1, \ldots, x_k}{x_1, \ldots, x_k}$ for some kernel matrix K, i.e.,

$$E_{\mathcal{A}} \sum_{\substack{x_1, \ldots, x_k \in \mathcal{A} \\ x_i \text{ distinct}}} f(x_1, \ldots, x_k) = \int_{R^k} f(x_1, \ldots, x_k) K \binom{x_1, \ldots, x_k}{x_1, \ldots, x_k} dx_1, \ldots, x_k.$$

One might compare and contrast the discrete with the continuous definition (or find the abstraction that covers both by just choosing the appropriate measure). The continuous case is often written in unordered fashion, but that is a minor detail as one can always divide the continuous equation by a factor of $k!$ on both sides. It might take getting used to the notion that the random $\mathcal{A}$ very likely has more than k elements, and then on the left we sum over all the subsets of size k with distinct elements.

References: https://arxiv.org/pdf/1207.6083.pdf Determinantal point processes for machine learning. Alex Kulesza Ben Taska.

Alexei Borodin and Eric Rains. Eynard-Mehta theorem, Schur process, and their pfaffian analogs. Journal of Statistical Physics, 121:291–317, 2005. ISSN 0022-4715. 10.1007/s10955- 005-7583-z.

Notes: Fermion point process (orginal name?)

Chapter 3

The Tracy Widom Distribution

Because of it's importance, this chapter will focus on the $\beta = 2$ Tracy Widom distribution.

maybe also 1 and 4

The story begins with the joint eigenvalue densities which may be found in ???????.

maybe also 1 and 4, cor 11.2 has the GOE

The next step is to transform the joint eigenvalue densities using the Cauchy-Binet formula ??? reference, should we discuss in context of the gram identity? but beta=1 adn 4 are not gram i guess?

as well as the beta=1 and 4 Andrèief variant of CB, does this appear anywhere?

3.1 Joint Densities as determinants

We start with

$$\prod (x_i - x_j)^2 \prod w(x_i)dx$$

and our goal is to recognize that this is a Cauchy Binet situation.

We have that

$$c \int_{x_1 < \cdots < x_n \in I} \prod (x_i - x_j)^2 \prod w(x_i)dx = \left| \int_I \pi_i \pi_j w(x) dx \right|_{i,j=1,\ldots,n},$$

where c is chosen to make the integral on the left 1 when $I = R$

Proof

Let V be indexed by $I \times 0 : n$ (so it is continuous in the row index and discrete in the column index.) Let W be the continuous analog of a diagonal matrix with $W(x,x) = w(x)$ for $x \in I$.

Therefore $V^T W V$ is an ordinary finite $n+1 \times n+1$ matrix. The act of multiplying out the matrices places univariate integrals in the entries of this matrix so that

$(V^T W V)_{ij} = \int_I \pi_i \pi_j w(x) dx$, for $0 \leq i, j \leq n$.

On the other hand, the Cauchy-Binet theorem in finite form states that if A is an $m \times n$ matrix with $m \geq n$ and D is an $m \times m$ matrix, then

$$\det(A^T DA) = \sum_{1 \le i_1 \le \cdots \le i_n \le m} A \begin{pmatrix} i_1 & \cdots & i_n \\ 1 & \cdots & n \end{pmatrix}^2 d_{i_1} \ldots d_{i_n}.$$

Going continuous, we then see that the multi-index sum becomes an exact multivariate integral:

$$\det(V^T \mathrm{diag}(w) V) =$$

$$\int_{x_1 \le \cdots \le x_n \in I} V \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & n \end{pmatrix}^2 W \begin{pmatrix} x_1 & \cdots & x_n \\ x_1 & \cdots & x_n \end{pmatrix} = \int_{x_1 \le \cdots \le x_n \in I} \prod (x_i - x_j)^2 \prod w(x_i) dx$$

(See exercise 12.3 (change to label))

* Explain the replacement monomial with orth org polynomials * Include the f's for general expections * Consider the 1 minus approach * mention the factorial if you unorder

*** Read Forrester's history

*** should we include beta=1 and 4 if it's already in lots of other places? or is this not worth doing

*** what about other beta's , what do we really know? selberg integral????? (proibably not)

*** has anyone stated when continous cuachy binet is legal (the analysis rigor that i never really care about)

*** det(I-A'A) = det(I-AA') turns a finite integral into a Fredholm integral or vice versa

3.2 Painleve and Jimbo form (for $\beta = 2$)

What are these limiting cases? bulk,soft,hard = sin, airy, bessel and which of hermite, laguerre, jacobi has which of these things KEY POINT: this is basically orthogonal to hermite, laguerre, jacobi however, historically and probably technically bulk and soft are easiest derived from hermite or probably circular

Everyone has a Fredholm Determinant Many have Painleve, but not all known??? Many have hypergeometric of matrix argument, but not all known??

Note : $E^{(n)}_{\mathrm{RMT}}(k; J) = \mathbb{P}(k$ eigenvalues of $n \times n$ RMT is in $J)$

3.2.1 Finite n : Hermite, Laguerre, and Jacobi

Hermite

$$E^{(n)}_{\mathrm{GUE}}(0; J) = \det(I - K^{\mathrm{Herm}}_n|_J) \qquad \text{(Fredholm determinant)}$$

$$\begin{array}{c} E^{(n)}_{\mathrm{GUE}}(0; (s,\infty)) = \exp(-\int_s^\infty \sigma(x)dx) \\ \sigma_{xx}^2 = 4(\sigma - x\sigma_x)^2 - 4\sigma_x^2(\sigma_x + 2n) \end{array} \qquad \text{(Painleve IV)}$$

Laguerre

$$E^{(n)}_{\mathrm{LUE}}(0; J, \alpha) = \det(I - K^{\mathrm{Lag}}_{\alpha,n}|_J) \qquad \text{(Fredholm determinant)}$$

$$\begin{array}{c} E^{(n)}_{\mathrm{LUE}}(0; (0,s), \alpha) = \exp(-\int_0^s \frac{\sigma(x)}{x}dx), \\ (x\sigma_{xx})^2 = (\sigma - \sigma_x(x + 2\sigma_x - 2n - \alpha))^2 - 4\sigma_x^2(\sigma_x - n)(\sigma_x - n - \alpha) \end{array} \qquad \text{(Painleve V)}$$

$$E^{(n)}_{\mathrm{LUE}}(0; (0,s), \alpha) = \text{[Forrested 1994 JMP Eq (2.13a)]} \qquad \text{(Hypergeometric fct)}$$

$$E^{(n)}_{\mathrm{LUE}}(0; (s,\infty), \alpha) = \text{[Dumitriu thesis Thm 10.2.1]} \qquad \text{(Hypergeometric fct)}$$

Jacobi

$$E^{(n)}_{\mathrm{JUE}}(0; (0,s), a, b) = \det(I - K^{\mathrm{Jac}}_{a,b,n}|_{(0,s)}) \qquad \text{(Fredholm determinant)}$$

$$E^{(n)}_{\mathrm{JUE}}(0; (0,s), a, b) = \text{[Borodin \& Forrester Eq (3.16)]} \qquad \text{(Hypergeometric fct)}$$

3.2.2 Infinite n, scaling limits

Bulk scaling limit: Hermite, Laguerre, Jacobi

Definitions

$$E^{\text{bulk}}(0;J) = \lim_{n\to\infty} E^{(n)}_{\text{GUE}}(0; \frac{\pi}{\sqrt{2n}} J)$$
$$E^{\text{bulk}}(0;J) = \lim_{n\to\infty} E^{(n)}_{\text{LUE}}(0; (w + \frac{\sqrt{w}\pi}{\sqrt{2n}}) J, \alpha)$$
$$E^{\text{bulk}}(0;J) = \lim_{n\to\infty} E^{(n)}_{\text{JUE}}(0; \frac{\pi}{n} J, a, b)$$

Representations

$$E^{\text{bulk}}(0;J) = \det(I - K_{\sin}|_J) \qquad \text{(Fredholm determinant)}$$

$$\begin{aligned} E^{\text{bulk}}(0;(0,s)) &= \exp(-\textstyle\int_0^{\pi s} \frac{\sigma(x)}{x} dx) \\ (x\sigma_{xx})^2 &= 4(\sigma - x\sigma_x)(x\sigma_x - \sigma - \sigma_x^2) \end{aligned} \qquad \text{(Painleve V)}$$

Soft-edge scaling limit: Hermite, Laguerre

Definitions

$$E^{\text{soft}}(0;J) = \lim_{n\to\infty} E^{(n)}_{\text{GUE}}(0; \sqrt{2n} + 2^{-1/2} n^{-1/6} J)$$
$$E^{\text{soft}}(0;J) = \lim_{n\to\infty} E^{(n)}_{\text{LUE}}(0; 4n + 2(2n)^{1/3} J, \alpha)$$

Representations

$$E_2^{(\text{soft})}(0;(s,\infty)) = \det(I - K_{\text{Ai}}|_{(s,\infty)}) \qquad \text{(Fredholm determinant)}$$

$$\begin{aligned} E_2^{(\text{soft})}(0;(s,\infty)) &= \exp(-\textstyle\int_s^{\infty} \sigma(x) dx) \\ \sigma_{xx}^2 &= -4\sigma_x(\sigma - x\sigma_x) - 4\sigma_x^3 \end{aligned} \qquad \text{(Painleve II)}$$

$$\begin{aligned} E_2^{(\text{soft})}(0;(s,\infty)) &= \exp(-\textstyle\int_s^{\infty}\int_y^{\infty} q(t)^2 dt dy) \\ q_{xx} &= 2q^3 + xq \end{aligned} \qquad \text{(Painleve II, non-Jimbo form)}$$

Hard-edge scaling limit: Laguerre, Jacobi

Definitions

$$E^{\text{hard}}(0;(0,s),\alpha) = \lim_{n\to\infty} E^{(n)}_{\text{JUE}}(0; (0, \frac{s}{4n^2}), \alpha, b)$$
$$E^{\text{hard}}(0;(0,s),\alpha) = \lim_{n\to\infty} E^{(n)}_{\text{LUE}}(0; (0, \frac{s}{4n}), \alpha)$$

Representations

$$E^{\text{hard}}(0;(0,s),\alpha) = \det(I - K_\alpha^{\text{Bess}}|_{(0,s)}) \qquad \text{(Fredholm determinant)}$$

$$\begin{gathered} E^{\text{hard}}(0;(0,s),\alpha) = \exp(-\textstyle\int_0^s \frac{\sigma(x)}{x}dx), \\ (x\sigma_{xx})^2 = \alpha^2\sigma_x^2 - \sigma_x(\sigma - x\sigma_x)(4\sigma_x - 1) \end{gathered} \qquad \text{(Painleve III)}$$

$$E^{\text{hard}}(0;(0,s),\alpha) = e^{-s/4}{}_0F_1^{(1)}(\alpha; x_1, \cdots, x_\alpha)|_{x_i=s/4} \qquad \text{(Hypergeometric fct)}$$

	Fredholm	Painlevè	Hypergeometric
Finite Hermite	H	IV	-
Finite Laguerre	L	V	
Finite Jacobi	J	-	
Bulk	Sine	V	-
Soft Edge	Airy	II	-
Hard Edge	Bessel	III	${}_0F_1^{(1)}$

Bibliography

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

M. Marcus. Determinants of sums. *The College Mathematics Journal*, 21(2):130–135, 1990.