6.881 Project Final Report: Entropy of Determinantal Point Processes

Yuzhou Gu

1 Introduction

Let $\mathcal{K}$ denote the space of $n \times n$ matrices K with $0 \preceq K \preceq I$. For $K \in \mathcal{K}$ we define the determinantal point process (DPP) $X(K)$, which is a random variable on the power set of $[n]$, such that for every $S \subseteq [n]$, we have

$$\mathbb{P}(S \subseteq X(K)) = \det(K_S)$$

where K_S is the submatrix of K with row and column indices in S.

One fundamental property of a distribution is its Shannon entropy. For a random variable X on a discrete set $\mathcal{X}$, its Shannon entropy defined is

$$H(X) := -\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) \log \mathbb{P}(X = x).$$

(All logs in this report are of base e.)

Lyons [Lyo03] conjectured that the entropy of determinantal point processes is concave in K, i.e., for any $t \in [0, 1]$ and matrices $K_1, K_2 \in \mathcal{K}$, we have

$$H(X((1-t)K_1 + tK_2)) \geq (1-t)H(X(K_1)) + tH(X(K_2)).$$

In this project we study towards Lyons's conjecture. Our main results are

1. Entropy of cardinality of $X(K)$ is concave in K (Theorem 1).
2. Lyons' conjecture is true when $K_2 = 0$ (Corollary 6).
3. Lyons' conjecture is true when $K_1 - K_2$ is of rank one (Theorem 7).

2 Concavity of entropy of cardinality

Let $|X(K)|$ denote the cardinality of $X(K)$. Then $|X(K)|$ is a random variable supported on $\{0, \dots, n\}$. In this section we prove that $H(|X(K)|)$ is concave in K.

Theorem 1. *For $t \in [0, 1]$, $K_1, K_2 \in \mathcal{K}$, we have*

$$H(|X((1-t)K_1 + tK_2)|) \geq (1-t)H(|X(K_1)|) + tH(|X(K_2)|).$$

Proof. For $a \in [0, 1]^n$, define $Y(a)$ to be the sum of n independent Bernoulli random variables $\mathrm{Ber}(a_i)$. Let $\lambda(K)$ denote the vector $(\lambda_1(K), \dots, \lambda_n(K))$ where $\lambda_i(K)$ is the i-th largest eigenvalue of K. It is known that $|X(K)|$ has the same distribution as $Y(\lambda(K))$ (see e.g. Hough et al. [HKPV06]).

Therefore it is natural to apply the Shepp-Olkin conjecture (Shepp and Olkin [SO81], proved by Hillion and Johnson [HJ17]), which states that $H(Y(a))$ is concave in a. Note that this in particular implies that

$$H(Y(ta)) \geq tH(Y(a)) + (1-t)H(Y(0)) = tH(Y(a))$$

for $0 \leq t \leq 1$.

Lidskii [Lid50] proved that for two Hermitian matrices A and B, we have

$$\lambda(A+B) \in \lambda(A) + \mathrm{conv}(\sigma(\lambda(B)) : \sigma \in S_n),$$

where conv denotes convex hull, and $\sigma(v) = (v_{\sigma(1)}, \dots, v_{\sigma(n)})$.

Now write

$$\begin{aligned}\lambda((1-t)K_1 + tK_2) &= \lambda((1-t)K_1) + \sum_{\sigma \in S_n} c_\sigma \sigma(\lambda(tK_2)) \\ &= (1-t)\lambda(K_1) + t \sum_{\sigma \in S_n} c_\sigma \sigma(\lambda(K_2)).\end{aligned}$$

where $c_\sigma \geq 0$ and $\sum_\sigma c_\sigma = 1$.

Then

$$\begin{aligned}H(|X((1-t)K_1 + tK_2)|) &= H(Y(\lambda((1-t)K_1 + tK_2))) \\ &= H(Y((1-t)\lambda(K_1) + t \sum_{\sigma \in S_n} c_\sigma \sigma(\lambda(K_2)))) \\ &\geq H(Y((1-t)\lambda(K_1))) + \sum_{\sigma \in S_n} H(Y(tc_\sigma \sigma(\lambda(K_2)))) \\ &\geq (1-t)H(Y(\lambda(K_1))) + t \sum_{\sigma \in S_n} c_\sigma H(Y(\sigma(\lambda(K_2)))) \\ &= (1-t)H(Y(\lambda(K_1))) + tH(Y(\lambda(K_2))).\end{aligned}$$

□

3 Concavity of entropy under thinning

In this section we view the determinantal point process as a distribution over $\{0,1\}^n \subseteq \mathbb{Z}^n_{\geq 0}$. We consider the thinning operation of Rényi [Rén56] and prove that entropy is concave under thinning, a multivariate generalization of a result of Yu and Johnson [YJ09].

Definition 2. Let X be a random variable supported on $\mathbb{Z}^n_{\geq 0}$. Let $t \in [0,1]$. Then the t-thinning of X is a random variable $T_t X$ supported on $\mathbb{Z}^n_{\geq 0}$ such that

$$\mathbb{P}(T_t X = b | X = a) = \prod_{1 \leq i \leq n} B_t(a_i, b_i)$$

where

$$B_t(m,n) = t^m (1-t)^{n-m} \binom{n}{m}.$$

That is, if we consider X_i as the number of particles of type i, then t-thinning is the operation of independently retaining each particle with probability t.

Definition 3. A distribution X on $\mathbb{Z}_{\geq 0}$ is called ultra log-concave (ULC) if the sequence $(\log(i! \mathbb{P}(X = i)))_{i \in \mathbb{Z}_{\geq 0}}$ is concave.

Theorem 4. *Let X and Y be distributions on $\mathbb{Z}^n_{\geq 0}$ whose marginals are ULC. Let $t \in [0,1]$. Then*

$$H(T_{1-t}X + T_t Y) \geq (1-t)H(X) + tH(Y).$$

Our approach is similar to Yu and Johnson [YJ09]. For $m \in \mathbb{R}^n_{\geq 0}$, we define the multivariate Poisson distribution $\mathrm{Po}(m)$ as the product distribution $\mathrm{Po}(m_1) \times \cdots \times \mathrm{Po}(m_n)$. Let X be a distribution supported on $\mathbb{Z}^n_{\geq 0}$ with finite mean. (Note that distributions with ULC marginals have finite mean.) Consider the decomposition

$$H(X) = -D(X) - L(X)$$

where

$$D(X) := -D(X || \mathrm{Po}(\mathbb{E}X))$$

and

$$L(X) := \mathbb{E}_X \log \mathrm{Po}(X, \mathbb{E}X)$$

where

$$\mathrm{Po}(x, m) := \mathbb{P}(\mathrm{Po}(m) = x) = \prod_{1 \le i \le n} \frac{m_i^{x_i} \exp(-m_i)}{x_i!}.$$

The following Lemma is a multivariate generalization of Yu [Yu09].

Lemma 5. *For any distribution X on $\mathbb{Z}_{\ge 0}^n$ with finite mean and $t \in [0, 1]$, we have*

$$D(T_t X) \le t D(X).$$

Proof. Note that $T_t \mathrm{Po}(m) = \mathrm{Po}(tm)$.

Consider the operator S acting on distributions Y on $\mathbb{Z}_{\ge 0}$ with finite mean, defined as

$$\mathbb{P}(SY = k) = \frac{(k+1)\mathbb{P}(Y = k+1)}{\mathbb{E}Y}.$$

Now let S_i ($i \in [n]$) be an operator acting on distributions on $\mathbb{Z}_{\ge 0}^n$ with finite mean, by applying S to the i-th coordinate.

Let us compute the derivative of $D(T_t X)$.

$$\begin{aligned}
\frac{d}{dt} D(T_t X) &= \frac{d}{dt} \sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) \log \frac{\mathbb{P}(T_t X = x)}{\mathrm{Po}(x, t\mathbb{E}X)} \\
&= \sum_{x \in \mathbb{Z}_{\ge 0}^n} (\frac{d}{dt} \mathbb{P}(T_t X = x)) \log \frac{\mathbb{P}(T_t X = x)}{\mathrm{Po}(x, t\mathbb{E}X)} \\
&+ \sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) \frac{d}{dt} \log \frac{\mathbb{P}(T_t X = x)}{\mathrm{Po}(x, t\mathbb{E}X)}.
\end{aligned}$$

Now

$$\sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) \frac{d}{dt} \log \mathbb{P}(T_t X = x) = \sum_{x \in \mathbb{Z}_{\ge 0}^n} \frac{d}{dt} \mathbb{P}(T_t X = x) = 0.$$

$$\begin{aligned}
&\sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) \frac{d}{dt} \log \mathrm{Po}(x, t\mathbb{E}X) \\
&= \sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) \sum_{1 \le i \le n} \frac{d}{dt} \log \mathrm{Po}(x_i, t\mathbb{E}X_i) \\
&= \sum_{1 \le i \le n} \sum_{x \in \mathbb{Z}_{\ge 0}^n} \mathbb{P}(T_t X = x) (\frac{x_i}{t} - \mathbb{E}X_i) \\
&= 0.
\end{aligned}$$

For $i \in [n]$, write e_i for the standard basis vector whose i-th coordinate is 1 and all other coordinates are 0.

Because

$$\frac{d}{dt} B_t(m, n) = n(B_t(m-1, n-1) - B_t(m, n-1)),$$

we have

$$
\begin{aligned}
&\frac{d}{dt}\mathbb{P}(T_tX=x)\\
&=\frac{d}{dt}\sum_{y\in\mathbb{Z}_{\geq 0}^n}\prod_{1\leq i\leq n}B_t(x_i,y_i)\mathbb{P}(X=y)\\
&=\sum_{y\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(X=y)\sum_{1\leq i\leq n}y_i(B_t(x_i-1,y_i-1)-B_t(x_i,y_i-1))\prod_{j\neq i}B_t(x_j,y_j)\\
&=\sum_{1\leq i\leq n}\sum_{y\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(X=y+e_i)(y_i+1)(B_t(x_i-1,y_i)-B_t(x_i,y_i))\prod_{j\neq i}B_t(x_j,y_j)\\
&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)\sum_{y\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(S_iX=y)(B_t(x_i-1,y_i)-B_t(x_i,y_i))\prod_{j\neq i}B_t(x_j,y_j).
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{d}{dt}D(T_tX)&=\sum_{x\in\mathbb{Z}_{\geq 0}^n}(\frac{d}{dt}\mathbb{P}(T_tX=x))\log\frac{\mathbb{P}(T_tX=x)}{\text{Po}(x,t\mathbb{E}X)}\\
&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)\sum_{y\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(S_iX=y)\sum_{x\in\mathbb{Z}_{\geq 0}^n}(\log\frac{\mathbb{P}(T_tX=x)}{\text{Po}(x,t\mathbb{E}X)})\\
&\cdot(B_t(x_i-1,y_i)-B_t(x_i,y_i))\prod_{j\neq i}B_t(x_j,y_j)\\
&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)\sum_{y\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(S_iX=y)\sum_{x\in\mathbb{Z}_{\geq 0}^n}\prod_{1\leq j\leq n}B_t(x_j,y_j)\\
&\cdot(\log\frac{\mathbb{P}(T_tX=x+e_i)}{\text{Po}(x+e_i,t\mathbb{E}X)}-\log\frac{\mathbb{P}(T_tX=x)}{\text{Po}(x,t\mathbb{E}X)})\\
&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)\sum_{x\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(T_tS_iX=x)\\
&\cdot(\log\frac{\mathbb{P}(T_tX=x+e_i)}{\text{Po}(x+e_i,t\mathbb{E}X)}-\log\frac{\mathbb{P}(T_tX=x)}{\text{Po}(x,t\mathbb{E}X)}).
\end{aligned}
$$

Now

$$
\begin{aligned}
&\log\frac{\mathbb{P}(T_tX=x+e_i)}{\text{Po}(x+e_i,t\mathbb{E}X)}-\log\frac{\mathbb{P}(T_tX=x)}{\text{Po}(x,t\mathbb{E}X)}\\
&=\log(\frac{x_i+1}{t\mathbb{E}X_i}\cdot\frac{\mathbb{P}(T_tX=x+e_i)}{\mathbb{P}(T_tX=x)})\\
&=\log\frac{\mathbb{P}(S_iT_tX=x)}{\mathbb{P}(T_tX=x)}.
\end{aligned}
$$

Note that S_i and T_t commute. So

$$
\begin{aligned}
\frac{d}{dt}D(T_tX)&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)\sum_{x\in\mathbb{Z}_{\geq 0}^n}\mathbb{P}(T_tS_iX=x)\log\frac{\mathbb{P}(S_iT_tX=x)}{\mathbb{P}(T_tX=x)}\\
&=\sum_{1\leq i\leq n}(\mathbb{E}X_i)D(T_tS_iX||T_tX).
\end{aligned}
$$

By data processing inequality, $D(T_tS_iX||T_tX)$ is non-decreasing in $t\in[0,1]$. So $D(T_tX)$ is convex in $t\in[0,1]$. This finishes the proof. □

Proof of Theorem 4. By data processing inequality and Lemma 5, we have

$$D((1-t)X+tY) \le D((1-t)X) + D(tY) \le (1-t)D(X) + tD(Y)$$

for distributions X, Y on $\mathbb{Z}_{\ge 0}^n$ with finite mean.

Next we consider $L(X)$. We have

$$\begin{aligned} L(X) &= \mathbb{E}_X \log \mathrm{Po}(X; \mathbb{E}X) = \mathbb{E}_X \sum_{i\in[n]} \log \mathrm{Po}(X_i; \mathbb{E}X_i) \\ &= \sum_{1\le i\le n} \mathbb{E}_X \log \mathrm{Po}(X_i, \mathbb{E}X_i) = \sum_{1\le i\le n} L(X_i). \end{aligned}$$

By Yu and Johnson [YJ09], for ULC distributions X, Y on $\mathbb{Z}_{\ge 0}$, we have

$$L(T_{1-t}X + T_tY) \le (1-t)L(X) + tL(Y).$$

Therefore this holds also for distributions on $\mathbb{Z}_{\ge 0}^n$ with ULC marginals.

So for distributions X, Y on $\mathbb{Z}_{\ge 0}^n$ with ULC marginals, we have

$$\begin{aligned} H(T_{1-t}X + T_tY) &= -D(T_{1-t}X + T_tY) - L(T_{1-t}X + T_tY) \\ &\ge -(1-t)D(X) - tD(Y) - (1-t)L(X) - tL(Y) \\ &= (1-t)H(X) + tH(Y). \end{aligned}$$

□

As a corollary, we derive concavity of entropy of a determinantal point process when one endpoint is 0.

Corollary 6. *For $t \in [0,1]$, $K \in \mathcal{K}$, we have*

$$H(X(tK)) \ge tH(X(K)).$$

Proof. Note that $T_tX(K) = X(tK)$ because for every $S \subseteq [n]$, we have

$$\begin{aligned} \mathbb{P}(S \subseteq T_tX(K)) &= t^{|S|}\mathbb{P}(S \subseteq X(K)) \\ &= t^{|S|} \det K_S = \det(tK)_S = \mathbb{P}(S \subseteq X(tK)). \end{aligned}$$

Then the result follows from Theorem 4 by taking $Y = 0$. □

4 Concavity of entropy along a rank one direction

In this section we prove entropy is concave along a rank one direction.

Theorem 7. *For $t \in [0,1]$, $K_1, K_2 \in \mathcal{K}$, if $\operatorname{rank}(K_1 - K_2) = 1$, then*

$$H(X((1-t)K_1 + tK_2)) \ge (1-t)H(X(K_1)) + tH(X(K_2)).$$

Proof. Restating the result, we would like to prove that for any $K \in \mathcal{K}$ and rank one matrix ΔK satisfying $K + t\Delta K \in \mathcal{K}$ for $t \ge 0$ small enough, we have $\frac{d^2}{dt^2}|_{t=0} H(K + t\Delta K) \le 0$.

For $S \subseteq [n]$, denote $f_S(t) = \mathbb{P}[X(K + t\Delta K) = S]$. Then

$$\begin{aligned} \frac{d^2}{dt^2}|_{t=0} H(K+t\Delta K) &= -\sum_{S\subseteq[n]} \frac{d^2}{dt^2}|_{t=0}(f_S(t)\log f_S(t)) \\ &= -\sum_{S\subseteq[n]} \frac{d}{dt}|_{t=0}(f_S'(t)(1+\log f_S(t))) \\ &= -\sum_{S\subseteq[n]} (\frac{(f_S'(0))^2}{f_S(0)} + f_S''(0)(1+\log f_S(0))) \\ &= -\sum_{S\subseteq[n]} (\frac{(f_S'(0))^2}{f_S(0)} + f_S''(0)\log f_S(0)). \end{aligned}$$

Note $\frac{(f'_S(0))^2}{f_S(0)} \geq 0$. Let us consider the second term.

Note that $\det((K + t\Delta K)_S) = \sum_{T \supseteq S} f_T(t)$. By inclusion-exclusion, we have

$$f_S(t) = \sum_{T \supseteq S} (-1)^{|T|-|S|} \det((K + t\Delta K)_T).$$

Now

$$\frac{\partial}{\partial K_{ij}} \det(K) = \det(K)(K^{-1})_{ji},$$
$$\frac{\partial^2}{\partial K_{ij} \partial K_{kl}} \det(K) = \det(K)((K^{-1})_{ji}(K^{-1})_{lk} - (K^{-1})_{jk}(K^{-1})_{li}).$$

So

$$\begin{aligned} &\frac{d^2}{dt^2}|_{t=0} \det((K + t\Delta K)_T) \\ &= \sum_{i,j,k,l \in T} \Delta K_{ij} \Delta K_{kl} \frac{\partial^2}{\partial K_{ij} \partial K_{kl}} \det(K_T) \\ &= \det(K_T) \sum_{i,j,k,l \in T} \Delta K_{ij} \Delta K_{kl} ((K_T^{-1})_{ji}(K_T^{-1})_{lk} - (K_T^{-1})_{jk}(K_T^{-1})_{li}). \end{aligned}$$

Because ΔK is of rank one, we have $\Delta K_{ij} \Delta K_{kl} = \Delta K_{il} \Delta K_{kj}$. So

$$\begin{aligned} &\sum_{i,j,k,l \in T} \Delta K_{ij} \Delta K_{kl} (K_T^{-1})_{jk}(K_T^{-1})_{li} \\ &= \sum_{i,j,k,l \in T} \Delta K_{il} \Delta K_{kj} (K_T^{-1})_{jk}(K_T^{-1})_{li} \\ &= \sum_{i,j,k,l \in T} \Delta K_{ij} \Delta K_{kl} (K_T^{-1})_{ji}(K_T^{-1})_{lk}. \end{aligned}$$

Hence $\frac{d^2}{dt^2}|_{t=0} \det((K + t\Delta K)_T) = 0$, and $f''_S(0) = 0$. □

References

[HJ17] R. Hillion and O. Johnson. A proof of the Shepp–Olkin entropy concavity conjecture. *Bernoulli*, 23(4B):3638–3649, 2017.

[HKPV06] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probability surveys*, 3:206–229, 2006.

[Lid50] B. V. Lidskii. The proper values of the sum and product of symmetric matrices. *Dokl. Akad. Nauk SSSR*, 74:769–772, 1950.

[Lyo03] R. Lyons. Determinantal probability measures. *Publications Mathématiques de l'IHÉS*, 98:167–212, 2003.

[Rén56] A. Rényi. A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 1(2):519–527, 1956.

[SO81] L. A. Shepp and I. Olkin. Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to probability*, pages 201–206. Elsevier, 1981.

[YJ09] Y. Yu and O. Johnson. Concavity of entropy under thinning. In *2009 IEEE International Symposium on Information Theory*, pages 144–148. IEEE, 2009.

[Yu09] Y. Yu. Monotonic convergence in an information-theoretic law of small numbers. *IEEE transactions on information theory*, 55(12):5412–5422, 2009.