Markov Properties of Discrete Determinantal Point Processes

Kayvan Sadeghi
University College London

Alessandro Rinaldo
Carnegie Mellon University

Abstract

Determinantal point processes (DPPs) are probabilistic models for repulsion. When used to represent the occurrence of random subsets of a finite base set, DPPs allow to model global negative associations in a mathematically elegant and direct way. Discrete DPPs have become popular and computationally tractable models for solving several machine learning tasks that require the selection of diverse objects, and have been successfully applied in numerous real-life problems. Despite their popularity, the statistical properties of such models have not been adequately explored. In this note, we derive the Markov properties of discrete DPPs and show how they can be expressed using graphical models.

1 Introduction

Determinantal point processes (DPPs) are stochastic processes for repulsion that model in a mathematically elegant and general way negative association. Originally developed in quantum physics, DPPs arise naturally in other areas, such as combinatorics, random matrix theory, probability and algebra; see, e.g., Lyons (2003), Borcea et al. (2009), Borodin (2011), Hough et al. (2006) Johansson (2005) and references therein. In particular, DPPs on finite base sets model the emergence of random subsets comprised by objects with a higher degree of mutual diversity that would have been otherwise observed through independent selections. In detail, for a given finite set V and a positive definite matrix K with eigenvalues bounded by 1, a discrete DPP Y is a random subset of V such that, for any $A \subset V$,

$$\mathbb{P}\left(A \subseteq Y\right) = \det(K_A),$$

where K_A is the principal minor of K corresponding to the coordinates in A and $\det(K_\emptyset) = 1$.

In their seminal work, Kulesza and Taskar (2012) have demonstrated the properties, expressive power, and computational ease of discrete DPPs for machine learning tasks aimed at producing diverse subset selections, and have exemplified their use in several real-life problems. Discrete DPPs have now been successfully employed in numerous areas, ranging from document summarization, signal processing, neuroscience, image segmentation, spatial statistics, survey sampling, and wireless network; see, e.g., Brunel et al. (2017) and references therein for a list of applications of discrete DPPs. Despite their popularity, the formal statistical properties of such models and their statistical inference remain largely unexplored; see Brunel et al. (2017) and Urschel et al. (2017) for recent works on maximum likelihood estimation, sample complexity guarantees and efficient algorithms in these models.

The main goal of this paper is to characterize the independence properties of discrete DPPs, and to express them trough the formalism of graphical modeling. By representing DPPs over finite sets as random binary vectors, we are able to easily derive the Markov properties of the corresponding distributions, using both bidirected and undirected graphical models. In particular, we make the following contributions:

1. We show that the distributions of discrete DPPs are Markov with respect to bidirected graphical models with edges corresponding to the non-zero entries of K. Such models are curved exponential families (Drton and Richardson, 2008, see, e.g.) and their Markov properties as well as other statistical properties are well-studied; see Kauermann (1996); Roverato et al. (2013).

2. For a *conditional* discrete DPPs Y, which is a DPP over a fixed non-empty subset A of V conditionally on $A^c \subset Y$, we further show that the corresponding conditional distribution is Markov with respect to an undirected graph whose edges instead correspond to the non-zero entries of K^{-1}. In particular, such conditional models are linear

exponential families. This particular type of independence properties are known as *context-specific* independence models, since they only apply to a distinguished subset of conditional distributions.

3. We further explore conditions that guarantee faithfulness in either one of the above cases.

Interestingly, we have found that DPP graphical models act similarly to the Gaussian graphical models, whereby the kernel matrix K plays a role analogous to that of the covariance matrix. Indeed, both K and K^{-1} express Markov properties that can be easily elicited from the corresponding bidirected and undirected graphs. To the best of our knowledge, results of this form are novel. Previous work on this topic include the results by Tadić (2014), which, among the other things, relate the conditional distribution of a discrete DPPs Y given $C \subset Y$ to the Schur complement of $\bar{K}$ in K.

2 Preliminaries

We first provide some background on determinantal point process on finite sets and review basic concepts and some recent results from the theory of probabilistic independence models and graphical models.

2.1 Determinantal point processes

Let V be a finite set of cardinality n, which, without loss of generality, may be assumed to coincide with $\{1, \ldots, n\}$. A *(discrete) determinantal point process (DPP)* Y on V is random variables taking values in 2^V, the set of all subsets of V, such that

$$\mathbb{P}(A \subseteq Y) = \det(K_A), \quad A \subseteq V, A \neq \emptyset, \qquad (1)$$

where $K \in \mathbb{R}^{V \times V}$ is a positive definite matrix called the *marginal kernel* or the simply *kernel* of the process. Here, for any non empty subset A of V, K_A denotes the restriction of K to the entries $A \times A$, whereby we set $\det(K_\emptyset) = 1$. Similarly, for any non-empty subsets A and B of V, $K_{A,B}$ is the restriction of K to the row entries indexed by A and column entries indexed by B. By (1), all principal minors of K must be nonnegative and no larger than 1. It is also easy to show that the eigenvalues of K are bounded above by 1. In fact, these requirements are also sufficient: any matrix $K \in \mathbb{R}^{V \times V}$ such that $0 \preceq K \preceq I$, where I is the identity matrix of the appropriate size, defines a DPP (see, e.g., Theorem 2.3 in Kulesza and Taskar, 2012).

For any pair i and j in A, (1) implies that

$$\mathbb{P}(\{i,j\} \in Y) = \det(K_{\{i,j\}}) = K_{\{i\}}K_{\{j\}} - K_{\{i,j\}}^2,$$

so that $\mathbb{P}(\{i,j\} \in Y) \leq \mathbb{P}(i \in Y)\mathbb{P}(j \in Y)$.

2.1.1 Marginalization, Conditioning and Complements

The class of probability distributions of DPPs over V is closed under marginalization, conditioning and complements. The first property is immediate: if Y is a DPP over V with kernel K and $A \subset V$ is non-empty, then $Y_A := Y \cap A$ is a DPP over A with kernel K_A. As for the conditioning, Tadić (2014) has showed that for any non-empty $A \subset V$, the conditional distribution of $Y_{\bar{A}}$ given that $Y_A = 1_A$, where $\bar{A} = V \setminus A$, is that of a DPP over $\bar{A}$ with kernel given by $K^A := K_{\bar{A}} - K_{\bar{A},A}K_A^{-1}K_{A,\bar{A}}$, the *Schur complement* of K_A in K. In addition, the *complement* $\bar{Y}$ of Y given by $\bar{Y} = V \setminus Y$ is a DPP over V with kernel $I - K$ (see, e.g. Kulesza and Taskar, 2012). Thats is,

$$\mathbb{P}(A \subseteq \bar{Y}) = \det(I - K_A), \quad A \subseteq V. \qquad (2)$$

2.1.2 DPPs as Random Binary Vectors

We will equivalently express the DPP Y on V as the n-dimensional binary random random vector $X_V = X = (X_1, \ldots, X_n)$ taking values on $\{0,1\}^V$ such that $X_i = 1$ if and only if $i \in Y$. Letting, for any non-empty $A \subset V$, X_A be the restriction of X to the coordinates in A, the condition (1) can be re-written as

$$q_A := \mathbb{P}(X_A = 1_A) = \det(K_A), \quad A \subseteq V, A \neq \emptyset, \quad (3)$$

where 1_A is a vector with all entries equal to 1 and indexed by A and we set $q_\emptyset = 1$. Similarly, (2) becomes

$$\mathbb{P}(X_A = 0_A) = \det(I - K_A).$$

The parameters $\{q_A, A \subseteq V\}$ are easily seen to be marginal probabilities and, borrowing the terminology from Drton and Richardson (2008), we will refer to them as the *Möbius parameter*. The joint probabilities of a finite DPP can be expressed in closed form using the so-called *L-ensembles* (see Kulesza and Taskar 2012), assuming that the spectrum of K does not contain 1. In fact, they can also be obtained from the Möbius parameters using the inclusion-exclusion principle as follows:

$$P(X_A = 1, X_{V \setminus A} = 0) = \sum_{B \subseteq V : A \subseteq B} (-1)^{|B \setminus A|} q_B,$$

for all $A \subset V$. Notice that the above expression holds true even of the largest eigenvalue of K is 1.

As remarked above, the set of distributions over $\{0,1\}^V$ that correspond to DPPs X_V over V is closed under marginalization, conditioning on events of the form $X_A = 1_A$, for any non-empty $A \subset V$, and complements, in the sense that $1_V - X_V$ also represents

a DPP over V. The above operations can be combined to generated other DPPs. In particular, this shows that DPPs are closed under conditioning on $X_A = 0_A$, the $|A|$-dimensional vector of all 0's, and consequently, since conditioning is commutative and transitive, DPPs are closed under conditioning on any event of the form $X_A = x_A$, for any non-empty $A \subset V$ and arbitrary $x_A \in \{0, 1\}^A$.

2.2 Independence models

An *independence model* $\mathcal{J}$ over a finite set V is a set of triples $\langle X, Y \,|\, Z \rangle$ (called *independence statements*), where X, Y, and Z are disjoint subsets of V; Z may be empty, and $\langle \varnothing, Y \,|\, Z \rangle$ and $\langle X, \varnothing \,|\, Z \rangle$ are always included in $\mathcal{J}$. The independence statement $\langle X, Y \,|\, Z \rangle$ is read as "X is independent of Y given Z". Independence models may in general have a probabilistic interpretation, but not necessarily. Similarly, not all independence models can be easily represented by graphs. For further discussion on general independence models, see Studený (2005).

2.2.1 Probabilistic independence models

In order to define probabilistic independence models, consider a set V and a collection of random variables $\{X_v\}_{v \in V}$ with state spaces $\{\mathcal{X}_v\}_{v \in V}$ and joint distribution P. We let $X_A = \{X_v\}_{v \in A}$ for each non-empty subset A of V. For disjoint subsets A, B, and C of V we use the short notation $A \perp\!\!\!\perp B \,|\, C$ to denote that X_A is *conditionally independent of X_B given X_C* (Dawid, 1979; Lauritzen, 1996), i.e. that for any measurable $\Omega \subseteq \mathcal{X}_A$ and P-almost all x_B and x_C,

$$P(X_A \in \Omega \,|\, X_B = x_B, X_C = x_C) = P(X_A \in \Omega \,|\, X_C = x_C)$$

We can now induce an independence model $\mathcal{J}(P)$ by letting

$\langle A, B \,|\, C \rangle \in \mathcal{J}(P)$ if and only if $A \perp\!\!\!\perp B \,|\, C$ w.r.t. P.

If A, B, or C has only one member $\{i\}$, $\{j\}$, or $\{k\}$, for better readability, we write $i \perp\!\!\!\perp j \,|\, k$. We also write $A \perp\!\!\!\perp B$ when $C = \varnothing$, which denotes the *marginal independence* of A and B.

A probabilistic independence model $\mathcal{J}(P)$ over a set V is always a *semi-graphoid* (Pearl, 1988), i.e., it satisfies the four following properties for disjoint subsets A, B, C, and D of V, which, as we shall see, makes them useful for graphical modeling:

1. $A \perp\!\!\!\perp B \,|\, C$ if and only if $B \perp\!\!\!\perp A \,|\, C$ (*symmetry*);

2. if $A \perp\!\!\!\perp B \cup D \,|\, C$ then $A \perp\!\!\!\perp B \,|\, C$ and $A \perp\!\!\!\perp D \,|\, C$ (*decomposition*);

3. if $A \perp\!\!\!\perp B \cup D \,|\, C$ then $A \perp\!\!\!\perp B \,|\, C \cup D$ and $A \perp\!\!\!\perp D \,|\, C \cup B$ (*weak union*);

4. if $A \perp\!\!\!\perp B \,|\, C \cup D$ and $A \perp\!\!\!\perp D \,|\, C$ then $A \perp\!\!\!\perp B \cup D \,|\, C$ (*contraction*).

Notice that the reverse implication of contraction clearly holds by decomposition and weak union. A semi-graphoid for which the reverse implication of the weak union property holds is said to be a *graphoid*; that is, it also satisfies

5. if $A \perp\!\!\!\perp B \,|\, C \cup D$ and $A \perp\!\!\!\perp D \,|\, C \cup B$ then $A \perp\!\!\!\perp B \cup D \,|\, C$ (*intersection*).

Furthermore, a graphoid or semi-graphoid for which the reverse implication of the decomposition property holds is said to be *compositional*, that is, it also satisfies

6. if $A \perp\!\!\!\perp B \,|\, C$ and $A \perp\!\!\!\perp D \,|\, C$ then $A \perp\!\!\!\perp B \cup D \,|\, C$ (*composition*).

If, for example, P has strictly positive density, the induced probabilistic independence model is always a graphoid; see e.g. Proposition 3.1 in Lauritzen (1996). See also Peters (2015) for a necessary and sufficient condition for P in order for the intersection property to hold. If the distribution P is a regular multivariate Gaussian distribution, $\mathcal{J}(P)$ is a compositional graphoid; e.g. see Studený (2005). Probabilistic independence models with positive densities are not in general compositional; this only holds for special types of multivariate distributions such as, for example, Gaussian distributions and the symmetric binary distributions used in Wermuth et al. (2009).

Another important property that is not necessarily satisfied by probabilistic independence models, is *singleton-transitivity* (also called *weak transitivity* in Pearl (1988), where it is shown that for Gaussian and binary distributions P, $\mathcal{J}(P)$ always satisfies it). For i, j, and k, single elements in V, and $C \subset V$,

7. if $i \perp\!\!\!\perp j \,|\, C$ and $i \perp\!\!\!\perp j \,|\, C \cup \{k\}$ then $i \perp\!\!\!\perp k \,|\, C$ or $j \perp\!\!\!\perp k \,|\, C$ (*singleton-transitivity*).

In addition, we have the two following properties:

8. if $i \perp\!\!\!\perp j \,|\, C$ then $i \perp\!\!\!\perp j \,|\, C \cup \{k\}$ for every $k \in V \setminus \{i, j\}$ (*upward-stability*);

9. if $i \perp\!\!\!\perp j \,|\, C$ then $i \perp\!\!\!\perp j \,|\, C \setminus \{k\}$ for every $k \in V \setminus \{i, j\}$ (*downward-stability*).

2.2.2 Context-specific independence models

In some situations, conditional independence may only hold when some variables are held fixed, an instance known as *context-specific independence* (Boutilier et al., 1996). More precisely, let A, B, D, and C be four disjoint sets of variables. A and B are conditionally independent given D in context $X_C = x_c$ if, for any measurable $\Omega \subseteq \mathcal{X}_A$,

$$P(X_A \in \Omega \,|\, X_B = x_B, X_C = x_C, X_D = x_D)$$

is equal to

$$P(X_A \in \Omega \,|\, X_C = x_C, X_D = x_D),$$

for P-almost all x_B and x_D such that $P(X_B = x_B, X_C = x_C, X_D = x_D) > 0$. Below we will express this property by writing $A \perp\!\!\!\perp B \,|\, D \cup (C = x_C)$. When D is empty, one simply says that A and B are independent in context $C = x_C$.

Now denote by $\mathcal{J}_{\cdot \,|\, X_C = x_C}(P)$ the *context-specific independence model* induced by P, i.e., the set of all context-specific independence statements above. The following result on proving that certain axioms are satisfied by context-specific independence models is a key property of context-specific independence models. See also Corander et al. (2016), for a more general discussion on the topic (although the following results are not explicitly proven in it).

Notice that the conditioning sets C in the conditional independence statements appearing in the axioms for context-specific independence models $\mathcal{J}_{\cdot \,|\, X_{C'} = x_{C'}}$ could be of form $C' \cup C''$, where we condition on the context-specific $X_{C'} = x_{C'}$ and on the general C''. For example, the weak union property is defined as follows (other axioms are translated from the original definitions in a similar fashion):

- if $A \perp\!\!\!\perp B \cup D \,|\, (C' = x_{C'}) \cup C''$ then $A \perp\!\!\!\perp B \,|\, (C' = x_{C'}) \cup C'' \cup D$ and $A \perp\!\!\!\perp D \,|\, (C' = x_{C'}) \cup C'' \cup B$ (*weak union*).

Lemma 1. *A context-specific probabilistic independence model* $\mathcal{J}_{\cdot \,|\, X_C = x_c}(P)$ *satisfies the symmetry, decomposition, and weak union properties.*

Proof. The proof is similar to that in the general case (see, e.g., Dawid 1979), where the value of the variables in conditioning set C are set to a fixed x_C (and in the weak-union case the values of the variables in D are also set to a fix value x_D). □

However, the contraction property is not necessarily satisfied for context-specific independence models even in the case of binary random variables:

Example 1. *Suppose that* A, B, C, D *are four binary random variables. Let* $P(C = 0) = 0$ *(i.e., C is deterministic and equal to 1 almost surely). For any* $x \in \{0, 1\}^3$, *set* $P(x) = P(A = x_1, B = x_2, D = x_3)$. *Next, define the probabilities*

$$
\begin{aligned}
p_1 &= P(0,0,0), p_2 = P(0,0,1), \\
p_3 &= P(0,1,0), p_4 = P(1,0,0), \\
p_5 &= P(0,1,1), p_6 = P(1,0,1), \\
p_7 &= P(1,1,0), p_8 = P(1,1,1).
\end{aligned}
$$

Now let $p_2 = p_5 = p_6 = p_8 = 1/8$ *and* $p_1 + p_3 = p_4 + p_7 = 1/4$, *but let* p_1, p_3, p_4, p_7 *be all distinct. It is now easy to check that* $A \perp\!\!\!\perp B \,|\, D \cup C = (1,1)$ *and* $A \perp\!\!\!\perp D \,|\, C = 1$, *but* $A \not\!\perp\!\!\!\perp B \,|\, C = 1$.

2.3 Graphical models

Graphical models (see, e.g., Lauritzen 1996) are statistical models expressing conditional independence clauses among a collection of random variables $X_V = (X_v, v \in V)$ indexed by a finite set V. A graphical model is determined by a graph $G = (V, E)$ over the indexing set V, and the edge set E (which may include edges of undirected, directed or bidirected type) encodes conditional independence relations among the variables, or *Markov properties*.

Using the formulation in (3), it is natural to define graphical models with node sets $X_1, \ldots, X_n$ for DPPs.

2.3.1 Undirected and bidirected graphical models

For non-empty subsets A and B of V and a (possibly empty) subset S of V, with A, B and S disjoint, we use the notation $A \perp\!\!\!\perp B \,|\, S$ as shorthand for the conditional independence relation $X_A \perp\!\!\!\perp X_B \,|\, X_S$, thus identifying random variables with their labels. When $S = \emptyset$ the independence relation is intended as marginal independence.

For an undirected graph G, where all edges are undirected ——, the *(global) Markov property* expresses that $A \perp\!\!\!\perp B \,|\, S$ when every path between A and B has a vertex in S or, in other words, S *separates* A from B in G. The *pairwise Markov property* expresses that if two nodes i and j are non-adjacent, i.e. there is no edge between them, then $i \perp\!\!\!\perp j \,|\, V \setminus \{i, j\}$. It is known that these two conditions are equivalent when the intersection property is satisfied; see Pearl (1988); Lauritzen (1996).

For a *directed acyclic graph* G, where all edges are directed ⟶ and there are no directed cycles, the (global) Markov property expresses that $A \perp\!\!\!\perp B \,|\, S$ when A and B are *d-separated* (Pearl, 1988) by S;

we refrain from discussing the notion of d-separation here and refer to Lauritzen (1996) for details.

We shall be interested in graphical models given by a *bidirected graph* G where all edges are bidirected $\leftrightarrow$. For such graphs the (global) Markov property (Cox and Wermuth, 1993; Kauermann, 1996) expresses that $A \perp\!\!\!\perp B \mid S$ when every path between A and B has a vertex outside $S \cup A \cup B$, i.e. $V \setminus (A \cup B \cup S)$ separates A from B. Note the obvious duality between this and the Markov property for undirected graphs. The *pairwise Markov property* expresses that if i and j are non-adjacent then $i \perp\!\!\!\perp j$. It is known that these two conditions are equivalent for distributions that satisfy the composition property; see Sadeghi and Lauritzen (2014).

For example, in the undirected graph of Figure 1(a), the pairwise Markov property implies that $i \perp\!\!\!\perp k \mid \{j, l\}$ and the global Markov property implies that $\{i, l\} \perp\!\!\!\perp k \mid j$, whereas in the bidirected graph of Figure 1(b), the pairwise Markov property implies that $i \perp\!\!\!\perp k$ and the global Markov property implies that $\{i, l\} \perp\!\!\!\perp k$.
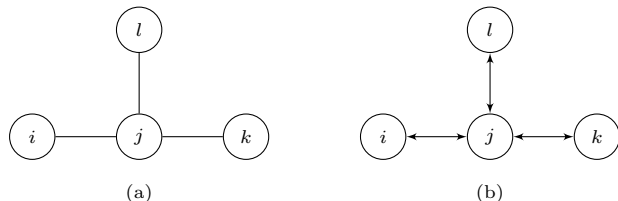

Figure 1: (a) An undirected graph. (b) a bidirected graph.

Another Markov property for bidirected graphs is the so-called *connected set Markov property* of Drton and Richardson (2008), which is equivalent to the global Markov property. Lemma 1 in Drton and Richardson (2008) states that for discrete random variables taking values in the set $\mathcal{I}$, the connected set Markov property is satisfied if and only if for every disconnected set $D \subseteq V$ (i.e. for D inducing a disconnected subgraph) such that $D = C_1 \cup C_2 \cup \cdots \cup C_r$, where $C_m, 1 \le m \le r$, are inclusion maximal connected sets, it holds that

$$
\begin{aligned}
P(X_D = i_D) = &P(X_{C_1} = i_{C_1}) \\
&\times P(X_{C_2} = i_{C_2}) \cdots P(X_{C_r} = i_{C_r}),
\end{aligned} \tag{4}
$$

for every $i \in \mathcal{I}$.

If, for P and undirected G, $A \perp_u B \mid C \iff A \perp\!\!\!\perp B \mid C$ then we say that P and G are *faithful*; Similarly, if, for P and bidirected G, $A \perp_b B \mid C \iff A \perp\!\!\!\perp B \mid C$ then P and G are *faithful*. Hence, faithfulness implies Markovness, but not the other way around.

For a given probability distribution P, we define the *skeleton* of P, denoted by $\text{sk}(P)$, to be the UG with

node set V such that nodes i and j are not adjacent if and only if there is *some* subset C of V so that $i \perp\!\!\!\perp j \mid C$. Thus, if P is Markov with respect to an undirected graph G then $\text{sk}(P)$ would be a subgraph of G (by considering all remaining nodes to be in C); whereas if P is Markov with respect to a bidirected graph G then $\text{sk}(P)$ is a subgraph of $\text{sk}(G)$ (by considering $C = \varnothing$).

In general a graph $G(P)$ is induced by P with skeleton $\text{sk}(P)$. For UGs, let $G_u(P) = \text{sk}(P)$, whereas for BGs, let $G_b(P)$ be $\text{sk}(P)$ with all edges being bidirected. We shall need the following results from Sadeghi (2017):

Lemma 2. *Probability distribution P and $G_u(P)$ are faithful if and only if P satisfies intersection, singleton-transitivity, and upward-stability.*

Lemma 3. *Probability distribution P and $G_b(P)$ are faithful if and only if P satisfies composition, singleton-transitivity, and downward-stability.*

2.3.2 Gaussian graphical models

Gaussian graphical models exemplify the properties describe above. Let P be a regular multivariate Gaussian distribution in $\mathbb{R}^V$ with covariance matrix Σ. It then holds that $\mathcal{J}(P)$ is a compositional graphoid. This follows from the fact that for such a distribution

$$
A \perp\!\!\!\perp B \mid C \iff (\Sigma_{ABC})^{-1}_{A,B} = 0, \tag{5}
$$

where $(\Sigma_{ABC})^{-1}$ is the *concentration matrix* of the distribution of $X_{A \cup B \cup C}$; for detailed proofs, see Studený (2005).

Define the undirected graph G_u as follows: there is no edge between nodes i and j if $\Sigma^{-1}_{ij} = 0$; and similarly define the bidirected graph G_b as follows: no edge between i and j if $\Sigma_{ij} = 0$. Hence, the pairwise Markov properties are satisfied with respect to G_u and G_b, and by the discussions above P is Markov to these graphs.

3 DPPs and conditional independence

In this section, we first focus on the context-specific conditional independence models where the conditioning set X_C is equal to the vector 1_C, but eventually we provide results for general independence as well. First, we prove the following linear algebraic result:

Lemma 4. *Let K be a symmetric positive semidefinite matrix K with row and column vector V. For disjoint, non-empty subsets A, B and C of V, the following statements are equivalent:*

    1. $\det(K_{ABC}) \det(K_C) = \det(K_{AC}) \det(K_{BC});$

    2. $(K_{ABC})^{-1}_{A,B} = 0;$

3. $K_{A,B} = K_{A,C}K_C^{-1}K_{C,B}$.

Simlarly, $\det(K_{AB}) = \det(K_A)\det(K_B)$ *if and only if* $(K_{AB})_{A,B} = (K_{AB})_{A,B}^{-1} = 0$.

Proof. Let K_{ABC}^C be the Schur complement of K_C in K_{ABC}. By the properties of the Schur complement, $(K_{ABC})_{AB}^{-1} = (K_{ABC}^C)^{-1}$. Thus, $(K_{ABC})_{A,B}^{-1} = 0$ if and only if $(K_{ABC}^C)_{A,B}^{-1} = 0$. Since K_{ABC}^C is symmetric and its entries are indexed by $A \cup B$, we see that, if $(K_{ABC}^C)_{A,B}^{-1} = 0$, then $(K_{ABC}^C)^{-1}$ must be block diagonal. In turn, this is equivalent to $(K_{ABC}^C)_{A,B} = 0$. By Theorem 15 in Tadić (2014), this is then equivalent to the events $\{X_A = 1_A\}$ and $\{X_B = 1_B\}$ being conditionally independent given $\{X_C = 1_C\}$, or, by definition, to the identity

$$\frac{\det(K_{ABC})}{\det(K_C)} = \frac{\det(K_{AC})}{\det(K_C)}\frac{\det(K_{BC})}{\det(K_C)},$$

which simplifies to $\det(K_{ABC})\det(K_C) = \det(K_{AC})\det(K_{BC})$. Thus, *1* and *2* are equivalent.

To prove that *2* is equivalent to *3* notice that, since $K_{ABC}^C = K_{AB} - K_{AB,C}K_C^{-1}K_{C,AB}$, we have that $(K_{ABC}^C)_{A,B} = K_{A,B} - K_{A,C}K_C^{-1}K_{C,B}$. Thus $(K_{ABC}^C)_{A,B}^{-1} = 0$ is equivalent to $K_{A,B} = K_{A,C}K_C^{-1}K_{C,B}$.

The final claim follows from Corollary 7 in Tadić (2014) and the fact that $(K_{AB})_{A,B} = 0$ if and only if $(K_{AB})_{A,B}^{-1} = 0$. $\square$

Condition 2. in the previous result is analogous to the right clause in the equivalence (5), with the kernel K playing the role of the covariance matrix Σ. This would suggest that, for a discrete DPP, condition 2. is equivalent to X_A being conditionally independent to X_B given X_C. This is in fact true provided that $X_C = 1_C$, as we demonstrate next.

The proof of the following result uses the same idea as the proof of Theorem 1 in Drton and Richardson (2008).

Theorem 1. *For a DPP with kernel K and any triplet A, B and C of non-empty disjoint subsets of V, $(K_{ABC}^{-1})_{A,B} = 0$ if and only if $X_A \perp\!\!\!\perp X_B \,|\, X_C = 1_C$.*

Proof. By Lemma 4, $(K_{ABC}^{-1})_{A,B} = 0$ is equivalent to $\det(K_{ABC})\det(K_C) = \det(K_{AC})\det(K_{BC})$. The proof of Lemma 4, in turn shows that this is equivalent to $P(X_A = 1_A, X_B = 1_B \,|\, X_C = 1_C) = P(X_A = 1_A \,|\, X_C = 1_C)P(X_B = 1_B \,|\, X_C = 1_C)$.

The last identity will provide the basis for an inductive argument that will show that, if $(K_{ABC}^{-1})_{A,B} = 0$,

then $P(X_A = i_A, X_B = i_B \,|\, X_C = 1_C) = P(X_A = i_A \,|\, X_C = 1_C)P(X_B = i_B \,|\, X_C = 1_C)$, for arbitrary $i_A \in \{0,1\}^A$ and $i_B \in \{0,1\}^B$. This is precisely the conditional probability clause $X_A \perp\!\!\!\perp X_B \,|\, X_C = 1_C$. The induction will be on the number of 0's in $(i_v)_{v \in A \cup B}$, with the base case – where there are no 0's – is proven above.

Thus, suppose that the result is true for less than r 0's, and assume that there are r 0's in $(i_v)_{v \in A \cup B}$. Suppose also that $i_v = 0$. Without loss of generality assume that $v \in A$. We have that

$$P(X_A = i_A, X_B = i_B \,|\, X_C = 1_C)$$
$$= P(X_{A\setminus\{v\}} = i_{A\setminus\{v\}}, X_B = i_B \,|\, X_C = 1_C)$$
$$-P(X_{A\setminus\{v\}} = i_{A\setminus\{v\}}, X_B = i_B, X_v = 1 \,|\, X_C = 1_C)$$
$$= (P(X_{A\setminus\{v\}} = i_{A\setminus\{v\}} \,|\, X_C = 1_C)$$
$$-P(X_{A\setminus\{v\}} = i_{A\setminus\{v\}}, X_v = 1 \,|\, X_C = 1_C))$$
$$P(X_B = i_B \,|\, X_C = 1_C)$$
$$= P(X_A = i_A \,|\, X_C = 1_C)P(X_B = i_B \,|\, X_C = 1_C),$$

where the second equality follows from the induction hypothesis. Thus, $(K_{ABC}^{-1})_{A,B} = 0$ implies that $X_A \perp\!\!\!\perp X_B \,|\, X_C = 1_C$.

To show the reverse implication, assume that $X_A \perp\!\!\!\perp X_B \,|\, X_C = 1_C$. Then, the events $X_A = 1_A$ and $X_B = 1_B$ are conditionally independent given $X_C = 1_C$, which is equivalent to $\det(K_{ABC})\det(K_C) = \det(K_{AC})\det(K_{BC})$. The result now follows from Lemma 4. $\square$

As an immediate corollary, we obtain the following context-specific pairwise Markov property.

Corollary 1. *For DPPs, $X_i \perp\!\!\!\perp X_j \,|\, X_R = 1_R$, where $R = V \setminus \{i,j\}$, if $(K^{-1})_{i,j} = 0$.*

Using the fact that the complement of a DPP with kernel K is a DPP with kernel $I - K$, we see that that the claims in both Theorem 1 and Corollary 1 hold true if the conditioning events are $X_C = 0_C$ and $X_R = 0_R$, respectively, provided K is replaced by $I - K$.

Next, by using very similar and in fact simpler arguments as in the proof of Theorem 1 for the case of $C = \emptyset$, we arrive at important result characterizing marginal independence in discrete DPPs based on the zero entries in K. See also Theorem 8 in Tadić (2014).

Theorem 2. *For DPPs, $X_A \perp\!\!\!\perp X_B$ if and only if $K_{A,B} = 0$. In particular, $X_i \perp\!\!\!\perp X_j$ if and only if $k_{ij} = 0$.*

Interestingly, since the correlation between any two variables X_i and X_j is

$$\frac{-K_{\{i,j\}}^2}{\sqrt{K_{\{i\}}(1 - K_{\{i\}})K_{\{j\}}(1 - K_{\{j\}})}},$$

Theorem 2 further implies that

Corollary 2. *For a DPP, $X_i \perp\!\!\!\perp X_j$ if and only if X_i and X_j are uncorrelated.*

Finally, we show that the context-specific conditional independence statements in Theorem 1 fulfill all the desiderata of a probabilistic model, much like the Gaussian graphical models.

Proposition 1. *For a DPP P, the context-specific independence model $\mathcal{J}_{\cdot \mid X_C = 1_C}(P)$ satisfies intersection, composition, and singleton-transitivity.*

Proof. Because of Theorem 1 and Equation (5), the proof is the same as the proof for the independence model induced by Gaussian distribution satisfying these properties. $\square$

Notice that because the complement of a DPP is a DPP, the above result also holds for the context-specific independence model $\mathcal{J}_{\cdot \mid X_C = 0_C}(P)$.

In fact, except intersection, the other two properties are also satisfied for the general DPP induced independence models (In fact singleton-transitivity is always satisfied by binary distributions; see Example 3.1.7 of Drton et al. (2009); but, here we provide a short proof for this special case):

Proposition 2. *For a DPP P, the independence model $\mathcal{J}(P)$ satisfies composition and singleton-transitivity.*

Proof. In order to prove composition, first notice that since DPPs are closed under marginalization, we only need to focus on $X_{A \cup B \cup C \cup D}$. Also, since DPPs are closed under conditioning on any binary vector, we only need to check that $A \perp\!\!\!\perp B$ and $A \perp\!\!\!\perp D$ imply $A \perp\!\!\!\perp B \cup D$. This follows from Theorem 2.

To prove singleton-transitivity, with the same argument as for composition, we only need to prove that $i \perp\!\!\!\perp j$ and $i \perp\!\!\!\perp j \mid l$ imply $i \perp\!\!\!\perp l$ or $j \perp\!\!\!\perp l$. We have, by Theorem 2, that $k_{ij} = 0$. If $l = 1$ then, by Lemma 4 and the inverse of a 3×3 matrix, $k_{ij} k_{ll} = k_{il} k_{jl}$. Therefore, either $k_{il} = 0$ or $k_{jl} = 0$, which implies the result. If $l = 0$ then the result follows from the fact that the complement of a DPP is a DPP. $\square$

However, notice that, for a DPP P, the independence model $\mathcal{J}(P)$ does not necessarily satisfy downward-stability. Examples for this are plentiful; for example, let

$$K = \begin{pmatrix} 0.11 & 0.04 & -0.10 \\ 0.04 & 0.29 & -0.22 \\ -010 & -0.22 & 0.54 \end{pmatrix}$$

This implies that $X_1 \not\perp\!\!\!\perp X_2$, but $K_{1,2}^{-1} \approx 0$, which implies that $X_1 \perp\!\!\!\perp X_2 \mid X_3$.

4 DPPs and Graphical Models

We will now use the results from the previous section to show how discrete DPPs are Markov to bidirected graphs corresponding to the non-zero entries of K and that context-specific discrete DPPs (conditioning on events of the form $X_C = 1_C$ or $X_C = 0_C$) are Markov with respect to the undirected graph corresponding to the non-zero entries of K^{-1}.

4.1 DPPs and Bidirected Graphical Models

Given a DPP P with kernel K, define $G_b(P)$ to be the bidirected graph with node set V where there is no edge between nodes i and j if and only if $K_{ij} = 0$ and vice versa.

By Theorem 2, we immediately have that

Proposition 3. *A DPP P is pairwise Markov to $G_b(P)$.*

In fact, we can obtain a stronger result.

Proposition 4. *A DPP P is Markov to $G_b(P)$.*

Proof. The proof is a direct corollary of Lemma 5 below since (4) and consequently the connected set Markov property is satisfied.

Lemma 5. *Let P be a DPP with kernel K. For every subgraph D of $G_b(P)$, it holds that $q_D = q_{C_1} q_{C_2} \cdots q_{C_r}$, where $C_1, \ldots, C_r$ are the maximal connected components of D.*

The proof of the previous claim follows from the fact that K_D is a block diagonal matrix with blocks $C_1, \ldots, C_r$, and hence, $\det(K_D) = \det(C_1) \ldots \det(C_r)$. $\square$

This global Markov property can also be proven using the pairwise Markov property (Proposition 3) and the composition property (Proposition 2).

Finally, making use of Lemma 3, we have the following result regarding faithfulness:

Proposition 5. *A DPP P is faithful to $G_b(P)$ if and only if $\mathcal{J}(P)$ satisfies downward-stability.*

As the example at the end of the previous section shows, downward-stability does not always hold, and, although DPPs are negatively associated, faithfulness does not immediately follow from Markovness. This result should be contrasted with the conditions for faithfulness in MTP_2 distributions, which are instead positively associated; see Section 6 in Fallat et al. (2017).

4.2 DPPs and Undirected Graphical Models

Given a DPP P with kernel K, define $G_u(P)$ to be the undirected graph on V where there is no edge between nodes i and j if and only if $K_{ij}^{-1} = 0$. By Corollary 1, we have the following:

Proposition 6. *For a DPP P, the context-specific independence model $\mathcal{J}_{\cdot \mid X_C = 1_C}(P)$ is pairwise Markov to $G_u(P)$.*

Like in the unconditional case, we caen easily strangthen the previous resul to obtain global Markov properties.

Proposition 7. *For a DPP P, the context-specific independence model $\mathcal{J}_{\cdot \mid X_C = 1_C}(P)$ is Markov to $G_u(P)$.*

Proof. The proof follows from the fact that $\mathcal{J}_{\cdot \mid X_C = 1_C}(P)$ satisfies symmetry and decomposition (Lemma 1) as well as the intersection property (Proposition 1). $\square$

By the use of Lemma 2, we also have the following:

Proposition 8. *For a DPP P, the context-specific independence model $\mathcal{J}_{\cdot \mid X_C = 1_C}(P)$ is faithful to $G_u(P)$ if and only if $\mathcal{J}(P)$ satisfies upward-stability.*

In a similar fashion to downward-stability, it is easy to find examples of DPPs whose induced context-specific independence model does not satisfy upward-stability.

5 Summary

We have shown that the bidirected DPP graphical model behaves similarly to the bidirected Gaussian graphical model. To us, this is noteworthy, as the only other known binary graphical model with the desirable properties of the Gaussian graphical model is the very restrictive class of symmetric binary models (Wermuth et al., 2009), where the marginal probabilities of 0 and 1 are equal. We have also illustrated that, for undirected graphs, it is the context-specific induced independence model by DPP given the vector of 1's (or 0's), which acts similarly to the independence models induced by the Gaussian distribution. This provides a nice example for the use of context-specific independence models as opposed to the general ones.

Although DPPs are negatively associated (Lyons, 2003), the faithfulness property in the bidirected graph case is not satisfied. Hence, DPPs are not (in some sense) dual to positively associated MTP$_2$ distributions (Karlin and Rinott, 1980) (which are always faithful to their corresponding undirected graphs (Fallat et al., 2017)). It would be interesting to find sufficient (and even necessary) conditions on the kernel K

so that downward-stability is satisfied. Such a result would imply that the restricted set of kernels based on these conditions generates DPPs that are faithful to their corresponding bidirected graph. Similarly, it would be interesting to find conditions on K so that the upward-stability is satisfied. This implies faithfulness of the context-specific independence model and the corresponding undirected graph.

In this note we only discuss undirected and bidirected graphs associated to DPPs, but the fact that context-specific independence models induced by discrete DPPs have essentially all the independence properties of the Gaussian distribution implies that graphical models arising from other types of graphs – such as directed acyclic graph or mixed graph (see Lauritzen and Sadeghi, 2018) – can be proposed for DPPs. The latter could cover the case where there are latent variables in a DPP graphical model.

Acknowledgements

References

Borcea, J., Brändén, P., and Liggett, T. (2009). Negative dependence and the geometry of polynomails. *Journal of the American Mathematical Society*, 22(2):521–567.

Borodin, A. (2011). *The Oxford handbook of random matrix theory*, chapter Determinantal point processes, pages 231–249. Oxford University Press.

Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. (1996). Context-specific independence in bayesian networks. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, pages 115–123, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Brunel, V. E., Moitra, A., Rigollet, P., and Urschel, J. (2017). Maximum likelihood estimation of determinantal point processes. *arXiv:1701.06501*.

Corander, J., Hyttinen, A., Kontinen, J., Pensar, J., and Väänänen, J. (2016). A logical approach to context-specific independence. In *Proceedings of the 23rd International Workshop on Logic, Language, Information, and Computation - Volume 9803*, pages 165–182, New York, NY, USA. Springer-Verlag New York, Inc.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, 8(3):204–218; 247–277.

Dawid, A. P. (1979). Conditional independence in

statistical theory (with discussion). *Journal of the Royal Statistical Society Series B*, 41(1):1–31.

Drton, M. and Richardson, T. S. (2008). Binary models for maginal independence. *Journal of the Royal Statistical Society Series B*, 41(2):287–309.

Drton, M., Sturmfels, B., and Sullivant, S. (2009). *Lectures on Algebraic Statistics*, volume 39 of *Oberwolfach Seminars*. Springer.

Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N., and Zwiernik, P. (2017). Total positivity in Markov structures. *Annals of Statistics.*, 45(3).

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. (2006). Determinantal processes and independence. *Probab. Surveys*, 3:206–229.

Johansson, K. (2005). Random matrices and determinantal processes. In *Mathematical Statistical Physics, Session LXXXIII: Lecture Notes of the Les Houches Summer School 2005*, pages 1–56.

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498.

Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scandinavian Journal of Statistics*, 23(1):105–116.

Kulesza, A. and Taskar, B. (2012). *Determinantal Point Processes for Machine Learning.* Now Publishers Inc., Hanover, MA, USA.

Lauritzen, S. L. (1996). *Graphical Models.* Clarendon Press, Oxford, United Kingdom.

Lauritzen, S. L. and Sadeghi, K. (2018). Unifying Markov properties for graphical models. *Annals of Statistics*, 46(5).

Lyons, R. (2003). Determinantal probability measures. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 98(1):167–212.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems : networks of plausible inference.* Morgan Kaufmann Publishers, San Mateo, CA, USA.

Peters, J. (2015). On the intersection property of conditional independence and its application to causal discovery. *Journal of Causal Inference*, 3:97–108.

Roverato, A., Lupparelli, M., and La Rocca, L. (2013). Log-mean linear models for binary data. *Biometrika*, 100:485–494.

Sadeghi, K. (2017). Faithfulness of probability distributions and graphs. *Journal of Machine Learning Reserch*, 18(148).

Sadeghi, K. and Lauritzen, S. L. (2014). Markov properties for mixed graphs. *Bernoulli,*, 20(2):676–696.

Studený, M. (2005). *Probabilistic Conditional Independence Structures.* Springer-Verlag, London, United Kingdom.

Tadić, T. (2014). Graphical structure of conditional independencies in determinantal point processes. *arXiv:1406.5577.*

Urschel, J., Brunel, V. E., Moitra, A., and Rigollet, P. (2017). Learning determinantal point processes with moments and cycles. *arXiv:1703.00539.*

Wermuth, N., Marchetti, G. M., and Cox, D. R. (2009). Triangular systems for symmetric binary variables. *Electronic Journal of Statistics*, 3:932–955.