Exchangeable random hypergraphs

By Danna Zhang and Peter McCullagh

Department of Statistics, University of Chicago

Abstract: A hypergraph is a generalization of a graph in which an edge may contain more than two vertices. Hereditary hypergraphs are particularly important because they arise in mathematics as the class of monotone subsets, in statistics as the class of factorial models, in topology as simplicial complexes and in algebra as the free distributive lattice. An exchangeable random hypergraph consists of a projective system and a probability distribution that is consistent with the projections in the system. The Indian buffet process is one instance. Three projective systems are defined. A specific sub-class of Poisson-generated hereditary hypergraphs is studied.

Keywords: Hypergraph, Projective system, Hereditary hypergraph, Exchangeability, Kolmogorov consistency, Probability distribution

1 Graphs and hypergraphs

We begin with the general definition of a hypergraph with vertex set V and its power set $\mathcal{P}(V)$.

Hypergraph: A function $G : \mathcal{P}(V) \to \mathcal{R}^+$, which assigns to each edge $e \subset V$ a weight $G(e) \geq 0$, is called a hypergraph, or a weighted hypergraph. See also Berge (1973).

Multi-hypergraph: A hypergraph taking non-negative integer values only is called a multi-hypergraph, and $G(e)$ is the multiplicity of edge e (Darling and Norris (2005)).

Boolean hypergraph: A function $G : \mathcal{P}(V) \to \{0, 1\}$ is called an ordinary hypergraph, or a Boolean hypergraph: in this setting the set of hyperedges $E \subset \mathcal{P}(V)$ need not include all singletons.

Hereditary hypergraph: A hypergraph $G : \mathcal{P}(V) \to \{0, 1\}$ is called hereditary if G is monotone, i.e., $G(e) = 1$ and $e' \subset e$ imply $G(e') = 1$.

Every hypergraph consists of some vertices V_1, some unordered pairs V_2, some unordered triples V_3 and so on. It can be seen that G is a set function but it may not be additive for disjoint sets and $G(\emptyset)$ is not necessarily equal to zero, so it is not a measure.

An undirected graph (Harary (1969)) is evidently a special case of a hypergraph, which consists of V_1 and V_2 only.

2 Projective systems

2.1 Projective systems and projective sub-systems

Let $\mathcal{H}_n$ be the set of hypergraphs with vertex set $[n]$. Three projective systems are defined on hypergraphs as follows.

Projective system I (PS I): For $m \leq n$, the selection map $\varphi : [m] \to [n]$, which is one-to-one, acts element-wise on the power sets $\varphi(e) = \{\varphi x : x \in e\}$ for $e \subset [m]$, and on hypergraphs G by composition or restriction: $\mathcal{P}([m]) \xrightarrow{\varphi} \mathcal{P}([n]) \xrightarrow{G} \mathcal{R}^+$. If $m = n$, φ is a permutation. The hypergraph $G\varphi$ induced by restriction has vertex set $[m]$, and edge weights $G(\varphi(e))$ for $e \subset [m]$. In essence, $\varphi^* : G \mapsto G\varphi$ is a mapping $\mathcal{H}_n \to \mathcal{H}_m$ on hypergraphs by *deletion of all edges* that are not subsets of $[m]$. If G is monotone, Boolean, or integer-valued, so also is $G\varphi$. We say the family $\mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1, \ldots\}$ of hypergraphs together with these permutation and restriction operators determines projective system I.

Projective systems IIa, IIb: The insertion map $\varphi : [m] \to [n]$ also acts on the power sets in the reverse direction $\mathcal{P}([n]) \to \mathcal{P}([m])$ by *deletion of vertices* $\varphi^*(e) = e \cap [m]$ for $e \subset [n]$. The general one-to-one selection map acts in a similar way by deletion and re-labelling $\varphi^*(e) = \varphi^{-1}(e \cap \varphi[m])$ for $e \subset [n]$. The induced graph is defined in one of two ways, either by maximization or summation over the inverse image $\varphi^{*-1}(e) \subset \mathcal{P}([n])$:

$$(\varphi^* G)(e) = \begin{cases} \max\{G(e') : \varphi^*(e') = e\}, & \text{PS IIa} \\ \sum_{\varphi^*(e')=e} G(e'), & \text{PS IIb} \end{cases} .$$

Then the family $\mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1, \ldots\}$ of hypergraphs with these maps is a projective system IIa or IIb, which means for each $m \leq n$ and one-to-one map $\varphi : [m] \to [n]$ there corresponds to a map $\varphi^* : \mathcal{H}_n \to \mathcal{H}_m$ such that $\mathcal{H}$ is closed.

In all three cases, the mapping $* : \varphi \mapsto \varphi^*$ preserves composition, but with reversal of the order, i.e., $(\psi\varphi)^* = \varphi^*\psi^*$ for arbitrary one-to-one composable ψ and φ.

Consider a sub-family $\mathcal{H}'_n \subset \mathcal{H}_n$ of hypergraphs, a subset of the hypergraphs defined by some rule with the vertex set $[n]$. We say the family $\mathcal{H}' = \{\mathcal{H}'_0, \mathcal{H}'_1, \ldots\}$ is a projective

sub-system of $\mathcal{H}$ if for every one-to-one map $\varphi : [m] \to [n]$, the induced map $\varphi^* : \mathcal{H}_n \to \mathcal{H}_m$ that determines $\mathcal{H}$ also sends $\mathcal{H}'_n$ to $\mathcal{H}'_m$, i.e., $\mathcal{H}'$ is closed under the maps φ^* that define $\mathcal{H}$ as a projective system, which is depicted below:

$$\begin{array}{ccccccc} \cdots & [m] & \xrightarrow{\varphi} & [n] & \xrightarrow{\psi} & [r] & \cdots \\ \cdots & \mathcal{H}_m & \xleftarrow{\varphi^*} & \mathcal{H}_n & \xleftarrow{\psi^*} & \mathcal{H}_r & \cdots \\ \cdots & \mathcal{H}'_m & \xleftarrow{\varphi^*} & \mathcal{H}'_n & \xleftarrow{\psi^*} & \mathcal{H}'_r & \cdots \end{array}$$

2.2 Random hypergraphs

A random hypergraph with vertex set $[n]$ is a probability distribution P_n on $\mathcal{H}_n$. Consider a sequence $\{P_n\}_{n\geq 0}$ corresponding to the collection $\mathcal{H} = \{\mathcal{H}_n\}_{n\geq 0}$, where $\mathcal{H}$ is one of the projective systems defined above. For some applications, two conditions on $\{P_n\}_{n\geq 0}$ are natural.

1. Exchangeability. A random hypergraph $G \sim P_n$ is finitely exchangeable if for every permutation $\sigma : [n] \to [n]$, the associated map $\sigma^* : \mathcal{H}_n \to \mathcal{H}_n$ in $\mathcal{H}$ is such that $G \sim P_n$ implies $\sigma^* G \sim P_n$.

2. Kolmogorov consistency. In order to generate an infinite hypergraph, the randomness at stage $n-1$ must be consistent with the randomness at stage n. That is to say, for every one-to-one map $\varphi : [m] \to [n]$, the associated map $\varphi^* : \mathcal{H}_n \to \mathcal{H}_m$ in $\mathcal{H}$ is such that $G \sim P_n$ implies $\varphi^* G \sim P_m$.

2.3 Examples

Erdős-Rényi random graphs

The Erdős-Rényi model which puts an edge between each pair of vertices with equal probability π_n, independently for all edges, was first introduced by Erdős and Rényi (1959). Kolmogorov consistency implies $\pi_n \equiv \pi$ for all n. As a result, the Erdős-Rényi random graphs are Boolean and exchangeable. If we delete a vertex, all the edges associated with that vertex are deleted. Thus, an Erdős-Rényi random graph is infinitely exchangeable and consistent with respect to PS I.

Exchangeable random partitions

For $n \geq 1$, a partition of the finite set $[n]$ is a a collection $B = \{b_1, \ldots\}$ of disjoint non-empty subsets, called blocks. The blocks are unordered and correspond to hyperedges of a Boolean hypergraph. Let $\mathcal{E}_n \subset \mathcal{H}_n$ be the set of partitions of $[n]$. Deletion of element n from the set $[n]$ determines a map $\mathcal{E}_n \to \mathcal{E}_{n-1}$ by element-wise deletion from the blocks, making the sets $\{\mathcal{E}_1, \mathcal{E}_2, \ldots\}$ into a projective sub-system in PS IIa or PS IIb, but not in PS I. In this case, all infinitely exchangeable random partitions are consistent with respect to PS IIa and PS IIb. Kingman's paintbox construction (Kingman (1978)) and the Chinese restaurant process (Aldous (1985)) are both generative models for a random partition.

An equivalent definition of a partition B of $[n]$ is an equivalence relation or Boolean function $B : [n] \times [n] \to \{0, 1\}$ that is reflexive, symmetric and transitive; or a symmetric Boolean matrix such that $B_{ij} = 1$ if i, j belong to the same block. This matrix defines a graph for which there is an edge between i and j if $B_{ij} = 1$. In this sense, element-wise deletion maps also make the sets $\{\mathcal{E}_1, \mathcal{E}_2, \ldots\}$ into a projective sub-system in PS I.

Random feature allocation

A multi-hypergraph $G : \mathcal{P}([n]) \to \mathcal{Z}^+$, where $\mathcal{Z}^+$ is the set of non-negative integers, consisting of certain possibly overlapping non-empty subsets may be generated by a feature map $F : \mathcal{F} \to \mathcal{P}([n])$ in which $\mathcal{F}$ is the set of feature labels, associating with each feature $e \in \mathcal{F}$ a subset $F(e) \subset [n]$ (Griffiths and Ghahramani (2005), Thibaux and Jordan (2007), Doshi et al. (2009), Broderick et al. (2013a,b)). The inverse image map ignoring feature labels $\#F^{-1} : \mathcal{P}([n]) \to \mathcal{Z}^+$ is a multi-hypergraph consisting of non-empty subsets of $[n]$. The Indian buffet process (IBP) (Griffiths and Ghahramani (2005)) is an example of an exchangeable random multi-hypergraph which is consistent with respect to PS IIb.

3 Hereditary hypergraphs

3.1 Monotone subsets

A collection of subsets $E \subset \mathcal{P}([n])$ is called *monotone* if $e \in E$ implies $\mathcal{P}(e) \subset E$ (Korshunov (1981), Wiedemann (1991)). In other words, E is monotone if $E = \bigcup_{e \in E} \mathcal{P}(e)$, i.e.,

E contains every subset of every element. Consider for example $n = 4$, and

$$E = \{\{1,2\},\{2,3\},\{3\}\},$$
$$E' = \bigcup_{e \in E} \mathcal{P}(e) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}\}.$$

We see that E' is monotone because it contains every subset of every element, but E is not. Note that E' does not contain the singleton $\{4\}$. It can be seen that the empty subset $\emptyset = 0$, the non-empty subset $\{\emptyset\} = 1$ and $\mathcal{P}([n])$ are all monotone. Moreover, for any collection $E \subset \mathcal{P}([n])$, the expanded collection $E' = \bigcup_{e \in E} \mathcal{P}(e)$ is monotone. Let $E \subset \mathcal{P}([n])$ be monotone. An element $e \in E$ is called *maximal* if there does not exist any other element $e' \in E$ such that $e \subset e'$. The set of maximal elements $e_1^*, e_2^*, \ldots$ is called the generating class of E, which is written as $\mathrm{Gen}(E) = \{e_1^*, e_2^*, \ldots\} \subset E$, and $E = \langle e_1^*, e_2^*, \ldots \rangle$. Note that $\langle \cdot \rangle = \emptyset$ is empty while $\langle \emptyset \rangle = \{\emptyset\}$ is non-empty. It can be easily seen that the monotone subset E' has two maximal elements $\{1,2\}$, $\{2,3\}$. Then we can specify the subset by its generating class as $E' = \langle \{1,2\}, \{2,3\} \rangle$ for simplicity.

A hereditary hypergraph (Berge (1973)) is a graph-theoretic construct equivalent to a monotone subset. A hypergraph $G = (V, E)$ with E consisting of a collection of subsets of V is called hereditary if it contains every subset of each hyperedge, which is also called an abstract simplicial complex in algebraic topology (Lee (2000)). As the above example shows, there may be vertices that do not occur in any hyperedge, which makes it awkward to distinguish graphically the hypergraph E' with vertex set [4] from the corresponding hypergraph with vertex set [3].

3.2 Factorial models

The set of factorial models is a class of subspaces arising in experimental design, analysis of variance and log-linear models for contingency tables. Some notable work includes Wilkinson and Rogers (1973), Nelder (1977), Dawid (1977, 1988), Tjur (1984), Tjur et al. (1991), Darroch et al. (1980), Speed (1987), Speed and Bailey (1987) and Bailey (1991, 1996). In the context of factorial models, $[n]$ is the set of factors, $\mathcal{P}([n])$ is the set of interactions, and a factorial model is a monotone subset of interactions. Monotonicity is equivalent to the marginality condition that each compound term (interaction) is accompanied by all of its subsets or marginal interactions in the sense of Nelder (1977), and

the subspaces automatically have this property. Thus, the set of monotone subsets of $[n]$, the set of hereditary hypergraphs with vertex set $[n]$ and the set of factorial models $\mathcal{F}_n$ with n factors are equivalent to each other. As a special case, the factorial model $\emptyset \equiv 0$ corresponds to the hypergraph $G = (V, \emptyset)$, i.e., $G(e) = 0$ for every $e \in \mathcal{P}(V)$. Meanwhile, the factorial model $\{\emptyset\} \equiv 1$ corresponds to the hypergraph $G = (V, \{\emptyset\})$, i.e., $G(e) = 0$ for every $e \in \mathcal{P}(V) \backslash \{\emptyset\}$ and $G(\emptyset) = 1$. In the literature on log-linear models such as Haberman (1970) and Darroch et al. (1980), these models are also called hierarchical. Here we list the generating classes associated with the factorial models $\mathcal{F}_n \subset \mathcal{H}_n$ with the number of factors $n \leq 2$.

$$
\begin{aligned}
&\mathcal{F}_0 : \emptyset, \quad \langle \emptyset \rangle; \\
&\mathcal{F}_1 : \emptyset, \quad \langle \emptyset \rangle, \quad \langle \{1\} \rangle; \\
&\mathcal{F}_2 : \emptyset, \quad \langle \emptyset \rangle, \quad \langle \{1\} \rangle, \quad \langle \{2\} \rangle, \quad \langle \{1\}, \{2\} \rangle, \quad \langle \{1, 2\} \rangle.
\end{aligned}
$$

The set of hereditary hypergraphs and the set of factorial models with the natural partial order of subset inclusion are equal to the free distributive lattice on n generators. The number of hereditary hypergraphs with n vertices is given by the Dedekind numbers. These numbers grow rapidly, and are known only for $0 \leq n \leq 8$; as reported by Wiedemann (1991), they are 2, 3, 6, 20, 168, 7581, 7828354, 2414682040998, 56130437228687557907788. Asymptotic approximations for large n are given by Korshunov (1981). Actually, the numbers above count the number of free distributive lattices in which the lattice operations are joins and meets of finite sets of elements, including the empty set. If empty joins and empty meets are disallowed, the resulting free distributive lattices have two fewer elements. The inclusion of the empty set or the zero vector space as a factorial model means that these numbers are one more than the values reported by Darroch et al. (1980).

In the application to linear models, the factors are usually denoted by capital letters $A, B, C, \ldots$. If the design has N units (e.g. objects, plots), a factor can be interpreted as a function from the sample to the factor levels $A : [N] \to \{\text{levels of factor } A\}$. More specifically, $A(i)$ is the level of factor A for the i-th sampling unit and $(AB)(i)$ is the ordered pair of levels for the two factors A, B. With each sampling unit labelled, the factor function A induces a partition of $[N]$. Each block of the partition is a labelled factor level or the collection of units with that level, and the number of blocks is either the number of levels of the factor or the number of levels that occur in the design. From another point of view,

each monotone subset represents a vector subspace in $\mathbb{R}^N$. Explicitly, $\emptyset = 0$ is the zero vector space, $\langle\emptyset\rangle = 1$ is the one-dimensional constant space, $\langle\{A\}\rangle$ represents the space of functions depending only on the value of A, and $\langle\{A\},\{B\}\rangle$ represents the subspace spanned by additive functions $A + B$, and so on. Therefore, each symbol represents both the factor as a function on the units and the associated vector subspace. In the compact notation of Wilkinson and Rogers (1973), we can represent factorial models in another way. For three factors, among all of the 20 factorial models, there are 3 of each of the five asymmetric types A, $A + B$, AB, $AB + C$, $AB + BC$ together with 0, 1, $A + B + C$, $AB+BC+AC$ and ABC. As can be seen, $AB+C$ is a compact notation for the monotone subset $\langle\{A, B\},\{C\}\rangle$ specified by its generating class.

3.3 Projective systems and sub-systems

The set of hereditary hypergraphs $\mathcal{F}_n$ is a sub-family of Boolean hypergraphs and is closed under PS I and PS IIa. Certain sub-families $\mathcal{F}'_n \subset \mathcal{F}_n$ that also form a projective sub-system in PS I and PS IIa with the same deletion operator as $\mathcal{F}_n$ are listed in the following.

- The set of hereditary hypergraphs $\mathcal{F}_n^{(k)}$ each of whose hyperedges is restricted to have degree no more than k. Thus a hereditary hypergraph $G = ([n], E)$ belongs to $\mathcal{F}_n^{(k)}$ if $e \in E$ implies $|e| \leq k$.

- The set of planar graphs $\mathcal{F}_n^{(2)}$ with or without the empty set in the edge set.

- The collection of hereditary hypergraphs for which each singleton in $[n]$ is an edge. In other words, it corresponds to a collection of factorial models including all the main effects.

- The collection of hereditary hypergraphs for which there exists an edge for each singleton in $[n]$ and an edge for each pair of vertices. For the corresponding factorial models, all the main effects and two-factor interactions are included.

- The set of hereditary hypergraphs with several disconnected complete sub-hypergraphs, i.e., the generating class is a partition of some subset of $[n]$. Specifically, if the maximal subsets of the hyperedge set are exhaustive, then its generating class is just a partition of $[n]$.

- The set of clique hypergraphs of planar graphs $G \in \mathcal{F}_n^{(2)}$ with or without the empty set in the hyperedge set. Darroch et al. (1980) define this type of factorial model as graphical models. For example, $\langle\{1,2\},\{2,3\},\{1,3\}\rangle$ is a planar graph model which is not graphical but its cliques determines the graphical model $\langle\{1,2,3\}\rangle$.

- The decomposable models in the sense of Haberman (1974) which is a proper subset of the graphical models. Andersen (1974) gives an example which is graphical but not decomposable, i.e. $\langle\{1,2\},\{2,3\},\{3,4\},\{1,4\}\rangle$.

4 Poisson-generated hereditary hypergraphs

4.1 Poisson-generated hypergraphs

For each $n \geq 0$, let Λ_n be a mean measure on $\mathcal{P}([n])$. Let X be a Poisson process with mean measure Λ_n associating with each hyperedge $a \in \mathcal{P}([n])$ a Poisson random variable $X_a \sim \text{Pois}(\lambda_n(a))$ with mean $\lambda_n(a) = \Lambda_n(\{a\})$, which need not be finite. Thus, X is a multi-hypergraph which assigns a multiplicity to each hyperedge, and $X_a = 0$ with probability $e^{-\lambda_n(a)}$ independently for each $a \in \mathcal{P}([n])$. Ignoring multiplicities induces a random Boolean hypergraph whose hyperedge set A is the collection of hyperedges with $X_a > 0$. Notice that A may not be monotone, but it has a least monotone cover $M \supset A$ consisting of all the subsets of all elements in A. As a result, it induces a hereditary hypergraph with vertex set $[n]$ and monotone hyperedge set $M \in \mathcal{F}_n$. The probability distribution on $\mathcal{F}_n$ is

$$P_n(G;\Lambda_n) = \text{pr}(M \subset G;\Lambda_n) = \exp(-\Lambda_n(\overline{G})); \tag{1}$$

$$p_n(G;\Lambda_n) = \text{pr}(M = G;\Lambda_n) = \exp(-\Lambda_n(\overline{G})) \prod_{a\in\text{Gen}(G)} (1 - e^{-\lambda_n(a)}), \tag{2}$$

for any monotone subset $G \in \mathcal{F}_n$ with complement $\overline{G} \subset \mathcal{P}([n])$ and generating class $\text{Gen}(G) \subset \mathcal{P}([n])$.

4.2 Exchangeability and Kolmogorov consistency

A permutation $\sigma : [n] \to [n]$ transforms Λ_n to $\sigma\Lambda_n$ and also induces an action $\sigma^* : \mathcal{F}_n \to \mathcal{F}_n$ on monotone subsets. In this way, it satisfies $p_n(G;\Lambda_n) = p_n(\sigma^*G, \sigma\Lambda_n)$. If $\sigma\Lambda_n = \Lambda_n$ for

every σ, which implies that $\lambda_n(a) = \lambda_n(|a|)$ is a function of the number of elements, the mean measure and P_n are both invariant under permutation and we say the monotone subset is finitely exchangeable.

When $m \leq n$, each one-to-one map $\phi : [m] \to [n]$ induces a deletion operator $\phi^* : \mathcal{F}_n \to \mathcal{F}_m$. The Kolmogorov consistency for the projective system $\{\mathcal{F}_n\}$ (PS I or PS IIa) requires that $p_m(\cdot; \Lambda_m)$ should be the marginal distribution of $p_n(\cdot; \Lambda_n)$ under ϕ^*, that is,

$$p_m(G; \Lambda_m) = p_n(\phi^{*-1}G; \Lambda_n) \tag{3}$$

for each $G \in \mathcal{F}_m$ and each sub-sample ϕ. Evidently, Kolmogorov consistency also imposes some conditions on the mean measures $\{\Lambda_n\}$.

Theorem 4.1. *For each $n \geq 0$, let Λ_n be a mean measure on $\mathcal{P}([n])$ and $\lambda_n(a) = \Lambda_n(\{a\}) = \lambda_n(|a|), a \in \mathcal{P}([n])$ be a function of the number of elements. Then the sufficient and necessary condition for the distribution (2) of the Poisson-generated hereditary hypergraphs to be infinitely exchangeable and consistent under deletion of elements is for every $n \geq 0$,*

$$\lambda_n(s) = \lambda_{n+1}(s) + \lambda_{n+1}(s+1), \text{ for } s = 0, 1, \ldots, n. \tag{4}$$

Proof. We first consider the very simple monotone subset $G = \emptyset$. Since $\phi^{*-1}\emptyset = \emptyset$, consistency (3) requires $\exp(-\Lambda_m(\overline{\emptyset})) = \exp(-\Lambda_n(\overline{\emptyset}))$ for $m \leq n$, which implies $\Lambda_m(\mathcal{P}([m])) = \Lambda_n(\mathcal{P}([n]))$. That means the total mass $\Lambda_n(\mathcal{P}([n]))$ should be the same for each n.

More generally, we consider a monotone subset $G = \langle a_1, a_2, \ldots, a_k \rangle \in \mathcal{F}_m$ where $a_1, a_2, \ldots, a_k$ are maximal in G and $[0, G]$ is the set of monotone subsets $G' \in \mathcal{F}_m$ such that $G' \subset G$. For $n = m + 1$ and the insertion map $\phi : [m] \to [m+1]$, the inverse image of $[0, G]$ under ϕ^* is

$$\phi^{*-1}[0, G] = [0, \langle a_1 \cup \{n\}, a_2 \cup \{n\}, \ldots, a_k \cup \{n\} \rangle]. \tag{5}$$

Here 0 is the empty set $\emptyset$ and $[0, G]$ is a sublattice of $\mathcal{P}([m])$. So consistency for $k = 1$ requires $\Lambda_m(\overline{[0, a]}) = \Lambda_{m+1}(\overline{[0, a \cup \{n\}]})$ for each interval $[0, a]$ in the binary lattice $\mathcal{P}([m])$. With fixed total mass, we then have $\Lambda_m([0, a]) = \Lambda_{m+1}([0, a \cup \{n\}])$. If we write $\lambda_m(a)$ in the form $\lambda_m(|a|)$, (4) is one necessary condition for the distribution (2) to be infinitely exchangeable and consistent under deletion of elements.

Condition (4) is also sufficient. First, (4) leads to $\Lambda_m(\mathcal{P}([m])) = \Lambda_{m+1}\mathcal{P}([m+1])$ by summing over all the subsets. Hence, consistency is satisfied for $G = \emptyset$. For each $0 \le r \le m$,

$$\sum_{i=0}^{r} \binom{r}{i} \lambda_m(i) = \sum_{i=0}^{r+1} \binom{r+1}{i} \lambda_{m+1}(i) \tag{6}$$

also follows from (4), which implies $\Lambda_m([0,a]) = \Lambda_{m+1}([0, a \cup \{n\}])$ for each interval $[0,a]$. By the same argument given above, consistency holds for any monotone subset with one maximal element in its generating class. For the most general case $G = \langle a_1, a_2, \ldots, a_k \rangle \in \mathcal{F}_m$, by the principle of inclusion-exclusion, we have

$$\begin{aligned} \Lambda_m(G) &= \sum_{i=1}^{k} \Lambda_m([0,a_i]) - \sum_{1 \le i < j \le k} \Lambda_m([0, a_i \cap a_j]) + \cdots + (-1)^{k-1} \Lambda_m([0, a_1 \cap \cdots \cap a_k]) \\ &= \sum_{h=1}^{k} (-1)^{h+1} \left(\sum_{1 \le i_1 < \cdots < i_h \le k} \Lambda_m([0, a_{i_1} \cap \cdots \cap a_{i_h}]) \right). \end{aligned} \tag{7}$$

Similarly, for $n = m+1$,

$$\Lambda_{m+1}(\phi^{*-1}G) = \sum_{h=1}^{k} (-1)^{h+1} \left(\sum_{1 \le i_1 < \cdots < i_h \le k} \Lambda_{m+1}([0, a_{i_1} \cap \cdots \cap a_{i_h} \cup \{n\}]) \right). \tag{8}$$

As has been verified, $\Lambda_m([0,a]) = \Lambda_{m+1}([0, a \cup \{n\}])$ for each interval $[0,a]$. So the summands in (7) and (8) are one-to-one equal. Hence, $\Lambda_m(G) = \Lambda_{m+1}(\phi^{*-1}G)$ and (3) follows for each $G \in \mathcal{F}_m$ and the insertion map $\phi : [m] \to [m+1]$. Evidently, consistency satisfies in any case of $m \le n$ since deletion can be conducted sequentially and we can define the deletion operator by removing the vertices in $[n] - [m]$ one by one in any order. □

Notice that (4) is also the necessary and sufficient condition for the Poisson-generated multi-hypergraphs to be infinitely exchangeable and consistent under PS IIb.

4.3 Examples

Finite total mass

In this subsection, we find explicit forms of $\lambda_n(s)$ such that (4) holds. We first consider the case where the total mass $\Lambda_n(\mathcal{P}([n]))$ is finite and independent of n.

Theorem 4.2. *Assume that $\Lambda_n(\mathcal{P}([n])) = 1$ for each n. Then the form of $\lambda_n(s)$ such that (4) holds must be*

$$\lambda_n(s) = \int_0^1 p^s(1-p)^{n-s} d\mu(p), \tag{9}$$

where μ is a distribution function on $[0,1]$.

Proof. By making the transformation $w_n(s) = 2^{-n}\lambda_n(s)$, we have the following equivalent system of equations

$$w_n(s) = \frac{1}{2}\omega_{n+1}(s) + \frac{1}{2}\omega_{n+1}(s+1), \text{ for } s = 0, 1, \ldots, n. \tag{10}$$

In the language of Markov chain theory, these equations state that w is a harmonic function for the random walk on the space of points (n, s) that moves according to the rule "if at (n, s) then jump to either$(n+1, s)$ or $(n+1, s+1)$ with probability 1/2". Martin boundary theory (see Theorem 1.2 and Remark (a) in Ney and Spitzer (1966)) indicates w has a unique representation

$$w_n(s) = 2^n \int_0^1 p^s(1-p)^{n-s} d\mu(p), \tag{11}$$

which immediately implies the form of $\lambda_n(s)$.

□

From another point of view, we consider the presence of each vertex as an infinite exchangeable sequence of Bernoulli-distributed random variables $X_1, X_2, \ldots, X_n, \ldots$. For each $a \in \mathcal{P}([n])$, define the probability function

$$\lambda_n(a) = P(X_i = 1, \text{ for } i \in a \text{ and } X_i = 0, \text{ for } i \in [n] - a). \tag{12}$$

Then the infinite exchangeability of the binary sequence and de Finetti's theorem also imply the form (9).

More generally, assume $\Lambda_n(\mathcal{P}([n])) = K$ for each n, where K is a finite positive constant independent of n, then $\lambda_n(s)$ is of the form

$$\lambda_n(s) = K \int_0^1 p^s(1-p)^{n-s} d\mu(p). \tag{13}$$

Corollary 4.3. *Consider the special case of (13) in which $\mu(1/2) = 1$, For a hereditary hypergraph $G \in \mathcal{F}_n$, we have*

$$p_n(G; \Lambda_n) \quad = \quad \alpha_n^{2^n - \#G}(1 - \alpha_n)^{\#\mathrm{Gen}(G)}, \tag{14}$$

where $\alpha_n = \exp(-K/2^n)$, $\#G$ is the number of elements in the monotone subset G and $\#\mathrm{Gen}(G)$ is the number of maximal elements of G.

For a large enough n, the most probable non-empty hereditary hypergraph for the distribution (14) is the full hypergraph $\langle [n] \rangle$ and the most probable non-empty equivalence class under permutation is the one with a single maximal element of size $\lfloor (n+1)/2 \rfloor$.

Let $\{\alpha_n\}_{n \geq 0}$ be an increasing sequence with $\alpha_n \in (0, 1)$, and let $\{P_n\}_{n \geq 0}$ be a sequence of probability distributions on hereditary hypergraphs defined by (14). Then the necessary and sufficient condition for Kolmogorov consistency under PS I and PS IIa is $\alpha_n = \alpha_{n+1}^2$.

Infinite total mass

As has been mentioned, the mean measure need not be finite. A non-trivial expression for $\lambda_n(s)$ satisfying (4) is

$$\begin{aligned} &\lambda_n(0) = \lambda_n(1) = \infty; \\ &\lambda_n(s) = (s-2)!/n^{\downarrow(s-1)} \text{ for } 2 \leq s \leq n, \end{aligned} \tag{15}$$

where $n^{\downarrow s} = n(n-1)\cdots(n-s+1)$ is the descending factorial symbol. With this mean measure, there must exist an edge for each singleton in $[n]$ for the induced hereditary hypergraph. The total mass on hyperedges of size 2 or more is

$$\sum_{s=2}^{n} \binom{n}{s} \lambda_n(s) = \sum_{s=2}^{n} \binom{n}{s} \frac{(s-2)!}{n^{\downarrow(s-1)}} = n - \sum_{s=1}^{n} \frac{1}{s} \tag{16}$$

which can be approximated by $n - \log(n)$ when n is large. In this case, when n is large enough, the most probable non-empty hereditary hypergraph and the most probable non-empty equivalence class under permutation are both the full hypergraph $\langle [n] \rangle$.

4.4 Conditional probability

If we are to construct a hereditary hypergraph sequentially by addition of vertices, the conditional probability is of interest. Consider $n = m + 1$ and the insertion map $\phi : [m] \to$

$[n]$. Since $\phi^{*-1}\emptyset = \emptyset$, we must have $p(G_n = \emptyset | G_m = \emptyset) = 1$. For $G_m = \langle \emptyset \rangle$, there are only two outcomes for $\phi^{*-1}G_m$: $\langle \emptyset \rangle$ and $\langle \{n\} \rangle$. Assume $\{\Lambda_m\}$ is a mean measure sequence such that (4) holds. Then

$$p(G_n | G_m = \langle \emptyset \rangle) = \begin{cases} \exp^{-\lambda_n(\{n\})}, & G_n = \langle \emptyset \rangle \\ 1 - \exp^{-\lambda_n(\{n\})}, & G_n = \langle \{n\} \rangle \end{cases}. \tag{17}$$

More generally, we consider the case where $G_m = \langle a_1, a_2, \ldots, a_k \rangle \in \mathcal{F}_m$ consisting of k generating subsets with all a_j's non-empty. If we add one more vertex, say $n = m+1$, the conditional probability given $G_m = \langle a_1, a_2, \ldots, a_k \rangle$ is

$$p(G_n | G_m) = \begin{cases} \prod_{a \in G_m} e^{-\lambda_n(a \cup \{n\})}, & G_n = G_m \\ \prod_{a \in G_m \setminus \{\emptyset\}} e^{-\lambda_n(a \cup \{n\})} \cdot (1 - \exp^{-\lambda_n(\{n\})}), & G_n = \langle a_1, a_2, \ldots, a_k, \{n\} \rangle \\ \prod_{a \in \Omega_1} e^{-\lambda_n(a)} \cdot \prod_{b \in \Omega_2} (1 - e^{-\lambda_n(b)}), & \text{otherwise} \end{cases} \tag{18}$$

where $\Omega_1 = \langle a_1 \cup \{n\}, a_2 \cup \{n\}, \ldots, a_k \cup \{n\} \rangle - G_n$ is the set of possible new hyperedges created (not already in G_n) and $\Omega_2 = \mathrm{Gen}(G_n) - \mathrm{Gen}(G_m) \cap \mathrm{Gen}(G_n)$ is the set of new maximal elements in G_n.

4.5 Marginal distribution

For each positive integer $k \leq n$, let $\mathcal{F}_n^{(k)} \subset \mathcal{F}_n$ be the set of monotone subsets $M \subset \mathcal{P}([n])$ such that each element of M is of size no more than k, i.e.,

$$\mathcal{F}_n^{(k)} = \{M \in \mathcal{F}_n : a \in M \text{ implies } |a| \leq k\}.$$

Evidently, $\mathcal{F}_n^{(1)} \subset \mathcal{F}_n^{(2)} \subset \cdots \mathcal{F}_n^{(n)} = \mathcal{F}_n$. In statistical language, $\mathcal{F}_n^{(k)}$ is the set of factorial models in which all of the interactions are of order no greater than k. Specifically, $\mathcal{F}_n^{(2)}$ is the set of undirected graphs. Each hereditary hypergraph $M \in \mathcal{F}_n$ can be reduced to another hypergraph $D_k M \in \mathcal{F}_n^{(k)}$ by deleting all hyperedges of degree $k+1$ or more. As a result, $D_1 M$ is the set of vertices that appear in M and $D_2 M$ is the graph consisting of vertices and undirected edges in M.

For each distribution P_n on $\mathcal{F}_n$, there corresponds a marginal distribution $D_k P_n$ on $\mathcal{F}_n^{(k)}$ defined by $(D_k P_n)(M) = P_n(D_k^{-1} M)$. If the distributions $\{P_n\}$ are infinitely exchangeable and consistent for $n = 1, 2, \ldots$, so also are the marginal distributions $\{D_k P_n\}$ for each

fixed k. In such cases, the Aldous-Hoover representation theorem (Aldous (1981), Hoover (1979)) applies, and indicates that the random graphs with distribution $D_2 P_n$ are dense in the sense that the expected number of edges is of order n^2. More generally, for a fixed k, the expected number of k-hyperedges (hyperedges of size k) is of order n^k as n goes to infinity.

References

Aldous, D. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598. 14

Aldous, D. (1985). *Exchangeability and Related Topics*. Springer. 4

Andersen, A. (1974). Multidimensional contingency tables. *Scandinavian Journal of Statistics*, pages 115–127. 8

Bailey, R. (1991). Strata for randomized experiments. *Journal of the Royal Statistical Society. Series B*, pages 27–78. 5

Bailey, R. (1996). Orthogonal partitions in designed experiments. *Designs, Codes and Cryptography*, 8(1-2):45–77. 5

Berge, C. (1973). *Graphs and Hypergraphs*, volume 7. North-Holland, Amsterdam. 1, 5

Broderick, T., Jordan, M., and Pitman, J. (2013a). Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3):289–312. 4

Broderick, T., Pitman, J., and Jordan, M. (2013b). Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 8(4):801–836. 4

Darling, R. and Norris, J. (2005). Structure of large random hypergraphs. *The Annals of Applied Probability*, 15(1A):125–152. 1

Darroch, J., Lauritzen, S., and Speed, T. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, pages 522–539. 5, 6, 8

Dawid, A. (1977). Invariant distributions and analysis of variance models. *Biometrika*, 64(2):291–297. 5

Dawid, A. (1988). Symmetry models and hypotheses for structured data layouts. *Journal of the Royal Statistical Society. Series B*, pages 1–34. 5

Doshi, F., Miller, K., Van Gael, J., and Teh, Y. (2009). Variational inference for the Indian buffet process. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pages 137–144. 4

Erdős, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297. 3

Griffiths, T. and Ghahramani, Z. (2005). Infinite latent feature models and the Indian buffet process. In *NIPS*, pages 475–482. MIT Press. 4

Haberman, S. (1970). *The general log-linear model*. Ph.D. dissertation, University of Chicago. 6

Haberman, S. (1974). *The Analysis of Frequency Data*. University of Chicago Press, Chicago. 8

Harary, F. (1969). *Graph Theory*. Addison-Wesley. 2

Hoover, D. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*. 14

Kingman, J. (1978). The representation of partition structures. *Journal of the London Mathematical Society*, 2(2):374–380. 4

Korshunov, A. (1981). The number of monotone Boolean functions. *Problemy Kibernet*, (38):5–108. 4, 6

Lee, J. (2000). *Introduction to Topological Manifolds*. Springer. 5

Nelder, J. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society. Series A*, 140(1):pp. 48–77. 5

Ney, P. and Spitzer, F. (1966). The Martin boundary for random walk. *Transactions of the American Mathematical Society*, pages 116–132. 11

Speed, T. (1987). What is an analysis of variance? *The Annals of Statistics*, pages 885–910. 5

Speed, T. and Bailey, R. (1987). Factorial dispersion models. *International Statistical Review/Revue Internationale de Statistique*, pages 261–277. 5

Thibaux, R. and Jordan, M. (2007). Hierarchical beta processes and the Indian buffet process. In *International conference on artificial intelligence and statistics*, pages 564–571. 4

Tjur, T. (1984). Analysis of variance models in orthogonal designs. *International Statistical Review*, pages 33–65. 5

Tjur, T., Sundberg, R., and Helland, I. (1991). Analysis of variance and design of experiments. *Scandinavian Journal of Statistics*, pages 273–322. 5

Wiedemann, D. (1991). A computation of the eighth Dedekind number. *Order*, 8(1):5–6. 4, 6

Wilkinson, G. and Rogers, C. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, pages 392–399. 5, 7