Acta Math. Hungar., **122 (1–2)** (2009), 131–152.
DOI: 10.1007/s10474-008-7240-5
First published online July 13, 2008

CONDITIONAL INDEPENDENCE RELATIONS AND LOG-LINEAR MODELS FOR RANDOM MATCHINGS

V. CSISZÁR

Loránd Eötvös University, Department of Probability, 1117 Budapest,
Pázmány P. sétány 1/C, Hungary
e-mail: villo@cs.elte.hu

(Received December 17, 2007; revised March 7, 2008; accepted March 10, 2008)

Abstract. We study probability distributions on all possible complete matchings in a complete bipartite graph, where the vertices in both sets admit a linear order. We define a family of distributions, and give its equivalent implicit and explicit (parametric) description: it is characterized implicitly by a collection of interesting conditional independence statements, or explicitly by the property that the distributions belonging to the family factorize into factors which depend on "local" properties of the matching. We also calculate the number of free parameters in this family.

1. Introduction

Let $G = (M, W, M \times W)$ denote a complete bipartite graph, where $|M| = |W| = n$ are the two sets of vertices. n is an arbitrary positive integer, which will be fixed throughout the paper (nontrivial results will hold only for $n \geqq 4$). We think of M as a set of men, W as a set of women. The set of edges is $M \times W$, i.e. there is an edge (m, w) in the graph for all pairs $(m, w) \in M \times W$. A complete matching is a subset $\delta \subset M \times W$ such that for every $m \in M$, there is exactly one $w \in W$ such that $(m, v) \in \delta$. Thus a

Key words and phrases: matching, random permutations, log-linear models, conditional independence.

2000 Mathematics Subject Classification: 62E10, 62H05, 60C05.

complete matching allocates a wife to each man. We will sometimes omit the adjective "complete", since we do not address incomplete matchings in this paper. We suppose that both men and women admit a linear order, which reflects the order of desirability of the individuals as marriage partners (or anything else). Thus we label both men and women with the elements of the set $[n]=\{1,\dots,n\}$: the smaller the label, the more desirable the individual is. This allows us to identify both M and W with the set $[n]$, so the edges become ordered pairs (i,j), where (i,j) denotes the edge between the ith man and the jth woman. We will think of the graph G drawn in the plane such that the vertices are arranged in two columns: men to the left, women to the right, and labels increasing from top to bottom. Edges are drawn as straight line segments.

A complete matching δ is equivalent to two permutations $\pi_W,\pi_M\in S_n$, where S_n denotes the symmetric group of all permutations of the set $[n]$. We denote an element of S_n as $\pi=\big(\pi(1),\pi(2),\dots,\pi(n)\big)$. Then π_W and π_M are given by

$$(1)\qquad \delta=\Big\{\big(i,\pi_W(i)\big)\,:\,1\leqq i\leqq n\Big\}=\Big\{\big(\pi_M(i),i\big)\,:\,1\leqq i\leqq n\Big\}.$$

For example, for $n=4$, if $\delta=\big\{(2,3),(1,4),(4,2),(3,1)\big\}$, then $\pi_W=(4312)$ is an ordering on the women, and $\pi_M=(3421)$ is an ordering on the men. Notice that π_W and π_M are each other's inverse in the group S_n. The three objects δ, π_W, π_M are equivalent.

Our aim is to define and study meaningful probability distributions on random complete matchings Δ, equivalently on random permutations Π_W or Π_M. Notice that we denote random variables (random matchings or random permutations) by uppercase letters. It will be convenient to work with distributions on S_n, denoted as $p=\big(p(\pi)\,:\,\pi\in S_n\big)$ where the numbers $p(\pi)$ are nonnegative, and their sum is one. If Δ is a random matching, we denote by p_W (resp. p_M) the distribution of Π_W (resp. Π_M) on S_n.

A probability distribution p on S_n can be specified in two ways: implicitly or explicitly. The implicit description is given by a set of relations, which the probabilities $p(\pi)$ must satisfy. The explicit description, on the other hand, is given by a parametric form for the probabilities. To make things clear, we illustrate this by a small example.

EXAMPLE 1. Let $n=4$. We are looking for distributions on complete matchings such that the following holds. Given that the first two men marry the first two women, the wife of the first man is independent of the wife of the third man. In other words, we require that $\Pi_W(1)$ and $\Pi_W(3)$ be conditionally independent, given that $\big\{\Pi_W(1),\Pi_W(2)\big\}=\{1,2\}$. Denoting the distribution of Π_W by p_W, this is easily seen to be equivalent to the relation

$$(2)\qquad p_W(1234)p_W(2143)=p_W(1243)p_W(2134).$$

Alternatively, we can postulate the existence of nonnegative parameters θ_i, $1 \leqq i \leqq 4$, such that

(3)
$$p_W(1234) = \theta_1\theta_3,\ p_W(1243) = \theta_1\theta_4,\ p_W(2134) = \theta_2\theta_3,\ p_W(2143) = \theta_2\theta_4,$$

while for the probabilities of the other 20 permutations in S_4, we make no restrictions. These two characterizations (2) and (3) are easily seen to be equivalent.

We will give equivalent implicit and explicit characterizations of families of distributions, as in Example 1. The implicit descriptions will consist of conditional independence statements. Similar conditional independence statements were studied previously in the context of random orderings of n objects (according to preference). More specifically, the property called L-decomposability of the distribution of a random ordering is equivalent to a collection of conditional independence statements. This property, introduced in [1], is well-known in the literature of mathematical psychology. The distribution of a random ordering $\Pi = \big(\Pi(1), \ldots, \Pi(n)\big)$ is called L-decomposable, if for all k and $C \subset [n]$ with $|C| = k$, conditional on the event

$$\big\{\Pi(1), \ldots, \Pi(k)\big\} = C,$$

the random orderings $\big(\Pi(1), \ldots, \Pi(k)\big)$ and $\big(\Pi(k+1), \ldots, \Pi(n)\big)$ are independent.

However, in the case of random matchings, the random orderings Π_W and $\Pi_M = \Pi_W^{-1}$ have symmetric role. Therefore we wish to identify random matchings Δ, where p_W and p_M are both L-decomposable: in this case we say that Δ, Π_W, Π_M (or their distributions) are bi-decomposable. We will prove that a strictly positive distribution p_W (or p_M) is bi-decomposable if and only if $p_W(\pi_W)$ (or $p_M(\pi_M)$) factorizes into factors, each of which depends on "local" properties of the corresponding matching δ. The following example illustrates what we mean.

Example 2. Let δ be a complete matching in G, and for an edge $e \in \delta$ let $T(e, \delta)$ denote the number of edges in δ which cross the edge e. We remind the reader that this definition makes sense, since we fixed how the graph is drawn in the plane. Formally, if $e = (m_e, w_e)$ (where $m_e, w_e \in [n]$), then

$$T(e,\delta) = \Big|\big\{ f \in \delta :\ (m_f < m_e \text{ and } w_f > w_e) \text{ or } (m_f > m_e \text{ and } w_f < w_e)\big\}\Big|.$$

A parametric family of probability distributions on matchings can be defined by the formula

$$P(\Delta = \delta) = P(\Pi_W = \pi_W) = p_W(\pi_W) = \prod_{e\in\delta} \theta_e\big(T(e,\delta)\big),$$

where $\theta_e(t)$ are arbitrary parameters for $e \in [n] \times [n]$, and $0 \leqq t \leqq n-1$. The statistic $T(e, \delta)$ can be called "local", since it is determined by what e "sees" of the matching.

REMARK 1. Our research is analogous, though much less general, to the theory of graphical models for contingency tables. An important result in that theory is the Hammersley–Clifford theorem, which we describe briefly. Let G be a graph with n vertices labelled $1, \dots, n$. Denote the set of cliques (maximal complete subgraphs) of G by $\mathcal{C}$, and for a vector $x = (x_1, \dots, x_n)$ and $A \subset [n]$, let $x_A = (x_i : i \in A)$. Let $X = (X_1, \dots, X_n)$ be an n-dimensional random variable taking values in a finite set $I = I_1 \times \dots \times I_n$. We say that a probability distribution p on I factorizes according to the cliques of G if

$$p(x) = P(X = x) = \prod_{C \in \mathcal{C}} \theta_C(x_C) \quad \forall x \in I, \tag{4}$$

where θ_C are suitable parameters. If we suppose in addition that p is strictly positive, (4) becomes an exponential family, or more specifically, a log-linear model, called the graphical model with graph G. The Hammersley–Clifford theorem states that (4) holds for a strictly positive p, if and only if p satisfies the global Markov property with respect to G: for any disjoint subsets $A, B, S \subset [n]$, X_A and X_B are conditionally independent, given X_S, if S separates A and B in the graph G. For more details, see e.g. [3]. Recent related research has been done in the context of toric statistical models [2, 5, 6].

The results obtained in this paper are mainly theoretic. However, we believe that these or similar models can successfully be applied to real datasets in the future. The paper is organised as follows. In Section 2, we review L-decomposability of distributions on S_n. Section 3 contains the main results. Section 3.1 introduces log-linear models for distributions on S_n in general, in analogy with log-linear models in the case of contingency tables. In Section 3.2 we show that strictly positive bi-decomposable distributions on S_n form a log-linear model, and calculate the number of free parameters of the model. Section 4 treats some issues of estimation, while Section 5 contains the proofs of some lemmas.

2. L-decomposability

For integers $i \leqq j$, $\{i \dots j\}$ will denote the set $\{k : i \leqq k \leqq j\}$. For any vector $v = \big(v(1), \dots, v(s)\big)$, we call $v(i)$ the ith element of v. For the set of the ith to jth elements, and for the subvector of the ith to jth elements of v, introduce the notations

$$v\{i \dots j\} = \big\{v(i), \dots, v(j)\big\}, \quad v(i \dots j) = \big(v(i), \dots, v(j)\big), \quad 1 \leqq i \leqq j \leqq s.$$

If $j < i$ then let $v\{i \dots j\}$ be the empty set. Let S_n stand for the symmetric group of all permutations π of $[n] = \{1 \dots n\}$. We denote a probability distribution on S_n by $p = \big\{ p(\pi) : \pi \in S_n \big\}$. Denote by $\Pi : \Omega \to S_n$ a random permutation on a probability space $(\Omega, \mathcal{A}, P)$ with distribution p, that is $P(\Pi = \pi) = p(\pi)$.

The idea of *L-decomposability* first appears in [1], motivated by Luce's ranking postulate [4]. A random permutation Π is L-decomposable, if for any k, the value of $\Pi(k+1)$ depends on $\Pi(1 \dots k)$ only through $\Pi\{1 \dots k\}$. Recall that the probability of a permutation can always be written in the product form

$$P(\Pi = \pi) = \prod_{k=0}^{n-1} P\big(\Pi(k+1) = \pi(k+1) \mid \Pi(1 \dots k) = \pi(1 \dots k)\big). \tag{5}$$

L-decomposability means that the conditions $\Pi(1 \dots k) = \pi(1 \dots k)$ can be replaced by the conditions $\Pi\{1 \dots k\} = \pi\{1 \dots k\}$ in (5). We formulate this in the following definition, in two different forms. For two permutations $\pi, \sigma \in S_n$, $\pi\sigma$ is their product in the group S_n, i.e. $(\pi\sigma)(i) = \pi\big(\sigma(i)\big)$.

Definition 1. Let Π be a random permutation with probability distribution p on S_n. p (or sometimes Π) is called *L-decomposable*, if either of the two equivalent conditions holds:

(i) For every $2 \leqq k \leqq n-2$,

$$\begin{aligned} &P\big(\Pi(k+1) = \pi(k+1) \mid \Pi(1 \dots k) = \pi(1 \dots k)\big) \\ &= P\big(\Pi(k+1) = \pi(k+1) \mid \Pi(1 \dots k) = (\pi\sigma)(1 \dots k)\big), \end{aligned} \tag{6}$$

for all $\pi \in S_n$ and $\sigma \in S_k$ such that both conditional probabilities are defined.

(ii) For every $2 \leqq k \leqq n-2$,

$$\begin{aligned} &P\big(\Pi(k+1) = \pi(k+1) \mid \Pi(1 \dots k) = \pi(1 \dots k)\big) \\ &= P\big(\Pi(k+1) = \pi(k+1) \mid \Pi\{1 \dots k\} = \pi\{1 \dots k\}\big), \end{aligned} \tag{7}$$

for all $\pi \in S_n$ such that the left hand side is defined.

We could formally include $k = 0, 1, n-1$ in the definition as well, but equations (6) and (7) are always satisfied for these k-values. It follows that for $n \leqq 3$, all distributions on S_n are L-decomposable.

The equivalence of the two formulations of Definition 1 is easily verified. The next proposition gives the implicit and the parametric (explicit) characterization of L-decomposability. Before we can formulate it, we introduce some more notation. A partition of the set $[n]$ into s disjoint subsets (also called atoms) is given by

$$\mathcal{Z} = (Z_1, \ldots, Z_s) : \cup_{i=1}^{s} Z_i = [n], \quad Z_i \cap Z_j = \emptyset \quad \forall i \neq j. \tag{8}$$

If none of the sets Z_i is empty, we call s the size of the partition. For an atom Z_i, the ordered, resp. unordered, restriction of $\pi \in S_n$ to Z_i is

$$\pi(Z_i) = \big(\pi(j) : j \in Z_i\big), \quad \pi\{Z_i\} = \big\{\pi(j) : j \in Z_i\big\}. \tag{9}$$

For the partition $\mathcal{Z}$, the coarsening of $\pi \in S_n$ to $\mathcal{Z}$ is

$$\big|\pi(\mathcal{Z})\big| = \big(\pi\{Z_i\} : i = 1, \ldots, s\big). \tag{10}$$

We will return to the idea behind this notation later. For example, suppose that δ is a complete matching in G, and π_W is the ordering on the women given by (1). In the graph $G_\delta = (M, W, \delta)$, substitute each set of vertices $Z_i \subset M$ by one vertex i^*, such that i^* inherits all edges adjacent to the old vertices $m \in Z_i$. Then i^* will be connected exactly with the vertices $\pi_W\{Z_i\} \subset W$. Thus $\big|\pi_W(\mathcal{Z})\big|$ describes the adjacencies in the new graph.

$\mathcal{Z}$ in (8) is called a consecutive partition, if the atoms are all intervals, i.e. $Z_i = \{a_i \ldots b_i\}$ for some $a_i < b_i$.

PROPOSITION 1. *The distribution p of a random permutation Π is L-decomposable if and only if, for every consecutive partition $\mathcal{Z}$ as in* (8)*, the ordered restrictions $\Pi(Z_i)$, $1 \leqq i \leqq s$ are conditionally independent, given $\big|\Pi(\mathcal{Z})\big|$. Equivalently, p is L-decomposable if and only if there exist a constant c, and a nonnegative function θ defined on pairs*

$$(x, C) : C \subset [n], \ x \notin C, \tag{11}$$

such that for all $\pi \in S_n$

$$p(\pi) = c \prod_{k=0}^{n-1} \theta\big(\pi(k+1), \pi\{1 \ldots k\}\big). \tag{12}$$

We leave the proof of the proposition to the reader, as the main ideas, as well as being standard, can be found in [1]. If (12) holds, the pair (θ, c) is called an L-decomposition of the distribution p. By (5) and (7), one L-decomposition of the L-decomposable distribution p is given by $c = 1$ and

$$\theta(x, C) = P\big(\Pi\big(|C| + 1\big) = x \mid \Pi\big\{1 \ldots |C|\big\} = C\big), \tag{13}$$

if the probability of the condition is positive, otherwise $\theta(x, C) = 0$. For strictly positive distributions p, the decomposition (13) is unique in the sense that it is the only decomposition satisfying $c = 1$ and $\sum_{x \notin C} \theta(x, C) = 1$ for all C. Therefore, for strictly positive L-decomposable distributions, the number of free parameters $\theta(x, C)$ is given by

$$\lambda_n = \sum_{k=0}^{n-1} \binom{n}{k} (n - k - 1) = 2^n (n/2 - 1) + 1.$$

Proposition 1 thus gives the equivalent implicit and explicit characterization of L-decomposable distributions on S_n. Notice that the equivalence is true not only for strictly positive distributions, but for any distribution on S_n. This is essentially because the family defined by (12) is closed (under pointwise convergence of probability distributions).

The main aim of the present paper is to characterize random matchings Δ, for which the distribution of both Π_M and Π_W is L-decomposable. We will call the distribution of such a random matching bi-decomposable. This concept can naturally be defined for distributions on S_n as well.

DEFINITION 2. The distribution of a random matching Δ is called *bi-decomposable,* if the distributions of Π_W and Π_M in (1) are both L-decomposable. The distribution of a random permutation Π is *bi-decomposable,* if it is L-decomposable, and the distribution of Π^{-1} is also L-decomposable.

Thus the distribution of Δ is bi-decomposable if and only if the distribution of Π_W (equivalently Π_M) is bi-decomposable.

Let $\mathcal{M}$ and $\mathcal{W}$ be two partitions of $[n]$ of sizes d and r respectively. Their product is a partition of $[n] \times [n]$ of size dr:

$$\mathcal{M} \times \mathcal{W} = (M_i \times W_j : 1 \leqq i \leqq d,\ 1 \leqq j \leqq r), \tag{14}$$

where $M_i \times W_j = \{ (m, w) : m \in M_i,\, w \in W_j \}$. The restriction of a matching δ to an atom of a product partition $\mathcal{M} \times \mathcal{W}$ is defined as

$$\delta(M_i \times W_j) = \big((m, w) \in \delta : m \in M_i, w \in W_j \big).$$

This is just the subgraph of $G_\delta = (M, W, \delta)$ induced on the vertex set (M_i, W_j).

Since the L-decomposability of both Π_W and Π_M can be characterized by conditional independence statements (Proposition 1), this immediately yields the implicit characterization of the bi-decomposability of Δ by conditional independence statements. However, the next proposition, which we prove in Section 5, is stronger.

Proposition 2. *The distribution of the random matching Δ is bi-decomposable if and only if for all pairs of consecutive partitions $\mathcal{M}$ and $\mathcal{W}$ of $[n]$ of sizes d and r respectively, the restrictions $\Delta(M_i \times W_j)$ for $1 \leqq i \leqq d$ and $1 \leqq j \leqq r$ are conditionally independent, given $\left|\Pi_W(\mathcal{M})\right|$ and $\left|\Pi_M(\mathcal{W})\right|$.*

The statement of Proposition 2 is immediate from Proposition 1 only if $\mathcal{M}$ or $\mathcal{W}$ is the trivial partition $\mathcal{Z} = \big([n]\big)$. An explicit (parametric) characterization of bi-decomposable random matchings is harder. In this paper, we will provide this characterization only for strictly positive distributions. The study of distributions where the probability of some matchings is allowed to be zero will be carried out in a subsequent paper.

3. Main results

In the rest of the paper, we focus our attention on strictly positive distributions, i.e. the case when all possible complete matchings have positive probability:

$$P(\Delta = \delta) = P(\Pi_W = \pi_W) = P(\Pi_M = \pi_M) > 0$$

for all δ, π_W, π_M. Suppose the random permutation Π has distribution p on S_n, then denote the distribution of Π^{-1} by p^{-1}:

$$p^{-1}(\pi) = p(\pi^{-1}), \quad \pi \in S_n.$$

Introduce the families of distributions on S_n

$$\mathcal{L} = \{p : p \text{ is strictly positive and } L\text{-decomposable}\},$$

$$\mathcal{L}^{-1} = \{p : p^{-1} \text{ is strictly positive and } L\text{-decomposable}\}.$$

We do not indicate the dependence on n, since n is always fixed, though arbitrary (nontrivial results are obtained for $n \geqq 4$). We want to characterize random matchings for which $p_W, p_M \in \mathcal{L}$, equivalently $p_W \in \mathcal{L} \cap \mathcal{L}^{-1}$ (remember that p_W is the distribution of Π_W), or equivalently $p_M \in \mathcal{L} \cap \mathcal{L}^{-1}$. The family

$$\mathcal{B} = \mathcal{L} \cap \mathcal{L}^{-1} \tag{15}$$

consists of all strictly positive bi-decomposable distributions on S_n. For a while, we leave matchings, and concentrate on the family $\mathcal{B}$ of distributions on S_n.

We will show that $p \in \mathcal{L}$ if and only if $\log p \in F$, where the logarithm is taken coordinate-wise, and F is a linear subspace of $\mathbb{R}^{n!}$. The same is true for $\mathcal{L}^{-1}$ with another subspace F^{-1}. Therefore we need to characterize the subspace $H = F \cap F^{-1}$. In the following sections, we will determine the dimensions (and some bases) of these three linear subspaces.

We will use orthogonality; here are the basic definitions and lemmas. Two subspaces U and V of a Hilbert space are called orthogonal, if every pair of vectors $u \in U$, $v \in V$ are orthogonal. The closed subspaces U and V intersect each other orthogonally, if the (orthogonal) projection of U on V equals $U \cap V$, or equivalently, the projection of V on U equals $U \cap V$. Denote the operator of orthogonal projection on U by Pr_U. Another equivalent condition for orthogonal intersection is that the projection operators Pr_U and Pr_V commute. Thus introducing the notation $\perp_\cap$ for orthogonal intersection,

$$U \perp_\cap V \iff \mathrm{Pr}_U V = U \cap V \iff \mathrm{Pr}_V U = U \cap V$$

$$\iff \mathrm{Pr}_U \mathrm{Pr}_V = \mathrm{Pr}_V \mathrm{Pr}_U .$$

We write $U = U_1 \oplus U_2$ for orthogonal decomposition, that is when $U = \mathrm{Span}\,(U_1, U_2)$ and U_1 and U_2 are orthogonal. $\mathrm{Span}\,(U_1, U_2)$ denotes the subspace spanned by U_1 and U_2.

LEMMA 1. *Suppose that* $U = \mathrm{Span}\,(U_i : i \in I)$, $V = \mathrm{Span}\,(V_j : j \in J)$ *are two subspaces, and* $U_i \perp_\cap V_j$ *for every pair* i, j. *Then* $U \perp_\cap V$, *and* $U \cap V = \mathrm{Span}\,(U_i \cap V_j : i \in I, j \in J)$.

LEMMA 2. *Let* $U = U_1 \oplus U_2$ *and* $V = V_1 \oplus V_2$ *be two subspaces with orthogonal decompositions. If* $U \perp_\cap V$, $U_1 \perp_\cap V_1$, $U \perp_\cap V_1$, *and* $U_1 \perp_\cap V$ *hold, then* $U_2 \perp_\cap V_2$ *is also true, and*

$$U \cap V = (U_1 \cap V_1) \oplus (U_1 \cap V_2) \oplus (U_2 \cap V_1) \oplus (U_2 \cap V_2).$$

The proofs of these lemmas are postponed to Section 5.

3.1. Log-linear models. In this section, we define and study log-linear models for distributions on S_n, whose generators are product partitions of $[n] \times [n]$ as in (14). The terms "log-linear model" and "generator" are borrowed from the theory of contingency tables. Let $X = (X_1, \ldots, X_n)$ be an n-dimensional random variable as in Remark 1. Let $\mathcal{A}$ be a set of subsets $A \subset [n]$. The log-linear model with generators $A \in \mathcal{A}$ consists of those (strictly positive) distributions p for which

$$p(x) = P(X = x) = \exp\left[\sum_{A \in \mathcal{A}} \theta_A(x_A)\right] \quad \forall x \in I,$$

for suitable parameters θ_A. The model is called graphical, if there is a graph G such that $\mathcal{A}$ is the set of cliques of G.

Let $\mathcal{D}$ (resp. $\mathcal{R}$) be a partition of $[n]$ of size d (resp. r), the coarsening of the permutation π to the product partition $\mathcal{D} \times \mathcal{R}$ is the $d \times r$ matrix

$$\left|\pi(\mathcal{D} \times \mathcal{R})\right| = (t_{ij}), \quad t_{ij} = \left|\left\{ 1 \leqq s \leqq n : s \in D_i, \ \pi(s) \in R_j \right\}\right|. \tag{16}$$

We changed the letters $\mathcal{M}$ and $\mathcal{W}$ to $\mathcal{D}$ and $\mathcal{R}$, because we now think of π as a function $[n] \to [n]$, and $\mathcal{D}$ is a partition of the domain, while $\mathcal{R}$ is a partition of the range of π. Notice the slight abuse of notation: $\left|\pi(\mathcal{Z})\right|$ is given by (10) if $\mathcal{Z}$ is a partition of $[n]$, while it is given by (16) if $\mathcal{Z} = \mathcal{D} \times \mathcal{R}$ is a product partition of $[n] \times [n]$.

Let us just remark that if δ is a complete matching, and $\mathcal{M}$ and $\mathcal{W}$ are two partitions of $[n]$, then $\left|\pi_W(\mathcal{M} \times \mathcal{W})\right|$ describes the multigraph obtained from $G_\delta = (M, W, \delta)$ by substituting each set of vertices $M_i \subset M$ by one vertex i^*, and substituting each set of vertices $W_j \subset W$ by one vertex j^*, such that the new vertices inherit the edges of the old vertices. Then t_{ij} is just the number of edges between i^* and j^* in the new graph.

A product partition $\mathcal{P} = \mathcal{D} \times \mathcal{R}$ of $[n] \times [n]$ defines a linear subspace $U_{\mathcal{P}} \subset \mathbb{R}^{n!}$. Here $\mathbb{R}^{n!}$ is a Euclidean space, its elements are vectors $v = \big(v(\pi) : \pi \in S_n\big)$. Then

$$U_{\mathcal{P}} = \left\{ v \in \mathbb{R}^{n!} : \left|\pi(\mathcal{P})\right| = \left|\sigma(\mathcal{P})\right| \Rightarrow v(\pi) = v(\sigma) \right\}, \tag{17}$$

i.e. $v \in U_{\mathcal{P}}$ if there exists a function θ on $d \times r$ integer matrices such that $v(\pi) = \theta(\left|\pi(\mathcal{P})\right|)$.

We will define a log-linear model by a set of product partitions, called the generators of the model.

Definition 3. Let $\mathcal{P}_1, \dots, \mathcal{P}_s$ be product partitions of $[n] \times [n]$. A (strictly positive) distribution p on S_n belongs to the *log-linear model with generators* $\mathcal{P}_1, \dots, \mathcal{P}_s$, denoted as $p \in \mathcal{L}(\mathcal{P}_1, \dots, \mathcal{P}_s)$, if

$$\log p(\pi) = \sum_{i=1}^{s} \theta_i\big(\left|\pi(\mathcal{P}_i)\right|\big) \quad \forall \pi \in S_n,$$

where the functions θ_i are arbitrary parameters, or, equivalently,

$$\log p \in \operatorname{Span}(U_1, \dots, U_s),$$

where we used the simplifying notation $U_{\mathcal{P}_i} = U_i$, see (17).

In the rest of this section, we give a sufficient condition, when the intersection of two log-linear models is itself a log-linear model, with directly identifyable generators. The proofs can be found in Section 5. The first lemma describes the relationship between conditional independence and orthogonal intersection.

LEMMA 3. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and denote by $L_2(\mathcal{A})$ the Hilbert space of square-integrable random variables on it. For a σ-algebra $\mathcal{S} \subset \mathcal{A}$, denote by $L_2(\mathcal{S})$ the closed linear subspace of $L_2(\mathcal{A})$ consisting of all $\mathcal{S}$-measurable random variables. Let $\mathcal{S}_1$, $\mathcal{S}_2 \subset \mathcal{A}$. Then $L_2(\mathcal{S}_1) \perp_\cap L_2(\mathcal{S}_2)$ if and only if $\mathcal{S}_1$ and $\mathcal{S}_2$ are conditionally independent, given $\mathcal{S}_1 \cap \mathcal{S}_2$.*

There is a partial ordering on the partitions of an arbitrary set. The partition $\mathcal{Z} = (Z_1, \dots, Z_s)$ is finer than $\mathcal{Z}' = (Z_1', \dots, Z_t')$ (or $\mathcal{Z}'$ is coarser than $\mathcal{Z}$) if for every i there exists a j such that $Z_i \subset Z_j'$. Write $\mathcal{Z} \succ \mathcal{Z}'$ if $\mathcal{Z}$ is finer that $\mathcal{Z}'$. If $\mathcal{P}$ and $\mathcal{P}'$ are product partitions of $[n] \times [n]$, then $\mathcal{P} \succ \mathcal{P}'$ clearly implies $U_{\mathcal{P}} \supset U_{\mathcal{P}'}$. By the application of Lemma 3, we get

LEMMA 4. *Let $\mathcal{D}' \succ \mathcal{D}$ and $\mathcal{R}' \succ \mathcal{R}$ be partitions of $[n]$. Then we have*

$$U_{\mathcal{D}\times\mathcal{R}'} \perp_\cap U_{\mathcal{D}'\times\mathcal{R}} \quad \textit{and} \quad U_{\mathcal{D}\times\mathcal{R}'} \cap U_{\mathcal{D}'\times\mathcal{R}} = U_{\mathcal{D}\times\mathcal{R}}. \tag{18}$$

As a direct corollary of Lemma 4 and Lemma 1, we obtain the main result of this section.

THEOREM 1. *Let $\mathcal{L}(\mathcal{D}_i \times \mathcal{R} : i = 1, \dots, s)$ and $\mathcal{L}(\mathcal{D} \times \mathcal{R}_j : j = 1, \dots, t)$ be two log-linear models, and suppose that $\mathcal{D} \succ \mathcal{D}_i$ and $\mathcal{R} \succ \mathcal{R}_j$ for all $1 \leqq i \leqq s$, $1 \leqq j \leqq t$. Then the intersection of the two models is the log-linear model $\mathcal{L}(\mathcal{D}_i \times \mathcal{R}_j : i = 1, \dots, s,\ j = 1, \dots, t)$.*

3.2. Decomposability as a log-linear model. In this section we prove as our main result that the family $\mathcal{B}$ of strictly positive bi-decomposable distributions on S_n is a log-linear model in the sense of Definition 3, with the number of free parameters equal to $\sum_{i=1}^{n-1} i^2$. We also identify the generators of this log-linear model.

First we need to define some special consecutive partitions of $[n]$. We define the kth *bold section*

$$\Sigma_k = \big(\{1 \dots k-1\}, \{k\}, \{k+1 \dots n\}\big), \quad 2 \leqq k \leqq n-1. \tag{19}$$

We will extend the notation Σ_k to $k = 1$ and $k = n$ for the sake of convenience. The (consecutive) partition which partitions $[n]$ into n sets is called the *full partition:*

$$\Phi = \big(\{1\}, \{2\}, \dots, \{n\}\big). \tag{20}$$

From the multiplicative form (12), it is straightforward that the family $\mathcal{L}$ is the log-linear model

$$\mathcal{L} = \mathcal{L}(\Sigma_k \times \Phi : 1 \leqq k \leqq n).$$

That is, the generators are the products of bold sections with the full partition. The reason is that the $0-1$ matrix $\big|\pi(\Sigma_k \times \Phi)\big|$ is equivalent to the coarsening of π to Σ_k, i.e. $\left|\pi(\Sigma_k)\right|$ defined in (10), which also explains the notation introduced there. And $\left|\pi(\Sigma_k)\right|$ is equivalent to the pair $\bigl(\pi\{1\dots k-1\}, \pi(k)\bigr)$.

The model $\mathcal{L}^{-1}$ can be described similarly as $\mathcal{L}$, only the domain and the range change roles when we consider π^{-1} instead of π. Thus

$$\mathcal{L}^{-1} = \mathcal{L}(\Phi \times \Sigma_k : 1 \leqq k \leqq n).$$

Define the (k,ℓ)*th bold cross-section* as

$$\Lambda_{k\ell} = \Sigma_k \times \Sigma_\ell, \tag{21}$$

where the component (domain and range) partitions are bold sections defined in (19). Now we are ready to state

THEOREM 2. *The family $\mathcal{B}$ of strictly positive bi-decomposable distributions is a log-linear model with generators $\Lambda_{k\ell}$, i.e.*

$$\mathcal{B} = \mathcal{L}(\Lambda_{k\ell} : 1 \leqq k, \ell \leqq n). \tag{22}$$

Moreover, the number of free parameters in the model is

$$\beta_n = \sum_{i=1}^{n-1} i^2. \tag{23}$$

The first statement in Theorem 2 is a direct corollary of Theorem 1. It remains to calculate the number of free parameters.

A submodel of $\mathcal{L}$ which we will use has coarser partitions as generators. Define the kth *thin section* as

$$\widetilde{\Sigma}_k = \bigl(\{1\dots k\}, \{k+1\dots n\}\bigr), \quad 1 \leqq k \leqq n-1. \tag{24}$$

Notice that $\Sigma_1 = \widetilde{\Sigma}_1$ and $\Sigma_n = \widetilde{\Sigma}_{n-1}$. We extend the notation $\widetilde{\Sigma}_k$ to $k = n$ for the sake of convenience. The log-linear model

$$\widetilde{\mathcal{L}} = \mathcal{L}\bigl(\widetilde{\Sigma}_k \times \Phi : 1 \leqq k \leqq n-1\bigr) \tag{25}$$

is a subfamily of $\mathcal{L}$.

Recall from (17) the definition of the subspace corresponding to a product partition, and for the sake of brevity introduce the notations

$$U_{\Sigma_k \times \Phi} = V_k, \quad U_{\widetilde{\Sigma}_k \times \Phi} = \widetilde{V}_k.$$

From the partial ordering of partitions, we get that $\widetilde{V}_k \subset V_k$; denote the orthogonal complement of $\widetilde{V}_k$ in V_k by F_k. By the same argument, $\widetilde{V}_k \subset V_{k+1}$ also holds. This yields that

$$\operatorname{Span}\left(V_k : 1 \leqq k \leqq n\right) = \operatorname{Span}\left(F_k : 1 \leqq k \leqq n, \widetilde{V}_n\right).$$

Since $\Sigma_1 = \widetilde{\Sigma}_1$, we get $F_1 = \{\mathbf{0}\}$. In addition, as $\widetilde{\Sigma}_n$ is the trivial partition, $\widetilde{V}_n = \operatorname{Span}(\mathbf{1})$, where $\mathbf{1} \in \mathbb{R}^{n!}$ is the vector whose every coordinate is 1. Thus the subspace belonging to the L-decomposable log-linear model $\mathcal{L}$ is

$$F := \operatorname{Span}(V_k : 1 \leqq k \leqq n) = \operatorname{Span}(F_k : 2 \leqq k \leqq n, \mathbf{1}). \tag{26}$$

In the next lemma, we show that the subspaces on the right hand side of (26) give an orthogonal decomposition of F. The proof is found in Section 5.

LEMMA 5. *The subspaces* F_k $(2 \leqq k \leqq n)$ *are orthogonal to each other and to the vector* $\mathbf{1}$.

Thus we can calculate again the number of free parameters in $\mathcal{L}$, which is the dimension of F minus one:

$$\lambda_n = \dim(F) - 1 = \sum_{k=2}^{n} \dim(F_k) = \sum_{k=2}^{n} \binom{n}{k}(k-1) = 2^n(n/2-1)+1,$$

where the dimension of F_k is easy to calculate.

The decomposition (26) simplifies the calculations regarding the dimension of the bi-decomposable model $\mathcal{B}$ as well. Define the subspaces V_ℓ^{-1}, $\widetilde{V}_\ell^{-1}$, F_ℓ^{-1}, F^{-1} belonging to the model $\mathcal{L}^{-1}$ just as V_k, $\widetilde{V}_k$, F_k, F were defined for $\mathcal{L}$, only interchange the role of domain and range in every partition. Using the notation introduced in (17),

$$U_{\Phi\times\Sigma_\ell} = V_\ell^{-1}, \quad U_{\Phi\times\widetilde{\Sigma}_\ell} = \widetilde{V}_\ell^{-1}.$$

Applying Lemma 5, the subspace corresponding to $\mathcal{L}^{-1}$ is

$$F^{-1} = \oplus_{\ell=2}^{n} F_\ell^{-1} \oplus \operatorname{Span}(\mathbf{1}). \tag{27}$$

By Lemma 4, for any pair

$$V \in \{V_k, \widetilde{V}_k : 2 \leqq k \leqq n\},\ V^{-1} \in \{V_\ell^{-1}, \widetilde{V}_\ell^{-1} : 2 \leqq \ell \leqq n\},$$

$V \perp_\cap V^{-1}$, since these subspaces correspond to product partitions, where in the V-partitions, the range-partition is as fine as possible, and in the V^{-1}-partitions, the domain-partition is as fine as possible. By Lemma 2, we get

$F_k \perp_\cap F_\ell^{-1}$ for all k, ℓ. By Lemma 1, the space $H = F \cap F^{-1}$ corresponding to bi-decomposable distributions has the orthogonal decomposition

$$H = \oplus_{2 \leqq k, \ell \leqq n} (F_k \cap F_\ell^{-1}) \oplus \mathbf{1}. \tag{28}$$

It remains to find the dimension and a basis of $F_k \cap F_\ell^{-1}$. For the time being, fix k and ℓ. By Lemma 4, the subspace corresponding to the (k, ℓ)th bold cross-section $\Lambda_{k\ell}$ is just $V_k \cap V_\ell^{-1}$. Observe that $F_k \cap F_\ell^{-1}$ consists of exactly those vectors of the space $V_k \cap V_\ell^{-1}$ which are orthogonal to both $\widetilde{V}_k$ and $\widetilde{V}_\ell^{-1}$. Recall that the πth coordinate of a vector in $V_k \cap V_\ell^{-1}$ depends only on its coarsening $\left|\pi(\Lambda_{k\ell})\right|$ defined by (16).

Obviously, the nine elements of the matrix $\left|\pi(\Lambda_{k\ell})\right| = (t_{ij})$ must satisfy row-sum and column-sum constraints: the row sums have to be $k-1$, 1, $n-k$, the column sums $\ell - 1$, 1, $n - \ell$. Therefore the matrix is determined by its coordinates for $i, j = 1, 2$. In fact, it is possible to specify the coarsening $\left|\pi(\Lambda_{k\ell})\right|$ by just two coordinates, which we will now introduce for notational convenience. Let

$$a^{k\ell}(\pi) = t_{11} + t_{12} + t_{21} + t_{22}, \tag{29}$$

and define $q^{k\ell}(\pi)$ by

(30)

$(t_{ij})_{1 \leqq i,j \leqq 2}$	$q^{k\ell}(\pi)$	$(t_{ij})_{1 \leqq i,j \leqq 2}$	$q^{k\ell}(\pi)$
$\begin{pmatrix} a-1 & 1 \\ 0 & 0 \end{pmatrix}$	1	$\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$	4
$\begin{pmatrix} a-2 & 1 \\ 1 & 0 \end{pmatrix}$	2	$\begin{pmatrix} a-1 & 0 \\ 0 & 1 \end{pmatrix}$	5
$\begin{pmatrix} a-1 & 0 \\ 1 & 0 \end{pmatrix}$	3		

Since the πth coordinate of a vector in $V_k \cap V_\ell^{-1}$ depends only on its coarsening $\left|\pi(\Lambda_{k\ell})\right|$, a basis of $V_k \cap V_\ell^{-1}$ is given by the indicator vectors of all the possible values of this coarsening. Therefore, for each a, q, we define this indicator vector

$$\rho_{aq}^{k\ell}(\pi) = \chi\left\{ a^{k\ell}(\pi) = a,\ q^{k\ell}(\pi) = q \right\}. \tag{31}$$

Of course, for many pairs a, q, these are zero vectors. For example, $\rho_{aq}^{k\ell}$ is a nonzero vector only if

$$\max(0, k + \ell - n) \leqq a \leqq \min(k, \ell).$$

q can usually be anything from 1 to 5, except when $a = 0, 1, k, \ell$. After all this preparation, we are ready for the proof of Theorem 2.

Proof of Theorem 2. We have to show (23). The number of free parameters of the bi-decomposable log-linear model is $\dim(H)-1$, since only those vectors v in H are allowed for which $p=e^v$ is a probability distribution. By (28), we only need to determine the dimension of each subspace $F_k\cap F_\ell^{-1}$, which consists of the vectors of $V_k\cap V_\ell^{-1}$ which are orthogonal to both $\widetilde{V}_k$ and $\widetilde{V}_\ell^{-1}$.

Let $u=\sum_{a,q} c_{aq}\rho_{aq}^{k\ell}$ be an arbitrary vector in $V_k\cap V_\ell^{-1}$; we find when it is orthogonal to $\widetilde{V}_k$ and $\widetilde{V}_\ell^{-1}$. First take a vector $v(\pi)=\chi\bigl(\pi\{1\dots k\}=C\bigr)$ in the basis of $\widetilde{V}_k$, and introduce the notation $\mid C\cap\{1\dots\ell\}\mid = a$. With $h=(k-1)!(n-k)!$, the scalar product is calculated as

$$(u,v)=\begin{cases} c_{a1}(k-a)h+c_{a2}(a-1)h+c_{a5}h & \text{if } \ \ell\in C\\ c_{a3}ah+c_{a4}(k-a)h & \text{if } \ \ell\notin C.\end{cases}$$

Similarly, if $v(\pi)=\chi\bigl(\pi^{-1}\{1\dots\ell\}=D\bigr)$ is a basis vector of $\widetilde{V}_\ell^{-1}$, $\mid D\cap\{1\dots k\}\mid = a$, and $g=(\ell-1)!(n-\ell)!$, then

$$(u,v)=\begin{cases} c_{a3}(\ell-a)g+c_{a2}(a-1)g+c_{a5}g & \text{if } \ k\in D\\ c_{a1}ag+c_{a4}(\ell-a)g & \text{if } \ k\notin D.\end{cases}$$

Thus $F_k\cap F_\ell^{-1}$ consists of the linear combinations of those vectors $\sum_{q=1}^{5} c_{aq}\rho_{aq}^{k\ell}$ for which the above four linear combinations of the coefficients c_{aq} are zero. Of the four constraints on the coefficients, only three are linearly independent, so in most cases there are two linearly independent solutions for the five coefficients. The cases $a=0,1,\min(k,\ell)$ must be treated separately, it is readily seen that in the case $a=0$, the only solution is zero, while in the cases $a=1,\min(k,\ell)$ there is one non-zero solution. Let $N_a^{k\ell}$ denote the number of linearly independent solutions, that is $N_a^{k\ell}$ is either zero, one or two. The following vectors form an orthogonal basis of $F_k\cap F_\ell^{-1}$ (with the exception that some vectors may be $\mathbf{0}$):

$$\mu_{a1}^{k\ell}=-\rho_{a2}^{k\ell}+(a-1)\rho_{a5}^{k\ell}, \tag{32}$$

$$\mu_{a2}^{k\ell}=-(\ell-a)a\rho_{a1}^{k\ell}+(k-a)(\ell-a)\rho_{a2}^{k\ell}-(k-a)a\rho_{a3}^{k\ell} + a^2\rho_{a4}^{k\ell}+(k-a)(\ell-a)\rho_{a5}^{k\ell}$$

Finally, since

$$\sum_{k,\ell}\dim(F_k\cap F_\ell^{-1})=\sum_{i\geqq 1}\Bigl|\bigl\{(k,\ell):\ \dim(F_k\cap F_\ell^{-1})\geqq i\bigr\}\Bigr|,$$

to finish the proof of the theorem, it suffices to show that for $1 \leqq i \leqq n$

$$\left|\left\{ (k,\ell) : \dim\left(F_k \cap F_\ell^{-1}\right) \geqq i \right\}\right| = (n-i)^2.$$

To this end, let us find those k, ℓ, for which $\dim\left(F_k \cap F_\ell^{-1}\right) \geqq 2j+2$. This happens if among the quantities $N_a^{k\ell}$ there are either two 1's and at least j 2's, or one 1 and at least $(j+1)$ 2's.

The first of these cases occurs when $\ell + k \leqq n+1$ and $\min\{k,\ell\} \geqq j+2$, while the second case occurs when $\ell + k \geqq n+2$ and $\max\{k,\ell\} \leqq n-j-1$. But if $\ell + k \leqq n+1$ and $k, \ell \geqq j+2$, then $k, \ell \leqq n-j-1$ also holds. Similarly, if $\ell + k \geqq n+2$ and $k, \ell \leqq n-j-1$, then at the same time $k, \ell \geqq j+3 > j+2$. Therefore, $\dim\left(F_k \cap F_\ell^{-1}\right) \geqq 2j+2$ holds if and only if $j+2 \leqq k, \ell \leqq n-j-1$, and there are $\left[n-(2j+2)\right]^2$ such pairs.

Let us find those k, ℓ, for which $\dim\left(F_k \cap F_\ell^{-1}\right) \geqq 2j+1$. This happens if among the quantities $N_a^{k\ell}$ there are either two 1's and at least j 2's, or one 1 and at least j 2's.

The first of these cases occurs when $\ell + k \leqq n+1$ and $\min\{k,\ell\} \geqq j+2$, while the second case occurs when $\ell + k \geqq n+2$ and $\max\{k,\ell\} \leqq n-j$. But if $\ell + k \leqq n+1$ and $k, \ell \geqq j+2$, then $k, \ell \leqq n-j-1 < n-j$ also holds. Similarly, if $\ell + k \geqq n+2$ and $k, \ell \leqq n-j$, then at the same time $k, \ell \geqq j+2$. Therefore, $\dim\left(F_k \cap F_\ell^{-1}\right) \geqq 2j+1$ holds if and only if $j+2 \leqq k, \ell \leqq n-j$, and there are $\left[n-(2j+1)\right]^2$ such pairs. $\square$

REMARK 2. In (32), we found an orthogonal basis $\{\mathbf{1}, \mu_{ai}^{k\ell}\}$ of the space H. This orthogonality is convenient for finding the parameters corresponding to a bi-decomposable distribution. There exists a basis consisting of indicator vectors as well, as follows. For any k, ℓ, a, let $\nu_a^{k\ell} = \sum_{q=1}^{5} \rho_{aq}^{k\ell}$, where $\rho_{aq}^{k\ell}$ was defined in (31). That is, $\nu_a^{k\ell}(\pi) = 1$ if $\left|\pi\{1 \ldots k\} \cap \{1 \ldots \ell\}\right| = a$, otherwise $\nu_a^{k\ell}(\pi) = 0$. We claim that the following vectors, together with $\mathbf{1}$, form a basis of H:

$$\nu_a^{k\ell} : 1 \leqq k, \ell \leqq n-1, \quad \max(0, k+\ell-n) < a \leqq \min(k,\ell),$$

$$\rho_{a5}^{k\ell} : 1 \leqq k, \ell \leqq n-1, \quad \max(1, k+\ell-n) < a \leqq \min(k,\ell).$$

This statement can be proved by induction; we omit the somewhat lengthy calculations.

Let us apply our results to matchings. Notice that it was relatively easy to show that $\mathcal{B}$ is a log-linear model, and to find its generators. To determine the number of free parameters was more difficult. In the language of matchings, we found that a strictly positive distribution on matchings is bi-decomposable if and only if there exist parameters $\theta_{k\ell}(t)$ such that

$$P(\Delta=\delta)=\prod_{(k,\ell)\in[n]\times[n]}\theta_{k\ell}\big(\big|\pi_W(\Lambda_{k\ell})\big|\big).$$

The point we wish to make is that the statistics $\big|\pi_W(\Lambda_{k\ell})\big|=(t_{ij})$ can be calculated "locally" from the matching: one needs to determine (i) whether the edge $\big(k,\pi_W(k)\big)$ runs above, below, or on the edge (k,ℓ), (ii) whether the edge $\big(\pi_M(\ell),\ell\big)$ runs above, below, or on the edge (k,ℓ), and (iii) the number of edges in δ which cross the edge (k,ℓ). To explain (iii), notice that the number of edges in δ which cross the edge (k,ℓ) is $t_{13}+t_{31}$, and t_{11} can be calculated from this sum, with the additional knowledge of the matrix entries t_{12}, t_{21}, t_{22}, t_{23}, t_{32}. This shows that the distributions in the parametric family defined in Example 2 are bi-decomposable.

4. Estimation

In this section, we address two statistical issues briefly. The first is maximum likelihood estimation in the model $\mathcal{B}$. Secondly, we consider the case when we do not know the linear orders of men and women, i.e. these two orders are also unknown parameters.

Denote by $\delta_1,\dots,\delta_N$ an iid sample of complete matchings taken from a positive bi-decomposable distribution p_W. The maximum likelihood estimate of the true distribution p_W does not appear to have an explicit form in general, the likelihood function has to be maximized by numerical methods. A natural option is iterative proportional fitting (IPFP), which is generally used for maximum likelihood estimation of log-linear models in the setting of contingency tables. This algorithm is guaranteed to converge to the maximum likelihood estimate, if it exists. Denote the empirical distribution of Π_W by r_W. The maximum likelihood estimate of p_W is a distribution p^*_W $\in\mathcal{B}$ such that the distributions of the coarsenings $\big|\Pi_W(\Lambda_{k\ell})\big|$ under p^*_W are the same as under the empirical distribution r_W for all pairs k,ℓ. There is at most one such $p^*_W\in\mathcal{B}$. In some cases, the maximum likelihood estimate does not exist, because no distribution in $\mathcal{B}$ gives the same distribution of the coarsenings as the empirical distribution (we say that the sample contains structural zeros). In these cases, a suitable p^*_W can only be found in the closure $\mathrm{cl}\,(\mathcal{B})$.

The IPFP algorithm proceeds by cyclically fitting the distributions of the individual coarsenings $\big|\Pi_W(\Lambda_{k\ell})\big|$ to that observed in the sample. Starting from an arbitrary $p^1_W\in\mathcal{B}$ (say the uniform distribution), the tth iteration step calculates

$$p_W^{t+1}(\pi) = \frac{\sum_{\sigma\colon |\sigma(\Lambda_{k\ell})|=|\pi(\Lambda_{k\ell})|} r_W(\sigma)}{\sum_{\sigma\colon |\sigma(\Lambda_{k\ell})|=|\pi(\Lambda_{k\ell})|} p_W^t(\sigma)} p_W^t(\pi),$$

where the pair (k, ℓ) runs cyclically over all possible values.

In contrast, in the L-decomposable family, the maximum likelihood estimate can be given explicitly. This family is parametrized by the conditional probabilities (13), and the maximum likelihood estimate of these conditional probabilities is given by the corresponding conditional probabilities under the empirical distribution.

Turning to the issue of the unknown linear ordering of the two sets M and W, we may treat these orders as unknown discrete parameters. Suppose the men and women are labeled arbitrarily, and we have an iid sample $\delta_1, \dots, \delta_N$ of complete matchings, which is assumed to come from a bi-decomposable distribution under some relabelings σ_M, σ_W (reflecting the true order) of the men and women. Let us work once again with Π_W and p_W. If we relabel the elements of W according to a permutation σ_W, and the elements of M according to a permutation σ_M, then we get the new random permutation $\Pi'_W = \sigma_W \Pi_W \sigma_M^{-1}$, and the new distribution $p'_W(\pi) = p_W(\sigma_W^{-1}\pi\sigma_M)$. Then one proceeds by taking all $(n!)^2$ possible relabelings σ_M, σ_W, and calculating the maximum likelihood

$$L_{\max}(\sigma_M, \sigma_W) = \max_{q\in\mathcal{B}} L(q, \sigma_M, \sigma_W) = \max_{q\in\mathcal{B}} \prod_{i=1}^{N} q(\sigma_W^{-1}\pi_W^i\sigma_M)$$

using IPFP, where π_W^i is obtained from the ith element of the sample, δ_i. Then one chooses the pair (σ_M, σ_W) for which $L_{\max}(\sigma_M, \sigma_W)$ is the largest. We show that this pair is never unique, in general there are $8 \cdot 8$ pairs of relabelings which yield the same maximum likelihood $L_{\max}(\sigma_M, \sigma_W)$. In particular, we show that if Δ is a bi-decomposable random matching, then it remains bi-decomposable if we interchange the labels of the first two men (or women), or if we reverse the whole labeling of the men (or women), but there is no other permutation of the labels (except for compositions of the above mentioned two), which preserves bi-decomposability in general.

For a distribution p on S_n and a $\sigma \in S_n$, let

$$p_{\circ\sigma}(\pi) = p(\pi\sigma), \quad p_{\sigma\circ}(\pi) = p(\sigma\pi).$$

Denote by $\sigma_{(12)}$ the permutation which exchanges 1 and 2 only, and by σ_r the reversing permutation which maps k to $n+1-k$. Denote the group generated by these permutations as

$$\mathcal{G} = \left\{\text{id},\ \sigma_r,\ \sigma_{(12)},\ \sigma_r\sigma_{(12)},\ \sigma_r\sigma_{(12)}\sigma_r,\ \sigma_{(12)}\sigma_r,\ \sigma_r\sigma_{(12)}\sigma_r\sigma_{(12)},\ \sigma_{(12)}\sigma_r\sigma_{(12)}\right\}.$$

Then we have the following theorem.

THEOREM 3. *Suppose p is L-decomposable. Then $p_{\sigma\circ}$ is L-decomposable for all $\sigma \in S_n$ and $p_{\circ\sigma}$ is L-decomposable for all $\sigma \in \mathcal{G}$. However, for any $\sigma \notin \mathcal{G}$, there exists a $p \in \mathcal{B}$ such that $p_{\circ\sigma}$ is not L-decomposable.*

PROOF. The first two invariances can be checked directly from the definitions. To prove the last statement, we will use the following property: let p be L-decomposable. Suppose that the probability of π_{11} and π_{22} is positive, and a is such that $\pi_{11}\{1\dots a\} = \pi_{22}\{1\dots a\}$. Define the "crossover" permutations:

$$\pi_{12}(k) = \begin{cases} \pi_{11}(k) & \text{if } \ k \leqq a \\ \pi_{22}(k) & \text{if } \ k > a \end{cases}, \qquad \pi_{21}(k) = \begin{cases} \pi_{22}(k) & \text{if } \ k \leqq a \\ \pi_{11}(k) & \text{if } \ k > a \end{cases}.$$

Then

$$p(\pi_{11})p(\pi_{22}) = p(\pi_{12})p(\pi_{21}). \tag{33}$$

If σ is not a member of the permutations in $\mathcal{G}$, then neither is its inverse, and there exists an $2 \leqq a \leqq n-2$, such that

$$\sigma^{-1}\{1\dots a\} \neq \{1\dots a\}, \{n-a+1\dots n\}.$$

Let a be such a number. Therefore there exist $c, e \in \{1\dots a\}$ and $d, f \notin \{1\dots a\}$, for which

$$c^* = \sigma^{-1}(c) > \sigma^{-1}(d) = d^*, \quad e^* = \sigma^{-1}(e) < \sigma^{-1}(f) = f^*.$$

For the numbers α, β, γ we say that α separates β and γ if $\beta < \alpha < \gamma$ or $\beta > \alpha > \gamma$. Now, if $d^* \geqq f^*$, then d^* (and f^* as well) separates c^* and e^*. If $d^* < f^*$, then either one of them separates c^* and e^*, or c^* (and e^* as well) separates d^* and f^*. Therefore, one of the following two cases holds:

1. $\exists\, c, e \in \{1\dots a\}, \quad d \notin \{1\dots a\} :$ d^* separates c^*, e^*
2. $\exists\, c \in \{1\dots a\}, \quad d, f \notin \{1\dots a\} :$ c^* separates d^*, f^*

The two cases can be treated in the same way. Let us deal with the first one. Let $f \notin \{1\dots a\}$, $f \neq d$ be arbitrary, with $f^* = \sigma^{-1}(f)$. Recall (31), and let $p = K(d^*)\exp\left\{\rho^{d^*d^*}_{d^*5}\right\}$, this is a positive bi-decomposable distribution. Let $\pi_{11} = \sigma^{-1}$, from which we obtain π_{22} by exchanging two pairs:

$$\pi_{22}(c) = e^*, \ \pi_{22}(e) = c^*, \ \pi_{22}(d) = f^*, \ \pi_{22}(f) = d^*.$$

Denote by π_{12} and π_{21} the crossover permutations. For these four permutations, $p_{\circ\sigma}$ does not satisfy (33). On the one hand, multiplying π_{11} by σ from the right, we get the identity permutation, for which $\rho^{d^*d^*}_{d^*5} = 1$. On the other hand, for both $\pi_{12}\sigma$ and $\pi_{21}\sigma$, $\rho^{d^*d^*}_{d^*5} = 0$, since for the first, d^* is not a fixed point, and for the second, there is an element greater than d^* among the first d^* elements. This completes the proof. □

Thus we have shown that the bi-decomposability of Π_W implies the bi-decomposability of $\Pi'_W = \sigma\Pi_W\rho^{-1}$ if and only if $\sigma, \rho \in \mathcal{G}$.

5. Proofs

PROOF OF PROPOSITION 2. If the stated conditional independences hold, the random matching Δ is clearly bi-decomposable. For the other direction, suppose Δ is bi-decomposable. By L-decomposability of Π_W, given $\left|\Pi_W(\mathcal{M})\right|$, the ordered restrictions $\Pi_W(M_i)$ are conditionally independent. Conditioning on $\left|\Pi_M(\mathcal{W})\right|$ as well does not ruin this independence, since the additional condition restricts the values of the $\Pi_W(M_i)$ one by one. Thus we proved that $\Pi_W(M_i)$ are conditionally independent, given $\left|\Pi_W(\mathcal{M})\right|$ and $\left|\Pi_M(\mathcal{W})\right|$. Using the L-decomposability of Π_M, the same conditional independence is true for the $\Pi_M(W_j)$. Let A be an atom in the σ-algebra generated by $\left(\left|\Pi_W(\mathcal{M})\right|, \left|\Pi_M(\mathcal{W})\right|\right)$, whose probability is positive, and let δ be a matching. Then

$$P(\Delta = \delta \mid A) = \prod_{i=1}^{d} P\big(\Pi_W(M_i) = \pi_W(M_i) \mid A\big),$$

and since

$$\Pi_W(M_i) = \big(\Delta(M_i \times W_j) : 1 \leqq j \leqq r\big),$$

where $\Delta(M_i \times W_j)$ is a function of $\Pi_M(W_j)$, also

$$P\big(\Pi_W(M_i) = \pi_W(M_i) \mid A\big) = \prod_{j=1}^{r} P\big(\Delta(M_i \times W_j) = \delta(M_i \times W_j) \mid A\big),$$

which proves the lemma. $\square$

PROOF OF LEMMA 3. Since $L_2(\mathcal{D}_1 \cap \mathcal{D}_2) = L_2(\mathcal{D}_1) \cap L_2(\mathcal{D}_2)$, the spaces $L_2(\mathcal{D}_1)$ and $L_2(\mathcal{D}_2)$ intersect orthogonally if and only if for any $f \in L_2(\mathcal{D}_1)$, $g \in L_2(\mathcal{D}_2)$,

$$E\big(\big[f - E(f \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big[g - E(g \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big) = 0.$$

If the conditional independence relation holds, then the following stronger equality holds:

$$E\big(\big[f - E(f \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big[g - E(g \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big] \mid \mathcal{D}_1 \cap \mathcal{D}_2\big) = 0.$$

In the other direction, if the spaces intersect orthogonally, then let $A_1 \in \mathcal{D}_1$, $A_2 \in \mathcal{D}_2$, and denote by C the event that

$$P(A_1 \cap A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2) - P(A_1 \mid \mathcal{D}_1 \cap \mathcal{D}_2)P(A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2) > 0.$$

With $f = \chi(A_1)\chi(C)$ and $g = \chi(A_2)\chi(C)$,

$$E\big(\big[f - E(f \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big[g - E(g \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big)$$

$$= E\big(E(fg \mid \mathcal{D}_1 \cap \mathcal{D}_2) - E(f \mid \mathcal{D}_1 \cap \mathcal{D}_2)E(g \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big)$$

$$= E\big(\chi(C)\big[P(A_1 \cap A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2) - P(A_1 \mid \mathcal{D}_1 \cap \mathcal{D}_2)P(A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2)\big]\big) = 0.$$

This is possible only if $P(A_1 \cap A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2) - P(A_1 \mid \mathcal{D}_1 \cap \mathcal{D}_2)P(A_2 \mid \mathcal{D}_1 \cap \mathcal{D}_2) \leqq 0$ with probability 1. The reverse inequality is obtained similarly, thus A_1 and A_2 are conditionally independent, given $\mathcal{D}_1 \cap \mathcal{D}_2$. □

Proof of Lemma 4. We apply Lemma 3 to S_n endowed with the uniform distribution. In this case, orthogonal intersection in the L_2-space is equivalent to orthogonal intersection in $\mathbb{R}^{n!}$. The second statement in (18) holds, because it is easy to check that $\sigma(\mathcal{D}' \times \mathcal{R}) \cap \sigma(\mathcal{D} \times \mathcal{R}') = \sigma(\mathcal{D} \times \mathcal{R})$. Concerning the first one, we have to prove that $\big|\Pi(\mathcal{D}' \times \mathcal{R})\big|$ and $\big|\Pi(\mathcal{D} \times \mathcal{R}')\big|$ are conditionally independent, given $\big|\Pi(\mathcal{D} \times \mathcal{R})\big|$ if Π is a uniformly distributed random permutation, which is again easy to check. □

Proof of Lemma 1. By supposition, $\Pr_{V_j} U_i \subset U_i$ for every i, j, therefore $\Pr_{V_j} U \subset U$ for every j, i.e. U intersects each V_j orthogonally. Consequently, $\Pr_U V_j \subset V_j$ for every j, which yields $\Pr_U V \subset V$, which was to be proved. On the other hand, let $Z = \operatorname{Span}(U_i \cap V_j : i \in I,\ j \in J)$. Then $\Pr_{V_j} U_i \subset Z$, furthermore $\Pr_U V_j = \Pr_{V_j} U \subset Z$, which leads to $\Pr_U V \subset Z$. □

Proof of Lemma 2. We use that U and V intersect orthogonally if and only if the projection operators onto them commute, that is $\Pr_U \Pr_V = \Pr_V \Pr_U$. Now

$$\Pr_{U_2} \Pr_{V_2} = (\Pr_U - \Pr_{U_1})(\Pr_V - \Pr_{V_1})$$

$$= \Pr_U \Pr_V - \Pr_{U_1} \Pr_V - \Pr_U \Pr_{V_1} + \Pr_{U_1} \Pr_{V_1},$$

and by supposition the operators in all four terms commute, yielding $\Pr_{V_2} \Pr_{U_2}$. The second statement follows from Lemma 1. □

Proof of Lemma 5. We use again that orthogonality in the L_2-space of random variables $f : S_n \to \mathbb{R}$, where S_n is endowed with the probability measure $P(\pi) = 1/n!$, is equivalent to orthogonality of vectors f in $\mathbb{R}^{n!}$. We denote by

$$\sigma_k = \sigma\big(\pi \mapsto \big|\pi(\Sigma_k \times \Phi)\big|\big), \quad \widetilde{\sigma}_k = \sigma\big(\pi \mapsto \big|\pi(\widetilde{\Sigma}_k \times \Phi)\big|\big)$$

the σ-algebras of subsets of S_n generated by the random variables in parentheses. An element of F_k is a difference $f_1 = f - E(f \mid \widetilde{\sigma}_k)$, where f is σ_k-measurable. Orthogonality to **1** means that $E(f_1) = 0$. For the other statement, let g_1 be an element of F_j, where $j > k$. It is easy to check that under the uniform distribution, σ_k and σ_j are conditionally independent, given $\widetilde{\sigma}_k$. Therefore,

$$E(f_1 g_1) = E\big[E(f_1 g_1 \mid \widetilde{\sigma}_k)\big] = E\big[E(f_1 \mid \widetilde{\sigma}_k)E(g_1 \mid \widetilde{\sigma}_k)\big] = 0,$$

since $E(f_1 \mid \widetilde{\sigma}_k) = 0$ with probability 1. $\square$

Acknowledgement. I would like to thank Gábor Tusnády for many valuable discussions.

References

[1] D. E. Critchlow, M. A. Fligner and J. S. Verducci, Probability models on rankings, *J. Math. Psych.,* **35** (1991), 294–318.

[2] D. Geiger, C. Meek and B. Sturmfels, On the toric algebra of graphical models, *Ann. Statist.,* **34** (2006), 1463–1492.

[3] S. Lauritzen, *Graphical Models,* Clarendon Press (Oxford, 1996).

[4] R. D. Luce, *Individual Choice Behavior,* Wiley (New York, 1959).

[5] F. Rapallo, Toric statistical models: parametric and binomial representations, *Annals of the Institute of Statistical Mathematics,* Online First (2006).

[6] S. Sullivant, *Toric Ideals in Algebraic Statistics,* PhD thesis, University of California, Berkeley (2005).