CONCEPTUAL 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 $n \geq 4$). 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
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 \([n] = \{1, \ldots, n\}\): 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 \(i\)th man and the \(j\)th 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 \(\delta\) 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 \(\pi = (\pi(1), \pi(2), \ldots, \pi(n))\). Then \(\pi_W\) and \(\pi_M\) are given by

\[
\delta = \{(i, \pi_W(i)) : 1 \leq i \leq n\} = \{(\pi_M(i), i) : 1 \leq i \leq n\}.
\]

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

Our aim is to define and study meaningful probability distributions on random complete matchings \(\Delta\), equivalently on random permutations \(\Pi_W\) or \(\Pi_M\). 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 \(p = (p(\pi) : \pi \in S_n)\) where the numbers \(p(\pi)\) are nonnegative, and their sum is one. If \(\Delta\) is a random matching, we denote by \(p_W\) (resp. \(p_M\)) the distribution of \(\Pi_W\) (resp. \(\Pi_M\)) 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 \(\{\Pi_W(1), \Pi_W(2)\} = \{1, 2\}\). Denoting the distribution of \(\Pi_W\) by \(p_W\), this is easily seen to be equivalent to the relation

\[
p_W(1234)p_W(2134) = p_W(1243)p_W(2314).
\]

\[132\]
V. Csiszár

*Acta Mathematica Hungarica* 122, 2009