

## VERIFYING THE INDEPENDENCE OF PARTITIONS OF A PROBABILITY SPACE

S.B. MULAY AND C.G. WAGNER

Let  $\{E_1, \dots, E_r\}$  and  $\{F_1, \dots, F_s\}$  be partitions of a probability space. We exhibit a natural bijection from the set of efficient ways of verifying the independence of such partitions to the set of spanning trees of the complete bipartite graph  $K_{r,s}$ .

### 1. INTRODUCTION

In what follows,  $(\Omega, \Sigma, p)$  is a probability space and  $[n] := \{1, 2, \dots, n\}$ . Partitions  $\{E_i : i \in [r]\}$  and  $\{F_j : j \in [s]\}$  of  $\Omega$ , with  $E_i, F_j \in \Sigma$ , are said to be *independent* (with respect to  $p$ ) if

$$(1) \quad p(E_i \cap F_j) = p(E_i)p(F_j)$$

for all  $(i, j) \in [r] \times [s]$ . Of course, one need not check all  $rs$  instances of (1) in order to verify independence. It is easy to see, for example, that if (1) holds for all  $(i, j) \in [r-1] \times [s-1]$ , then the partitions in question are independent. Let us call a subset  $\mathcal{N}$  of  $[r] \times [s]$  *negligible* when, if (1) holds for all  $(i, j) \in \mathcal{N}^c$ , then it holds for all  $(i, j) \in \mathcal{N}$  as well. We show in this note that there is a natural bijection from the family of maximal negligible subsets of  $[r] \times [s]$  to the family of spanning trees of the complete bipartite graph  $K_{r,s}$ . It follows that there are  $r^{s-1}s^{r-1}$  efficient ways to verify the independence of the aforementioned partitions.

### 2. NEGLIGIBILITY AND LINEAR INDEPENDENCE

For all  $i \in [r]$ , let  $X_i$  be the  $(r+s)$ -dimensional unit column vector with a one in the  $i$ th position and zeros elsewhere, and for all  $j \in [s]$ , let  $Y_j$  be the  $(r+s)$ -dimensional unit column vector with a one in the  $(r+j)$ th position and zeros elsewhere. Then

$$(2) \quad \sum_{(i,j) \in [r] \times [s]} p(E_i \cap F_j)(X_i + Y_j) = \sum_{i \in [r]} p(E_i)X_i + \sum_{j \in [s]} p(F_j)Y_j,$$

---

Received 31st May, 1999

Research supported by the National Science Foundation (SBR-9528893).

---

Copyright Clearance Centre, Inc. Serial-fee code: 0004-9727/00 \$A2.00+0.00.

as a consequence of the familiar formulas for the marginal probabilities  $p(E_i)$  and  $p(F_j)$ .

Suppose that  $\mathcal{N} \subset [r] \times [s]$ , and that  $p(E_i \cap F_j) = p(E_i)p(F_j)$  for all  $(i, j) \in \mathcal{N}^c$ . Then (2) becomes

$$(3) \quad \sum_{(i,j) \in \mathcal{N}} p(E_i \cap F_j)(X_i + Y_j) = \sum_{i \in [r]} p(E_i)X_i + \sum_{j \in [s]} p(F_j)Y_j - \sum_{(i,j) \in \mathcal{N}^c} p(E_i)p(F_j)(X_i + Y_j).$$

Regard the quantities  $p(E_i \cap F_j)$ , where  $(i, j) \in \mathcal{N}$ , as unknowns. It is clear that  $p(E_i \cap F_j) = p(E_i)p(F_j)$  for all  $(i, j) \in \mathcal{N}$  furnishes a solution of (3).  $\mathcal{N}$  is negligible if and only if this is the only solution of (3), and the latter condition clearly obtains if and only if  $\{X_i + Y_j : (i, j) \in \mathcal{N}\}$  is a linearly independent subset of  $V$ , the subspace of  $\mathbb{R}^{r+s}$  spanned by  $\{X_i + Y_j : (i, j) \in [r] \times [s]\}$ . Consequently,  $\mathcal{N}$  is a maximal negligible subset of  $[r] \times [s]$  if and only if  $\{X_i + Y_j : (i, j) \in \mathcal{N}\}$  is a basis of  $V$ .

It is easy to see that the dimension of  $V$  is  $r+s-1$ . In particular, the set of column vectors  $\{X_1 + Y_s, X_2 + Y_s, \dots, X_{r-1} + Y_s, X_r + Y_1, X_r + Y_2, \dots, X_r + Y_s\}$  is a basis of  $V$ . This set spans  $V$  since  $X_i + Y_j = (X_i + Y_s) + (X_r + Y_j) - (X_r + Y_s)$ . It is linearly independent as a simple consequence of the linear independence of  $\{X_1, \dots, X_r, Y_1, \dots, Y_s\}$ . In the next section we present a graphical characterisation of bases of  $V$  consisting of vectors of the form  $X_i + Y_j$ .

### 3. A NATURAL BIJECTION

We assume in this section familiarity with the basic terminology and elementary results of graph theory, as described, for example, in [3]. In particular, we use the fact that if all vertices of a graph have degree at least two, then the graph contains a cycle [3, Lemma 1.2.18], and the fact that a graph with  $n$  vertices is a tree if and only if it is acyclic and has  $n - 1$  edges [3, Theorem 2.13].

Consider the complete bipartite graph  $K_{r,s}$  with vertex set  $\mathcal{V} = \{X_1, \dots, X_r, Y_1, \dots, Y_s\}$  and edge set  $\mathcal{E} = \{\{X_i, Y_j\} : (i, j) \in [r] \times [s]\}$ . To each  $\mathcal{S} \subset [r] \times [s]$  we associate the subgraph of  $K_{r,s}$  having vertex set  $\mathcal{V}(\mathcal{S}) = \bigcup_{(i,j) \in \mathcal{S}} \{X_i, Y_j\}$  and edge set  $\mathcal{E}(\mathcal{S}) = \{\{X_i, Y_j\}; (i, j) \in \mathcal{S}\}$ . The map  $\mathcal{S} \mapsto (\mathcal{V}(\mathcal{S}), \mathcal{E}(\mathcal{S}))$  is clearly an injection from  $2^{[r] \times [s]}$  into the set of all subgraphs of  $K_{r,s}$ .

**LEMMA.** *The set of column vectors  $S(\mathcal{S}) = \{X_i + Y_j : (i, j) \in \mathcal{S}\}$  is linearly dependent if and only if  $(\mathcal{V}(\mathcal{S}), \mathcal{E}(\mathcal{S}))$  contains a cycle.*

**PROOF:** Sufficiency. Suppose, with no loss of generality, that  $(\mathcal{V}(\mathcal{S}), \mathcal{E}(\mathcal{S}))$  contains the cycle  $X_{i_1}, Y_{i_1}, X_{i_2}, Y_{i_2}, \dots, X_{i_n}, Y_{i_n}, X_{i_1}$ . Since then  $X_{i_1} + Y_{i_1}, X_{i_2} + Y_{i_1}, X_{i_2} +$

$Y_{i_2}, \dots, X_{i_n} + Y_{i_n}$  and  $X_{i_1} + Y_{i_n} \in S(\mathcal{S})$  and  $(X_{i_1} + Y_{i_1}) - (X_{i_2} + Y_{i_1}) + (X_{i_2} + Y_{i_2}) - \dots + (X_{i_n} + Y_{i_n}) - (X_{i_1} + Y_{i_n}) = 0$ , it follows that  $S(\mathcal{S})$  is linearly dependent.

Necessity. Suppose that  $S(\mathcal{S})$  is linearly dependent. Then there exists a nonempty subset  $\mathcal{S}^+$  of  $\mathcal{S}$  and, for each  $(i, j) \in \mathcal{S}^+$ , a nonzero real number  $\alpha_{ij}$  such that

$$(4) \quad \sum_{(i,j) \in \mathcal{S}^+} \alpha_{ij}(X_i + Y_j) = 0.$$

Consider the graph  $(\mathcal{V}(\mathcal{S}^+), \mathcal{E}(\mathcal{S}^+))$ . Clearly, every vertex in  $\mathcal{V}(\mathcal{S}^+)$  has degree at least one. Suppose some vertex, say  $X_{i^*}$ , has degree one, belonging only to the edge  $\{X_{i^*}, Y_{j^*}\}$ . Then  $X_{i^*}$  occurs just once in (4), with the nonzero coefficient  $\alpha_{i^*j^*}$ . This implies that  $X_{i^*}$  is a linear combination of  $\{X_1, \dots, X_r, Y_1, \dots, Y_s\} \setminus \{X_{i^*}\}$ , contradicting the linear independence of  $\{X_1, \dots, X_r, Y_1, \dots, Y_s\}$ . Hence every vertex in  $\mathcal{V}(\mathcal{S}^+)$  has degree at least two, and so  $(\mathcal{V}(\mathcal{S}^+), \mathcal{E}(\mathcal{S}^+))$ , and thus  $(\mathcal{V}(\mathcal{S}), \mathcal{E}(\mathcal{S}))$ , contains a cycle.  $\square$

Students of matroid theory will not be surprised by the above lemma. Indeed, it establishes a special case of a much more general result, namely the fact that the cycle matroid of every graph has a vectorial representation [2, Section 9.5]. We may now establish the main result of this note.

**THEOREM.** *The map  $\mathcal{N} \mapsto (\mathcal{V}(\mathcal{N}), \mathcal{E}(\mathcal{N}))$  is a bijection from the family of all maximal negligible subsets of  $[r] \times [s]$  to the set of all spanning trees of  $K_{r,s}$ .*

PROOF: If  $\mathcal{N}$  is a maximal negligible subset of  $[r] \times [s]$ , then, as shown in Section 2 above,  $S(\mathcal{N}) = \{X_i + Y_j : (i, j) \in \mathcal{N}\}$  is a basis of  $V$ , the  $(r + s - 1)$ -dimensional subspace of  $\mathbb{R}^{r+s}$  spanned by  $\{X_i + Y_j : (i, j) \in [r] \times [s]\}$ . By the preceding lemma,  $(\mathcal{V}(\mathcal{N}), \mathcal{E}(\mathcal{N}))$  is acyclic. Clearly,  $|\mathcal{E}(\mathcal{N})| = |S(\mathcal{N})| = r + s - 1$ . Also,  $\mathcal{V}(\mathcal{N}) = \{X_1, \dots, X_r, Y_1, \dots, Y_s\}$ , for if not,  $S(\mathcal{N})$  would not span  $V$ . Hence,  $|\mathcal{V}(\mathcal{N})| = r + s$ . It follows that  $(\mathcal{V}(\mathcal{N}), \mathcal{E}(\mathcal{N}))$  is a tree with the same vertex set as  $K_{r,s}$ , and edge set contained in the edge set of  $K_{r,s}$ , that is, a spanning tree of  $K_{r,s}$ .

As a restriction of the injective map  $\mathcal{S} \mapsto (\mathcal{V}(\mathcal{S}), \mathcal{E}(\mathcal{S}))$ , the map  $\mathcal{N} \mapsto (\mathcal{V}(\mathcal{N}), \mathcal{E}(\mathcal{N}))$  is injective. It remains only to show that this map is surjective. Let  $(\mathcal{V}, \mathcal{E})$  be a spanning tree of  $K_{r,s}$ , so that  $\mathcal{V} = \{X_1, \dots, X_r, Y_1, \dots, Y_s\}$  and  $\mathcal{E} \subset \{\{X_i, Y_j\} : (i, j) \in [r] \times [s]\}$ . Then  $|\mathcal{E}| = r + s - 1$ . Suppose that  $\mathcal{E} = \{\{X_i, Y_j\} : (i, j) \in \mathcal{N}\}$  where  $\mathcal{N} \subset [r] \times [s]$ . Clearly,  $\mathcal{V}(\mathcal{N}) = \mathcal{V}$  and  $\mathcal{E}(\mathcal{N}) = \mathcal{E}$ . By the lemma,  $S(\mathcal{N}) = \{X_i + Y_j : (i, j) \in \mathcal{N}\}$  is linearly independent since  $(\mathcal{V}(\mathcal{N}), \mathcal{E}(\mathcal{N}))$  is acyclic. Since  $|S(\mathcal{N})| = |\mathcal{N}| = |\mathcal{E}| = r + s - 1$ ,  $S(\mathcal{N})$  is a basis of  $V$ . Hence by the results of Section 2 above,  $\mathcal{N}$  is a maximal negligible subset of  $[r] \times [s]$ , which completes the proof of surjectivity.  $\square$

Since the complete bipartite graph  $K_{r,s}$  has  $r^{s-1}s^{r-1}$  spanning trees [1], it follows that there are  $r^{s-1}s^{r-1}$  efficient ways to verify the independence of partitions

$\{E_1, \dots, E_r\}$  and  $\{F_1, \dots, F_s\}$ .

REMARK 1. The foregoing analysis could have been carried out with  $\{X_1, \dots, X_r, Y_1, \dots, Y_s\}$  being any set of distinct indeterminates over  $\mathbb{R}$ .

REMARK 2. Our characterisation in Section 2 above of the efficient ways of verifying the independence of two partitions of a probability space may be generalised to the case of three or more partitions. In the case of partitions  $\{E_1, \dots, E_r\}$ ,  $\{F_1, \dots, F_s\}$ , and  $\{G_1, \dots, G_t\}$ , for example, maximal negligible subsets of  $[r] \times [s] \times [t]$  correspond to bases of the vector space generated by  $\{X_i + Y_j + Z_k : (i, j, k) \in [r] \times [s] \times [t]\}$  comprised of vectors of the form  $X_i + Y_j + Z_k$ . The problem of enumerating bases of this type has, as far as we know, not been solved. The vectors comprising such bases correspond in a natural way to edges of a hypergraph, but it is not clear what sorts of hypergraphs arise in this way, or whether they facilitate the enumeration in question.

#### REFERENCES

- [1] H. Scoins, 'The number of trees with nodes of alternate parity', *Proc. Cambridge Philos. Soc.* **58** (1962), 12–16.
- [2] D. Welsh, *Matroid theory* (Academic Press, London, 1976).
- [3] D. West, *Introduction to graph theory* (Prentice Hall, Upper Saddle River, NJ, 1996).

Department of Mathematics  
The University of Tennessee  
Knoxville, TN 37996-1300  
United States of America