NEGATIVE DEPENDENCE AND SRINIVASAN'S SAMPLING PROCESS

JOSH BROWN KRAMER, JONATHAN CUTLER, AND A.J. RADCLIFFE

Abstract. In [1] Dubhashi, Jonasson, and Ranjan study the negative dependence properties of Srinivasan's sampling processes (SSPs), random processes which sample sets of a fixed size with prescribed marginals. In particular they prove that linear SSPs have conditional negative association, by using the Feder-Mihail theorem [3] and a coupling argument. We consider a broader class of SSPs that we call tournament SSPs (TSSPs). These have a tree-like structure and we prove that they have conditional negative association. Our approach is completely different from that of Dubhashi, Jonasson, and Ranjan. We give an abstract characterization of TSSPs, and use this to deduce that certain conditioned TSSPs are themselves TSSPs. We show that TSSPs have negative association, and hence conditional negative association. We also give an example of an SSP that does not have negative association.

1. Introduction

The theory of negative dependence has been a subject of great interest for mathematicians of late. Intuitively, a sequence of random variables is negatively dependent if the event that some subset of them are "large" tends to make the values of other variables "small." In [6], Pemantle calls for a general theory of negative dependence corresponding to that of positive dependence. One step towards this goal had already been achieved by Reimer [7] in proving the van den Berg-Kesten conjecture [10].

The examples of negative dependence properties with which this paper is concerned are that of negative association and conditional negative association (defined below). While these concepts were introduced by Joag-Dev and Proschan [5], they were studied earlier by Feder and Mihail [3] in the context of balanced matroids. Since then negative association has been well-studied, see, for example, [2], [4], and [8]. In the definition that follows, note that negative association is a much stronger condition than pairwise negative correlation.

Definition 1. Given a sequence $A = (A_x)_{x\in I}$ of random variables we write A_J for the subsequence $(A_x)_{x\in J}$. We say that A is *negatively associated* if for every pair of disjoint subsets $J, K \subseteq I$ and all nondecreasing functions f, g,

$$\mathbb{E}\left[f(A_J)g(A_K)\right] \le \mathbb{E}\left[f(A_J)\right]\mathbb{E}\left[g(A_K)\right].$$

The variables $(A_x)_{x\in I}$ are *conditionally negatively associated* if for any subset $J \subseteq I$ and sequence $a = (a_y)_{y\in J}$ the sequence $(A_{I\setminus J} \mid A_J = a)$ is negatively associated.

In this paper, we will be concerned with the negative association and conditional negative association of Srinivasan's sampling process (SSP), a method of producing a random k-subset of an n-element set [9]. A random k-subset of an n-element set can be thought of as a sequence of binary random variables $A = (A_x)_{x\in I}$, where $|I| = n$, in which $A_x = 1$ if x is in the k-set and $A_x = 0$ if not. Since we are choosing a k-subset, we have $\sum_{x\in I} A_x = k$. We freely switch between saying $x \in A$ and $A_x = 1$. One simple example of such a process is *uniform sampling of k-subsets*, where k-subsets are chosen uniformly from all subsets of the n-element set of size k. In applications,

Date: 12/3/09.

including integer linear programming, it is desirable to be able to prescribe the marginals, i.e., $\mathbb{P}(A_x = 1) = p_x$ where $0 \le p_x \le 1$ for all $x \in I$. This rules out uniform sampling. One way to address this issue is with rejection sampling. We can pick independent samples from I according to given marginals and reject samples not having size k. In practice, however, this method is slow.

Srinivasan's sampling process was introduced as a method of producing random k-sets from an n-element set quickly and with given marginals. The question of whether SSPs have negative association and conditional negative association was considered by Dubhashi, Jonasson, and Ranjan [1].

Before we define SSPs, we introduce some results related to negative association and conditional negative association. Feder and Mihail [3] gave a somewhat surprising sufficient condition for CNA, involving the notion of variables of positive influence, defined as follows.

Definition 2. Let $A = (A_x)_{x\in I}$ be real-valued random variables and F be an A-measurable random variable (i.e., some function of the A_x). Then we say A_y is a *variable of positive influence* for F if

$$\mathbb{E}\left[F \mid A_y = t\right]$$

is a nondecreasing function of t.

Feder and Mihail showed that the relatively weak property of conditional pairwise negative correlation along with the existence of a variable of positive influence, gives conditional negative association.

Theorem 1 (Feder and Mihail [3])**.** *Let $(A_x)_{x\in I}$ be binary random variables such that for any $J \subseteq I$, and any $a = (a_y)_{y\in J}$, the random sequence $(B_x)_{x\in I\setminus J} = (A_x \mid A_J = a)_{x\in I\setminus J}$ satisfies the following:*

- *Every nondecreasing B-measurable F has a variable of positive influence;*
- *The B_x are pairwise negatively correlated.*

Then the variables $(A_x)_{x\in I}$ are conditionally negatively associated.

Thus, when studying random k-sets, it is enough to show that they have conditional pairwise negative correlation in order to show that they have CNA. In [1], a rather complicated coupling argument is used to show this in the special case of so-called *linear SSPs*. In this paper we discuss a broader class of SSPs called *tournament SSPs*. We show that this class of SSPs can be described by a slightly different random process that we call a *tournament sample*. This allows us to give an abstract characterization of tournament SSPs, and use this characterization to show that certain conditioned tournament SSPs are, in fact, tournament SSPs themselves. This in turn allows us to deduce conditional negative association for tournament SSPs directly from the much simpler fact that they have negative association, bypassing the Feder-Mihail theorem.

Further, we disprove a conjecture of Dubhashi, Jonasson and Ranjan by exhibiting an SSP that does not even have negative association. This is in fact a counterexample to Theorem 5.1 in their paper. The proof they give is correct for linear SSPs, but does not apply, as they claim, to arbitrary SSPs.

2. Srinivasan's sampling process

Given a finite index set I, Srinivasan's sampling process [9] is a probability distribution on $\binom{I}{k}$ (the set of all k-subsets of I) for some $0 \le k \le |I|$. It is determined by a sequence $(p_x)_{x\in I}$ of probabilities satisfying $\sum_{x\in I} p_x = k$, together with a total ordering (which we refer to as the "match ordering") on $\binom{I}{2}$. If A is a random variable with this distribution then for every $x \in I$ we have $\mathbb{P}(x \in A) = p_x$. From another perspective we are defining binary random variables $(A_x)_{x\in I}$ such that $\mathbb{P}(A_x = 1) = p_x$ and $|\{x \in I : A_x = 1\}|$ is always k.

In outline the distribution is defined iteratively as follows: we initialize variables $(w_x)_{x\in I}$ to $w_x := p_x$. For each pair $\{x, y\}$ in turn we look at the current values of w_x and w_y. We play a "match" between x and y and, based on the (random) outcome of the match, we change the values of w_x and w_y in such a fashion that one of them is set to either 0 or 1 and their sum remains constant. We carefully ensure that during this procedure any value w_x that is currently set to 0 or 1 is never subsequently changed. Thus at the end of the procedure all of the w_x are either 0 or 1 (and exactly k of the w_x are 1). We then let the random variables A_i be defined by $A_x = w_x$. Our process also guarantees that $\mathbb{P}(A_x = 1) = p_x$. A careful definition is best given inductively.

Definition 3. Let I be a finite index set, let $p = (p_x)_{x\in I}$ a family of probabilities such that $\sum_x p_x =: k$ is an integer, and let $<$ be an ordering on $\binom{I}{2}$. Then a random subset A is *generated by the Srinivasan sampling process $SSP(I, p, <)$* if it is obtained in the following inductive fashion. If $I = \emptyset$ then we set $A = \emptyset$ with probability 1. If $I = \{x\}$ then (recalling that $p_x = k$ is either 0 or 1) we set A to be the unique k-set contained in I with probability 1. If neither of these trivial cases apply, so $|I| \geq 2$, we consider the $<$-first pair $\{x, y\} \in \binom{I}{2}$. There are three cases, depending on the magnitude of $p_x + p_y$.

(a) $0 < p_x + p_y \leq 1$: In this case we play a match between x and y and the loser is marked as not being in A. The winner is chosen randomly (and this random choice is independent of other random choices in the algorithm). Denoting the winner by W and the loser by L we have $W = x$ with probability $p_x/(p_x + p_y)$ and $W = y$ with probability $p_y/(p_x + p_y)$. Then we generate A' by the Srinivasan sampling process with parameters $I' = I \setminus \{L\}$, $<'=<|_{I'}$, and p' given by $p'_W = p_i + p_j$ and $p'_x = p_x$ for $x \neq W$. Note that $\sum_{x\in I'} p'_x = \sum_{x\in I} p_x = k$. Then set $A = A'$.

(b) $1 < p_x + p_y < 2$: In this case we play a match between x and y and the **loser** is marked as being in A. [For a discussion about the choice of this terminology see after this definition.] Again denoting the winner by W and the loser by L we have

$$\mathbb{P}(W = x) = \frac{1 - p_x}{2 - p_x - p_y} \qquad \mathbb{P}(W = y) = \frac{1 - p_y}{2 - p_x - p_y}.$$

Then we generate A' by the Srinivasan sampling process with parameters $I' = I \setminus \{L\}$, $<'=<|_{I'}$, and p' given by $p'_W = p_i + p_j - 1$ and $p'_x = p_x$ for $x \neq W$. Note that $\sum_{x\in I'} p'_x = \left(\sum_{x\in I} p_x\right) - 1$. Then set $A = A' \cup \{L\}$.

(c) If $p_x = p_y = 0$, or similarly $p_x + p_y = 2$, then we can deal with both x and y simultaneously. We generate A' by the Srinivasan sampling process with parameters $I' = I \setminus \{x, y\}$, $<'=<|_{I'}$, and $p' = p\,|_{I'}$. Then we set $A = A'$ (if $p_x = p_y = 0$) or $A = A' \cup \{x, y\}$ (if $p_x = p_y = 1$).

We abuse terminology slightly and use $\mathrm{SSP}(I, p, <)$ to denote both the random variable described above and also its distribution.

Our terminology concerning winners and losers is consistent in the sense that in each match the loser's status becomes fixed and the winner continues on to play in further matches. In case (c) above we say that both x and y are losers of the match. This allows us to talk more smoothly about the progress of the SSP.

Definition 4. If A is generated by an $SSP(I, p, <)$ then we refer to the SSP that is used after the j^{th} match to generate the remainder of A as *the j^{th} subprocess*. Note that the parameters of the j^{th} subprocess are random variables. We categorize the matches that are played in the the Srinivasan sampling process in the following way. If the match falls under (a) of Definition 3 we call it a *loser-out* match. If it falls under (b) we call it a *loser-in* match, and it if falls under (c) we call it a *double loser-out* or *double loser-in* match, as appropriate.

It was conjectured in [1] that all SSPs have conditional negative association. This conjecture turns out to be false. We give here a counterexample. In fact it is a counterexample to Theorem 5.1 of [1], since it doesn't even have negative association[1].

Example 1. Define $p_i = 4/7$ for $i = 1, 2, \ldots, 7$. Let $<$ be the ordering on $\binom{[7]}{2}$ given by:

$$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{1,7\}, \{3,4\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\}, \\ \{2,7\}, \{3,5\}, \{3,6\}, \{3,7\}, \{4,5\}, \{4,6\}, \{4,7\}, \{5,6\}, \{5,7\}, \{6,7\}.$$

Note that this ordering is the lexicographic ordering, except that match $\{3,4\}$ is played immediately after all matches involving 1. We let $A \sim \mathrm{SSP}([7], p, <)$. The probabilities of the 4-sets of [7] being chosen are displayed in the table below. Sets that are not listed have probability 0 of being chosen.

4-set	Probability	4-set	Probability
$\{1,2,3,6\}$	1/56	$\{1,4,5,6\}$	1/14
$\{1,2,3,7\}$	1/56	$\{1,4,5,7\}$	1/14
$\{1,2,4,6\}$	9/280	$\{1,4,6,7\}$	1/28
$\{1,2,4,7\}$	9/280	$\{2,3,4,6\}$	2/35
$\{1,2,5,6\}$	3/175	$\{2,3,4,7\}$	2/35
$\{1,2,5,7\}$	3/175	$\{2,3,5,6\}$	12/175
$\{1,2,6,7\}$	3/350	$\{2,3,5,7\}$	12/175
$\{1,3,4,6\}$	1/28	$\{2,3,6,7\}$	6/175
$\{1,3,4,7\}$	1/28	$\{2,4,5,6\}$	2/35
$\{1,3,5,6\}$	1/14	$\{2,4,5,7\}$	2/35
$\{1,3,5,7\}$	1/14	$\{2,4,6,7\}$	1/35
$\{1,3,6,7\}$	1/28		

A routine calculation shows that

$$\mathbb{P}(\{2,3,4\} \subseteq A) - \mathbb{P}(2 \in A)\,\mathbb{P}(\{3,4\} \subseteq A) = \frac{2}{245},$$

and so A fails to be negatively associated since

$$\mathbb{E}(\mathbb{1}(2 \in A)\mathbb{1}(\{3,4\} \subseteq A)) > \mathbb{E}(\mathbb{1}(2 \in A))\,\mathbb{E}(\mathbb{1}(\{3,4\} \subseteq A)).$$

One can understand this by considering the question of whether 1 wins the first match (in which case 2 is definitely in A). If 1 wins then the next match is the losing match 1 vs. 3, biased 4 to 1 in favor of 3. Then, assuming that 3 is not eliminated, 3 plays 4 with respective probabilities 5/7 and 4/7, giving a probability of 2/7 that they are both in A. On the other hand if 2 wins the first match then 3 and 4 play each other immediately, both with probabilities 4/7 and there is a 1/7 probability that they are both in A.

There is a natural class of SSPs whose structure is more regular. These are the linear SSPs, in which, in every match after the first, a new contender plays against the winner of the previous match.

[1] In the proof of Theorem 5.1 of [1] the authors introduce a random variable Z representing the outcome of the first match and claim that Z is independent of the random variables A_x for all x not involved in the first match. This is correct for linear SSPs, and, as we prove later, for tournament SSPs, but is not true in Example 1.

Definition 5. Suppose that $I = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{N}$ has $x_1 < x_2 < \cdots < x_n$. Let $<_L$ be the lexicographic order on $\binom{I}{2}$ defined by $A <_L B$ if $\min(A \triangle B) \in A$. When we generate a random subset of I by the Srinivasan sampling process $\mathrm{SSP}(I, p, <_L)$, we start by playing a match between x_1 and x_2. Next we play x_3 against the winner of the first match (whichever was *not* fixed to 0 or 1). Then x_4 plays against the winner of the second match, and so on. Due to this behavior this type of SSP has been called a *linear SSP*; we denote it by $\mathrm{LSSP}(I, p)$.

There are several ways in which linear SSPs are easier to understand than general SSPs. The essential reason is that for a linear SSP the general outline of the process is known from the beginning. To be precise the parameters of the j^{th} subprocess do not depend on the prior random choices. In advance of playing match i we know whether it will be an loser-out match or an loser-in match; we know that one of the contestants is x_{i+1} and that its opponent is one of $x_1, x_2, \ldots, x_i$.

Theorem 2 (Dubhashi, Jonasson, and Ranjan [1])**.** *The distribution produced by an linear SSP has conditional negative association.*

3. Tournament SSPs

In this section we present the main results of the paper. We generalize Theorem 2 to a class of SSPs we call *tournament SSPs* or TSSPs. These are SSPs for which the match schedule has a tree structure. In the construction of a TSSP we imagine the various potential elements of our random set A as competing in a tournament. Leaves of the tree correspond to elements of the ground set I; internal vertices correspond to matches. For a given tournament structure there are many edge orderings which implement it (the same is true for linear SSPs). In Section 4 we discuss these orderings. In this section we give an abstract characterization of TSSPs. This allows us to deduce that certain conditioned tournament TSSPs are themselves TSSPs. This immediately implies that TSSPs have conditional negative association, since we prove that TSSPs have negative association. Our theorems for tournament SSPs apply in particular to linear SSPs.

In order to make the definition of a TSSP clear, we first need to define the notion of a tournament tree, and also the reduction of a tree (the tournament structure corresponding to the situation after one match has been played).

Definition 6. A *tournament tree* is a rooted binary tree. If T is a tournament tree with root r, and x and y are leaves of T with a common parent m, we define the *reduction of T at m*, denoted T_m, to be T with leaves x and y deleted. This is a tournament structure, which carries no information about the winner of the first match. When we need to record such information we talk about the *reduction of T in which x beats y*, denoted $T_{x/y}$, which is simply T_m with x replacing m. The root of $T_{x/y}$ is r unless $r = m$, in which case the root of $T_{x/y}$ is x. We call the non-leaf vertices of T *matches*, since they correspond to matches in the tournament. There is a natural (partial) order on the vertices of T in which $a \geq_T b$ if a is on the unique $b - r$ path in T. In fact $(V(T), \geq_T)$ is a join semi-lattice; for all $a, b \in V(T)$ there exists a unique least upper bound $a \vee b$ such that $c \geq a, b$ iff $c \geq a \vee b$. In particular, if $x, y \in I$ are leaves of T then $x \vee y$ is the unique match of the tournament in which x might meet y. For this reason we define $\mathrm{match}(x, y) = x \vee y$.

Now we define tournament SSPs precisely. We define, inductively, the notion of an edge ordering implementing a tournament structure.

Definition 7. Let T be a tournament tree with root r and set of leaves I. Let $<$ be a total order on $\binom{I}{2}$. We say that $<$ *implements the tournament structure* T if firstly, the $<$-first pair $\{x, y\}$ contains two leaves of T with a common parent m, and secondly, we have both that $<|_{I \setminus \{y\}}$ implements the tournament structure $T_{x/y}$, and $<|_{I \setminus \{x\}}$ implements $T_{y/x}$. The empty ordering implements the unique one vertex tournament tree.

Our aim now is to prove that if two match orderings, $<$ and $<'$ say, both implement the tournament structure T then the two distributions $\mathrm{SSP}(I, p, <)$ and $\mathrm{SSP}(I, p, <')$ are the same. We'll show this by giving a description of this distribution that is clearly independent of the ordering. To this end we introduce another distribution that we call a tournament sample.

Definition 8. Let T be a tournament tree with set of leaves I and let $p = (p_x)_{x \in I}$ be a family of probabilities with $\sum_{x \in I} p_x = k$, an integer. We start by extending p to be a function on all the vertices of T. For a match m of T define

$$k_m = \sum_{\substack{x \in I \\ x <_T m}} p_x$$
$$p_m = k_m - \lfloor k_m \rfloor .$$

Suppose now that m is a match with children a and b. [The children of m might of course be either leaves or matches.] We say that m is an *loser-out match* if $p_a + p_b \leq 1$, otherwise we say it is an *loser-in* match. Let Z_m be a two valued random variable whose possible values are a and b. The Z_m are chosen to be independent, and to satisfy

$$\mathbb{P}(Z_m = a) = \begin{cases} p_a/(p_a + p_b) & m \text{ is an loser-out match} \\ (1 - p_a)/(2 - p_a - p_b) & m \text{ is an loser-in match,} \end{cases}$$

$$\mathbb{P}(Z_m = b) = \begin{cases} p_b/(p_a + p_b) & m \text{ is an loser-out match} \\ (1 - p_b)/(2 - p_a - p_b) & m \text{ is an loser-in match.} \end{cases}$$

In the extreme cases when $p_a + p_b = 0, 2$ we allow Z_m to have an arbitrary distribution on $\{a, b\}$, and will prove later that in these cases the distribution of Z_m does not affect the distribution of the tournament sample which we now define. For $x \in I$ let $a_0, a_1, a_2, \ldots, a_q$ be the unique $x - r$ path in T (where $x_0 = x$ and $x_q = r$). Define

$$\ell(x) = \begin{cases} \min\{i \geq 1 \,:\, Z_{a_i} \neq a_{i-1}\} & \text{if this set is non-empty} \\ q + 1 & \text{otherwise,} \end{cases}$$

and

$$m(x) = \begin{cases} a_{\ell(x)} & \ell(x) \leq q \\ \infty & \text{otherwise.} \end{cases}$$

Thus $m(x)$ is the first match lost by x. (The symbol ∞ represents winning the final—the match r.) Finally we define the random set S by

$$S = \{x \in I \,:\, m(x) \text{ is an loser-in match or } m(x) = \infty \text{ and } r \text{ is a double loser-in match.}\}$$

This random variable S is the *tournament sample with parameters T and p*. We denote its distribution by $\mathrm{Tourn}(T, p)$. If we wish also to specify the Z_m we write $S = \mathrm{Tourn}(T, p, Z)$.

Example 2. Consider the tournament structure below (Figure 1), with $p_1 = p_2 = 2/3$, $p_3 = p_4 = 1/6$ and $p_5 = 1/3$. We have $k = 2$ and loser-in matches are marked with a $+$, loser-out matches with a $-$. The values of the Z_a are indicated by the arrows; there is an arrow from x to m if $Z_m = x$. In this case $A = \{1, 5\}$.

We will prove some basic facts about tournament samples and, with these in hand, establish that tournament SSPs are in fact tournament samples. Following that we prove an abstract characterization of tournament samples that will allow us to deduce rather quickly that certain conditioned

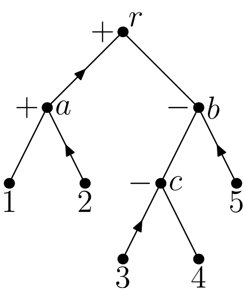

FIGURE 1. An example of a tournament sample; here $A = \{1, 5\}$

tournament samples are tournament samples, and hence that they have conditional negative association.

Lemma 3. *Suppose T is a tournament tree with set of leaves I and $p = (p_x)_{x \in I}$ is as in Definition 8. Let m be a match in T with children x, y that are leaves, and set $S = \mathrm{Tourn}(T, p, Z)$. We define S' to be the tournament sample $S' = \mathrm{Tourn}(T_m, p, Z)$. [Here p_m is defined, as in Definition 8, to be $p_x + p_y - \lfloor p_x + p_y \rfloor$ and we drop Z_m from the list of match results.] Then*

(a) S is given in terms of S' by

$$S = \begin{cases} S' & m \text{ is an loser-out match, } m \notin S' \\ S' \cup \{Z_m\} \setminus \{m\} & m \text{ is an loser-out match, } m \in S' \\ S' \cup \{x, y\} \setminus \{m\} & m \text{ is an loser-in match, } m \in S' \\ S' \cup \{x, y\} \setminus \{Z_m\} & m \text{ is an loser-in match, } m \notin S'. \end{cases}$$

(b) For any $x \in I$ we have $\mathbb{P}(x \in S) = p_x$. Moreover, if some match m' has children a, b with $p_a + p_b \in \{0, 2\}$ then the distribution of $Z_{m'}$ does not affect the distribution of S.

(c) S has size $k = \sum_{x \in I} p_x$.

Proof. To prove (a) let us write z for the winner of match m. Then $z \in S$ if and only if $m \in S'$. On the other hand the fate of the loser of match m depends only on whether m is an loser-in match or an loser-out match.

We prove (b) by induction, noting that it is trivially correct when T is a one vertex tree. First we note that if $p_x + p_y = 0$ then, by induction $\mathbb{P}(m \in S') = 0$ so $S = S'$. The value, and hence the distribution, of Z_m is irrelevant to the distribution of S. Similarly if $p_x + p_y = 2$ then $\mathbb{P}(m \in S') = 1$ and $S = S' \cup \{x, y\} \setminus \{m\}$ independently of the value of Z_m. To verify the probabilities note that if $z \in I \setminus \{x, y\}$ we have $\mathbb{P}(z \in S) = \mathbb{P}(z \in S') = p_z$. On the other hand for $z \in \{x, y\}$ we split the cases where $0 < p_x + p_y < 2$ according to whether m is an loser-out match or an loser-in match. When m is an loser-in match we have:

$$\begin{aligned} \mathbb{P}(z \in S) &= \mathbb{P}(m \in S', Z_m = z) \\ &= \mathbb{P}(m \in S')\mathbb{P}(Z_m = z) \\ &= p_m \frac{p_z}{p_x + p_y} \\ &= (p_x + p_y) \cdot \frac{p_z}{p_x + p_y} = p_z. \end{aligned}$$

Similarly, if m is an loser-out match we have:

$$\begin{aligned}\mathbb{P}(z \in S) &= \mathbb{P}(m \in S' \text{ or } Z_m \neq z) \\ &= 1 - \mathbb{P}(m \notin S', Z_m = z) \\ &= 1 - \mathbb{P}(m \notin S')\mathbb{P}(Z_m = z) \\ &= 1 - (1 - p_m)\frac{1 - p_z}{2 - p_x - p_y} \\ &= 1 - (1 - (p_x + p_y - 1)) \cdot \frac{1 - p_z}{2 - p_x - p_y} \\ &= p_z.\end{aligned}$$

The fact that the distribution of S is unaffected by the distribution of $Z_{m'}$ for any other double loser match m' is a straightforward consequence of the fact that, by induction, the distribution of S' is similarly unaffected. Finally we have, for (c),

$$\begin{aligned}|S| &= \left|S'\right| + \begin{cases} 0 & m \text{ is an loser-out match} \\ 1 & m \text{ is an loser-in match} \end{cases} \\ &= p_m + \sum_{z \in I \setminus \{x,y\}} p_z + \begin{cases} 0 & m \text{ is an loser-out match} \\ 1 & m \text{ is an loser-in match} \end{cases} \\ &= \sum_{z \in I} p_z.\end{aligned}$$

□

Now we show that tournament SSPs in fact generate tournament samples. It is easy to translate the match results in the running of the SSP into corresponding match results for the tournament sample. The only (minor) issue is that the match results for an SSP take values that are contestants, whereas those for a tournament sample refer instead to other (earlier) matches. The following definition provides the requisite translation process.

Definition 9. Suppose we are given an ordering $<$ implementing a tournament structure T and a collection of random variables W_{xy} corresponding to the running of $\text{SSP}(I, p, <)$, where W_{xy} is the winner of the match between x and y (so of course W_{xy} will only be defined for some pairs x, y.) Then for m a match of T with children a and b define

$$Z_m = \begin{cases} a & \text{there exist } x, y \text{ with } x \in T_a,\ y \in T_b \text{ and } W_{xy} = x \\ b & \text{there exist } x, y \text{ with } x \in T_a,\ y \in T_b \text{ and } W_{xy} = y. \end{cases}$$

Similarly if $\text{Tourn}(T, p, Z)$ is a tournament sample we define inductively

$$W_m = \begin{cases} x & \text{if } m = x \text{ is a leaf} \\ W_{Z_m} & \text{if } m \text{ is a match.} \end{cases}$$

This "fills in the brackets"; every match is labelled with the match winner. We will also write L_m for the loser of match m, defined equivalently. Then we define, for $x, y \in I$,

$$W_{xy} = \begin{cases} x & \exists \text{ match } m \text{ with children } a, b \text{ s.t. } x = W_a,\ y = W_b \text{ and } Z_m = a \\ y & \exists \text{ match } m \text{ with children } a, b \text{ s.t. } x = W_a,\ y = W_b \text{ and } Z_m = b \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Lemma 4. *If $<$ implements the tournament structure T, then a random set A distributed as* $\mathrm{SSP}(I, p, <)$ *has the same distribution as a tournament sample S distributed as* $\mathrm{Tourn}(T, p)$*. To be precise we have $A = \mathrm{Tourn}(T, p, Z)$, where Z_m is defined as in the previous definition.*

Proof. We proceed by induction on the number of matches in T, the result being trivial for a one vertex tree. Let $\{x, y\}$ be the $<$-first pair, and let m be the common parent of x and y (whose existence is guaranteed by the fact that $<$ implements T). The definition of a tournament SSP makes clear that $A' = A \setminus \{L_m\}$ is generated by a tournament SSP on the tournament tree T_{W_m/L_m} with

$$p_m = \begin{cases} p_x + p_y & m \text{ is an loser-out match} \\ p_x + p_y - 1 & m \text{ is an loser-in match.} \end{cases}$$

By induction

$$A' = \mathrm{Tourn}\left(T_{W_m/L_m}, p, Z\right).$$

Now, noting that

$$A = \begin{cases} A' & m \text{ is an loser-out match} \\ A' \cup \{L_m\} & m \text{ is an loser-in match,} \end{cases}$$

and comparing with Lemma 3 (equation a) we see that $A = \mathrm{Tourn}(T, p, Z)$. □

Knowing the result of this lemma we can define a tournament SSP as follows.

Definition 10. A *tournament SSP with tournament structure T and probabilities $p = (p_x)_{x \in I}$*, denoted $\mathrm{TSSP}(T, p)$, is a process $\mathrm{SSP}(I, p, <)$ such that $<$ implements T. Equivalently it is a random variable whose distribution is that of a tournament sample with parameters T and p.

Our goal is to prove that tournament SSP's have conditional negative association. We start by proving that they have negative association, closely following the proof of Theorem 5.1 in [1].

Theorem 5. *If T is a tournament structure with set of leaves I and $p = (p_x)_{x \in I}$ is a family of probabilities with $\sum_x p_x = k \in \mathbb{N}$ and $S \sim \mathrm{Tourn}(T, p)$ then S has negative association.*

Proof. We may suppose that $S = \mathrm{Tourn}(T, p, Z)$. Suppose that J, K are disjoint subsets of I and that f, g are increasing functions on $\{0,1\}^J$ and $\{0,1\}^K$ respectively. We need to show that

$$\mathbb{E}\left[f(S_J) g(S_K)\right] \leq \mathbb{E}\left[f(S_J)\right] \mathbb{E}\left[g(S_K)\right].$$

Pick a leaf match m (I.e., a match both of whose children, x and y say, are leaves). Then we have, by the conditional covariance formula, that

$$\mathrm{cov}(f(S_J), g(S_K)) = \mathbb{E}[\mathrm{cov}(f(S_J), g(S_K) \mid Z_m)] + \mathrm{cov}(\mathbb{E}[f(S_J) \mid Z_m], \mathbb{E}[g(S_K) \mid Z_m]).$$

For the first term, note that both $(S \setminus \{y\} \mid Z_m = x)$ and $(S \setminus \{x\} \mid Z_m = y)$ are tournament samples (on $T_{x/y}$ and $T_{y/x}$ respectively). Thus, by induction,

$$\mathrm{cov}(f(S_J), g(S_K) \mid Z_m) \leq 0,$$

hence

$$\mathbb{E}[\mathrm{cov}(f(S_J), g(S_K) \mid Z_m)] \leq 0.$$

As far as the second term is concerned, note that $S_{I \setminus \{x,y\}}$ is independent of Z_m, so the second term is 0 unless one of x, y belongs to J and the other to K. Without loss of generality, let us suppose that $x \in J$ and $y \in K$. Suppose, firstly, that m is an loser-out match. Let $Z = \mathbb{1}(Z_m = x)$. Note that, as in Lemma 3, if we set $S' = \operatorname{Tourn}(T_m, p, Z)$ then S' is independent of Z_m and

$$S = \begin{cases} S' & m \notin S' \\ S' \cup \{Z_m\} \setminus \{m\} & m \in S'. \end{cases}$$

This makes it clear that $\mathbb{E}[f(S_J) \mid Z]$ is increasing in Z since possible values of S_J are either equal for $Z = 0$ and $Z = 1$, or equal apart from having $S_x = Z$. Similarly $\mathbb{E}[g(S_K) \mid Z]$ is decreasing in Z, from which it follows immediately that

$$\operatorname{cov}(\mathbb{E}[f(S_J) \mid Z_m], \mathbb{E}[g(S_K) \mid Z_m]) \leq 0.$$

The case where m is an loser-in match is similar. This time we set $Z = \mathbb{1}(Z_m = y)$. We have

$$S = \begin{cases} S' \cup \{x, y\} \setminus \{m\} & m \in S' \\ S' \cup \{x, y\} \setminus \{Z_m\} & m \notin S'. \end{cases}$$

Again the possible values of S_J for a given S' are either equal or differ in that $S_x = Z$. Again we also have that $\mathbb{E}[g(S_K) \mid Z]$ is decreasing in Z, and

$$\operatorname{cov}(\mathbb{E}[f(S_J) \mid Z_m], \mathbb{E}[g(S_K) \mid Z_m]) \leq 0.$$

Having shown that both terms in the conditional covariance formula are non-positive we have $\operatorname{cov}(f(S_J), g(S_K)) \leq 0$, hence S has negative association. □

We now establish some not-quite-so-basic properties of tournament samples, properties that, it turns out, characterize them.

Theorem 6. *Suppose that $S = \operatorname{Tourn}(T, p, Z)$ is a tournament sample. For m a match of T define (recalling the definition of W_m from Definition 9)*

$$\begin{aligned} D(m) &= \{z \in I \,:\, z <_T m\} \\ S^+(m) &= \begin{cases} S \setminus D(m) & W_m \notin S \\ (S \cup \{m\}) \setminus D(m) & W_m \in S \end{cases} \\ S^-(m) &= S_{D(m)} \setminus \{W_m\} \\ N_m &= |S \cap D(m)|, \end{aligned}$$

and recall that

$$\begin{aligned} k_m &= \sum_{\substack{x \in I \\ x <_T m}} p_x \\ p_m &= k_m - \lfloor k_m \rfloor. \end{aligned}$$

Then for all matches m in T,

(a) $S^+(m) = \operatorname{Tourn}(T_m, p, Z)$, where we abuse notation and write p and Z for the restrictions of p and Z to the leaves and matches of Z_m respectively.

(b) If m's children are both leaves, x and y say, then S can be reconstructed from $S^+(m)$ and Z_m as

$$S = \begin{cases} S^+(m) & m \text{ is an loser-out match and } m \notin S^+(m) \\ (S^+(m) \setminus \{m\}) \cup \{Z_m\} & m \text{ is an loser-out match and } m \in S^+(m) \\ S^+(m) \cup \{x, y\} \setminus \{Z_m\} & m \text{ is an loser-in match and } m \notin S^+(m) \\ S^+(m) \cup \{x, y\} \setminus \{m\} & m \text{ is an loser-in match and } m \in S^+(m). \end{cases} \tag{1}$$

(c) $\mathbb{E}(N_m) = k_m$.
(d) The random variables N_m *satisfy*

$$\lfloor k_m \rfloor = \lfloor \mathbb{E}N_m \rfloor \leq N_m \leq \lceil \mathbb{E}N_m \rceil = \lceil k_m \rceil .$$

(e) $S_{D(m)}$ *and* $S_{I \setminus D(m)}$ *are conditionally independent given* N_m.

Proof. Firstly, (a) is straightforward; $\mathrm{Tourn}(T_m, p, Z)$ contains no elements of $D(m)$ and contains other elements of I, or the leaf m of T_m, precisely if their final loss is an loser-in match. This is exactly what $S^+(m)$ does also. In the other direction, if we know $S^+(m)$ then the only additional pieces of information we need to reconstruct S are S_x and S_y. The fate of the loser of match m is determined by whether m is an loser-out match or an loser-in match. On the other hand the fate of the winner is, by construction, the same as that of m in $S^+(m)$. This is precisely what (1) states, and thus (b) is proved. The next claim, (c), is simply linearity of expectation.

To prove (d) note that

$$N_m = \left|S^-(m)\right| + \mathbb{1}(W_m \in S).$$

The first term is exactly the number of loser-in matches below m, so N_m can take at most two values, and the result follows. It only remains to prove the conditional independence of $S_{D(m)}$ and $S_{I \setminus D(m)}$ given N_m. It is clear that $S^+(m)$ and $S^-(m)$ are independent since they are determined by disjoint subsets of the Z variables. Also, since from part (c) we know that $N_m - \mathbb{1}(W_m \in S)$ is a constant, we wish to show that $S_{D(m)}$ and $S_{I \setminus D(m)}$ are conditionally independent given the random variable $\mathbb{1}(W_m \in S)$. Given subsets A, B of $D(m)$ and $I \setminus D(m)$ respectively we have,

$$\begin{aligned}
&\mathbb{P}(W_m \in S)\mathbb{P}(S_{D(m)} = A, S_{I \setminus D(m)} = B, W_m \in S) \\
&\quad = \mathbb{P}(W_m \in S) \sum_{x \in A} \mathbb{P}(S^-(m) = A \setminus \{x\}, S^+(m) = B \cup \{m\}, W_m = x, x \in S) \\
&\quad = \mathbb{P}(W_m \in S) \sum_{x \in A} \mathbb{P}(S^-(m) = A \setminus \{x\}, W_m = x)\mathbb{P}(S^+(m) = B \cup \{m\}) \\
&\quad = \mathbb{P}(S^+(m) = B \cup \{m\}) \sum_{x \in A} \mathbb{P}(W_m \in S)\mathbb{P}(S^-(m) = A \setminus \{x\}, W_m = x) \\
&\quad = \mathbb{P}(S^+(m) = B \cup \{m\}) \sum_{x \in A} \mathbb{P}(S^-(m) = A \setminus \{x\}, W_m = x, x \in S) \\
&\quad = \mathbb{P}(S_{I \setminus D(m)} = B, W_m \in S)\mathbb{P}(S_{D(m)} = A, W_m \in S),
\end{aligned}$$

where the penultimate inequality is a consequence of the fact that the event $\{W_m \in S\}$ is independent of the event $\{S^-(m) = A \setminus \{x\}, W_m = x\}$ since they are functions of disjoint sets of the Z variables. This is exactly the condition that $S_{D(m)}$ and $S_{I \setminus D(m)}$ are conditionally independent given the event $Z_m \in S$. A similar calculation establishes that

$$\begin{aligned}
&\mathbb{P}(W_m \in S)\mathbb{P}(S_{D(m)} = A, S_{I \setminus D(m)} = B, W_m \notin S) \\
&\qquad = \mathbb{P}(S_{D(m)} = A, W_m \notin S)\mathbb{P}(S_{I \setminus D(m)} = B, W_m \notin S),
\end{aligned}$$

completing the proof. □

We are now almost ready to state a theorem that characterizes those random processes that are tournament samples. Before we can do that however we need to clarify the relationship between the random variables S_x and the various Z_m where $S = \mathrm{Tourn}(T, p, Z)$ is a tournament sample. Note that there is more information in the random variables Z than in the sample S. Some match results are rendered irrelevant by the results of later matches. In the next lemma we prove that this "irrelevant" information can, in a certain sense, be taken from an arbitrary source. To be more precise we would like to show that we can choose the Z_m so that they are $(\tilde{Z}, S)$-measurable, where $\tilde{Z}$ is any "suitable" source of randomness. The following lemma provides the details.

Lemma 7. *Let $S = \mathrm{Tourn}(T, p, Z')$ and let $\tilde{Z} = (\tilde{Z}_m)$ be any family of independent two-valued random variables with $\mathbb{P}(\tilde{Z}_m = a) = \mathbb{P}(Z'_m = a)$ for all matches m and all vertices a, and in addition $\tilde{Z}$ is independent of Z'. Define, for $a \in V(T)$, the random variable $X_a = N_a - \lfloor k_a \rfloor$. If we let, for every match m with children a, b,*

$$Z_m = \begin{cases} \tilde{Z}_m & X_a = X_b \\ a & X_a = 1, X_b = 0, m \textit{ is a loser-out match} \\ b & X_a = 0, X_b = 1, m \textit{ is a loser-out match} \\ b & X_a = 1, X_b = 0, m \textit{ is a loser-in match} \\ a & X_a = 0, X_b = 1, m \textit{ is a loser-in match,} \end{cases}$$

then the (Z_m) are independent, $(\tilde{Z}, S)$-measurable, and $S = \mathrm{Tourn}(T, p, Z)$.

Proof. Notice that we have rigged the match results so as to achieve the same outcome as S; only when a match result was irrelevant did we refer to the value of $\tilde{Z}_m$. The measurability result is obvious, since $(X_a)_{a \in V(T)}$ is S-measurable. The result is also clearly true when T has at most one match. In that case either $k = 0, 2$ and the value of Z_m is always irrelevant, or $k = 1$ and Z_m is defined in such a way that $\mathrm{Tourn}(T, p, Z) = \mathrm{Tourn}(T, p, Z')$. Moreover in the latter case

$$\mathbb{P}(Z_m = a) = \mathbb{P}(S = \{a\}) = \mathbb{P}(Z'_m = a).$$

Suppose now that T has at least two matches, and that $\widehat{m}$ is a leaf match of T. Consider $S^+ = S^+(\widehat{m})$. By Theorem 6(a) we have $S^+ = \mathrm{Tourn}(T_{\widehat{m}}, p, Z')$. We claim that also $S^+ = \mathrm{Tourn}(T_{\widehat{m}}, p, Z)$. First note that for all $a \in T_{\widehat{m}}$ we have

$$N_{S^+}(a) = \begin{cases} N_S(a) & \text{if } \widehat{m} \text{ is an loser-out match} \\ N_S(a) - 1 & \text{if } \widehat{m} \text{ is an loser-in match} \end{cases}$$

and hence

$$N_{S^+}(a) - \lfloor \mathbb{E} N_{S^+}(a) \rfloor = N_S(a) - \lfloor \mathbb{E} N_S(a) \rfloor = X_a.$$

Therefore we can apply induction, since the definition of Z_m will be unchanged. By Theorem 6 we know how to recover S from S^+; we simply replace $\widehat{m}$ by $W'_{\widehat{m}}$ if it appears, and add $L'_{\widehat{m}}$ if $\widehat{m}$ is an loser-in match. (We write W'_m and L'_m for the variables defined in Definition 9 corresponding the Z'.) We would like to check that it produces the same set if we perform the same operation, but based on Z instead. Well, if both (or neither) of the children are in S then everything is fine since the value of $Z_{\widehat{m}}$ is irrelevant. On the other hand if only one is in S then by construction $W_{\widehat{m}} = W'_{\widehat{m}}$.

It remains to show that the Z_m are independent. As above, we have that the $(Z_m)_{m \neq \widehat{m}}$ are independent. Let the children of $\widehat{m}$ be x and y. Let ξ be a sequence with $\mathbb{P}((Z_m)_{m \neq \widehat{m}} = \xi) > 0$. Now there are two cases, depending on whether $X_x = X_y$, *which is determined by* ξ. Firstly if $X_x = X_y$ then

$$\begin{aligned} \mathbb{P}(Z_{\widehat{m}} = x, (Z_m)_{m \neq \widehat{m}} = \xi) &= \mathbb{P}(\tilde{Z}_{\widehat{m}} = x, (Z_m)_{m \neq \widehat{m}} = \xi) \\ &= \mathbb{P}(\tilde{Z}_{\widehat{m}} = x)\mathbb{P}((Z_m)_{m \neq \widehat{m}} = \xi), \end{aligned}$$

since $\tilde{Z}_m$ is independent of S and $(Z_m)_{m\neq\widehat{m}}$ by hypothesis. The other case is that $X_x \neq X_y$. If $\widehat{m}$ is a loser-out match then

$$\begin{aligned}\mathbb{P}(Z_{\widehat{m}} = x, (Z_m)_{m\neq\widehat{m}} = \xi) &= \mathbb{P}(x \in S, (Z_m)_{m\neq\widehat{m}} = \xi)\\ &= \mathbb{P}(x \in S, (Z_m)_{m\neq\widehat{m}} = \xi, N_{\widehat{m}} = 1)\\ &= \mathbb{P}(x \in S, (Z_m)_{m\neq\widehat{m}} = \xi \mid N_{\widehat{m}} = 1)\mathbb{P}(N_{\widehat{m}} = 1)\\ &= \mathbb{P}(x \in S \mid N_{\widehat{m}} = 1)\mathbb{P}((Z_m)_{m\neq\widehat{m}} = \xi \mid N_{\widehat{m}} = 1)\mathbb{P}(N_{\widehat{m}} = 1) \qquad (*)\\ &= \frac{p_x}{p_x + p_y}\mathbb{P}((Z_m)_{m\neq\widehat{m}} = \xi, N_{\widehat{m}} = 1)\\ &= \mathbb{P}(Z_{\widehat{m}} = x)\mathbb{P}((Z_m)_{m\neq\widehat{m}} = \xi).\end{aligned}$$

The equality $(*)$ follows from the facts that S_x is conditionally independent of $S_{I\setminus D(\widehat{m})}$ given $N_{\widehat{m}}$, and that $(Z_m)_{m\neq\widehat{m}}$ is $((\tilde{Z}_m)_{m\neq\widehat{m}}, S_{I\setminus D(\widehat{m})})$-measurable by induction. The case where $X_x \neq X_y$ and $\widehat{m}$ is a loser-in match is similar. □

We are now ready to prove our characterization of tournament SSPs. Essentially the proof is a straightforward induction, but we need to be careful in moving up from a smaller case that we have enough independence between our inductively established tournament sample and the behavior of our SSP. This is where Lemma 7 comes in.

Theorem 8. *Let A be a random variable with values in $\{0,1\}^I$. Define $p_i = \mathbb{P}(A_i = 1)$, $p = (p_i)_{i\in I}$ and, for m any match in T, set $N_m = \sum_{i<_T m} A_i$. Suppose that T be a tournament structure with set of leaves I. Then $A \sim \mathrm{Tourn}(T,p)$ if and only if*

(1) For all matches m of T,

$$\lfloor \mathbb{E}(N_m) \rfloor \leq N_m \leq \lceil \mathbb{E}(N_m) \rceil\,,$$

(2) For any match m the variables $A_{D(m)}$ and $A_{I\setminus D(m)}$ are conditionally independent given N_m.

Proof. The implication in the forward direction follows from Theorem 6 (d) and (e).

For the backward direction, the proof proceeds by induction on the number of leaves of T. The result is trivial if T has only one vertex. Suppose then that T has at least two leaves, and let $\widehat{m}$ be a match both of whose children, x and y say, are leaves. We define a random subset B of the leaves of $T_{\widehat{m}}$ as follows. We'll describe the $\{0,1\}$-valued variables B_z that specify whether each leaf z of $T_{\widehat{m}}$ is in or out. For leaves of the original tree we just set $B_z = A_z$. For the special leaf $\widehat{m}$ we compare $|A \cap \{x,y\}|$ with its expectation and include $\widehat{m}$ if we have exceeded the expected value. I.e.,

$$B_z = \begin{cases} N_{\widehat{m}} - \lfloor \mathbb{E}(N_m) \rfloor & z = \widehat{m} \\ A_z & \text{otherwise.} \end{cases}$$

Note first that since $\lfloor \mathbb{E}(N_{\widehat{m}}) \rfloor \leq N_{\widehat{m}} \leq \lceil \mathbb{E}(N_{\widehat{m}}) \rceil$ we have $B_{\widehat{m}} \in \{0,1\}$. We'll show that B and $T_{\widehat{m}}$ satisfy the hypotheses of the theorem. To see this let us write, for m a match in $T_{\widehat{m}}$,

$$M_m = \sum_{z <_{T_{\widehat{m}}} m} B_z$$

Note that all matches m of $T_{\widehat{m}}$ are also matches of T and moreover $M_m = N_m$. Thus (1) is trivially satisfied. The required conditional independence follows from that of the A_z. Thus, by induction, $Y \sim \mathrm{Tourn}(T_m, q)$ where $q_{\widehat{m}} = p_x + p_y - \lfloor p_x + p_y \rfloor$ and otherwise $q_z = p_z$. Thus there exist random variables Z_m for each match of $T_{\widehat{m}}$ such that $B = \mathrm{Tourn}(T_{\widehat{m}}, q, Z)$. By Lemma 7 we can assume that there are random variables $\tilde{Z}$ independent of A such that Z is $(\tilde{Z}, B)$-measurable. We will now make a careful choice of a Bernoulli random variable $Z_{\widehat{m}}$ such that $A = \mathrm{Tourn}(T, p, Z)$. We start by letting Z_0 be a random variable, independent of $\tilde{Z}$ and B, distributed as

$$Z_0 \sim \begin{cases} \text{Bernoulli}\left(\frac{p_x}{p_x+p_y}\right) & p_x + p_y < 1 \\ \text{Bernoulli}\left(\frac{1-p_x}{2-p_x-p_y}\right) & p_x + p_y \geq 1. \end{cases}$$

Then we set

$$Z_{\widehat{m}} = \begin{cases} Z_0 & \text{if } N_{\widehat{m}} = 0 \text{ or } N_{\widehat{m}} = 2 \\ x & \text{if } A_x = 1,\ A_y = 0 \text{ and } \widehat{m} \text{ is a loser-out match} \\ y & \text{if } A_x = 0,\ A_y = 1 \text{ and } \widehat{m} \text{ is a loser-out match} \\ y & \text{if } A_x = 1,\ A_y = 0 \text{ and } \widehat{m} \text{ is a loser-in match} \\ x & \text{if } A_x = 0,\ A_y = 1 \text{ and } \widehat{m} \text{ is a loser-in match.} \end{cases}$$

For this $Z_{\widehat{m}}$ it is straightforward to check, from (1), that $A = \mathrm{Tourn}(T, p, Z)$; in the cases where $Z_{\widehat{m}}$ is defined to be Z_0 its value is ignored in the calculation of $\mathrm{Tourn}(T, p, Z)$, in the other cases it is chosen to have the correct value to ensure that $A = \mathrm{Tourn}(T, p, Z)$.

It remains of course to prove that $Z_{\widehat{m}}$ is independent of the other Z_m and has the right distribution. For any fixed sequence $\zeta = (\zeta_m)_{m \neq \widehat{m}}$ we wish to show that $\mathbb{P}(Z_{\widehat{m}} = x \mid (Z_m)_{m \neq \widehat{m}} = \zeta)$ has the appropriate value. We split into two cases according to whether $\widehat{m}$ is a loser-in or a loser-out match. If it a loser-out match then we have

$$\begin{aligned} \mathbb{P}(Z_{\widehat{m}} = x \mid (Z_m)_{m \neq \widehat{m}} = \zeta) &= \mathbb{P}(Z_{\widehat{m}} = x \mid (Z_m)_{m \neq \widehat{m}} = \zeta, N_{\widehat{m}} = 1)\mathbb{P}(N_{\widehat{m}} = 1) + \\ &\qquad \mathbb{P}(Z_{\widehat{m}} = x \mid (Z_m)_{m \neq \widehat{m}} = \zeta, N_{\widehat{m}} = 0)\mathbb{P}(N_{\widehat{m}} = 0) \\ &= \mathbb{P}(A_x = 1 \mid (Z_m)_{m \neq \widehat{m}} = \zeta, N_{\widehat{m}} = 1)\mathbb{P}(N_{\widehat{m}} = 1) + \\ &\qquad \mathbb{P}(Z_0 = 1)\mathbb{P}(N_{\widehat{m}} = 0) \\ &= \mathbb{P}(A_x = 1 \mid N_{\widehat{m}} = 1)\mathbb{P}(N_{\widehat{m}} = 1) + \mathbb{P}(Z_0 = 1)\mathbb{P}(N_{\widehat{m}} = 0) \\ &= \frac{p_x}{p_x + p_y}(\mathbb{P}(N_{\widehat{m}} = 1) + \mathbb{P}(N_{\widehat{m}} = 0)) \\ &= \frac{p_x}{p_x + p_y}. \end{aligned}$$

The third equality follows from the conditional independence of A_x and A_y from $A_z, z \notin \{x, y\}$, together with the independence of $\tilde{Z}$ from A and the $(\tilde{Z}, (A_z)_{z \neq x,y})$-measurability of Z. The fourth is a simple calculation: we have

$$\mathbb{P}(A_x = 1 \wedge N_{\widehat{m}} = 1) = \mathbb{P}(A_x = 1) = p_1,$$

and, since $N_{\widehat{m}}$ is a Bernoulli random variable,

$$\mathbb{P}(N_{\widehat{m}} = 1) = \mathbb{E}(N_{\widehat{m}}) = \mathbb{E}(A_x) + \mathbb{E}(A_y) = p_1 + p_2.$$

The other case is when $\widehat{m}$ is a loser-in match; this case is proved entirely analogously. ☐

Corollary 9. *If $S \sim \mathrm{Tourn}(T, p)$, J is a subset of I, and $\zeta = (\zeta_j)_{j \in J}$ then the random variable $S' = (S \mid S_J = \zeta)$ is distributed as $\mathrm{Tourn}(T, q)$ where $q_x = \mathbb{P}(i \in S \mid S_J = \zeta)$.*

Proof. We need merely check the conditions of Theorem 8. For a match m in T we write $M_m = \sum_{x <_T m} S'_x$. Since S is a tournament sample M_m takes at most two adjacent values, so certainly S' satisfies (1). Now we need to prove that for m a match of T we have $S'_{D(m)}$ and $S'_{I \setminus D(m)}$ conditionally independent given M_m. Of the "fixed" values $S_J = \zeta$ some are in $D(m)$ and some are not. Let's write $J_1 = D(m) \cap J$, $J_2 = J \setminus D(m)$ and ζ_1, ζ_2 for the restriction of ζ to J_1, J_2

respectively. Suppose then that ξ, ψ are possible values for $S'_{D(m)}$ and $S'_{I\setminus D(m)}$ respectively. [In particular we know that ξ restricted to J_1 equals ζ_1 and similarly ψ restricted to $J_2 = \zeta_2$.]

$$\begin{aligned}
\mathbb{P}(S'_{D(m)} = \xi \mid S'_{I\setminus D(m)} = \psi, M_m = k) &= \frac{\mathbb{P}(S'_{D(m)} = \xi, S'_{I\setminus D(m)} = \psi, M_m = k)}{\mathbb{P}(S'_{I\setminus D(m)} = \psi, M_m = k)} \\
&= \frac{\mathbb{P}(S_{D(m)} = \xi, S_{I\setminus D(m)} = \psi, N_m = k)/\mathbb{P}(S_J = \zeta)}{\mathbb{P}(S_{J_1} = \zeta_1, S_{I\setminus D(m)} = \psi, N_m = k)/\mathbb{P}(S_J = \zeta)} \\
&= \frac{\mathbb{P}(S_{D(m)} = \xi, N_m = k)}{\mathbb{P}(S_{J_1} = \zeta_1, N_m = k)} \\
&= \mathbb{P}(S'_{D(m)} = \xi \mid M_m = k).
\end{aligned}$$

This proves the conditional independence of $S'_{D(m)}$ and $S'_{I\setminus D(m)}$ given M_m, which, by Theorem 8, proves the corollary. □

It is now immediate to deduce the main result of our paper, that tournament SSPs have conditional negative association.

Theorem 10. *TSSPs have conditional negative association.*

Proof. By Corollary 9 we know that conditioned TSSPs are themselves TSSPs. These have negative association by Theorem 5. Thus TSSPs have conditional negative association. □

4. Orderings implementing a given tournament structure

Finally in this section we characterize those edge orderings that implement a given tournament structure T.

Theorem 11. *An ordering $<$ on $\binom{I}{2}$ implements a tournament structure T if and only if for all $x, y, z \in I$ with $\mathrm{match}(x, y) <_T \mathrm{match}(y, z)$, we have $\{x, y\} < \{y, z\}$.*

Before proving the theorem we state and prove a lemma. For this purpose it will be useful to give a name to the property in the theorem.

Definition 11. Given an ordering $<$ on I and a tournament structure T with set of leaves I we say that $<$ is *compatible* with T if for all $x, y, z \in I$ we have

$$\mathrm{match}(x, y) <_T \mathrm{match}(y, z) \quad \Longrightarrow \quad \{x, y\} < \{y, z\}\,.$$

Lemma 12. *Suppose $<$ is an ordering compatible with a tournament structure T. Let $\{x, y\}$ be the $<$-first pair. Then x and y have a common parent, $<|_{I\setminus\{y\}}$ is compatible with $T_{x/y}$, and $<|_{I\setminus\{x\}}$ is compatible with $T_{y/x}$.*

Proof. For the first claim, suppose that x and y do not have a common parent. Let $m = \mathrm{match}(x, y)$. At most one of x, y is a child of m, so let us suppose, without loss of generality, that y is not an immediate child of m. Thus the parent p of y is not m. The sub-tree below p has at least two leaves, and thus there is some leaf other than y below p. Pick one such and call it z. Now we have

$$\mathrm{match}(z, y) = p <_T m = \mathrm{match}(y, x),$$

but $\{z, y\} \not< \{y, x\}$, contradicting the compatibility of $<$.

We will show that $<|_{I\setminus\{y\}}$ is compatible with $T_{x/y}$; the other case is clearly identical. Let us write p for the common parent of x and y. The ordering $<_{T_{x/y}}$ is a restriction of the ordering $<_T$, bearing in mind that both y and p have been removed to get to $T_{x/y}$. Moreover for all leaves $u, w \neq y$ we have $\mathrm{match}_{T_{x/y}}(u, w) = \mathrm{match}_T(u, w)$. Consider then a triple $a, b, c \in I$ of vertices of the reduced

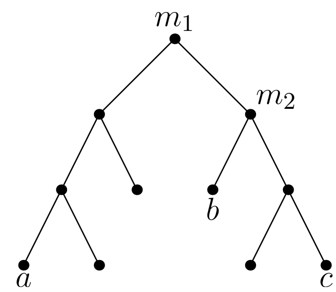

FIGURE 2. In this tournament structure we have

$$\text{match}(a, c) = m_1 >_T m_2 = \text{match}(b, c).$$

Any match ordering $<$ implementing T must have $\{b, c\} < \{a, c\}$.

tree $T_{x/y}$. We must show that if $\text{match}_{T_{x/y}}(a, b) <_{T_{x/y}} \text{match}_{T_{x/y}}(b, c)$, then $\{a, b\} < \{b, c\}$. Clearly none of a, b, c is y or p, so $\text{match}_T(a, b) <_T \text{match}_T(b, c)$, therefore $\{a, b\} < \{b, c\}$, as required. □

Proof of Theorem 11. The lemma above establishes, by induction, that if $<$ is compatible with a tournament structure T then $<$ implements T. (The base case is trivial.)

For the other direction, suppose that $<$ implements a tournament structure T with leaf set I, and yet, for some $x, y, z \in I$ with $\text{match}(x, y) >_T \text{match}(y, z)$, we have $\{x, y\} < \{y, z\}$. We may suppose that our counterexample is vertex minimal, in which case $\{x, y\}$ must be the $<$-first pair (otherwise we could consider $T_{u/v}$, where $\{u, v\}$ is $<$-first). But x, y do not have a common parent; to be precise $\text{match}(y, z) <_T \text{match}(x, y)$. This contradicts the definition of $<$ implementing T. □

5. Further Directions

There are still a large number of natural open questions in this area. In general it would certainly be interesting to know whether there are other classes of SSPs that have conditional negative association, or indeed negative association. However the most interesting question concerns a related process called a *random ordering SSP*. In this process we start by picking an ordering from some distribution on the set of all orderings on $\binom{I}{2}$, and then we run an SSP using this ordering. The techniques in this paper seem to offer very little traction in this more general setting. Even proving that a random ordering SSP that starts by picking a random linear ordering on I has negative association seems a difficult task.

References

1. Devdatt Dubhashi, Johan Jonasson, and Desh Ranjan, *Positive influence and negative dependence*, Combin. Probab. Comput. **16** (2007), no. 1, 29–41. MR MR2286510 (2008h:62035)
2. Devdatt Dubhashi and Desh Ranjan, *Balls and bins: a study in negative dependence*, Random Structures Algorithms **13** (1998), no. 2, 99–124. MR MR1642566 (99k:60048)
3. Tomás Feder and Milena Mihail, *Balanced matroids*, STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing (New York, NY, USA), ACM, 1992, pp. 26–38.
4. G. R. Grimmett and S. N. Winkler, *Negative association in uniform forests and connected graphs*, Random Structures Algorithms **24** (2004), no. 4, 444–460. MR MR2060630 (2004m:60014)
5. Kumar Joag-Dev and Frank Proschan, *Negative association of random variables, with applications*, Ann. Statist. **11** (1983), no. 1, 286–295. MR MR684886 (85d:62058)
6. Robin Pemantle, *Towards a theory of negative dependence*, J. Math. Phys. **41** (2000), no. 3, 1371–1390, Probabilistic techniques in equilibrium and nonequilibrium statistical physics. MR MR1757964 (2001g:62039)

7. David Reimer, *Proof of the van den Berg-Kesten conjecture*, Combin. Probab. Comput. **9** (2000), no. 1, 27–32. MR MR1751301 (2001g:60017)
8. Charles Semple and Dominic Welsh, *Negative correlation in graphs and matroids*, Combin. Probab. Comput. **17** (2008), no. 3, 423–435. MR MR2410396
9. Aravind Srinivasan, *Distributions on level-sets with applications to approximation algorithms*, 42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001), IEEE Computer Soc., Los Alamitos, CA, 2001, pp. 588–597. MR MR1948748
10. J. van den Berg and H. Kesten, *Inequalities with applications to percolation and reliability*, J. Appl. Probab. **22** (1985), no. 3, 556–569. MR MR799280 (87b:60027)

Mathematics and Computer Science Department, Illinois Wesleyan University, Bloomington, IL
E-mail address: jbrownkr@iwu.edu

Department of Mathematical Sciences, Montclair State University, Montclair, NJ
E-mail address: jonathan.cutler@montclair.edu

Department of Mathematics, University of Nebraska-Lincoln, Lincoln, NE
E-mail address: aradcliffe1@math.unl.edu