## Operations Research

## Random Knockout Tournaments

Ilan Adler, Yang Cao, Richard Karp, http://orcid.org/0000-0002-7459-7562Erol A. Peköz,
Sheldon M. Ross

Please scroll down for article—it is on subsequent pages

# Random Knockout Tournaments

**Ilan Adler,[a] Yang Cao,[b] Richard Karp,[c] Erol A. Peköz,[d] Sheldon M. Ross[b]**

[a] Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720;
[b] Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90033; [c] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, California 94720; [d] Questrom School of Business, Boston University, Boston, Massachusetts 02215
**Contact:** adler@ieor.berkeley.edu (IA); cao573@usc.edu (YC); karp@cs.berkeley.edu (RK); pekoz@bu.edu,
http://orcid.org/0000-0002-7459-7562 (EAP); smross@usc.edu (SMR)

**Abstract.** We consider a random knockout tournament among players $1, \ldots, n$, in which each match involves two players. The match format is specified by the number of matches played in each round, where the constitution of the matches in a round is random. Supposing that there are numbers $v_1, \ldots, v_n$ such that a match between $i$ and $j$ will be won by $i$ with probability $v_i/(v_i + v_j)$, we obtain a lower bound on the tournament win probability for the best player, as well as upper and lower bounds for all of the players. We also obtain additional bounds by considering the best and worst formats for player 1 in the special case $v_1 > v_2 = v_3 = \cdots = v_n$.

## 1. Introduction

We consider a tournament among players $1, \ldots, n$, in which each match involves two players. The tournament is assumed to be of knockout type in that the losers of matches are eliminated and do not move on to the next round, and the tournament continues until all but one player is eliminated, with that player being declared the winner of the tournament. The match format is specified by the set of positive integers $r$, $m_1, \ldots, m_r$ with the interpretation that there are a total of $r$ rounds, with round $i$ consisting of $m_i$ matches, $\sum_{i=1}^{r} m_i = n - 1$. Because $\sum_{j=1}^{i-1} m_j$ players have been eliminated by the end of round $i - 1$, we must have that $m_i \leqslant (n - \sum_{j=1}^{i-1} m_j)/2$.

We suppose that the constitution of the matches in a round is totally random—that is, for instance, the $2m_1$ players that play in round 1 are randomly chosen from all $n$ players and then randomly arranged into $m_1$ match pairs. The winners of these $m_1$ matches, along with the $n - 2m_1$ players that did not play a match in round 1, then move to round 2, and so on.

We suppose that the players have respective values $v_1, \ldots, v_n$ and that a match involving players $i$ and $j$ is won by player $i$ with probability $v_i/(v_i + v_j)$. We let $P_i$ be the probability that player $i$ wins the tournament, $i = 1, \ldots, n$. In Section 3, we derive a lower bound on the probability that the strongest player (e.g., the one with the largest $v$) wins the tournament and an upper bound on the probability that the weakest player wins the tournament. In Section 4, we derive upper and lower bounds on the win probabilities $P_i, i \geqslant 1$, and

also show that if $v_1 \geqslant v_2 \geqslant \cdots \geqslant v_n$, then $P_1 \geqslant P_2 \geqslant \cdots \geqslant P_n$. In Section 5, we consider the special case where $v_1 > v_2 = \cdots = v_n$ and show that when $n = 2^s + k, 0 \leqslant k < 2^s$, the best format for the strongest player is the so-called *balanced format* that has $k$ matches in the first round and then has all remaining players competing in each subsequent round. We also show that whenever the number of remaining players( say, $t$) is even, there is an optimal (from the point of view of the best remaining player) format that calls for $t/2$ matches in the next round. We also show, for the section's special case, that the worst format for the best player is to have exactly one match each round. Analogous results for the worst player are also shown. Although we do not have a proof, we conjecture that, among all possible formats, the balanced format maximizes and the one-match-per-round format minimizes the best player's probability of winning the tournament even in the case of general $v_i$. (We show by a counterexample that the format that calls for $t/2$ matches when an even number $t$ of players remain is not optimal for the best player in the general case.)

## 2. Literature Review

Most papers in the literature on random knockout tournaments consider more structured formats than the ones we are considering, which suppose that the number of matches in each round is fixed and that the game participants in each round are randomly chosen from those that remain. In these papers, the structure

**Figure 1.** Two Balanced Structures with Six Players



(a) First balanced structure    (b) Second balanced structure
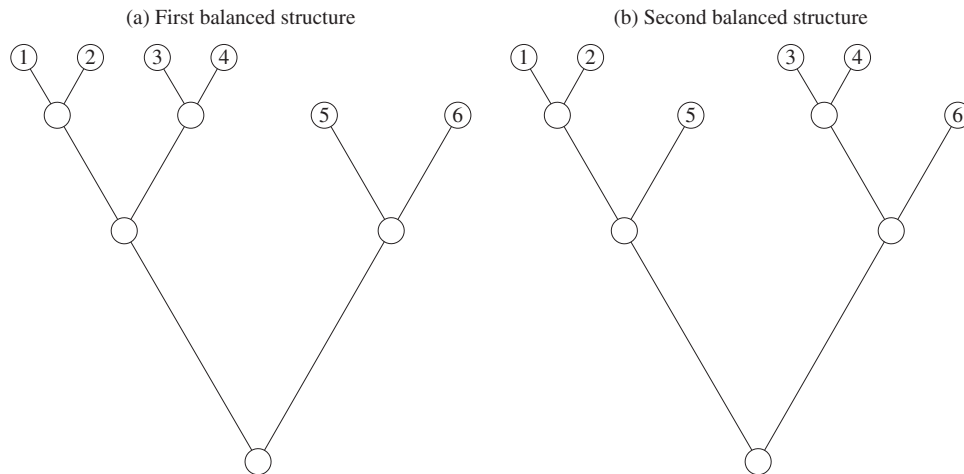
of the tournament is fixed, and players are randomly assigned to positions of the structure. For example, a structure with six players having a balanced format (meaning two matches in round 1, two matches in round 2, and one match in round 3) might in round 1 have a match between those in positions 1 and 2 and one between those in positions 3 and 4, with the winners then playing each other in round 2 and five playing six in round 2. Another possibility would be to have the winners of round 1 each play in round 2 one of those that did not play in round 1. Panels (a) and (b), respectively, in Figure 1 indicate these two structures.

An example of a nonbalanced structure with six players is one where there is a single match in each round, with the winner of a match playing in the following round someone who has not yet played.

Maurer (1975) proved for random structured formats, when $v_1 > v = v_2 = v_3 = \cdots = v_n$, that the win probability of player 1 is maximized under the balanced structures. Because, as noted by Maurer, a random format can be expressed as a mixture of random structured formats, this also establishes the result for random formats. In Section 5, we give another proof of this result, which is both quite elementary and also shows, under the same condition, that when there are $n = 2^s + k$ remaining players, the optimal random format can schedule either $k$ or $n/2$ games if $n$ is even and $k$ games if $n$ is odd. We also prove, when $v_1 > v = v_2 = v_3 = \cdots = v_n$, that the win probability of player 1 is uniquely minimized by the random format that has exactly one match in each round.

Chung and Hwang (1978) proved that for any structured format, if $v_i$ is decreasing in $i$, then so is the winning probability, which also proves the result for any random unstructured format. In Corollary 1 of Section 4, we give a proof of this result that also proves, when $v_i$ is decreasing in $i$, that the probability player $i$ reaches round $s$ is decreasing in $i$ for all $s$.

There are other papers studying knockout tournaments—for example, David (1959), Glenn (1960), Hennessy and Glickman (2016), Horen and Riezman (1985), Hwang (1982), Marchand (2002), Khatibi et al. (2015), and Jacobson and King (2009). The paper by Ross and Ghamami (2008) discusses efficient ways of simulating win probabilities for such tournaments.

## 3. Lower Bound on the Strongest Player's Win Probability

In this section, we establish a lower bound on the probability that the strongest player wins the tournament and an upper bound on the probability that the weakest player wins.

**Theorem 1.** *If $v_i$ is decreasing in $i$, then*

$$P_1 \geqslant \frac{v_1}{\sum_{j=1}^n v_j}, \quad P_n \leqslant \frac{v_n}{\sum_{j=1}^n v_j}.$$

Preliminary to proving the theorem, we define the *weight* of a player as follows. Say that a player is alive if that player has not been eliminated. If $S$ is the current set of alive players, then, for $i \in S$, the weight of player $i$ is defined to equal $v_i / \sum_{j \in S} v_j$. If $i$ is no longer alive, its weight is defined to be 0.

The following lemma is needed to prove the theorem.

**Lemma 1.** *Suppose at the start of a round that players $A$ and $B$ are alive and have respective weights $x$ and $y$, where $x \geqslant y$. If $X$ and $Y$ are random variables denoting the respective weights of $A$ and $B$ after the round, then $y\mathrm{E}[X] \geqslant x\mathrm{E}[Y]$.*

**Proof.** The result is immediate if $y = 0$, so suppose that $x \geqslant y > 0$. Define a random variable $Z$ whose value is determined by the results of the following round, as follows:

- $Z = 0$ if $A$ and $B$ both lose,
- $Z = 1$ if $A$ and $B$ both advance,

- $Z = 2$ if $A$ and $B$ play each other,
- $Z = (i, j)$ if one of the players $A$ or $B$ defeats $i$ whereas the other loses to $j$, and
- $Z = (0, j)$ if one of the players $A$ or $B$ advances by not being selected to play whereas the other loses to $j$.

By considering the possible values of $Z$, we now show that $y\mathrm{E}[X \mid Z] \geqslant x\mathrm{E}[Y \mid Z]$.

1. $\mathrm{E}[X \mid Z = 0] = \mathrm{E}[Y \mid Z = 0] = 0$.
2. Conditional on $Z = 1$, let $R$ represent the sum of the weights of those aside from $A$ and $B$ that advance. Then,

$$\mathrm{E}[X \mid Z = 1] = \mathrm{E}\left[\frac{x}{x + y + R}\right], \quad \mathrm{E}[Y \mid Z = 1] = \mathrm{E}\left[\frac{y}{x + y + R}\right],$$

showing that $y\mathrm{E}[X \mid Z = 1] = x\mathrm{E}[Y \mid Z = 1]$.

3. Conditional on $Z = 2$, let $R$ represent the sum of the weights of those aside from $A$ and $B$ that advance (and note that $R$ is independent of who wins the game between $A$ and $B$). Then,

$$\mathrm{E}[X \mid Z = 2] = \frac{x}{x + y}\mathrm{E}\left[\frac{x}{x + R}\right],$$

$$\mathrm{E}[Y \mid Z = 2] = \frac{y}{x + y}\mathrm{E}\left[\frac{y}{y + R}\right].$$

Hence,

$$y\mathrm{E}[X \mid Z = 2] - x\mathrm{E}[Y \mid Z = 2] = \frac{xy}{x + y}\mathrm{E}\left[\frac{x}{x + R} - \frac{y}{y + R}\right] \geqslant 0,$$

where the inequality follows because $x \geqslant y$ implies that $x/(x + r) \geqslant y/(y + r)$ for any $r \geqslant 0$.

4. With $w$ and $v$ representing, respectively, the weights of $i$ and $j$ before the round, and using the fact that $A$ is equally likely to play $i$ or $j$, we have

$$P(A \text{ advances} \mid Z = (i, j)) = c\left(\frac{x}{x + w}\right)\left(\frac{v}{v + y}\right),$$

where $1/c = (x/(x + w))(v/(v + y)) + (y/(y + w)) \cdot (v/(v + x))$. Let $R$ represent the sum of the weights of those, aside from $A$, $B$, and $j$, that advance, and note that $R$ is independent of which of $A$ or $B$ loses. Then,

$$\mathrm{E}[X \mid Z = (i, j)] = c\left(\frac{x}{x + w}\right)\left(\frac{v}{v + y}\right)\mathrm{E}\left[\frac{x}{R + x + v}\right],$$

and similarly,

$$\mathrm{E}[Y \mid Z = (i, j)] = c\left(\frac{y}{y + w}\right)\left(\frac{v}{v + x}\right)\mathrm{E}\left[\frac{y}{R + y + v}\right].$$

Hence, letting $D = y\mathrm{E}[X \mid Z = (i, j)] - x\mathrm{E}[Y \mid Z = (i, j)]$, we see that

$$D = cxyv\mathrm{E}\left[\frac{x}{(x + w)(y + v)(R + x + v)} - \frac{y}{(y + w)(x + v)(R + y + v)}\right].$$

As it is easy to check that for every $R \geqslant 0$ and $x \geqslant y$

$$x(y + w)(x + v)(R + y + v) \geqslant y(x + w)(y + v)(R + x + v),$$

it follows that $y\mathrm{E}[X \mid Z = (i, j)] - x\mathrm{E}[Y \mid Z = (i, j)] \geqslant 0$ for $x \geqslant y$.

5. That $y\mathrm{E}[X \mid Z = (0, j)] - x\mathrm{E}[Y \mid Z = (0, j)] \geqslant 0$ follows from the preceding result by setting $w = 0$.

Hence, we have shown that $y\mathrm{E}[X \mid Z] \geqslant x\mathrm{E}[Y \mid Z]$, and the result follows on taking expectations of both sides of this inequality.  □

**Proof of Theorem 1.** We give the proof that $P_1 \geqslant v_1/\sum_{j=1}^n v_j$. The proof that $P_n \leqslant v_n/\sum_{j=1}^n v_j$ is similar. Let $W_j(k)$ be the weight of player $j$ after $k$ rounds have been played. Also, let $H_k$ be the history of all results through the first $k$ rounds. We claim that

$$\mathrm{E}[W_1(k + 1) \mid H_k] \geqslant W_1(k).$$

Because the claim is true when $W_1(k) = 0$, assume that $W_1(k) > 0$. Let $A_k$ denote the set of alive players after round $k$. Now, from Lemma 1, we have, for $j \in A_k$, that

$$W_j(k)\mathrm{E}[W_1(k + 1) \mid H_k] \geqslant W_1(k)\mathrm{E}[W_j(k + 1) \mid H_k].$$

Hence, if $\mathrm{E}[W_1(k + 1) \mid H_k] < W_1(k)$, then for any $j \in A_k$,

$$W_j(k) > \mathrm{E}[W_j(k + 1) \mid H_k],$$

which is a contradiction since $1 = \sum_{j \in A_k} W_j(k) = \sum_{j \in A_k} \mathrm{E}[W_j(k + 1) \mid H_k]$. Hence, $\mathrm{E}[W_1(k + 1) \mid H_k] \geqslant W_1(k)$, and taking expectations of both sides gives

$$\mathrm{E}[W_1(k + 1)] \geqslant \mathrm{E}[W_1(k)], \quad k \geqslant 0.$$

If the tournament has $r$ rounds, then the preceding yields that $\mathrm{E}[W_1(r)] \geqslant \mathrm{E}[W_1(0)]$, which gives the result since $\mathrm{E}[W_1(r)] = P_1$, whereas $W_1(0) = v_1/\sum_{j=1}^n v_j$.  □

**Remark 1.** The preceding argument shows that $W_1(k)$, $k \geqslant 0$ is a submartingale, and that $W_n(k), k \geqslant 0$ is a supermartingale.

## 4. Bounds on Win Probabilities

Let

$$p_i = \frac{1}{n - 1}\sum_{j \neq i}\frac{v_i}{v_i + v_j}$$

be the probability that $i$ would win a match against a randomly chosen opponent. In this section, we prove that $P_i$ is smaller than it would be if it were the case that $i$ would win each game it plays with probability $p_i$. That is, we will prove the following.

**Theorem 2.** *If the tournament format is* $(r, m_1, \dots, m_r)$, *then*

$$P_i \leqslant \prod_{s=1}^r\left(\frac{2m_s}{r_s}p_i + 1 - \frac{2m_s}{r_s}\right),$$

*where* $r_s = n - \sum_{j=1}^{s-1} m_j$ *is the number of players that advance to round* $s$.

To prove the preceding theorem, we will need a couple of lemmas. Before giving these lemmas, we introduce the following notation. We let $R_{i,s}$ be the event that player $i$ reaches round $s$, $s = 1, \ldots, r$, and we let $R_{i,r+1}$ be the event that $i$ wins the tournament. If a player receives a bye in a round (that is, if it reaches that round but is not chosen to play a match), say that it plays player 0 in that round. Also, let $p_{ij} = v_i/(v_i + v_j)$, $i \neq j$, be the probability that $i$ beats $j$ in a game.

Lemma 2 is easily proven by a coupling argument.

**Lemma 2.** *For all $s = 1, \ldots, r+1$, $P(R_{i,s}R_{j,s})$ is an increasing function of $v_j$. (When $i = j$, this states that $P(R_{i,s})$ is an increasing function of $v_i$.)*

**Lemma 3.** *Assume $v_1 \geqslant v_2 \geqslant \cdots \geqslant v_n$. For fixed $i$, $P(R_{i,s}R_{j,s})$ decreases in $j$ for $j \neq i$.*

**Proof.** Fix $i$, $j$, $k$, where $j < k$ and $j, k \neq i$. The proof is by induction on $n$. For $n = 3$, the only format is to play one game each round. The result holds in this case because $v_j \geqslant v_k$ implies that

$$3P(R_{i,2}R_{j,2}) = \frac{v_i}{v_i + v_k} + \frac{v_j}{v_j + v_k} \geqslant \frac{v_i}{v_i + v_j} + \frac{v_k}{v_j + v_k}$$
$$= 3P(R_{i,2}R_{k,2}).$$

Assume the results holds for up to $n - 1$ players and for all formats. We now consider the $n$ player case. Let $\pi$ be an arbitrary format. Define a random vector $Z$ whose value is determined by the results of the first round of the tournament under $\pi$, as follows:

- $Z = (1, u)$ if $i$ plays against $j$ and $k$ plays against $u$, or $i$ plays against $k$ and $j$ plays against $u$;
- $Z = (2, u)$ if $i$ plays against $u$, and $j$ plays against $k$;
- $Z = (3, u, v, w)$ if $i$ plays against $u$, $j$ beats $v$, and $k$ beats $w$;
- $Z = (4, u, v, w)$ if $i$ plays against $u$, $j$ loses to $v$, and $k$ loses to $w$;
- $Z = (5, u, v, w)$ if either (a) $i$ plays against $u$, $j$ beats $v$, and $k$ loses to $w$ or (b) $i$ plays against $u$, $k$ beats $v$, and $j$ loses to $w$.

By considering the possible values of $Z$, we now show that $P(R_{i,s}R_{j,s} \mid Z) \geqslant P(R_{i,s}R_{k,s} \mid Z)$.

1. Let $p_{zy} = v_z/(v_z + v_y)$. Because $p_{ik} \geqslant p_{ij}$ and $p_{ju} \geqslant p_{ku}$, it follows that

$$P(R_{i,2}R_{j,2} \mid Z = (1, u)) = p_{ik}p_{ju}/2 \geqslant p_{ij}p_{ku}/2$$
$$= P(R_{i,2}R_{k,2} \mid Z = (1, u)).$$

Let $T$ denote the set of players other than $i$, $j$, and $k$, that reach round 2. Given that $i$ and $j$ have reached round 2, the probability of $i$ and $j$ reaching round $s > 2$ is equal to the probability that $i$ and $j$ reaches round $s - 1$ in a new tournament that begins with players $i$, $j$, and $T$ and follows the same format as $\pi$ after round 1. Let $P_{i,j,T}(s - 1)$ denote the probability that $i$ and $j$ reach round $s - 1$ in the new tournament. Similarly, let $P_{i,k,T}(s - 1)$ denote the probability that $i$ and $k$ reach round $s - 1$ in a tournament that begins with players $i$, $k$, and $T$, and follows the same format as $\pi$ after round 1. Then, by Lemma 2, $P_{i,j,T}(s - 1) \geqslant P_{i,k,T}(s - 1)$. Therefore, for $s > 2$,

$$P(R_{i,s}R_{j,s} \mid Z = (1, u))$$
$$= P(R_{i,2}R_{j,2} \mid Z = (1, u))\mathrm{E}[P_{i,j,T}(s - 1)]$$
$$\geqslant P(R_{i,2}R_{k,2} \mid Z = (1, u))\mathrm{E}[P_{i,k,T}(s - 1)]$$
$$= P(R_{i,s}R_{k,s} \mid Z = (1, u)),$$

where the preceding used that $T$ is independent of the event $R_{i,s}R_{j,s}$ and of the event $R_{i,s}R_{k,s}$.

2. Since $v_j \geqslant v_k$, we have that

$$P(R_{i,2}R_{j,2} \mid Z = (2, u)) = p_{iu}p_{jk} \geqslant p_{iu}p_{kj}$$
$$= P(R_{i,2}R_{k,2} \mid Z = (2, u)).$$

Let $T$ denote the set of players other than $i$, $j$, and $k$ that reach round 2. With the same definition and argument as when $Z = (1, u)$, it can be shown that for $s > 2$,

$$P_{i,j,T}(s) \geqslant P_{i,k,T}(s),$$

which implies that

$$P(R_{i,s}R_{j,s} \mid Z = (2, u)) \geqslant P(R_{i,s}R_{k,s} \mid Z = (2, u)).$$

3. In this case,

$$P(R_{i,2}R_{j,2} \mid Z = (3, u, v, w)) = p_{iu}$$
$$= P(R_{i,2}R_{k,2} \mid Z = (3, u, v, w)).$$

Now, given that $i$, $j$, and $k$ have reached round 2, the induction hypothesis implies that the probability of $i$ and $j$ reaching round $s$ is greater than or equal to the probability of $i$ and $k$ reaching round $s$ for $s > 2$. Therefore, for $s > 2$,

$$P(R_{i,s}R_{j,s} \mid Z = (3, u, v, w)) \geqslant P(R_{i,s}R_{k,s} \mid Z = (3, u, v, w)).$$

4. In this case,

$$P(R_{i,s}R_{j,s} \mid Z = (4, u, v, w))$$
$$= P(R_{i,s}R_{k,s} \mid Z = (4, u, v, w)) = 0.$$

5. In this case,

$$P(R_{i,2}R_{j,2} \mid Z = (5, u, v, w)) = cp_{iu}p_{jv}p_{wk},$$
$$P(R_{i,2}R_{k,2} \mid Z = (5, u, v, w)) = cp_{iu}p_{kv}p_{wj},$$

where $1/c = p_{jv}p_{wk} + p_{kv}p_{wj}$.

Since $p_{jv} \geqslant p_{kv}$ and $p_{wk} \geqslant p_{wj}$, we have that

$$P(R_{i,2}R_{j,2} \mid Z = (5, u, v, w)) \geqslant P(R_{i,2}R_{k,2} \mid Z = (5, u, v, w)).$$

Let $T$ denote the set of players other than $i$, $j$, and $k$ that reach round 2, and note that $T$ has the same distribution whether (a) or (b) resulted. With the same

definition and argument as when $Z = (1, u)$, it can be shown that for $s > 2$,

$$P_{i,j,T}(s) \geqslant P_{i,k,T}(s),$$

which implies that

$$P(R_{i,s}R_{j,s} \mid Z = (5, u, v, w)) \geqslant P(R_{i,s}R_{k,s} \mid Z = (5, u, v, w)).$$

Hence, we have shown that

$$P(R_{i,s}R_{j,s} \mid Z) \geqslant P(R_{i,s}R_{k,s} \mid Z),$$

and the result follows on taking expectations of both sides of this inequality.    □

**Proof of Theorem 2.**  Assume that $v_j$ is decreasing in $j$. Now, given that $i$ reaches round $s$, the conditional probability that $j$ also reaches round $s$ is

$$P(R_{j,s} \mid R_{i,s}) = \frac{P(R_{j,s}R_{i,s})}{P(R_{i,s})}.$$

Hence, from Lemma 3, it follows that $P(R_{j,s} \mid R_{i,s})$ is a decreasing function of $j$, $j \neq i$. Now,

$$P_i(v_1, \ldots, v_n) = P(R_{i,2} \ldots R_{i,r+1}) = \prod_{s=1}^{r} P(R_{i,s+1} \mid R_{i,s}).$$

Let $C_{i,s}$ be the event that $i$ competes in round $s$ (that is, it is the event that $i$ reaches round $s$ and then plays a match in that round). With $Q_j = P(i$ plays $j$ in round $s \mid C_{i,s})$, $j \neq i$, we have that

$$\begin{aligned} Q_j &= P(i \text{ plays } j \text{ in round } s \mid C_{i,s}R_{j,s})P(R_{j,s} \mid C_{i,s}) \\ &= P(i \text{ plays } j \text{ in round } s \mid C_{i,s}R_{j,s})P(R_{j,s} \mid R_{i,s}) \\ &= \frac{1}{r_s - 1}P(R_{j,s} \mid R_{i,s}), \end{aligned}$$

where $r_s = n - \sum_{j=1}^{s-1} m_j$. Hence, $Q_j$, $j \neq i$ is a decreasing function of $j$. Letting $Y$ be a random variable such that $P(Y = j) = Q_j$, $j \neq i$, it thus follows that $Y$ is stochastically smaller than the random variable $X$ having $P(X = j) = 1/(n-1)$, $j \neq i$. Therefore,

$$\begin{aligned} P(R_{i,s+1} \mid R_{i,s}) &= 1 - \frac{2m_s}{r_s} + \frac{2m_s}{r_s}, P(R_{i,s+1} \mid C_{i,s}) \\ &= 1 - \frac{2m_s}{r_s} + \frac{2m_s}{r_s}\sum_{j \neq i}\frac{v_i}{v_i + v_j}Q_j \\ &= 1 - \frac{2m_s}{r_s} + \frac{2m_s}{r_s}\mathrm{E}\left[\frac{v_i}{v_i + v_Y}\right] \\ &\leqslant 1 - \frac{2m_s}{r_s} + \frac{2m_s}{r_s}\mathrm{E}\left[\frac{v_i}{v_i + v_X}\right] \\ &= 1 - \frac{2m_s}{r_s} + \frac{2m_s}{r_s}\frac{1}{n-1}\sum_{j \neq i}\frac{v_i}{v_i + v_j}, \end{aligned}$$

where the preceding used that $v_i/(v_i + v_j)$ is an increasing function of $j$ and that $X$ is stochastically larger than $Y$ to conclude that $\mathrm{E}[v_i/(v_i + v_Y)] \leqslant \mathrm{E}[v_i/(v_i + v_X)]$.    □

**Remark 2.**  Remark 2. It follows from Lemma 2 that for all random formats, $P_i = P(R_{i,r+1})$ is an increasing function of $v_i$. On the other hand, it seems intuitive that $P_i$ is a decreasing function of $v_j$ for $j \neq i$. However, while this is true for $n = 3$, it is not true for $n \geqslant 4$. The argument, when $n = 3$, uses that

$$\begin{aligned} P_1 = \frac{1}{3}&\left(\frac{v_2}{v_2 + v_3}\frac{v_1}{v_1 + v_2} + \frac{v_3}{v_2 + v_3}\frac{v_1}{v_1 + v_3}\right) \\ &+ \frac{2}{3}\frac{v_1}{v_1 + v_2}\frac{v_1}{v_1 + v_3}. \end{aligned}$$

This gives

$$\begin{aligned} \frac{\partial P_1}{\partial v_2} &= \frac{1}{3}v_1\frac{\partial}{\partial v_2}\frac{v_2}{(v_2 + v_3)(v_1 + v_2)} + \frac{1}{3}\frac{v_1v_3}{v_1 + v_3}\frac{\partial}{\partial v_2}\frac{1}{v_2 + v_3} \\ &\quad + \frac{2}{3}\frac{v_1^2}{v_1 + v_3}\frac{\partial}{\partial v_2}\frac{1}{v_1 + v_2} \\ &= \frac{1}{3}\frac{v_1}{(v_1 + v_2)^2(v_2 + v_3)^2(v_1 + v_3)} \\ &\quad \cdot (-v_1v_3^2 - 2v_2^2v_3 - 3v_1v_2^2 - 6v_1v_2v_3) \\ &\leqslant 0. \end{aligned}$$

For $n = 4$, a counterexample can be constructed as follows. Consider the balanced format, and let $P_i(v_1, v_2, v_3, v_4)$ denote the probability that $i$ wins the tournament when player $j$ has value $v_j$, $j = 1, 2, 3, 4$. Conditioning on whether or not player 1 first plays against player 4, we have

$$\begin{aligned} P_1(2, 1, 1, x) = \frac{1}{3}&\left(\frac{2}{2 + x} \cdot \frac{2}{3}\right) + \frac{2}{3} \\ &\cdot \frac{2}{3}\left(\frac{1}{1 + x} \cdot \frac{2}{3} + \frac{x}{(1 + x)}\frac{2}{(2 + x)}\right) \end{aligned}$$

and thus

$$\begin{aligned} P_1\left(2, 1, 1, \frac{1}{100}\right) &= \frac{31{,}600}{60{,}903} \approx 0.518858\cdots < 0.518861\cdots \\ &\approx \frac{7{,}744}{14{,}925} = P_1\left(2, 1, 1, \frac{1}{99}\right).    □ \end{aligned}$$

Using Lemma 3, it is easy to show that if $v_i$ is decreasing in $i$, then so is $P(R_{i,s})$.

**Corollary 1.**  If $v_1 \geqslant \cdots \geqslant v_n$, then $P(R_{1,s}) \geqslant P(R_{2,s}) \geqslant \cdots \geqslant P(R_{n,s})$, $s = 1, \ldots, r + 1$.

**Proof.**  Suppose $i < j$. Because $r_s$ is the number of players that are still alive at the start of round $s$, it follows that given $i$ reaches round $s$, the expected number of others that also reach round $s$ is $r_s - 1$. Consequently,

$$r_s - 1 = \sum_{k \neq i}P(R_{k,s} \mid R_{i,s}).$$

Hence,

$$(r_s - 1)P(R_{i,s}) = \sum_{k \neq i,j} P(R_{i,s}R_{k,s}) + P(R_{i,s}R_{j,s}),$$

$$(r_s - 1)P(R_{j,s}) = \sum_{k \neq j,i} P(R_{j,s}R_{k,s}) + P(R_{j,s}R_{i,s}),$$

which by Lemma 3 shows the desired result for $s \leqslant r$. In addition, we have

$$P(R_{i,r+1}) = \sum_{k \neq j,i} P(R_{i,r}R_{k,r}) \frac{v_i}{v_i + v_k} + P(R_{j,r}R_{i,r}) \frac{v_i}{v_i + v_j},$$

$$P(R_{j,r+1}) = \sum_{k \neq j,i} P(R_{j,r}R_{k,r}) \frac{v_j}{v_j + v_k} + P(R_{i,r}R_{j,r}) \frac{v_j}{v_j + v_i},$$

which by Lemma 3, and by the assumption that $v_i \geqslant v_j$, completes the proof. $\square$

We now give lower bounds on the win probabilities.

**Theorem 3.** *Suppose* $v_1 \geqslant v_2 \geqslant \cdots \geqslant v_n$. *Let* $X_1, \ldots, X_r$ *be independent with* $P(X_j = 1) = 2m_j/r_j = 1 - P(X_j = 0)$, $j = 1, \ldots, r$, *and let* $N = \sum_{j=1}^{r} X_j$. *Then*

$$P_i \geqslant \sum_{g=1}^{i-1} P(N = g) \prod_{k=1}^{g} \frac{v_i}{v_i + v_k}$$

$$+ \sum_{g=i}^{r} P(N = g) \prod_{k=1}^{i-1} \frac{v_i}{v_i + v_k} \prod_{k=i+1}^{g+1} \frac{v_i}{v_i + v_k}.$$

**Proof.** Define the surrogate of $i$ as follows: The initial surrogate of $i$ is $i$ itself; anyone who beats a surrogate of $i$ becomes the current surrogate of $i$. Note that at any time there is exactly one player who is the current surrogate of $i$. If we let $X_j$ be the indicator of whether the surrogate of $i$ plays a match in round $j$, it follows that $X_1, \ldots, X_r$ are independent with $P(X_j = 1) = 2m_j/r_j = 1 - P(X_j = 0)$. Also, let $N = \sum_{j=1}^{r} X_j$ be the number of games played by surrogates of $i$ (while they are the current surrogate), and let $O = \{J_1, \ldots J_N\}$ be their set of opponents in these games. Because $P(i \text{ wins the tournament } | N, O) = \prod_{k=1}^{N}(v_i/(v_i + v_{J_k}))$, it follows that

$$P_i = E\left[\prod_{k=1}^{N} \frac{v_i}{v_i + v_{J_k}}\right]$$

$$\geqslant E\left[\prod_{k=1}^{\min(N, i-1)} \frac{v_i}{v_i + v_k} \prod_{k=i+1}^{N+1} \frac{v_i}{v_i + v_k}\right],$$

where $\prod_{k=i+1}^{N+1}(v_i/(v_i + v_k))$ is equal to 1 when $N < i$. Hence,

$$P_i \geqslant \sum_{g=1}^{i-1} P(N = g) \prod_{k=1}^{g} \frac{v_i}{v_i + v_k}$$

$$+ \sum_{g=i}^{r} P(N = g) \prod_{k=1}^{i-1} \frac{v_i}{v_i + v_k} \prod_{k=i+1}^{g+1} \frac{v_i}{v_i + v_k}. \quad \square$$

**Remark 3.** Remark 3. The probability mass function of $N$, which does not depend on $i$, is easily obtained by solving recursive equations (see example 3.24 of Ross 2014). In the special case where $n = 2^s + k$, $0 \leqslant k < 2^s$, and where there are $k$ matches in round 1 and afterwards all remaining players have matches in each subsequent round, we have

$$P(N = s + 1) = \frac{2k}{n} = 1 - P(N = s).$$

**Remark 4.** Remark 4. A weaker bound than the one provided in Theorem 3 is obtained by noting that, for $i > 1$, $v_i/(v_i + v_{J_k}) \geqslant v_i/(v_i + v_1)$, and so for $i > 1$,

$$P_i \geqslant E\left[\left(\frac{v_i}{v_i + v_1}\right)^N\right]$$

$$= E\left[\left(\frac{v_i}{v_i + v_1}\right)^{X_1 + \cdots X_r}\right]$$

$$= \prod_{j=1}^{r} E\left[\left(\frac{v_i}{v_i + v_1}\right)^{X_j}\right]$$

$$= \prod_{j=1}^{r} \left(\frac{2m_j}{r_j} \frac{v_i}{v_i + v_1} + 1 - \frac{2m_j}{r_j}.\right)$$

## 5. Special Case Best and Worst Formats for the Strongest and Weakest Players

**Theorem 4.** *Suppose* $v_1 > v = v_2 = v_3 = \cdots = v_n$. *If* $n = 2^s + k$, $0 \leqslant k < 2^s$, *then the balanced format that schedules* $k$ *matches in round* 1 *and then has all remaining players competing in each subsequent round leads to the maximal possible value of* $P_1$.

**Proof.** Let $p = v_1/(v_1 + v) > 0.5$ be the probability that player 1 wins in a match against another player. The proof is by induction. As there is nothing to prove when $n = 2$, let $n = 2^s + k$, $1 \leqslant k \leqslant 2^s$, and suppose that the result is true for all smaller values of $n$. Consider a format that calls for $j$ matches in the first round, where $j < k$. Then, by the induction hypothesis, the format of this type that is best for player 1 will call for $k - j$ matches in the second round. The probability that player 1 is among the final $2^s$ players under these conditions is

$$f_1(p) \equiv \left[\frac{2j}{n}p + 1 - \frac{2j}{n}\right]\left[\frac{2(k-j)}{n-j}p + 1 - \frac{2(k-j)}{n-j}\right].$$

However, the probability that player 1 is among the final $2^s$ players if the first round has $k$ matches is

$$f_2(p) \equiv \frac{2k}{n}p + 1 - \frac{2k}{n}.$$

Now, $f_1(p) - f_2(p) = 0$ when $p$ is either $1/2$ or 1. Because $f_1(p) - f_2(p)$ is easily seen to be strictly convex, this implies that $f_1(p) - f_2(p) < 0$ when $1/2 < p < 1$. Thus, in

searching for the best format for player 1, we need not consider any format that calls for less than $k$ matches in round 1.

Now consider any format that calls for $k + i$ matches in round 1 where $i > 0$. By the induction hypothesis, it will call for $2^{s-1} - i$ matches in round 2. This will result in player 1 being among the final $2^{s-1}$ players with probability

$$g_1(p) \equiv \left[\frac{2k+2i}{n}p + 1 - \frac{2k+2i}{n}\right]\left[\frac{2^s - 2i}{2^s - i}p + 1 - \frac{2^s - 2i}{2^s - i}\right].$$

However, the format that calls for $k$ matches in round 1 and $2^{s-1}$ matches in round 2 leads to player 1 being among the final $2^{s-1}$ players with probability

$$g_2(p) \equiv \left(\frac{2k}{n}p + 1 - \frac{2k}{n}\right)p.$$

Now, $g_1(p) - g_2(p) = 0$ when $p$ is either 1/2 or 1. Because

$$g_1''(p) = 2\frac{2k+2i}{n}\frac{2^s - 2i}{2^s - i}, \quad g_2''(p) = 2\frac{2k}{n},$$

it follows that $g_1''(p) \geqslant g_2''(p)$ is equivalent to

$$(k + i)(2^s - 2i) \geqslant k(2^s - i),$$

which is easily seen to be equivalent to

$$2^s \geqslant k + 2i,$$

which holds because having $k + i$ two-person matches implies that $2^s + k \geqslant 2(i + k)$. Hence, $g_1(p) - g_2(p)$ is convex, which shows that there is an optimal format that initially has $k$ matches in round 1. The induction hypothesis then proves the result. □

**Remark 5.** Remark 5. Because $g_1(p) - g_2(p)$ is strictly convex unless $2^s + k = 2(i + k)$, it follows that the format specified in Theorem 4 is uniquely optimal (in the sense of maximizing $P_1$) when $n$ is odd, whereas when $n$ is even, there is also an optimal format that schedules $n/2$ matches in round 1.

**Remark 6.** Remark 6. In the case of general $v_i$, the format that calls for $n/2$ matches when an even number $n$ of players remain is not optimal for the best player. For a counterexample, consider a knockout tournament with six players having values $v_1 = 6$, $v_2 = 4$, $v_3 = 3$, $v_4 = v_5 = v_6 = 1$. Under the balanced format, the probability of player 1 winning the tournament is $P_1 = 0.4422$. Under a format that plays three games in round 1 and then one game in each round, the probability of player 1 winning the tournament is $P_1' = 0.4412 < P_1$.

**Theorem 5.** *Suppose $v_1 > v = v_2 = v_3 = \cdots = v_n$. The unique format that minimizes $P_1$ is the one that has exactly one match in each round.*

**Proof.** The proof is by induction. Suppose it is true for all tournaments with fewer than $n$ players, and now suppose there are $n$ players. Consider any format that calls for $s$ matches in the first round, where $s > 1$. The probability that player 1 is still alive when there are only $n - s$ alive players is

$$f_1(p) = \frac{2s}{n}p + 1 - \frac{2s}{n}.$$

On the other hand, the probability that player 1 is still alive when there are only $n - s$ alive players when the format is one match per round is

$$f_2(p) = \prod_{j=0}^{s-1}\left(\frac{2}{n-j}p + 1 - \frac{2}{n-j}\right).$$

Because $f_2(p)$ is a polynomial whose coefficients are all positive, it follows that $f_2''(p) > 0$. As $f_1''(p) = 0$, this implies that $f_1(p) - f_2(p)$ is strictly concave, which, since $f_1(p) - f_2(p) = 0$ when $p = 1/2$ or $p = 1$, enables us to conclude that $f_1(p) - f_2(p) > 0$ when $1/2 < p < 1$. Hence, by the induction hypothesis, the format that calls for one match in each round results in a win probability for player 1 that is strictly smaller than what it is under any format that calls for $s > 1$ matches in round 1. □

The following theorem gives the analogous results for the weakest player. The proofs are similar and thus omitted.

**Theorem 6.** *Suppose $v_1 = v_2 = v_3 = \cdots = v_{n-1} > v_n$. The format that results in the highest win probability for player $n$ is the one that has one match in each round. If $n = 2^s + k$, $1 \leqslant k \leqslant 2^s$, then the format that minimizes $P_n$ is the one that schedules $k$ matches in round 1 and then has all remaining players competing in each subsequent round.*

We now use the preceding results to obtain universal (that is, they hold for all formats) upper bounds on $P_i$. Recall that $p_i = (1/(n-1))\sum_{j\neq i} v_i/(v_i + v_j)$.

**Lemma 4.** *Suppose $n = 2^s + k$, $0 \leqslant k < 2^s$.*
*(i) If $p > \frac{1}{2}$, then $\prod_{j=1}^{r}((2m_j/r_j)p + 1 - 2m_j/r_j)$ is maximized by the balanced format, and its value is $((2k/n)p + 1 - 2k/n)p^s$.*
*(ii) If $p = \frac{1}{2}$, then $\prod_{j=1}^{r}((2m_j/r_j)p + 1 - 2m_j/r_j) = 1/n$ for all eligible formats.*
*(iii) If $p < \frac{1}{2}$, $\prod_{j=1}^{r}((2m_j/r_j)p + 1 - 2m_j/r_j)$ is maximized by the format where at each round there is exactly one match, and its value is $\prod_{j=1}^{n-1}((2/(n-j+1))p + 1 - 2/(n-j+1))$.*

**Proof.** Note that for a tournament where the players have respective values $(v_1, v \ldots, v)$, $P_1 = \prod_{j=1}^{r}((2m_j/r_j) \cdot p_1 + 1 - 2m_j/r_j)$. Now, (i) and (iii) are direct corollaries of Theorems 2 and 4, respectively, while (ii) is true because if $v_1 = v$, then all players have equal probability of winning the tournament regardless of the format. □

**Theorem 7.** *Suppose $n = 2^s + k$, $0 \leqslant k < 2^s$. Then*

$$
P_i \leqslant \begin{cases}
\left( \dfrac{2k}{n} p_i + 1 - \dfrac{2k}{n} \right) p_i^s & \text{if } p_i > \dfrac{1}{2}, \\[2ex]
\dfrac{1}{n} & \text{if } p_i = \dfrac{1}{2}, \\[2ex]
\displaystyle\prod_{j=1}^{n-1} \left( \dfrac{2}{n-j+1} p_i + 1 - \dfrac{2}{n-j+1} \right) & \text{if } p_i < \dfrac{1}{2}.
\end{cases}
$$

**Proof.** The proof follows directly from Theorem 2 and the preceding lemma. ☐

**Corollary 2.** *For a player with $p_i < \frac{1}{2}$, the probability of winning the tournament under any format is less than $1/n$.*

### References

Chung FRK, Hwang FK (1978) Do stronger players win more knockout tournaments? *J. Amer. Statist. Assoc.* 73(363):593–596.

David HA (1959) Tournaments and paired comparisons. *Biometrika* 46(1/2):139–149.

Glenn WA (1960) A comparison of the effectiveness of tournaments. *Biometrika* 47(3/4):253–262.

Hennessy J, Glickman M (2016) Bayesian optimal design of fixed knockout tournament brackets. *J. Quant. Anal. Sports* 12(1):1–15.

Horen J, Riezman R (1985) Comparing draws for single elimination tournaments. *Oper. Res.* 33(2):249–262.

Hwang FK (1982) New concepts in seeding knockout tournaments. *Amer. Math. Monthly* 89(4):235–239.

Jacobson SH, King DM (2009) Seeding in the NCAA men's basketball tournament: When is a higher seed better? *J. Gambling Bus. Econom.* 3(2):63–87.

Khatibi A, King DM, Jacobson SH (2015) Modeling the winning seed distribution of the NCAA Division I men's basketball tournament. *Omega* 50(January):141–148.

Marchand É (2002) On the comparison between standard and random knockout tournaments. *J. Roy. Statist. Soc. Ser. D* 51(2):169–178.

Maurer W (1975) On most effective tournament plans with fewer games than competitors. *Ann. Statist.* 7(7):717–727.

Ross SM (2014) *Introduction to Probability Models*, 11th ed. (Academic Press, Oxford, UK).

Ross SM, Ghamami S (2008) Efficient simulation of a random knockout tournament. *J. Indust. Systems Engrg.* 2(2):88–96.

**Ilan Adler** is a professor in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley. His research interests are in optimization theory, financial engineering, and combinatorial probability models.

**Yang Cao** is a PhD student in the Department of Industrial and Systems Engineering at the University of Southern California. He received his MS in industrial and systems engineering from the University of Southern California and his B.S. in theoretical and applied mechanics from Peking University.

**Richard Karp** has been a professor at the University of California, Berkeley, from 1968 to 1994 and 1999 to the present. From 1988 to 1995 and 1999 to the present he has been a research scientist at the International Computer Science Institute in Berkeley. His current activities center on algorithmic methods in genomics and computer networking. His honors and awards include the U.S. National Medal of Science, the Turing Award, the Fulkerson Prize, the Harvey Prize (Technion), the Centennial Medal (Harvard), the Lanchester Prize, the Von Neumann Theory Prize, the Von Neumann Lectureship, the Distinguished Teaching Award (Berkeley), the Faculty Research Lecturer (Berkeley), the Miller Research Professor (Berkeley), the Babbage Prize, and eight honorary degrees.

**Erol A. Peköz** is a professor in the Department of Operations and Technology Management at the Boston University Questrom School of Business. His research is focused in three areas: probability approximations and network science, healthcare quality and provider profiling, and queueing models and reliability for operations. He is the author of *The Manager's Guide to Statistics.*

**Sheldon M. Ross** is the Daniel J. Epstein Chair and Professor of Industrial Systems Engineering at the University of Southern California. He is the author of *Introduction of Mathematical Finance: Options and Other Topics, Simulation, A First Course in Probability, Probability Models for Computer Science.* He is an INFORMS Fellow.