STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. 11, March 2023, pp 332–344.
Published online in International Academic Press (www.IAPress.org)

Winning a Tournament According to Bradley-Terry Probability Model

Shuyang Gao* and Hosam Mahmoud

Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A

Abstract We analyze the chances of winning a tournament under the assumption that the probabilities of winning individual matches follow Bradley-Terry model [2]. We present an exact solution and show a few examples of its use. The examples are from California volleyball tournaments, the round of sixteen in the World Cup and the Champions League, the group stage of the Association of Tennis Professionals tournament, and the volleyball SuperLega in Italy.
The computational complexity of the solution grows exponentially fast with the number of teams and we seek approximations via multivariate Gaussian laws.

Keywords Bradley-Terry model, Multivariate Gaussian law, Combinatorial probability, Quantitative analysis of tournaments

AMS 2010 subject classifications 60C05, 60F05

DOI: 10.19139/soic-2310-5070-1490

1. The setup

This is a problem in the mathematics of sports that was graciously communicated to the authors by Professor Sheldon Ross. This manuscript follows up on [6].

The problem is concerned with the probability of a team (or a person in individual sports) winning a competition, when we have information about the values (strengths) and the schedule of matches for each team (or a person in individual sports). We cast the paper in the context of "teams," and its understood how to switch to the language of individual sports simply by replacing "team" with "person."

There are many tournament structures. We develop the combinatorial formulas for general tournaments with a predesignated schedule of matches, where the individual matches have a win-lose format (no draws). The most common form of this tournament structure is the round-robin, where every contestant plays every other contestant. Also common in tournaments is to play several rounds (typically two), with each round in the round-robin format. We focus the evaluation of the formulas by restricting to round-robin tournaments with multiple-rounds.

We analyze the chances of winning such a tournament under the assumption that the probabilities of winning individual matches follow Bradley-Terry model [2] (more on this below). In [6], Ross is concerned with structural issues in these tournaments; here we look into computational questions. The number of potential winners in a tournament that follows Bradley-Terry model has been studied in [3]. In [3], the team values are allowed to be random variables. Compared to the asymptotic studies in [3], our study is more on the applied perspective and focuses more on related real world examples. In a recent study, Bradley-Terry model is used to forecast tennis match results in [4]. Compared to the predictive model

*Correspondence to: Shuyang Gao (Email: gshuyang@gwmail.gwu.edu). Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

in [4], our study is not limited to tennis and provide some asymptotic analysis. To generalize the match format, a Bradley-Terry type model that allows multiple outcomes is considered in [7]. While the study in [7] focuses on the winning probability of an individual match, our scope is on finding overall probability of winning a tournament. Compared to our study, one advantage of the study in [7] is to allow for ties. This is a direction where we can further generalize our model.

There are r teams, numbered $1, \ldots, r$, participating in a tournament. The tournament has a fixed schedule specifying $N_{i,j}$, the number of times Team i plays Team j, for $1 \le i, j \le r$, with $i \ne j$. Then, of course, $N_{i,j} = N_{j,i}$ over the range of indices. The results of individual matches are independent. For example, for single round-robin tournaments, we have $N_{i,j} = 1$, for $1 \le i < j \le r$.

2. Real-world examples

Since the general combinatorial formulas we develop work for tournaments that are not round-robin, we show some round-robin examples and one example that is not round-robin.

In volleyball, matches are in the form win-lose, with no possible ties. Southern California Municipal Athletic Federation has published rules for breaking the ties in certain volleyball tournaments by having the top teams compete in a one-round tournament in the round-robin style. The teams that qualify for the final, typically $r = 3$ or 4, play a round-robin tournament, where each team meets every other team once ($N_{i,j} = 1$), for $i \ne j$.

The round of sixteen of the World Cup as well as the Champions League is made up of only win-lose matches. The teams are numbered, say $1, 2, \ldots, 16$, and listed such that Team $2i - 1$ is paired against Team $2i$, for $i = 1, 2, \ldots, 8$. According to the rules, this round has a knockout format. In this example, we have $N_{2i-1,2i} = 1$ for $i = 1, 2, \ldots, 8$, and otherwise $N_{i,j} = 0$.

The Association of Tennis Professionals (ATP) tournament is a competition among top ranked professional tennis players. The players are organized in two groups, and each group goes into a round-robin competition. The top two players in a group advance for a semifinal stage. In the semifinal, the top player in the first group is paired for a match against the runner-up of the second group, and vice versa. The losers of the semifinals are eliminated, and the top two players advance to the final. The group stage fits our setup with $r = 4$ (in each of the two groups) and $N_{i,j} = 1$, for $1 \le i < j \le 4$.

SuperLega is the highest level men's volleyball club competition in Italy. The tournament consists of two phases: a regular season round-robin competition, where each pair of teams plays twice and the standings are determined based on points gained (win-lose format), and a playoff tournament among the first eight teams in the standings. The regular season phase is an example for our setup, where $r = 12$ (there were $r = 13$ teams previously and the number is reduced to 12 in 2021) and $N_{i,j} = 2$, for $1 \le i < j \le 12$.

3. Bradley-Terry model

To evaluate the winning probabilities of the teams of a tournament, one needs the winning probabilities of the teams of each match. One model that fits this setup is the Bradley-Terry probability model.

In a Bradley-Terry model, according to training, skill, and quality of the players, each team has a value (strength). Let $v_i \in \mathbb{R}^+$ be the value of Team i, for $i = 1, \ldots, r$. Let the symbol $\rhd$ stand for "beats." For example, the terminology $i \rhd j$ means Team i beats Team j in a scheduled meeting in the tournament. In this model, the probability that Team i defeats Team j is

$$p_{i,j} := \mathbb{P}(i \rhd j) = \frac{v_i}{v_i + v_j}, \qquad 1 \le i, j \le r;$$

we take i and j to be distinct. Certainly, we have $p_{j,i} = 1 - p_{i,j}$. However, as we shall see, the use of $p_{j,i}$ "symmetrizes" the formulas and makes them compact.

For instance, in the case of the Champions League, we can use the published "market values" for the teams as measures of strength. In 2020, the top three values are 1.18 (Manchester City), 1.13 (Liverpool) and 0.8816 (Chelsea); the unit of measurement is billions of dollars.

An interesting question is What is the probability that Team i wins the tournament? This question has two interpretations: Team i is the sole winner, or shares the top position with other teams (perhaps more than one). For example, in the round of sixteen of the Champions League, the tournament is structured for eight teams to share the top position of this round and advance to the next round. Such a question can be addressed via joint distributions. In this study, we show combinatorial expression for both the probability of being the sole winner, and the probability of being the winner and sharing it with other teams. Since the computation of these two probabilities are similar, for the exact case, we show the probability of being the sole winner, and for the asymptotics, we show the probability of being the winner and sharing it with other teams. In each case, the alternative probability can be computed similarly.

4. Exact joint distribution of wins

In this section, we develop the exact joint distribution of wins. Although we start with a combinatorial expression that works for tournaments that are not necessarily round-robin, as later needed in the asymptotic analysis, we confine the rest of the paper to tournaments of n rounds, with each round in the round-robin format, that is $N_{i,j} = n$, for all $i \neq j$. The quantity $\sum_{\substack{j=1 \\ j\neq i}}^{r} N_{i,j} = n\binom{r}{2} := Q$ is a fixed number. In this setup, the team who wins the most number of games wins the competition, with no advantage in the count of wins for a team by playing more games than other teams. Let W_i be the number of times Team i wins by the end of the tournament. We construct a formula for

$$\mathbb{P}(W_1 = w_1, \dots, W_r = w_r).$$

In vector notation, using T for the transpose, we write the vector $(W_1, \dots, W_r)^T$ as $\mathbf{W}$, and the vector $(w_1, \dots, w_r)^T$ of target values as $\mathbf{w}$. So, the latter probability is reduced to the compact form $\mathbb{P}(\mathbf{W} = \mathbf{w})$.

We need to address the question at a feasible vector $(w_1, \dots, w_r)^T$. The probabilities are 0 outside the range of feasibility. Such a feasible vector puts a restriction on each component, and on the overall structure of how they are inter-related. To avoid the exclusion of the indices in the sums (such as repeatedly writing $i \neq j$), most of the time we write the full range of the indices from 1 to r, with the interpretation that $N_{i,i} = 0$ and $p_{i,i} = 0$. We must have

$$0 \le w_i \le \sum_{j=1}^{r} N_{i,j} = n(r-1), \qquad \text{for } i = 1, \dots, r,$$

together with

$$w_1 + w_2 + \cdots + w_r = \sum_{1 \le i < j \le r} N_{i,j} = n\binom{r}{2} = Q,$$

where n is the number of times each pair of teams meet. We shall see other restrictions in the joint distribution formula.

We use the notation $\mathrm{Bin}(n, p)$ to denote a binomial random variable counting successes in n independent identically distributed trials each of success probability p.

Let $K_{i,j}$ be the (random) number of times Team i beats Team j. And so, we have

$$W_i = K_{i,1} + K_{i,2} + \cdots + K_{i,r},$$

with the interpretation $K_{i,i} = 0$. The random variable $K_{i,j}$ is distributed like $\mathrm{Bin}(N_{i,j}, p_{i,j})$, for $i \neq j$. Then, we have $K_{j,i} = N_{i,j} - K_{i,j}$. We use the wiggly notation $\mathcal{W}_i$ for the event that Team i wins the tournament (possibly sharing the top position with other teams), and $\mathcal{W}_i^*$ for the event that Team i is the sole winner. Note that the sum of the probabilities $\mathbb{P}(\mathcal{W}_i)$ may exceed 1, accounting for the ties at the top position.

The realization $K_{i,j} = k_{i,j}$ occurs with probability $p_{i,j}^{k_{i,j}} p_{j,i}^{k_{j,i}} \binom{N_{i,j}}{k_{i,j}}$. By independence, the joint probability follows:

$$\mathbb{P}(W_1 = w_1, \ldots, W_r = w_r) = \sum \prod_{1 \le i < j \le r} p_{i,j}^{k_{i,j}} p_{j,i}^{k_{j,i}} \binom{N_{i,j}}{k_{i,j}}, \tag{1}$$

where the sum is taken over all feasible realizations $k_{i,j}$'s of the $K_{i,j}$'s. That is, realizations satisfying the constraints

$$\sum_{s=1}^{r} k_{m,s} = \Big(\sum_{s=1}^{m-1} N_{s,m} - \sum_{s=1}^{m-1} k_{s,m} \Big) + \sum_{s=m+1}^{r} k_{m,s} = w_m, \qquad m = 1, \ldots, r,$$

and

$$0 \le k_{i,j} \le N_{i,j}, \qquad 1 \le i, j \le r, \quad i \ne j.$$

This is a general formula that works even when the $N_{i,j}$'s are not all equal.

From the joint distribution, we get the winning probabilities for Team i (possibly sharing the top position with other teams):

$$\mathbb{P}(\mathcal{W}_i) = \sum_{w_i = \max(\mathbf{w})} \mathbb{P}(W_1 = w_1, \ldots, W_r = w_r), \tag{2}$$

the sum is taken over every feasible $\mathbf{w} = (w_1, \ldots, w_k)$ in which w_i is a maximal component.

To find the probability $\mathcal{W}_i^*$, we compute (1), then proceed with

$$\mathbb{P}(\mathcal{W}_i^*) = \sum_{\substack{w_i > w_j \\ j \ne i}} \mathbb{P}(W_1 = w_1, \ldots, W_r = w_r). \tag{3}$$

This probability is equal to or smaller than then one given in Eq. (2). The formulas (1)–(3) are computationally demanding. However, they are amenable to hand or computer evaluation for a small number of teams, and a small number of matches played by each team. When r or $N_{i,j}$ is in the middle range or is large, we resort to asymptotic approximations and simulations.

4.1. Cases amenable to computation

As an example, take $r = 4$, and the four teams compete in a round-robin tournament of only one round ($N_{i,j} = 1$). In the exact formula, all $k_{i,j} \le N_{i,j} = 1$ are in $\{0, 1\}$. Hence, all the binomial coefficients are 1. In expanded form, the probability of Team 1 being the sole winner of the tournament is

$$\begin{aligned}
\mathbb{P}(\mathcal{W}_1^*) &= p_{1,2}p_{1,3}p_{1,4}p_{2,3}p_{2,4}p_{3,4} + p_{1,2}p_{1,3}p_{1,4}p_{3,2}p_{2,4}p_{3,4} + p_{1,2}p_{1,3}p_{1,4}p_{3,2}p_{4,2}p_{4,3} \\
&\quad + p_{1,2}p_{1,3}p_{1,4}p_{3,2}p_{4,2}p_{3,4} + p_{1,2}p_{1,3}p_{1,4}p_{2,3}p_{2,4}p_{4,3} + p_{1,2}p_{1,3}p_{1,4}p_{2,3}p_{4,2}p_{4,3} \\
&\quad + p_{1,2}p_{1,3}p_{1,4}p_{2,3}p_{3,4}p_{4,2} + p_{1,2}p_{1,3}p_{1,4}p_{3,2}p_{2,4}p_{4,3} \\
&= \frac{v_1^3 \big(v_2 v_3 (v_2 + v_3) + v_2 v_4 (v_2 + v_4) + v_3 v_4 (v_3 + v_4) + 2 v_2 v_3 v_4 \big)}{(v_1 + v_2)(v_1 + v_3)(v_1 + v_4)(v_2 + v_3)(v_2 + v_4)(v_3 + v_4)}.
\end{aligned}$$

Similar (symmetrical) formulas can be written for the probabilities of the events $\mathcal{W}_2^*$, $\mathcal{W}_3^*$ and $\mathcal{W}_4^*$.

This applies directly to the groups of the ATP tournament. The ATP keeps track of the players via "points" accrued throughout the playing history of a player. The point calculation is a combination of the number of wins in major tournaments and the total prize money earned. We use the player's points as his value. The 2021 ranking and points are published on the ATP web page. A summary of the ranks and points is in the following table:

rank	player	points
1	Novak Djokovic	12,030
2	Rafael Nadal	9,850
3	Daniil Medvedev	9,735
4	Dominic Thiem	9,125

Assuming these players appear in a group in the ATP competition, the probabilities associated with these values are

$$\begin{aligned} p_{1,2} &= 0.5498171846, & p_{2,1} &= 0.4501828154,\\ p_{1,3} &= 0.5527222605, & p_{3,1} &= 0.4472777395,\\ p_{1,4} &= 0.5686598913, & p_{4,1} &= 0.4313401087,\\ p_{2,3} &= 0.5029359203, & p_{3,2} &= 0.4970640797,\\ p_{2,4} &= 0.5191040843, & p_{4,2} &= 0.4808959157,\\ p_{3,4} &= 0.5161717922, & p_{4,3} &= 0.4838282078. \end{aligned}$$

The probability that Djokovic is at the top of the group having won more games than anyone else in the group is only 0.1728135784. The probability is small in view of the closeness of all the values of the player in the group.

Although we restrict the scope of this study to round-robin tournaments, the combinatorial expression works for a general tournament in which the number of matches that a pair of teams play is prespecified. An example with light computation is a knockout round of sixteen, as most $N_{i,j}$' s are 0. In the proposed notation in Section 2, each W_i is a simple Bernoulli random variable (coming in dependent pairs). For $i = 1, 2, \dots, 8$, we have $W_{2i-1} = \text{Bernoulli}(p_{2i-1,2i})$, and $W_{2i} = 1 - W_{2i-1} = \text{Bernoulli}(p_{2i,2i-1})$. Prediction of the bracket in the round of sixteen comes with a lot of excitement in the world. The formula (1) comes down to

$$\begin{aligned} &\mathbb{P}(W_1 = w_1, \dots, W_{16} = w_{16})\\ &\quad = \sum_{k_{1,2}=w_1} \sum_{k_{3,4}=w_3} \sum_{k_{5,6}=w_5} \sum_{k_{7,8}=w_7} \sum_{k_{9,10}=w_9} \sum_{k_{11,12}=w_{11}} \sum_{k_{13,14}=w_{13}} \sum_{k_{15,16}=w_{15}} p_{1,2}^{k_{1,2}} p_{2,1}^{k_{2,1}}\\ &\qquad \times p_{3,4}^{k_{3,4}} p_{4,3}^{k_{4,3}} p_{5,6}^{k_{5,6}} p_{6,5}^{k_{6,5}} p_{7,8}^{k_{7,8}} p_{8,7}^{k_{8,7}} p_{9,10}^{k_{9,10}} p_{10,9}^{k_{10,9}} p_{11,12}^{k_{11,12}} p_{12,11}^{k_{12,11}}\\ &\qquad \times p_{13,14}^{k_{13,14}} p_{14,13}^{k_{14,13}} p_{15,16}^{k_{15,16}} p_{16,15}^{k_{16,15}} \end{aligned}$$

(subject to the constraints $k_{1,2} + k_{2,1} = 1, \dots, k_{15,16} + k_{16,15} = 1$), which is not hard to compute, for any consistent set of w_i's all in $\{0, 1\}$. We have, for example,

$$\mathbb{P}(W_{2i-1,i} = 1, W_{2i} = 0, \quad i = 1, \dots, 8) = p_{1,2} p_{3,4} p_{5,6} p_{7,8} p_{9,10} p_{11,12} p_{13,14} p_{15,16}.$$

4.2. Complexity of the exact solution

The calculation of the exact probability requires too many operations. The multiplicand has a fixed number of multiplications. Let us take that as the unit of computing time. The product in (1) has $\binom{r}{2}$ terms. Then, we sum over the equations

$$\begin{aligned} w_1 &= k_{1,1} + k_{1,2} + k_{1,3} + \cdots + k_{1,r};\\ w_2 &= k_{2,1} + k_{2,2} + k_{2,3} + \cdots + k_{2,r};\\ &\vdots\\ w_r &= k_{r,1} + k_{r,2} + k_{r,3} + \cdots + k_{r,r}, \end{aligned}$$

subject to the constraints $0 \le k_{i,j} \le n$ and $k_{i,j} = n - k_{j,i}$. (Recall that $k_{i,i} = 0$, for $i = 1, \dots, r$.) The term $k_{i,j}$ can be $0, 1, \dots,$ or n. So, the pair $(k_{i,j}, k_{j,i})$ can be chosen in $N_{i,j} + 1$ ways, because the value $k_{i,j}$ determines $k_{j,i}$. Hence, the equations have a total of $\prod_{1 \le i < j \le r} (N_{i,j} + 1)$ solutions. To determine the probability in (1), one needs

$$\prod_{1 \le i < j \le r} (N_{i,j} + 1) = (n+1)^{\binom{r}{2}}$$

time units. For instance, if $r = 20$ and $N_{i,j} = 2$, for $i \ne j$, one needs

$$3^{\binom{20}{2}} = 4.498196225 \times 10^{90}$$

formidable time units.

4.3. Asymptotics for round-robin tournaments

For a round-robin tournament with n rounds, it is evident that for large r or large n, the exact formula may be intractable. In these cases, we resort to asymptotic approximations. As it is set up, $K_{i,j}$ is distributed like $\mathrm{Bin}(N_{i,j}, p_{i,j})$. We have

$$W_i = K_{i,1} + K_{i,2} + \dots + K_{i,r}, \tag{4}$$

and W_i is a convolution (sum of independent random variables). Here, we have $K_{i,j} = \mathrm{Bin}(n, p_{i,j})$. Note that $N_{i,i} = 0$, $p_{i,i} = 0$, and the corresponding binomial $K_{i,i} = \mathrm{Bin}(N_{i,i}, p_{i,i}) = \mathrm{Bin}(0, 0) \equiv 0$.

We denote an m-component multivariate normally distributed random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ by $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Customarily, the index m is dropped when it is one, in which case $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are written as scalars.

For large n, the binomial distribution of $\mathrm{Bin}(n, p)$ is well approximated by the normal distribution of $\mathcal{N}(np, np(1-p))$. Thus, for n in mid-range or is large, W_i can be well approximated by a sum of independent normal random variates. The sum of independent normal variates is itself normally distributed, with the means adding up as the collective mean, and the variances adding up as the collective variance. Namely, Eq. (4) can be approximated by normal distributions:

$$W_i \approx \mathcal{N}\Big(n \sum_{j=1}^{r} p_{i,j}, n \sum_{j=1}^{r} p_{i,j} p_{j,i}\Big),$$

when the parameter n is large enough to allow it. Note again that $N_{i,i} = 0$, $p_{i,i} = 0$, and $\mathcal{N}(N_{i,i}, p_{i,i}) \equiv 0$.

All the (dependent) wins W_i, for $i = 1, \dots, r$, are approximated by normal variates, and their joint distribution is a multivariate normal. However, such a distribution is improper, in view of the constraint

$$W_1 + W_2 + \dots + W_r = \sum_{1 \le i < j \le r} N_{i,j} = n \binom{r}{2}.$$

The rank of the covariance matrix is $r - 1$, and the matrix is not invertible.

However, any $r - 1$ components of $\mathbf{W}$ have a proper multivariate distribution, with an invertible covariance matrix. We consider the vector comprised of the first $r - 1$ components, which has an approximate multivariate distribution:

$$\begin{pmatrix} W_1 \\ \vdots \\ W_{r-1} \end{pmatrix} \approx \mathcal{N}_{r-1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\mu} = n \begin{pmatrix} \sum_{j=1}^{r} p_{1,j} \\ \sum_{j=1}^{r} p_{2,j} \\ \vdots \\ \sum_{j=1}^{r} p_{r-1,j} \end{pmatrix},$$

and $\mathbf{\Sigma}$ is an $(r-1)\times(r-1)$ covariance matrix. The diagonal elements of the covariance matrix are the variances $\mathbb{V}\mathrm{ar}[W_i] = \sum_{j=1}^{r} np_{i,j}p_{j,i}$, for $i = 1, \dots, r-1$. The off-diagonal elements are determined as follows:

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[W_i, W_j] &= \mathbb{C}\mathrm{ov}[K_{i,1} + K_{i,2} + \cdots + K_{i,r}, \quad K_{j,1} + K_{j,2} + \cdots + K_{j,r}] \\
&= \mathbb{C}\mathrm{ov}[K_{i,j}, K_{j,i}] + \sum_{\substack{1\le k,\ell\le r \\ (k,\ell)\neq(j,i)}} \mathbb{C}\mathrm{ov}[K_{i,k}, K_{j,\ell}] \\
&= \mathbb{C}\mathrm{ov}[K_{i,j}, N_{i,j} - K_{i,j}] + 0 \qquad \text{(by independence)} \\
&= -\mathbb{C}\mathrm{ov}[K_{i,j}, K_{i,j}] \\
&= -\mathbb{V}\mathrm{ar}[K_{i,j}] \\
&= -np_{i,j}p_{j,i}.
\end{aligned}$$

For concreteness, we write the entire covariance matrix:

$$\mathbf{\Sigma} = n \begin{pmatrix}
\sum_{j=1}^{r} p_{1,j}p_{j,1} & -p_{1,2}p_{2,1} & -p_{1,3}p_{3,1} & \cdots & -p_{1,r-1}p_{r-1,1} \\
-p_{2,1}p_{1,2} & \sum_{j=1}^{r} np_{2,j}p_{j,2} & -p_{2,3}p_{3,2} & \cdots & -p_{2,r-1}p_{r-1,2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-p_{r-1,1}p_{1,r-1} & -p_{r-1,2}p_{2,r-1} & -p_{r-1,3}p_{3,r-1} & \cdots & \sum_{j=1}^{r} p_{r-1,j}p_{j,r-1}
\end{pmatrix}.$$

A mutivariate normal vector $\mathbf{Y} = (Y_1, \dots, Y_{r-1})^T$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ has the joint density

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1,\dots,Y_r}(y_1, \dots, y_{r-1}) = \frac{e^{-(\mathbf{y}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}}{\sqrt{(2\pi)^{r-1}|\mathbf{\Sigma}|}}; \tag{5}$$

here $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$.

Now, we are in a position to compute the approximate probabilities of winning (the case of sole winner can be computed similarly by Eq. (3)):

$$\begin{aligned}
\mathbb{P}(\mathcal{W}_i) &= \mathbb{P}(W_1 \le W_i, W_2 \le W_i, \dots, W_{i-1} \le W_i, W_{i+1} \le W_i, \dots, W_r \le W_i) \\
&= \mathbb{P}\Big(W_1 \le W_i, W_2 \le W_i, \dots, W_{i-1} \le W_i, W_{i+1} \le W_i, \dots, \\
&\qquad\qquad Q - \sum_{k=1}^{r-1} W_k \le W_i\Big) \\
&\approx \underset{\mathcal{V}}{\int\int\cdots\int} f(w_1, \dots, w_{r-1})\, dw_1 \dots dw_{r-1};
\end{aligned}$$

the integration is taken over the volume $\mathcal{V} \in \mathbb{R}^{r-1}$ defined by the inequalities

$$\begin{aligned}
Q - \sum_{\substack{j=1\\ j\neq i}}^{r-k-1} w_j - (k+1)w_i \quad &\le \quad w_{r-k} \quad \le \quad w_i, \qquad k = 1, \dots, r-i-1; \\
Q - \sum_{\substack{j=1\\ j\neq i}}^{r-k-1} w_j - kw_i \quad &\le \quad w_{r-k} \quad \le \quad w_i, \qquad k = r-i+1, \dots, r-1.
\end{aligned}$$

To discern how the limits on the integrals come about, we construct these inequalities in a bottom-up bootstrapping sequence. Here, we have $Q := n\binom{r}{2}$ is the total number of matches in the tournament. For the event $\mathcal{W}_i$ to occur the following bounds should be satisfied:

$$\begin{aligned} w_1 &\le w_i;\\ w_2 &\le w_i;\\ &\vdots\\ w_{i-1} &\le w_i;\\ w_{i+1} &\le w_i;\\ &\vdots\\ w_r &\le w_i. \end{aligned}$$

The bottom inequality is

$$w_r = Q - (w_1 + w_2 + \cdots + w_{r-1}) \le w_i.$$

Rearranging, we get

$$Q - (w_1 + w_2 + \cdots + w_{i-1} + 2w_i + w_{i+1} + \cdots + w_{r-2}) \le w_{r-1} \le w_i.$$

The limits of the innermost integration are now determined. We then move up the chain of inequalities. Plugging this in the penultimate inequality, we get

$$Q - (w_1 + w_2 + \cdots + w_{i-1} + 3w_i + w_{i+1} + \cdots + w_{r-3}) \le w_{r-2} \le w_i.$$

Iterating this $k < r - i$ steps, we get

$$Q - (w_1 + w_2 + \cdots + w_{i-1} + (k+1)w_i + w_{i+1} + \cdots + w_{r-k-1}) \le w_{r-k} \le w_i.$$

We repeat this for $k = 1, \ldots, r - i - 1$. Going further steps back, we establish the pattern of similar inequalities—we only need to observe a shift in indexing starting at the inequality $w_{i-1} \le w_i$.

Furthermore, w_i cannot be the maximum, unless it is at least Q/r. So, the outer integral on w_i runs from Q/r to infinity. As an instance, take the case where r equals 4, and all $N_{i,j}$ are 30, we have

$$\mathbb{P}(\mathcal{W}_2) \approx \int_{w_2=45}^{\infty} \int_{w_1=180-3w_2}^{w_2} \int_{w_3=180-w_1-2w_2}^{w_2} f(w_1, w_2, w_3)\, dw_1\; dw_2\, dw_3.$$

4.4. Small r and large n

We have successfully tried the approach in Section 4.3 on approximations for small r and large n, $1 \le i, j \le r$. We provide a worked out example.

Take the case $r = 3$, all $N_{i,j} = n = 30$, for $1 \le i, j \le 3$, and team values

$$v_1 = 100, \qquad v_2 = 102, \qquad v_3 = 112.$$

In view of the constraint

$$W_1 + W_2 + W_3 = 90,$$

we are looking at the reduced vector $\binom{W_1}{W_2}$, with mean

$$\boldsymbol{\mu} = \begin{pmatrix} \frac{155250}{5353} \\ \\ \frac{318240}{10807} \end{pmatrix} = \begin{pmatrix} 29.00242854 \\ \\ 29.44758027 \end{pmatrix},$$

and covariance matrix

$$\begin{pmatrix} \frac{429109500}{28654609} & -\frac{76500}{10201} \\ -\frac{76500}{10201} & \frac{1749870180}{116791249} \end{pmatrix}.$$

Hence, we have the approximate bivariate distribution

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \mathcal{N}_2\left(\begin{pmatrix} \frac{155250}{5353} \\ \frac{318240}{10807} \end{pmatrix}, \begin{pmatrix} \frac{429109500}{28654609} & -\frac{76500}{10201} \\ -\frac{76500}{10201} & \frac{1749870180}{116791249} \end{pmatrix}\right).$$

The density of such a bivariate distribution is

$$\begin{aligned} f(x,y) = {} & \frac{572771\sqrt{612876894}}{367726136400\,\pi} \\ & \times \exp\Big(-\frac{1606335077}{36051582000}x^2 - \frac{32160241}{721031640}yx - \frac{3275249777}{73545227280}y^2 \\ & \qquad + \frac{93684867}{24034388}x + \frac{94128863}{24034388}y - \frac{2744477955}{24034388}\Big). \end{aligned}$$

We get

$$\mathbb{P}(\mathcal{W}_1) \approx \int_{w_1=30}^{\infty}\int_{w_2=90-2w_1}^{w_1} f(w_1,w_2)\,dw_1\,dw_2 \approx 0.2439864192.$$

Similarly, we get

$$\begin{aligned} \mathbb{P}(\mathcal{W}_2) &\approx 0.2767548809, \\ \mathbb{P}(\mathcal{W}_3) &\approx 0.47925869991. \end{aligned}$$

These approximations are quite reasonable in comparison with the exact probabilities, which we obtained by computing (1)–(2) using a computer. The exact probabilities are

$$\begin{aligned} \mathbb{P}(\mathcal{W}_1) &= 0.2714855085\ldots, \\ \mathbb{P}(\mathcal{W}_2) &= 0.3057791749\ldots, \\ \mathbb{P}(\mathcal{W}_3) &= 0.5132831108\ldots. \end{aligned}$$

Remark 4.1
The three probabilities $\mathbb{P}(\mathcal{W}_i)$, for $i = 1, 2, 3$, are not mutually exclusive. As such, their sum exceeds 1, since, for example, $\mathbb{P}(\mathcal{W}_1 \cap \mathcal{W}_2) > 0$. To obtain $\mathbb{P}(\mathcal{W}_3)$, we looked at the reduced vector $\binom{W_1}{W_3}$, and reran the computer program. The alternative would be a somewhat unpleasant inclusion-exclusion approach.

Remark 4.2
The results for integrations are obtained symbolically. If we use Monte Carlo integration with sample size 10^5, we get $\mathbb{P}(\mathcal{W}_1) \approx 0.2426$, $\mathbb{P}(\mathcal{W}_2) \approx 0.2763$ and $\mathbb{P}(\mathcal{W}_3) \approx 0.4784$. Since the difference between symbolic integration and numerical integration is small, we can use numerical integration when r is moderately large.

Remark 4.3
Since for a fixed r, all $N_{i,j}$, for $1 \le i, j \le r$, are the same and are equal to n. Consider now the scaled wins

$$\frac{W_i}{n} = \frac{K_{i,1}}{n} + \frac{K_{i,1}}{n} + \cdots + \frac{K_{i,1}}{n}.$$

Recall that $K_{i,j}$ are independent and binomially distributed. By the strong law of large numbers, we have $K_{i,j} \xrightarrow{a.s.} p_{i,j}$, for $i \neq j$ (recall that $K_{i,i} = 0$). So, we have

$$\frac{W_i}{n} \xrightarrow{a.s.} p_{i,1} + p_{i,2} + \cdots + p_{i,r} =: w_i^*.$$

Next, assume $w_i^* < w_j^*$ Then, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{W}_i) &= \mathbb{P}(W_1 \le W_i, W_2 \le W_i, \ldots, W_{i-1} \le W_i, W_{i+1} \le W_i, \ldots, W_r \le W_i) \\
&\le \mathbb{P}(W_j \le W_i) \\
&= \mathbb{P}\Big(\frac{W_j}{n} \le \frac{W_i}{n}\Big) \\
&\to \mathbb{P}(w_j^* \le w_i^*) \\
&= 0.
\end{aligned}$$

Unless w_i^* is the highest, the corresponding $\mathcal{W}_i$ has 0 limiting probability. In contrast, if w_i^* is one of the highest w^*'s we have

$$\begin{aligned}
\mathbb{P}(\mathcal{W}_i) &= \mathbb{P}\Big(\frac{W_1}{n} \le \frac{W_i}{n}, \frac{W_2}{n} \le \frac{W_i}{n}, \ldots, \frac{W_{i-1}}{n} \le \frac{W_i}{n}, \frac{W_{i+1}}{n} \le \frac{W_i}{n}, \ldots, \frac{W_r}{n} \le \frac{W_i}{n}\Big) \\
&\to \mathbb{P}\Big(w_1^* \le w_i^*, w_2^* \le w_i^*, \ldots, w_{i-1}^* \le w_i^*, w_{i+1}^* \le w_i^*, \ldots, w_r^* \le w_i^*) \\
&= 1.
\end{aligned}$$

Indeed, the strongest team (and there can be more than one), has a chance to shine over a very large number of matches and recover from losing streaks.

4.5. Large r and small n

The case in Subsection 4.4 (small r and large $N_{i,j}$, for all $1 \le i, j \le r$) is of theoretical interest but may not be of practical value. There are not many tournaments that have small r and large $N_{i,j}$. The world chess championship title is one of few competitions that used to follow this structure in the past, with $r = 2$ and $n = 24$.

More common is a case where a medium or large number of teams participate in a tournament in which each team faces every other team a few number of times.

In this case, W_i is a sum of a large number of binomial random variables, each on a small number of experiments. Namely, we have

$$W_i = \mathrm{Bin}(n, p_{i,1}) + \mathrm{Bin}(n, p_{i,2}) + \cdots + \mathrm{Bin}(n, p_{i,r});$$

all the binomial variables in this expression are independent. Under the right conditions, the central limit theorem ascertains that, for large r, the sum can be approximated by a normally distributed random variable:

$$W_i \approx \mathcal{N}\Big(n \sum_{j=1}^{r} p_{i,j}, n \sum_{j=1}^{r} p_{i,j} p_{j,i}\Big).$$

The central limit theorem here is in a general form (say Lindeberg's form (Theorem 27.2 in [1]) that deals with possibly non-identically distributed but independent variables. The "right conditions" here are

$$s_i^2(r) := n \sum_{j=1}^{r} p_{i,j} p_{j,i} \to \infty,$$

and for any $\varepsilon > 0$, we have

$$\frac{1}{s_i^2(r)} \sum_{j=1}^{r} \sum_{\left|\mathrm{Bin}(n,p_{i,j}) - np_{i,j}\right| > \varepsilon s_i(r)} \left|\mathrm{Bin}(n, p_{i,j}) - np_{i,j}\right|^2 \to 0.$$

Such conditions are satisfied, if $p_{i,j}$'s are "close." By close, we mean none of them is dominant. The closeness condition is satisfied, if

$$\frac{\max\limits_{1\leq j\leq r} np_{i,j}p_{j,i}}{\sum\limits_{1\leq j\leq r} np_{i,j}p_{j,i}} \to 0, \qquad \text{as } r\to\infty,$$

which is Feller's condition.

For instance, in a league with $N_{i,j}=2$, with equal team values, we have

$$s_i^2(r) := 2\sum_{\substack{j=1\\ j\neq r}}^{r} \frac{1}{2}\times\frac{1}{2} = \frac{1}{2}(r-1)\to\infty, \qquad \text{as } r\to\infty.$$

Lindeberg's condition is also satisfied, because for large r, the factor $s_i(r)$ is large, while

$$\left|\text{Bin}(2,p_{i,r}) - 2p_{i,r}\right| = \left|\text{Bin}\left(2,\frac{1}{2}\right) - 2\times\frac{1}{2}\right| \leq 2+1=3.$$

So, for large $N_{i,j}$, the sum is empty (equal to 0), and $s_i(r)\to 0$.

The asymptotic probability is to be computed by an r-fold multiple integral. This remains to be a formidable computational challenge.

Remark 4.4
This approach does not provide a good approximation for the first round in a ladder with $2r$ (large) teams. In this case, we have

$$s_{2i-1}^2(r) = p_{2i-1,2i}p_{2i,2i-1} < \frac{1}{4},$$

and $s_{2i-1}^2(r)$ does not diverge to infinity when we have $n=1$. Fortunately, the case is amenable to direct computation from the exact joint distribution in (1), as discussed in Section 2.

One possible application that fits this setup is the Italian SuperLega, which is the highest level championship in the Italian Male Volleyball League. As of the 2020–2021 season, the SuperLega has 12 competing teams, and each team plays twice against every other team. This is an example with $r=12$ and $n=2$. To calculate the winning probabilities for the 2020–2021 season according to Bradley-Terry model, we use the points gained in last season as the values:

rank	team	points
1	Lube	53
2	Modena	52
3	Perugia	51
4	Trentino	45
5	Milano	36
6	Porto Robur Costa	26
7	Kioene Padova	25
8	Verona	24
9	Volley Milano	23
10	Gas Sales Piacenza	21
11	Top Volley Latina	16
12	Volley Callipo	16

4.6. Simulation of SuperLega

As an alternative to a possibly tough integration, we can obtain results via simulation. We conducted a Monte Carlo simulation on the SuperLega of Italy, with the number of points scored in the 2020 season as values. The Monte Carlo simulation is implemented in the statistical programming langauge R with sample size $s = 10^5$. The multivariate normal samples are generated using the package "mvtnorm"[5]. The empirical probabilities are obtained from frequency counts. For example, to empirically compute $\mathbb{P}(\mathcal{W}_1)$, we generate samples from the 11-dimensional multivariate normal specified in Subsection 4.3, and count the number of times that the value of the first coordinate is the maximum.

The algorithmic steps to compute the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ from the values are straightforward implementations of the formulas presented in Subsection 4.3. These computations, together with the desired number of simulations, s, are fed as parameters into the a subroutine shown below in pseudocode. The built-in function rmvnorm receives as input the dimensions of a random vector and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and returns a multivariate normally distributed random vector, which we call W. After such vector is generated, we add 1 to a counting vector at each position corresponding to a winner.

The simulation algorithm assumes the existence of an external function argmax that takes in an integer and a vector of that dimension, and returns a vector Whoiswinning that flags (with 1's) the indeces that attain the maximal value in the input vector. The indeces corresponding to teams that did not win are flagged with 0. At the end of the computation, the frequencies of winning are recorded in counter. These frequencies are turned into probabilities stored in the vector $\mathbb{P}$, which is returned to the caller of the subroutine.

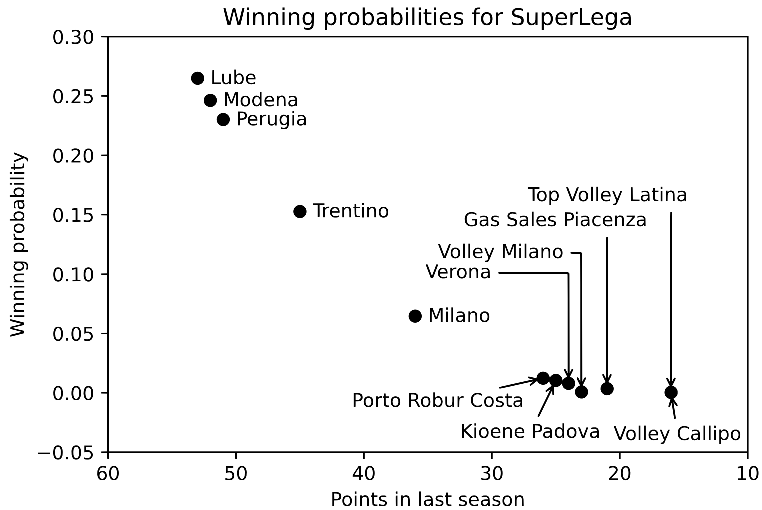
```
Input parameters: n, μ, Σ, s
Primitives function: rmvnorm
Local array: W[1 .. n − 1], counter[1 .. n − 1],
                 Whoiswinning[1 .. n − 1]
Local control variable: i, j

      counter ← 0
      for i = 1 .. s do
            W ← rmvnorm(n − 1, μ, Σ)
            call argmax(n − 1, W, Whoiswinning)
            for j ← 1 .. n − 1 do
                  if Whoiswinning [j] = 1 then
                        counter[j] ← counter[j] + 1
      for i = 1 .. n − 1 do
            ℙ[i] := counter[i]/s
      return ℙ
```

We present the results in the figure below.

Via this simulation the winning probabilities (of possibly multiple winners) are

$$
\begin{array}{lll}
\mathbb{P}(\mathcal{W}_1) \approx 0.26495, & \mathbb{P}(\mathcal{W}_2) \approx 0.24626, & \mathbb{P}(\mathcal{W}_3) \approx 0.23019, \\
\mathbb{P}(\mathcal{W}_4) \approx 0.15272, & \mathbb{P}(\mathcal{W}_5) \approx 0.06464, & \mathbb{P}(\mathcal{W}_6) \approx 0.01224, \\
\mathbb{P}(\mathcal{W}_7) \approx 0.01034, & \mathbb{P}(\mathcal{W}_8) \approx 0.00797, & \mathbb{P}(\mathcal{W}_9) \approx 0.00072, \\
\mathbb{P}(\mathcal{W}_{10}) \approx 0.00348, & \mathbb{P}(\mathcal{W}_{11}) \approx 0.00047, & \mathbb{P}(\mathcal{W}_{12}) \approx 0.00031.
\end{array}
$$

We see that the top three teams from the past season share high probabilities to be a winner in terms of points in this season, while the last six teams have a very slim chance to get to the top position. The accuracy of the results depends on the rate of convergence of approximation by normal distributions and the rate of convergence of the numerical integration. The results get more accurate when the number of



teams or the rounds in the tournament is relatively large. The error from the numerical integration is usually ignorable when the sample size taken is large.

Acknowledgment

The authors are indebted to Sheldon Ross for sharing a preprint on the subject and to Hossam Elgindy for numerous insightful discussions that enhanced our intuition and our overall understanding of the scope of the problem and the limitation of certain approaches. The authors also thank the anonymous referees for comments and suggestions.

REFERENCES

1. Billingsley, P. (2012). Probability and Measure, Anniversary Ed., Wiley.
2. Bradley, R. and Terry, M. (1952). Rank analysis of incomplete block designs: I. The Method of Paired Comparisons. Biometrika 39, 324–345.
3. Chetrite, R., Diel, R. and Lerasle, M. (2017). The number of potential winners in Bradley-Terry model in random environment. The Annals of Applied Probability, 27(3), 1372–1394.
4. Fayomi, A., Majeed, R., Algarni, A., Akhtar, S., Jamal, F. and Nasir, J. (2022). Forecasting Tennis Match Results Using the Bradley-Terry Model. International Journal of Photoenergy.
5. Genz, A., Bretz, F., Miwa T., Mi, X., Leisch F., Scheipl, F. and Hothorn, T. (2020). mvtnorm: Multivariate Normal and t Distributions. R package version 1.1–1.
6. Ross, S. (2022). Team's seasonal win probabilities. Probability in the Engineering and Informational Sciences, 36(4), 988–998.
7. Whelan, J. and Klein, J. (2021). Bradly-Terry modeling with multiple games outcomes with applications to college hockey. Mathematics for Applications, 10, 157–177.